

Predictive analytical model for chloride concentrations in the Port of Rotterdam

For analysing the effect of human interventions in the Rhine-Meuse Delta



HydroLogic BV
P.O. Box 2177
3800 CD Amersfoort
+31 33 4753535
hydrologic.com

MSc. Thesis
R.H. Linneman
September 2019

Cover image: New Waterway. <https://beeldbank.rws.nl>, Rijkswaterstaat / Joop van Houdt

Predictive analytical model for chloride concentrations in the Port of Rotterdam

For analysing the effect of human interventions in the Rhine-Meuse delta

by
R.H. Linneman

In partial fulfilment of the requirements for the degree of Master of Science in Water Engineering and Management at the University of Twente.

September 2019,
Amersfoort

Contact:
ralflinneman@hotmail.com

Graduation Committee:
Prof. Dr. Kathelijne M. Wijnberg
Dr. ir. Erik M. Horstman
Glenn M. Morvan MSc
Ir. Matthijs van den Brink

University of Twente
University of Twente
HydroLogic
HydroLogic

HydroLogic

UNIVERSITY OF TWENTE.

Preface

First, I would like to thank Glenn for all the time and effort that he put into supporting and guiding me during this research. Our numerous meetings on how to improve and expand this research made this thesis into something that I am very proud of. Also, I would like to thank Matthijs for keeping sight on the bigger picture of this complex subject. Whenever you joined the conversation, I was always better able to structure the thesis. Also, I would like to thank all my colleagues at HydroLogic for the very pleasant time during my stay.

I owe many thanks to Erik for his major support in structuring and focussing this thesis during our weekly Skype meetings and improving my writing skills with the use of the reviews he provided. I would also like to thank Kathelijne for her feedback which helped a lot in presenting the advantages of this research.

I hope you find this report informative and I hope that you enjoy reading it.

*Ralf Linneman
Amersfoort, September 2019*

Summary

Fresh water from estuaries is widely used, from drinking water production to agricultural use. The water quality standards for these various applications are regulated. One of these standards concerns the chloride concentration. Alterations in estuaries, such as deepening, may affect the chloride concentration inside the estuary. The Rhine-Meuse delta is such an estuary in which the fresh water is widely used for e.g. shipping but also for drinking water production and cooling. Therefore, predicting chloride concentrations in estuaries is important. Based on previously obtained measurement data, an analytical model is developed which provides insight in the importance and influence of boundary conditions on chloride concentrations.

Chloride concentrations within the estuary are affected by many processes, which can be summarized in three main factors; the inflow of salt water due to tides; the inflow of fresh water due to river discharge and the mixing processes between these inflows. Previous research indicated that deepening of the New Waterway and Botlek may lead to increased chloride concentrations in the Rhine-Meuse delta. In this research daily averaged values were used. Due to the dependence of the inflow of salt water on the tidal water movement, however, this analysis is best performed at the time scale of the in- and outflow of the tidal wave. The inflow of fresh water in the Rhine-Meuse delta originates from the Waal, Meuse and Lek rivers, of which the discharge volumes are measured upstream of the estuary. These discharges take a certain amount of time to reach the measurement locations for chloride concentrations in the estuary. Similarly, the inflow of salt water with the tidal wave, measured as the water level at the mouth of the estuary, takes time to propagate into the estuary and reach the chloride concentration measurement locations. These time lags are determined, with the use of a cross-correlation analysis between the observed boundary conditions and the chloride concentrations, at four different locations in the estuary. Resulting time lags vary from 110 minutes to 280 minutes regarding the tide and 750 minutes to 1900 minutes regarding the discharges of the Waal, Meuse and Lek.

Variations in chloride concentrations at all four examined measurement locations are best explained with a non-linear analytical model, including parameters that describe the autocorrelation of the input parameters with a moving weighted average. Performance of the developed predictive analytical model of Lekhaven on the training dataset was determined at a R^2 value of 0.87 and a RMSE value of 469.4 mg/L and on the validation dataset at a R^2 value of 0.80 and a RMSE value of 579.1 mg/L. Similar results were found for the three other measurement locations.

For the analysis of the effects of human interventions on chloride concentrations in the estuary of the Rhine-Meuse Delta, such as deepening of the New Waterway and Botlek, the developed analytical predictive models can be applied on post human-intervention gathered data. This analysis on measurement data can be used to validate results of theoretical models, and as indication on how relations between input parameters have changed due to human intervention in the Rhine-Meuse Delta. Furthermore, the developed prediction models can be used for predictions of chloride concentrations with the use of expected values for the discharge of the Rhine and the astronomical tide.

Contents

Preface	iii
Summary	iv
1 Introduction	4
1.1 Background		5
1.2 The Rhine-Meuse Delta		6
1.2.1 Salt intrusion dynamics		7
1.2.2 Chloride concentration measurement locations		11
1.3 Effects of deepening		12
1.4 Objective and research questions		13
1.5 Research approach and reading guide		14
2 Optimization of dataset	16
2.1 Available measurements		17
2.2 Analysis period		18
2.3 Correlation method		19
2.4 Time lag of boundary conditions		21
2.4.1 Water level		21
2.4.2 Discharge		22
2.5 Sampling interval		26
2.6 Summary		28
3 Chlorinity predictor model development	30
3.1 Regression model building and validation methodology		31
3.1.1 Training and validation datasets		32
3.1.2 Parameter normalization and non-linear weighting		32
3.1.3 Autocorrelation of boundary conditions		34
3.1.4 Parameter selection		35
3.1.5 Sensitivity analysis		37
3.1.6 Uncertainty analysis		37
3.2 Lekhaven prediction model		38
3.2.1 Parameter selection with multi-step analysis		38
3.2.2 Performance on training dataset		40
3.2.3 Performance on validation dataset		41
3.2.4 Sensitivity analysis		41
3.2.5 Uncertainty analysis		42
3.2.6 Model selection		44
3.3 Spijkenisserbrug prediction model		45

3.3.1	Parameter selection with multi-step analysis	45
3.3.2	Performance on training dataset	47
3.3.3	Performance on validation dataset	48
3.3.4	Sensitivity analysis	48
3.3.5	Model selection	49
3.4	Brienoordbrug and Beerenplaat prediction models	49
3.5	Summary	50
4	Model application	52
4.1	Methodology for analysis of residuals	54
4.2	Example on synthetic timeseries	56
5	Discussion	60
5.1	Gained insight on salt intrusion in the Port of Rotterdam	61
5.2	Impact of model assumptions	62
5.3	Identifying salinity changes in the Port of Rotterdam	64
5.4	Benefits and other applications of model	65
6	Conclusion and recommendations	66
6.1	Conclusion	66
6.1.1	Optimization of dataset	66
6.1.2	Predictive analytical model development	66
6.1.3	Future application of predictive analytical models	67
6.2	Recommendations	68
7	Bibliography	69
Annex A	Discharge Rhine and chloride concentrations	72
Annex B	Lekhaven - model parameter coefficients	74
Annex C	Spijkenisserbrug - model parameter coefficients	75
Annex D	Model development Brienoordbrug	76
Annex E	Model development Beerenplaat	79

1 Introduction



1.1 Background

The Port of Rotterdam is Europe's largest seaport and the New Waterway forms its connection to the North Sea. It is a shipping canal especially designed for sea-going vessels of which yearly over 15.000 pass along to reach the port (Port of Rotterdam, 2016). The New Waterway and adjoining ports are constantly dredged to maintain navigability. In order to handle ships with a draught of up to 15 meters, and to match with international standards, the New Waterway and adjacent Botlek were deepened. The New Waterway is located in the Rhine-Meuse Delta, forming the final terminal of these rivers before they discharge into the North Sea (Figure 1).

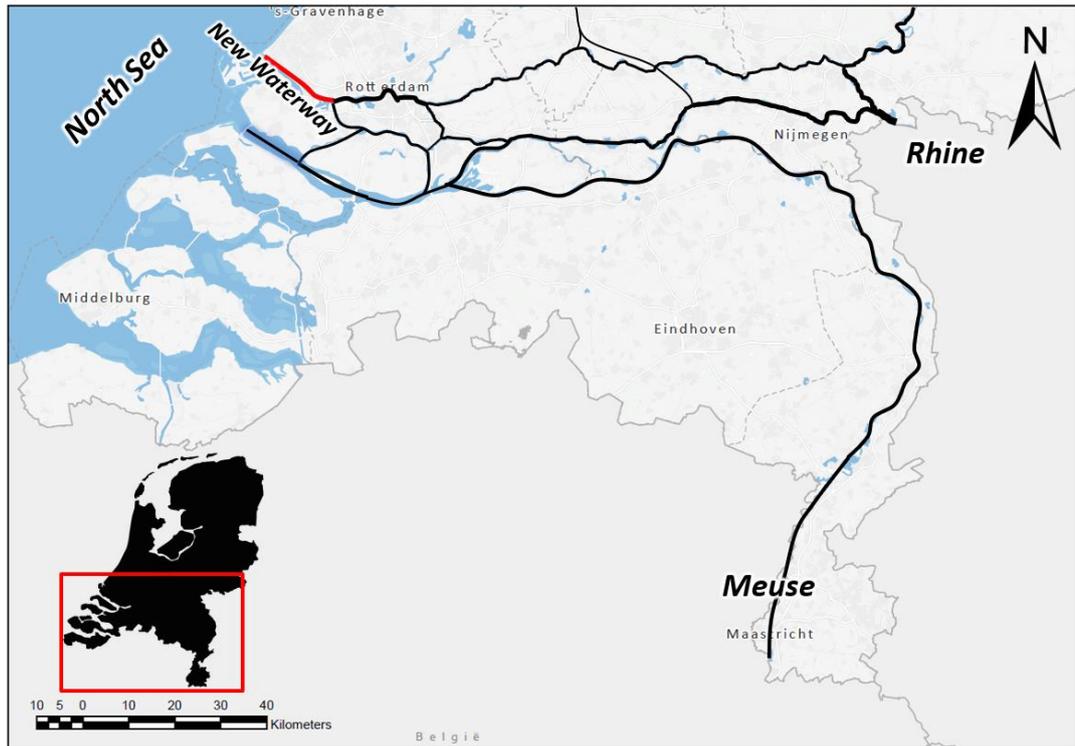


Figure 1. Rhine-Meuse Delta and New Waterway in the Netherlands.

A delta or estuary is the transition between a river and a sea (Nguyen, 2008). The salinity of the estuarine water is the result of two opposing fluxes: a saltwater flux, and a freshwater flux. The saltwater flux is driven by the tidal motion of the sea and the freshwater flux is driven by the river that discharges freshwater into the estuary (Savenije, 2012). Chloride concentrations in estuaries are the result of interaction between these two opposing fluxes. Savenije (2012) states both fluxes are strongly dependant on the estuary topography: “.. the salt water flux because the amount of water entering the estuary depends on the surface area of the estuary; and the fresh water flux, because the cross-sectional area of the estuary determines the efficiency of the fresh water flow to push back the salt”.

Alterations to the estuary, such as deepening, affect the estuaries topography, which in turn influences the interaction between the saltwater and freshwater fluxes.

Water in the estuary of the Rhine-Meuse Delta is widely used. Drinking water companies take in fresh water from these rivers for the production of drinking water. Water quality is regulated by law and for the production of drinking water, the maximum chloride concentration is 150 mg/l (Ministerie van Infrastructuur en Milieu, 2019).

Similarly, maximum chloride concentrations are determined for industrial and agricultural use and for areas marked as Natura 2000 areas (HydroLogic, 2015a).

Interventions are necessary to extend the navigability in the Port of Rotterdam to maintain the global economic position. However, these interventions may have a negative effect on the production of drinking water, agricultural and industrial use of fresh water and on natural habitats. Therefore, studying the effect of human interventions on chloride concentrations is important. For that reason, salinity concentrations have been monitored at several locations in the Rotterdam harbour area, since 2011. At each of these locations the chloride concentration is measured at various depths. Changes of chloride concentrations at these measurement locations could also be an indicator of changes further upstream in the delta. Complex processes in deltas can be approximated with the use of analytical models (van Rijn, 2011 and Xu, et al. 2017). This study focuses on developing an analytical model predicting chloride concentrations at these measurement locations in the Port of Rotterdam.

1.2 The Rhine-Meuse Delta

The study area is part of the complex Rhine-Meuse Delta, consisting of several bifurcations and convergences. Within the system, weirs and dams are constructed to control water levels as to facilitate shipping, but these constructions are also obstructing free flow of river water into the North Sea.

In the east of the study area, the Rhine enters the Netherlands at Lobith. It is then called the Upper Rhine. At Pannerdense Kop, the Upper Rhine bifurcates into the Waal and the Pannerden Canal (Figure 2). Further downstream, the Pannerden Canal bifurcates into the IJssel, which flows into the Lake IJssel, and the Lower Rhine. At Hagestein, water is let into the Amsterdam-Rhine Canal. After this bifurcation of the Lower Rhine the river is called the Lek. The southern part of the Rhine delta, the Waal, reaches the New Meuse through the Lower Merwede and Noord. Via the New Merwede it converges with the Meuse.

The New Meuse is fed by water from the Lek and Waal. The Old Meuse is fed by water from the Waal and the Meuse. Water from the Meuse and Waal (through the New Merwede) flows into the Old Meuse either through the Dordtsche Kil in the east or through the Spui in the west. The New Meuse and Old Meuse converge into the New Waterway which flows into the North Sea. Water from the Rhine-Meuse Delta may also be discharged through the Haringvliet sluices, located south of the New Waterway. In contrast to the discharge through the New Waterway, discharge through the Haringvliet sluices is controlled.

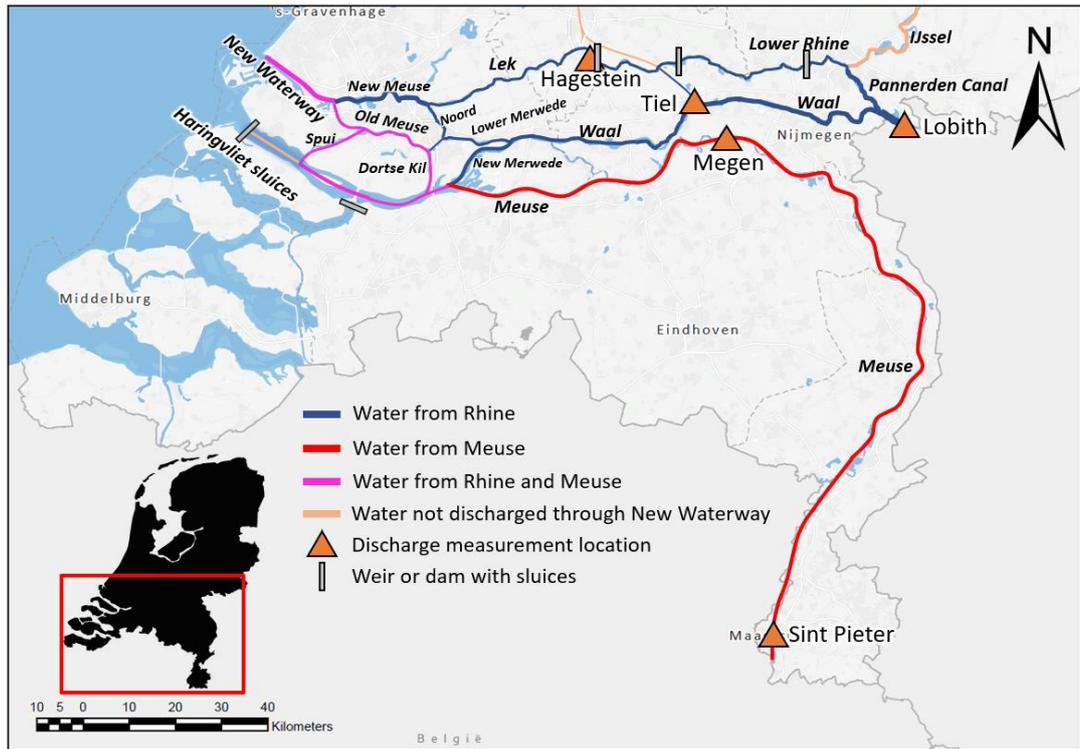


Figure 2. Overview of Meuse and Rhine water in the Netherlands, location of discharge measurements and locations of weirs or dams with sluices (Rijkswaterstaat, 2015).

1.2.1 Salt intrusion dynamics

Salt intrusion is a dynamic interaction between two opposing fluxes: a saltwater flux, and a freshwater flux. The saltwater flux, driven by tidal motion, and the freshwater flux, driven by the river discharge, are subjected to many influencing processes outside the estuary. Inside the estuary, the saltwater and freshwater fluxes meet under the influence of several mixing processes (Figure 3). The main factors influencing these fluxes and mixing processes are described below.

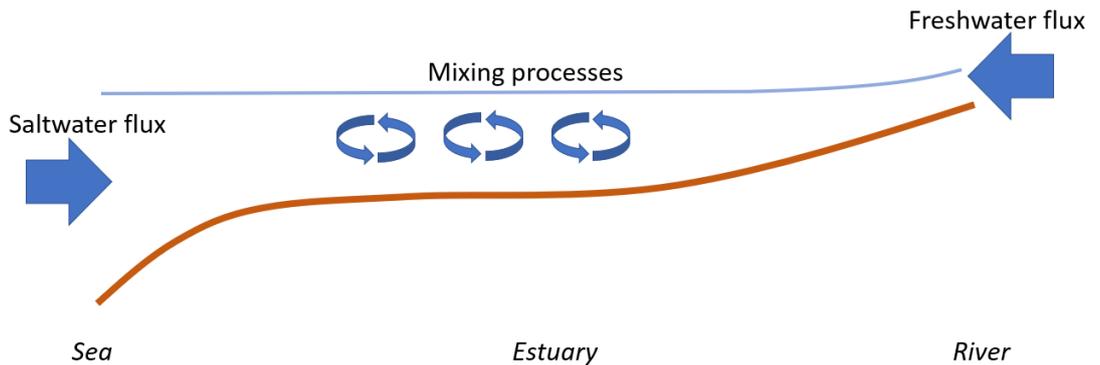


Figure 3. Schematics visualization of saltwater flux, freshwater flux and mixing processes in the transition from river to sea.

Saltwater flux

The saltwater flux at the mouth of the estuary has two characteristics, volume and salinity (Savenije, 2012). The volume of the flux varies, with constant geometry, with the water level.

The water level at the mouth of the estuary is the sum of the astronomical tide and the effect of wind.

Tide is caused by gravitational interactions in the planetary system and is the main driver of the saltwater flux. At the mouth of the Rhine-Meuse Delta, at Hoek van Holland, the tidal amplitude varies between 1.4 and 2 meters with a period of 12 hours and 25 minutes (Deltares, 2014).

The tidal amplitude varies due to the spring/neap cycle (Figure 4), with a period of approximately 14 days.

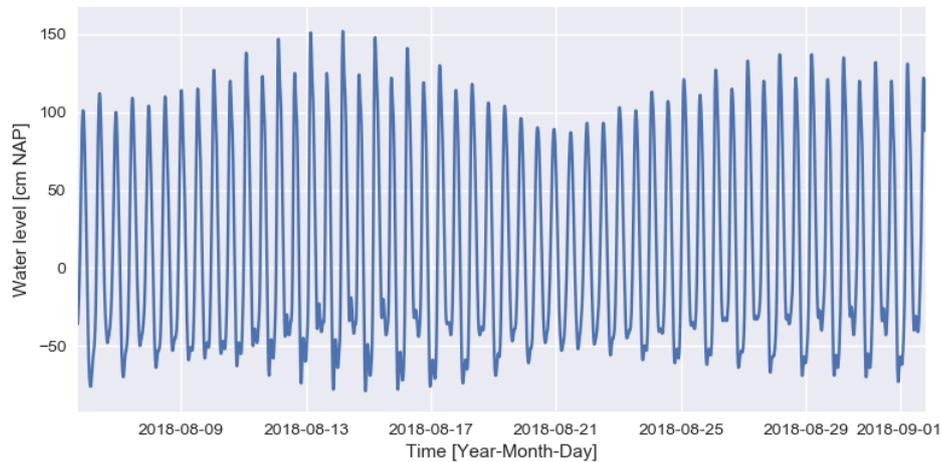


Figure 4. Astronomical tidal water level fluctuation at Hoek van Holland, showing the spring/neap cycle.

Wind on the surface of the North Sea may either increase or decrease water levels at the mouth of the estuary, generally referred to as wind setup or setdown, depending on the wind direction.

The salinity of the saltwater flux at the mouth of the estuary is dependent on the chloride concentration of the North Sea and the inflow of freshwater along the coast.

In- and outflow of tide

In- and outflow of the tidal wave is caused by the water level difference between the water level at the mouth of the estuary and the water level further upstream in the estuary (Deltares, 2016). The water level at the mouth of the estuary is influenced by the height of the astronomical tide and wind, as explained above. The water level upstream of the estuary is dependent on the discharge volume of the river, and thus the freshwater flux.

Freshwater flux

The freshwater flux is driven by the inflow of fresh river water, and similar to the saltwater flux, has two characteristics; volume and salinity. The volume is equal to the discharge. The most downstream discharge measurement locations in the Rhine-Meuse Delta are situated at Tiel, Hagestein and Megen, measuring the discharge of the Waal, Lek and Meuse, respectively (Figure 2).

The freshwater flux is also affected by precipitation and evapotranspiration directly at the water surface of the estuary and in the hinterland, downstream of the discharge measurement locations. During times of large precipitation events in the hinterland, water is being discharged via the regional water system into the Rhine-Meuse Delta, through

various pumping stations. During times of precipitation shortages, mainly during the growth season, water is withdrawn from the Rhine-Meuse Delta for, for example, agricultural usage. Especially during periods of relatively low discharge, this water withdrawal may affect the chloride concentrations in the delta significantly.

Due to the multi-channel layout of the Rhine-Meuse Delta, and lack of discharge measurement stations at every channel, exact discharge volumes via each branch are difficult to determine. Water in the system is discharged into the North Sea trough the New Waterway and, dependant on the open sluice area, through the Haringvliet sluices.

Discharge through Haringvliet sluices

Before deepening of the New Waterway and Botlek in 2018, the opening of the Haringvliet sluices occurred based on the LPH'84 policy. In this policy the sluice opening of the Haringvliet sluices is set based on the discharge of the Rhine at Lobith (Figure 6). During flood, the sluices are closed. During ebb, below a discharge of 1100 m³/s all sluices are closed, between 1100 m³/s and 1700 m³/s the total sluice opening is equal to 25m², with discharges of the Rhine at Lobith above 1700 m³/s the sluice opening increases with increasing discharge (Deltares, 2016).

The actual discharge through the Haringvliet sluices in not measured. However, model simulations have been performed within SOBEK from which a relation between Rhine discharge at Lobith and discharge through the sluices was composed (Figure 5) (Rijkswaterstaat, 2011).

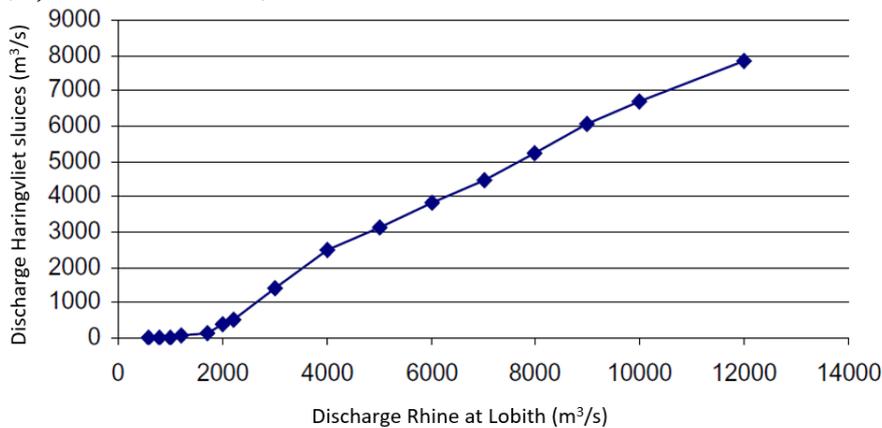


Figure 5. Relation between discharge of Rhine at Lobith and discharge through Haringvliet sluices corresponding with the LPH'84 policy (Rijkswaterstaat, 2011).

On the 15th of November 2018 the 'Kierbesluit' was set in motion. This meant the opening of the Haringvliet sluices during high tide (Figure 6), based on the discharge quantity of the Rhine measured at Lobith. This way, saline water from the North Sea can enter the Haringvliet and migratory fish can enter. During ebb, the opening of the sluices is increased as well, in order to discharge the saltwater that entered during high tide back into the North Sea (Deltares, 2017).

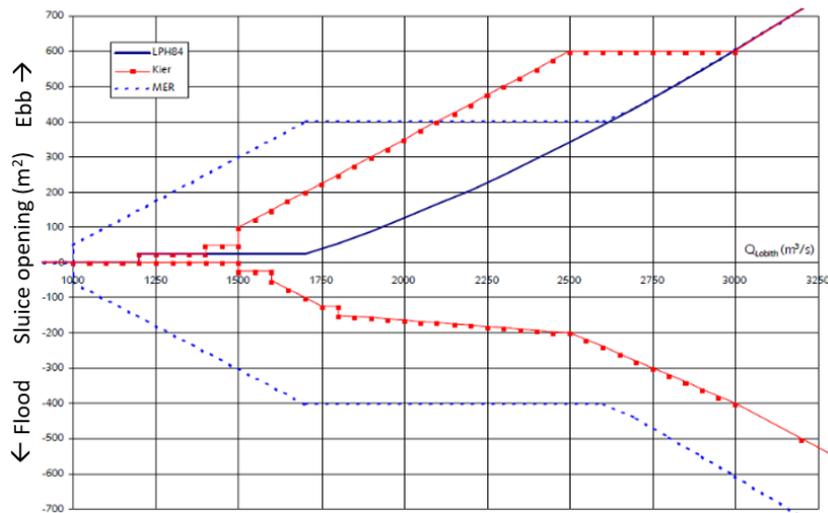


Figure 6. Opening of Haringvliet sluices based on Rhine discharge at Lobith before 'Kierbesluit' (LPH'84, solid blue) and after 'Kierbesluit' in 2018 (Kier, red) (Deltares, 2017).

The size of the sluice opening with the LPH'84 and Kier policies is very similar below a Rhine discharge of 1500 m³/s at Lobith. Below this limit, the Haringvliet sluices are almost completely closed in order to direct all freshwater discharge through the New Waterway and minimize salinization in the Port of Rotterdam. If a period of low discharge of the Rhine is expected, the Haringvliet is flushed with fresh water during several tidal periods to maintain a fresh Haringvliet as long as possible (Deltares, 2017).

Salinity of freshwater flux

The freshwater inflow also contains a certain amount of chloride, the background concentration. This chloride concentration is dependent on the volume of discharge (Kranenbrug, et al., 2015).

Mixing processes

“There is virtually no limit to the number of mixing processes that can be identified” stated Savenije (2012). However, three main factors were identified which cause mixing and dispersion in an estuary; tidal flow, river flow and wind stresses.

Mixing by tidal flow is probably the most important factor (Savenije, 2012) and is dependent on the salt flux. Mixing due to river flow is dependent on the freshwater flux. Mixing due to wind stresses have little influence compared to the other main factors (Savenije, 2012), and is therefore neglected in this research.

The saltwater flux, freshwater flux and the mixing processes lead to a certain vertical and horizontal distribution of chloride concentrations in the estuary (Figure 7, left panels, blue lines are isohalines), often referred to as the salt wedge (Savenije, 2012). With increasing river discharge or decreasing tidal range, the vertical salinity gradient increases.

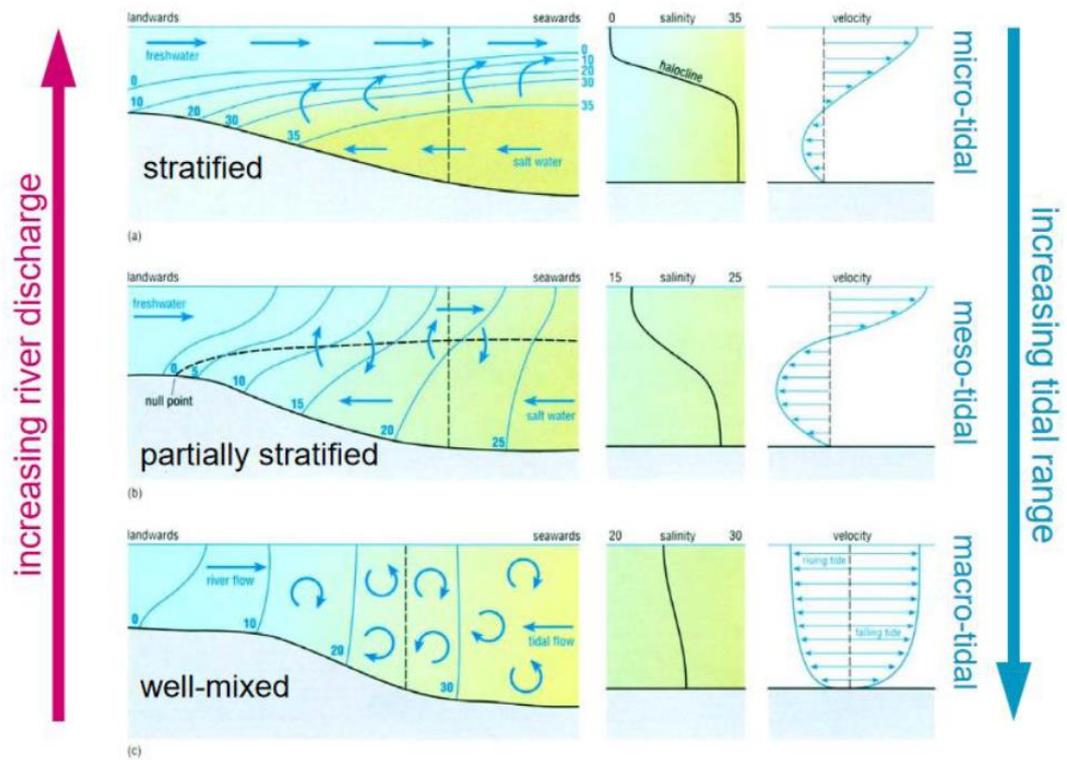


Figure 7. Schematic representation of water circulation, salinity distribution and velocity gradients in the estuary from stratified (top), through partially stratified or mixed (centre), to well-mixed (bottom) under increasing river discharge and increasing tidal range. The broken horizontal lines in the left panels indicate the positions of the salinity distributions and the velocity profiles (adapted from: Open University. Oceanography Course Team, 1999).

1.2.2 Chloride concentration measurement locations

In order to monitor salt intrusion in the Rhine-Meuse Delta, chloride concentrations are being monitored at several locations within the system. Measurement locations Lekhaven and Brienoordbrug (Figure 8), are situated in the New Meuse on the north side of the study area. At the southern part of the Port of Rotterdam, in the Old Meuse, measurement locations Spijkenisserbrug and Beerenplaat are situated. Under normal river discharges of the Rhine and Meuse, daily fluctuations in chloride concentrations are measured at Lekhaven and Spijkenisserbrug (Annex A). During dry periods, with decreased river discharges, increased chloride concentrations are measured at Brienoordbrug and Beerenplaat.



Figure 8. Locations of chloride measurement stations in the New Meuse and Old Meuse.

At Spijkenisserbrug and Lekhaven chloride concentrations are measured at three different depths. At Brienoordbrug, chloride concentrations are measured at two depths and at Beerenplaat at one depth (Table 1).

Table 1. Measuring depths of chloride concentrations at each of the four measurement locations.

Measurement location	Measuring depth [m NAP]
Lekhaven	-2.5, -5.0, -7.0
Spijkenisserbrug	-2.5, -4.5, -9.0
Brienoordbrug	-2.5, -6.5
Beerenplaat	-2.0

1.3 Effects of deepening

No research has been performed on the effects of deepening on chloride concentrations at a specific location in an estuary. The intrusion length of the salt wedge, which can be used as indication of chloride concentrations at a specific location, has been widely examined with the use of analytical models (van den Burgh, 1972; Savenije, 1993; Nguyen, 2008). Cai et al. (2012) derived a tidally averaged analytical model based on Savenije et al. (2008) for the effects of river discharge and channel deepening on the tidal amplitude and tidal wave travel time in the riverine Modaomen Estuary in China. It proved to be efficient and effective. With the use of this model, effects of dredging were calculated under constant discharges. Deepening of an estuary by dredging, increased the tidal wave propagation which in turn lead to increased chloride concentrations, and decreased the tidal wave travel time.

Alterations to the Rhine-Meuse Delta, such as deepening of the New Waterway, change the morphological characteristics of the delta, potentially causing changes in the intrusion length of the salt flux. Increased chloride concentrations at the measurement locations in the Old Meuse and New Meuse may be an indicator of changes in chloride concentrations further upstream where fresh river water is used for drinking water production, agricultural and industrial processes and Natura 2000 areas are present.

During relatively dry periods, in which the discharge of the Rhine at Lobith is below 1500 m³/s, discharge distributions in the Rhine-Meuse Delta are assumed to be stable due to the closing of the Haringvliet sluices. During these dry periods, at high tide saltwater intrudes up to all four chloride concentration measurements stations. Changes in morphology due to deepening of the New Waterway and the effects on chloride concentrations are especially of interest during these dry periods. Previous research has disregarded the effect of tides and wind setup above 0.15 meter, deepening of the New Waterway and adjacent Botlek potentially affected the influence of these processes on chloride concentrations in the Port of Rotterdam.

During its most recent deepening, the New Waterway was deepened by approximately 1.5 meter along its entire length to facilitate accessibility of ships with a draught of up to 15 meters (Port of Rotterdam, 2016). The dredging works for this deepening started in March 2018 and were finished at the end of 2018.

Prior to the start of the dredging project, Svašek Hydraulics performed a model analysis on the potential effects of this deepening on the salinity concentrations in the Rotterdam harbour area (Svasek Hydraulics, 2015). From this work, HydroLogic deduced a synthetic dataset of chloride concentrations after deepening. These synthetic data for the situation after deepening were compared to the measured chloride concentrations prior to the deepening. With a z-score test for analysing different statistical means (Blaas & van den Boogaard, 2006), HydroLogic concluded that at the location Lekhaven a significant difference in chloride concentrations was to be expected due to the deepening ($z = 6.1$). A significant difference in chloride concentrations at Spijkenisserbrug could not be proven with this analysis ($z = 1.25$). Within this analysis, day-averaged data was used and this analysis was restricted to situations in which the discharge of the Rhine at Lobith was below 1500 m³/s. Situations at which the recorded wind setup at Hoek van Holland were above 0.15 meter were disregarded as well (HydroLogic, 2015a). By using day-averaged data the correlation with discharges were optimized but effects of the tidal variations were disregarded.

1.4 Objective and research questions

Currently it is unknown how deepening of the New Waterway and Botlek has affected chloride concentrations in the Port of Rotterdam and further upstream. Model simulations show that chloride concentrations are expected to increase due to deepening of the New Waterway and Botlek. However, these expectations are not validated with measurements of chloride concentrations post-deepening. The use of an analytical model, developed with the use of measurements, can provide this validation.

The salt intrusion process is mostly determined by the independent boundary conditions of the system: river discharge, intruding tidal wave and wind setup. This research intends to

use these parameters to build improved analytical models for the chloride concentration at each of the four measurement locations in the Port of Rotterdam. By comparing these new analytical models with measurement data collected post human interventions, the effect of human alterations in the Rhine-Meuse delta on chloride concentrations in the Port of Rotterdam can be assessed. As the construction of the Maasvlakte 2 potentially had an influence on the relation between the salt intrusion processes due to geometrical change of the mouth of the New Waterway (Blaas & van den Boogaard, 2006), the period after completion of the Maasvlakte 2 in 2011, to present is examined.

The research objective is stated as follows:

How can measurement of hydrodynamic conditions best be used in an analytical model for predicting chloride concentrations in the Port of Rotterdam, and how can this model be applied for analysing effects of human interventions in the Rhine-Meuse delta?

1. How do monitored boundary conditions relate to chloride concentrations in the Rhine-Meuse basin and how can these data best be used as an input for the analytical model?
2. What relation between salinity, at each of the four measurement locations, and the boundary conditions can be composed from measurement data obtained before deepening of the New Waterway and Botlek?
3. How can effects of human interventions on chloride concentrations in the Port of Rotterdam be analysed by application of the analytical model?

1.5 Research approach and reading guide

To answer each of the research questions an overview of the research approach is provided in Figure 9. Chapter 2 answers the first research question in which the correlation between boundary conditions and chloride concentrations is optimized in four steps. Firstly, the availability of measurement data is elaborated on. Secondly, the most optimal correlation method is determined in order to correctly relate boundary conditions to chloride concentrations. This is done by visual interpretation of scatter diagrams of the boundary conditions in relation to chloride concentrations and of the distributions of the boundary conditions. Thirdly, measurements of the boundary conditions are performed up- or downstream of the chloride concentration measurement locations. Measurements at one location take time to propagate to and affect parameters at another location, a time lag. The time lag of the boundary conditions is determined by calculating the correlation coefficient between each boundary condition and the chloride concentrations at various time shifts of the boundary conditions. With the use of the time lag analysis, multiple time series can be aligned to optimize correlation. Finally, the sampling interval of the dataset is optimized. Within this optimization, the effect of three sampling intervals on the correlation coefficients between the boundary conditions and chloride concentrations is analysed. From this analysis, the sampling interval with the highest correlation coefficients is selected.

In Chapter 3 the analytical models are developed. Firstly, the applied models in this study and corresponding optimization techniques is elaborated on. Secondly, the training- and

validation datasets are determined from the optimized dataset obtained in Chapter 2. Thirdly, in order to describe the ‘memory’ that exists in the system an autocorrelation analysis is performed on the boundary conditions. From this analysis new parameters are determined. Fourthly, the optimal set of boundary conditions in relation to observed chloride concentrations is determined by evaluating the added value of each parameter. The result are trained analytical models for each of the four measurement locations, which are evaluated with a sensitivity analysis and an uncertainty analysis. The sensitivity analysis provides insights into the importance of the boundary conditions in relation to chloride concentrations at each of the four measurement locations. The uncertainty analysis, performed for Lekhaven, is applied to analyse the effect of uncertainty in the discharge on chloride concentrations.

Chapter 4 contains a methodology for analysis of model residuals followed by an example analysis with the use of a synthetical dataset.

Chapter 5 contains the discussion of the applied methodology and outcomes. Finally, the conclusions and recommendations are provided in Chapter 6.

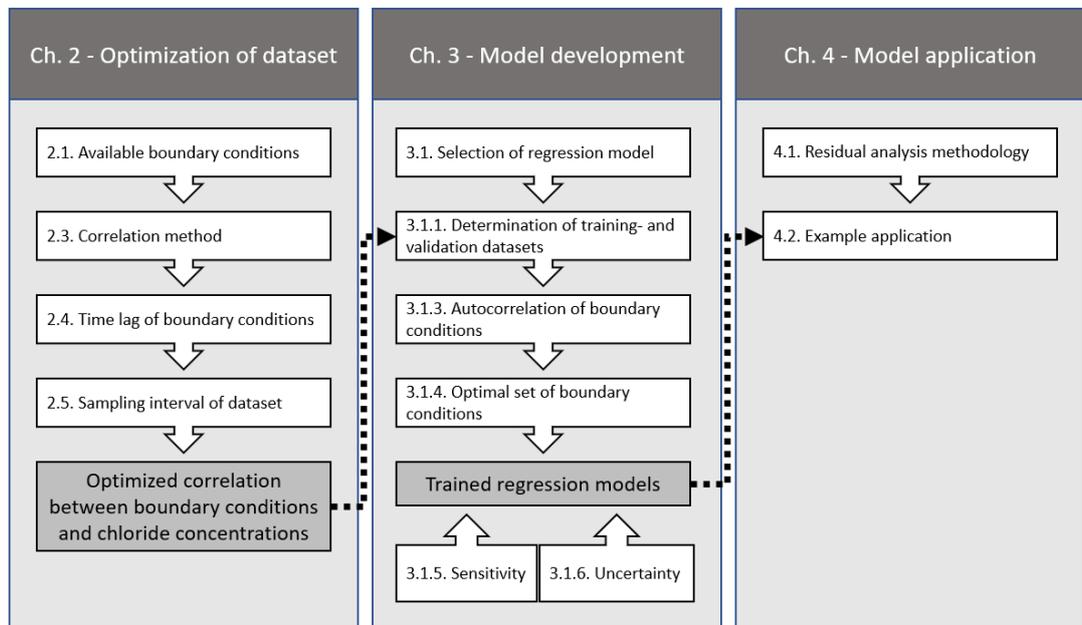


Figure 9. Flow chart of research approach.

2 Optimization of dataset



Optimization of the dataset is performed with the use of the correlation between the boundary conditions and chloride concentration measurements. The optimization consists of four parts. First, the availability of each factor affecting the salt- and freshwater flux is described. Second, the correlation method is determined by examining the type of relation between individual boundary conditions and chloride measurements. Third, the time lag of boundary conditions relative to the chloride concentration measurement, due to data monitoring at different distant locations, is optimized. Finally, the optimal sampling interval is determined by examining several time intervals for analysis.

2.1 Available measurements

Not all processes are continuously measured in the Rhine-Meuse Delta from 2011 to present. Therefore, not all processes described in Section 1.2.1. can be included in the analysis. Regarding the saltwater flux and the freshwater flux, each available parameter is briefly described. Finally, the available dataset on chloride concentrations is described.

Processes affecting the saltwater flux

At Hoek van Holland, the water level is measured at a 10-minute interval. This water level measurement can be translated in two components; the astronomical tide, which is predicted based on interactions between the planetary movements, and the wind setup, by subtracting the astronomical tide from the observed water level. The chloride concentration of the incoming seawater is mostly constant over time and is not included in this study.

Processes affecting the freshwater flux

The main inflow of freshwater is measured at a 10-minute interval by the measurement stations at Tiel, Hagestein and Megen, measuring the discharge of the Waal, Lek and Meuse, respectively. By examining observations in which the Haringvliet sluices are (almost) completely closed, corresponding to a discharge of the Rhine of 1500 m³/s at Lobith, variation in discharge distribution through the lower branches of the Rhine-Meuse Delta is assumed to be constant.

Lateral inflow or outflow by pumping stations connecting the Rhine-Meuse delta with the surrounding hinterland is not continuously measured and is therefore disregarded.

Chloride concentration measurements

As mentioned in Section 1.2.2., chloride concentrations are measured at several depths, except for Beerenplaat. Nguyen (2008) classified the New Waterway as a partially mixed estuary, where chloride concentrations gradually vary in the horizontal and vertical direction. At Lekhaven the shape of the salt wedge is very similar under various discharge conditions during low tide (Figure 10, top panels). During high tide and with increasing discharge, the vertical variation of chloride concentration decreases, by which the estuary can be classified as well-mixed. At Lekhaven, the estuary can be classified as partially mixed or well-mixed, slight variations are observed based on the quantity of discharge. Similarly, at Spijkenisserbrug the vertical variation in chloride concentrations retains a similar shape

under various discharge volumes (Figure 10, bottom panels). Due to the presence of little vertical variation in concentration, a depth-averaged chloride concentration is determined, in order to obtain a single time-dependant observation. This is applied to each chloride concentration measurement location.

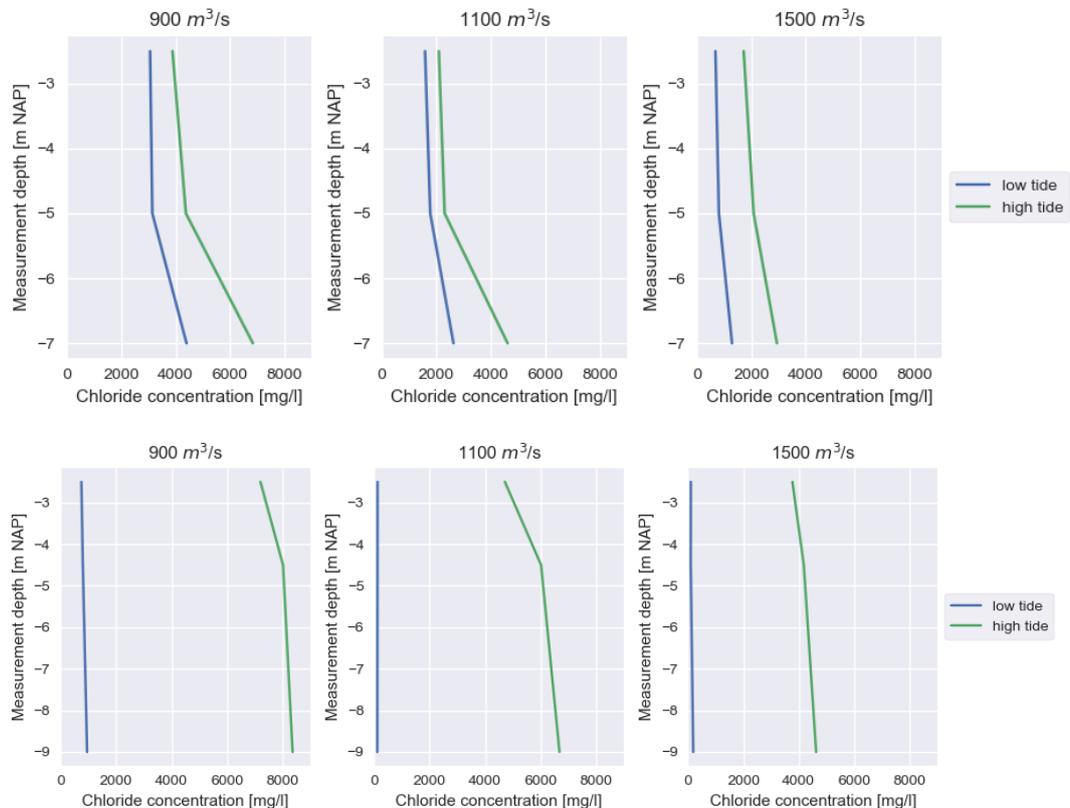


Figure 10. Shape of the salt wedge at Lekhaven (top panels) and Spijkensisserbrug (bottom panels), for three different discharge conditions of the Rhine measured at Lobith, indicated with the chloride concentration at various depths.

2.2 Analysis period

During droughts, in which the discharge of the Rhine at Lobith is below 1500 m³/s, salinization occurs at each of the four measurement locations. Determination of the time lag, which describes the propagation time of the boundary conditions to each of the measurement locations, is performed on a long period of drought in the spring of 2011. This period is selected because of the absence of long-lasting extreme wind setup events, resulting in a 'clean' signal for correlation with tide and discharge.

During this period of drought from 27th of March 2011 until 23rd of June 2011 (Figure 11), discharge of the Waal varied from 740 m³/s to 1250 m³/s and discharge of the Lek varied from 0 m³/s to 110 m³/s. Discharge of the Meuse varied from 19 m³/s to 260 m³/s. In this period substantial wind setup (>80 cm) only occurs for short time around the 24th of May 2011. Chloride concentrations at Lekhaven did not return to the background concentration of the Rhine (80-130 mg/l), indicating constant salinization at this measurement location. At the other measurement locations, the chloride concentrations did return to Rhine background concentrations, indicating the river discharge was able to flush out the intruded salt wedge.

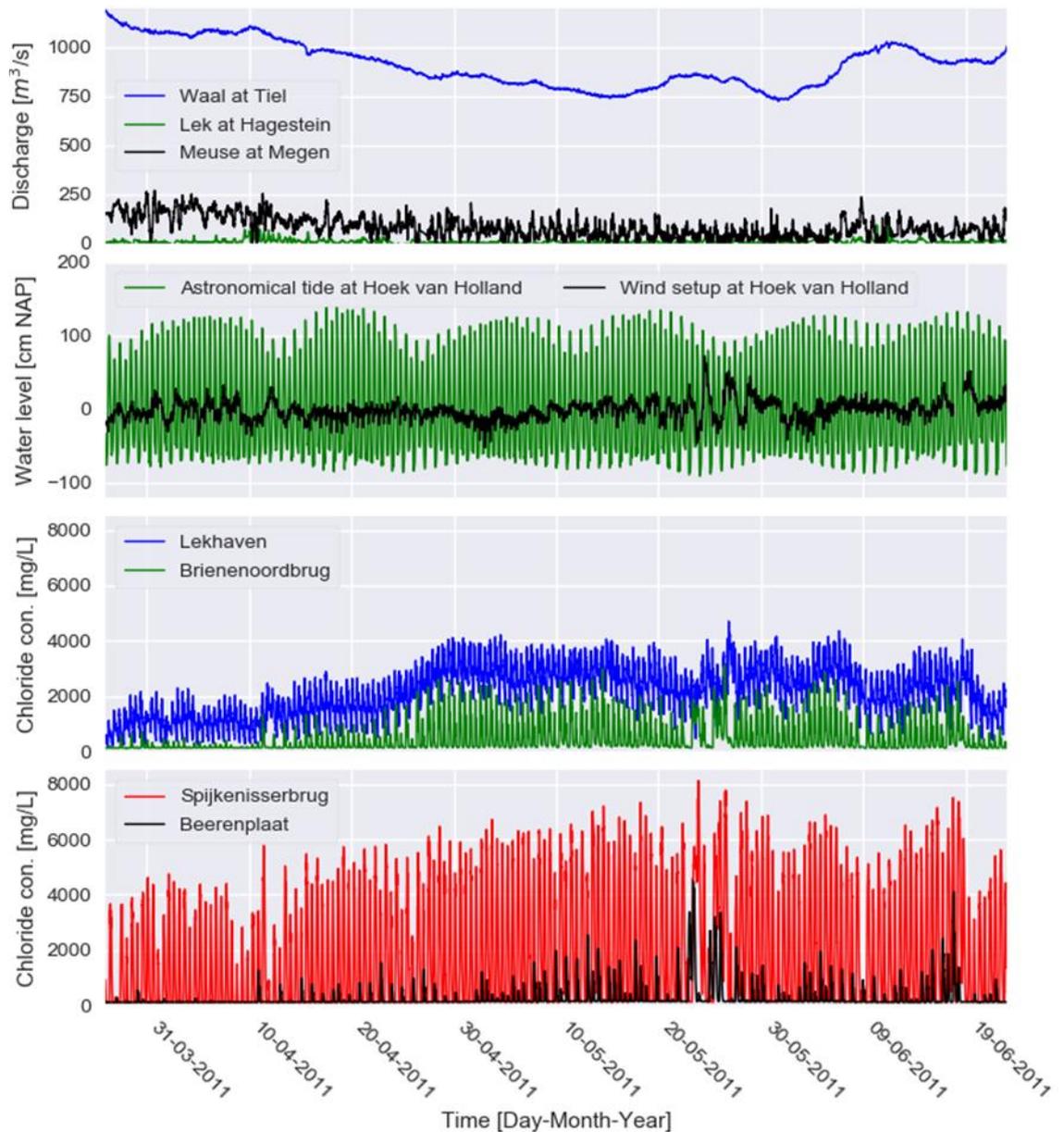


Figure 11. Overview of boundary conditions and chloride concentration measurements from 27-03-2011 until 23-06-2011.

2.3 Correlation method

Assessing correlations between water quality parameters, such as chloride concentration, and hydrodynamic processes, is a common practice in the field of hydrology (Shrestha & Kazama, 2007). Widely used correlation coefficients are the Pearson coefficient and Spearman R coefficient. The Pearson coefficient is best applied to parameters that have a normal distribution and show a linear relation between parameters. Spearman R coefficient can also handle non-normal distributed parameters and non-linear relations between parameters. Spearman R coefficient is similar to Pearson correlation, except that it is computed from ranked data (Alberto, et al., 2002).

The most basic determination of a suitable correlation method is with the use of a scatter diagram, a scatter plot of the variables. If a clear linear relation is visually detectable, the

Pearson correlation coefficient is applicable. If a clear non-linear relation is detectable or a linear relation cannot be observed the Spearman correlation method can be applied. In this study, determination of suitable correlation methods is performed based on a visual interpretation of a scatter plot of the individual boundary conditions and chloride measurements during the analysis period (Figure 12).

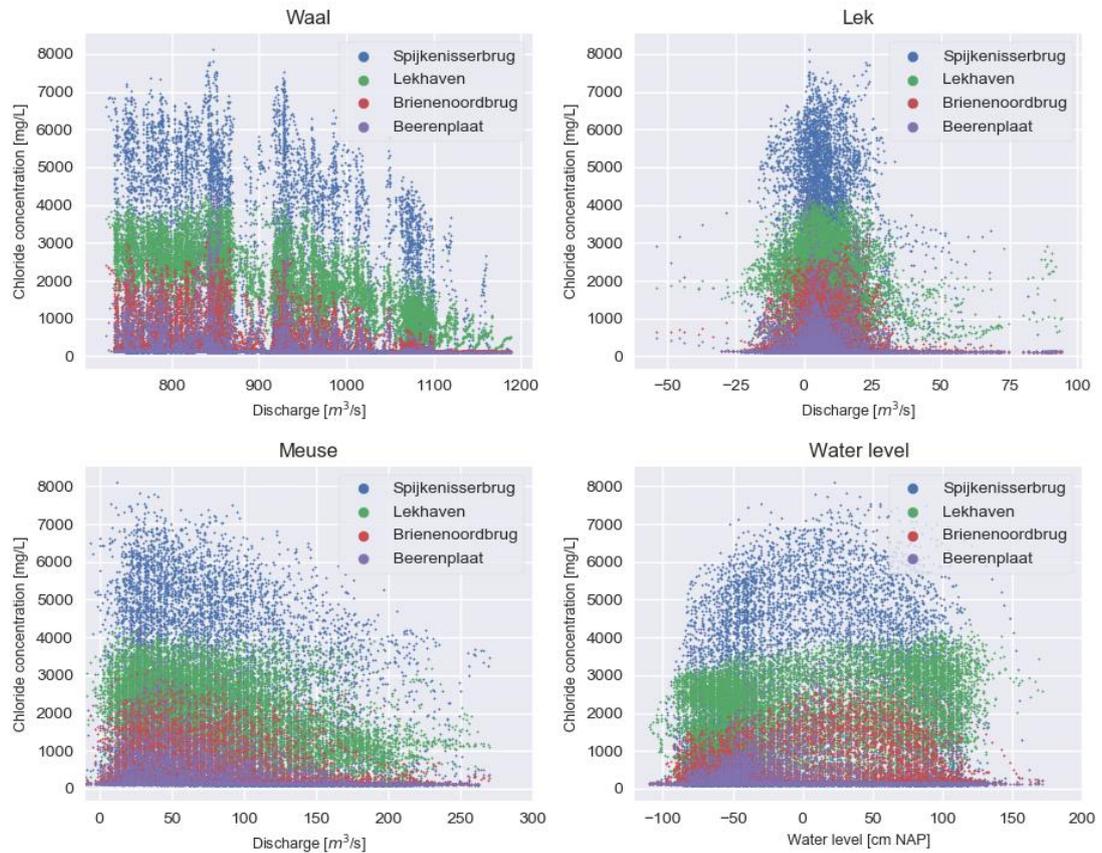


Figure 12. Scatter diagram of boundary conditions with chloride measurement locations in the analysis period from 27th of March 2011 until 23rd of June 2011.

No clear relations can be observed from the scatter diagrams in Figure 12, correlating the chloride concentrations with the boundary conditions for all measurement locations and boundary conditions simultaneously. Therefore, the distribution of each parameter is examined individually. All boundary conditions show a clear non-normal distribution (Figure 13). As the Spearman R-coefficient is capable of assessing correlation between non-normally distributed parameters, it is applied for all further analysis.

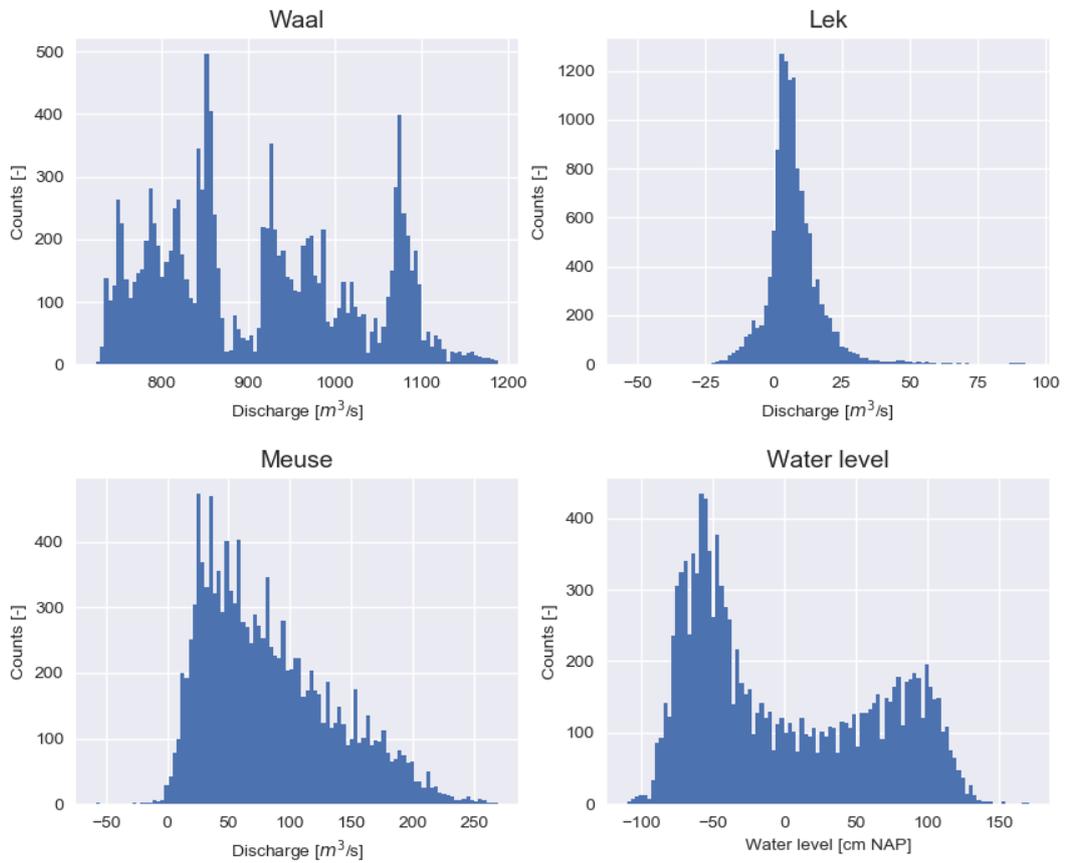


Figure 13. Distribution of boundary conditions in the analysis period from 27th of March 2011 until 23rd of June 2011. As non-normal distributions are observed, Spearman R-coefficient is applied for further analysis.

2.4 Time lag of boundary conditions

In time series analysis with a spatial orientation, observations of influencing processes at one location take time to propagate to and affect parameters at another location: a time lag. Discharges measured at Hagestein, Lek and Megen have a certain travel time before they reach the chloride measurement locations at the Port of Rotterdam. Equally, the intruding tide, measured as a water level at the mouth of the river, takes time to reach the measurement locations. With the use of time lag analysis, multiple time series can be aligned to optimize Spearman correlation. Common practice for analysing time lags of correlated variables is with a Spearman cross correlation function (CCF).

2.4.1 Water level

Correlation of the water level, measured at Hoek van Holland (HvH), and chloride concentrations at each of the four measurement locations all show a similar pattern of a sinusoidal wave when varying the time lag of the water level (Figure 14). During the spring of 2011, the highest correlation coefficients are observed at Spijkenisserbrug, in the Old Meuse, at a time lag of 190 min. Further upstream on the Old Meuse, at Beerenplaat, 90 minutes later, at a time lag of 280 min the highest correlation is found between water level at HvH and chloride concentrations at this measurement location. In the New Meuse, effects

of the intruding tide reach Lekhaven first at a time lag of 110 min, and further upstream at Brienoordbrug at 200 min.

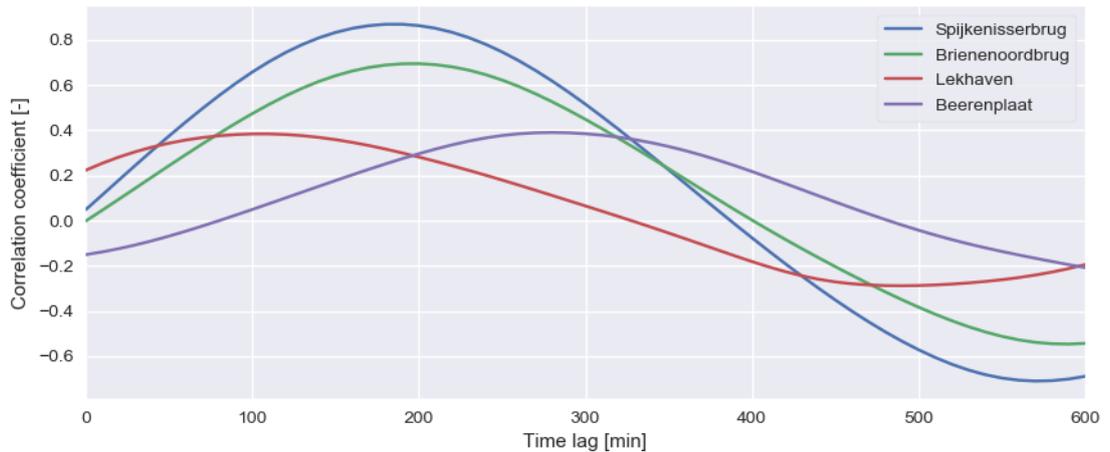


Figure 14. Spearman cross correlation function diagram of water level at Hoek van Holland and chloride concentrations at measurement locations during analysis period from 27th of March 2011 until 23rd of June 2011.

2.4.2 Discharge

Discharge through the port of Rotterdam mostly consists of discharge from the Waal, followed by Meuse discharge and Lek discharge (Section 2.2). Firstly, the time lag of Waal discharge with each of the four measurement locations is determined with a Spearman cross-correlation function (CCF). Secondly, the time lag of Meuse discharge is, simultaneously with discharge of the Waal, determined with the use of a two-dimensional CCF. Finally, with a similar methodology, the time lag of Lek discharge is determined. In order to determine time lags of discharges, a 24-hour average of the chloride concentrations, as well as the discharge, are computed. This averaging is performed after the application of each time shift.

Waal discharge

The optimum correlation between Waal discharge, measured at Tiel, and chloride concentrations at Brienoordbrug, is found at a time lag of 780 minutes. Further downstream, at Lekhaven, an optimal time lag of 1170 minutes is observed (Figure 15). Regarding the measurement location Beerenplaat, an optimum time lag is observed at 1200 minutes. Further downstream, at Spijkenisserbrug, the optimum is observed at 1070 minutes. The time lags on the Old Meuse, at Beerenplaat and Spijkenisserbrug, are unexpected, as one would expect the discharge of the Waal to reach Beerenplaat first, followed by Spijkenisserbrug sometime later. This might be caused by the relatively low correlation of both locations with Waal discharge, compared to locations on the New Meuse (Figure 15).

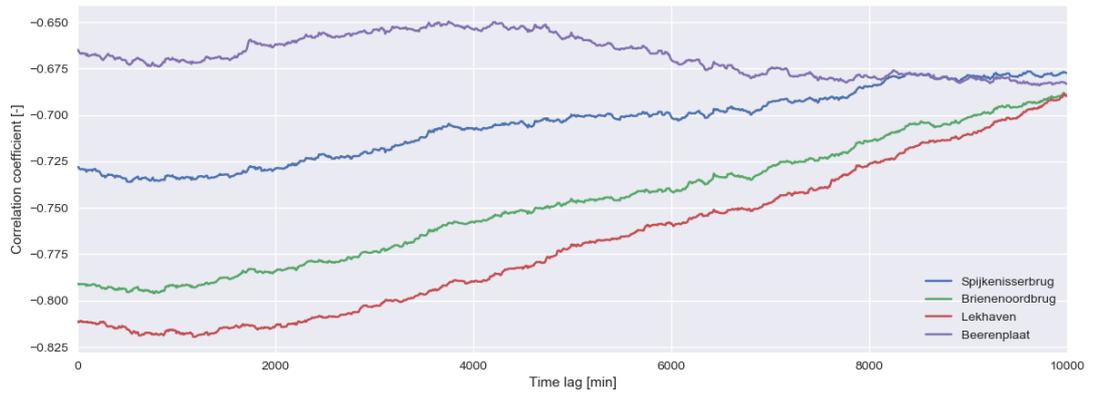


Figure 15. Spearman cross correlation function (CCF) of Waal discharge measured at Tiel and the chloride concentration measurement locations using Spearman correlation, during analysis period from 27th of March 2011 until 23rd of June 2011.

Addition of Meuse discharge

By varying the time lags for both the Waal and the Meuse discharge, a two-dimensional Spearman cross correlation function is created. With this two-dimensional CCF the optimum time lag, at which the maximum correlation exists between discharges of the Waal and the Meuse and the chloride concentrations at the measurement locations, can be determined.

The optimum for Spijkenisserbrug, with a correlation coefficient of -0.771, is observed with a Waal time lag of 1250 min and a Meuse time lag of 1650 min (Figure 16, left panel). However, an optimum range can be observed in which the correlation coefficient does not differ much from the maximum value, indicated in black and blue shades. Similarly, an optimal range is observed at measurement location Beerenplaat (Figure 16, right panel). As the highest correlation between discharge and chloride concentration is observed at Spijkenisserbrug (Table 2), this observed optimum is assumed to be most representative for the time lag. From a physical perspective, discharge from the Waal and Meuse will reach the more upstream measurement location Beerenplaat first, before reaching Spijkenisserbrug. As the distance between Beerenplaat and Spijkenisserbrug is around 10 percent of the total distance from the measurement location of Waal discharge, at Tiel, to Beerenplaat. Based on this, the time lag of the Waal and Meuse for Beerenplaat are estimated to be 1100 min and 1500 min, respectively. These estimated time lags are within the optimum range of Beerenplaat (Figure 16, right panel, indicated in blue/black)

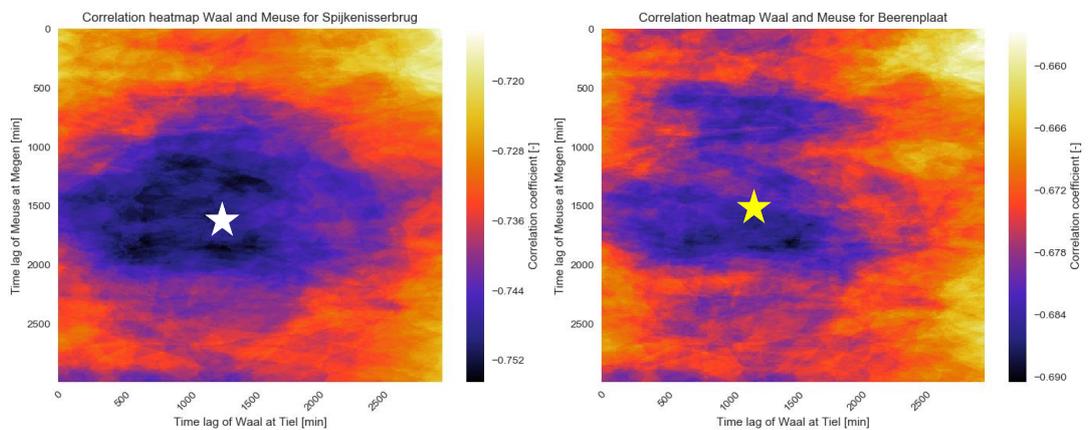


Figure 16. Correlation heatmaps of time lag between discharges of the Waal and Meuse and salinity at measurement locations Spijkenisserbrug (left) and Beerenplaat (right), both situated on the Old Meuse.

Optimum at Spijkenisserbrug indicated with white star. Estimated time lags at Beerenplaat indicated with yellow star.

Similarly to the time lag determination on the Old Meuse, time lag determination of the measurement locations on de the New Meuse is guided by the most downstream measurement location, Lekhaven. At Lekhaven an optimum is found for a Waal time lag of 1150 min and Meuse time lag of 1900 min (Figure 17). Again, based on distance of measurement locations, the time lags at Brienoordbrug are estimated at 1000 min regarding the Waal and 1750 min regarding the Meuse.

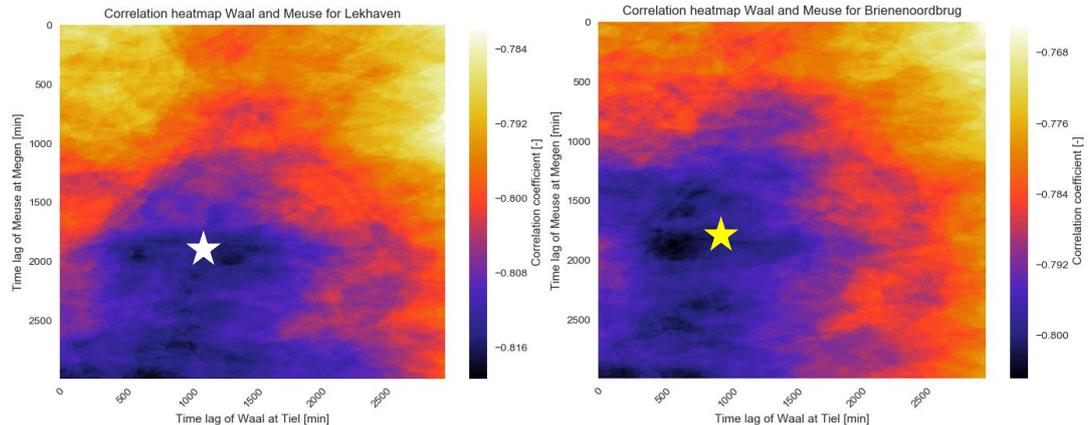


Figure 17. Correlation heatmaps of time lag between discharges of the Waal and Meuse and chloride concentrations at measurement locations Lekhaven (left) and Brienoordbrug (right), both situated on the New Meuse. Optimum at Lekhaven indicated with white star. Estimated time lags at Brienoordbrug indicated with yellow star.

The addition of Meuse discharge improves the correlation compared to an analysis based on just the Waal discharge, especially at the measurement locations on the Old Meuse: Spijkenisserbrug and Beerenplaat (Table 2). This is to be expected based on discharge distribution as mentioned in Section 1.3.1. At Lekhaven, no change in correlation is observed, and at Brienoordbrug only a small change occurs due to the addition of the Meuse.

Addition of Lek discharge

Similarly to the addition of the Meuse, discharge of the Lek is added to the Waal discharge. At Spijkenisserbrug (Figure 18, left panel) a wide range of time lags for both the Waal, indicated with a wide spread in the x-direction, as well as the Lek, indicated with a wide spread in the y-direction, is observed. The optimal correlation is found at a Waal time lag of 1200 min and a Lek time lag of 1750. The corresponding Spearman R-coefficient is -0.730. At Beerenplaat (Figure 18, right panel) no clear range of Lek discharge can be observed. For the Waal the correlation is optimal for time lags below 1500 min. Again, the time lag of the Waal and Lek at Beerenplaat is estimated based on the distance between measurement locations (Table 2).

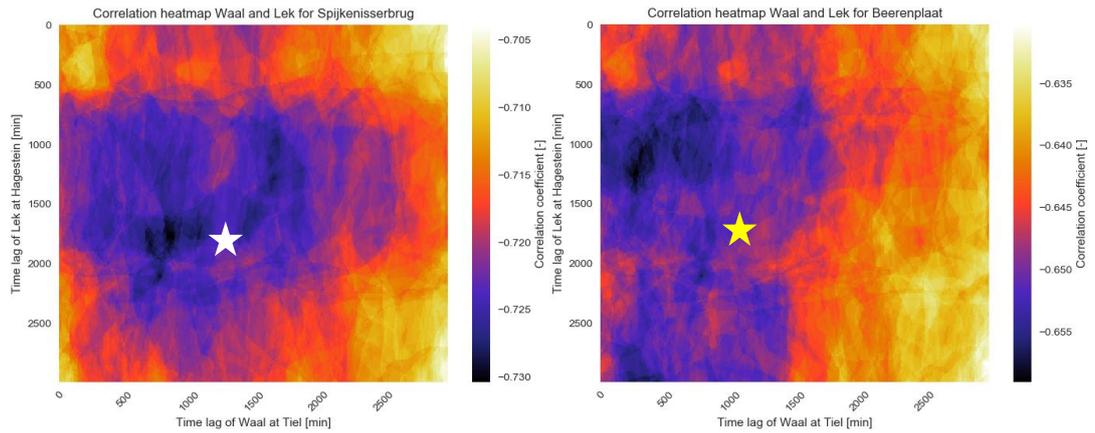


Figure 18. Correlation heatmaps of time lag between discharges of the Waal and Lek and chloride concentrations at measurement locations Spijkennisbrug (left) and Beerenplaat (right), both situated on the Old Meuse. Optimum at Spijkennisbrug indicated with white star. Estimated time lags at Beerenplaat indicated with yellow star.

At Lekhaven (Figure 19, left panel) as well as at Brienoordbrug (Figure 19, right panel) again a wide range of time lags of the Lek can be observed, possibly caused by the low discharge during the analysis period. The time lag of the Waal at Lekhaven is optimal at 1150 min and of the Lek at 900 min. The time lags of the Waal and Lek at Brienoordbrug are again estimated based on the distance between measurement locations.

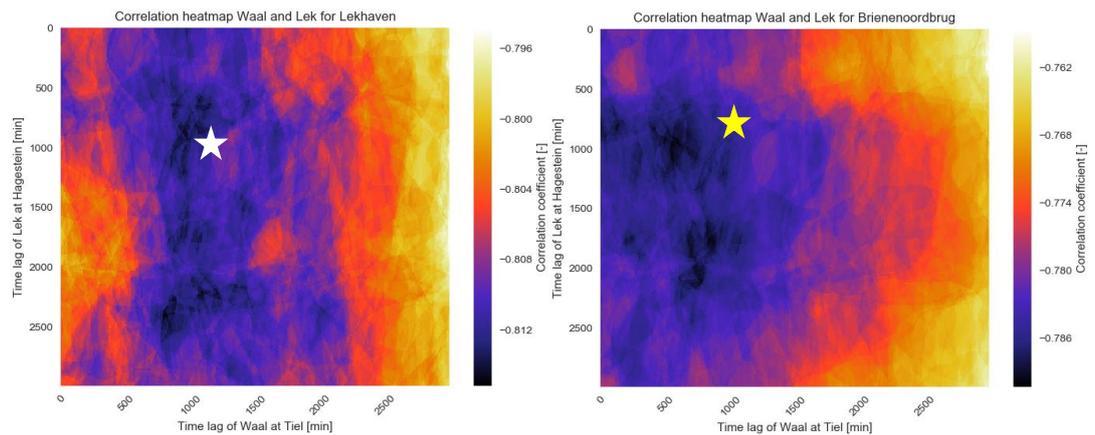


Figure 19. Correlation heatmaps of time lag between discharges of the Waal and Lek and chloride concentrations at measurement locations Lekhaven (left) and Brienoordbrug (right), both situated on the New Meuse. Optimum at Lekhaven indicated with white star. Estimated time lags at Brienoordbrug indicated with yellow star.

At none of the measurement locations the addition of Lek discharge improves the Spearman R-coefficient compared to an analysis only including the Waal discharge (Table 2). Contrary, at all locations the correlation between discharge and chloride concentrations decrease. The observed and estimated time lag of the Waal discharge is similar to the time lag when only considering the Waal or when considering the Waal with addition of the Meuse (Table 2).

Table 2. Spearman R-coefficients (Corr.) and time lags for Waal discharge and for Waal discharge with the addition of Meuse and Lek discharge, at each of the four measurement locations.

Measurement location:	Waal discharge		Waal and Meuse			Waal and Lek		
	Corr. [-]	Time lag [min]	Corr. [-]	Time lag [min]		Corr. [-]	Time lag [min]	
				Waal	Meuse		Waal	Lek
<i>Old Meuse</i>								
Spijkenisserbrug	-0.736	1070	-0.771	1250	1650	-0.730	1250	1750
Beerenplaat	-0.660	1200	-0.708	1100	1500	-0.651	1100	1600
<i>New Meuse</i>								
Lekhaven	-0.820	1170	-0.820	1150	1900	-0.815	1150	900
Brienoordbrug	-0.796	780	-0.805	1000	1750	-0.784	1000	750

2.5 Sampling interval

Previous research used 24-hour averaged values for the determination of the Rhine discharge time lag (HydroLogic, 2015a). In section 2.4.2 again 24-hour averaged values were used to determine correlations between Waal, Meuse and Lek discharges in relation to chloride concentrations. For the determination of time lags of the astronomical tide in relation to chloride concentration measurements, in Section 2.4.1., the original data interval of 10 minutes was applied. Considering the most appropriate averaging interval, it needs to be considered that in order to incorporate tides, wind and discharges into a single analysis, the sampling interval may not exceed the duration of half a tidal cycle, as this will cause information loss from the tidal signal.

Three time intervals are examined, the original 10-minute interval of the data, an hourly average and data sampling based on half a tidal cycle. The tidal sampling is based on peaks and troughs in the tide signal (Figure 20). The cycle is split in two parts, from low water level to high, the incoming tidal wave, and from high to low water, the outgoing tidal wave. During the incoming tidal wave (or flood), the minimum chloride concentration and minimum water level and tidal water level are selected. An average over half a cycle is taken of the discharge and wind setup during the flood period. During the outgoing tidal wave (or ebb), the maximum chloride concentration and maximum water level and tidal water level are taken. Again, the average is taken of the discharge and wind setup during the ebb period. The Spearman correlation method and time lags determined in previous sections are used.

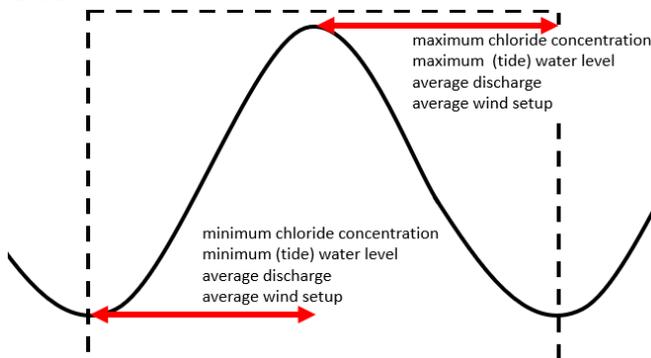


Figure 20. Applied sampling for tidal sampling technique based on the peaks and troughs in the astronomical tide signal.

Compared to the correlation of the 10 min data, an hourly average of water level measured at Hoek van Holland shows a slight improvement in correlation coefficient between the water level and chloride concentrations (Table 3). Applying the tidal sampling technique further improves the correlation between chloride concentrations and water level measurements for all measurement locations except Spijkenisserbrug. Regarding Spijkenisserbrug, a slight decrease of the Spearman R coefficient is observed. Wind setup and astronomical tidal water level show a similar pattern (Table 4 and Table 5). By taking an hourly average, the Spearman R-coefficient slightly increases, and by applying tidal sampling the coefficient further increases, except for Spijkenisserbrug.

Table 3. Spearman R correlation coefficients between water level, measured at Hoek van Holland, and chloride concentrations at each of the four measurement locations for three sampling intervals.

Measurement location	10 min (original data)	1 hour	Tidal sampling
Brienoordbrug	0.6924	0.7027	0.8005
Lekhaven	0.3835	0.3885	0.4933
Spijkenisserbrug	0.8674	0.8796	0.8456
Beerenplaat	0.3900	0.4075	0.6523

Table 4. Spearman R correlation coefficients between wind setup, derived from water level measured at Hoek van Holland, and chloride concentrations at each of the four measurement locations for three sampling intervals.

Measurement location	10 min (original data)	1 hour	Tidal sampling
Brienoordbrug	0.1074	0.1019	0.1243
Lekhaven	- 0.0049	- 0.0003	0.0401
Spijkenisserbrug	0.0933	0.0908	0.1813
Beerenplaat	0.1528	0.1554	0.1748

Table 5. Spearman R correlation coefficients between astronomical tide determined at Hoek van Holland and chloride concentrations at each of the four measurement locations for three sampling intervals.

Measurement location	10 min (original data)	1 hour	Tidal sampling
Brienoordbrug	0.6555	0.6665	0.7690
Lekhaven	0.3792	0.3840	0.4929
Spijkenisserbrug	0.8333	0.8472	0.8011
Beerenplaat	0.3619	0.3772	0.5946

Correlating hourly averages of the discharge has no influence on the Spearman R coefficient (Table 6, Table 7 and Table 8) compared to the correlation coefficients of the 10 minute data. When applying the tidal sampling on Waal, Waal + Lek and Waal + Meuse the correlation decreases, except for a minor improvement at Spijkenisserbrug.

Table 6. Spearman R correlation coefficients between Waal discharge measured at Tiel and chloride concentrations at each of the four measurement locations for three sampling intervals.

Measurement location	10 min (original data)	1 hour	Tidal sampling
Brienoordbrug	- 0.5483	- 0.5479	- 0.4247
Lekhaven	- 0.7185	- 0.7309	- 0.7628
Spijkenisserbrug	- 0.3130	- 0.3034	- 0.3524
Beerenplaat	- 0.6686	- 0.6640	- 0.5450

Table 7. Results of correlations for three sampling intervals regarding Waal and Lek discharge.

Measurement location	10 min (original data)	1 hour	Tidal sampling
----------------------	------------------------	--------	----------------

Brienoordbrug	- 0.5261	- 0.5260	- 0.4214
Lekhaven	- 0.6983	- 0.7105	- 0.7669

 Table 8. Results of correlations for three sampling intervals regarding Waal and Meuse discharge.

Measurement location	10 min (original data)	1 hour	Tidal sampling
Spijkenisserbrug	- 0.3125	- 0.3029	- 0.3525
Beerenplaat	- 0.6551	- 0.6523	- 0.5325

Although correlations between discharges and chloride concentrations generally decrease with tidal sampling, the correlations of all other parameters show a greater increase. Therefore, tidal sampling is applied for further analysis.

2.6 Summary

Due to the non-normal distributed hydrodynamic input parameters (Figure 13), correlations with chloride concentrations at each of the four measurement locations is best described with the Spearman R coefficient.

From the time lag analysis, the propagation time of hydrodynamic boundary conditions water level and discharge are determined (Table 9). As the effect of wind setup depends on the intruding or outgoing tidal wave (Section 1.2.1), the time lags of wind setup regarding each measurement locations are assumed equal to the time lags regarding tide.

The sampling interval analysis shows that applying a tidal sampling interval provides the best correlations for chloride concentrations and water level, astronomical tide and wind setup at each measurement location. Although the tidal sampling interval causes a decrease in correlation coefficient of discharge of the Waal, Lek and Meuse with the chloride concentrations within this study area, this decrease is less significant. Therefore, for further analysis tidal sampling is applied.

Table 9. Time lags of tide and Waal, Meuse and Lek discharges regarding each of the four measurement locations.

	Tide	Waal	Meuse	Lek
<i>Old Meuse</i>				
Spijkenisserbrug	190 min	1250 min	1650 min	1750 min
Beerenplaat	280 min	1100 min	1500 min	1600 min
<i>New Meuse</i>				
Lekhaven	110 min	1150 min	1900 min	900 min
Brienoordbrug	200 min	1000 min	1750 min	750 min

3 Chlorinity predictor model development



3.1 Regression model building and validation methodology

Development of a chlorinity predictor is performed by a regression analysis. A linear and a non-linear model are created based on a limited dataset prior to deepening, the training set. The residuals of the model predictions for the training are examined. Validation of the model is performed with a different dataset of measurements under similar conditions prior to deepening, the validation dataset.

Regression analysis is performed with the open source machine learning library Scikit-Learn in Python (Pedregosa, et al., 2011). Two models are selected for this regression analysis, a linear model and a non-linear model. The linear regression model makes use of an ordinary least squares (OLS) optimization. The non-linear model consists of a linear model, but uses polynomial input features created from the selected parameters. The non-linear model uses a technique of least absolute shrinkage and selection operator (Lasso) which performs both variable selection and regularization in order to enhance the accuracy and prevent overfitting. In order to reduce the number of parameters in the regression an extended Lasso-model with cross-validation is applied (LassoCV). The addition of cross-validation reduces overestimation of the model (Chetverikov & Liao, 2016). In further analyses both models are run simultaneously. The linear OLS model is more simplistic and uses fewer parameters compared to the non-linear LassoCV model, which is potentially more accurate.

Performance of the linear and non-linear model is tested with the use of the coefficient of determination, the r-squared (Eq. 1). The coefficient of determination is the proportion of the variance in the dependant variable, the chloride concentration in this study, that is predicted by the independent variables, the boundary condition parameters. The r-squared value varies between 0 (no predictive value) and 1 (perfect prediction).

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (\text{Eq. 1})$$

where \hat{y}_i is the prediction value of y for observation i , \bar{y} is the mean of y and y_i is the y value for observation i .

The error of the model is indicated with the root-mean squared error (RMSE). The RMSE (Eq. 2) is the standard deviation of the residuals (Barnston, 1992). The RMSE is used to indicate the spread of the residuals around the line of best fit, and has the unit of the dependant prediction variable, thus in mg/L.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (\text{Eq. 2})$$

where n is the number of observations.

The derivation of the model consists of several steps. First, the input parameters are normalized, with the use of scaling, in order to be able to compare coefficients of the various parameters. The influence of parameters, such as discharge or wind setup, might not be linear in relation to observed chloride concentrations. Non-linear weighting is applied to several parameters based on physical relationships for each relevant process. Second, the training and validation datasets are elaborated on. Third, based on autocorrelation analysis of the input parameters, new parameters are derived with the goal of incorporating the 'memory' of the system. Fourth, the multi-step analysis is explained to determine the most

suitable parameters for describing the chloride concentrations at each of the four measurement locations. Finally, effect uncertainty in the input boundary conditions on predicted chloride concentrations is examined.

3.1.1 Training and validation datasets

Validation of both regression models is performed according to the hold-out method (Devroye & Wagner, 1979). In the hold-out method part of the dataset is not used for model training, but for model validation. This fundamental model validation method is best applied when particular sequences within datasets are used for either training or validation (Arlot & Celisse, 2010).

The training dataset consists of three long periods of Rhine discharge at Lobith below 1500 m³/s (Figure 21, indicated in green). The training set consists of data measurements gathered throughout all four seasons (Table 10). The validation data set consists of all data points prior to the deepening in 2018, below a Rhine discharge at Lobith of 1500 m³/s, excluding the training dataset (Figure 21, indicated in black). Similarly, to the training dataset, the validation data set contains datapoints throughout all seasons.

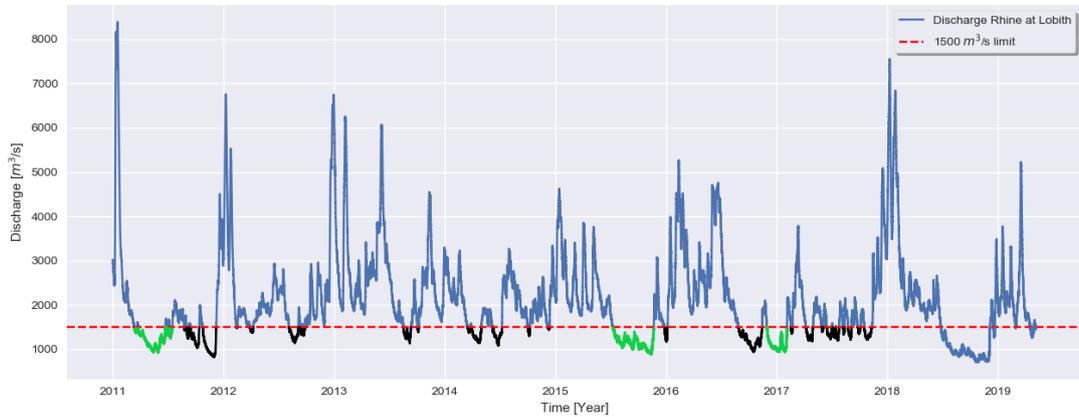


Figure 21. Discharge of Rhine at Lobith from 2011 until start of 2019. The training dataset is indicated in green, the validation dataset consists of all data points prior to the deepening started in March 2018 and below a discharge of the Rhine at Lobith of 1500 m³/s (indicated in black).

Table 10. Distribution of training and validation datapoints by season.

Season	Number of datapoints	
	Training set	Validation set
Spring	254	296
Summer	274	544
Autumn	313	894
Winter	252	123
<i>Total:</i>	1093	1857

3.1.2 Parameter normalization and non-linear weighting

Normalization of boundary condition parameters is applied to compare the influence of each boundary condition on chloride concentrations individually. This is done by analysing the coefficients the linear or non-linear model assigns to each individual parameter. The boundary conditions are therefore normalized to a range between 0 and 1. Except for wind setup, which is normalized between -0.5 and 0.5 since negative values have an opposite effect. The values corresponding with the normalized -0.5 and 0.5 or 0 and 1 are provided

in Table 11 for each parameter. Astronomical tide is normalized such that the original mean value of 0.30 m NAP corresponds with the normalized value of 0.5. Similarly, wind setup is normalized such that the normalized value 0, corresponds with no wind setup. The corresponding original values (Table 11) correspond with the minimum and maximum values is the combined training and validation dataset.

Table 11. Criteria for normalization of input parameters.

Parameter	Normalized minimum value	Corresponding original value	Normalized maximum value	Corresponding original value
Astronomical tide	0	-100 [cm NAP]	1	160 [cm NAP]
Discharge Waal	0	0 [m ³ /s]	1	1500 [m ³ /s]
Wind setup	-0.5	-121 [cm]	0.5	121 [cm]
Discharge Meuse	0	0 [m ³ /s]	1	522 [m ³ /s]
Discharge Lek	0	0 [m ³ /s]	1	175 [m ³ /s]

Wind setup not only affects the height of the tidal wave, it can also be applied as indicator of mixing processes inside the estuary, as mentioned in Section 1.2.1. Also, in situations with extreme wind setup, no water level difference is present between the mouth of the estuary and further upstream in the estuary, preventing the tidal wave from extruding (Deltares, 2016). Therefore, a non-linear weighting is applied on to the normalized value of the wind setup parameter. Similarly, the incoming tide affects not only the amount of saline water intruding in the estuary, it also affects mixing processes. Therefore, also a non-linear weighting is applied to the astronomical tide. Squaring of the normalized parameter, which is performed by default when compiling polynomial features, gives more weight to higher values and decreases the weight of lower values (Figure 22, green line). However, a desired weighting would be an exponentially increasing weight as seen from the 0.5 normalized value and higher values, representing increasing weight for more positive values, and an exponentially decreasing weight as seen from the 0.5 normalized values and lower, representing increasingly negative values in the original data. An inverted Smootherstep distribution exactly describes this distribution (Figure 22, red line):

$$f(x) = x + (x - (x^3 * (6x - 15) + 10)) \tag{Eq. 3}$$

A similar non-linear weighted distribution is applied to the normalized wind setup between -1 and 1, however now the original ‘no wind setup’ value corresponds with a normalized value of 0.

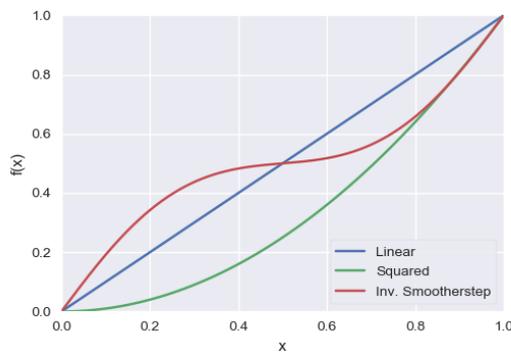


Figure 22. Distribution functions applied for creating polynomial input features LassoCV non-linear model. x-axis indicating the original normalized value, f(x) representing the non-linear weighted value of x.

3.1.3 Autocorrelation of boundary conditions

Autocorrelation analysis is applied to examine to what extent samples of a series, in this case the boundary conditions, are related in time and to what extent a ‘memory’ exists in the system (Blaas & van den Boogaard, 2006). By computing parameters based on the autocorrelation analysis, observed chloride concentrations can be indicated as independent from previous observations. Thus, eliminating autocorrelation in the predicted chloride concentrations. Within this autocorrelation analysis the threshold of ‘memory’ in the system is set at a correlation equal to the e-folding value (i.e. $e^{-1} \approx 0.367$), which is often applied in analyses with hydrological parameters (Blaas & van den Boogaard, 2006; Gerberet al., 2008; Park et al., 2018). The extent of the system ‘memory’ is indicated as the autocorrelation time (τ_{at}).

The auto-correlogram of the wind setup indicates an autocorrelation time $\tau_{at} = 2$ sampling periods (Figure 23), which is equal to one tidal cycle or about 12.4 hours. A parameter describing the ‘memory’ of wind setup is added based on a moving weighted average (MWA), which weighs previous observations, in this case two, according to a linear series between 0 and 1 (Appendix B). The new parameter thus contains a weighted average of the two previous wind setup measurements.

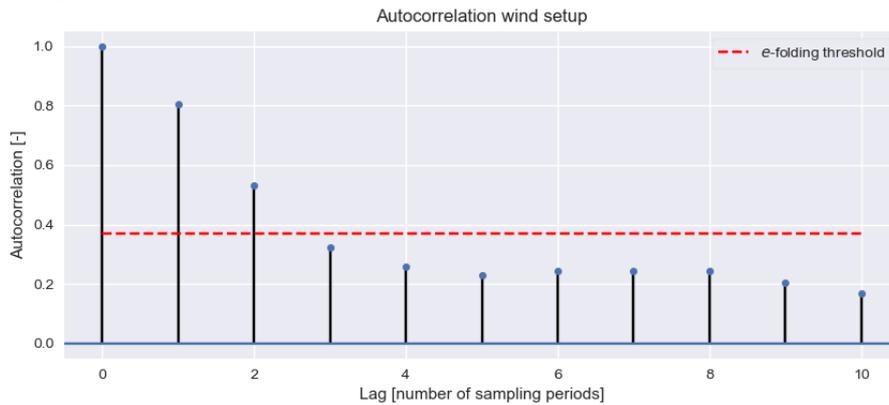


Figure 23. Autocorrelation of wind setup of tidal sampled data.

A similar approach is applied to the discharge time series of the Waal, Meuse and Lek. From this an autocorrelation time of the Waal discharge is determined at $\tau_{at} = 94$ sampling periods, about 24 days (Figure 24). Autocorrelation times of the Meuse and Lek are much smaller, $\tau_{at} = 27$ sampling periods (~7 days) and $\tau_{at} = 4$ sampling periods (~1 day), respectively (Appendix 2).

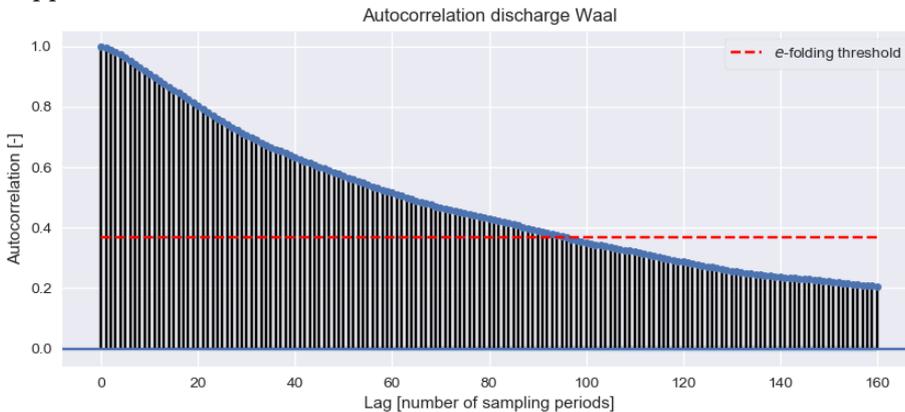


Figure 24. Autocorrelation of Waal discharge of tidal sampled data.

Also in the astronomical tidal signal autocorrelation is present (Figure 25) due to the neap-spring cycle. The autocorrelation of the astronomical tide $\tau_{at} = 90$, about 12 days.

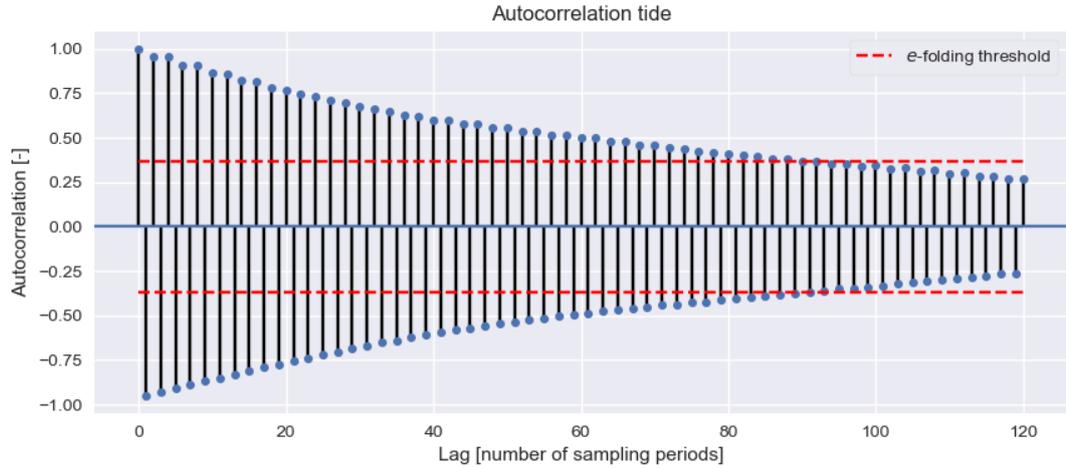


Figure 25. Autocorrelation of astronomical tide of tidal sampled data.

Finally, a parameter is added describing the absolute change in water level from low tide to high tide and visa-versa, the tidal amplitude. This change in water level is relevant to the amount of saline water flowing in or out of the estuary.

3.1.4 Parameter selection

Model development is a trade-off between model performance on the training dataset and model performance on the validation dataset (Figure 26). An increase in model complexity, by for example the addition of parameters, generally increases the model performance on the training dataset. A too complex model, however, can cause over-fitting on the training dataset, which leads to a decrease of model performance on the validation dataset and increased variance, indicated with the RMSE. A too simplistic model has a low performance on the training dataset as well as on the validation dataset. The model is optimized by selection of the model complexity corresponding with the highest model performance on the validation dataset .

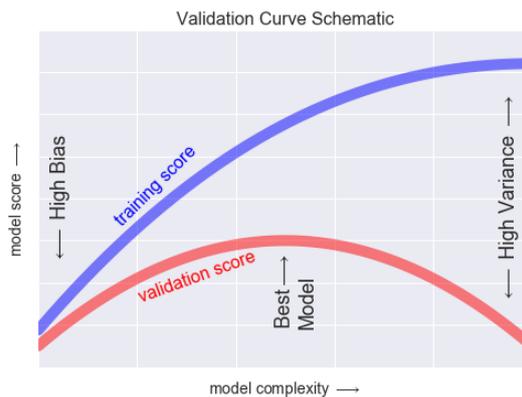


Figure 26. Schematic trade-off between model complexity and model performance.

Model complexity and performance is assessed with the use of a multi-step analyses. The multi-step analysis is designed to assess the effect of each additional parameter on the model performance indicators R^2 and RMSE, on the training dataset as well as on the validation dataset.

For each of the four chloride concentration measurement locations, a multi-step analysis is performed to determine the most suitable set of parameters to describe chloride

concentrations at that specific location. The multi-step analysis regarding the linear OLS model evaluates the addition of five independent input parameters (i.e. tide, Waal/Meuse/Lek discharge and wind setup), with addition of the moving weighted averages (MWS) of these parameters developed using the autocorrelation time (Section 3.1.3) and the tidal amplitude (Table 12). The non-linear LassoCV model incorporates, in addition to the first 11 parameters, each cross-product as a parameter. For example, when evaluating the first three parameters the additional cross-products: [Tide * Discharge Waal], [Tide * Wind setup] and [Discharge Waal * Wind setup] are added. These parameters are added in order to describe processes, as described in Section 1.2.1, which are dependent on multiple input parameters, such as the inflow of the tidal wave. The non-linear model also incorporates the non-linear weighted parameters (1b, 3b, 9b and 10b in Table 12) as described in Section 3.1.1.

Table 12. Parameters for multi-step analysis.

Parameter number	Parameter name	Parameter unit
<i>Independent input parameters:</i>		
1	Astronomical Tide	[cm NAP]
2	Discharge Waal	[m ³ /s]
3	Wind setup	[cm]
4	Discharge Lek	[m ³ /s]
5	Discharge Meuse	[m ³ /s]
<i>MWA of autocorrelation of parameter:</i>		
6	Discharge Waal	[m ³ /s]
7	Discharge Meuse	[m ³ /s]
8	Discharge Lek	[m ³ /s]
9	Wind setup	[cm]
10	Astronomical Tide	[cm NAP]
11	Tidal amplitude	[cm]
<i>Smootherstep of:</i>		
1b	Astronomical tide	[cm NAP]
3b	Wind setup	[cm]
9b	MWA wind setup	[cm]
10b	MWA astronomical tide	[cm NAP]

Based on the outcome of the multi-step analysis, the most suitable parameters for each of the four measurement locations are selected and applied for further analysis.

The criteria that are applied for incorporating any addition parameter are:

- Inclusion of the parameter increases the model performance on the training dataset, without decreasing model performance on the validation dataset, or:
- Inclusion of the parameter increases the model performance on the validation dataset.

3.1.5 Sensitivity analysis

A sensitivity analysis is performed based on the SALib-package (Usher, et al., 2016). SALib is an open source library written in Python for performing sensitivity analysis. It can be applied to Python based models such as the applied linear OLS model and non-linear LassoCV model.

The applied technique for computing the sensitivity of the model to the input parameters is the Sobol' sensitivity analysis (Sobol', 2001), a variance-based sensitivity analysis. Based on the spread of the input parameters, this technique determines the sensitivity of the model output based on variation of the input parameters (Saltelli, et al., 2010).

The computed sensitivity indices refer to the fraction of variance in the output. Main advantage of the Sobol' sensitivity is the division of first and higher order sensitivity indices. First order sensitivity indices indicate the fraction of variance explained by the input variable. Higher order sensitivity indices indicate the explained variance of the output from interactions between input parameters and non-linearity of input parameters, such as applied in the non-linear LassoCV model. The output of the Sobol' sensitivity analysis is a total order sensitivity index, which is a combination of first and higher order indices (Zhang, et al., 2015). Application of this Sobol' sensitivity analysis makes it possible to compare sensitivity of both models to each of the input parameters, by comparing the total order sensitivity index computed for each model.

3.1.6 Uncertainty analysis

Measurement of the input parameters is not performed at the location of chloride concentration measurements. For example, the discharge of the Waal may change between Hagestein and Lekhaven due to water withdrawal or lateral inflow from pumping stations along the river.

Especially during spring and summer, water is needed for agricultural use. For measurement location Lekhaven, on the New Meuse, we assess the effect of water withdrawal and discharge from Boezemgemaal Gouda on chloride concentrations. This pumping station is capable of pumping up to 50 m³/s (HydroLogic, 2015b).

3.2 Lekhaven prediction model

Computation of the Lekhaven prediction models consists of several steps. Firstly, the most suitable parameters are selected with the use of the multi-step analysis. Secondly, results of this parameter selection are used to train the model based on the training dataset. Model performance on this training dataset is further elaborated on. Thirdly, performance of the model is further specified by examining predictions for the validation dataset and analysis of the residuals. Furthermore, a sensitivity analysis is applied to both models. Finally, the effect of uncertainty in the discharge input of the Waal on chloride concentration predictions is assessed. Based on comparisons of the model performances, one model is selected for further analysis.

3.2.1 Parameter selection with multi-step analysis

Linear OLS model

From the multi-step analysis it can be concluded that the addition of the Lek (as regular parameter or as MWA parameter) has no influence on the performance of the model (Figure 27). Similarly, the addition of a parameter based on the autocorrelation of the Waal (MWA Discharge Waal) has no influence on model performance. These three parameters are therefore disregarded for further computation of the Lekhaven prediction model prior to deepening. The addition of the parameter describing the autocorrelation of the astronomical tide improves the model performance on the training dataset, however performance on the validation dataset decreases. Therefore, the parameter 'MWA Tide' is also not incorporated in further analysis.

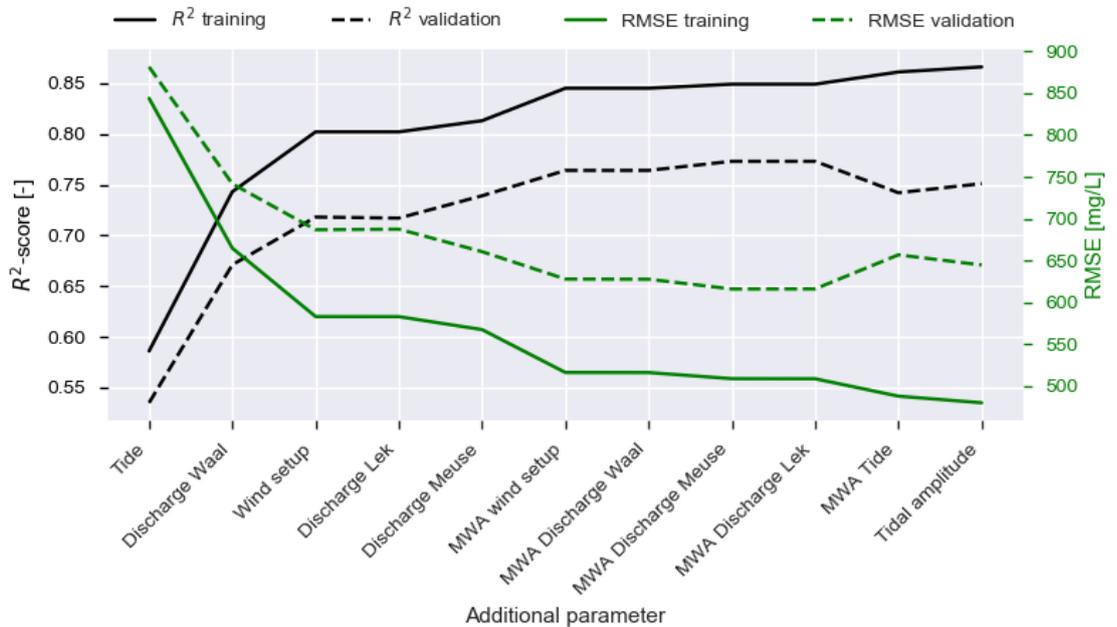


Figure 27. Results of multi-step analysis for parameter selection of the linear OLS model regarding measurement location Lekhaven. Results consist of model performance indicators R^2 and RMSE.

Nonlinear LassoCV model

Similarly, the multi-step analysis has been applied to the non-linear model (Figure 28). Again, the addition of Lek discharge has no effect on model performances. Addition of the

parameters ‘MWA Discharge Waal’ and ‘Smootherstep tide’ also have no influence on model performance on the training dataset as well as on the validation dataset. The addition of ‘MWA tide’ and ‘Smootherstep MWA tide’ improve the model performance on the training dataset, however, have a negative effect on the model performance on the validation dataset, and thus are disregarded.

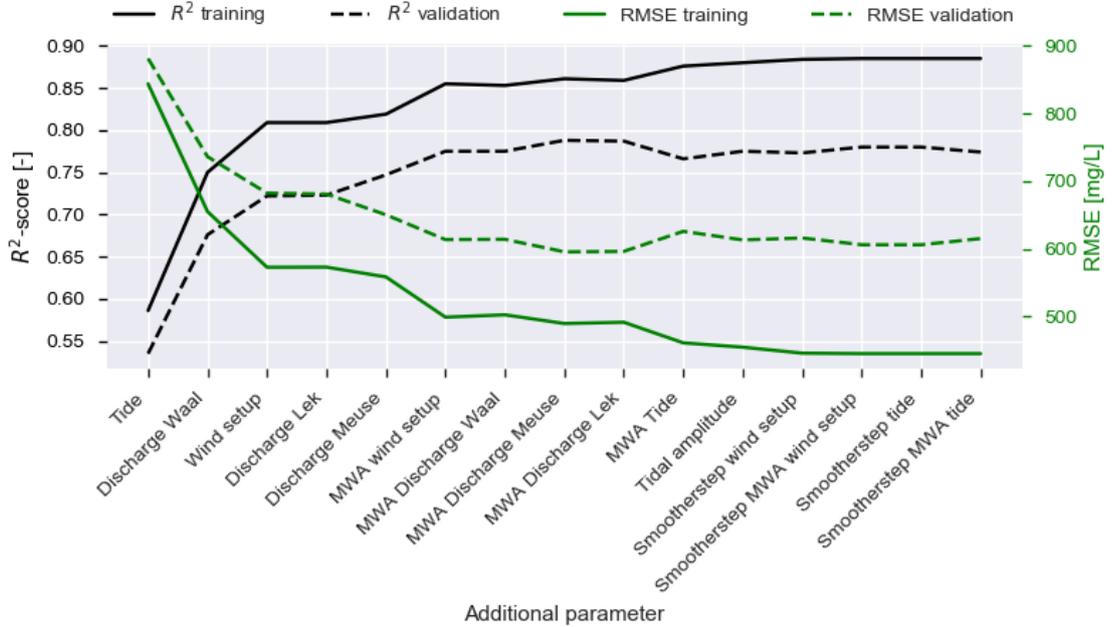


Figure 28. Results of multi-step analysis for parameter selection of the non-linear LassoCV model regarding measurement location Lekhaven. Results consist of model performance indicators R2 and RMSE.

Table 13 shows the selected parameters for predicting chloride concentrations at Lekhaven regarding the linear and non-linear model which are determined with the multi-step analysis. Model coefficients for each of the parameters are presented in Annex B.

Table 13. Parameters selected for computation of the linear OLS model and the non-linear LassoCV model with the use of the multi-step analysis for location Lekhaven.

Parameter name	linear OLS model	nonlinear LassoCV model
Astronomical Tide	x	x
Discharge Waal	x	x
Wind setup	x	x
Discharge Lek		
Discharge Meuse	x	x
Tidal amplitude	x	x
MWA Discharge Waal		
MWA Discharge Meuse	x	x
MWA Discharge Lek		
MWA Wind setup	x	x
MWA Astronomical Tide		
Smootherstep Astronomical tide		
Smootherstep Wind setup		x
Smootherstep MWA wind setup		x
Smootherstep MWA astronomical tide		

3.2.2 Performance on training dataset

The linear OLS and non-linear LassoCV model are trained based on the training dataset of Lekhaven chloride concentration observations and parameters as selected with the use of the multi-step analysis. With these models, expected chloride concentrations can be derived based on the input boundary conditions. The predicted chloride concentrations are compared to the measured chloride concentrations (Figure 29). The coefficient of correlation (R^2) and error (RMSE) of the model are based on these predicted and measured chloride concentrations (Table 14).

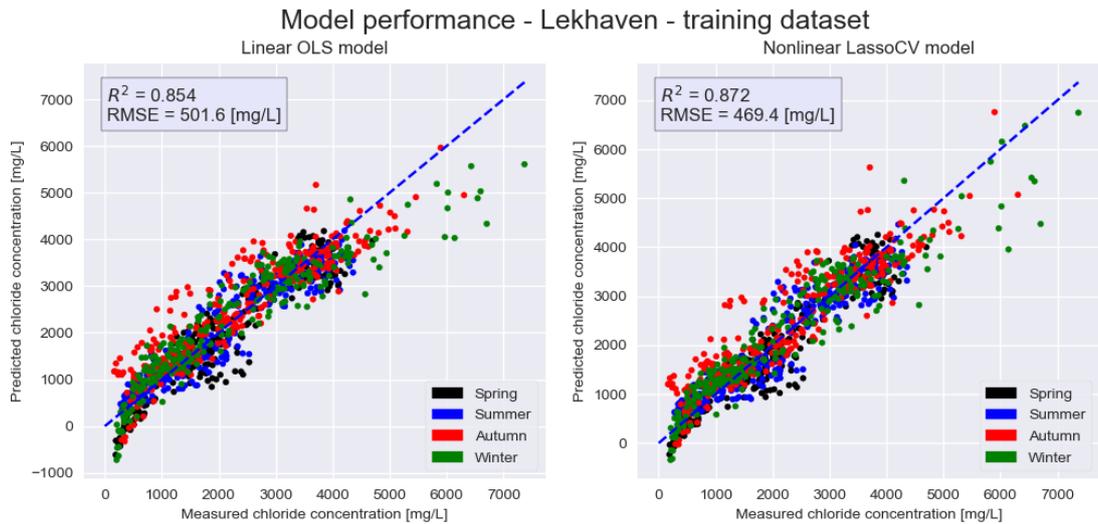


Figure 29. Measured and predicted chloride concentrations with linear OLS model and non-linear LassoCV including model performance parameters on training dataset.

Table 14. Performance of linear and non-linear model per season on training dataset.

Season	number of points	<i>linear OLS model</i>		<i>nonlinear LassoCV model</i>	
		R^2 [-]	RMSE [mg/L]	R^2 [-]	RMSE [mg/L]
Spring	254	0.858	446.7	0.882	407.0
Summer	274	0.878	407.2	0.878	406.4
Autumn	313	0.821	563.6	0.839	534.2
Winter	252	0.864	561.8	0.891	502.7
All	1093	0.854	501.6	0.872	469.4

Performance of both models is further specified per season (Table 14). Regarding the linear OLS model, variance of chloride concentrations at Lekhaven can best be explained during summer ($R^2 = 0.878$), in which also the standard deviation of the error is lowest (RMSE = 406.9 mg/L). During autumn the performance of the model decreases, the variance that can be explained by the independent input parameters decreases ($R^2 = 0.821$) and the standard deviation of the errors increase (RMSE = 563.6 mg/L).

Regarding the nonlinear LassoCV model, performance of the model is best during spring ($R^2 = 0.882$ and RMSE = 407.0 mg/L) and least optimal during autumn ($R^2 = 0.839$ and RMSE = 534.2 mg/L). Overall the performance of the nonlinear LassoCV model surpasses that of the linear OLS model. Both models score best during spring and summer. During autumn and winter performance of both models decreases, potentially due to seasonal fluctuations in water withdrawal and extreme weather frequency.

3.2.3 Performance on validation dataset

Performance of the model on the validation dataset shows a larger spread around the line of perfect fit (Figure 30, blue line) which is represented with an increase in the RMSE. Also, the predicted chloride concentrations are generally lower compared to the measured chloride concentrations. The linear model has an average prediction error of -288.0 mg/L and the non-linear model the average prediction error is -264.4 mg/L.

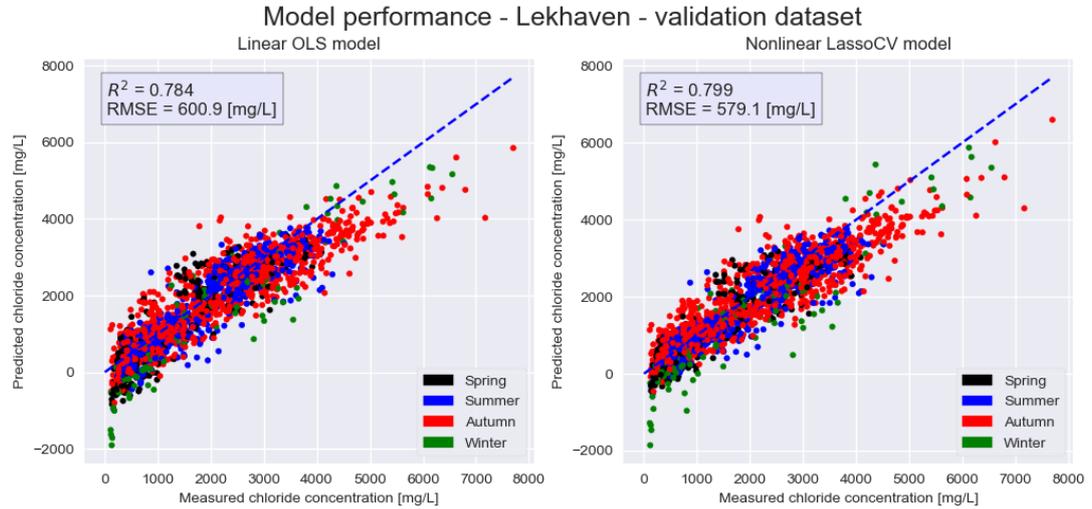


Figure 30. Measured and predicted chloride concentrations with linear OLS model and nonlinear LassoCV including model performance parameters on validation dataset.

Proportion of variance explained (R^2) by the model in the validation dataset is quite constant over all seasons regarding the linear model (Table 15). The estimation error (RMSE) increases drastically for datapoints in autumn and winter. The non-linear model performs best for datapoints in spring, datapoints collected in winter again show a substantial increase in estimation error.

Table 15. Performance of linear and nonlinear model per season on validation dataset

Season	Number of points	<i>linear OLS model</i>		<i>nonlinear LassoCV model</i>	
		R^2 [-]	RMSE [mg/L]	R^2 [-]	RMSE [mg/L]
Spring	296	0.717	549.9	0.773	492.2
Summer	544	0.787	503.7	0.810	476.2
Autumn	894	0.786	640.4	0.798	622.6
Winter	123	0.749	787.8	0.733	812.3
All	1857	0.784	600.9	0.799	579.1

3.2.4 Sensitivity analysis

Sensitivity of the linear OLS model is fully described by the first order sensitivity indices, which is to be expected as the model does not make use of interdependent parameters (Figure 31Figure 32). The linear OLS model is most sensitive to changes in the discharge of the Waal, followed by the MWA of the wind setup and astronomical tide.

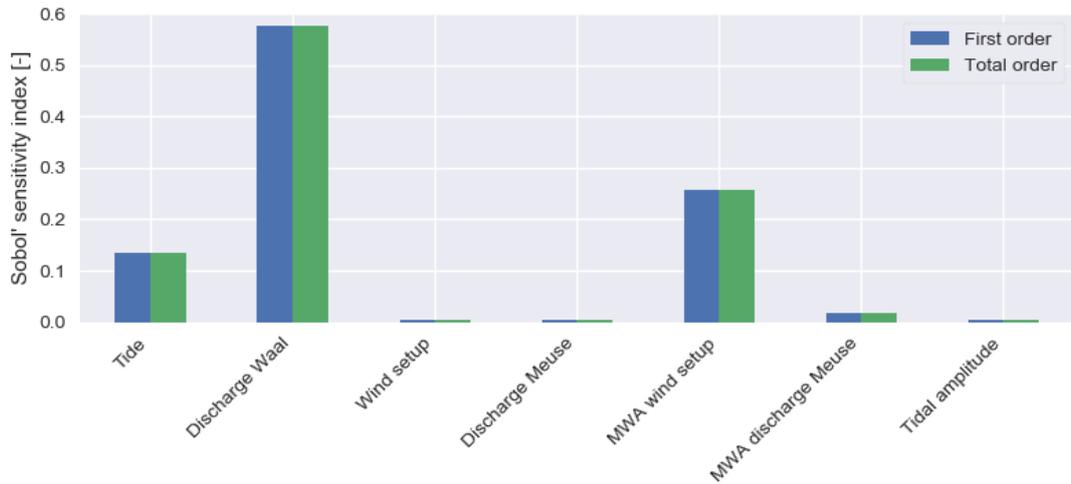


Figure 31. Linear OLS model sensitivity to each parameter visualized as first and total order sensitivity indices.

The nonlinear LassoCV model is most sensitive to changes in discharge of the Waal, followed by changes in the MWA of wind setup, astronomical tide and wind setup (Figure 32). Compared to the linear model, the nonlinear model for most parameters a difference between the first and total order sensitivity indices is present. The difference is caused by interdependence between variables, incorporated as cross-products of input parameters, and non-linear weighting.

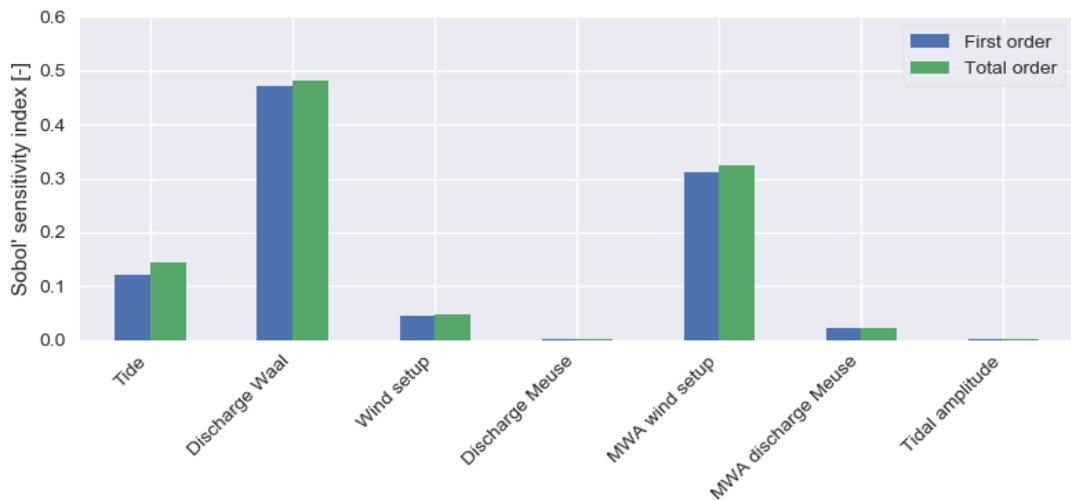


Figure 32. Nonlinear LassoCV model sensitivity to each parameter visualized as first and total order sensitivity indices.

3.2.5 Uncertainty analysis

Plotting the residuals of the non-linear LassoCV model over time during the first training period (Figure 33) shows multiple periods of underestimation of the chloride concentrations (i.e. negative residuals). During these periods, the boundary conditions used to predict chloride concentrations at Lekhaven show no extreme values (Figure 33, bottom plot), which could indicate unexpected observed chloride concentrations. Therefore, this is possibly caused by external factors affecting these boundary conditions, or influencing parameters which have not been incorporated in this analysis. The analysis period in Figure 33 is renowned for the precipitation shortage (KNMI, 2011), which potentially caused increased

water withdrawal from the area between the discharge measurement locations Hagestein and Lekhaven.

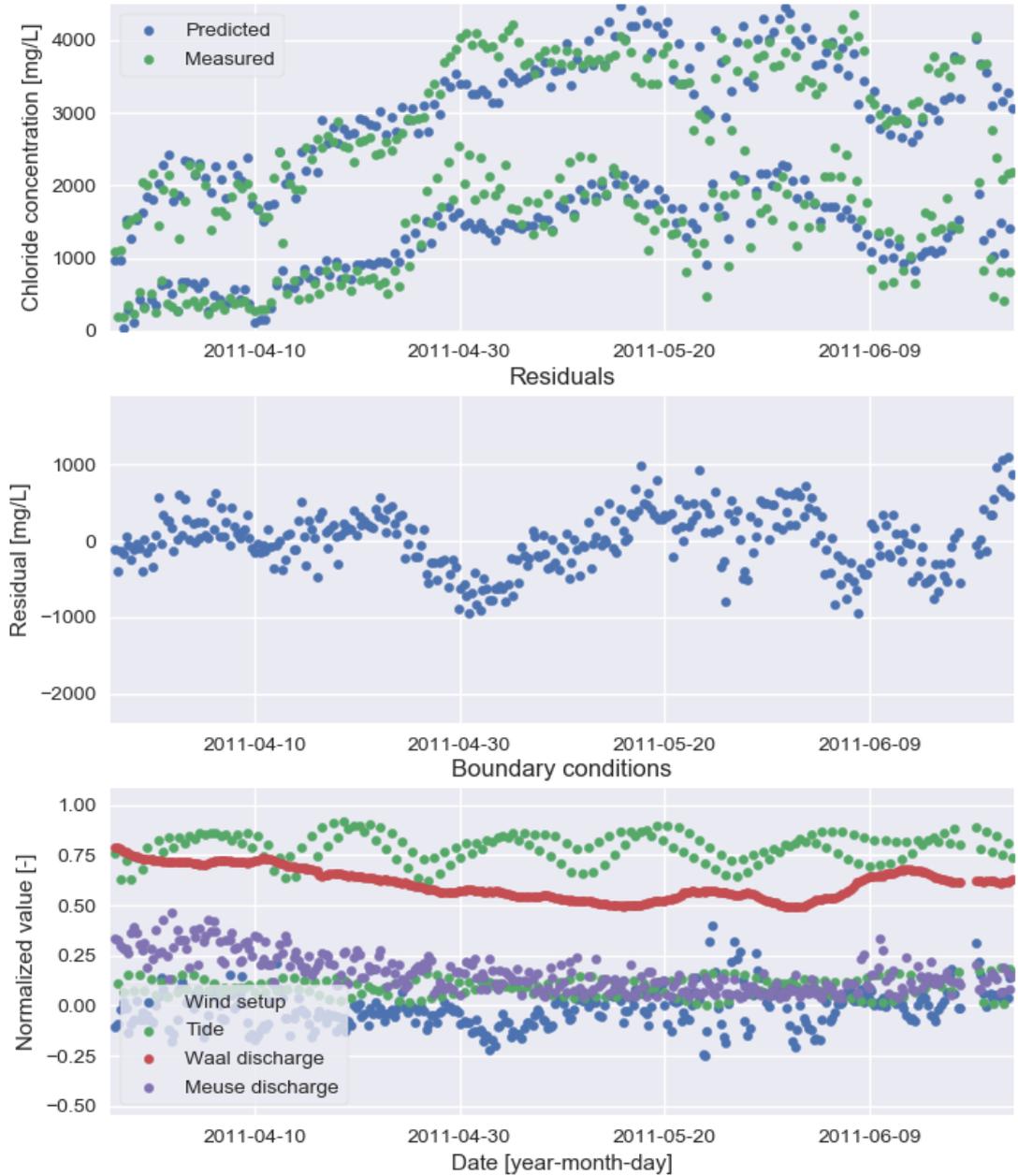


Figure 33. Predicted and measured chloride concentrations, residuals and normalized input independent boundary condition of nonlinear LassoCV model during first training period (2011).

Upon further inspection it turns out that during the period of model underestimation, the Boezemgemaal Gouda withdraws water for several consecutive sampling periods (Figure 34).

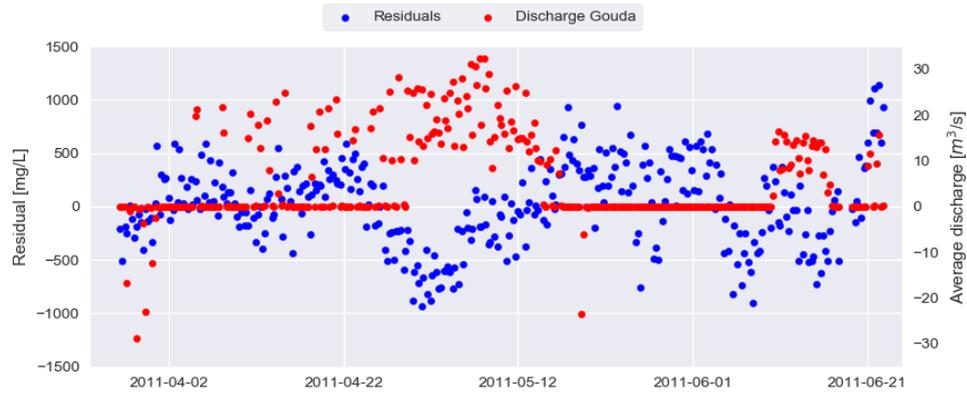


Figure 34. Residuals from nonlinear LassoCV model versus average discharge of Boezemgemaal Gouda per sampling period during the first training period (2011).

The effect of the discharge from Boezemgemaal Gouda on residuals is analysed by adjusting the discharge input of the Waal with the observed withdrawal (positive values) or discharge (negative values) of the pumping station. The influence of Boezemgemaal Gouda on the non-linear model is dependent on whether this occurs during high tide or low tide due to higher order dependencies, as indicated in the sensitivity analysis (Figure 32), which results in two ‘lines’ of change in residual regarding the nonlinear model (Figure 35). The effect on residuals of the linear model is very similar to that of the nonlinear model.

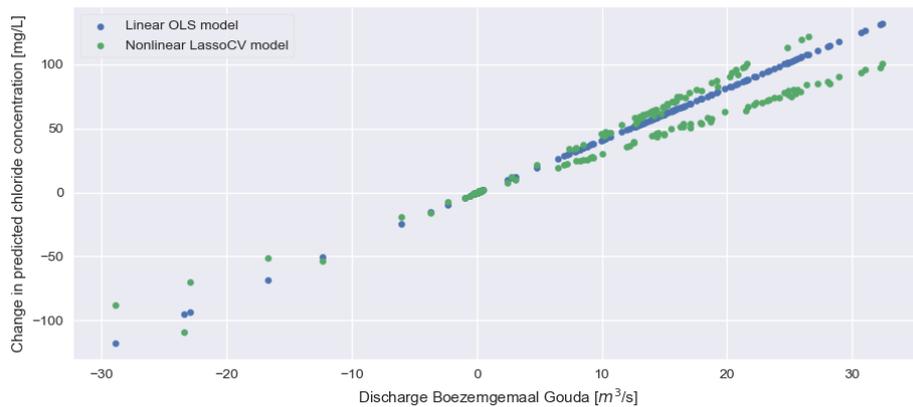


Figure 35. Influence of water withdrawal and discharge of Boezemgemaal Gouda on residuals of the linear and nonlinear model.

Extrapolation of the results obtained from this analysis indicates that an underestimation of 1000 mg/L, the maximum underestimation during the spring of 2011, would indicate a potential water withdrawal of approximately 300 m³/s. This seems unrealistic as the discharge of the Waal in this period is between 750 and 1000 m³/s (Figure 11).

3.2.6 Model selection

Overall the non-linear LassoCV model performs better on the training dataset (Table 14) in all seasons. On the validation dataset, the non-linear model performs slightly less during winter, however, performs better during all other seasons (Table 15). Furthermore, the non-linear model only differs from the linear model due to the addition of the parameters ‘Smootherstep wind setup’ and ‘Smootherstep MWA wind setup’ (Table 13), which does not increase the model complexity much. Therefore, for further application of predicting chloride concentrations at Lekhaven, only the non-linear LassoCV model is applied. Parameter coefficients are provided in Annex B.

3.3 Spijkenisserbrug prediction model

For development of a prediction model for chloride concentrations at Spijkenisserbrug a similar analysis as Lekhaven is performed. However, Spijkenisserbrug differs from Lekhaven in one major characteristic: at Spijkenisserbrug the chloride concentration reduces to the background concentration of the Rhine during almost every low tide (Figure 11), resulting in little variation in the low tide chloride concentrations. Hence, the observed and predicted chloride concentrations during low tide are much smaller in relation to high tide: 100 mg/L and 2000 - 10000 mg/L, respectively.

3.3.1 Parameter selection with multi-step analysis

Linear OLS model

Five parameters do not meet the selection criteria as stated in Section 3.1.4, which are; 'discharge Lek', 'MWA discharge Meuse', 'Tidal amplitude', 'MWA discharge Lek' and 'MWA tide' (Figure 36). These parameters are therefore disregarded for further analysis.

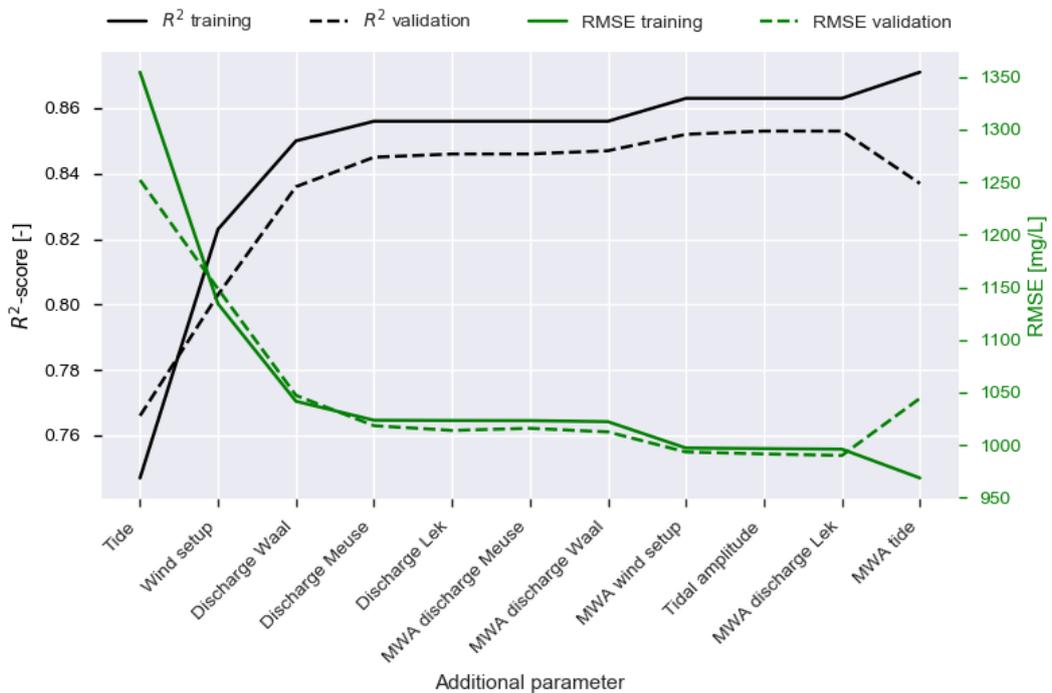


Figure 36. Results of multi-step analysis for parameter selection of the linear OLS model regarding measurement location Spijkenisserbrug when only using high tide observations.

Nonlinear LassoCV model

Regarding the non-linear LassoCV model, 'Discharge Lek', 'MWA discharge Lek', 'MWA tide' and the non-linear weighted parameters 'Smootherstep wind setup', 'Smootherstep MWA wind setup' and 'Smootherstep MWA tide' do not meet the selection criteria (Figure 37). These parameters are disregarded for further analysis.

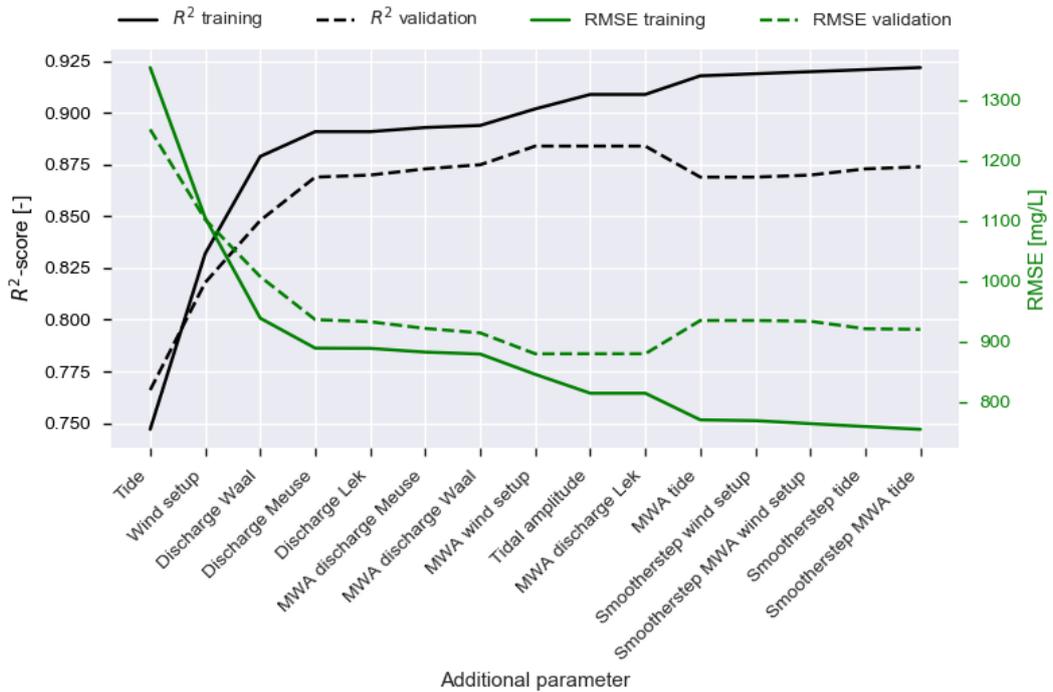


Figure 37. Results of multi-step analysis for parameter selection of the nonlinear LassoCV model regarding measurement location Spijkenisserbrug when only using high tide observations.

Table 16 shows the selected parameters for predicting chloride concentrations at Spijkenisserbrug regarding the linear and non-linear model, which are determined with the multi-step analysis.

Table 16. Parameters selected for computation of the linear OLS model and the nonlinear LassoCV model with the use of the multi-step analysis for location Spijkenisserbrug.

Parameter name	linear OLS model	nonlinear LassoCV model
Astronomical Tide	x	x
Discharge Waal	x	x
Wind setup	x	x
Discharge Lek		
Discharge Meuse	x	x
MWA Discharge Waal	x	x
MWA Discharge Meuse		x
MWA Discharge Lek		
MWA Wind setup	x	x
MWA Tide		
Tidal amplitude		x
Smootherstep tide		x
Smootherstep Wind setup		
Smootherstep MWA wind setup		
Smootherstep MWA tide		

3.3.2 Performance on training dataset

The linear OLS and nonlinear LassoCV model are trained based on the training dataset of Spijkenisserbrug chloride concentration observations and parameters as selected with the use of the multi-step analysis (Table 16) (Figure 38) (Figure 29). An overview of the corresponding parameter coefficients can be found in Annex C.

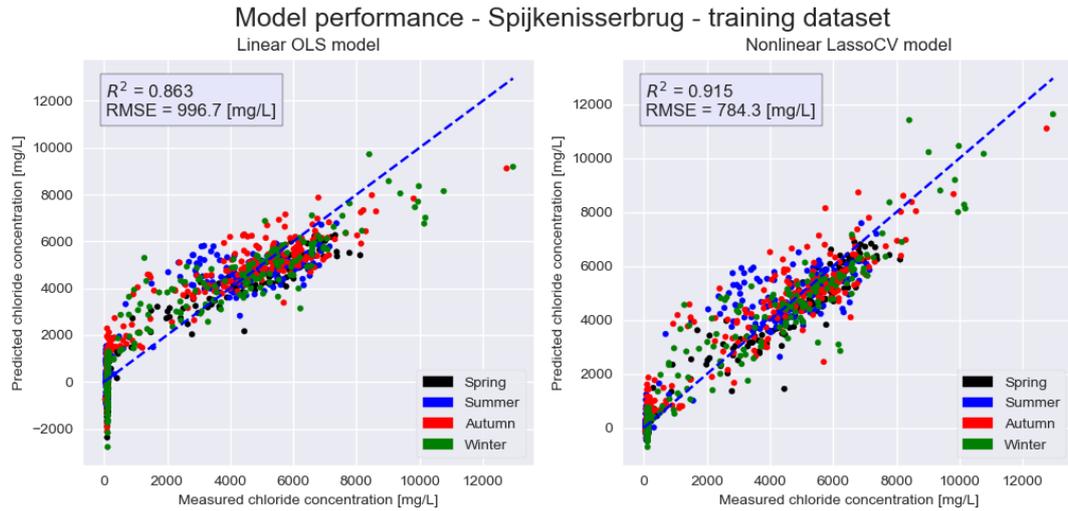


Figure 38. Measured and predicted chloride concentrations with linear OLS model and nonlinear LassoCV including model performance parameters on training dataset for location Spijkenisserbrug with only high tide observations.

Overall performance of the non-linear model exceeds the linear model performance (Figure 38). When specifying model performance per season (Table 17), the non-linear LassoCV model also outperforms the linear OLS model each season.

Table 17. Performance of Spijkenisserbrug linear and nonlinear model per season on training dataset.

Season	number of points	<i>linear OLS model</i>		<i>nonlinear LassoCV model</i>	
		R ² [-]	RMSE [mg/L]	R ² [-]	RMSE [mg/L]
Spring	255	0.889	877.2	0.953	573.3
Summer	274	0.854	940.3	0.884	839.9
Autumn	312	0.861	1025.5	0.910	823.5
Winter	238	0.849	1132.6	0.913	857.7
All	1079	0.863	996.7	0.915	784.3

3.3.3 Performance on validation dataset

Performance of the model on the validation dataset shows a larger spread around the line of perfect fit (Figure 39) which is represented with an increase in the RMSE.

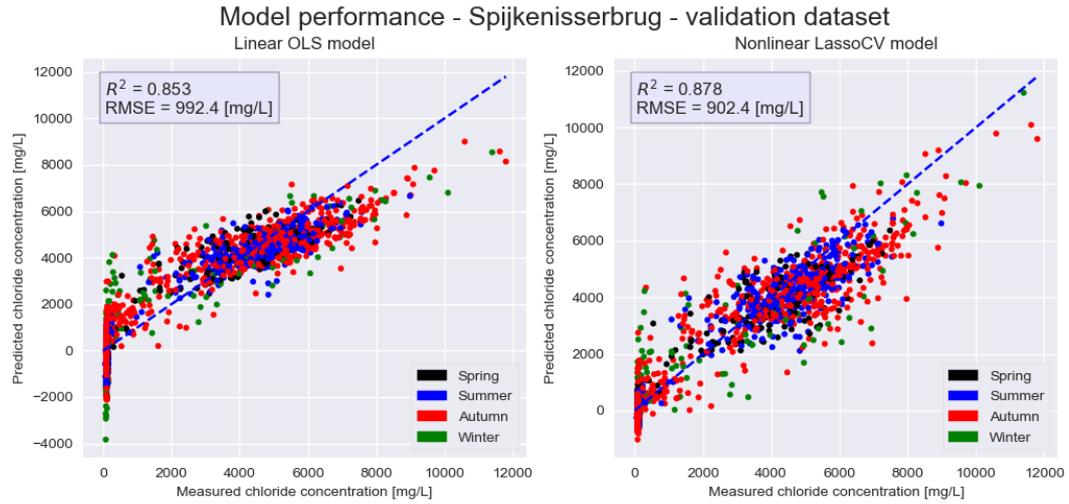


Figure 39. Measured and predicted chloride concentrations with linear OLS model and nonlinear LassoCV including model performance parameters on validation dataset for location Spijkenisserbrug.

The overall performance of the non-linear LassoCV model exceeds that of the linear OLS model on the validation dataset (Figure 39). If model performance is specified per season (Table 18), performance of the non-linear model, again, outperforms the linear model.

Table 18. Performance of Spijkenisserbrug linear and nonlinear model per season on validation dataset.

Season	number of points	<i>linear OLS model</i>		<i>nonlinear LassoCV model</i>	
		R ² [-]	RMSE [mg/L]	R ² [-]	RMSE [mg/L]
Spring	296	0.847	886.6	0.894	735.8
Summer	510	0.891	797.8	0.914	710.7
Autumn	846	0.864	1010.4	0.877	959.2
Winter	120	0.656	1643.4	0.738	1434.2
All	1772	0.853	992.4	0.878	902.4

3.3.4 Sensitivity analysis

The linear OLS model is most sensitive to variation in the discharge of the Waal, followed by variation in wind setup and tide (Figure 40).

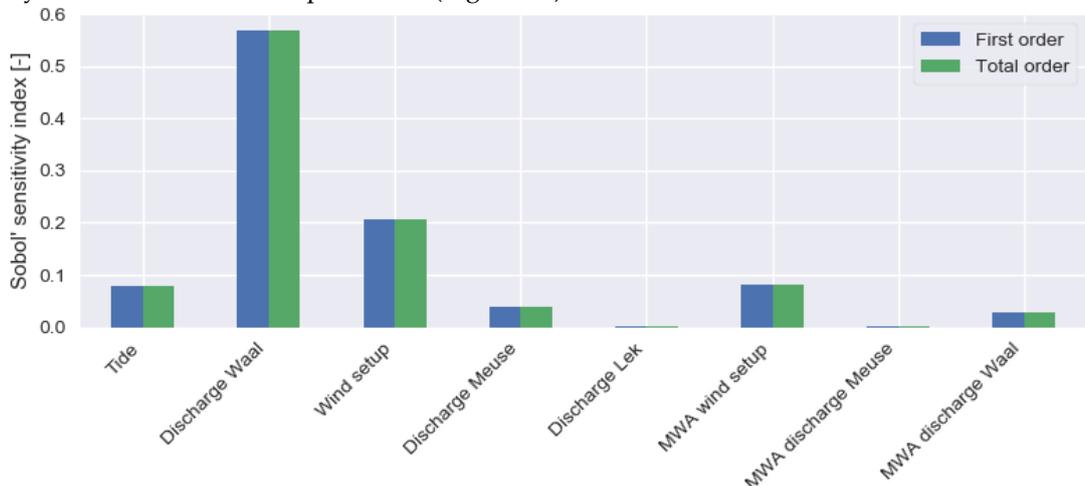


Figure 40. Spijkenisserbrug linear OLS model sensitivity to each parameter visualized as first and total order sensitivity indices.

The nonlinear LassoCV model is most sensitive to variation of Waal discharge, closely followed by variation in wind setup (Figure 41). The nonlinear model is much more sensitive to variation in wind setup, compared to the linear model. Also, the sensitivity regarding Waal discharge shows a large decrease, while simultaneously the sensitivity to the MWA discharge Waal shows a large increase.

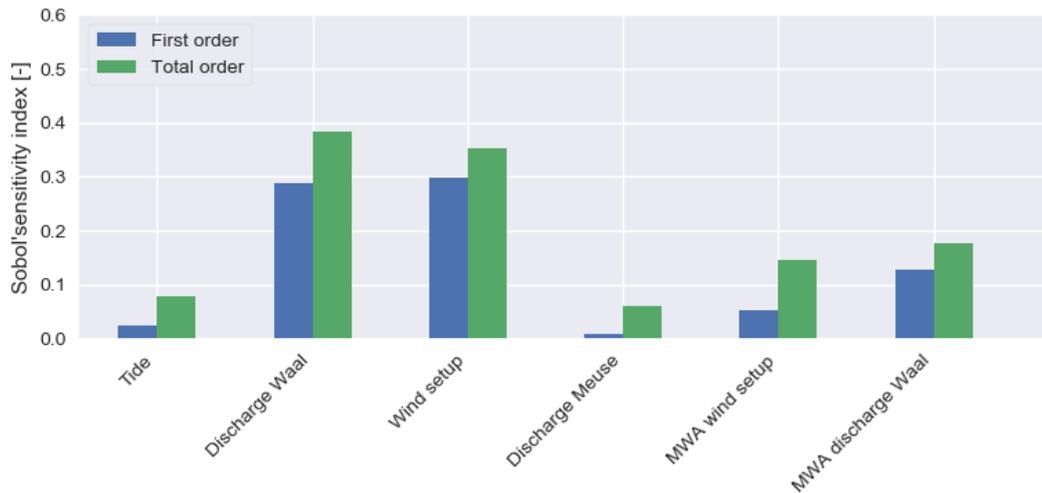


Figure 41. Spijkenisserbrug nonlinear LassoCV model sensitivity to each parameter visualized as first and total order sensitivity indices.

3.3.5 Model selection

Overall the non-linear LassoCV model for predicting the chlorinity at Spijkenisserbrug performs better on the training dataset than the linear model (Figure 20Table 14) in all seasons. On the validation dataset, the non-linear model also performs better during all seasons (Figure 22). Furthermore, the non-linear model only differs from the linear model due to the addition of the parameters 'MWA discharge Meuse', 'Tidal amplitude' and 'Smootherstep tide' (Figure 19), which does not increase the model complexity much. Therefore, for further application of predicting chloride concentrations at Lekhaven, only the non-linear LassoCV model is applied. Parameter coefficients are provided in Annex C.

3.4 Brienoordbrug and Beerenplaat prediction models

Equal to Lekhaven and Spijkenisserbrug, for measurement locations Brienoordbrug and Beerenplaat the non-linear LassoCV model outperforms the linear OLS model (Annex D and Annex E).

3.5 Summary

For comparison between individual models, the prediction error has been normalized by dividing the RMSE with the mean observed value (NRMSE). Similarly to the RMSE, a lower value for the NRMSE indicates a better model performance.

Highest factor of variance explained by the input parameters (R^2) regarding the validation dataset, on which the model type is selected, is found at the most downstream measurement locations Spijkenisserbrug and Lekhaven (Table 19), on the Old Meuse and New Meuse respectively. Further upstream at Beerenplaat and Brienoordbrug, where the influence of the tidal wave decreases, the R^2 -values decrease and the estimation errors (NRMSE) increase.

Table 19. Selected model type regarding each measurement location and corresponding model performance on validation and training dataset.

Measurement location	Model	Training dataset			Validation dataset		
		R^2 [-]	RMSE [mg/L]	NRMSE [-]	R^2 [-]	RMSE [mg/L]	NRMSE [-]
Lekhaven	non-linear	0.872	469.4	0.21	0.799	579.1	0.31
Spijkenisserbrug	non-linear	0.915	784.3	0.30	0.878	902.4	0.37
Brienoordbrug	non-linear	0.813	370.5	0.50	0.699	413.1	0.69
Beerenplaat	non-linear	0.725	541.1	1.07	0.620	531.2	1.42

The inclusion of non-linear parameters and the inclusion of cross-products improves the explained variance in chloride concentrations by variance in the input parameters, indicated with an increase in R^2 -value. Also, the estimation error (RMSE) decreases with application of the non-linear LassoCV model. Therefore, chloride concentrations in the Port of Rotterdam are best explained with a non-linear model, as described in Section 3.1.

4 Model application



This chapter elaborates on potential applications of the developed prediction models (Chapter 3), that have been developed in this research. Two applications of the models can be indicated; impact of human intervention and chloride concentration prediction.

Assessing impact of human intervention

The main application of the models is the analysis of the effect of human interventions on chloride concentrations in the Port of Rotterdam. This application is based on the comparison of predicted chloride concentrations, computed with the developed models, with the measured chloride concentrations after a human intervention. The analysis of effects of human interventions on chloride concentrations can be performed according to the flow chart as provided in Figure 42. This flow chart can be applied once the prediction model has been trained such as described in Chapter 3. Step 1 describes the computation of chloride concentrations for the reference period. Step 2 describes the computation of expected chloride concentrations using the model that was trained on the period before human interventions, with use of the input of boundary conditions after human intervention. Step 3 holds the computation of residuals before and after human intervention and the analysis of these residuals, which is further elaborated on in Section 4.1. An example of the analysis of the impact of potential human interventions is provided in Section 4.3.

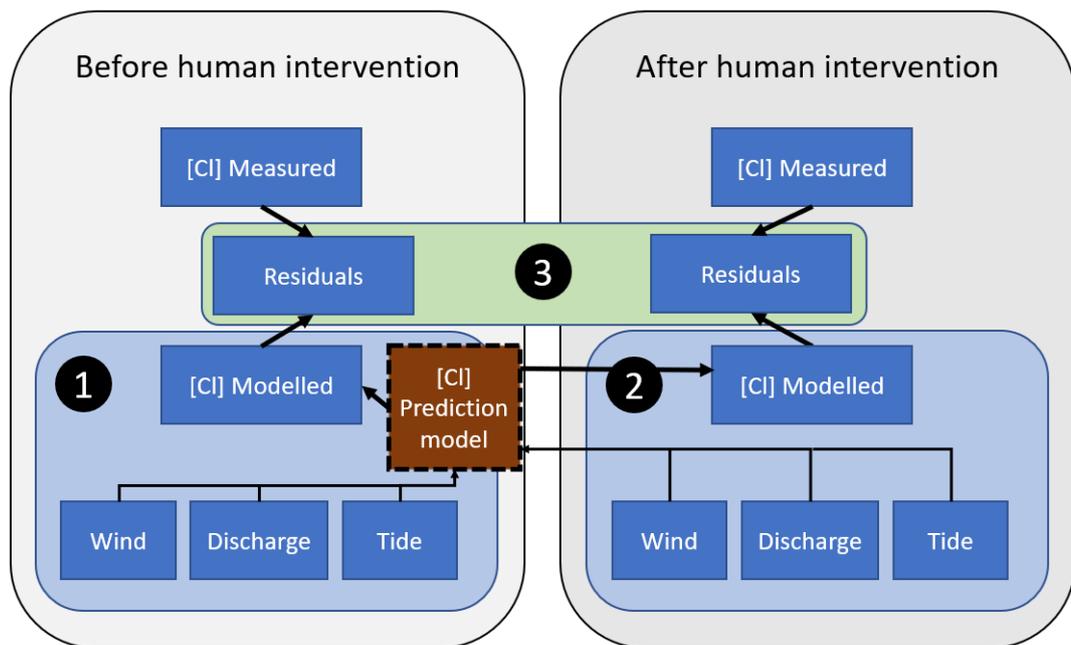


Figure 42. Flow chart for analysing effects of human intervention on chloride concentrations [Cl]. With the trained prediction model indicated in brown.

Effect of human intervention on chloride concentrations can be further examined by comparison between the model sensitivity before and after human intervention. First, the model is trained on data before human intervention (Figure 43, Step 1). Followed by the training of an alternative model trained with data obtained after human intervention. This training is performed with the same set of input parameters as used before human intervention (Figure 43, Step 2). Next, the sensitivity indices before human intervention and after human intervention can be compared (Figure 43, Step 3). This provides information on which processes became more, or less, influential due to the human intervention.

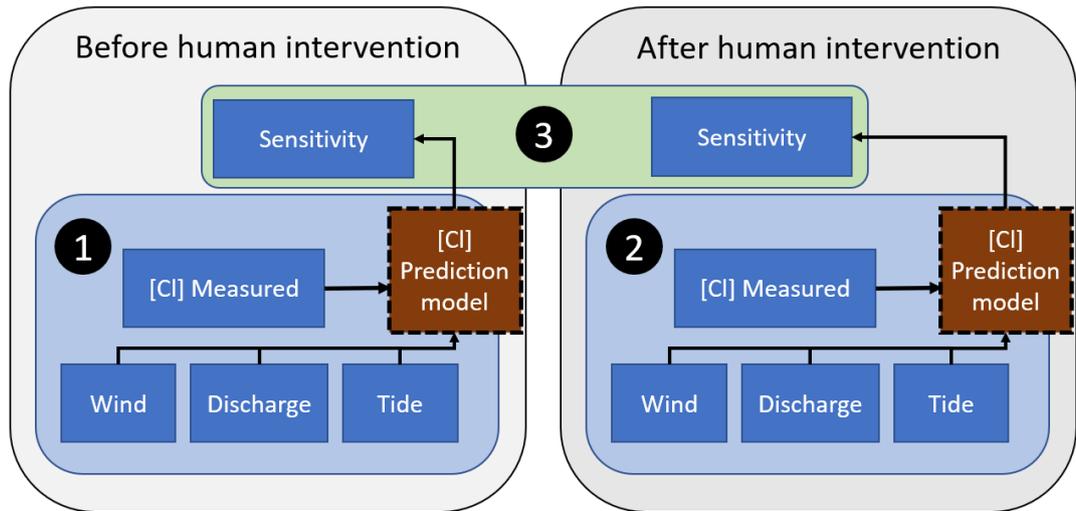


Figure 43. Flow chart for analysis of changes in model sensitivity due to human intervention.

Chloride concentration prediction

Another application of the developed chloride concentration prediction model is the prediction of chloride concentrations with the use of expected values of the input boundary conditions (Figure 44).

The discharge boundary condition is currently predicted 4 days in advance. Astronomical tide is predicted several months in advance. Step 1 contains the prediction of chloride concentrations with the use of a trained prediction model. Step 2 contains the validation of the prediction which can be performed as a hindcast for model improvement.

These predictions, which are currently not performed, can be useful to water managers, as this information can potentially be used for planning of water extraction by agriculture or industry.

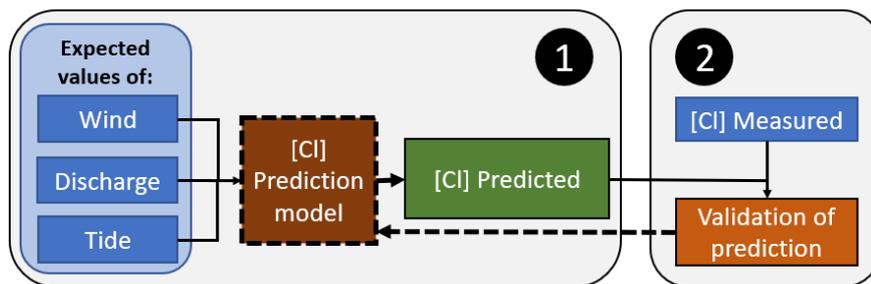


Figure 44. Flow chart for predicting chloride concentrations based on predictions of boundary conditions.

4.1 Methodology for analysis of residuals

For future application of the analytical prediction models to analyse the impact of human interventions on chloride concentrations in the Port of Rotterdam, a methodology is determined for analysis of residuals. This methodology may be applied to the residuals collected during high and low tide separately or combined. However, especially for measurement locations: Spijkenisserbrug, Beerenplaat and Brienoordbrug, a separate

analysis on residuals would be preferable due to the decrease of the chloride concentration to Rhine background concentrations during low tide.

For increased validity of residuals calculated after human intervention, the application of the prediction model must be performed within the range of the input parameters before human intervention (Table 20).

Table 20. Range of input parameters prior to deepening for which the model has been trained and validated.

Parameter	Range	Unit
astronomical tide	-91 to 160	[cm NAP]
wind setup	-90 to 122	[cm]
discharge Waal	670 to 1250	[m ³ /s]
discharge Meuse	13 to 522	[m ³ /s]
discharge Lek	0 to 174	[m ³ /s]

Statistical testing for difference in means

Previous analysis of HydroLogic (2018) applied the z -score test for testing for equal means (Eq. 4).

$$z_{score} = \frac{X - \mu}{\sigma} \quad (\text{Eq. 4})$$

where X is the mean residual prior to human intervention, μ is the mean residual post human intervention and σ is the standard deviation prior to human intervention. The z -score is used to test the hypothesis:

$$\begin{aligned} H_0: & \text{Mean residual post human intervention} = \text{Mean residual before human intervention} \\ H_1: & \text{Mean residual post human intervention} \neq \text{Mean residual before human intervention} \end{aligned}$$

The z -score for different means is a quick tool for assessing changes in the mean residual. However, it does not consider the number of observations. The more advanced Welch's t -test (Eq. 5) does take the number of observations in the datasets into account (Welch, 1947). For testing of equal means, as stated in H_0 , a two-tailed test is applied, of which the critical score is determined based on the significance level (α) and the degrees of freedom (ν), which can be estimated with Eq. 6. The Welch's t -test can handle equal and unequal variances and sample sizes, providing a uniform testing methodology (Zimmerman, 2004).

$$t_{score} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \quad (\text{Eq. 5})$$

where \bar{X}_1 is the mean of the residuals prior to human intervention, \bar{X}_2 is the mean of the residuals post human intervention, N_1 is the number of independent observations prior to human intervention, N_2 is the number of independent observation post human intervention, s_1^2 is the variance of residuals prior to human intervention and s_2^2 is the variance of residuals post human intervention. The degrees of freedom needed for determining the critical t -score is determined with (Derrick & White, 2016):

$$v \approx \frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{S_1^4}{N_1^2 v_1} + \frac{S_2^4}{N_2^2 v_2}} \quad (\text{Eq. 6})$$

where $v_1 = N_1 - 1$ and $v_2 = N_2 - 1$.

4.2 Example on synthetic timeseries

All dredging work for deepening of the New Waterway has been completed in January 2019. Around 10 percent of the total dredged volume (6 million m³) has been used to fill erosion pits in the Old Meuse. Between measurement location Lekhaven and the divergence point of the New and Old Meuse, dredging has not been completed yet due to the presence of pipes and cables in the riverbed (Figure 45). This final part of the deepening project is scheduled in the fourth quarter of 2019. The presence of these pipes and cables have an unknown influence on the final effect of deepening of the New Waterway and Botlek on chloride concentrations as they currently form a sill with a height of approximately 1.5 meter.

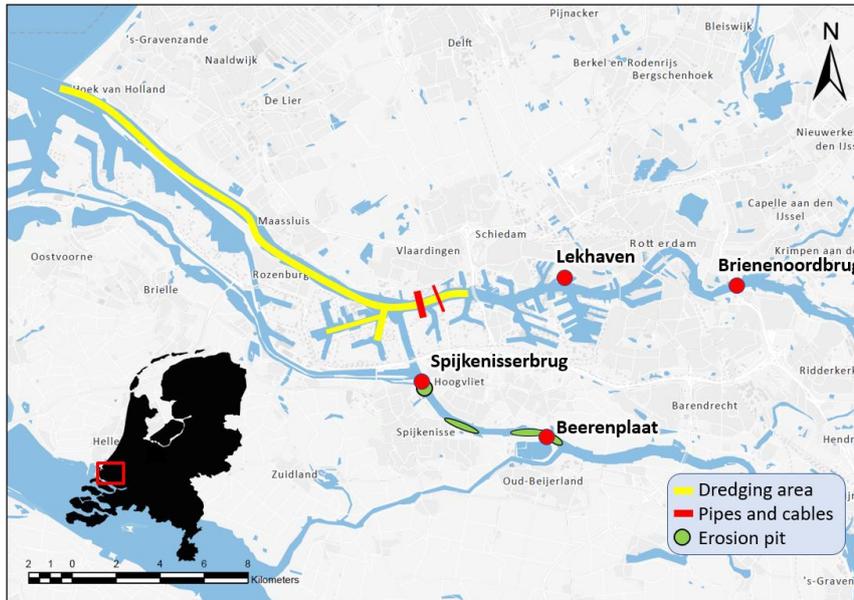


Figure 45. Deepening of the New Waterway and Botlek and location of erosion pits and pipes and cables.

Synthetic time series

As the deepening of the New Waterway and Botlek has not been fully completed yet, an example analysis on effects of human interventions on chloride concentrations is performed on a synthetic period of low discharges. The corresponding measured chloride concentrations for measurement location Lekhaven have been altered based on the outcome of previous research performed by HydroLogic (2018). This analysis by HydroLogic, based on calculations performed by Svasek Hydraulics (2015), resulted in a percentual change in chloride concentrations with a Rhine discharge dependency (Figure 46).

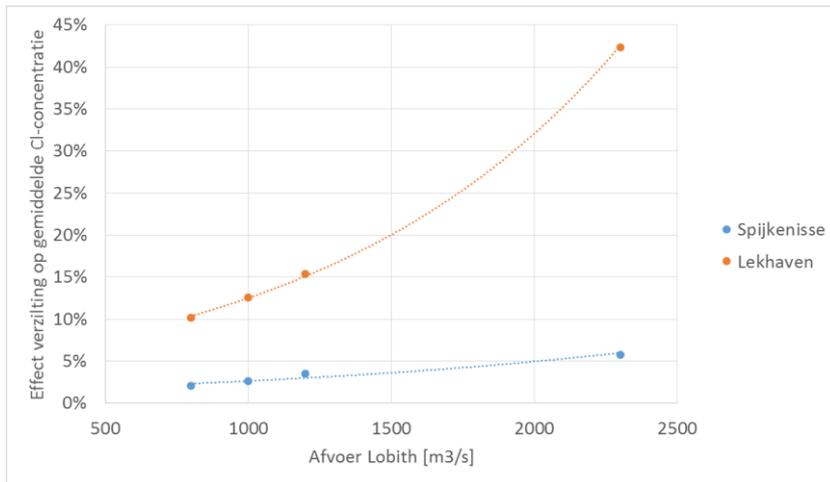


Figure 46. Expected relative change at measurement locations Lekhaven and Spijkenisserbrug with a discharge dependency (HydroLogic, 2018)

The measured chloride concentrations (Figure 47, green line) correspond with the “[Cl] measured” boxes in Figure 42. The measured chloride concentrations have been altered (Figure 47, blue line) based on the percentual change as provided in Figure 46, and correspond with the “[Cl] measured” box in Step 2 of Figure 42. With the use of the Lekhaven non-linear prediction model chloride concentrations have been predicted based on discharges of the Waal and Meuse, wind setup and astronomical tide (Figure 47, bottom plot).

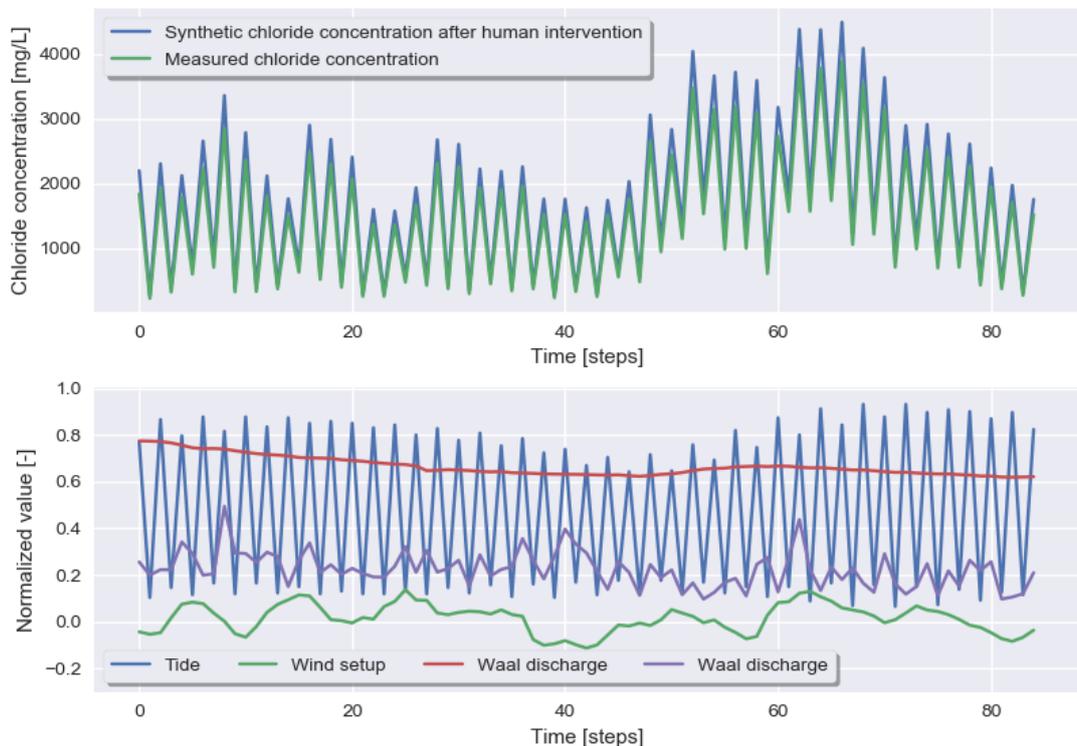


Figure 47. Measured chloride concentrations (top plot, green line), synthetic chloride concentrations based on HydroLogic (2018) (top plot, blue line) and overview of normalized input boundary conditions (bottom plot).

Figure 48 shows the synthetic chloride concentrations versus the predicted chloride concentrations, determined with the use of the Lekhaven prediction model. The trend of the observations is indicated with the black dotted line. An increase in trend corresponds with

higher synthetic chloride concentrations, compared to the expected chloride concentrations. This suggests a change in chloride concentrations, which is statistically tested by comparing residuals obtained from the synthetic dataset and the predicted dataset, relative to the reference situation (Figure 42).

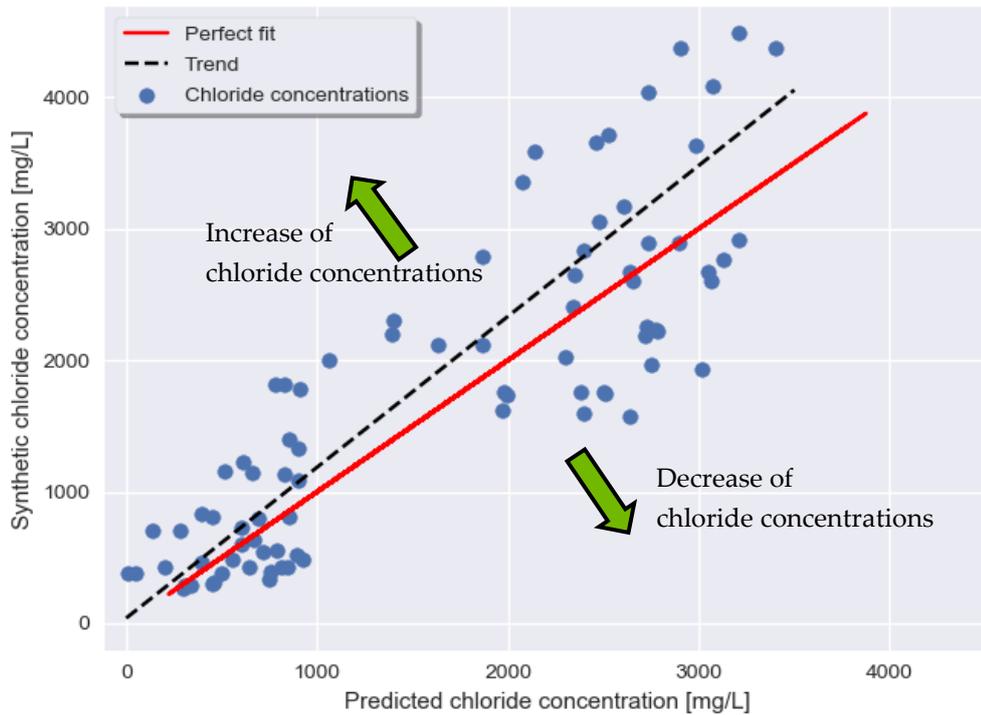


Figure 48. Scatter plot of predicted and synthetic chloride concentrations for example period. Change in chloride concentrations can be indicated by shift in trend line relative to the perfect fit line.

A summary of statistics of the residuals is provided in Table 21. A negative mean of the synthetic residuals is observed, indicating an underestimation of the Lekhaven prediction model caused by an increase in chloride concentrations.

Table 21. Number of observations, means and standard deviations of measured and synthetic residuals.

	Observations	Mean [mg/L]	Standard deviation [mg/L]
Measured residuals	85	71.8	549.8
Synthetic residuals	85	-168.2	608.4

Statistical testing for difference in means

The statistical test, as elaborated on in Section 4.1, is performed on the residual’s statistics provided in Table 21. Result the Welch’s t-test, given in Equation 5, is provided in Table 22.

Table 22. Outcome of statistical test between synthetic and expected residuals.

	Synthetic series
Degrees of freedom (df)	166
Significance level (α)	0.05
critical t_{score}	± 1.984
computed t_{score}	2.70
H_0	rejected

The statistical test indicates that residuals from the predicted and altered time series cannot be regarded as equal (computed $t_{score} >$ critical t_{score}). This indicates a significant change in chloride concentrations due to the alteration of the time series. This is in line with the conclusion from the analysis of HydroLogic (2018).

5 Discussion

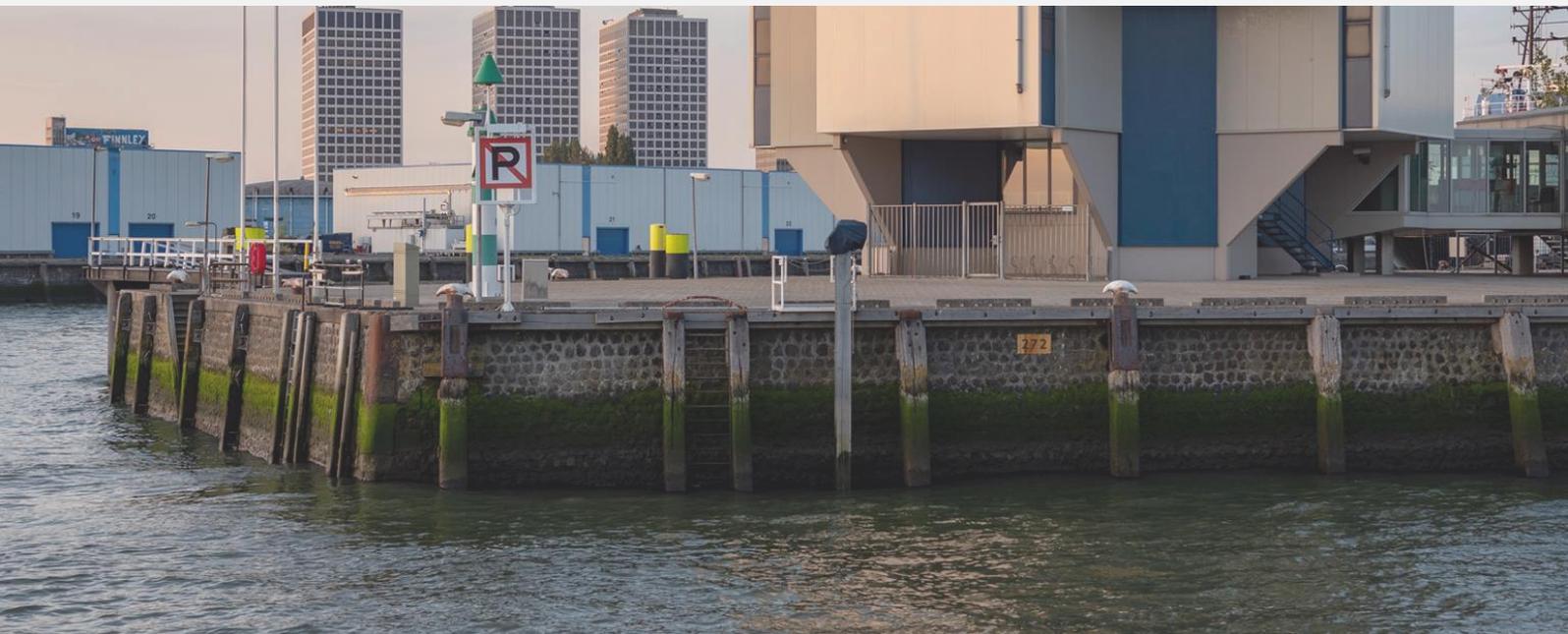


image: Lekhaven, chloride concentration measurement location / Erik Plaggenmars

5.1 Gained insight on salt intrusion in the Port of Rotterdam

Previous analysis by HydroLogic on salt intrusion in the Port of Rotterdam focused on predicting chloride concentrations at Lekhaven and Spijkenisserbrug with day-averaged discharges of the Rhine at Lobith below a wind setup of 0.15 meter at Hoek van Holland. A time lag of one day was applied to the Rhine discharge (HydroLogic, 2015a). With the use of the cross-correlation analyses, as described in Section 2.4, time lags of the water level at Hoek van Holland and the discharges of the Waal, Lek and Meuse at Tiel, Hagestein and Megen, respectively, are determined at a 10 minute resolution, regarding each of the four measurement locations.

The reduction of the sampling interval, from one day to the frequency of half a tidal cycle (e.g. 6.2 hours), facilitates the incorporation of the tidal water level variation of each ebb and flood period. This water level variation is split in the components astronomical tide and wind setup. Deviating from previous analyses, all occurring values of wind setup, within the analysis period are now incorporated in the analysis. This incorporation of tide and wind setup proves to be efficient as, for example at Spijkenisserbrug, variation in chloride concentration is much more sensitive to variation in wind setup compared to variation in tide (Figure 41).

With the auto-correlation analysis, as described in Section 3.1.3, the auto-correlation in each of the five boundary conditions, astronomical tide, wind setup and discharges of the Waal, Lek and Meuse, are incorporated mean-weighted averages in the analysis in order to describe chloride concentration observations independent from previous observations. At Lekhaven, this results in a high sensitivity of the non-linear model to the autocorrelation parameter of the wind setup (Figure 32, MWA wind setup).

Application of a multi-step analysis, for selecting parameters when describing chloride concentrations at each of the four measurement locations, provides insight in the added value of each parameter. For example, the discharge of the Lek is not included in any of the models, as it does not increase the ability of the models to predict chloride concentrations. Also, with the use of the validation dataset in the multi-step analysis parameters are excluded that would have been included in the analysis when only using the training dataset as an indicator. An example is the exclusion of MWA tide in the Lekhaven prediction model (Figure 28). Here, the performance of the model on the training dataset increases with inclusion of the parameter MWA tide, but it decreases the performance on the validation dataset.

The increased performance of the non-linear model compared to the linear model at each of the four measurement locations indicates that processes affecting salt intrusion are better described with the inclusion of cross products of the input parameters and the autocorrelation parameters (Chapter 3.4). At Lekhaven, the wind setup is no longer incorporated in the non-linear model as a single parameter (parameter coefficient = 0) but only in the cross product parameters (Annex B). Similar results are obtained from the parameters coefficients at the three other measurement locations (Annex C, Annex D and Annex E).

From the specification of model performance per season, variation in model performance is observed. Regarding Lekhaven and Spijkenisserbrug, performance is best during summer and least optimal during winter (Table 15 and Table 18). This may indicate a seasonal variation in the uncertainty of the input parameters, such as water withdrawal or discharge

from pumping stations, or in the influencing factors not incorporated in the analysis, such as the salt concentration of the intruding sea water (Chapter 2.1). Further analysis on seasonally fluctuating processes in the Rhine-Meuse delta or incorporation of (expected) seasonal water withdrawal or discharge may reduce the seasonal performance fluctuation of the developed models.

5.2 Impact of model assumptions

Section 2.1. shows the variation in chloride concentrations at several depths. Based on a visual interpretation of the salinity profiles (Figure 10), a depth-averaged value is applied. A slight difference in salt wedge profile is observed from the analysis of salt wedges corresponding with three discharge volumes. From theory we know, that the salinity profile, and thus the vertical variation of chloride concentrations, is dependent on the river discharge volume and tidal range (Savenije, 2012; Open University, 1999). As the tidal range will not differ much, the vertical variation of chloride concentrations is mainly influenced by the river discharge. However, application of a depth-averaged value, disregards the variation of the vertical chloride concentrations distribution.

Although the shape of the vertical variation does not differ much under various discharge conditions, a difference, regarding high and low tide, is observed between the measurement locations Lekhaven and Spijkenisserbrug (Figure 10). Measurement location Lekhaven is located inside a harbour basin which causes trapping of saline water at the bottom (Savenije, 2012). Also, it limits the influence of the fresh water to flush out the salt during low tide, causing increased chloride concentrations during low tide. Water movement inside the harbour basin is dependent on the flow velocity in the main river (Langedoen, 1992) and density currents due to salinity differences (De Nijs, 2012). At Lekhaven, chloride concentrations inside the harbour basin are dependent on the exchange of water within the basin with water in the main river under influence of tidal movement. At Spijkenisserbrug, which is situated in the main river, this dependency does not occur, and the measured chloride concentration is that of the water in the main river.

Although this distinct difference in dynamics of salt intrusion, due to the location of the chloride concentration measurement devices, the analytical models are very capable of predicting chloride concentrations at each of the measurement locations. Therefore, a depth-averaged chloride concentration proves to be a good indicator of salinity under various conditions and at different measurement locations.

Optimization of correlation between boundary conditions and chloride concentration only is performed on the dataset from spring 2011, as described in Section 2.2. This can possibly be improved by applying the training set used in model building in Section 3.1.1, as a wider range of events will be included.

The correlation method, time lags and sampling interval are determined based on a single period of low discharges in 2011 (Figure 11). During this analysis period, the influence of discharge of the Lek on chloride concentrations is insignificant, resulting in a wide range of time lags (Figure 18 and Figure 19) and no increase in the correlation coefficients when added to the discharge of the Waal (Table 2Table 6). Therefore, the time lag of the discharge of the Lek is estimated based on time lags of the Waal and Meuse. For exact determination

of the time lag of the Lek, an analysis period with reasonable discharge of the Lek will provide a more data-based time lag.

The determination of the training and validation dataset is done by selecting certain periods of low discharge, the hold-out method. Because of this methodology, the remaining observations assigned to the validation dataset are mostly obtained from shorter periods of drought (Figure 21). Influence of this hold-out method on model performance is unknown and can be examined with a comparison to a randomly selected training and validation dataset. Also, the size of both the training and validation dataset is user determined. This might influence the model performance and optimal computation, according to the model learning curve principle (Figure 49). A larger training dataset will, according to the principle, increase the model's goodness of fit. It should however be noted that with increasing training dataset, the size of the validation dataset decreases.

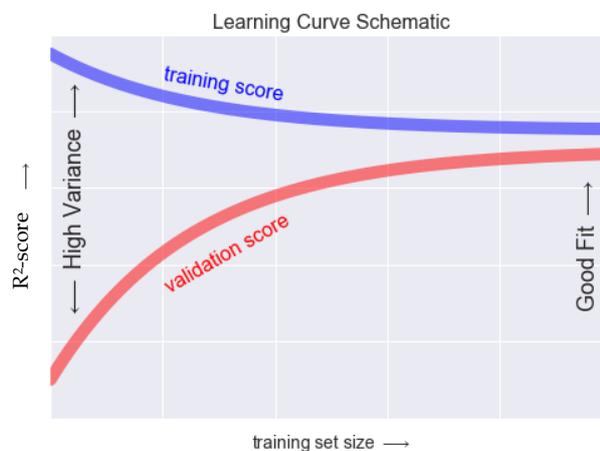


Figure 49. The learning curve principle. With increasing training set size the model validation score increases.

Normalization of parameters is all performed to values between 0 and 1, except for wind setup, which is normalized between -1 and 1. This increases the range for wind setup compared to the other parameters. The Sobol' sensitivity analysis compensates for varying ranges of input parameters when determining the sensitivity indices, as a range must be set for each parameter (Saltelli, et al., 2010). The difference in ranges is therefore not affecting the result of the sensitivity analysis. However, the parameter coefficients, as provided in Annex B to D, should not be compared directly as these are influenced by the range of the input parameters.

The application of non-linear parameters regarding the tide and wind setup is performed to incorporate mixing processes influenced by these parameters. These non-linear parameters however do not have a direct physical substantiation. This makes interpretation of effects of these parameters more difficult. Also, the shape of the applied smoother step function is arbitrary, as this is based on expectations of effects.

The discharge into the estuary of the Rhine-Meuse Delta is described with the inflow of freshwater from the Waal, Meuse and Lek. This discharge is however also influenced by the withdrawal and discharge of pumping stations along the rivers downstream of the discharge measurement locations. The withdrawal and discharge of these pumping stations

in not continuously measured during the analysis period described in Section 2.2. and therefore not incorporated in the present model. It was found that an underestimation of chloride concentrations at Lekhaven of 1000 mg/L would indicate a potential water withdrawal of 300 m³/s. Although this seems unrealistic, as the discharge ranged from 750 to 1000 m³/s during that period, under- and overestimation of the predictive analytical models can be reduced by correcting the input discharge with water withdrawal or discharge by pumping stations.

Although wind velocity and direction are measured at the mouth of the estuary, effects of wind on the water surface in the estuary are not included. As Savenije (2012) states, mixing inside the estuary is dominated by the salt- and freshwater flux. Therefore, mixing due to wind is not incorporated in order to decrease the number of input parameters.

5.3 Identifying salinity changes in the Port of Rotterdam

Salinization at the selected measurement locations occurs on a daily frequency at Lekhaven and Spijkenisserbrug (Annex A) below a 1500 m³/s discharge of the Rhine. At Beerenplaat and Brienoordbrug, salinization occurs during low discharges in combination with wind setup. Although salinization at these locations is not regarded as problematic, the measurement locations used in this study are selected because of the higher frequency of salinization compared to more upstream locations. Therefore, within the scope of this research salinity changes in the Port of Rotterdam are examined at the four selected measurement locations. However, the described methodology in Chapter 2 and Chapter 3 may also be applicable to measurement locations further upstream in the estuary.

The stated methodology in Chapter 4 is one of the many possible analysis methods of residuals computed post human intervention. As these residuals are not available yet, the optimal methodology for analysis of these residuals is yet unclear.

For example, based on the systematic difference between low and high tide observations at Spijkenisserbrug, Brienoordbrug and Beerenplaat, where chloride concentrations return to background concentrations of the Rhine during low tide, separate analysis is more useful. Difference in means, as described in Section 4.2 can be applied on high tide observations, while computing an exceedance frequency, based on a certain threshold, might provide more insight on changes during low tide.

When assessing effects of human interventions to the estuary, such as deepening, which alter the morphology, some results may have to be revised, such as time lags. Cai et al. (2012) concluded that deepening of an estuary decreased the wave travel time due to an increased tidal amplitude. As wave travel time at each measurement location, indicated with the time lag, is assumed constant, the effect of alterations in the estuary on wave travel time is currently disregarded. The wave celerity in an estuary can be estimated by: $c = \sqrt{gh}$, where g is the gravitational acceleration and h is the water depth (Savenije, 2012). An increase of 10 percent of the water depth, would result in a 5 percent increase in the wave velocity. However, the wave travel time is also dependant on the friction and discharge volume (Nguyen, 2008). Theoretically, if the wave travel time and, consequently, the time lag would decrease with 5 percent, this would result in a decrease of 5 minutes regarding the time lag

of the tide for Lekhaven. This would not affect results significantly as the cross-correlation function shows a wide peak of optimal time lags (Figure 14) and is followed by the tidal sampling in which a period of around 6.2 hours is considered.

5.4 Benefits and other applications of model

Developed models can be fairly easily implemented in real-time monitoring systems and applied for chloride concentrations predictions. Input parameters: astronomical tide, wind setup and discharge of the Waal, Lek and Meuse are easily accessible for daily application of the model, on collected measurements post deepening. The obtained residuals may be tested for deviation of the mean prior to deepening, as described in Section 4.1, at, for example, a monthly interval. This provides a regular update as indication of effects due to the deepening of the New Waterway and Botlek. A real-time implementation can also be used for further calibration of developed models, as performance of the models can be assessed on a continuous basis.

Currently, the methodology is based on the semidiurnal tidal cycle observed at the Rhine-Meuse basin. However, the sampling methodology, from peak-to-trough and trough-to-peak, as described in Section 2.5, may also be applied to other tidal cycles, such as diurnal or mixed semidiurnal tidal cycles. This makes the developed methodology interesting for predicting chloride concentrations with an analytical model for other estuaries.

6 Conclusion and recommendations

6.1 Conclusion

6.1.1 Optimization of dataset

Available measurements in the analysis period of 2011 to present, comprise the discharges of the Waal, Meuse and Lek and the water levels at the mouth of the Rhine-Meuse estuary, which are split in an astronomical tide and a wind setup component. Due to the non-normal distribution of these parameters, the Spearman R correlation is applied to assess changes in correlation with varying time lags.

Optimal correlation between the water level at Hoek van Holland and the chloride concentration measurement locations Lekhaven, Spijkenisserbrug, Brienenoordbrug and Beerenplaat is found at a time lag of 110, 190, 200 and 280 minutes, respectively.

Optimal correlation between the discharge of the Waal at Tiel and the chloride concentration measurement locations Brienenoordbrug, Beerenplaat, Lekhaven and Spijkenisserbrug is found at a time lag of 1000, 1100, 1150 and 1250 minutes, respectively. Regarding the Meuse discharge measured at Megen time lags of 1500, 1650, 1750 and 1900 minutes optimize the correlation with chloride concentrations at Beerenplaat, Spijkenisserbrug, Brienenoordbrug, Lekhaven, respectively. Time lags of the discharge of the Lek could not be determined with the cross-correlation analysis and were determined based on expert judgement at 750, 900, 160 and 1750 minutes for Brienenoordbrug, Lekhaven, Beerenplaat and Spijkenisserbrug, respectively.

Analysis of the above stated input parameters and their relation to chloride concentrations at each measurement location, is best performed at the interval of the in- and outflow of the tidal wave. This is achieved by separating the tidal wave signal in a trough-to-peak and peak-to-trough section (Figure 20). Corresponding to each section the average discharge and wind setup are calculated. Regarding the trough-to-peak section, the minimum chloride concentration and tidal water level are determined. For the peak-to-trough section, the maximum chloride concentration and tidal water level are determined. For inclusion of tide, the tidal sampling method proves to be an improvement over the 10-minute or hourly sampling interval.

6.1.2 Predictive analytical model development

At all four measurement locations the non-linear LassoCV model performs better than the linear OLS model (Annex D and Annex E). The highest factor of variance explained by the input parameters (R^2) regarding the validation dataset, on which the model type is selected, is found at the most downstream measurement locations Spijkenisserbrug and Lekhaven (Table 23), on the Old Meuse and New Meuse, respectively. Further upstream at Beerenplaat and Brienenoordbrug, where the influence of the tidal wave decreases, the R^2 -value decreases. Also, the normalized estimation error (NRMSE) increases, which is caused by relatively lower average observed chloride concentrations at these locations compared to other locations.

Table 23. Selected model type regarding each measurement location and corresponding model performance on validation dataset.

Measurement location	Model	Validation dataset		
		<i>R</i> ² [-]	<i>RMSE</i> [mg/L]	<i>NRMSE</i> [-]
Lekhaven	non-linear	0.799	579.1	0.31
Spijkenisserbrug	non-linear	0.878	902.4	0.37
Brienoordbrug	non-linear	0.699	413.1	0.69
Beerenplaat	non-linear	0.620	531.2	1.42

With the use of the multi-step analyses a selection of parameters was determined to describe chloride concentration at each of the measurement locations (Table 24). Discharge of the Lek is not incorporated in any of the predictive analytical models, as it did not improve model performance on the validation dataset. The addition of parameters describing the autocorrelation of the input parameters (MWA) proves to be a valuable addition as salinity observations can be examined independently.

Table 24. Overview of selected parameters, with the use of the multi-step analysis, to describe chloride concentrations at each of the four measurement locations. X indicates a selected parameter.

Parameter name	Lekhaven	Spijkenisser brug	Brienoordbrug	Beerenplaat
Astronomical Tide	x	x	x	x
Discharge Waal	x	x	x	x
Wind setup	x	x	x	x
Discharge Lek				
Discharge Meuse	x	x	x	
MWA Discharge Waal		x		
MWA Discharge Meuse	x	x	x	
MWA Discharge Lek				
MWA Wind setup	x	x	x	x
MWA Tide				
Tidal amplitude	x	x	x	x
Smotherstep tide		x		x
Smotherstep Wind setup	x			
Smotherstep MWA wind setup	x			
Smotherstep MWA tide			x	

All models were found to be sensitive to the set of parameters, astronomical tide, discharge of the Waal and wind setup. And far less sensitive to the discharge of the Meuse.

6.1.3 Future application of predictive analytical models

Application of the developed predictive analytical model must be performed on data gathered within the ranges of the input parameters used in the training and validation dataset. The proposed application of the t-test for analysing difference in mean residuals prior and post human interventions is a more extensive analysis compared to the z-score analysis applied in previous research, as the t-test incorporates the sample size and variance of both datasets.

The developed chloride prediction models can be applied to analyse effects of human interventions on chloride concentrations within the Port of Rotterdam. Real-time

implementation of the models will facilitate continuous updates of these effects. A real-time implementation may also be applied to predict chloride concentrations, based on expected discharges of the Rhine and the astronomical tide, which are currently not available.

6.2 Recommendations

Based on this research on chloride concentrations in the Port of Rotterdam, recommendations can be made regarding the input of: boundary condition discharge, the applied methodology and the application of the model.

Decreasing the uncertainty of the discharge must be performed. The largest uncertainty of the input boundary conditions lies in the discharge at each measurement location, due to the uncertainty of the discharge distribution within the estuary and lacking information on withdrawal or discharge at pumping stations. For decreasing the uncertainty, it is best to perform continuous discharge measurements in the New Meuse and Old Meuse. Although these will be influenced by tide, tidally averaged discharge data will provide valuable insights on the distribution along branches and make the use of regional discharge data or data on the withdrawal by pumping stations unnecessary. This would have the advantages that, in contrast to adding parameter such as upstream water withdrawal, it does not increase the amount of input parameters of the analytical models. Also, the dependence on the discharge limit below 1500 m³/s would no longer be present, increasing the application range of the methodology and the model.

The selection of the training and validation datasets should be randomized. The application of a random determination of training and validation datasets is fairly easily applicable.

Real-time implementation of the model is fairly easily executable and will facilitate further calibration. This directly makes it possible to analyse model performance on discharge above the current threshold of 1500 m³/s. Also, real-time implementation can be used for predictions of chloride concentrations based on predictions of the input parameters. Predictions of chloride concentrations are currently non-existent.

Finally, the main recommendation is the application of the model to analyse the effect of deepening of the New Waterway and Botlek on chloride concentrations in the Port of Rotterdam, once the deepening has been fully completed.

7 Bibliography

- Alberto, W., María del Pilar, D., María Valeria, A., Fabiana, P., Cecilia, H., & María de los Ángeles, B. (2002, 7 25). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study. *Water Research*, 35(12), 2881-2894.
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv.*, 4, 40-79.
- Barnston, A. (1992, 12 1). Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score. *Weather and Forecasting*, 7(4), 699-709.
- Blaas, M., & van den Boogaard, H. (2006). *Statistical methods to asses the impact of MV2 on SPM along the Dutch coast*. WL | Delft Hydraulics Report Z4046_a.
- Bolaños, R., Brown, J., Amoudry, L., Souza, A., Bolaños, R., Brown, J., . . . Souza, A. (2013, 1 28). Tidal, Riverine, and Wind Influences on the Circulation of a Macrotidal Estuary. *Journal of Physical Oceanography*, 43(1), 29-50.
- Cai, H., Savenije, H., Yang, Q., Ou, S., & Lei, Y. (2012, 10). Influence of River Discharge and Dredging on Tidal Wave Propagation: Modaomen Estuary Case. *Journal of Hydraulic Engineering*, 138(10), 885-896.
- Chetverikov, D., & Liao, Z. (2016). *On Cross-Validated Lasso*. The Institute for Fiscal Studies, Department of Economics, UCL.
- De Nijs, M. (2012). *On sedimentation processes in a stratified estuarine system*. Delft: TU Delft. Retrieved from <https://doi.org/10.4233/uuid:fb0ec18d-dcfd-468d-af4a-455fbf02cac1c>
- Deltares. (2014). *Evaluatie van het OSR-model voor zoutindringing in de Rijn-Maasmonding (I) Onderdeel KPP B&O Waterkwaliteitsmodelschematisaties*.
- Deltares. (2016). *Systeemanalyse Rijn-Maasmonding: analyse relaties noord- en zuidrand en gevoeligheid stuurknoppen*. Deltares.
- Deltares. (2017). *Het Haringvlies na de Kier. Samenvatting van hydrologische prognoses ten behoeve van effectinschatting op vis en vogels*. Deltares.
- Derrick, B., & White, P. (2016). Why Welch's test is Type I error robust. *The Quantitative Methods for Psychology*, 12(1), 30-38.
- Devroye, L., & Wagner, T. (1979, 9). Distribution-free performance bounds for potential function rules. *IEEE Transactions on Information Theory*, 25(5), 601-604.
- Hickey, B., Zhang, X., & Banas, N. (2002, 10 1). Coupling between the California Current System and a coastal plain estuary in low riverflow conditions. *Journal of Geophysical Research*, 107(C10), 3166.
- HydroLogic. (2015a). *Verziltling door verdieping Nieuwe Waterweg en Botlek (P716)*. Havenbedrijf Rotterdam.
- HydroLogic. (2015b). *Inventarisatie Slim Watermanagement Rijn-Maasmonding (P720)*. Amersfoort: HydroLogic BV.
- HydroLogic. (2018). *Tijdreeksanalyse Verdieping NWW (P928)*. Amersfoort.
- Iooss, B., & Lemaitre, P. (2015). *A review on global sensitivity analysis methods*. Springer.
- Jochen, K. (2016). The Functioning of Coastal Upwelling Systems. In K. Jochen, *Upwelling Systems of the World: A Scientific Journey to the Most Productive Marine Ecosystems* (pp. 31-65). Springer International Publishing.

- KNMI. (2011, June 1). *Lente 2011*. Retrieved from Lente 2011, extreem droog en zonnig, zeer zacht: <https://www.knmi.nl/nederland-nu/klimatologie/maand-en-seizoensoverzichten/2011/lente>
- Kranenbrug, W., Mens, M., Buschman, F., Wesselius, C., Huismans, Y., Ter Maat, J., & Diermanse, F. (2015). *Systeemanalyse van de Rijn-Maasmonding Voor Verzilting (Factsheets Proceskennis, Systeemkennis, Modelinstrumentarium En Statistiek)*. Deltares.
- Langedoen, E. (1992). *Flow patterns and transports of dissolved matter in tidal harbours*. Ph.D. Thesis. Delft, The Netherlands: Delft University of Technology.
- Ministerie van Infrastructuur en Milieu. (2019, April 18). *Besluit kwaliteitseisen en monitoring water 2009*. Retrieved from <https://wetten.overheid.nl/BWBR0027061/2017-01-01>
- Nguyen, A. (2008). *Salt Intrusion, Tides and Mixing in Multi-Channel Estuaries*. Taylor & Francis.
- Ocean Motion. (n.d.). *Ocean Motion : Definition : Wind Driven Surface Currents - Upwelling and Downwelling*. Retrieved from <http://oceanmotion.org/html/background/upwelling-and-downwelling.htm>
- Open University. Oceanography Course Team. (1999). *Waves, tides and shallow-water processes*. Milton Keynes, England: Pergamon Press, in association with Open University.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011, 11). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12, 2825-2830.
- Port of Rotterdam. (2016). *Preparations underway for deepening of Nieuwe Waterweg | Port of Rotterdam*. Retrieved from <https://www.portofrotterdam.com/en/news-and-press-releases/preparations-underway-for-deepening-of-nieuwe-waterweg>
- Reyes-Merlo, M., Díez-Minguito, M., Ortega-Sánchez, M., Baquerizo, A., & Losada, M. (2013, 13). On the relative influence of climate forcing agents on the saline intrusion in a well-mixed estuary: Medium-term Monte Carlo predictions. *Journal of Coastal Research*, 165(sp2), 1200-1205.
- Rijkswaterstaat. (2011). *Beschrijving huidige situatie Haringvliet*. Rijkswaterstaat.
- Rijkswaterstaat. (2015). *Watermanagement en het stuwenensemble Nederrijn en Lek. Voldoende zoetwater, bevaarbare rivieren*. Rijkswaterstaat.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010, 2 1). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259-270.
- Savenije, H. (1993). Predictive model for salt intrusion in estuaries. *Journal of Hydrology*, 203-218.
- Savenije, H. (2012). *Salinity and tide in alluvial estuaries*. Amsterdam: Elsevier.
- Savenije, H., Toffolon, M., Haas, J., & Veling, E. (2008). Analytical description of tidal dynamics in convergent estuaries. *J. Geophys. Res. Oceans*, 113(C10), 18.
- Shrestha, S., & Kazama, F. (2007, 4). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling and Software*, 22(4), 464-475.
- Sobol', I. (2001, 2 15). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3), 271-280.
- Svasek Hydraulics. (2015). *Effectbepaling verdieping Nieuwe Waterweg op verzilting in benedenrivierengebied*. Havenbedrijf Rotterdam.

- Usher, W., Herman, J., Whealton, C., Hadka, D., xantares, Rios, F., . . . Engelen, J. (2016, 10 11). SALib/SALib: Launch!
- van den Burgh, P. (1972). *Ontwikkeling van een methode voor het voorspellen van zoutverdeling in estuaria, kanalen en zee (in Dutch)*. Rijkswaterstaat Rapport.
- van Rijn, L. (2011, 11 10). Analytical and numerical analysis of tides and salinities in estuaries; part I: tidal wave propagation in convergent estuaries. *Ocean Dynamics*, 61(11), 1719-1741.
- Wan, H., Xia, J., Zhang, L., She, D., Xiao, Y., & Zou, L. (2015). Sensitivity and interaction analysis based on Sobol' method and its application in a distributed flood forecasting model. *Water (Switzerland)*, 7(6), 2924-2951.
- Welch, B. (1947, 1 1). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1-2), 28-35.
- Xu, Y., Zhang, W., Zhu, Y., & Zheng, J. (2017, 10 10). Analytical solution for salt intrusion in multiple-freshwater-source estuaries: application to Humen Estuary. *Environmental Earth Sciences*, 76(19), 661.
- Zimmerman, D. (2004, 5 1). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1), 173-181.
- Zu, T., & Gan, J. (2015, 7 1). A numerical study of coupled estuary–shelf circulation around the Pearl River Estuary during summer: Responses to variable winds, tides and river discharge. *Deep Sea Research Part II: Topical Studies in Oceanography*, 117, 53-64.

Annex A Discharge Rhine and chloride concentrations

On the north side of the system, on the New Meuse, when chloride concentration measurement are plotted against the Rhine discharge measurements at Lobith, three distinct states in the system show up (Figure 50). With discharges above 4850 m³/s at Lobith and wind setup below 15 cm, the tide causes no increased chloride concentrations at both measurement locations (stage 3 in Figure 50). Between Rhine discharge at Lobith of 2350 m³/s and 4850 m³/s, intruding tide does not cause increased chloride concentrations at Brienoordbrug but does at Lekhaven (stage 2 in Figure 50). Below a Rhine discharge of 2350 m³/s at Lobith the intruding tide causes increased chloride concentrations both at Lekhaven and Brienoordbrug (stage 1 in Figure 50).

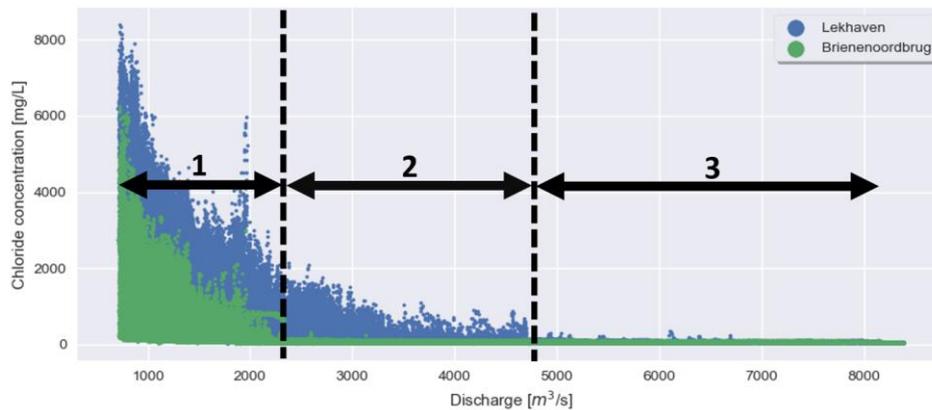


Figure 50. Chloride concentration measurements in the New Meuse and discharge at Lobith (with delay of 1 day), filtered for situation with wind setup below 15 cm.

On the south side of the system, on the Old Meuse, a similar classification can be made (Figure 51). With discharges above 6250 m³/s at Lobith and wind setup below 15 cm the tide causes no increased chloride concentrations at both measurement locations. Between Rhine discharge at Lobith of 3000 m³/s and 6250 m³/s, the intruding tide does not cause increased chloride concentrations at Beerenplaat but does so at Spijkenisserbrug. Below a Rhine discharge of 3000 m³/s at Lobith the intruding tide causes increased chloride concentrations both at Spijkenisserbrug and Beerenplaat.

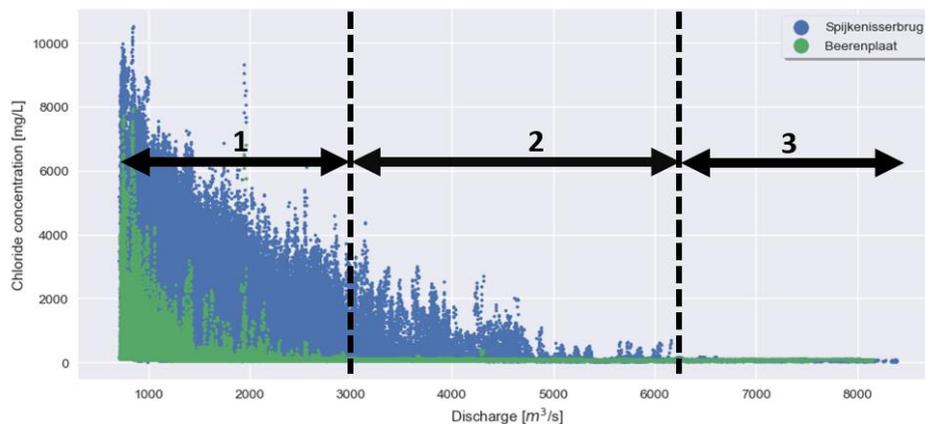


Figure 51. Chloride concentration measurements in the Old Meuse and discharge at Lobith (with delay of 1 day), filtered for situation with wind setup below 15 cm.

The observations from Figure 50 and Figure 51 can be translated to an intrusion length of the salt wedge on the New Meuse and Old Meuse based on Rhine discharge measurements at Lobith (Figure 53).

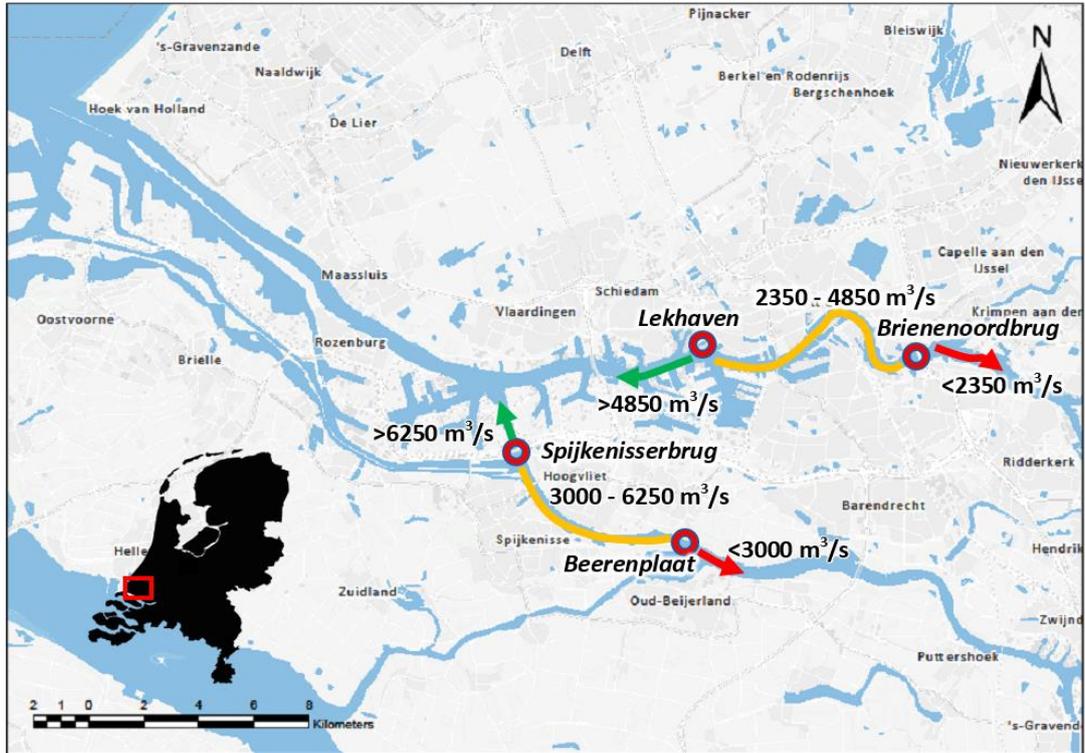


Figure 52. Intrusion of salt wedge based on Rhine discharge measured at Lobith with wind setup at Hoek van Holland below 0.15cm.

Annex B Lekhaven - model parameter coefficients

Parameter	Coefficients LassoCV model
Tide	5227.9
Discharge Waal	-4262.2
Wind setup	0
Discharge Meuse	-403.7
MWA wind setup	853.2
MWA discharge Meuse	-941.2
Tidal amplitude	-310.9
Tide * Discharge Waal	-3056.0
Tide * Wind setup	-862.7
Tide * Discharge Meuse	0
Tide * MWA wind setup	1645.9
Tide * MWA discharge Meuse	-756.7
Tide * Tidal amplitude	-444.5
Discharge Waal * Wind setup	0
Discharge Waal * Discharge Meuse	0
Discharge Waal * MWA wind setup	0
Discharge Waal * MWA discharge Meuse	0
Discharge Waal * Tidal amplitude	0
Wind setup * Discharge Meuse	0
Wind setup * MWA wind setup	-246.2
Wind setup * MWA discharge Meuse	0
Wind setup * Tidal amplitude	0
Discharge Meuse * MWA wind setup	0
Discharge Meuse * MWA discharge Meuse	0
Discharge Meuse * Tidal amplitude	0
MWA wind setup * MWA discharge Meuse	0
MWA wind setup * Tidal amplitude	0
MWA discharge Meuse * Tidal amplitude	0
Smootherstep Wind setup	1820.7
Smootherstep MWA wind setup	1120.9

Annex C Spijkenisserbrug - model parameter coefficients

Parameter	linear OLS model	non-linear LassoCV model
Tide	6818.9	3250.9
Discharge Waal	-7235.9	-1648.8
Wind setup	2719.9	2633.5
Discharge Meuse	-2103.9	0
MWA discharge Waal	1518.5	14.2
MWA wind setup	2071.3	1732.0
MWA discharge Meuse	623.1	0
Tidal amplitude	222.5	-67.9
Tide * Discharge Waal		-10001.8
Tide * Wind setup		2406.8
Tide * Discharge Meuse		-3364.3
Tide * MWA discharge Waal		0
Tide * MWA wind setup		1261.4
Tide * MWA discharge Meuse		-3117.8
Tide * Tidal amplitude		1671.2
Discharge Waal * Wind setup		0
Discharge Waal * Discharge Meuse		0
Discharge Waal * MWA discharge Waal		0
Discharge Waal * MWA wind setup		-856.9
Discharge Waal * MWA discharge Meuse		0
Discharge Waal * Tidal amplitude		0
Wind setup * Discharge Meuse		0
Wind setup * MWA discharge Waal		0
Wind setup * MWA wind setup		3984.1
Wind setup * MWA discharge Meuse		0
Wind setup * Tidal amplitude		-2804.9
Discharge Meuse * MWA discharge Waal		0
Discharge Meuse * MWA wind setup		-1725.6
Discharge Meuse * MWA discharge Meuse		0
Discharge Meuse * Tidal amplitude		0
MWA discharge Waal * MWA wind setup		0
MWA discharge Waal * MWA discharge Meuse		0
MWA discharge Waal * Tidal amplitude		0
MWA wind setup * MWA discharge Meuse		0
MWA wind setup * Tidal amplitude		0
MWA discharge Meuse * Tidal amplitude		0
Smotherstep Tide		7435.2

Annex D Model development Brienenoordbrug

Results of the multi-step analysis regarding the linear and non-linear model are provided in Figure 53 and Figure 54. An overview of the selected parameters is provided in Table 25.

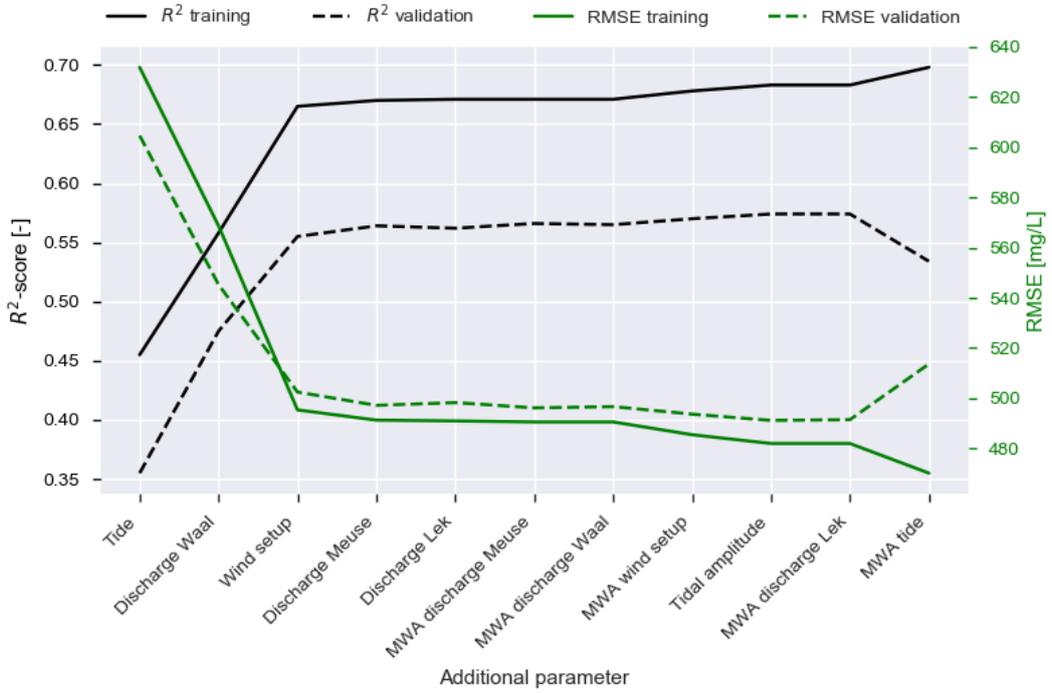


Figure 53. Result of multistep analysis of linear model regarding Brienenoordbrug.

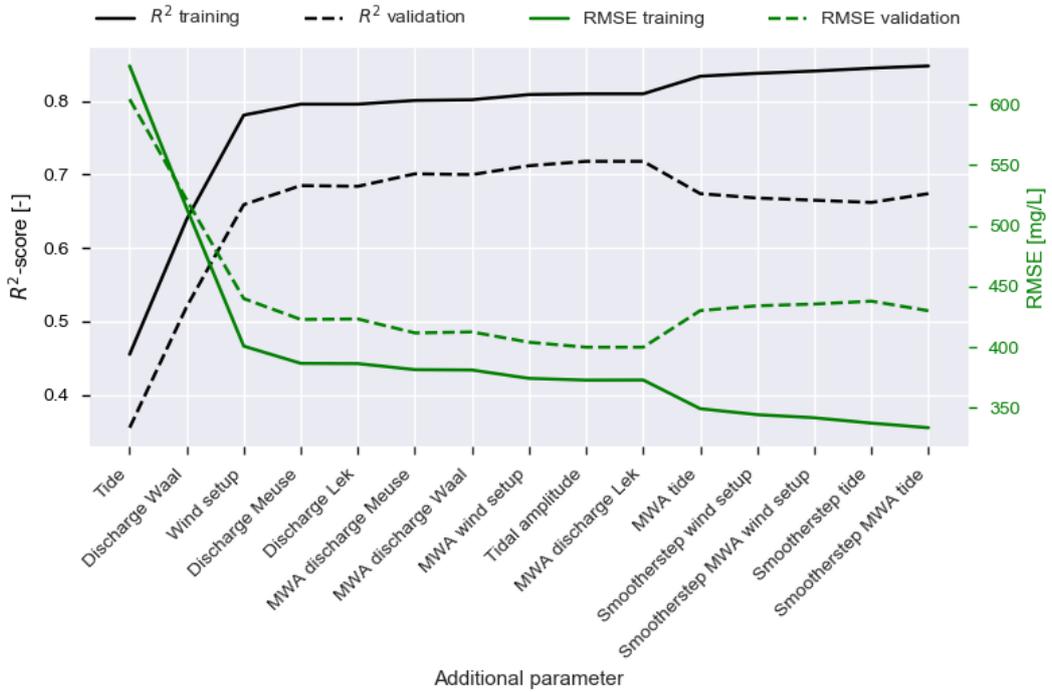


Figure 54. Result of multi-step analysis of nonlinear model regarding Brienenoordbrug.

Table 25. Select parameters for both models based on the outcome of the multi-step analysis.

Parameter name	linear OLS model	nonlinear LassoCV model
Astronomical Tide	x	x
Discharge Waal	x	x
Wind setup	x	x
Discharge Lek		
Discharge Meuse	x	x
MWA Discharge Waal		
MWA Discharge Meuse	x	x
MWA Discharge Lek		
MWA Wind setup	x	x
MWA Tide		
Tidal amplitude	x	x
Smootherstep tide		
Smootherstep Wind setup		
Smootherstep MWA wind setup		
Smootherstep MWA tide		x

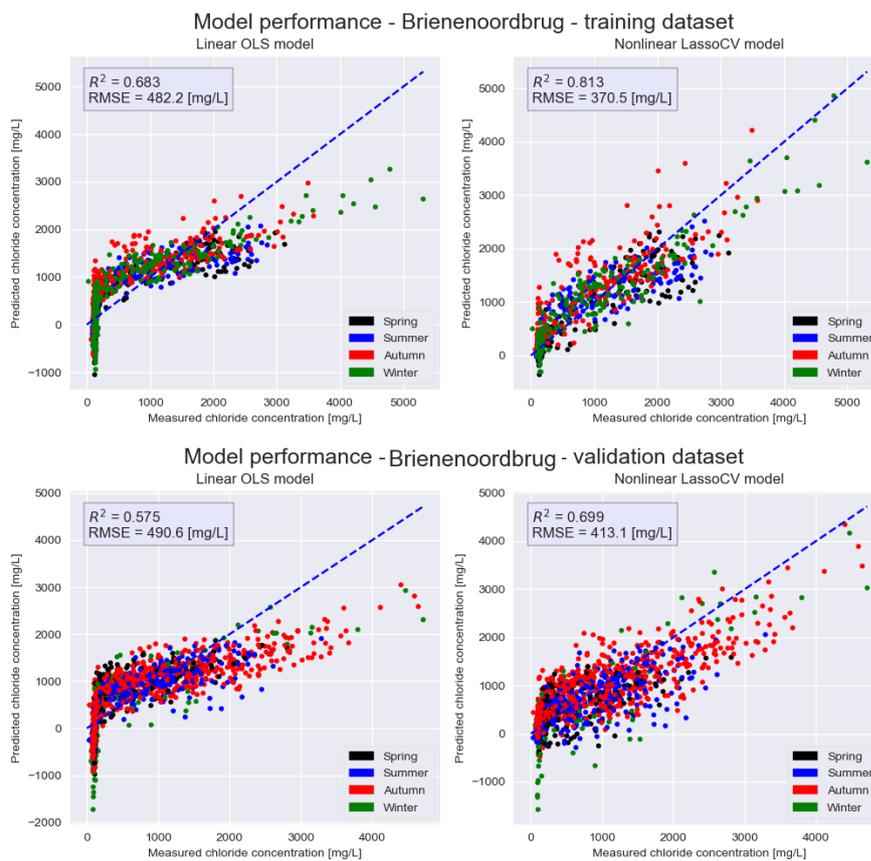


Figure 55. Model performance on the training dataset (top panels) and the validation dataset (bottom panels).

Based on the outperformance of the nonlinear LassoCV model over the linear model, the sensitivity analysis is only performed on the nonlinear model (Figure 56).

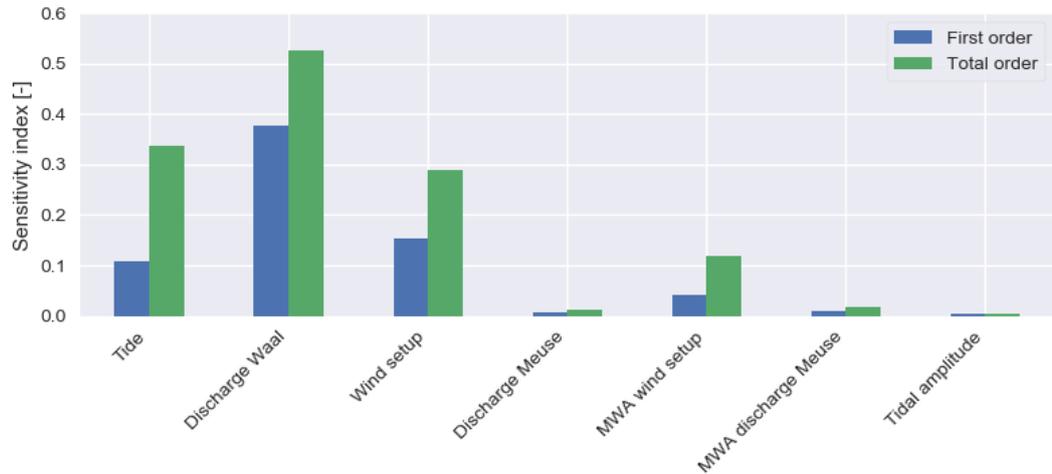


Figure 56. Result of sensitivity analysis of non-linear LassoCV model.

Table 26. Parameter coefficients of nonlinear LassoCV model

Parameter	Coefficients LassoCV model
Tide	6252
Discharge Waal	0
Wind setup	0
Discharge Meuse	0
MWA wind setup	161
MWA discharge Meuse	0
Tidal amplitude	0
Tide * Discharge Waal	-6767
Tide * Wind setup	2215
Tide * Discharge Meuse	-891
Tide * MWA wind setup	694
Tide * MWA discharge Meuse	-1532
Tide * Tidal amplitude	18
Discharge Waal * Wind setup	0
Discharge Waal * MWA discharge Meuse	549
Discharge Waal * Tidal amplitude	0
Wind setup * Discharge Meuse	-240
Wind setup * MWA wind setup	1180
Wind setup * MWA discharge Meuse	0
Wind setup * Tidal amplitude	0
Discharge Meuse * MWA discharge Meuse	0
Discharge Meuse * Tidal amplitude	0
MWA wind setup * MWA discharge Meuse	0
MWA wind setup * Tidal amplitude	0
MWA discharge Meuse * Tidal amplitude	0
Smootherstep tide	205

Annex E Model development Beerenplaat

Results of the multi-step analysis regarding the linear and non-linear model are provided in Figure 58 and Figure 59. An overview of the selected parameters is provided in Table 28.

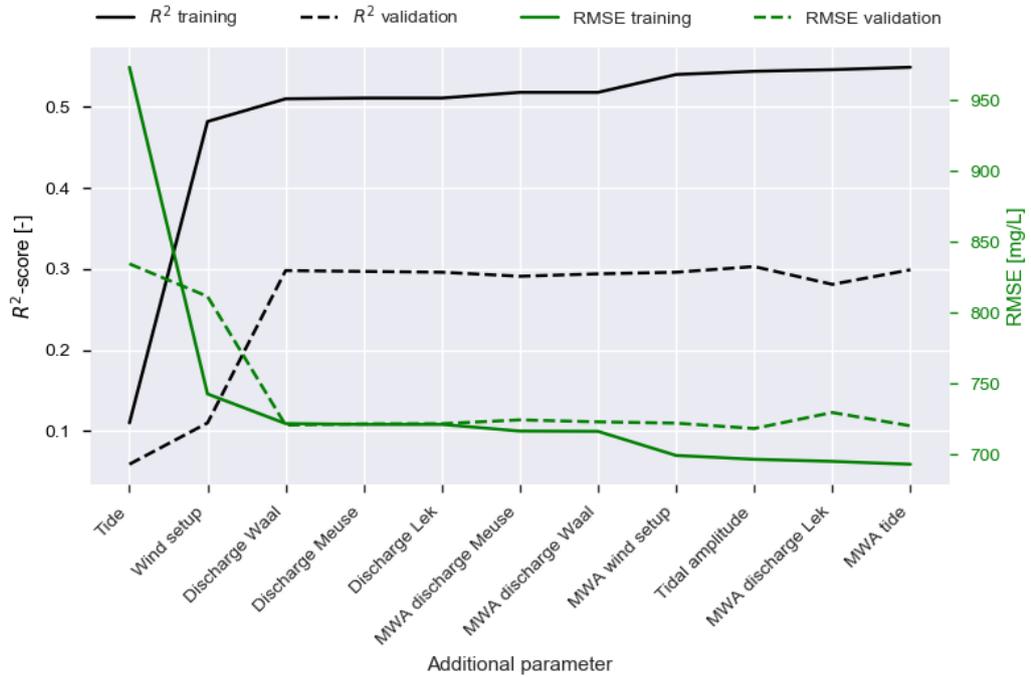


Figure 57. Result of multistep analysis of linear model regarding Beerenplaat.

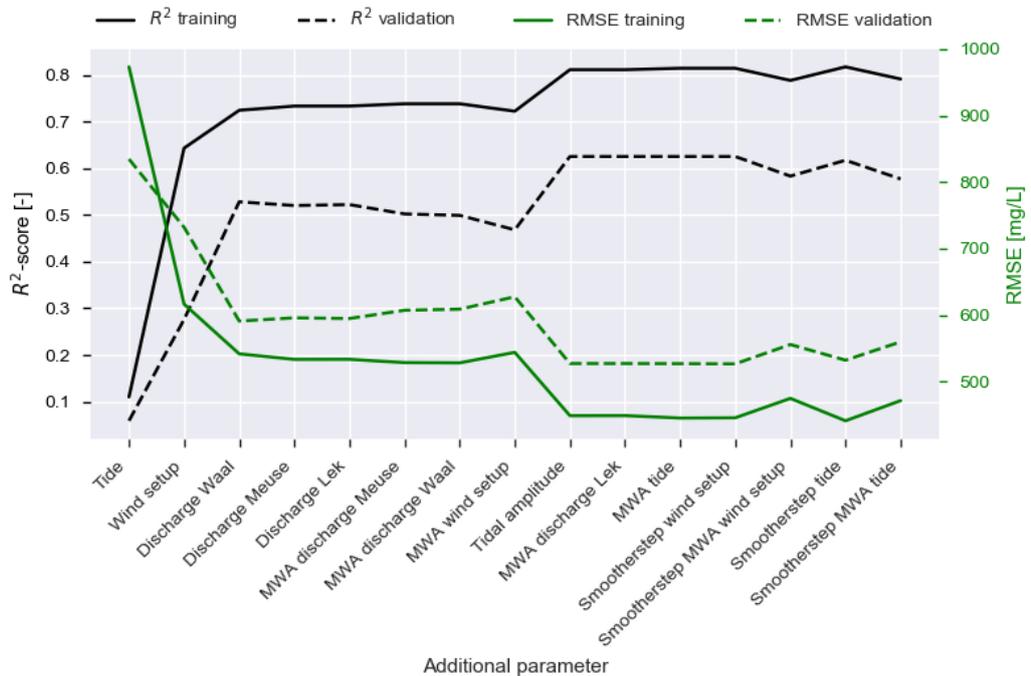


Figure 58. Result of multistep analysis of non-linear model regarding Beerenplaat.

Table 27. Select parameters for both models based on the outcome of the multi-step analysis.

Parameter name	linear OLS model	nonlinear LassoCV model
Astronomical Tide	x	x
Discharge Waal	x	x
Wind setup	x	x
Discharge Lek		
Discharge Meuse		
MWA Discharge Waal		
MWA Discharge Meuse		
MWA Discharge Lek		
MWA Wind setup	x	x
MWA Tide	x	
Tidal amplitude		x
Smootherstep tide		x
Smootherstep Wind setup		
Smootherstep MWA wind setup		
Smootherstep MWA tide		

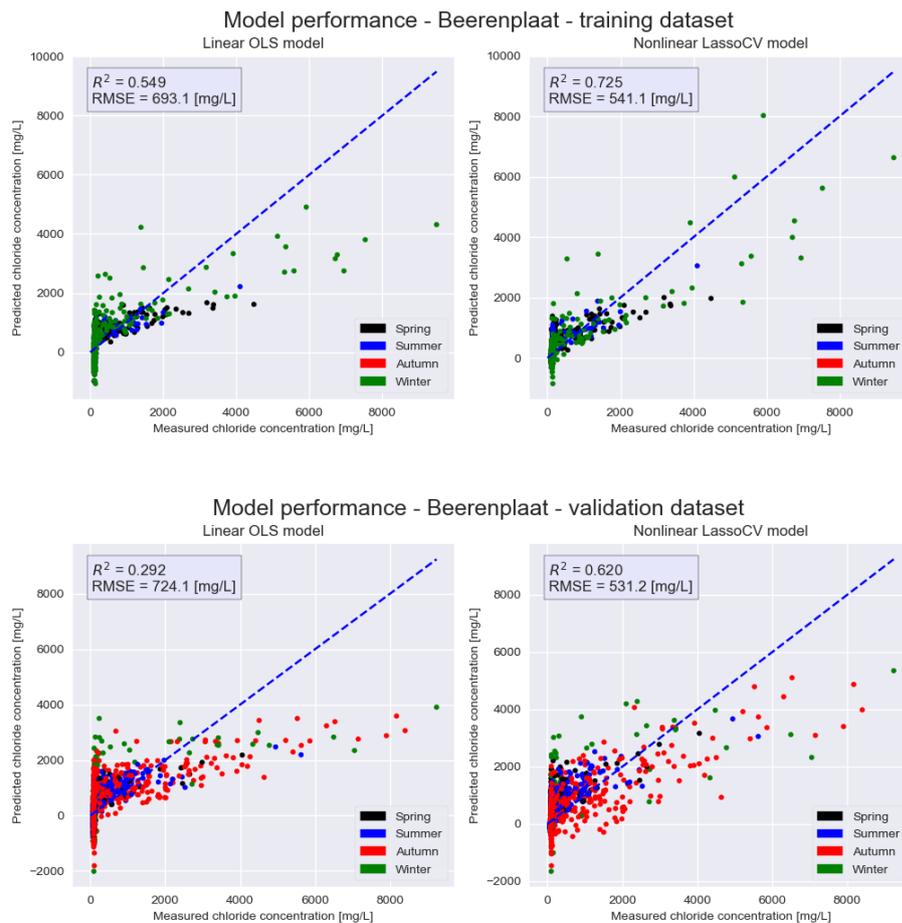


Figure 59. Model performance on the training dataset (top panels) and the validation dataset (bottom panels).

Based on the outperformance of the nonlinear LassoCV model over the linear model, the sensitivity analysis is only performed on the nonlinear model (Figure 54Figure 56).

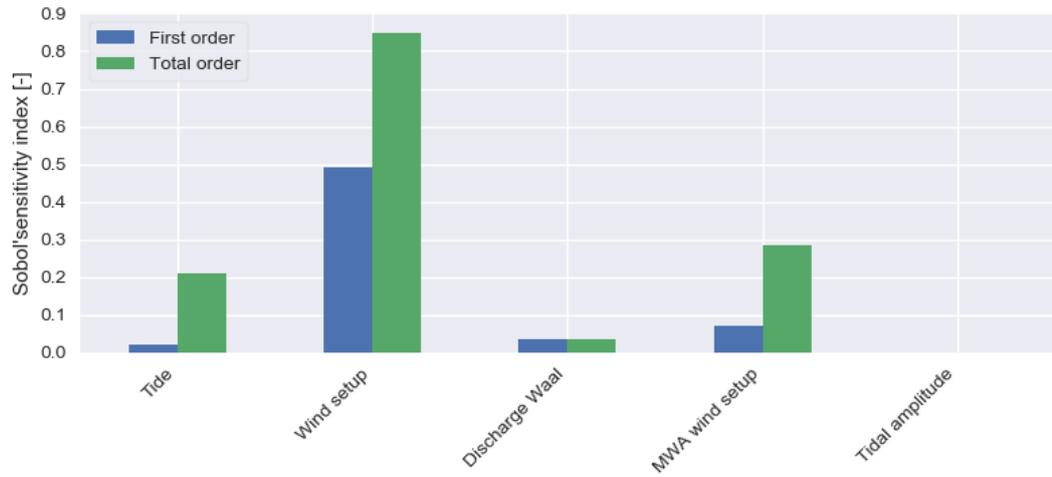


Figure 60. Result of sensitivity analysis of non-linear LassoCV model.

Table 28. Parameter coefficients of nonlinear LassoCV model

Parameter	Coefficients LassoCV model
Tide	0
Discharge Waal	-1305.4
Wind setup	0
MWA wind setup	0
Tidal amplitude	0
Tide * Discharge Waal	0
Tide * Wind setup	4781.5
Tide * MWA wind setup	1804.5
Tide * Tidal amplitude	303.4
Discharge Waal * Wind setup	0
Discharge Waal * MWA wind setup	0
Wind setup * MWA wind setup	2587.1
Wind setup * Tidal amplitude	0
MWA wind setup * Tidal amplitude	0
MWA discharge Meuse * Tidal amplitude	0
Smootherstep Tide	663.5