

2022

Using Random Forest Machine learning to estimate the impact of hydrological drought on the shipping industry

MASTER THESIS
J.A.H VAN DE VEN



Using Random Forest Machine learning to estimate the impact of hydrological drought on the shipping industry

J.A.H van de Ven

Master thesis Utrecht University

2022

Abstract

Hydrological droughts can have a severe negative effect on inland shipping. For example, the 2018 drought is estimated to have caused damages up to 345 million euros for the shipping industry in the Netherlands alone. To better estimate the economic impact of drought on inland shipping we need to identify the drivers and mechanisms that result in severe shipping impacts. Therefore, it is essential that we improve our understanding of the impact of hydrological drought on shipping and identify the methods capable of achieving this goal. Conventionally, this is done by means of numerical modelling, but this method is very complex, requires understanding of the driving mechanisms and is often time consuming. In this paper I present an alternative method to estimate the impact of hydrological drought on shipping, by implementing machine learning methods. Machine learning methods are data driven instead of process driven and are often fast, easy to use and can deal with large amounts of data. In this paper a Random Forest Machine Learning model was developed which is able to accurately predict sailing depth and the economic damages resulting from drought for the Rhine branches and the Meuse. The model is able to reconstruct historic data and make predictions for a 2°C climate change scenario. The prediction made by this model can then be used to estimate the economic damages to shipping industry due to drought. Future work is however needed as model performance can vary between locations, which can result in relatively large errors in damage predictions for locations where the model's performance is not optimal. Notwithstanding this, the random forest models are often fast, easy to use and still accurate enough. The results of this paper show that machine learning is well suited for the study of the impact of hydrological drought. The new random forest model developed in this paper promises to be a very useful tool for water managers and researchers alike.

Table of content

Abstract	1
Table of content.....	2
1. Introduction	4
1.1 Droughts	4
1.2 (Hydrological) Drought impact	4
1.3 Hydrological modelling.....	5
1.4 Machine learning modelling.....	5
1.5 Research questions and hypothesis	6
2. Materials and methods.....	7
2.1 Study area.....	7
2.1.1 Rhine catchment	7
2.1.2 Meuse catchment	7
2.1.3 Locations of interest	7
2.2 Data	8
2.2.1 Input data	8
2.2.2 Pre-processing.....	8
2.3 ML method selection.....	8
2.4 Model development.....	10
2.4.1 Location specific model (LocSpe)	10
2.4.2 General model (Isildur).....	10
2.4.3 Sensitivity analysis of Random forests tuning parameters	11
2.4.4 Model training and evaluation	11
2.5 Damage estimation.....	11
2.6 Future predictions.....	12
3. Results	13
3.1 Sensitivity analysis and model development.....	13
3.1.1 Tuning parameters of location specific model	13
3.1.2 Tuning parameters of Isildur.....	14
3.1.3 Transformer selection.....	14
3.2 Location specific model results	15
3.2.1 Location used for training/familiar locations	15
3.2.2 Unfamiliar locations.....	15
3.3 Isildur results	17
3.3.1 Locations used for training	17
3.3.2 Unfamiliar locations.....	18
3.4 Damage estimation results.....	18

3.5 Future predictions.....	19
4. Discussion	23
4.1 Reconstruction of historic data.....	23
4.2 Damage estimation.....	24
4.3 Future predictions.....	25
4.3.1 Future sailing depth predictions	25
4.3.2 Future damage predictions	25
4.4 Improvements, issues and limitations.....	26
4.5 link to previous work.....	26
4.6 Recommendations	27
4.6 Synthesis.....	27
5. Conclusion.....	28
6. Acknowledgements	29
7. References	29
Appendix	32
Statement of originality of the MSc thesis	34

1. Introduction

Rivers perform a multitude of important socio-economic functions in a society, including but not limited to drinking water supply, inland shipping and recreation (Middelkoop et al., 2001). For the Netherlands inland shipping provides a vital connection from the port of Rotterdam to the hinterland which includes the highly industrialised Ruhr valley (van Slobbe et al., 2016). Change to this shipping system can result in disrupted transport networks and large economic damages (Hekman et al., 2019). Therefore, it is essential to better understand processes which could affect this vital connection. Droughts in particular are one of these processes which can severely impact the navigability of the Rhine river by lowering water levels in substantial parts (van Slobbe et al., 2016).

1.1 Droughts

Droughts are a type of natural hazard characterized by a lower availability of water than usual in a given area for a certain period (Mishra & Singh, 2010). There is however no universal definition for droughts because these definitions need to take local climatic conditions and the affected sector into account (Wilhite et al., 2007). In this paper I will use the definition proposed by Tallaksen & Van Lanen (2004) “*Drought is a sustained period of below-normal water availability. It is a recurring and worldwide phenomenon, with spatial and temporal characteristics that vary significantly from one region to another (p.4)*”. Droughts are often classified into four different types, meteorological drought, agricultural drought, hydrological drought and socio-economic drought (Mehran et al., 2015; Van Loon, 2015; Wilhite & Glantz, 1985). Meteorological drought refers to a period where little precipitation occurs in an area. Directly linked to this is agricultural drought which considers the impact of drought on agriculture. Hydrological drought refers to a period of time when streamflow and (sub)surface flow is abnormally low. Finally socio-economic drought considers the impact of the before mentioned types of droughts on a society. For the purposes of this paper, I am mainly interested in hydrological drought, as this relates to discharge and streamflow and consequently also to overland shipping.

The reasons for the occurrence of hydrological drought are manifold and highly dependent on local conditions (Van Loon, 2015). Hydrological droughts are a function of climatic processes, catchment characteristics and anthropogenic influences (Van Lanen et al., 2013; Van Loon & Van Lanen, 2012). Climatic processes include decreases in precipitation, warm winters resulting in little snow accumulation or cool summers resulting in little meltwater generation (Van Loon, 2015). Catchment characteristics which relate to hydrological drought include soil composition, geology, (groundwater) responsiveness and storage capacity (Van Lanen et al., 2013; Van Loon, 2015). Anthropogenic influences impact hydrological drought mainly by reservoirs and water extraction (Wanders & Wada, 2015). Reservoirs may dampen peak flow and buffer low flow conditions by saving water in wetter periods and releasing it during drought, which could reduce the impact of hydrological drought. On the other hand, water abstraction both from groundwater and from surface water decreases water availability and enhances hydrological drought. For more in-depth information on hydrological drought, I recommend the review by Van Loon (2015).

1.2 (Hydrological) Drought impact

Droughts have been identified as one of the most severe natural disasters both in terms of impact it has on quality of life and economic damages (Mishra & Singh, 2010). They are often referred to as ‘creeping disasters’ because they arise slowly and undetected and can last for multiple years (Van Loon, 2015). Droughts negatively impact water supply, water quality, agriculture, hydropower generation, ecosystems, recreational activity and river discharge (Mishra & Singh, 2010; Van Loon, 2015; van Vliet & Zwolsman, 2008). These negative effects can result in economic, ecological and human health damages (Sugg et al., 2020).

The Netherlands experienced severe drought in 2018, 2019 and 2020 (Witte et al., 2020). This drought had severe impacts on Western Europe, including reducing the streamflow and river depth of one of the most important waterways in Europe, the Rhine River (Jong, 2019). Due to reduced river depth barges were unable to travel at full capacity or were not able to travel at all (Hekman et al., 2019). This resulted in damages up to 345 million euros for the overland shipping in 2018 alone (Hekman et al., 2019). With climate change the regime of the Rhine is expected to become more rainwater dominated than it currently is (Pfister et al., 2004; van Slobbe et al., 2016). It is expected that the streamflow at Lobith in summer and autumn will decrease by 15-30% while in winter and spring streamflow is expected to increase by 30% (Hurkmans et al., 2010; van der Wiel et al., 2019). Climate change can therefore increase the risk of future hydrological droughts in the Netherlands.

1.3 Hydrological modelling

To study hydrological drought, we need to study the hydrology of a river. Research on rivers and streamflow is traditionally performed by means of numerical modelling (Srivastava et al., 2006). Numerical hydrological models work by solving the governing and empirical equations for processes in the hydrological cycle (Moges et al., 2021). Catchment hydrological models work by combining several smaller sub models for different individual processes, such as rainfall or runoff (Daniel et al., 2011). Numerical hydrologic models have excellent control over boundary conditions and allow for high degrees of customization.

Numerical models however do have significant disadvantages. Firstly, numerical models are heavily reliant on the accuracy of the chosen boundary conditions. Secondly, because nature is very complex and not all processes are well understood, simplifications must be made, which may result in inaccuracies in the model outcome (Cheng et al., 2020; Moges et al., 2021; Srivastava et al., 2006). Thirdly, numerical models require copious amounts of data and computer power (Daniel et al., 2011; Srivastava et al., 2006). Because of the previously mentioned issues new methods are needed to improve model results.

1.4 Machine learning modelling

A recent development in the field of modelling is modelling based on machine learning (ML) (Daniel et al., 2011). These methods are data oriented rather than process oriented (Daniel et al., 2011; Lange & Sippel, 2020). The basic principle behind most ML algorithms is finding functional relationships between different data sets (Lange & Sippel, 2020). This is done by training and testing the model on provided data (Boehmke & Greenwell, 2019; Burrell, 2016). The advantages of ML methods include: the ability to adapt to new data (Boehmke & Greenwell, 2019), the ability to capture nonlinear processes (Cheng et al., 2020), ease of use and computational efficiency (Loganathan & Mahindrakar, 2020). Hydrological models are well suited for the use of ML algorithms because there are clear inputs and outputs in a predictable direction (Lange & Sippel, 2020).

Like numerical models, ML methods also have disadvantages. Firstly, ML methods are heavily dependent on the quality and quantity of data (Boehmke & Greenwell, 2019; Hauswirth et al., 2021). Secondly, models may perform well during training but poorly during testing, which is called underfitting. Or vice versa, the model may perform poorly on the training data set but performs well on general cases (Boehmke & Greenwell, 2019). However, these challenges can be overcome.

Previous research has shown that ML methods can be used for modelling rainfall-runoff, streamflow and water quality among others (Lange & Sippel, 2020; Loganathan & Mahindrakar, 2020; Shamshirband et al., 2020; Tyralis & Papacharalampous, 2019). However little research has been performed on the impact of drought on river and sailing depth predictions by means of ML methods.

1.5 Research questions and hypothesis

According to the Dutch meteorological institute it is likely that droughts will become more frequent (KNMI, 2014; KNMI, 2018). In this research I want to gain a better understanding of the processes and consequences of hydrological droughts in the Netherlands. Having a better understanding of these processes and potential damages allows better strategies to be constructed to reduce the negative impact of these changes in the climate. The aim of this research is gaining insight in whether ML methods are a viable tool for the study of hydrological droughts, with a focus on the impact on overland shipping. The main research question to be answered within this paper is:

- To what degree can ML methods be used to study and predict the impact of hydrological drought on overland shipping

To achieve this, I aim to answer the following three sub questions:

- To what degree are ML methods capable of simulating historic data of river depth during hydrological droughts?
- To what degree are ML models suited to estimate the economic damages of hydrological drought on overland shipping?
- How reliable are ML based predictions of future river depths during droughts?

I hypothesize that ML methods will be well suited for the study of hydrological drought impact on overland shipping. Hydrological models often require vast amounts of data and, as mentioned previously, ML methods are well suited for process with large amounts of data. Furthermore, hydrological models have clear in and outputs along a predictable direction, making them well suited for the use of ML. I expect that the utilisation of ML for hydrological drought modelling can improve sailing depth predictions.

2. Materials and methods

This study aims to find out how feasible the use of ML methods is to model the impact of hydrological drought on the shipping industry in the Netherlands. To this end a Random Forest (RF) model was constructed to predict the sailing depth of the main rivers in the Netherlands. This was first constructed for the chokepoints along the busiest river, the Waal. Afterwards this model was expanded to also cover different locations along the river network of the Netherlands.

The predictions for sailing depth resulting from the RF model were then used to further estimate the direct economic damages of a drought by means of a RF. The model results were then compared to historical data to estimate the reliability of using RFs for modelling and estimating the impact of hydrological drought.

2.1 Study area

The study area includes the Rhine River, its distributaries and the Meuse River in the Netherlands (Figure 1). The Upper Rhine and Waal rivers are considered the most important shipping routes within the Netherlands (Hekman et al., 2019; Jong, 2019). Transport along the Waal and Upper Rhine is responsible for over half of all shipping transport costs in the Netherlands (Jong, 2019). The shipping importance of these rivers in combination with abundant data availability creates an ideal case study.

2.1.1 Rhine catchment

The Rhine river catchment (Figure 2) is a highly urbanised and industrialised river basin (Hurkmans et al., 2010). It is the river with the highest traffic density in Europe and a length of approximately 1230 kilometers (Hurkmans et al., 2010). The average discharge of the Rhine on entering the Netherlands at Lobith is 2200 m³/s with low flows of 600 m³/s and high flows of 16000 m³/s (Warmink et al., 2013). The Rhine originates in the Alpine regions of Switzerland and discharges into the North Sea. The Rhine drains parts of Switzerland, France, Germany, Austria, Luxembourg and the Netherlands.

In the Netherlands, the Rhine splits into the Waal and the Pannerdensch Kanaal at the Pannerdensch Kop bifurcation. Shortly downstream at IJssellkop, the Pannerdensch Kanaal further splits into the IJssel and the Nederrijn Rivers (Figure 1). Of the Rhine distributaries the Waal is the largest, carrying on average 71% of discharge from the Rhine. The Nederrijn and IJssel carry 14% and 15% of the discharge of Rhine, respectively (Frings et al., 2019; Warmink et al., 2013).

The Rhine is a combined rainwater and snowmelt fed river, however due to climate change it is expected to become more rainfall dominated (Adger et al., 2007; Jonkeren et al., 2011). This would result in higher water levels during winter and lower water levels during summer than they are currently (Jonkeren et al., 2011).

2.1.2 Meuse catchment

The Meuse River drains parts of France, Belgium and the Netherlands (Figure 2). The Meuse has a length of 885 km and an average discharge of 200 m³/s with low flows of 60 m³/s and high flows of 3000 m³/s. Whereas the Rhine is a combined rainfall and snowmelt fed river, the Meuse River is a primarily rainfall dominated river. Consequently, the river has considerable fluctuations in streamflow between years and seasons and quick response times. Typically low flows occur in summertime, while high flows occur in wintertime (Descy, 2009).

2.1.3 Locations of interest

There are multiple channels between the Meuse and Waal, and the Waal and Nederrijn. Furthermore, both the Nederrijn and Meuse have weirs in their channels. Based on this, the rivers can be subdivided

into smaller subsections between a bifurcation and a channel, a channel and a channel, a channel and a weir, or a weir and a weir. To represent each of these subsections, the following locations along the Rhine were used to develop the model: Amerongen, Arnhem, Driel, lower Hagestein, Katerveer, Nijmegen, Pannerdensche kop, Pannerdensche kanaal, Sint-Andries, Tiel and Zutphen. Additionally, the following locations along the Meuse were used: Belfeld, Grave, Heel, Roermond and Sambeek. These locations are highlighted in [Figure 1](#) with red and blue dots.

2.2 Data

2.2.1 Input data

To be able to simulate the sailing depth and shipping damage due to low flow the following parameters were considered as input for the ML model: sailing depth at target location, precipitation at De Bilt, discharge of the Rhine at Lobith, discharge of the Meuse at Eijsden and distance along the river of the target locations. Water Height at that location was used to estimate the sailing depth for locations with insufficient data on sailing depth water. Sailing depth is the target variable while the other mentioned parameters are predictor variables.

Data on discharge, water height and sailing depth were gathered from Rijkswaterstaat and precipitation data was gathered from the Royal Dutch Meteorological Institute (KNMI).

2.2.2 Pre-processing

For most locations used in developing the model, insufficient data is available for sailing depth. Data on sailing depth is only measured and made public during drought or low water levels. Since droughts occur infrequently, this data is sparse. Therefore, two methods were used to append this data and gather a sufficient dataset for training and testing the model.

1. Calculating sailing depth from water height measurements. This data is often readily available. On dates where both the water height and the sailing depth are known it is possible to calculate the difference between the two. This difference can then be used to convert water height into sailing depth. The average difference was used to make a good estimate of the water height-sailing depth difference. By applying this method data coverage could be improved. A similar method was also suggested in [van der Mark et al. \(2020\)](#).
2. Calculating the relationship between discharge and flow depth (a QH relationship) for the different locations. For locations and periods where the discharge and sailing depth are known it is possible to determine a QH relationship. Such relation are expressed in a $H = aQ^{(-)B}$ format. To get the exponents for this relationship a power function was fit to the sailing depth data and discharge data at a nearby point. The QH relation was used to fill in the gaps in the dataset after the water height method was used.

2.3 ML method selection

Recently RF have gained popularity in the field of hydrology (e.g. [Adnan et al., 2021](#); [Cheng et al., 2020](#); [Loganathan & Mahindrakar, 2020](#); [Shamshirband et al., 2020](#)). RF are a supervised ML method based on constructing multiple bagged random decision trees and taking the average result ([Breiman, 2001](#); [Hastie et al., 2009](#)). It works by adding another layer of randomness to bagging, which increases accuracy ([Adnan et al., 2021](#); [Liaw & Wiener, 2002](#)). The extra randomness decreases correlation between the individual trees and consequently decreases the variance of the predictions ([Tyralis & Papacharalampous, 2019](#)). These models are known to give high prediction accuracy and good-to-fit performances ([Adnan et al., 2021](#)). Furthermore RF are able to capture low flow events ([Hauswirth et al., 2021](#)). RF possess the same advantages as ML mentioned in the introduction, on top of that, RFs do not overfit ([Breiman, 2001](#); [Tyralis & Papacharalampous, 2019](#)). Because of the reasons mentioned above RF were selected for the purposes of this study.



Figure 1: The study area consisting of the Rhine (Rijn), Nederrijn, Waal and Meuse (Maas) rivers. The dark blue accented rivers are the most important rivers in the study area. The red dots represent locations used for training of the model, the blue dots represent locations used for testing and the yellow-red dots represent locations used for training and testing. The Red lines represent the major weirs in the rivers and the red arrows indicate the flow direction. The grey dots represent some major cities contained within the study area. The yellow box in the top left corner shows a schematic overview of the Rhine and Meuse River basins. (Derived from Rijkswaterstaat n.d).

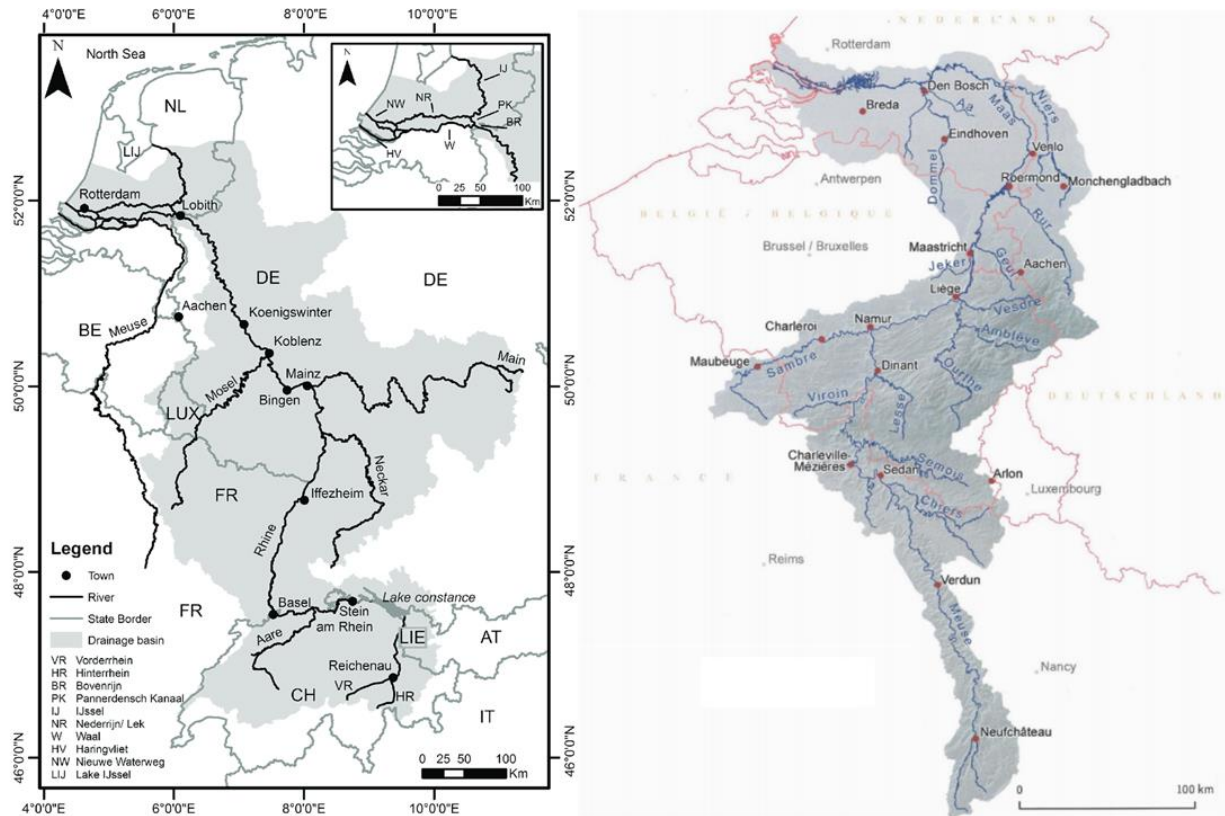


Figure 2: Left: Map of the rhine catchment (Frings et al., 2019). Right: map of the Meuse catchment (Internationale Maascommissie, 2013).

2.4 Model development

Two approaches were taken in model development, a location specific model (LocSpe) and a general model (Isildur). Pre-processing of the training and testing data for RFs both use a powertransformer to normalise the data. Furthermore, in both cases the outliers were removed from the dataset. Additionally, the training and testing datasets were both split into a target dataset and a predictor dataset. The differences in pre-processing the data mainly relates to the way the data was split into a training and a testing dataset.

2.4.1 Location specific model (LocSpe)

The approach taken in the location specific model was to construct an individual RF model for each individual location. The RF model was trained and tested on data for each individual location. In other words, each model was developed using one single location. The data was split chronologically into a training and testing dataset according to a distribution value. This distribution value was determined based on the results of the sensitivity analysis. The training dataset was made up of the first portion of data and the testing data was made up of the latter part of the data. These models are likely to perform well on the locations it was trained on, but poor on locations it has never encountered before.

2.4.2 General model (Isildur)

For the construction of a general model, a “one model to rule them all” approach was taken. Meaning that a single large RF was constructed which could make predictions for locations it had never encountered before. The training and testing datasets of Isildur are constructed by combining all data for multiple locations into a single file. For the testing set one location was selected per river (Figure 1 blue dots), all other locations were used for training. This model is likely to result in poorer predictions for individual locations but generalizes better on locations it was not trained on compared to LocSpe.

2.4.3 Sensitivity analysis of Random forests tuning parameters

The tuning parameters for RF are the number of trees constructed, the maximum depth of each tree, training and testing dataset distribution, and the minimal split value. To select the optimal values for these tuning parameters a sensitivity analysis was performed iteratively for each parameter. These sensitivity analyses were performed for two criteria, the RMSE between the test data set and the model predictions, and a timeseries split cross validation. Based on the results of the sensitivity analysis on the two mentioned criteria the optimal value was chosen for the tuning parameters.

2.4.4 Model training and evaluation

To evaluate the performance of LocSpe an RF was developed for each location indicated in Figure 1 except the blue dots. For each location the RMSE and the r^2 score were calculated. To test how well LocSpe performs on unfamiliar locations, four locations were selected. These target locations are: Driel (T2), Grave (T4), Katerveer (T1) and Sint-Andries (T3). The number in brackets indicates the location on Figure 1. Predictions for these four locations were then made by RF models already trained on different locations. These models were developed for Amerongen(D2), Sambeek (D4), Tiel (D3) and Zutphen (D1). Again, for the four test locations the RMSE and the r^2 score were determined to evaluate the model.

To evaluate the performance of Isildur an RF model was trained on all locations indicated in red on Figure 1. Similarly to LocSpe, Isildur was evaluated on familiar and unfamiliar locations. The unfamiliar locations were the same ones used to test LocSpe, so the blue dots. For familiar locations Amerongen(D2), Sambeek (D4), Tiel (D3) and Zutphen (D1) were used. Once more for each test location the RMSE and the r^2 score were determined.

2.5 Damage estimation

To estimate the damage of a drought a simple relation between sailing depth and cargo transport was used derived by Jong (2019) which I shall now briefly explain. When sailing depth is low, barges travelling along the river may need to take on a reduced load. To still be able to transport all cargo, barges will then need to sail more often. This may cause the shipping costs to increase by a factor M according to the following relationship:

When

$$\begin{aligned} sd \geq d + kv & M = 1 \\ sd < d + kv & M = \frac{t-t_0}{sd-kv-t_0} \end{aligned}$$

$$sd - kv < t_0 \quad M = 0$$

sd = sailing depth, d = ship depth category, kv = keel variation, t = depth of a ship, t_0 = depth of an empty ship, M = multiplier

The extra sailing costs can be calculated by multiplying M by the average shipping costs per depth category per day in a year unaffected by low water and subtracting the normal sailing costs.

To estimate the damages of a drought in terms of shipping costs, the predicted sailing depth is used as input for the formulas described above. For the other variables in this function the following values were used. Ship depth category (d) is taken from data containing all shipping that occurs at a given location. kv is taken to be 0.2 m, t_0 is taken to be 1.3 m and t is taken to be equal to d .

In the above described method, it is assumed that sailing will always take place. However, this is not the case. If M has a value of X , it means that each ship can only transport $1/X$ amount of cargo compared to a normal situation. At certain threshold of M (M_{max}) it no longer is viable to sail. If M supersedes M_{max} , it is assumed no shipping will take place at all. To put this into a formula:

When

$$\begin{aligned} M > M_{max} & M = 0 & L = 1 \\ M < M_{max} & M = M & L = 0 \end{aligned}$$

Multiplying the limit (L) by the transported cargo per depth category results in all non-transported cargo per day due to reduced sailing depth. This cargo can then either be transported differently such as by train or by road or the cargo could not be transported at all, which is called a modal shift (Schasfoort et al., 2019). The added transport costs due to the modal shift causes prices to increase and in turn are counted as economic damages, however these economic damages for consumers are not considered in this paper. The value of cargo that is not transported at all can be considered as direct economic damage. Indirect economic damages due to no transport occurring is not considered in this paper.

By focussing primarily on the major chokepoints on the shipping routes a good estimation of the damages for overland shipping can be made. These chokepoints are Nijmegen and Sint Andries for the Waal. Since almost every cargo ship needs to pass by these chokepoints, it was not necessary to model the entire river network of the Netherlands to estimate the damages to the shipping industry. The main differences between these two points are shipping costs and the cargo transported by depth category.

2.6 Future predictions

To make sailing depth predictions for the future artificial data was used as input for the model, instead of historical data. Most artificial data on discharge predictions is calculated for monthly discharge data, which is not suited for the RF model. Therefore, data will be generated using a delta method. For a climate change scenario of 2 degrees global warming it is predicted that discharge of the Rhine and Meuse will on average be 15% lower compared to present values (van der Wiel et al., 2019). To estimate the impact of a future drought the discharge of both the Rhine and Meuse are lowered by 15%. This method is in line with current climate scenario calculations done by the KNMI. For this purpose, the recent drought of 2018 was used to study the impact of a similar drought in a 2 degree Celsius climate change scenario.

3. Results

3.1 Sensitivity analysis and model development

3.1.1 Tuning parameters of location specific model

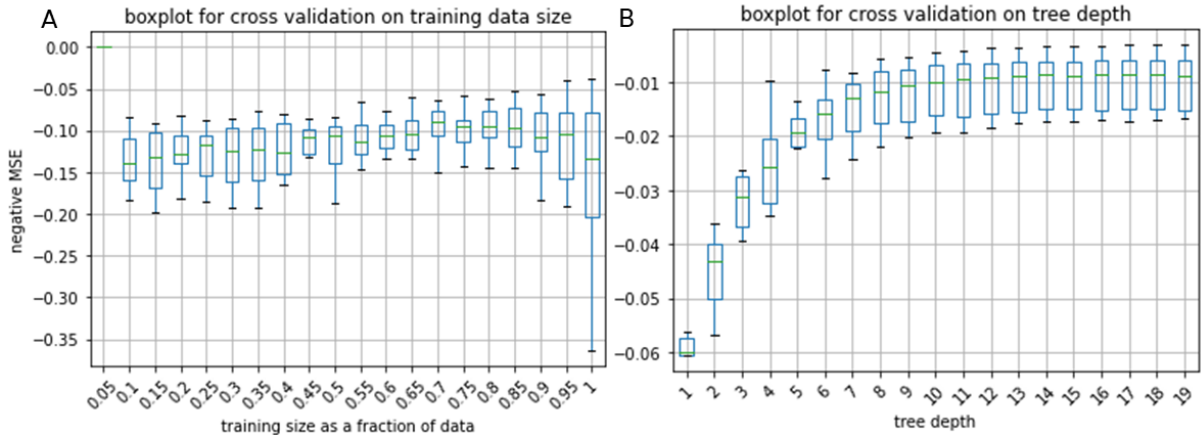


Figure 3: Results of the sensitivity analysis on the tuning parameters of the location specific model RF displayed as boxplots. The following tuning parameters are displayed in the following graphs: A) training data size fraction, B) tree depth, The results are given in negative mean squared error, where values closer to zero represent a better fit.

A sensitivity analysis was performed for the tuning parameters of LocSpe (Figure 3). All results are achieved from a fivefold timeseries split cross validation sensitivity analysis and are given in the unit of negative MSE. The sensitivity analysis was performed iteratively for each tuning parameter. After a sensitivity analysis was performed the optimal result was used in the model for the next sensitivity analysis. The analyses were performed in the following order, training size, tree depth, number of trees and minimal sample split.

For the training size fraction, the lowest error is achieved for a fraction of 0.7 (Figure 3A). The largest relative difference in median error for the performed sensitivity analysis was achieved for tree depth (Figure 3 B). A rapid decrease can be seen for tree depths up to 12 after which the error remains stable. For the number of trees there is significant variation in median error up to a value of 180 trees, after which the median error becomes somewhat stable. The lowest relative difference in median error was found in the analysis on minimal sample split (Appendix Figure 1). The error remains roughly constant although slight increases for higher minimal sample split values occur. The differences in absolute error are not important since these are dependent on the order in which the sensitivity analyses were performed.

The interquartile ranges for the sensitivity analysis on both the number of trees and minimal sample split remain roughly constant. Whereas the interquartile ranges for the sensitivity analysis on both the number of trees and minimal sample split remain roughly constant, the interquartile range for the other tuning parameters vary. For the training data size fraction, the lowest interquartile range results from intermediate values of the tuning parameter, while the highest ranges occur for both the lowest and highest value of training data size fraction. For tree depth the interquartile range varies slightly up to a tree depth of 12, after which the ranges remain constant.

In the end the tuning parameters were selected which resulted in the lowest error and smallest interquartile range (Table 1)

3.1.2 Tuning parameters of Isildur

As with the LocSpe model a sensitivity analysis was performed on the tuning parameters of Isildur. The model was trained and tested on the entire datasets for multiple locations instead of splitting the data of a single location, as has been done for the LocSpe model. The results were achieved by means of a fivefold timeseries split cross validation (Figure 4). No sensitivity analysis was performed on training size fraction because this tuning parameter is not used in Isildur.

The largest reduction in median error and interquartile range in both absolute and relative terms was achieved for tree depth (Figure 4A). Here the median error and the interquartile range both decrease following a log like pattern, until stabilizing after reaching a tree depth of 16.

For the minimal sample split and number of trees the median errors remain roughly constant with increasing parameter values (Appendix Figure 2). For the minimal sample split the median error as well as the interquartile range tends to increase with increasing parameter values. For the number of trees constructed the interquartile range and the median error remain roughly constant, while stabilizing after 200 constructed trees.

The tuning parameters chosen for Isildur are based on the results in which the smallest overall model median error and the smallest interquartile range was observed (Table 1).

Table 1: Selected tuning parameters for both the location specific model (LocSpe) and Isildur.

Tuning parameter	LocSpe	Isildur
Training data size fraction	0.70	-
Tree depth	12	16
Number of trees	180	200
Minimal sample split	3	3

3.1.3 Transformer selection

Two types of transformers are available to alter input data of an RF to be gaussian, these transformers are quantile and powertransformer. The best transformer results in the closest match between the observations and the model predictions. To this end a power transformer using the Yeo-Johnsen method was selected as the match between the observations and the predictions are more accurate (Figure 5).

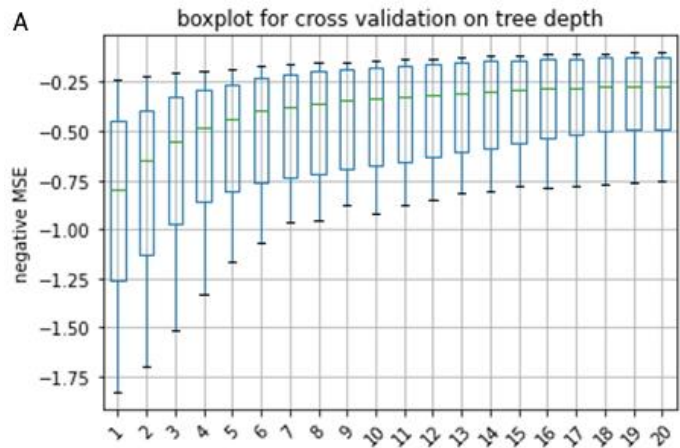


Figure 4: Results of the sensitivity analysis on Tree depth for Isildur. The results are given in negative mean squared error, where values closer to zero represent a better fit.

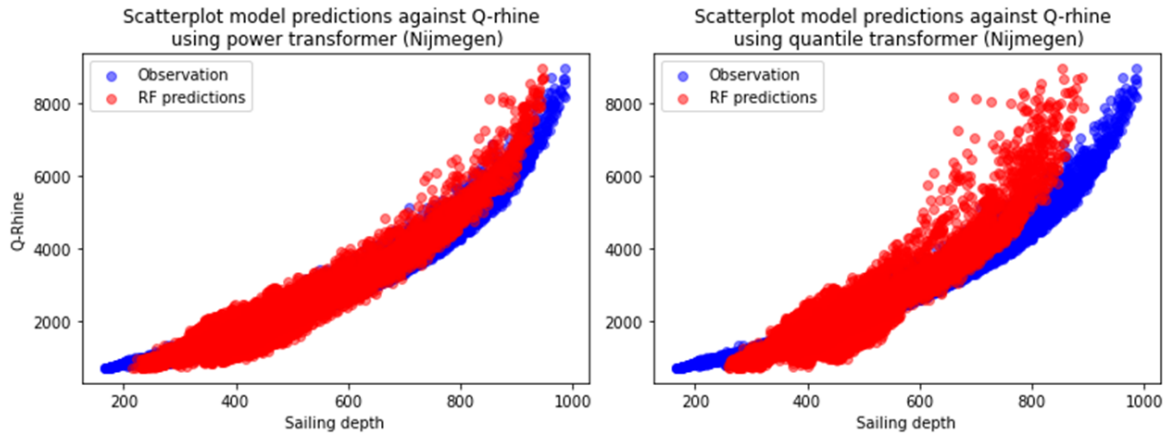


Figure 5: Scatterplots comparing the model results and the observations to the discharge of the Rhine for Nijmegen (which is not connected to Meuse)

3.2 Location specific model results

3.2.1 Location used for training/familiar locations

The LocSpe model was trained on a single location on the chronological first 70% of the data and tested on the remaining 30%. This distribution is based on the results of the sensitivity analysis on training size as can be found above (Table 1). For the target locations, the model is tested for time periods the model was trained on and for time periods it had not yet encountered.

For every location an RF model was trained and tested. Each RF model was developed on average of 12760 datapoints where 8932 datapoints were used for training and the remaining data was used for testing. The average model score for all locations where an RF model was developed is $r^2 = 0.8094 (\pm 0.100)$, the average RMSE between the model predictions and target value is equal to 31.57 cm ($\pm 8,16$). These metrics were calculated for the entire model, so including both the periods the model was trained for and where periods the model was not trained for.

Results for familiar target locations and for a period the model was also trained on are accurate (Figure 6). The model tends to accurately predict the sailing depth. However, the most extreme values, so the highest and lowest values tend to be underestimated. For locations the model was trained on but for a period the model had not encountered before the model performances are less accurate

The model predictions tend to be more accurate for locations in close proximity to a sluice or weir. For both Amerongen and Sambeek the sailing depth predictions match better with the observed results compared with Zutphen and Nijmegen. The former two locations are located near a lock whereas Zutphen and Nijmegen are not. This observation holds for predictions made for trained and untrained periods.

3.2.2 Unfamiliar locations

To test how well the LocSpe model performs on unfamiliar locations, four different locations, not specifically trained for, were selected. These target locations are: Driel, Grave, Katerveer and Sint-Andries. Predictions for these four locations were then made by RF models already trained on different locations. These models were developed for Amerongen, Sambeek, Nijmegen and Zutphen. For each target location four different RF predictions were made using different RF models.

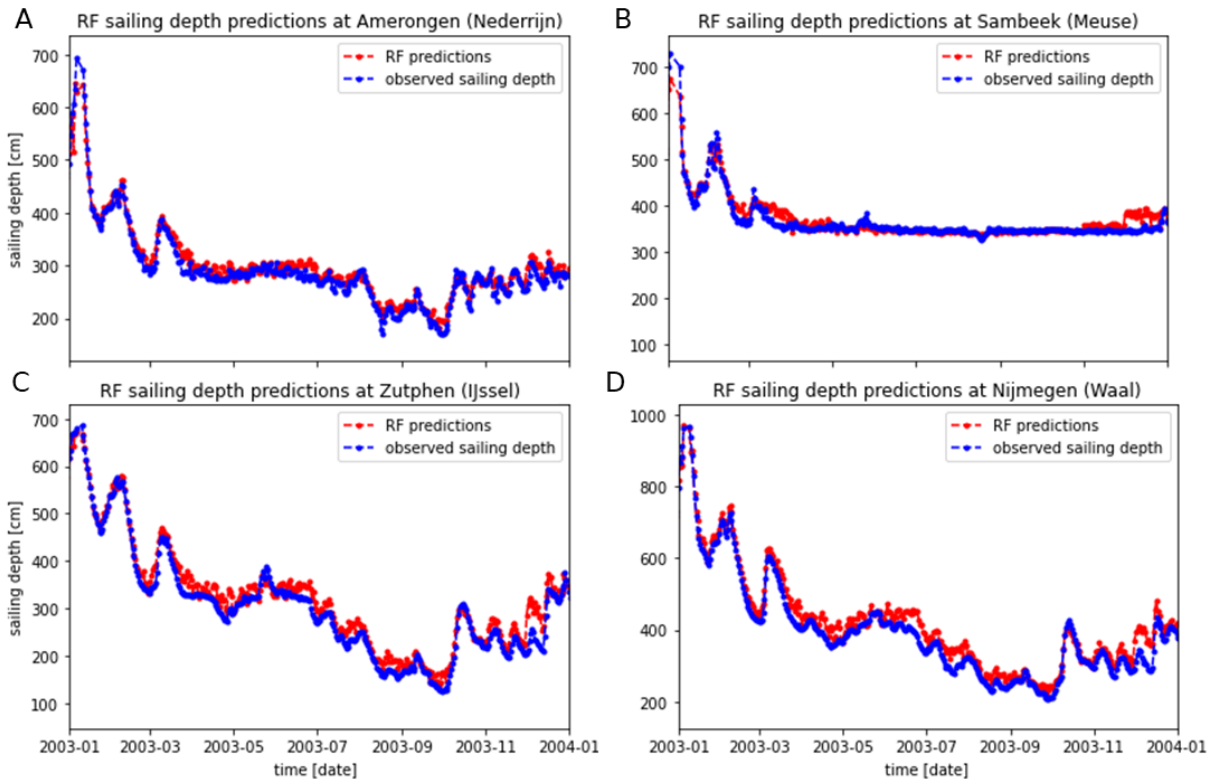


Figure 6 LocSpe Sailing depth for A) Amerongen B) Sambeek C) Zutphen and D) Nijmegen. For each location a separate RF model was trained. The results shown in the plot are for a period the model was trained on, so it has encountered this data before. In red are the results of the RF model results, in blue the observed sailing depths. The depths are measured in centimetres.

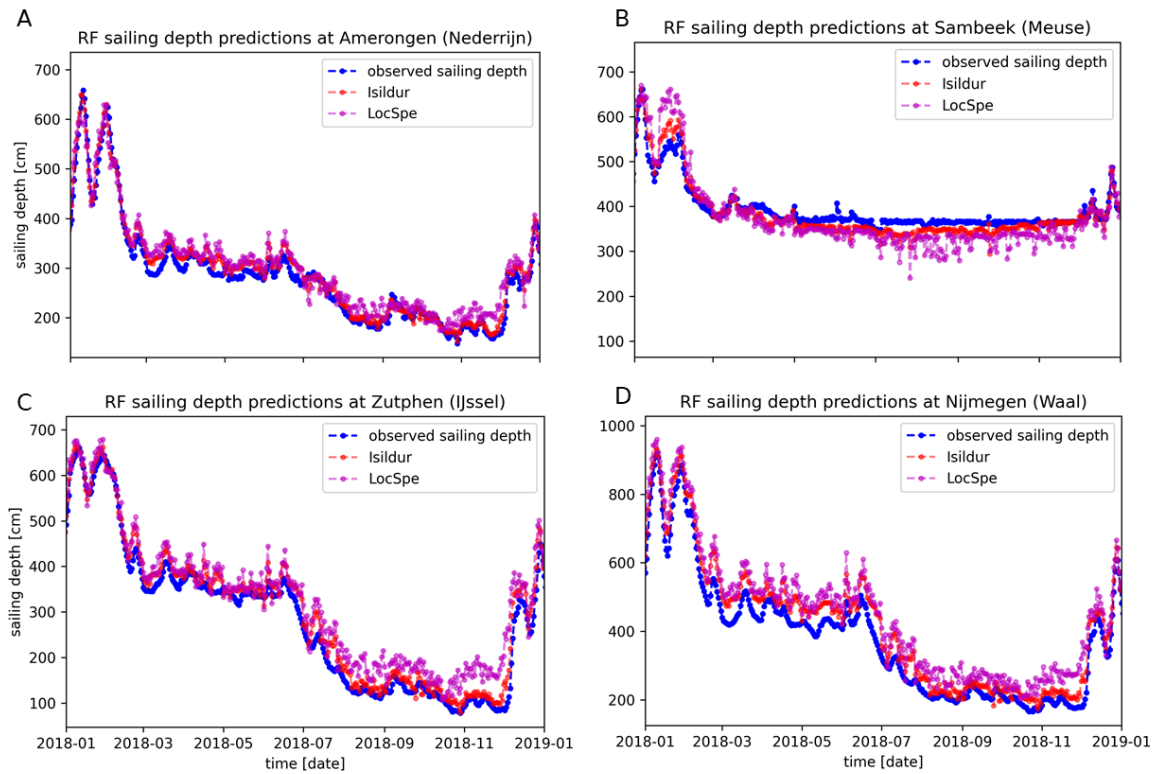


Figure 7: LocSpe and Isildur results for locations included in the training data set. For LocSpe the plots are shown for a period outside of the validation period. In blue the observed sailing depth, in red the Isildur predictions and in purple the LocSpe predictions are given.

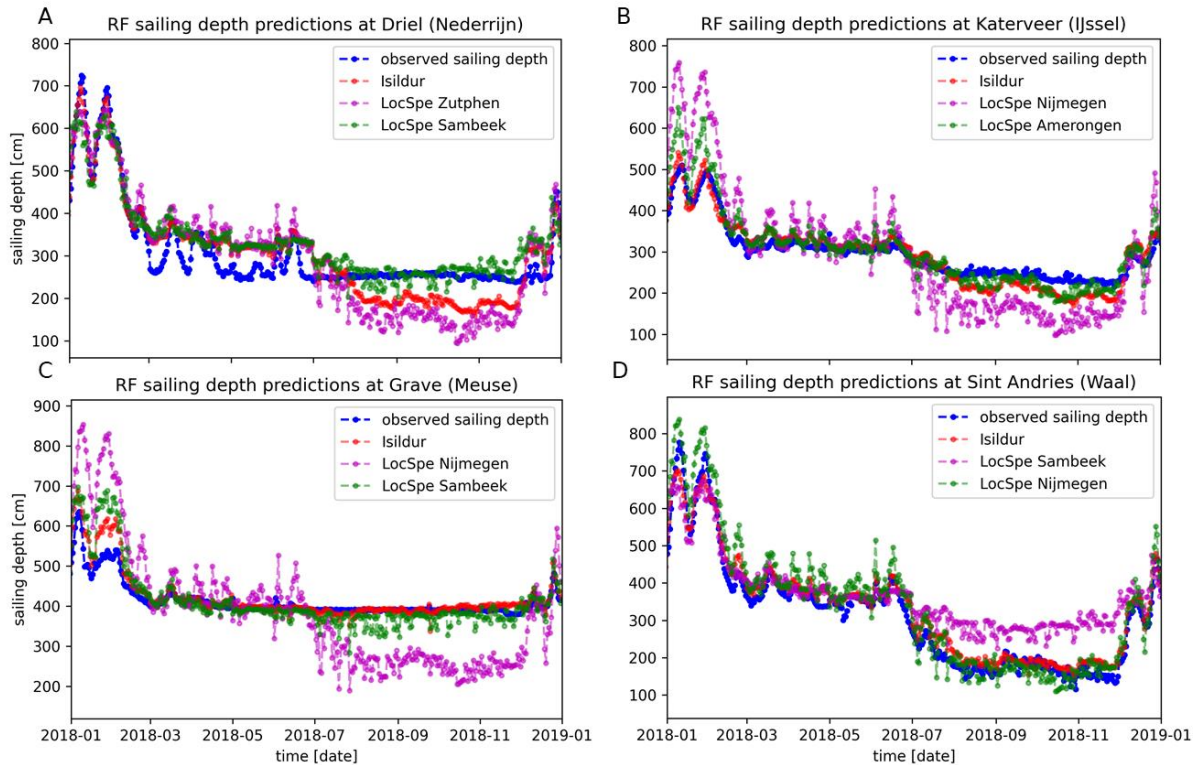


Figure 8: *LocSpe* and *Isildur* results for locations excluded in the training data set. For *LocSpe* the plots are shown for a period outside of the validation period. In blue the observed sailing depth, in red the *Isildur* predictions and in purple and green the *LocSpe* predictions are given.

There is great variation in the model performance based on which model was used to make predictions at each location (Table 2, Figure 8). The difference between the best and worst performance is often a factor of two or more for both the r^2 score and RMSE. The best model results are comparable to the results of a model trained on the target location (Figure 8). While the worst results often have a poor fit to the target data. The best results are achieved by using a model developed on a location in the same river or river branch. The poorest results are the result of using a model developed on a totally different river. For instance, the best results for Sint Andries (T3) are achieved by using the model developed for Nijmegen (D3) (Figure 1).

The difference between the worst performing and second worst performing model-location combination is often relatively small. For Sint Andries for example the RMSE differed only 2,9 cm depending on whether predictions were made by the model trained on Amerongen or Sambeek. Similarly for the best results a comparable observation is made where the best and second best result also differ slightly. Again, models used to predict targets located in the same river or a connected river branch often resulted in closer matches. The best result for Grave, which was achieved by using the model trained on Sambeek, are at least twice as accurate as the other predictions, which were trained on the Rhine branches.

3.3 *Isildur* results

3.3.1 Locations used for training

Isildur was trained using the entire dataset for 10 locations, adding up to about 117402 data points. The average model score is $R^2 = 0.8742$, the average RMSE between the model predictions and target value for locations it was trained on is equal to 29,03 ($\pm 10,25$) cm. The match between the RF predictions and the observed data is close for each location (Figure 7). The RF model is already familiar with input data for these locations both in terms of the location itself and the period. The match is comparable to the match between the observed values and RF predictions shown in Figure 6. At Nijmegen the model has the poorest fit to the observed data, however it is still a close match (Figure 7D).

Table 2: Results of the location specific model on locations the RF model is unfamiliar with.

Testing Location	RMSE (cm)		r ² score		Training location	
	Lowest	Highest	Highest	Lowest	Lowest error	Highest error
Driel	32,36	69,26	0,7794	0,3893	Nijmegen	Sambeek
Grave	17,96	32,5	0,6731	0,1398	Sambeek	Amerongen
Katerveer	19,46	37,06	0,7970	0,2515	Zutphen	Amerongen
Sint-Andries	33,25	79,83	0,9360	0,4146	Nijmegen	Sambeek

3.3.2 Unfamiliar locations

Isildur was tested on the same four locations used for testing LocSpe, which are Driel, Grave, Katerveer and Sint-Andries. These four locations are each located along a different major waterway in the Netherlands (Figure 1). The RF model was not trained on these four locations. In total these locations add up to 50342 datapoints, which is equal to 30,01% of the entire dataset used in testing and training the model. The model score is equal to $r^2 = 0.7792$. The average RMSE is equal to 32,92 ($\pm 10,25$) cm for locations the model has not encountered before.

The prediction made at Driel also underestimate the sailing depth, (Figure 8A). Here there is an intersection between rivers flowing through the Amsterdam-Rijn-channel and the Nederrijn nearby. Driel is the only location located near such a crossing in either the training or testing data set. At such a crossing of rivers the water height needs to be controlled tightly. The predictions for Sint Andries and Grave both match well with the observations (Fout! Verwijzingsbron niet gevonden. B&D). Initially at Katerveer the RF prediction significantly underestimated the sailing depth compared to the observed values (Fout! Verwijzingsbron niet gevonden. C). The observations at Katerveer are made near a sluice. The observations at the other location used for training the model for the IJssel were not. Although there is no sluice located in the IJssel itself, adding one in the predictor data yielded significantly better results.

3.4 Damage estimation results

Using the method described in Jong (2019) the damages due to lower sailing depth could be calculated. These calculations were performed for Sint-Andries and Nijmegen because these two locations are the major chokepoints along the busiest sailing route in the Netherlands. As input for the method used, as described in the methods sections, the outputs of both the LocSpe and general model were used. For the predictions made by Isildur for Nijmegen, the location was taken out of the training data, and replaced with a different location for the Waal. Additionally, the measured sailing depth was also used to calculate a true damage estimate, to which the estimates based on ML predictions can be compared (Figure 9A).

The estimates for cargo transported by means of shipping are relatively accurate for Sint Andries (Figure 9 & Table 3). The best predictions are made by LocSpe at Sint-Andries (Figure 9C). Both the error and the r^2 score are similar regardless of the model used (Table 3). The shape of the sailing costs curve matches closer to LocSpe predicted sailing costs than to Isildur predicted sailing costs (Figure 9). Cargo estimates at Nijmegen are less accurate than those for Sint Andries. Again, the best results are for LocSpe.

For all models the sailing costs calculations are less accurate than the cargo not shipped estimations (Figure 9 & Table 3). For Nijmegen no model has a positive r^2 score, which indicates a poor match. For Sint-Andries the best results are achieved using LocSpe trained on a different location.

Table 3: Results of damage calculations in terms of the RMSE and r^2 score for different modelling approaches.

Model	Nijmegen				Sint-Andries			
	Sailing costs		Tonnage		Sailing costs		Tonnage	
	RMSE	r^2	RMSE	r^2	RMSE	r^2	RMSE	r^2
Isildur	1,218	-0,515	0,138	-0,245	0,092	0,212	0,01	0,888
LocSpe familiar	1,107	-0,251	0,135	-0,186	0,096	0,145	0,009	0,908
LocSpe unfamiliar	1,145	-0,338	0,122	0,037	0,091	0,231	0,01	0,895

The cumulative sailing costs at Nijmegen for the damages based on observed water depth values is equal to 1243,49 million euros and in total 114,19 million tons of goods were transported by means of shipping. These values are 53.85 million euros higher in sailing costs and 36,68 million tons less cargo transported compared to a year without drought. No RF based predictions matched these values for Nijmegen (Table 4). The models overestimate the shipping costs and the shipped cargo. For Sint Andries the LocSpe model predictions are off by only a few percent for sailing costs and cargo shipped. The results based on Isildur are also more accurate for Sint Andries than for Nijmegen.

Table 4: Results of the damage estimation for the observed values, the values predicted based on Isildur and the location specific model. The bottom two rows are for a 15% decline in discharge due to climate change.

Cumulative Damages In millions	Observed		Isildur		LocSpe			
	Nijmegen	Sint Andries	Nijmegen	Sint Andries	Trained on location		Not trained on location	
					Nijmegen	Sint Andries	Nijmegen	Sint Andries
Sailing costs (euros)	1243,49	76,92	1335,19	87,63	1307,5	78,44	1325,64	70,49
Cargo shipped (tons)	114,19	15,01	140,56	16,35	139,99	15,49	139,07	14,9
Climate change sailing costs (euros)	/	/	1360,93	82,16	1342,99	73,83	1337,57	68,84
Climate change cargo shipped (tons)	/	/	138,58	15,61	137,09	14,89	135,45	14,47

Note: No observed data is displayed for the climate change scenario, because this data does not exist. More information about the climate change scenario results can be found in chapter 3.5 Future predictionsA

3.5 Future predictions

By incorporating the artificially generated data by means of the delta method into both the LocSpe and Isildur models predictions for sailing depth and damages can be made for a climate change scenario. These predictions were made for Nijmegen and Sint Andries, as these locations have the largest impact on shipping in the Netherlands.

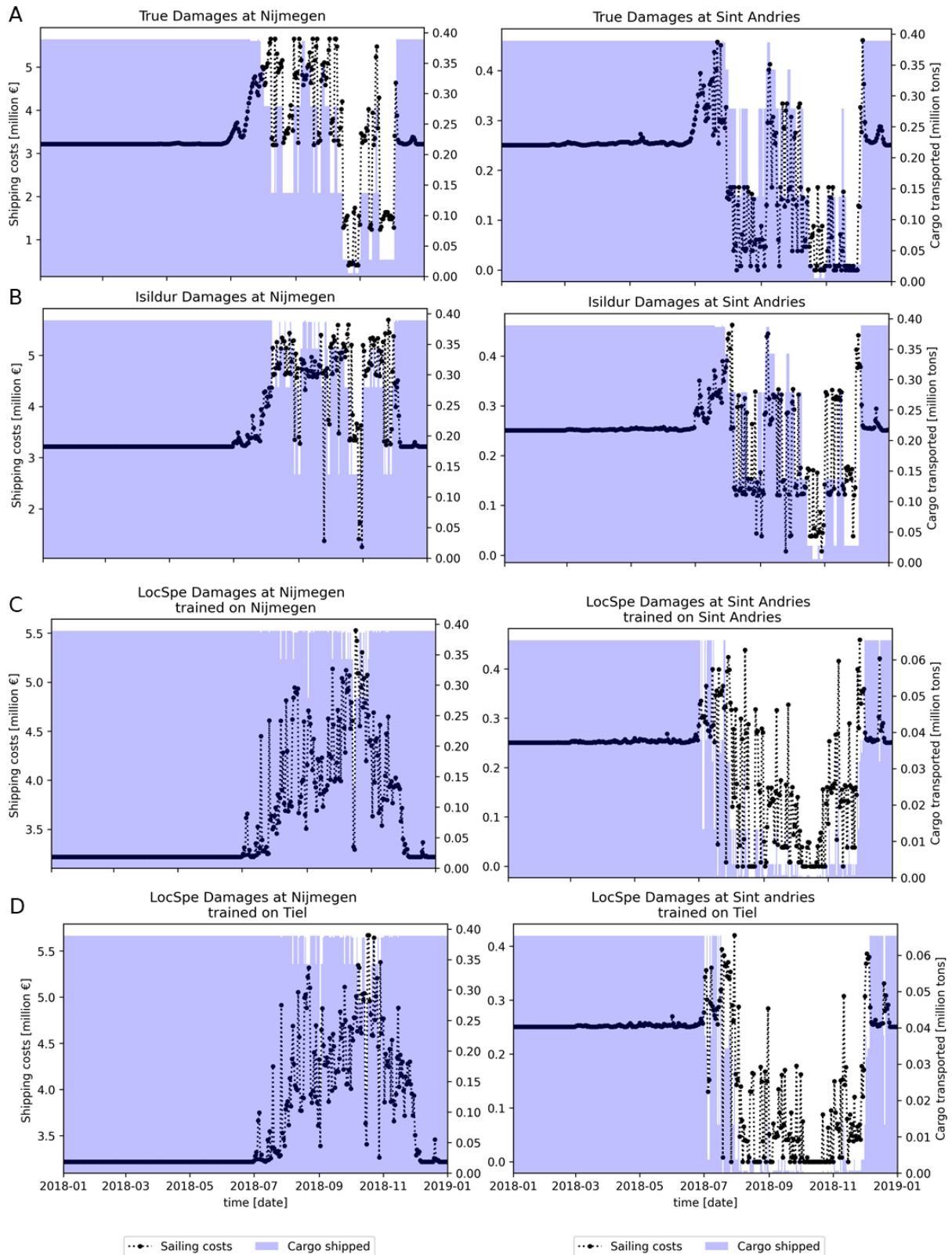


Figure 9: Estimations of the damages due to the hydrological drought of 2018 in the Netherlands. A) shows the estimation based on the observed depth data. B, C and D show the estimations of the damages in terms of sailing costs and tonnage not shipped for Nijmegen and Sint Andries. B) results from Isildur, C) results from the location specific model trained on target location, D) results from the location specific model trained on different location.

All sailing depth predictions for drought in a 2 degrees Celsius climate change scenario are lower than the original estimates (Figure 10). The difference between the normal and the climate change predictions are lowest for Isildur. Here the climate change predictions are only slightly lower compared to the original predictions. Both the LocSpe predictions estimate significantly lower sailing depths in future droughts. Remarkably for Sint Andries the LocSpe trained on target locations predicts lower results than the LocSpe trained on Tiel while the opposite is true for Nijmegen.

The estimates of the damage due to drought all increase in severity in the climate change scenario. At Sint Andries the number of days that sailing is not possible at all increases according to both the LocSpe models (Figure 11 B&C). Accordingly, the cumulative shipping costs also drop significantly while the cargo not transported increases (Table 4). For Nijmegen the most realistic results are from LocSpe, however still not matching the original damage of 2018. Both LocSpe models predict an increase in both cargo not shipped and sailing costs.

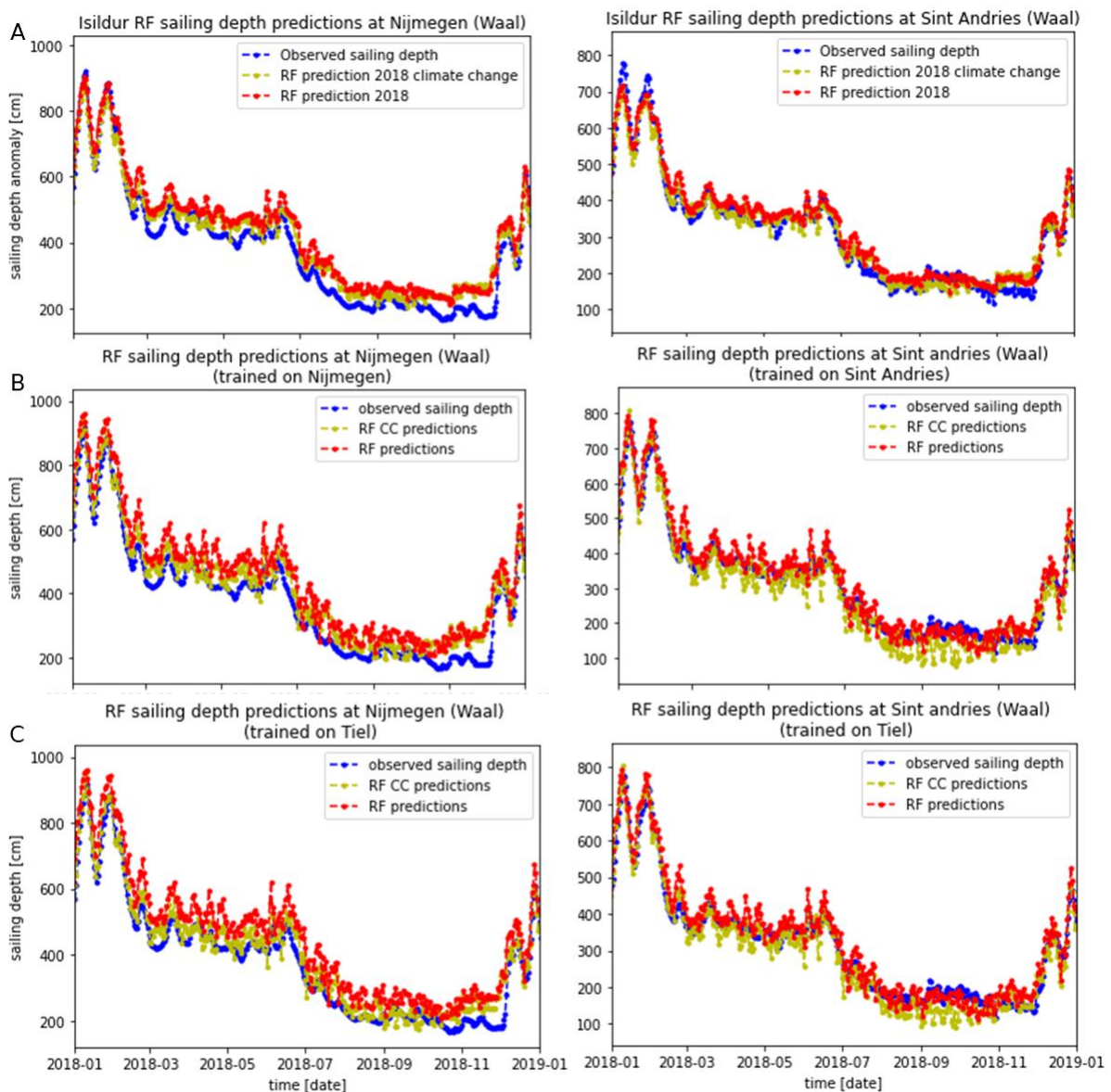


Figure 10: Future estimates of sailing depth at Nijmegen and Sint-Andries for a 2 degree Celsius increase in temperature. Yellow: Future RF predictions. Red: Current RF predictions. Blue: Observed sailing depth 2018. A) General model B) LocSpe trained on location C) LocSpe trained on different location.

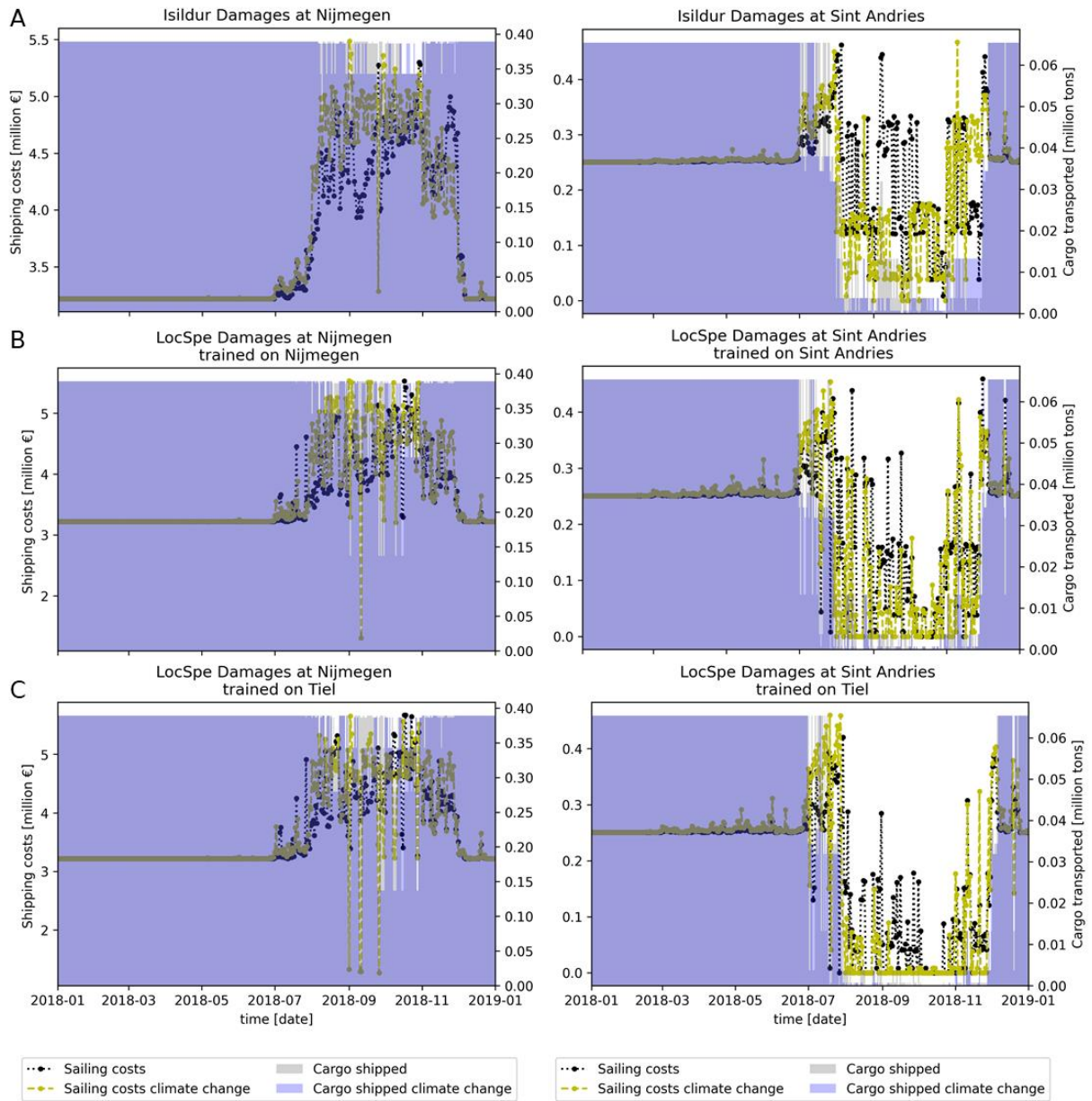


Figure 11: Damage estimate for the climate change scenarios for Nijmegen and Sint Andries compared to the damage estimation for the normal climate scenario predictions. A) Isildur B) LocSpe trained on location C) LocSpe trained on different location

4. Discussion

The aim of this study was to investigate to what degree ML methods can be used to study the impact of hydrological drought on the shipping industry. To that end two RF models were constructed to model the sailing depth, which has a large impact on the shipping industry. The first model is based on constructing individual RF models for individual locations and for the second model one single large model was developed for multiple locations. First the ability to reconstruct historic data was checked for both models. Later this was expanded to include predictions based on climate scenarios by the KNMI. The Sailing depth predictions were then used to predict the potential damages for the shipping industry.

The two RF approaches used in this paper are relatively quick, a single run of the program took only a few minutes, which is in line with previous observations (Adnan et al., 2021; Hauswirth et al., 2021; Tyralis & Papacharalampous, 2019). This relatively short runtime makes adapting the model easier compared to traditional numerical models which take much longer. Which in turn makes it easier to use for monitoring and research purposes compared to traditional hydrological models.

4.1 Reconstruction of historic data

For LocSpe, the best results were achieved for locations the model was trained on when reconstructing historic data included in the training set (Figure 6). These results almost perfectly match the historic data. This is not surprising, since the RF is already familiar with the target data. It is similar to giving the same test to a student as the practice and official exam, the student will be familiar with the material and will perform better than it normally would. Therefore this is of little interest for the aims of this study.

Both modelling approaches yield close matches between the target data and the predicted data. LocSpe yields better predictions for locations where a model was developed compared to Isildur. However, Isildur yields better results than LocSpe on locations where no LocSpe model was developed. In other words, Isildur generalizes better than LocSpe.

The LocSpe model struggles more when it comes to reconstructing the lowest values than Isildur (Figure 7 C&D). This is the case when the target depth is outside of the training data range. For example, at Nijmegen the lowest sailing depths occur in 2018, and LocSpe struggles to reconstruct these as this data is excluded from the training data set. For Isildur these values are included in the training data set, and this model can therefore better reconstruct these low values.

Previous work suggests that RF struggles with reconstructing the extremely low or high values (Hauswirth et al., 2021). By implementing different methods of training data selection, the accuracy of the predictions for the extreme values were improved, although the model still had difficulty with the lowest and highest values.

The accuracy of predictions made by LocSpe for unfamiliar locations relies heavily on whether the target location is along the same river (Table 2). Furthermore, predictions also tend to be better for connected river branches than for separate rivers. These results are likely because the study area is heavily regulated by humans. For the Rhine branches included in this study, there is a rather constant ratio between the discharges of the different branches (Frings et al., 2019). This means that for a model it is relatively easy to make predictions on unfamiliar locations because the relative changes in discharge and streamflow are roughly equal across the different Rhine branches, while the absolute difference may not be. Consequently, the predictor data will be very similar in shape to the training data set. Additionally, all target locations are positioned in a relatively small area where regional differences in climatic conditions are minimal. Furthermore, the catchments of the Meuse and Rhine rivers are also in a similar climate. This means that even though they are two different rivers, there will be some

similarities in the trends in discharges of the rivers. And these similarities also make it easier for a ML model to make predictions on different locations without being trained on them.

The results show that while outcomes of LocSpe may be better for locations it was developed on, Isildur outperforms LocSpe when it comes to predicting locations the model has not yet encountered (Figure 7). Because Isildur generalizes better than LocSpe, this approach may be preferable for a general monitoring set-up. That is not to say that it is not possible to use the latter for monitoring. The predictions made by LocSpe tend to be accurate for other locations along the same river. Therefore, it is possible to set up a predictive monitoring scheme by training one RF model for each river of interest. This will however require more work compared to developing a similar monitoring scheme using Isildur for all rivers.

Another option is to use a combined approach. Use an Isildur to monitor the water transport network on a large scale. Then use LocSpe for the most important choking points. Isildur's prediction can then be used to broadly monitor the network. LocSpe can then be used on the most important locations, such as Nijmegen and Sint Andries, to make more accurate predictions for the shipping industry.

4.2 Damage estimation

LocSpe yields better results when it comes to predicting the damages of a hydrological drought. This is likely because LocSpe makes better predictions for locations it was trained on than Isildur. These small differences in model performance result in large differences in sailing costs or tonnage not shipped (Figure 12). Changes in available sailing depth changing from 4 to 3 meters will result in much less lost cargo carrying capacity than a sailing depth change from 3 to 2 meters (Figure 12).

The negative relation between sailing costs and cargo transported found in Table 4 for Sint Andries can be explained by the following process. As the sailing depth decreases the sailing costs initially increases, however at a certain threshold, ships of a given depth category cannot sail at all. Therefore, these ships make no sailing costs. (Appendix Figure 1 and 2). This process also explains the low extra sailing costs at Nijmegen according to the observed data compared to the model results. As here the sailing depth is overestimated by both model approaches.

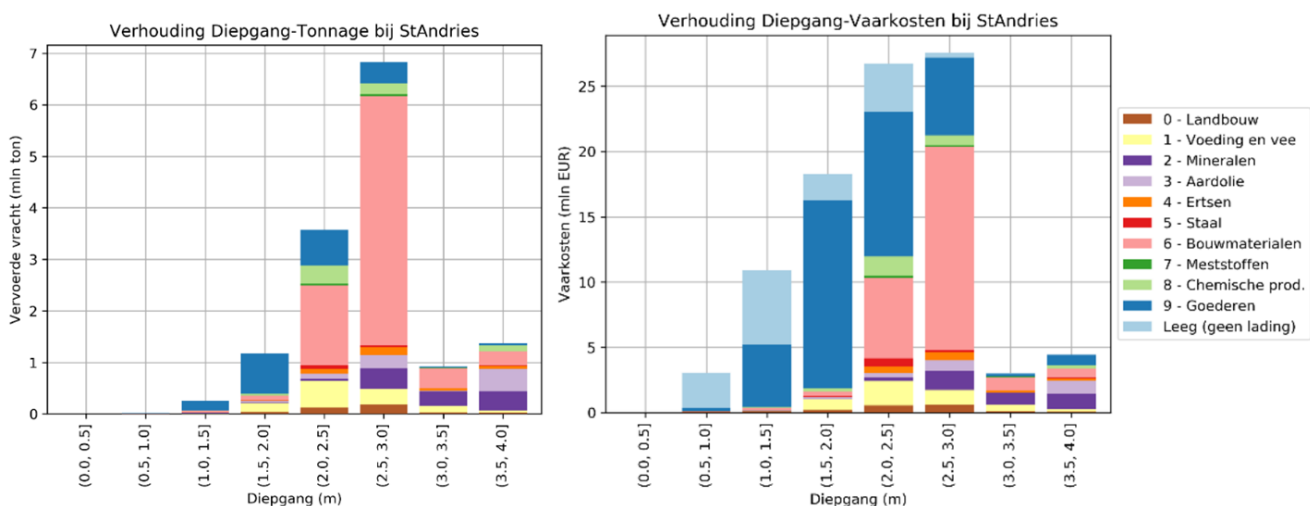


Figure 12: Yearly tonnage shipped (left) and yearly sailing costs (right) per depth category at Sint Andries. Type of good transported is displayed in both bar charts (Jong, 2019)

The total damage for the shipping industry for the drought of 2018 is estimated to be between 160 and 345 million euros. For the upper estimate about 165 million euros is due to direct economic damages while the remaining damages are due to indirect effects (Hekman et al., 2019). The method used in this paper estimates that for the Upper Rhine-Waal actual damages are 53.85 million euros in extra shipping costs and 36.68 million tons of cargo is not shipped (Table 4). Keep in mind that these damages are only

for the Upper Rhine-Waal, and do not consider the other rivers in the study area. The Upper Rhine-Waal is responsible for half of all shipping costs in the Netherlands (Schasfoort et al., 2019), so in reality the estimated sailing costs will be twice as high.

The most accurate predictions of damage based on ML outcomes are 119.16 million euros in extra sailing costs and 11.12 million tons not shipped. These predictions are made using LocSpe trained on target location. The large difference between the observed values and calculated damages is mainly due to inaccuracies in sailing depth predictions at Nijmegen.

As stated above the predictions for Sint Andries are rather accurate. This means that it could be possible, with future work, to improve the results for Nijmegen and increase prediction accuracies. If this could be achieved, the development of a damage estimation program based on the method described in this paper is feasible. Such a model would be a very useful tool for water managers and researchers alike.

This paper does not consider all measures that the shipping industry could take to lessen the impact of hydrological drought. The main difference is that modal shift is not considered in the damage estimations (Jong, 2019; Schasfoort et al., 2019). It is assumed that if shipping becomes unfeasible or unprofitable that the cargo is simply not shipped, while different means of transport could also be used, such as transport by road or by train. These different modes of transport will be able to reduce the amount that is not shipped. Other factors that are not included in the damage estimation are: increasing sailing efficiency, storing of goods not transported, changing fleet composition, decreases in sailing speed due to congestion and changing sailing times. Implementation of these measures would yield more accurate results in predicting sailing damage.

4.3 Future predictions

4.3.1 Future sailing depth predictions

Both LocSpe and Isildur RF approaches yield lower sailing depths as a result of reduced discharge due to climate change. This means that the approach of using RF to study the impact of future hydrological droughts is possible. The degree of reduction of sailing depth does vary between the different modelling approaches. The difference in sailing depth predictions between the normal and climate change scenarios is relatively minimal for Isildur, while the difference is larger for the LocSpe predictions. Since it is not possible to benchmark the future, it is difficult to ascertain which approach yields better results.

Van der Mark et al. (2020) presents a traditional numerical method for predicting sailing depth in the Rhine branches. This method is aimed at predicting future water depths for a six to eight week period. It has similar inputs as the models described in this paper. It takes the predicted discharge of the Rhine at Lobith based on model predictions (BfG) and uses this to eventually predict the sailing depth. This is however where the similarities end. The method described in Van der Mark et al. (2020) calculates the expected sailing depth by using multiple models to predict the water height (SOBEK3 and RVWG), which is then combined with the measured and predicted depth profile over the river to estimate the future sailing depths.

The RF models utilised in this paper require less work and less data to make predictions on future sailing depth compared to the traditional approach. However, the RF models presented in this paper are currently not able to match the resolution offered by the traditional hydrological model.

4.3.2 Future damage predictions

All model approaches used in this paper predict that the shipping industry will be negatively affected by climate change. There will be an increase in the cargo that cannot be transported, which can be double that of 2018 (Table 4). Accordingly the number of days with water levels below the target depths of 280 cm also increases (Kroekenstoel, 2014). Subsequently the number of days where sailing is not possible will increase during a hydrological drought in a 2 degrees Celsius climate change scenario compared to the current climate.

Even with a reduced discharge of the Rhine and Meuse the scale of damages are still underestimated by both RF models at Nijmegen. This difference can largely be explained by the extreme low water in 2018 at the end of the drought around November. No predictions match this period of extreme low water. As during this time, the cargo that cannot be transported will be highest, so inaccuracies in these predictions will also result in the highest difference.

These estimates are based on the current amount of traffic along the Upper-Rhine and Waal and do not consider changes in traffic. According to [De Jong et al. \(2019\)](#) it is expected that tonnage that will be transported along the Upper-Rhine and Waal will almost double by 2050. So, in consideration of this increase, damages will be much higher than the estimates made in this paper.

4.4 Improvements, issues and limitations

During the development of both RF model approaches a problem was encountered when transforming the predictor data to be gaussian, however it occurred more frequently with LocSpe. Normally data is transformed to be gaussian and have values between -2 and 2. However it sometimes occurred that all predictor values for certain variables, such as the discharge of the Rhine, were set to zero. This means the RF will not consider this variable during predictions, which leads to significantly worse results.

Related to the above mentioned encountered issue is a general problem with using RFs. RF and many other ML methods are grey or black box algorithms, which means it is difficult to understand what is going on within the program. This usually is not an issue as long as no strange problems arise. If these problems do arise, it can be difficult to find out where the error lies.

It is a known problem of RFs that they struggle to predict the lowest extreme values, especially if they would fall outside of the training data set ([Hauswirth et al., 2021](#); [Hengl et al., 2018](#)). The lowest sailing depth values in the entire training and testing data can be found in Nijmegen for 2018. To fairly compare the damage estimation based on Isildur the target location was excluded from the training data. This means that the data containing the lowest sailing depths was excluded for Nijmegen and consequently the model fails to reconstruct the lowest sailing depth values. However, including this data may lead to unrealistically good results.

All work done in this paper focusses on gauged locations, meaning ungauged locations are not considered. This means that the current results can only be used at certain points along a river or stream instead of modelling the entire stream. [Hengl et al. \(2018\)](#) shows that RF can be used to extrapolate that in spatial or spatiotemporal framework. In the current paper this was not applied, however such a spatial RF (RFsp) could be used to make better predictions at ungauged locations.

Finally, a small note has to be made on the methods used to expand sailing depth training data results in a proxy of the actual sailing data. It would be better to train exclusively on measured and observed training data instead of a calculated value. For example, at Sint-Andries this method results in extremely low water depths for 2018, which may be unrealistic, which then results in unrealistic estimations of damages.

4.5 link to previous work

[Hauswirth et al. \(2021\)](#) investigated the viability of using data driven approaches to model hydrological extremes. They found that more complex ML algorithms, such as RF, can model drought events but struggle to model the most extreme drought events. It was suggested that by changing the training data selection model performance could be improved. The Results of this paper seem to suggest that this is correct. The RF predictions made by both Isildur and LocSpe yield better results for the extreme drought of 2018 compared to what was achieved in [Hauswirth et al. \(2021\)](#). The main difference is that in [Hauswirth et al. \(2021\)](#) data was split randomly while in the current paper it is split chronologically.

Tyralis & Papacharalampous (2019) provide an excellent review of the use of RF in the study of hydrology. The main advantages of RF mentioned there were also found in this paper. RFs were found to be fast, capture non-linear dependencies, were able to handle large datasets, were easy to use and were accurate. The disadvantages of RF were also encountered in this paper, mainly the inability of RF to extrapolate outside of the training data (Hengl et al., 2018; Tyralis & Papacharalampous, 2019).

Lange & Sippel (2020) reviewed multiple ML methods in hydrology and suggest that one of the main reasons for using RF is its resistance against overfitting the target data (Breiman, 2001). In this paper no problem was encountered with overfitting, which is in line with previous observations.

Adnan et al. (2021) found that RF may struggle with predicting or reconstructing data compared to Long-Short-Term-Memory networks (LSTM) or Extreme Learning Machines (ELM). They suggest that in their paper this is likely due to the linear structure of RF. In this paper few problems were encountered with RF and periodicity. It would however be interesting to see how LSTM or ELM compare to the RF results in this paper.

4.6 Recommendations

The aim of this paper was to model hydrological drought, however most of the time there is no drought present, meaning the input data of the RF model is relatively sparse on drought data. Currently, there does not exist an easy way for a RF to focus on this sparse data when it comes to regression. Algorithms have been designed for classification to focus on sparse data and to artificially expand it. However, few of these exist for regression. There are some, for example SMOTE for regression (Torgo et al., 2013). Implementation of such methods could improve RF predictions.

This study used the Rhine and Meuse rivers in the Netherlands as a case study to see if ML methods can be used to study hydrological drought. Based on the findings above the RF model shows great promise to be a good tool in studying the impact of hydrological drought. One advantage of ML methods is that they are flexible and are able to easily learn new relationships (Hauswirth et al., 2021). This means that it is likely that this approach can be applied to different areas than just the one used in this paper.

Results in this paper show that RF can predict future sailing depths and consequently estimate damage to the shipping industry. A next step before RF could be used in practice as a water management tool is to test this by making short term predictions during drought and test whether they are accurate. If this proves to be the case the resulting model could be used to make estimations to compare operational scenarios. These calculations can be done on the fly due to the relatively short runtimes of the program which gives it a significant advantage over numerical models.

Feature importance is not considered in this paper. This metric can give very useful information about what variables are the driving forces behind the model. By taking feature importance into consideration the RF model can also be used to better understand and quantify which processes are most important in hydrological drought (Albers et al., 2016).

There is no single best ML method for modelling hydrological processes (Lange & Sippel, 2020). This is also the case for the subject of this study. As stated above RFs are capable to reconstruct historic data and predict possible future droughts. That this method works well does not mean other ML methods should not be investigated. LSTMs are another promising ML method based on Artificial Neural Networks (Adnan et al., 2021; Hauswirth et al., 2021; Kratzert et al., 2019). These methods are able to take into account previous timesteps when making predictions for the next timestep and could therefore prove a very useful ML method for the prediction of sailing depth.

4.6 Synthesis

The results presented in this paper indicate that ML methods show great promise in the study and management of hydrological droughts. RFs are shown to be able to accurately reconstruct historic sailing

depths, which in turn could yield good estimates of damages to the shipping industry. Estimates of future sailing depth based on a 15% decrease in discharge due to climate change also gave realistic outcomes. Two approaches were considered in this paper to model sailing depths in the Meuse and the Rhine branches with RFs. The location specific approach yielded better predictions on individual locations than a general approach, but the location specific approach generalizes worse. The implementation of a combined set up using a locations specific model for the most important locations and using the general model for the other locations could be a very useful tool for water managers and researchers alike.

5. Conclusion

The main aim of this paper was to investigate to what degree machine learning methods can be used to study the impact of hydrological drought on inland shipping. To this end two random forest machine learning models were developed, based on two different approaches; An approach utilising location specific RF models that were trained for individual locations (LocSpe) and an approach where one single model to rule them all was constructed based on data from all locations (Isildur). The models are trained on a limited number of variables; discharge at Lobith, discharge at Eijsden, precipitation at De Bilt and distance along the river. By keeping the model inputs simple, this model can also be more easily transferred to operational use. These models were then tested to reconstruct historical data and make future prediction in a 2°C climate change scenario. The results of the models were then used to estimate the economic damages to the shipping industry.

Results show that both RF approaches were able to accurately reconstruct historical sailing depth data, although model performance does vary per location. The estimates of future sailing depth with decreased discharges due to climate change all predict a decrease in sailing depth. This in turn will increase the economic damages to shipping industry, as insufficient sailing depths will occur more frequently.

It was found that Isildur generalizes better than LocSpe, while the predictions of LocSpe are more accurate for individual locations. This is especially the case for locations where the LocSpe model was developed on. The performance of LocSpe for locations it had not yet encountered depends largely on whether the target location is along the same river or a connected river branch as the predictor location. It is suggested that the optimal application may be a combined model, where Isildur is used for general monitoring while LocSpe can be used for the major chokepoints.

The estimates of economic damages to the shipping industry are heavily reliant on the accuracy of the sailing depth predictions. For locations with a good match between the observed sailing depth and predicted sailing depth, the economic damage estimates are also accurate. While for locations with poorer matches in sailing depth, the resulting economic damage also provides a poor estimate. Especially mismatches in predictions for the extremely low sailing depths can result in large differences between observed and estimated damages, as these are the conditions when economic impacts occur. Small differences here can result in large differences in economic damages due to the fleet composition.

There are some issues with the use of RF or ML in general. The main disadvantage found is that RFs struggle to predict the most extreme values and are unable to extrapolate outside of the training data range. Data that has never been observed is very hard to predict for the RF models. This means that the model is heavily dependent on quality and volume of the training data. Secondly, the grey box nature of RF makes understanding the internal processes of the model difficult, which makes it difficult to identify causal relationships.

Nonetheless, machine learning methods show great promise as a useful tool in the water managers or water researcher's inventory. The random forest based results found in this paper show the great potential of the implementation of machine learning programs. Compared to traditional numerical

models they are faster and easier to use, while still being accurate. This makes ML methods well suited for monitoring, predicting and research of hydrological drought.

6. Acknowledgements

I would like to thank my supervisors Niko Wanders and Sandra Hauswirth for their help and useful feedback. I would further like to thank Sven Krook, Namera Nous and Anniek Vermeijlen for proofreading this work.

The name Isildur is trademarked under middle earth enterprises.

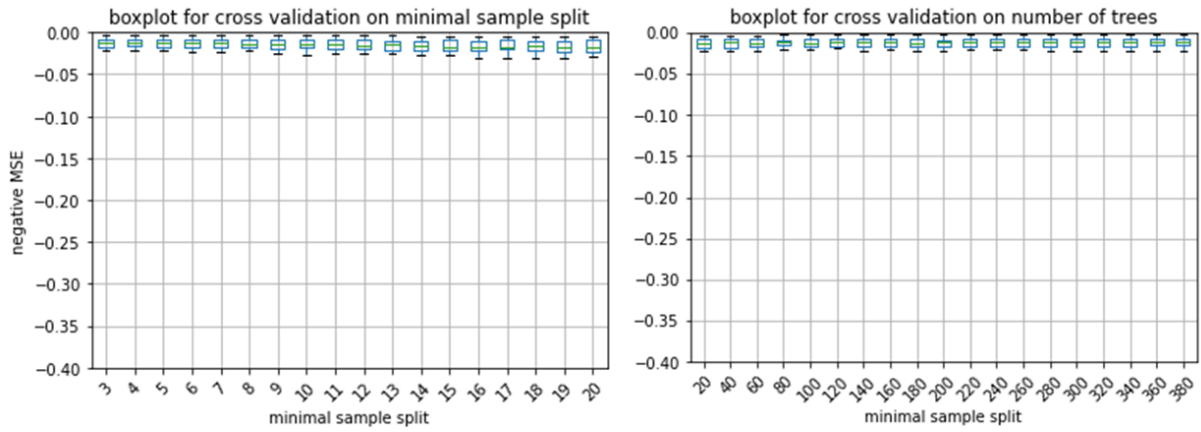
7. References

- Adger, N., Aggarwal, P., Agrawala, S., Alcamo, J., Allali, A., Anisimov, O., Arnell, N., Boko, M., Canziani, O., Carter, T., Casassa, G., Confalonieri, U., Cruz, R. V., Alcaraz, E. D. A., Easterling, W., Field, C., Fischlin, A., Fitzharris, B., García, C. G., ... Vaughan, D. (2007). Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Summary for Policymakers Drafting Authors. *Assessment*, 37(April 2007), 7–22. <https://doi.org/10.2134/jeq2008.0015br>
- Adnan, R. M., Zounemat-Kermani, M., Kuriqi, A., & Kisi, O. (2021). *Machine Learning Method in Prediction Streamflow Considering Periodicity Component* (Issue August).
- Albers, S. J., Déry, S. J., & Peticrew, E. L. (2016). Flooding in the Nechako River Basin of Canada: A random forest modeling approach to flood analysis in a regulated reservoir system. *Canadian Water Resources Journal*, 41(1–2), 250–260. <https://doi.org/10.1080/07011784.2015.1109480>
- Boehmke, B., & Greenwell, B. (2019). Hands-On Machine Learning with R. In *Hands-On Machine Learning with R*. <https://doi.org/10.1201/9780367816377>
- Breiman, L. (2001). Random forests. *Random Forests*, 1–122. <https://doi.org/10.1201/9780429469275-8>
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data and Society*, 3(1), 1–12. <https://doi.org/10.1177/2053951715622512>
- Cheng, M., Fang, F., Kinouchi, T., Navon, I. M., & Pain, C. C. (2020). Long lead-time daily and monthly streamflow forecasting using machine learning methods. *Journal of Hydrology*, 590(July), 125376. <https://doi.org/10.1016/j.jhydrol.2020.125376>
- Daniel, E. B., Camp, J. V., LeBoef, E. j., Penrod, J. R., Dobbins, J. P., & Abkowitz, M. D. (2011). Watershed Modeling and its Applications: A State-of-the-Art Review. *The Open Hydrology Journal*, 5(1), 26–50. <https://doi.org/10.2174/1874378101105010026>
- Descy, J. P. (2009). Continental Atlantic Rivers. In *Rivers of Europe* (Issue January). <https://doi.org/10.1016/B978-0-12-369449-2.00005-9>
- Frings, R. M., Hillebrand, G., Gehres, N., Banhold, K., Schriever, S., & Hoffmann, T. (2019). From source to mouth: Basin-scale morphodynamics of the Rhine River. *Earth-Science Reviews*, 196(January). <https://doi.org/10.1016/j.earscirev.2019.04.002>
- Hastie, T., Tibsharani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Journal of the Royal Statistical Society* (Vol. 103, Issue 2). <https://doi.org/10.2307/2980421>
- Hauswirth, S. M., Bierkens, M. F. P., Beijl, V., & Wanders, N. (2021). The potential of data driven approaches for quantifying hydrological extremes. *Advances in Water Resources*, 155(August), 104017. <https://doi.org/10.1016/j.advwatres.2021.104017>
- Hekman, A., Läkamp, R., van der Kooij, S., van de Velde, I., & van Hussen, K. (2019). *Economische schade door droogte in 2018 | Rapport | Rijksoverheid.nl*. 28–31. <https://www.rijksoverheid.nl/documenten/rapporten/2019/12/18/bijlage-1-rapport-economische-schade-door-droogte-in-2018>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 2018(8). <https://doi.org/10.7717/peerj.5518>
- Hurkmans, R., Terink, W., Uijlenhoet, R., Torfs, P., Jacob, D., & Troch, P. A. (2010). Changes in streamflow dynamics in the Rhine basin under three high-resolution regional climate scenarios.

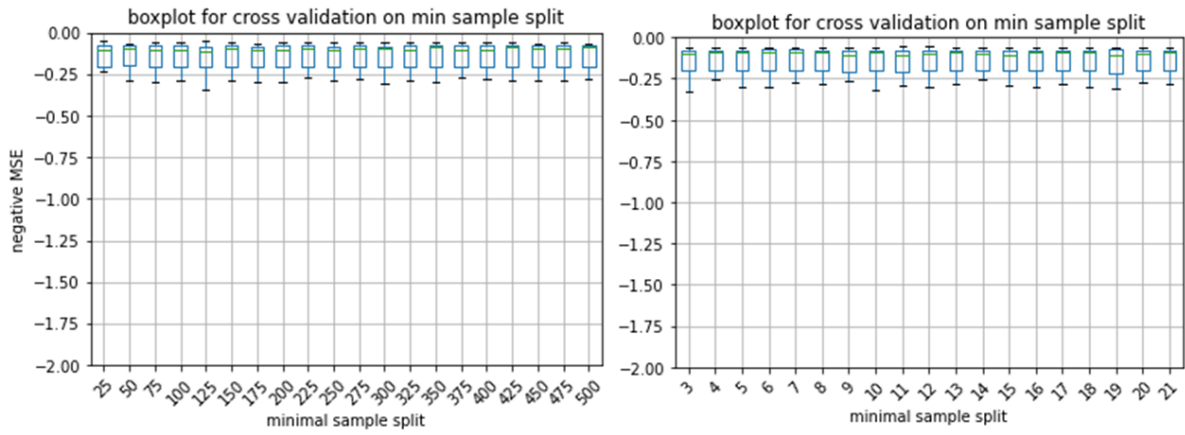
- Journal of Climate*, 23(3), 679–699. <https://doi.org/10.1175/2009JCLI3066.1>
- Jong, J. de. (2019). Eenvoudige scheepvaartrelatie voor vaarkosten Waal door afname vaardiepte. *Deltares Memo 11203734-012-ZWS-0007*, 29.
- Jonkeren, O., Jourquin, B., & Rietveld, P. (2011). Modal-split effects of climate change: The effect of low water levels on the competitive position of inland waterway transport in the river Rhine area. *Transportation Research Part A: Policy and Practice*, 45(10), 1007–1019. <https://doi.org/10.1016/j.tra.2009.01.004>
- KNMI, Attema, J., Bakker, A., Beersma, J., Bessembinder, J., Boers, R., Brandsma, T., van den Brink, H., Drijfhout, S., Eskes, H., Haarsma, R., & others. (2014). *KNMI'14: Climate Change scenarios for the 21st Century—A Netherlands perspective*. May. http://www.klimaatsscenarios.nl/brochures/images/KNMI_WR_2014-01_version26May2014.pdf
- Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089–5110. <https://doi.org/10.5194/hess-23-5089-2019>
- Kroekenstoel, D. (2014). *Rivierkundig beoordelingskader voor ingrepen in de Grote Rivieren*. 108. <http://publicaties.minienm.nl/documenten/rivierkundig-beoordelingskader-voor-ingrepen-in-de-grote-rivieren-versie-3-0>
- Lange, H., & Sippel, S. (2020). *Machine Learning Applications in Hydrology*. June 2021, 233–257. https://doi.org/10.1007/978-3-030-26086-6_10
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22.
- Loganathan, P., & Mahindrakar, A. B. (2020). Intercomparing the robustness of machine learning models in simulation and forecasting of streamflow. *Journal of Water and Climate Change*, 1–14. <https://doi.org/10.2166/wcc.2020.365>
- Maascommissie, & Internationale. (2013). *Jaarverslag 2012*.
- Mehran, A., Mazdiyarni, O., & Aghakouchak, A. (2015). A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand. *Journal of Geophysical Research Atmospheres*, 175(4449), 238. <https://doi.org/10.1038/175238c0>
- Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J. C. J., Lang, H., Parmet, B. W. A. H., Schädler, B., Schulla, J., & Wilke, K. (2001). Impact of climate change on hydrological regimes and water resources management in the Rhine basin. *Climatic Change*, 49(1–2), 105–128. <https://doi.org/10.1023/A:1010784727448>
- Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, 391(1–2), 202–216. <https://doi.org/10.1016/j.jhydrol.2010.07.012>
- Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2021). Review: Sources of hydrological model uncertainties and advances in their analysis. *Water (Switzerland)*, 13(1), 1–23. <https://doi.org/10.3390/w13010028>
- Pfister, L., Kwadijk, J., Musy, A., Bronstert, A., & Hoffmann, L. (2004). Climate change, land use change and runoff prediction in the Rhine-Meuse basins. *River Research and Applications*, 20(3), 229–241. <https://doi.org/10.1002/rra.775>
- Rijkswaterstaat. (n.d) "*Rivieren*". <https://www.rijkswaterstaat.nl/water/waterbeheer/beheer-en-ontwikkeling-rijkswateren/rivieren#:~:text=RUSTIGE%20MUZIEK%20VOICE%2DOVER%3A%20Rijkswaterstaat,voor%20vlotte%20en%20veilige%20scheepvaart>.
- Schasfoort, F., De Jong, J., & Meijers, E. (2019). *Effectmodules in het Deltaprogramma Zoetwater - Van hydrologisch effect naar economisch effect van droogte*.
- Shamshirband, S., Hashemi, S., Salimi, H., Samadianfard, S., Asadi, E., Shadkani, S., Kargar, K., Mosavi, A., Nabipour, N., & Chau, K. W. (2020). Predicting Standardized Streamflow index for hydrological drought using machine learning models. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 339–350. <https://doi.org/10.1080/19942060.2020.1715844>
- Srivastava, P., McNair, J. N., & Johnson, T. E. (2006). Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed. *Journal of the American Water Resources Association*, 42(3), 545–563. <https://doi.org/10.1111/j.1752-1688.2006.tb04475.x>

- Sugg, M., Runkle, J., Leeper, R., Bagli, H., Golden, A., Handwerger, L. H., Magee, T., Moreno, C., Reed-Kelly, R., Taylor, M., & Woolard, S. (2020). A scoping review of drought impacts on health and society in North America. *Climatic Change*, 162(3), 1177–1195. <https://doi.org/10.1007/s10584-020-02848-6>
- Tallaksen, L. M., & Van Lanen, H. A. J. (2004). *Hydrological drought: processes and estimation methods for streamflow and groundwater*.
- Torgo, L., Ribeiro, R. P., Pfahringer, B., & Branco, P. (2013). SMOTE for regression. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8154 LNAI(September), 378–389. https://doi.org/10.1007/978-3-642-40669-0_33
- Tyralis, H., & Papacharalampous, G. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*.
- van der Mark, R., Den Toom, M., van der wijk, R., & Sloff, K. (2020). *Verwachting waterdiepte Rijntakken*.
- van der Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. *Geophysical Research Letters*, 46(4), 2093–2102. <https://doi.org/10.1029/2019GL081967>
- Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., & Van Loon, A. F. (2013). Hydrological drought across the world: Impact of climate and physical catchment structure. *Hydrology and Earth System Sciences*, 17(5), 1715–1732. <https://doi.org/10.5194/hess-17-1715-2013>
- Van Loon, A. F. (2015). Hydrological drought explained. *WIREs Water*, 2(4), 359–392. <https://doi.org/10.1002/wat2.1085>
- Van Loon, A. F., & Van Lanen, H. A. J. (2012). A process-based typology of hydrological drought. *Hydrology and Earth System Sciences*, 16(7), 1915–1946. <https://doi.org/10.5194/hess-16-1915-2012>
- van Slobbe, E., Werners, S. E., Riquelme-Solar, M., Bölscher, T., & van Vliet, M. T. H. (2016). The future of the Rhine: stranded ships and no more salmon? *Regional Environmental Change*, 16(1), 31–41. <https://doi.org/10.1007/s10113-014-0683-z>
- van Vliet, M. T. H., & Zwolsman, J. J. G. (2008). Impact of summer droughts on the water quality of the Meuse river. *Journal of Hydrology*, 353(1–2), 1–17. <https://doi.org/10.1016/j.jhydrol.2008.01.001>
- Wanders, N., & Wada, Y. (2015). Human and climate impacts on the 21st century hydrological drought. *Journal of Hydrology*, 526, 208–220. <https://doi.org/10.1016/j.jhydrol.2014.10.047>
- Warmink, J. J., Booij, M. J., Van der Klis, H., & Hulscher, S. J. M. H. (2013). Quantification of uncertainty in design water levels due to uncertain bed form roughness in the Dutch river Waal. *Hydrological Processes*, 27(11), 1646–1663. <https://doi.org/10.1002/hyp.9319>
- Wilhite, D. A., & Glantz, M. H. (1985). Understanding the drought phenomenon: The role of definitions. *Planning for Drought: Toward A Reduction of Societal Vulnerability*, 11–27. <https://doi.org/10.4324/9780429301735-2>
- Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resources Management*, 21(5), 763–774. <https://doi.org/10.1007/s11269-006-9076-5>
- Witte, J.-P., De Louw, P., Van Ek, R., Bartholomeus, R., Van den Eertwegh, G., Gilissen, H. K., Van Rijswick, M., Beugelink, G., Ruijtenberg, R., & Van der Kooij, W. (2020). Aanpak droogte vraagt transitie waterbeheer. *Water Governance*, 3, 120–131. <https://edepot.wur.nl/538739>

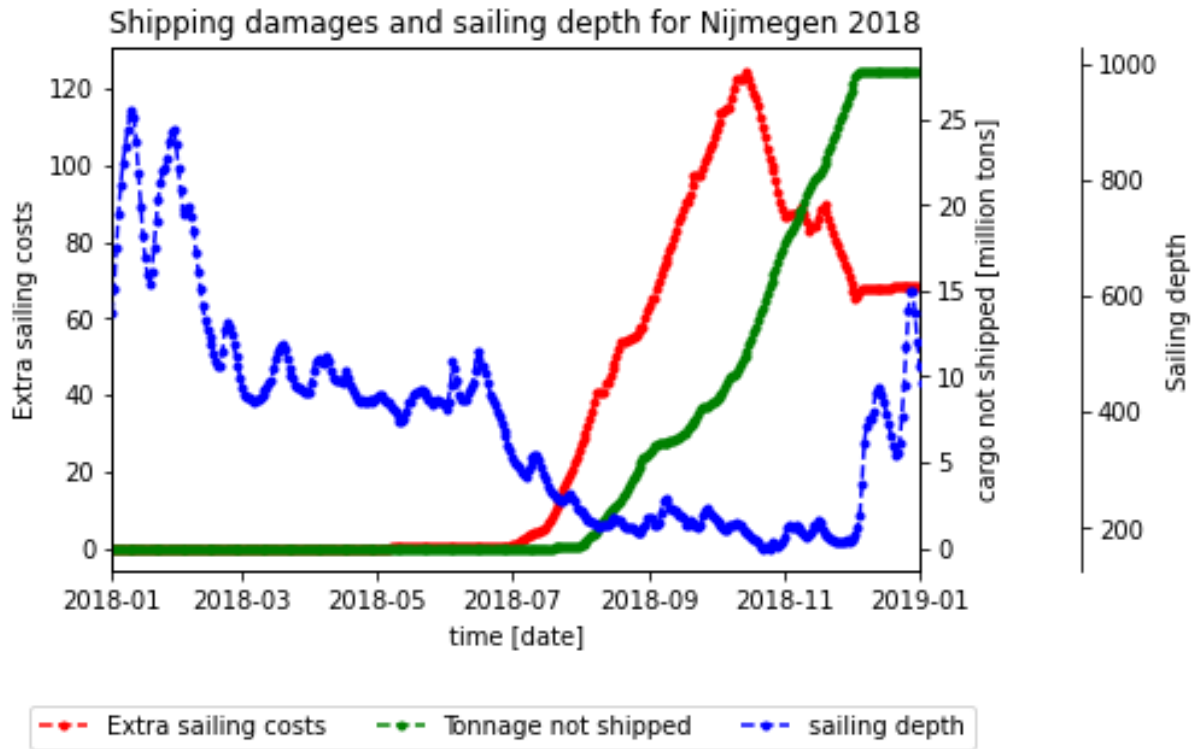
Appendix



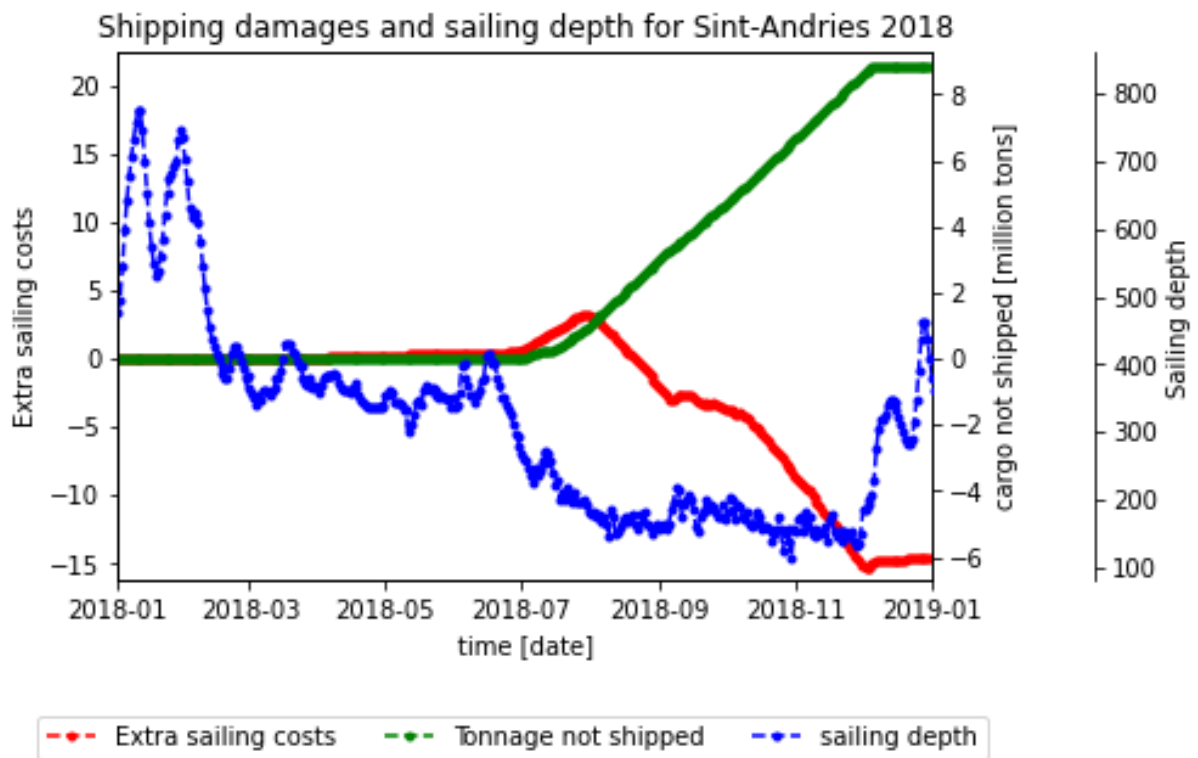
Appendix Figure 1: Results for minimal sample split and number of trees from the SA of the LocSpe model. Values closer to zero represent better matches between the predictions and the targets.



Appendix Figure 2 Results for minimal sample split and number of trees from the SA of the Isildur model. Values closer to zero represent better matches between the predictions and the targets.



Appendix Figure 3: Cumulative cargo not transported and sailing costs compared to sailing depth for Nijmegen



Appendix Figure 4: Cumulative cargo not transported and sailing costs compared to sailing depth for Sint Andries

Statement of originality of the MSc thesis

I declare that:

1. this is an original report, which is entirely my own work,
2. where I have made use of the ideas of other writers, I have acknowledged the source in all instances,
3. where I have used any diagram or visuals, I have acknowledged the source in all instances,
4. this report has not and will not be submitted elsewhere for academic assessment in any other academic course.


Student data:

Name: Jordy van de Ven

Registration number: 5859700

Date: 11-03-22

Signature:

A handwritten signature in blue ink is written over a horizontal dotted line. The signature is stylized and appears to be 'J. van de Ven'.