Optimizing the conjunctive use of surface water and groundwater resources for drinking water production under uncertain climatic conditions

Master thesis by:
Jorine Oosting
University of Twente
18-08-2022

Graduation committee:
Dr. Maarten Krol
Dr. ir. Rick Hogeboom
Ir. Martijn Deenen (Royal HaskoningDHV)
Dr. ir. Martijn Bakker (Royal HaskoningDHV)

Preface

Dear all,

This thesis marks the end of my seven year studying period wandering in and around Enschede. Although I do not yet fully realize it, I am happy and proud that I have travelled all the way down this road to arrive at a new starting point in life. Big shout-out to all my friends and family for always supporting and inspiring me!

With a bachelor's degree in Biomedical Engineering, I started my master's degree in Civil Engineering in February 2020, just before the Covid crisis. This was not the happiest time for anyone living at our planet, but I am astonished by each and everyone's resilience against this unexpected event. Although my master's in Civil Engineering has mostly taken place from behind my desk at home, I enjoyed diving into this cool field. I feel comfortable and expect a long and interesting career in water management and its related fields.

First, I owe many thanks to Martijn Deenen for the intense supervision you provided me with during the past months. Our weekly discussions were partly dedicated to my research, but also left enough space to discuss other interesting topics such as national and world politics. You would always lift me up with your never-ending enthusiasm and relativizing feedback, when I was once and again disappointed in myself for not being a super(wo-)man ("Je bent net een mens").

Also many thanks to Martijn Bakker, who has arranged this assignment for me and helped me adjust the contents to my own wishes. You were extremely helpful in the beginning of my project to guide the research into a relevant direction, by discussing practical considerations and asking stimulating questions. I enjoyed our trip to De Watergroep together with Pieter and Martijn, a great experience into company-life!

Besides the Martijn-duo, my other colleagues at Royal HaskoningDHV and especially at AquaSuite have added a lot to my graduation experience. A special thanks to Haochen, Nicole, Raul, Zixi, Konstantinos, and Hermen, it was great getting to know you and to exchange our research progress! You have been of great help.

Lastly, I want to thank my UT supervisors for guiding me through this process. I experienced some struggles with finding the sweet spot between the consultancy-mindset and the actual academic research view. Maarten and Rick, you have helped me structure my research and focus on the academic relevance. Thank you for giving me the chance to stand above the research the way I do now.

Looking forward to the future,

Jorine

Utrecht, 11-08-2022

“All models are wrong, but some are useful.”

– George Box (1976)

Summary

Drinking water is a basic human need that is under constant and increasing threat of climate change. Specific changes in the climatic system are uncertain, which leads to uncertain surface and groundwater availability for drinking water production. Surface water supplies are highly variable and vulnerable to climatic processes, while the groundwater system changes more gradual and is less vulnerable. The resilience of drinking water production to climate change can be increased with the conjunctive use of surface and groundwater resources. This research focuses on the role of uncertainty in climatic conditions in the conjunctive use of surface and groundwater resources for drinking water production. A case study for De Watergroep, a drinking water company in Belgium that implements a conjunctive use strategy, was used to implement the proposed methodology. An adaptive forecast-based production strategy for one of the reservoirs was developed using a water quality forecasting model. Three different drought years were evaluated based on their performance on yearly groundwater use and the risk of water shortages.

Machine learning models were developed to forecast five water quality parameters that are decisive in the conjunctive use of surface and groundwater resources in the case study: nitrate, phosphate, sulfate, conductivity and bentazon. The autoregressive behavior of all compounds was captured with their lagged values and rolling statistics. External features that were investigated were calendar effects, precipitation, temperature, discharge, and land use. Six different algorithms were tested with 20 different feature sets, predicting one to three week(s) ahead for all water quality parameters. For nitrate prediction, the autocorrelation and calendar effects were of main importance, while all other water quality parameters perform best when precipitation, discharge, and temperature are included as well. Interdependencies of the climatic processes highly influence their predictive power.

An adaptive forecast-based production strategy was developed to find the optimal conjunctive use of surface and groundwater resources under uncertain climatic conditions. The goal of this new production strategy was to meet water demand with the trade-off of minimizing groundwater use and minimizing the risk of water shortages, under the physical constraints of the operating reservoir and surface water availability. The research shows that relatively simple models using water quality data and open-source climatic data can be used to develop an adaptive forecast-based approach with improved reservoir performance. The data-driven approach enables modelling without the need to investigate all individual relations in the climate-water system. The forecast-based strategy was evaluated by comparing the yearly groundwater use and the risk of water shortages to a traditional production strategy. It was found that the new production strategy averts the risk of water shortages, which sometimes comes at the cost of increased groundwater use. Improving the forecasting model results in a lower risk of water shortages.

Samenvatting

Drinkwater is een van de eerste levensbehoeften van mensen en staat onder een toenemende druk van klimaatverandering. Specifieke veranderingen in het klimaatsysteem zijn onzeker, wat leidt tot onzekerheden in de beschikbaarheid van oppervlakte- en grondwater voor drinkwaterproductie. Oppervlaktewater is variabel en gevoelig voor klimaatprocessen, terwijl het grondwatersysteem geleidelijk verandert en minder kwetsbaar is. Het gelijktijdig gebruik van oppervlakte- en grondwaterbronnen in drinkwaterproductie is een beheerstrategie die de weerbaarheid van een watersysteem vergroot. Dit onderzoek richt zich op de rol van onzekerheid in klimatologische condities in het gelijktijdig gebruik van oppervlakte- en grondwaterbronnen in drinkwaterproductie. Er wordt gebruik gemaakt van een case study van De Watergroep, een drinkwaterbedrijf in België dat een strategie van gelijktijdig gebruik implementeert. Aan de hand van waterkwaliteitsvoorspelmodellen is een flexibele productiestrategie ontwikkeld. De nieuwe productiestrategie is geëvalueerd op basis van het jaarlijkse grondwatergebruik en het risico op watertekorten in drie verschillende droogtejaren.

Machine learning modellen zijn ontwikkeld om vijf verschillende waterkwaliteitsparameters die belangrijk zijn in het gelijktijdig gebruik van oppervlakte- en grondwaterbronnen in de case study te voorspellen: nitraat, fosfaat, sulfaat, geleidbaarheid en bentazon. Het autoregressieve gedrag van alle parameters is in het model meegenomen aan de hand van hun afgelopen waarden en de interne statistiek. Externe features zijn de kalendereffecten, neerslag, temperatuur, debiet en landgebruik. Zes verschillende algoritmen zijn getest met 20 verschillende feature sets, die 1-3 weken vooruit voorspellen voor alle verschillende waterkwaliteitsparameters. De autocorrelatie en kalendereffecten waren het belangrijkst bij het voorspellen van nitraat. Voor alle andere waterkwaliteitsparameters geldt een beste voorspelling met neerslag, debiet en temperatuur. De onderlinge afhankelijkheid van klimaatprocessen heeft een grote invloed op hun voorspellende waarde.

Een flexibele productiestrategie is ontwikkeld om het optimale gelijktijdige gebruik van oppervlakte- en grondwaterbronnen onder onzekere klimaatomstandigheden te vinden. Het doel van deze nieuwe productiestrategie was om te voldoen aan de watervraag met een trade-off van het minimaliseren van grondwatergebruik en het risico op watertekorten, onder de fysieke grenzen van het reservoir en de beschikbaarheid van oppervlaktewater. Simpele modellen die gebruik maken van waterkwaliteitsdata en open-source klimaatdata kunnen worden gebruikt om een voorspelling gebaseerde strategie met verbeterde performance te ontwikkelen. De datagedreven benadering maakt het onnodig om alle individuele klimaat-waterrelaties van het systeem te modelleren. De op voorspelling gebaseerde strategie werd vergeleken met een traditionele productiestrategie, op basis van jaarlijks grondwatergebruik en het risico op watertekorten. De nieuwe productiestrategie zorgt voor een lager risico op watertekorten, wat soms leidt tot een toename in grondwatergebruik. Verbeteringen in het voorspelmodel kunnen zorgen voor een nog lager risico op watertekorten.

Table of Contents

1. Introduction

Clean drinking water is one of the basic needs of human society (Zhang, 2015). Climate change threatens drinking water production because of its deteriorating effects on the quantity and quality of water resources (Bloetscher et al., 2014; Kundzewicz & Döll, 2009; O'Connell, 2017). Surface water and groundwater are both resources that can be used to produce drinking water. Quality surface and groundwater resources are limited and shrinking because of overexploitation, urbanization, and climate change impacts (Singh, 2014). Specific future climatic impacts are difficult to predict due to complex weather dynamics and site-specificity, which leads to much uncertainty. Uncertainty in climatic variability will even increase under a changing climate. In general, it can be stated that weather and temperature extremes become more common. In Europe, temperatures rise and precipitation patterns change (Maiolo et al., 2017). Increased temperatures cause decreased soil moisture content, precipitation deficits and higher water demand. Changes in rainfall pattern are highly region specific. In general, wet areas are projected to get wetter and dry areas are projected to get drier (O'Connell, 2017). Climate change is therefore expected to exacerbate regional and global water scarcity. Appropriate reservoir operation strategies are needed to ensure reliable water supplies for drinking water production under uncertain climatic conditions with increased water demand and reduced surface and groundwater availability (Bloetscher et al., 2014; Maiolo et al., 2017; Renwick, 2018).

The conjunctive use of surface and groundwater resources is an example of an integrated water management approach that increases a system's resilience against climatic variability and improves the security of water supplies under climatic change (Renwick, 2018). Conjunctive use is necessary because availability of one source of water may not be sufficient to fulfill the system's requirements (Singh, 2014). Surface water availability, as defined by its quantity and quality, is uncertain under a changing climate (Kundzewicz & Döll, 2009). Groundwater can be used as a replacement for surface water when surface water availability is low. The groundwater system changes more gradually and can be used as a buffer against the high degree of variability and vulnerability of surface water resources. It allows withdrawal during dry seasons when surface water is often more polluted than groundwater. Groundwater use has increased in recent decades relative to surface water use.

Natural and anthropogenic processes can cause low surface water availability through water quality problems (Baker, 2003; Delpla, 2009; Rostami et al., 2018; Simeonov et al., 2003). Precipitation, geology, soil type, vegetation cover, and discharge are all conditions that influence the water system and are highly region specific. Water pollution can originate from point or diffuse sources (Delpla, 2009). Point source pollution enters a river at a specific point, such as industrial effluents or a pipe discharging wastewater. Diffuse sources are pollutants that leach into the water system as a result of hydrological processes. Agricultural applications are well-known sources of diffuse pollution.

Conjunctive use of surface and groundwater resources has been evaluated in literature and shows advantages when surface water availability is low. Khare et al. (2006) showed conjunctive use reduces losses in irrigation supply when surface water resources are limited and groundwater is used as a supplementary source. Das et al. (2015) also used groundwater to complement surface water in times of low surface water availability. Conjunctive management can also be used when shortages of surface water supply and depletion of groundwater resources occur simultaneously (Sarwar & Eggers, 2006).

In conjunctive use, a trade-off exists between the risk of water shortages due to low surface water availability and the supplementary use of groundwater resources (Scanlon et al., 2016). Water shortages can be caused by climatic processes that deteriorate the surface water quality such that intake is not possible. Groundwater resources can then be used to avert the risk of a water shortage. Consequences of a water shortage are low water pressure and low drinking water quality (De Watergroep, 2021b). Groundwater can be strictly limited through groundwater permits, of which exceedance can become costly. Groundwater use should therefore also be limited, which leads to conflicting objectives.

Conjunctive use can be implemented using (a combination of) predefined long-term strategies or adaptive real-time production strategies (Ahmadi et al., 2015). Production strategies should prescribe decision rules and constraints for surface water and groundwater production rates. A predefined approach is based on historical long-term series or experiences. An adaptive real-time approach includes current system variables and requirements to which the operator responds. Uncertainties in climatic conditions make long-term operational strategies less favorable and stress the need for an adaptive real-time approach (Li et al., 2010). A forecast-based approach uses the knowledge of future events to operate the reservoir adaptively and anticipative, which can lead to a reduction of both the risk of water shortages and groundwater use. Adaptive forecast-based reservoir operation approaches were implemented before by (amongst others) Dagli & Miles (1980), Alemu et al. (2011), and Gavahi et al. (2019). Forecasting models predict input parameters based on which a set of decision rules and constraints regarding the output parameter are developed.

Production from surface water resources is limited through riverine water quality. Water quality measurements can be used for a predefined approach, while water quality forecasting is needed to develop an adaptive forecast-based production strategy. Forecasting surface water quality indicates what the expected period of intake is and when groundwater should be used supplementary. The water quality system and its dependencies on climatic processes is highly site-specific, which complicates understanding the underlying physical processes needed for conventional physically-based water quality modelling (Ahmed et al., 2019; Fukushima et al., 2000). While physically-based models are more transparent and easy to interpret, they are not able to capture the complex, site-specific relations that occur in the climate-water quality system. Data-driven approaches are more accurate in capturing system's specific behavior and therefore promising for water quality modelling (Solomatine & Ostfeld, 2008; Wu et al., 2017). Data-driven modelling is based on the analysis of concurrent input and output series. Only a limited number of assumptions about the physical behaviour of the system is needed. Machine learning models are data-driven modelling approaches that find relations between input and output data through different types of algorithms with different levels of complexity.

An adaptive forecast-based approach is more flexible and resilient to uncertain climatic conditions than a predefined approach and can reduce the risk of water shortages or groundwater use when surface water availability is limited through riverine water quality problems. Uncertainties in climatic conditions can, however, limit the advantages of forecast-based approaches due to inaccurate forecasting. The aim of this research is therefore **to explore the role of uncertainty in climatic conditions in the conjunctive use of surface and groundwater resources for drinking water production by developing an adaptive forecast-based production strategy.**

The following research questions will be answered to fulfill the aim of this research:

1. How can conjunctive use of surface and groundwater resources be decided on under uncertain climatic conditions?
    a. How do climatic processes influence the system parameters that are decisive in the conjunctive use of surface and groundwater resources?
    b. How can these climatic processes be used to forecast the conjunctive use of surface and groundwater resources under uncertain climatic conditions?
2. How can an adaptive forecast-based production strategy be developed for the optimal conjunctive use of surface and groundwater resources under uncertain climatic conditions?
    a. How can the conjunctive use of surface and groundwater resources be simulated?
    b. How can a predefined and an adaptive forecast-based production strategy be modelled?
    c. How does an adaptive forecast-based approach compare to a predefined production strategy?
    d. What are the trade-offs in optimizing the conjunctive use of surface and groundwater resources under uncertain climatic conditions?

A case study of a drinking water company in Western Flanders that adopts the conjunctive use of surface and groundwater resources is used to compare a predefined production strategy and an adaptive forecast-based strategy. Machine learning models are used to forecast water quality which defines water availability for surface water production in one of the surface water reservoirs. A reservoir model is developed with which both strategies are compared. A retrospective analysis of three historic years was done to compare the performance in risk of water shortages and groundwater use for both strategies.

2. Case study

De Watergroep is a Belgian drinking water company that implements a conjunctive use approach in their drinking water production strategy. Their approach is used as a case study for the present research. De Watergroep experienced problems in producing and supplying sufficient drinking water to their supply area as an effect of climate change. The dry summers of 2018 – 2020 led to low water quality and increased water demand (De Watergroep, 2021b). These summers are expected to be a prelude to future climate change effects, such as heavy droughts. De Watergroep therefore strives for an optimal production strategy in the conjunctive use of surface and groundwater resources in which groundwater use and the risk of water shortages is limited. This section explains the case study and a possible solution pathway.

2.1 Study area

De Watergroep België is the largest drinking water company in Flanders, with 3.3 million clients in over 150 Flamish municipalities (De Watergroep, 2021a). The supply area covers most of the West Flanders region in Belgium. Drinking water is produced from both surface water resources from the river Yser and groundwater resources. The river Yser emerges from small streams in France (Heylen, n.d.; Srubbe, 2005). The catchment knows relatively long dry periods and the river easily floods during heavy rainfall. When rainfall is low, discharge is much lower than with high rainfall. The average winter discharge is therefore much higher than the average summer discharge. 80% of the Yser catchment is covered with agriculture, horticulture and grasslands (Coordinatiecommissie Integraal Waterbeleid, 2016). Figure 1 shows a map of the supply area with the river Yser and the most important surface water reservoir: the Blankaart.

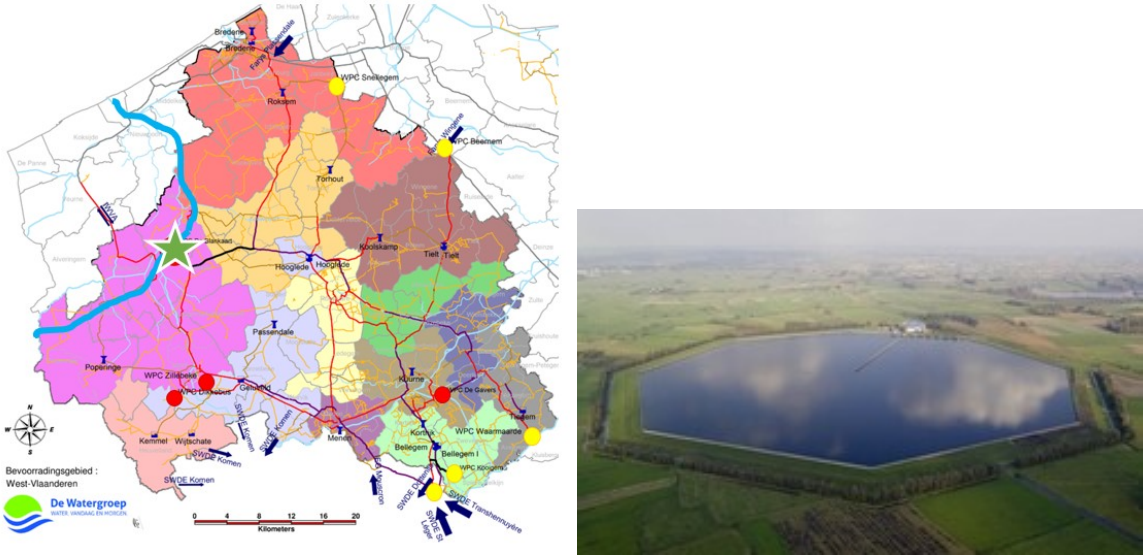


Figure 1: A) Supply area of de Watergroep. Red dots: surface water production centers. Yellow dots: groundwater production centers. Arrows: external purchases (De Watergroep, 2021b). Green star: WPC De Blankaart. Blue line: river Yser. B) Reservoir of WPC De Blankaart.

The red dots in Figure 1 indicate surface water production centers (WPC), the yellow dots indicate groundwater production. Surface water production centers have large reservoirs that are filled with raw river water before it is treated into high-quality drinking water, of which the Blankaart reservoir is the largest and most important. The Blankaart is located centered and relatively isolated. Switching off the Blankaart in summer can cause drinking water problems and creates the risk of water shortages.

2.2 Production strategies

Intake of river water to the Blankaart reservoir is regulated through water quality measurements in the river Yser. Different water quality compounds with alarm concentrations are measured that describe river water quality. If the water quality becomes too low (exceeds alarm concentration), intake of river water to the reservoir is stopped. If the water quality is sufficiently high, the company decides to start intake from the river again. The residual water demand is supplied from groundwater resources or neighboring companies. Both options are expensive and uncertain (De Watergroep, 2021b). Purchasing water from neighboring companies generates uncertainty since these companies might experience similar production issues during periods of low water quality. Increased groundwater uptake is expensive and does not comply with governmental ecological requirements. The Belgian government has decided to set strict permits to limit intake from groundwater resources because of extensive historic industrial groundwater use.

2.2.1 Current production strategy

The current production strategy is an approach that is reactive to water quality measurements and the coinciding potential intake. The operator can either switch directly to a low production rate or remain a high production rate although intake is not possible. This choice results in a trade-off between groundwater use and the risk of water shortages, which is shown in Figure 2. The company currently decides to either use maximum production capacity or half of the production capacity. The company plans to adopt a different system in five years, in which the production rate can be controlled linearly. This increases flexibility in designing a new production strategy.

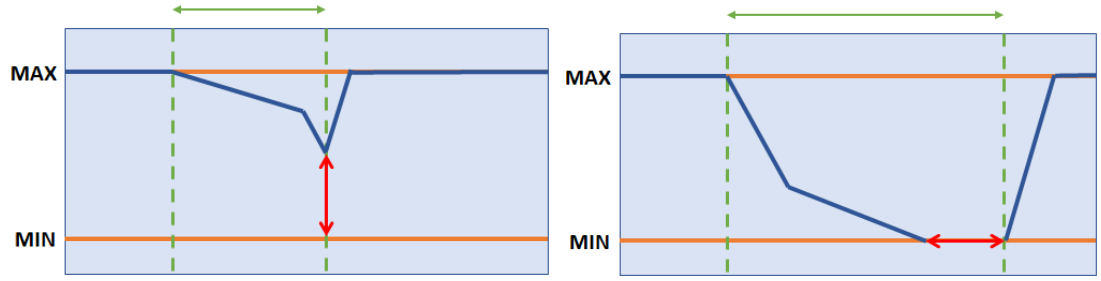


Figure 2: Trade-off in reservoir operation. Green lines: period of low water quality. Left-hand side: directly switch to low production. Right-hand side: remain high production rate.

The left-hand side in Figure 2 shows that the risk of water shortages can be averted by switching to low production early on in the dry period. This approach comes at the cost of increased groundwater use and exceedance of groundwater permits, because there is a large remaining volume in the reservoir at the end of the low water quality period. The right-hand side approach shows a high risk of water shortages, but limited groundwater use. There is no remaining water in the reservoir for which groundwater has been used as replacement.

2.2.2 New production strategy

Two optimal operational decision paths in which both the risk of water shortages and groundwater use are minimized are shown in Figure 3. The reservoir approaches its minimum volume just at the end of the low water quality period. This means that no unnecessary groundwater has been used and the reservoir has never reached the minimum volume which means that it can easily be filled. A newly developed production strategy should approach these optimal operations.

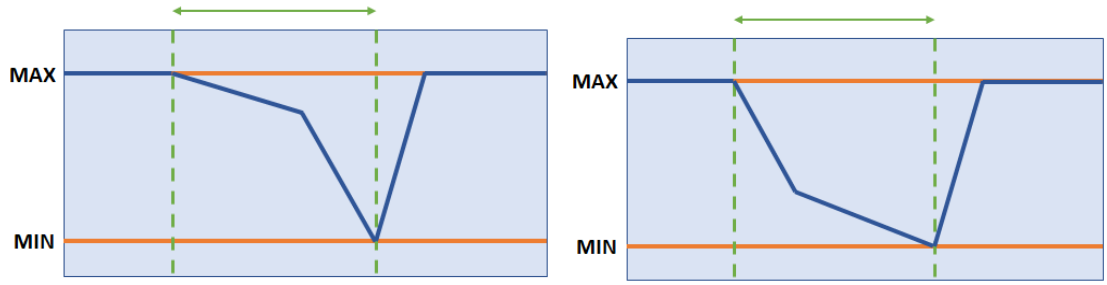


Figure 3: Optimal operation of the surface water reservoir.

Defining the end of a low water quality period comes with much uncertainty. If the end of a dry period can be clearly defined, the operator can take more risk and empty the reservoir such that less water remains in the reservoir before the end of the dry period. The production strategy can be adapted during the drought period itself according to river water quality forecasts. Accurate water quality forecasting helps to make operational decisions and optimize water production from surface water resources.

3. Methodology

Reservoirs can be operated with predefined strategies or adaptive forecast-based strategies. Intake is driven by surface water availability, which depends on water quality and/or water quantity. The Blankaart intake rate is quality-dependent and therefore driven by the concentration of several pollutants. The selected parameters in this research are nitrate (NO3), (ortho-)phosphate (oPO4), sulfate (SO4), conductivity and bentazon (see Appendix 1A). These compounds represent nutrients, pesticides, and the salination of water which are all important parameters for water quality. Climatic and anthropogenic processes influence river water quality and can be used as predictors for water quality forecasting with machine learning models. The data-driven approach of machine learning models allows capturing complex relations between predictors and water quality compounds without the need to model the complete climate-water system. The relation between climatic/anthropogenic processes and the water quality compounds can differ per parameter and forecast horizon. Individual models were therefore made for each water quality parameter and forecast horizon. An adaptive forecast-based reservoir production strategy is expected to improve performance in uncertain climatic conditions because the approach is more proactive instead of reactive.

This chapter proposes a methodology for forecasting river water quality and designing an adaptive forecast-based approach for reservoir operation. Research question 1 was covered by developing machine learning models for each water quality compound and experimenting with different feature sets including and excluding specific climatic processes. To answer research question 2, a reservoir model and a predefined and adaptive forecast-based production strategy were developed, and their performance is compared based on several key performance indicators. Figure 4 shows the outline of the methodology.

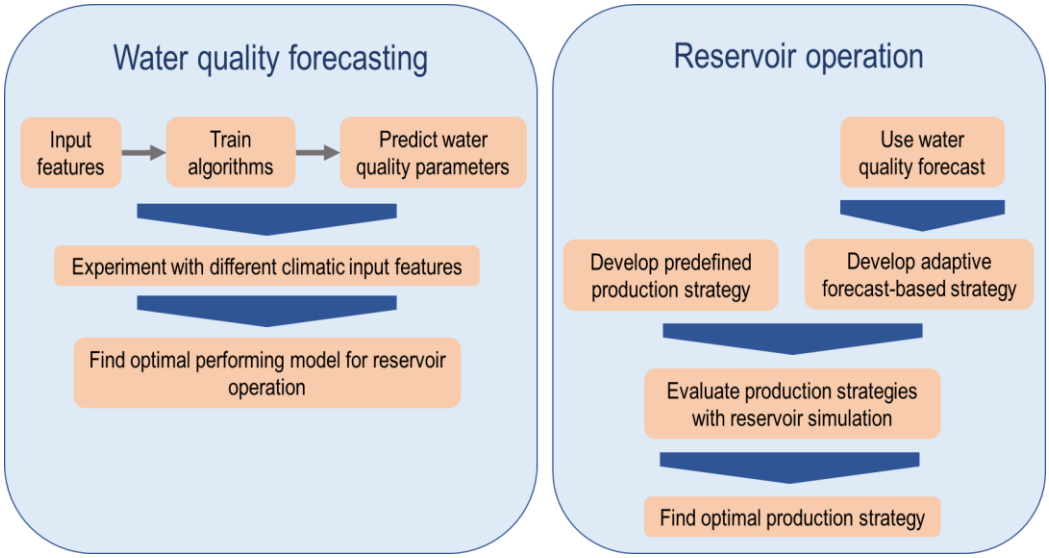


Figure 4: Methodology outline.

3.1 Water quality forecasting

3.1.1 Machine learning for water quality forecasting

Machine learning models can be used for time series forecasting to develop a predictive model on data with a temporal relationship between observations (Lazzeri, 2020; Nielsen, 2019). The models are data-driven and focus on identifying the patterns and relations between these data without the need for complex programming (Asadollah et al., 2021). Supervised learning is a machine learning approach in which the problem's input and output variables are both known and used in the learning process. Nitrate, phosphate, sulfate, conductivity, and bentazon are all influenced by climatic and anthropogenic processes. Relevant processes can differ per water quality parameter and forecast horizon, so separate dataframes were made for each combination of water quality parameter, forecast horizon, and set of input features (direct multi-step forecasting) (Lazzeri, 2020; Nielsen, 2019). All algorithms were tested on each dataframe. The columns of each dataframe represent the input features and one target variable. The input features were transformations of relevant climatic processes (precipitation, temperature, discharge, and land use), which were obtained through feature engineering. The target variable depended on the forecast horizon: the concentration of each water quality parameter of either 1, 2 or 3 weeks ahead was to be predicted. The rows of each dataframe are the observations of each input feature and target variable over time. An example snippet of a dataframe is shown in Figure 5.

Date	lagged_C_1	lagged_C_2	jan	feb	lagged_P_1	lagged_T_1	grassland	C
12-01-2015	36	35	1	0	15	4	0.33	38
19-01-2015	38	36	1	0	13	3.5	0.33	42
26-01-2015	42	38	1	0	8	3	0.33	44
02-02-2015	44	42	0	1	5	2	0.33	42

Figure 5: Example snippet dataframe. All rows are 'observations'. All columns except column 'C' are input features. Column 'C' is the target variable.

Each dataframe was split into a training (years 2011-2018) and testing (years 2019-2021) subset, for which both the input features and target variables are known (supervised learning). With each preceding time step, data from the testing set is added to the training set such that all currently available information is used for prediction. The data was not shuffled due to the temporal dependency. The training phase is used for modelling and the testing phase is used to evaluate the accuracy of the models by comparing forecasts with actual observations. Each input feature was scaled between 0 and 1, such that all input features fall within the same range of numerical values. The selected algorithm studies the training data to find relations between input and target variables and to be able to make predictions about future data points. The predictions based on unseen data from the test set are compared to actual observations from the test set to find the performance of the predictive model.

3.1.2 Model performance

Model performance of all models was assessed with the R^2 score (coefficient of determination), which indicates the degree of correlation between the observed and forecasted values (Equation 1) (Asadollah et al., 2021).

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2} \quad (1)$$

The R^2 score measures the amount of variability explained by the model based on the predictions and the actual observations in the test set. $R^2 = 0$ means no correlation, $R^2 = 1$ means perfect correlation. The score gives the overall model performance and is easy to compare to other studies that use machine learning for forecasting purposes. Following the study of Alnahit et al. (2022), the following classification for model

performance was used: weak performance = $R^2 < 0.25$, moderate performance = $0.25 < R^2 < 0.75$, and strong performance = $R^2 > 0.75$.

For each combination of water quality parameter/forecast horizon, a baseline model was made to which the model performance of all forecasting models was compared. This baseline model was based on the principle of persistence, in which the last measured concentration is assumed to be the predictor for the next value. This is a rough approximation of a predefined production strategy that does not use a forecasting model.

3.1.3 Analysis

The model performance was assessed through three different pathways. The effect of all climatic processes was assessed by comparing model performance between feature sets that include each specific process and the ones that exclude the process. The average, minimum and maximum performance for including or excluding the climatic process are important indicators of the predictive value of each climatic process. The weighted optimal feature set represents which set of features overall performs best on all forecast horizons of a water quality parameter. The top 5 performing models are scored for how well they perform. The best performing feature set gets 5 points, the least one 1 point. This ranking is done for each forecast horizon. Then, the weighted ranking is obtained per feature set by adding all scores and dividing this by 3 (forecast horizons). The highest scoring feature set is the weighted optimal feature set. The optimal performing model was lastly found for each water quality parameter and forecast horizon, that were used for reservoir operation.

3.1.4 Machine learning algorithms

Water-related problems are often characterized by noisy and poor quality data (Solomatine & Ostfeld, 2008). The performance of algorithms for water quality forecasting is not evident and rather case-dependent. Tiyasha et al. (2020) provide an extensive overview. Different algorithms can result in different performance and it is therefore necessary to apply various modelling techniques and compare the results. Six algorithms were used and compared in this research (see Table 1). All algorithms were implemented with the open-source Python machine learning package scikit-learn (Pedregosa et al., n.d.). Scikit-learn is an easy to use package that consists of a range of machine learning, preprocessing and visualization algorithms that is easy to integrate with other useful package (Lazzeri, 2020). Algorithms can be optimized by tuning the hyperparameters (model settings). The high number of possible combinations of hyperparameters makes this process labor expensive, so the standard values of the hyperparameters were used.

*Table 1: Machine learning algorithms used in the current research. *See: https://scikit-learn.org/ .*

Algorithm	Algorithm (sklearn)*
Linear Regression	LinearRegression
Decision Tree	DecisionTreeRegressor
Support Vector Machine	LinearSVR, SVR
K-Nearest Neighbors	KNeighborsRegressor
Random Forest	RandomForestRegressor

Simple or Multiple Linear Regression (SLR/MLR) linearly estimates the relationship of dependent variables to independent variables by fitting a line (Koranga et al., 2022; Lepot et al., 2017). Linear regression models try to minimize the sum of squared errors between the measured and modelled data (Yildiz et al., 2017). The models are simple and easy to interpret. A disadvantage is that their focus is more on finding a general solution over the complete dataset than on specific subsets, and do not handle nonlinearities sufficiently.

Decision Trees (DT) are a hierarchical structure where the prediction of an output variable depends on all higher-level data attributes (Koranga et al., 2022). Each node of the tree represents a characteristic which distinguishes different decision paths. The model learns through following top-to-bottom decision rules. Decision

trees are simple to understand and interpret and require little data preparation (Pedregosa et al., n.d.). They work well for data with many nonlinearly interacting features (Singh et al., 2016; Yildiz et al., 2017).

Support Vector Machines (SVM/SVR) find a hyperplane to separate n-dimensional spaces into classes (Koranga et al., 2022; Sattari et al., 2016). Most of the training subset values should be within a margin from the hyperplane, which is defined by a certain tolerance level. SVM's are customizable to fit the case purpose by using specific settings. They are suitable for nonlinear regression problems and able to predict quantities forward in time based on training from past data and is suitable for extrapolation (Mosavi et al., 2018; Yildiz et al., 2017). A disadvantage is that the algorithm is complex and can be time costly.

The k-Nearest Neighbors (KNN) approach uses different metrics of windows of surrounding or preceding data to forecast the window of interest for the target variable (Lepot et al., 2017). The algorithm computes a weighted average of the number of k nearest neighbors which are inversely distance weighted (Sattari et al., 2016). KNN can be used for cases with continuous data labels, where the label to the query point is computed based on the mean of the labels of its nearest neighbors (Pedregosa et al., n.d.). Extrapolation is based on the nearest past k neighbors, which makes this algorithm not the most favorable for extrapolation purposes.

A Random Forest (RF) is an ensemble of decision trees that selects the best output prediction by aggregating the outcomes of multiple decision trees (Koranga et al., 2022). Random forests have the advantages of decision trees and the additional advantage that errors are cancelled out by taking the average of all decision tree predictions (Pedregosa et al., n.d.). Random forests are robust to noise and do not overfit to data, but they can be slow for real-time prediction (Singh et al., 2016).

3.1.5 Input features

Climatic variability affects water quality compounds through natural and anthropogenic processes such as precipitation, temperature, discharge, and land use (Baker, 2003; Rostami et al., 2018; Simeonov et al., 2003). Many processes depend on meteorological and hydrological conditions, which are highly region specific (Fukushima et al., 2000). Site-specific investigation is therefore needed to capture the predictive value of each climatic process. The climate processes covered in this research are precipitation, air temperature, discharge, and land use.

Different input feature sets were used in combination with the machine learning algorithms to find the predictive value of the climatic processes on the water quality parameters. Six algorithms and 20 different feature sets (see Table 2) were altered for five water quality parameters and 3 forecast horizons. Experiments with different feature sets including or excluding relevant climatic processes were iteratively used to find the optimal performing model, and the influence of each climatic process in the prediction of each water quality parameter for each forecast horizon. The optimal performing models for forecast horizons of 1, 2, and 3 weeks were used in the adaptive forecast-based strategy for reservoir operation.

Table 2: Feature sets.

#	Feature set		#	Feature set
1	Base		11	Calendar
2	Base + P		12	Calendar + P
3	Base + T		13	Calendar + T
4	Base + Q		14	Calendar + Q
5	Base + LU		15	Calendar + LU
6	Base + P +T		16	Calendar + P + T
7	Base + P + Q		17	Calendar + P + Q
8	Base + T + Q		18	Calendar + T + Q
9	Base + P + T + Q		19	Calendar + P + T + Q
10	Base + P + T + Q + LU		20	Calendar + P + T + Q + LU

3.1.4.1 Base features

In time series forecasting, all previous time steps can be used as input to predict the next time step as output of the model, using a sliding window. Base features are therefore included to capture the autoregressive behavior of all water quality variables. These include the lagged concentrations of each water quality parameter, and the rolling statistics (mean, maximum, minimum, standard deviation). Each water quality parameter has different autoregressive behavior. Autocorrelations with a Pearson correlation coefficient (R) of >=0.5 are considered to have enough predictive power to contribute to the forecasting model performance (see Appendix 1B). The rolling statistics are determined over a rolling antecedent window of which the size is determined by the highest lag for which R>=0.5 holds.

3.1.4.2 Calendar features

Calendar features were added to the model to capture the seasonal behavior of the system. Each month of the year is a distinct feature. One-hot encoding is used to convert the month classes into numerical values. An example is shown in Figure 6.

	Input features						
Date	Jan	Feb	Mar	Apr	May	Jun	Jul
01-03-2022	0	0	1	0	0	0	0
01-04-2022	0	0	0	1	0	0	0
01-05-2022	0	0	0	0	1	0	0

Figure 6: Calendar features, one-hot encoded.

3.1.4.3 Precipitation

Precipitation influences river water quality either direct through wash-off or indirect through dilution (Rostami et al., 2018). Water pollutants from diffuse sources are prone to wash-off, while dilution is more important for point sources. Wash-off occurs when non-point source pollutants are accumulated on land surface, for example in soils or leaves, and end up in the river by surface runoff generated from precipitation. Many non-point sources, like nitrate, phosphate, and pesticides, are used in agricultural applications and end up in the river after intense rainfall (Fukushima et al., 2000; Kalkhoff et al., 2016). For point-sources, such as industrial effluent, dilution is a more important effect of precipitation. Their concentration decreases through dilution (Rostami et al., 2018).

The mean areal precipitation over the study area was determined with three spatial interpolation methods, which are explained in Appendix 1C. The methods with the strongest correlations were used for each water quality parameter. The mean areal precipitation time series will be transformed into cumulative antecedent precipitation, similar to the research conducted by Rostami et al. (2018).

Wash-off is captured with the 1 – 10 days past precipitation. Dilution is expected to take place over a longer lagged time scale. The Pearson correlation (95% C.I.) between the cumulative antecedent precipitation and the water quality parameters was determined to identify appropriate time scales to capture dilution, using an expanding window of window size 10 – 100 days. The optimal lag has the highest Pearson correlation (see Appendix 1D).

3.1.4.4 Temperature

High air temperature affects water quality through enhanced nutrient cycling rates and eutrophication (Benítez-Gilabert et al., 2010; Fukushima et al., 2000). Pesticide concentrations are expected to increase with increased temperature through increased crop growth. Conductivity increases with water temperature and is therefore related to air temperature (Hayashi, 2004; Ozaki et al., 2003). Sulfate release from soil can increase with increased temperature (Zhu et al., 2019).

Temperature is expected to influence water quality mostly on a longer time scale. A longer period of high daily maximum temperature can cause water quality problems, while one day of high temperature is not

expected to highly influence the water quality variables. Temperature drives slow processes that affect water quality on a longer time scale, such as evaporation.

The correlation between the average antecedent daily maximum temperature and the water quality parameters was obtained. An expanding window was used to calculate the lagged average daily maximum temperature for a window size of 10 to 100 days. The Pearson correlation (95% C.I.) between the average lagged daily maximum temperature and the current concentration was determined. The optimal lag is the window with the highest Pearson correlation (see Appendix 1D).

3.1.4.5 Discharge

Discharge influences river water quality through its composition of different flow components (Kalkhoff et al., 2016). Low flow mainly consists of base flow while higher flow has a greater contribution from surface runoff and overland flow, which may contain more contamination from agricultural soils. High discharges are therefore expected to have a higher concentration of nutrient pollutants than low discharges. This holds for nitrate, phosphate and pesticides which all originate from diffuse agricultural applications. Point sources are expected to negatively relate to discharge, due to the dilution effect. The relation between discharge and physicochemical water quality parameters such as conductivity and salination are complex and site-specific. Low discharges could lead to more seawater intrusion and therefore discharge might be negatively related to conductivity and salination. Effects like dilution might also play an important role here.

Discharge is expected to have a short-term predictive value. The current discharge and the average discharge over the past 1-4 days is used as input features to the model.

3.1.4.6 Land use

Many water quality parameters are related to land use (Ahearn et al., 2005; Tong & Chen, 2002). Runoff from agricultural land may be enriched with nutrients and sediments, while runoff from urbanized areas can contain larger compounds and heavy metals. Processes such as evapotranspiration, interception and infiltration can differ per crop type. Land use characteristics such as fertilizer input and cropping patterns play an important role (Kalkhoff et al., 2016). Pesticide concentration and timing are crop-dependent and therefore expected to be influenced by crop type.

The percentage of catchment area covered per crop type was used as a yearly feature to define the land use in the area. Each crop type is defined as a distinct feature. In total, thirteen crop types were distinguished: grassland, corn, 'grains, seeds and legumes', 'vegetables and spices', 'potatoes', 'agricultural infrastructure', 'sugarbeets', 'cattle feed', 'fruits and nuts', water, hemps, 'woodlike crops', and 'other crops'.

3.1.4.7 Summary

All above mentioned variables are transformed into useful features for modelling. A summary of the engineered features is listed in Table 3 below. A total of (n+57) input features is used in each model, in which n is determined by the autoregressive behavior of each water quality variable.

Table 3: Feature engineering for each variable.

Variable	Features [# of features]	Notes
Base	Lagged concentration [n]	n is determined by autocorrelation
	Rolling mean (window = n days) [1]	
	Rolling min (window = n days) [1]	
	Rolling max (window = n days) [1]	
	Rolling stdev (window = n days) [1]	
Calendar	Month [12]	One-hot encoded
Precipitation	Current precipitation [1]	*j* is optimal lag determined by correlation analysis
	Cumulative antecedent precipitation over the previous 0-10 days [10]	

	Cumulative antecedent precipitation over the previous (j-2)-(j+2) days [5]	
Temperature	Current temperature [1]	*k* is optimal lag determined by correlation analysis
	Average antecedent temperature over the previous (k-2)-(k+2) days [5]	
Discharge	Current discharge [1]	
	Average discharge over the previous 1-4 days [4]	
Land use	Crop type [13]	

3.2 Reservoir operation

Conjunctive use of surface and groundwater resources is a complex decision-making problem involving many decision variables and multiple objectives (Ahmadi et al., 2015; Alemu et al., 2011). This research is limited to surface water availability for conjunctive use, defined by the water quality compounds nitrate, phosphate, sulfate, conductivity, and bentazon. Reservoir operation can be decided on by simulating a reservoir, constructing production strategies, and finding the optimal solution based on performance assessment of key performance indicators (Li et al., 2010). A simulation model examines how the water system behaves under a given set of control actions (Lin & Rutten, 2016). Production strategies are sets of rules that define the values of the production rate from the reservoir. Retrospective analysis can be used to compare the performance of different production strategies using a modelled reservoir (Alemu et al., 2011; Allawi et al., 2019). A complete pattern of intake and production can be reconstructed with a reservoir simulation model. A predefined (production rate based on water quality measurements) and an adaptive forecast-based strategy (production rate based on water quality forecast) were tested with a modelled reservoir to compare the performance on two predefined key performance indicators. Drought is defined as days where intake is not possible. Summer is defined as 1 April – 1 October, winter is 1 October – 1 April.

3.2.1 Modelling the conjunctive use of surface and groundwater resources

The Blankaart reservoir model is based on the mass balance equation, which is a widely used approach for modelling simple reservoirs (Allawi et al., 2019; Lin & Rutten, 2016). The model consists of a reservoir, river intake, river outlet, reservoir outlet, and groundwater resource. The numerical scheme is defined by Equations 2-6 of which the components are explained in Table 4. The numerical scheme and the simulation were implemented with Python programming language.

$$V(t) = V(t-1) + \left(Q_{in}(t) - Q_{out}(t)\right) * dt \quad (2)$$

$$Q_{in}(t) = Q_{in,max} * f(x) + Q_{groundwater,extra}(t) \quad (3)$$

$$f(x) = \begin{cases} 1 \text{ if WQ is sufficient} \\ 0 \text{ if WQ is insufficient} \end{cases} \quad (4)$$

$$Q_{out}(t) = Q_{out,production}(t) + Q_{out,river}(t) \quad (5)$$

$$Q_{groundwater}(t) = D(t) - Q_{out,production}(t) + Q_{groundwater,extra}(t) \quad (6)$$

*Table 4: Components of the numerical scheme used to model the Blankaart reservoir. ** Artificially added.*

Symbol	Explanation	Unit
V	Volume of water in reservoir at the end of the day	m^3
Q_{in}	Inflow from river into the reservoir	m^3/day
Q_{out}	Total outflow out of the reservoir	m^3/day
$Q_{in,max}$	Pumping capacity inflow	m^3/day
$f(x)$	Step function to define when intake is possible	-
$Q_{groundwater,extra}$	Extra groundwater inflow to reservoir to keep minimum volume constraint **	m^3/day
$Q_{out,production}$	Production from the reservoir	m^3/day
$Q_{out,river}$	Extra outflow from reservoir to river to keep maximum volume constraint **	m^3/day
$Q_{groundwater}$	Total groundwater use per time step	m^3/day
D	Water demand	m^3/day

The time step of the numerical scheme is one day. The volume represents the volume at the end of the day (after the ingoing and outgoing fluxes of that day took place). Q_{in} is defined with a step function using the water quality measurements. If one of the water quality parameters exceed their limit concentration, intake is 0 m^3/day.

If intake is possible, the intake rate equals the intake pumping capacity. The hard constraints of the reservoir model are listed in Table 5 and should be satisfied at all times (Kendall, 1975; Lin & Rutten, 2016).

Table 5: Hard constraints for reservoir model de Blankaart.

Symbol	Constraint	Value
V_{min}	Minimum reservoir volume	600 000 m^3
V_{max}	Maximum reservoir volume	3 000 000 m^3
$Q_{in,min}$	Minimum inflow	0 m^3/day
$Q_{in,max}$	Pumping capacity inflow	142 000 m^3/day
$Q_{out,production,min}$	Minimum production rate	0 m^3/day
$Q_{out,production,max}$	Production capacity	40 000 m^3/day

Modelling the constraints is done through constraint rules, that are defined as follows:

- If Q_{prod} (t) > $Q_{prod,max}$:
  - $Q_{out,river}$ (t) = Q_{prod} (t) – $Q_{prod,max}$-> *let abundant water 'flow back' to river*
  - Q_{prod} (t) = $Q_{prod,max}$
- If Q_{prod}(t) < $Q_{prod,min}$:
  - $Q_{groundwater,\ extra}$(t) = Q_{prod}(t) – $Q_{prod,min}$ -> *take groundwater into reservoir*
  - Q_{prod}(t) = $Q_{prod,max}$

Two artificial in-/outflows are added to the model that do not exist in reality: $Q_{groundwater,extra}$ and $Q_{out,river}$. These fluxes are added to meet the model's hard constraints. In reality, if the reservoir tends to overflow because the intake is much higher than the production capacity, the intake rate would be lowered. In the model however, water can 'flow back' to the river through the additional outlet modelled as $Q_{out,river}$. Secondly, if the production rate tends to become smaller than the minimum production rate (0 m^3/day), groundwater should be taken into the reservoir. The production rate is set to 0 m^3/day , and the difference between the actual production rate and the desired production rate should be provided for from the groundwater source. This is done such that no "extra" water will be formed, and the water balance is closed. This extra groundwater source adds up to the actual groundwater used as supplementary source when production from the reservoir is not sufficient to meet the water demand (see Equation 6).

The initial conditions are listed in Table 6. The initial volume and initial production rate are defined for only the first day of the year. Volume and production rate are calculated through the reservoir operation rules for the rest of the year. The intake rate is externally defined but is chosen to be maximum for the first seven days of each year to overcome problems with the hard constraints for the reservoir model.

Table 6: Initial conditions for reservoir model de Blankaart.

Symbol	Condition	Value
V_{init}	Initial volume	$7*10^5$ m^3
$Q_{prod,init}$	Initial production rate	$4*10^4$ m^3/day
$Q_{in,init}$	Initial intake rate	142 000 m^3/day for first 7 days

3.2.2 Developing production strategies

Reservoir operation can be based on a combination of historical long-term series and real-time rules (Ahmadi et al., 2015). In real-time operation, the production rate (or any other decision variable) depends on other variables that hold in the current time step. The operator makes a final operational decision considering all important variables. The traditional production strategy is a predefined strategy based on past experiences of operating the studied reservoir and represents the currently adopted production strategy. The strategy is based

on water quality measurements rather than a forecast, so no water quality predictions are used in reservoir operation.

Uncertainties in climatic conditions, caused by climatic change, make predefined operational strategies less favorable and stresses the need for a real-time adaptive approach (Li et al., 2010). Adaptive forecast-based approaches include the formulation of reservoir operation rules based on a forecast of future system variables, which is updated at regular intervals leading to adjustments in the operational plan (Alemu et al., 2011; Dagli & Miles, 1980; Gavahi et al., 2019). The second production strategy is an adaptive forecast-based approach, always using the latest available information such that an optimal strategy can be achieved. The operational rules are based on predictions of the input parameter over a certain planning horizon. The approach is an anticipative rather than a reactive one. Using water quality forecasting models, the operator knows when a period of low water quality (no intake) starts and when the period ends. The production rate can be adjusted accordingly, such that optimal performance can be achieved.

The two production strategies were developed with a set of rules that define the production rate for each time step, driven by either the water quality measurements (predefined strategy) or a combination of water quality measurements and forecasts (adaptive forecast-based strategy).

3.2.3 Comparing production strategies

The optimal production strategy can be found by comparing the production strategies with defined simulation goals and constraints (Pereira et al., 2021; Snyman & Wilke, 2018). The hard constraints are the reservoir boundary conditions mentioned in section 3.2.1. When multiple simulation goals are defined, trade-offs between goals can occur. The simulation goals of this research were:

1) Meet water demand
2) Minimize groundwater use
3) Minimize risk of water shortages

The first goal is to always meet the water demand in the area. This goal can be achieved by either producing from surface water reservoir or from the groundwater resource. The second goal applies to the ratio between production from either surface water or groundwater. As much as possible water should be produced from the surface water resources, such that groundwater use is limited. The third goal prescribes that water shortages should be averted. Having low volume in the reservoir has negative effects on the water quality and intensifies the subsequent production steps. Water shortages can cause a complete shutdown of the Blankaart reservoir, which has negative effects on the drinking water quality.

The groundwater resources in the model are unlimited which means that the first simulation goal will always be achieved. The performance of the strategies is therefore compared with two key performance indicators that correspond to the latter two simulation goals.

3.2.3.1 Performance assessment

Hashimoto et al. (1982) define three criteria for evaluating the performance of water resource systems that are widely accepted and used in other water resources research: reliability, vulnerability, and resilience (Ahmadi et al., 2015; Allawi et al., 2019; Bolouri-Yazdeli et al., 2014).

Reliability is a measure for how likely a system is to fail (unsatisfactory performance) (Hashimoto et al., 1982). In the current research, unsatisfactory performance is defined as the time in which the production rate from the Blankaart does not equal the production capacity, and additional groundwater is needed in replacement ($Q_{prod} \neq Q_{prod,max}$). Reliability is therefore described with the total groundwater use (**TGWU**) in m^3, which corresponds to the second simulation goal: minimize groundwater use. The daily groundwater use is calculated with Equation 6. The TGWU is the sum of all values for the daily groundwater use.

Vulnerability represents the impact of consequences of failure (Hashimoto et al., 1982). It is described with the second KPI: days of minimum volume (DMV) in %, which corresponds to the third simulation goal: minimizing the risk of water shortages. It is the percentage of summer (1 April – 1 October) days that the reservoir volume exceeds the minimum drought volume ($V_{min,drought}$). The minimum drought volume is a soft constraint that describes the risk of water shortages. An elaboration of the DMV indicator can be found in Appendix 1E.

Resilience is a measure for how quickly a system recovers from failure. Resilience is not considered in the current research since it mostly depends on hard system constraints that cannot be altered with a new production strategy.

3.2.3.2 Retrospective analysis

Figure 7 shows the average, minimum and maximum summer production rates for 2011 – 2021.

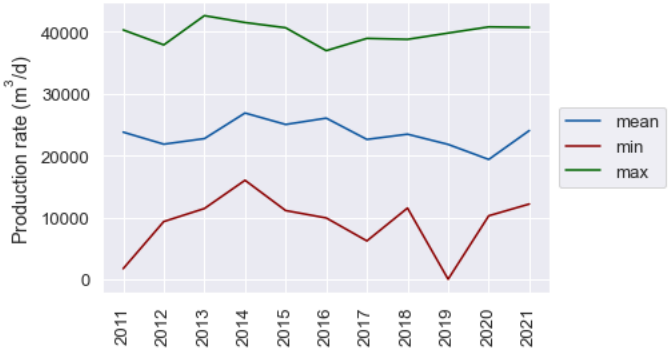


Figure 7: Average, minimum and maximum summer (1 April – 1 October) production rates from the Blankaart for 2011 -2021.

2018, 2019 and 2020 were experienced as severely dry years with many problems in drinking water production (De Watergroep, 2021b). 2019 shows a minimum summer production of zero, which indicates that the Blankaart was completely switched off due to water quality problems. 2020 has the lowest average summer production. 2021 was a year with less water quality problems. Data from 2011 – 2018 are used to train the water quality model and therefore not suitable for evaluation. The years 2019 and 2020 are used to evaluate the production strategies under dry circumstances. 2021 is used to assess the performance under normal water quality conditions.

Figure 8, 9, and 10 show the potential intake patterns for respectively the years 2019, 2020 and 2021. The intake patterns are based on the water quality measurements that were done during all years. If all considered water quality parameters (NO3, oPO4, SO4, conductivity and bentazon) allow intake, the intake rate is always maximum ($14.2*10^4$ m^3/day). If one of the water quality parameters exceeds their limit values, the intake rate is zero. All days where intake is not possible are considered 'drought days'.

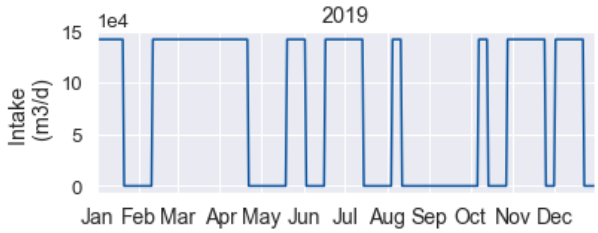


Figure 8: Potential intake pattern for the year 2019. If water quality is sufficient, intake is maximum. If water quality is insufficient, intake is zero.

The year 2019 has multiple short drought periods spread over the year, with a longer dry period in the end of summer. The total number of drought days is 170. This means that nearly half of the year, intake is not possible and $Q_{in} = 0$. The drought days are spread over eight periods that differ in length.

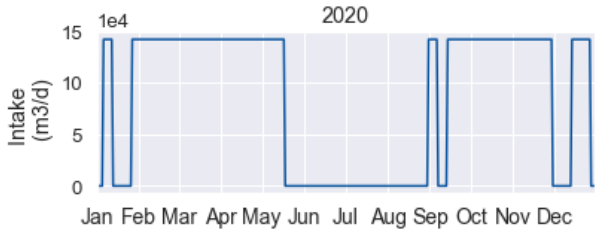


Figure 9: Potential intake pattern for the year 2020.

The year 2020 has one long period of consecutive drought days. Intake is not possible during the summer months of June, July and August. The total number of drought days is 149, which is less than in 2019. However, the long dry period in summer causes higher risk of water shortages than multiple short dry periods.

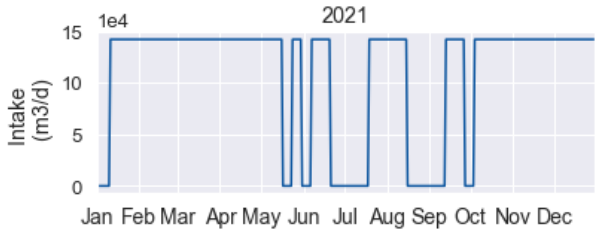


Figure 10: Potential intake pattern for the year 2021.

2021 is less dry than the other years and shows some short drought periods in summer. The total number of drought days is 87.

3.2.5 Sensitivity analysis

Model constraints influence model outcome. A sensitivity analysis is performed to assess how all initial and boundary conditions influence the key performance indicators TGWU and DMV for the adaptive forecast-based strategy. The considered constraints are the initial volume and initial production rate, the minimum and maximum reservoir volume, the production capacity, and the intake rate. Each parameter is altered individually within a sensitivity range with all other parameters fixed on their set value.

Table 7: Sensitivity analysis. All parameters are individually altered, with all other parameters fixed on their set value.

	Symbol	Set value	Sensitivity range	Unit
Initial volume	V_{init}	$7*10^5$	$7*10^5 – 3*10^6$	m^3
Initial production rate	Q_{init}	$4*10^4$	$0 – 4*10^4$	m^3/d
Minimum reservoir volume	V_{min}	$7*10^5$	$1*10^5 – 1*10^6$	m^3
Maximum reservoir volume	V_{max}	$3*10^6$	$1*10^6 – 1*10^7$	m^3
Production capacity	$Q_{prod,max}$	$4*10^4$	$2*10^4 – 6.5*10^4$	m^3/d
Intake rate	Q_{in}	$1.42*10^5$	$2.2*10^4 – 2.2*10^5$	m^3/d

3.2.6 Uncertainty analysis

An uncertainty analysis is performed to show what the implications of the water quality forecasting model performance are for the key performance indicators of the reservoir operation. Prediction intervals are

constructed as a measure for model uncertainty. Higher model performance should lead to smaller prediction intervals. An approximation of the uncertainty is found with the standard deviation between the predictions and measurements for each month of the year. A confidence level of 95% is used because of its widely accepted use. The upper and lower bounds of the prediction intervals are used to calculate the extremes and to show how these influence the key performance indicators.

4. Data

This chapter describes the water quality and climatic data that was used in the research.

4.1 Water quality data

The Blankaart intake strategy is quality-dependent and driven by the concentration of several compounds which are measured on several measuring stations along the river Yser. The parameters considered in this research are nitrate (NO3), (ortho-)phosphate (oPO4), sulfate (SO4), conductivity and bentazon, measured as closely as possible to the intake point of river water to the reservoir. These compounds represent nutrients, pesticides and the salination of water which are all important parameters for water quality. Appendix 1A explains the intake strategy. The availability of data differed per water quality parameter (see Table 8). All data was provided by De Watergroep.

Table 8: Data availability for water quality parameters.

	Available time period	Measurement frequency
Nitrate	2011-2021	2011 = monthly, other years = semi-weekly
Phosphate	2013-2021	2013 – July 2016 = monthly, other years = semi-weekly
Sulfate	2011-2021	2011 = monthly, other years = semi-weekly
Conductivity	2011-2021	2011 = monthly, other years = semi-weekly
Bentazon	2011-2021	All years = semi-weekly

4.1.1. Data preprocessing

All duplicate values (measured on the same day) were replaced by the average of the double measurement. Some data contained outliers (Table 9). These peaks were incidentally caused and do not represent normal system behavior and are thus not useful for modelling. This concerns peak concentrations of NO3 in June 2016 and peak concentrations of SO4 and conductivity in April 2011. The SO4 and conductivity peak was caused by incidental industrial disposal. Outliers were replaced with NAN values.

Table 9: Outlier detection water quality data.

	Parameter	Date	Value	Reason
1	NO3	06-06-2016	75 mg/l	Unexpected increase and peak, very uncommon in time of the year.
2	SO4	07-06-2011 till 05-07-2011	210 – 230 mg/l	Peak caused incidentally by industrial dump. Identified by client.
3	Conductivity	07-06-2011 till 05-07-2011	3900 – 4200 µS/cm	Peak caused incidentally by industrial dump. Identified by client.

NAN (Not A Number) values are datapoints with no available data. No NAN values were detected before removing outliers. After replacing the above-described outliers by NAN values, the data contained 5 NAN values (NO3 = 1 NAN, SO4 = 2 NAN's, Conductivity = 2 NAN's). NAN values were replaced with linear interpolation.

The water quality data were irregularly spaced time series, meaning that the interval between the observations is not constant. The measurement frequency was not equal for all water quality parameters over the complete time series of 2011 – 2020. To overcome this problem, all water quality series were first upsampled and linearly interpolated to daily data. The daily time series was resampled with a weekly frequency that optimally matched the original time series such that as much as possible of the original weekly data remained. Figure 11 shows a made-up example of this resampling strategy.

Original series
02-04
09-04
14-04
21-04
01-05

Fri	Sat	Sun	Mon	Tue	Wed	Thu
01-04	02-04	03-04	04-04	05-04	06-04	07-04
08-04	09-04	10-04	11-04	12-04	13-04	14-04
15-04	16-04	17-04	18-04	19-04	20-04	21-04
22-04	23-04	24-04	25-04	26-04	27-04	28-04
29-04	30-04	01-05	02-05	03-05	04-05	05-05

Figure 11: Made-up example of resampling strategy. Saturday and Thursday are the optimal resampling weekdays because most of the original data is maintained. Green cells represent datapoints in which the original data is remained.

4.1.2 Nitrate

Figure 12 shows the long-term time series and seasonal plot of nitrate concentration before data preprocessing. Figure 12A clearly shows the development of nitrate problems over the past years. In the years 2016-2020, the alarm concentration was exceeded every year. Figure 12B shows the seasonal behavior of nitrate. Contrary to the other water quality parameters, the nitrate concentration increases in the winter months and is low in summer.

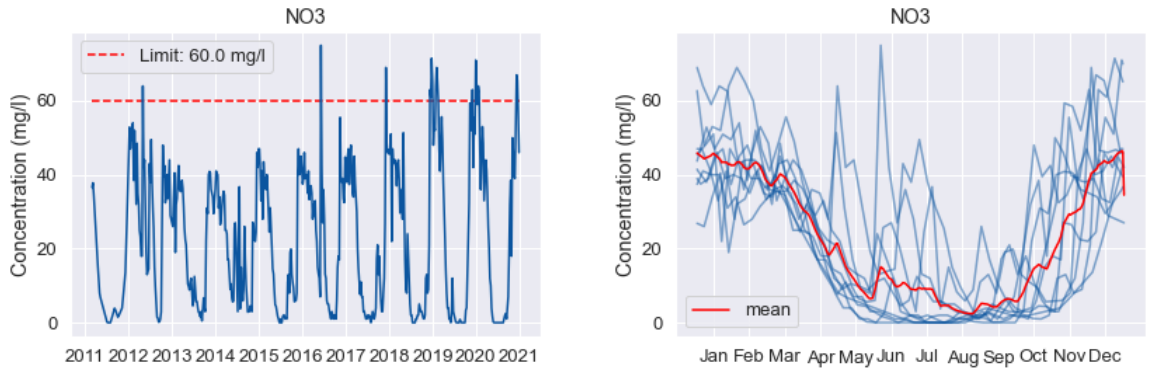


Figure 12: A) Long-term time series (2011-2021). B) Seasonal plot (2011-2021). Intake from the river is interrupted when the measured NO3 concentration exceeds 60 mg/l.

Western Flanders and especially the Yser catchment is one of the most nitrate-polluted areas of Belgium (Vlaamse Milieumaatschappij, 2021a). Emissions of nitrate to the surface water system are mainly caused by agricultural applications (Vlaamse Milieumaatschappij, 2022). Nitrate is a common substance of fertilizer used in agriculture. Limited crop growth during dry periods leads to less uptake of nitrate from applied fertilizer. This causes more wash-off of nitrate to the groundwater and surface water systems when precipitation increases during winter (Vlaamse Landmaatschappij, 2021).

4.1.3 Phosphate

Figure 13 shows the long-term series and seasonal plot of (ortho-)phosphate concentration before data preprocessing. The long-term time series shows that problems with oPO4 have increased over the past years. The seasonal plot shows that most water quality problems regarding oPO4 occurred in the summer months June, July, August, September and October. Data between 2011-2013 were not available.

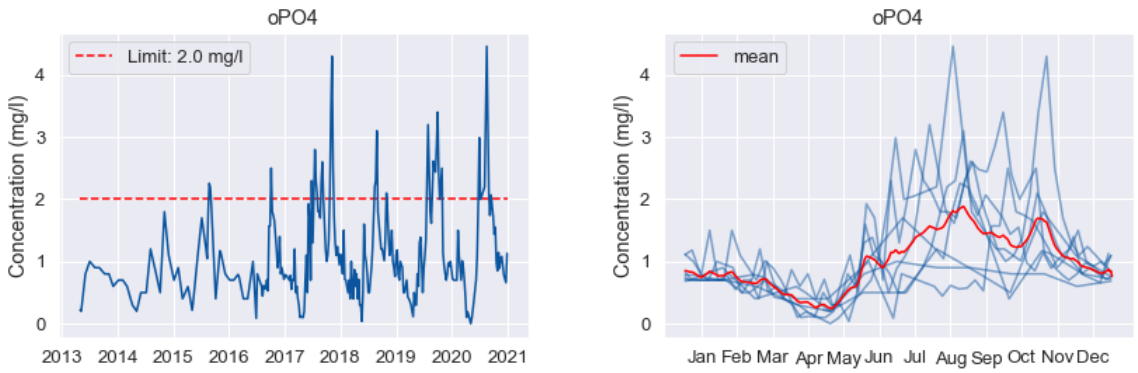


Figure 13: A) Long-term time series (2013-2021). B) Seasonal plot (2013-2021). Intake from the river is interrupted when the measured oPO4 concentration exceeds 2 mg/l.

Similar to nitrate, phosphate is an important substance of manure (Vlaamse Milieumaatschappij, 2021b). Current phosphate problems are mainly caused by historical over-usage of manure (De Watergroep, 2021b). Manure contains high amounts of phosphate which may highly exceed the uptake capacity of crops. This historically led to phosphate accumulation in soils. When soil is saturated with phosphate, the surplus of phosphate will wash off to the surface water system. Phosphate problems mainly occur in summer (Baken et al., 2016).

4.1.4 Sulfate

Figure 14 shows the long-term time series and seasonal plot of the sulfate concentration before data preprocessing. Figure 14A shows a large peak in SO4 concentration in 2011. This peak was caused by incidental industrial disposal (De Watergroep, 2021b). Peak concentrations of SO4 have increased in the past years. Figure 14B shows that the SO4 concentration in the river is relatively stable over the year. In some years, extreme values of the concentration are visible, which mainly occur during summer.

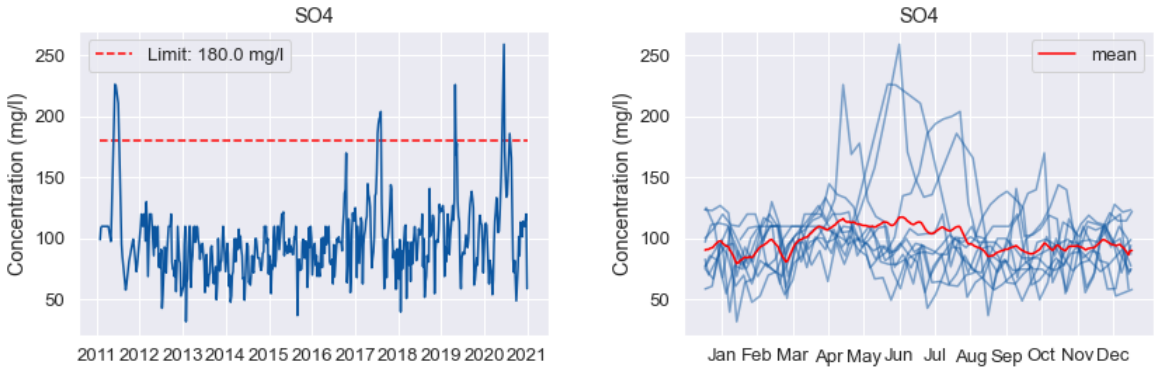


Figure 14: A) Long-term time series (2011-2021). B) Seasonal plot (2011-2021). Intake from the river is interrupted when the measured SO4 concentration exceeds 180 mg/l.

Industrial wastewater contains high salt concentrations, such as sulfate (Silva et al., 2002). Besides agricultural activity, the West Flanders region contains an industrialized area, of which the wastewater gets discharged in the river Yser and the surrounding channels. Seawater intrusion is expected to influence sulfate concentrations as well.

4.1.5 Conductivity

Figure 15 shows the long-term time series and seasonal plot of conductivity before data preprocessing. Similar to the SO4 concentration, the conductivity peak in 2011 was caused incidentally by an industrial dump (De Watergroep, 2021b). Conductivity correlates with SO4, showing similarities in the occurrence of limit concentration exceedance. The seasonal plot shows that conductivity mainly causes problems in summer and not in winter.

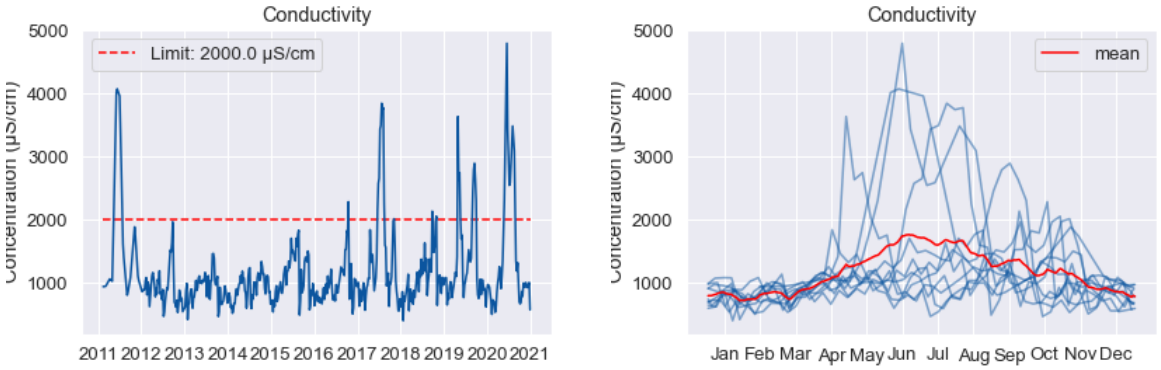


Figure 15: A) Long-term time series (2011-2021). B) Seasonal plot (2011-2021). Intake from the river is interrupted when the measured conductivity exceeds 2000 µS/cm.

Conductivity approximates the salinity in water. The Yser catchment is located in a coastal area, naturally more vulnerable to salinization problems (Coordinatiecommissie Integraal Waterbeleid, 2016). Groundwater in the area is salinated. Seepage of groundwater to surface water can cause elevated conductivity during periods of drought (Coordinatiecommissie Integraal Waterbeleid, 2016). Similar to SO4, conductivity increases due to the dumping of industrial wastewater.

4.1.6 Bentazon

Figure 16 shows the long-term time series and seasonal plot of bentazon before data preprocessing. The long-term time series plot shows that bentazon has been causing water quality troubles for a long time. The peak concentrations of bentazon have decreased in previous years as a result of agricultural policies. The seasonal plot shows that the bentazon concentration is highly seasonal. The concentration is very low in winter while it increases severely during summer. Almost every year, the limit concentration of bentazon is exceeded.

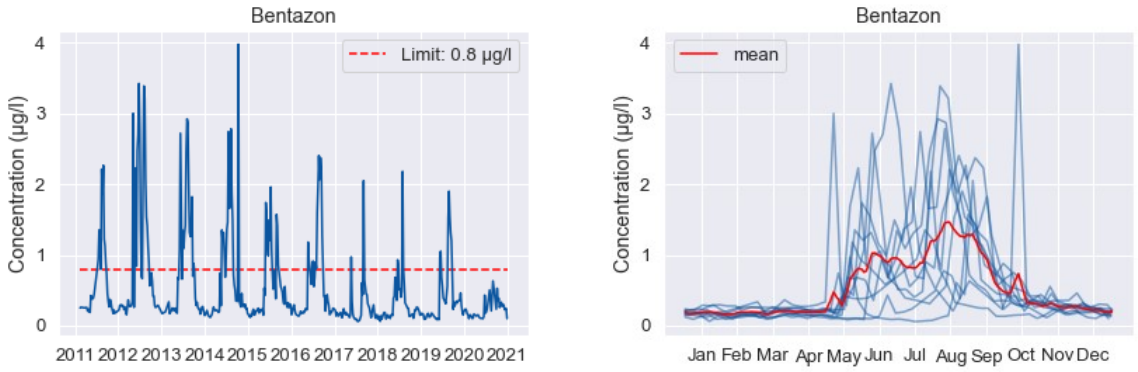


Figure 16: A) Long-term time series (2011-2021). B) Seasonal plot (2011-2021). Intake from the river is interrupted when the measures bentazon concentration exceeds 0.8 µg/l.

Bentazon is a herbicide applied to weeds at the start of the growing season in the production of maize, peas, beans and grass (KIWA NV, 1990; Rosenboom et al., 2021; Vlaamse Milieumaatschappij, 2017). In past decades, bentazon was found in surface water and groundwater systems and caused problems in drinking water production (Coordinatiecommissie Integraal Waterbeleid, 2016). Of all studied compounds, bentazon is the only one that does not appear naturally in the water system. All bentazon that ends up in the river originates from an artificial source. High pesticide concentrations can cause aquatic mortality and infertility. Intense purification steps are needed to produce safe drinking water from pesticide-polluted raw water. The use of bentazon in maize production was therefore prohibited by the Belgian government from 2018 onwards (Vlaamse Milieumaatschappij, 2017). Bentazon is not the only dangerous pesticide in the area. Of all measured pesticides, it has the highest data availability and is therefore chosen as the representative pesticide.

4.2 Climatic data

Precipitation, temperature, discharge, and land use are all expected to influence river water quality. The data availability for the climatic variables is summarized in Table 10. The data did not contain NAN values or any artificial outliers that needed extra data preprocessing steps.

Table 10: Data availability for climatic variables.

Parameter	Configuration	Source
Precipitation	Daily total precipitation for several rainfall stations	Vlaamse Milieu Maatschappij
Temperature	Daily maximum temperature for weather stations	Vlaamse Milieu Maatschappij
Discharge	Daily discharge measurements for discharge stations	Vlaamse Milieu Maatschappij
Land use	Yearly land use per agricultural plot	Geopunt Vlaanderen

Figure 17 shows the Yser catchment upstream of the Blankaart intake point, the Blankaart reservoir itself, and the rainfall and discharge stations used in this research.

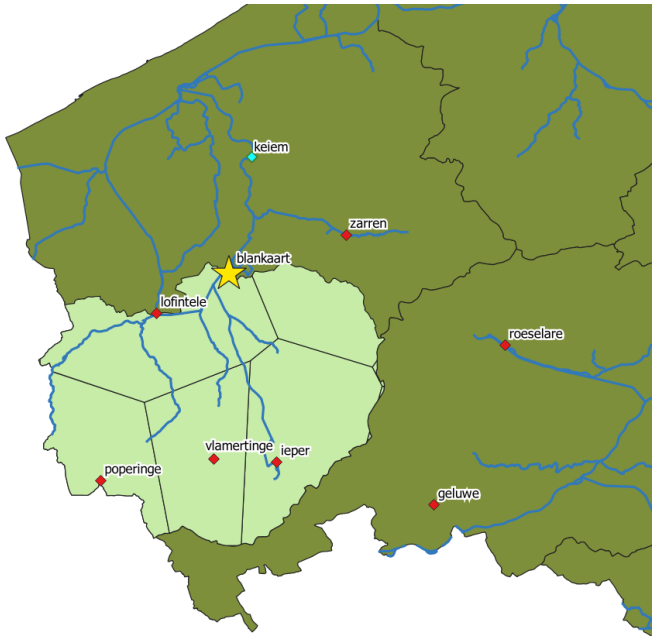


Figure 17: Yser catchment (light green) upstream of the Blankaart intake point and corresponding rainfall stations. Red diamonds = rainfall stations, blue diamond = discharge station, yellow star = Blankaart WPC.

4.2.1 Precipitation

Figure 18 shows a time series of the daily precipitation measured at the Lo-Fintele rainfall station and the rolling mean over 365 days to extract climatic trend. Boxplots were made to compare data from 2011-2018 to data from 2019-2021. Precipitation was measured at each rainfall station on a daily time scale.

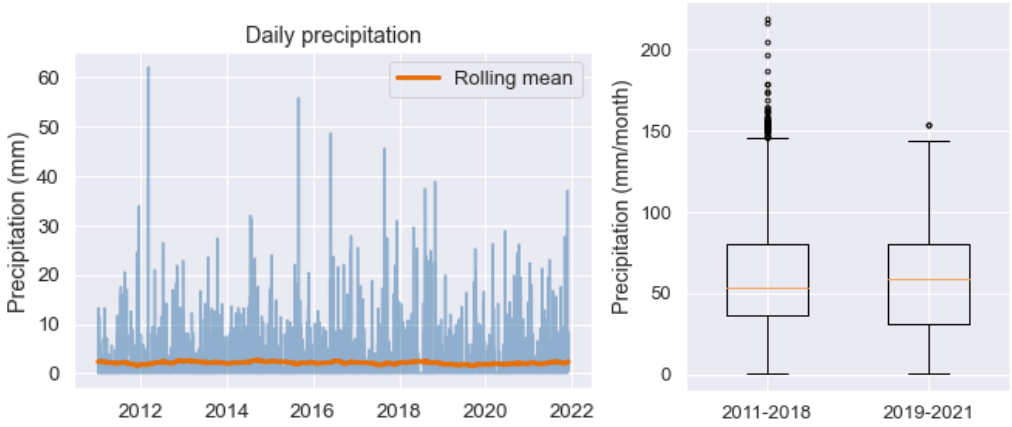


Figure 18: A) Time series and rolling mean (365 days) of daily precipitation in the Yser catchment measured at rainfall station "Lo-Fintele". B) Boxplot of monthly precipitation for resp. 2011-2018 and 2019-2021. Monthly precipitation was used for visibility purposes.

Much variability in the daily precipitation pattern can be found. Extreme precipitation events occur throughout the whole series (2012, 2015, 2016, 2017), and many days hold zero precipitation. The rolling mean shows that the average yearly precipitation remains relatively stable over the complete time series. The boxplots show that the period until 2019 has a lower median but more extreme precipitation values.

4.2.2 Temperature

The daily maximum temperature was obtained from temperature data that was measured at a 15-minute interval at weather station "Zarren", see Figure 17. Figure 19 shows the time series, climatic trend, and box plots for daily maximum temperature.

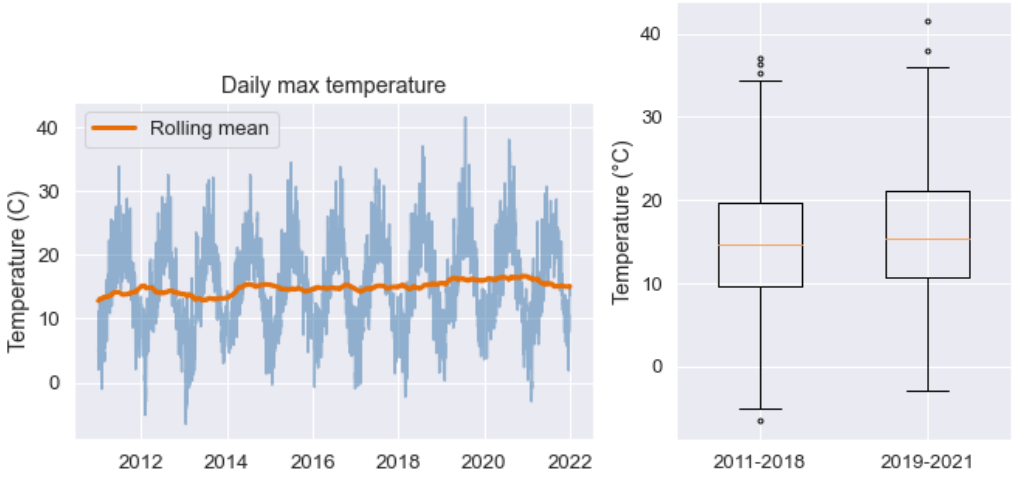


Figure 19: A) Time series and rolling mean (365 days) of daily maximum temperature in the Yser catchment measured at weather station "Zarren". B) Boxplot of daily maximum temperature for resp. 2011-2018 and 2019-2021.

The temperature shows a clear yearly pattern, with high temperatures in summer and low temperatures in winter. Most extreme values can be found in the summers of 2018-2020. The rolling mean shows a slight increase in daily maximum temperature over the time series. The boxplots show similar results, with a slightly higher median for the period 2019-2021 compared to 2011-2018. The minimum and maximum have also increased for the latest period.

4.2.3 Discharge

Discharge measured at station Keiem (see Figure 17) was used as input data for the discharge features of the model. The discharge station is located downstream of the intake point. Figure 20 shows the time series, climatic trend and box plots for daily maximum temperature.

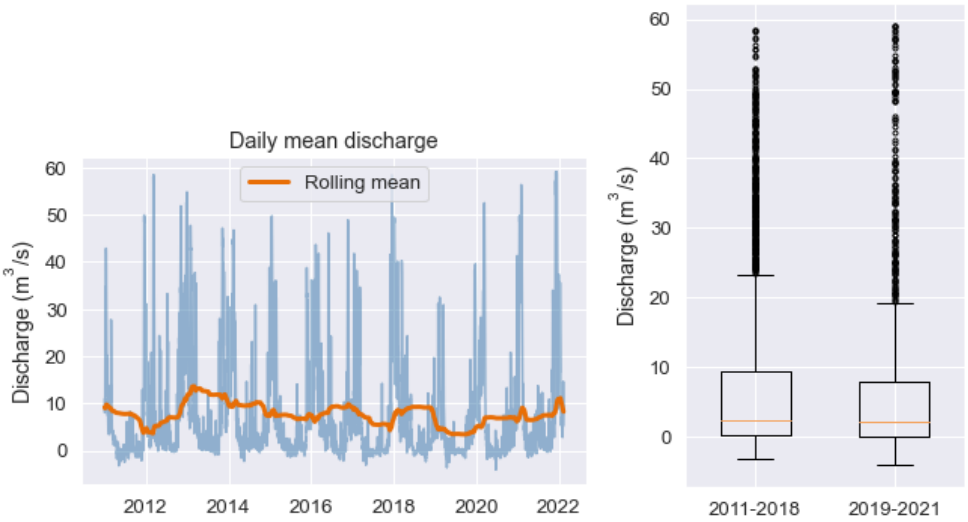


Figure 20: A) Time series and rolling mean (365 days) of daily mean discharge in the river Yser measured at discharge station Keiem. B) Boxplot of daily mean discharge for resp. 2011-2018 and 2019-2021.

The discharge time series shows a yearly pattern, with peak discharges in December and January of each year. No yearly trend can be observed from the rolling mean. The boxplots do not show a clear difference between the first and later studied periods.

4.2.4 Land use

The last climatic variable assessed in this research is the crop type percentage of land area of the Yser catchment upstream of the Blankaart intake point (see Figure 17). The catchment area covered the complete Belgian part of the Yser catchment upstream of the Blankaart intake point. Yearly data per agricultural plot is available through the geoportal of the Flemish government (Vlaamse Milieumaatschappij, n.d.).

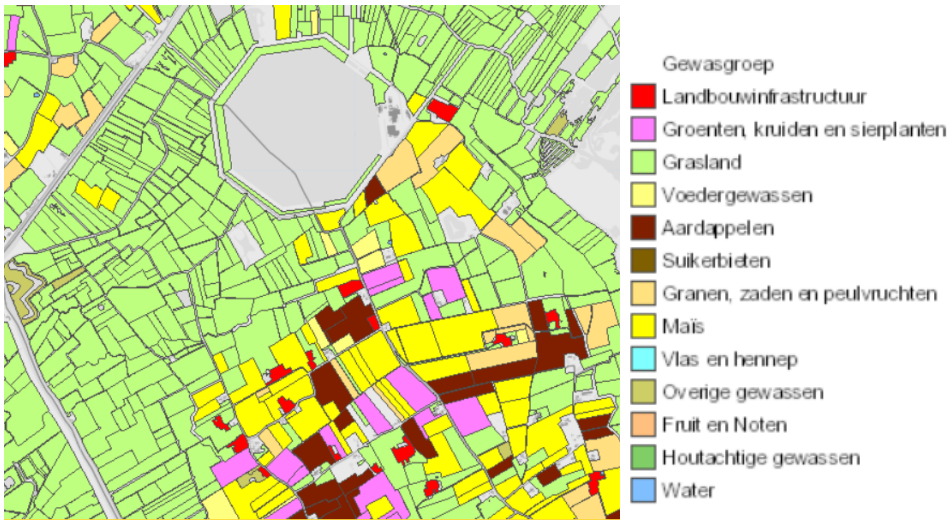


Figure 21: Example of agricultural land use per crop type (Vlaamse Milieumaatschappij, n.d.).

Land use data was obtained as geographical information from the GIS platform of the Belgian government (Vlaamse Milieumaatschappij, n.d.). The open source QGIS software was used to obtain the percentage of catchment area covered per crop type on a yearly basis, which is shown in Figure 22.

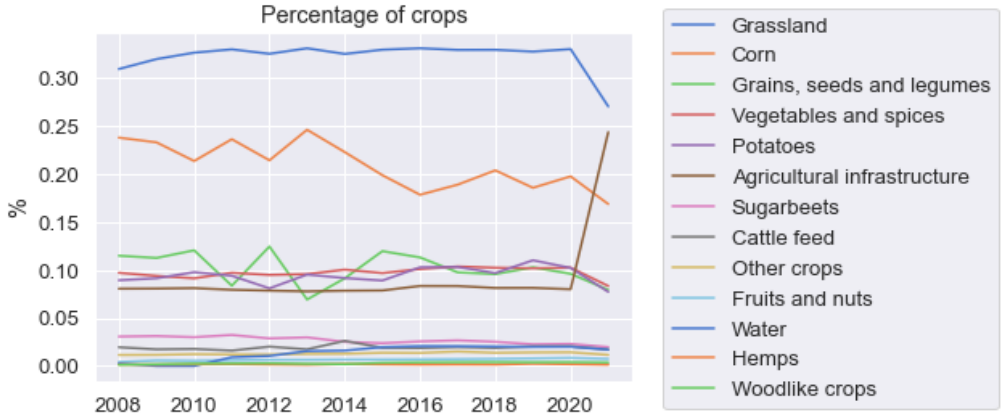


Figure 22: Percentage of catchment area covered per crop type.

Some crops cover only a very small land area and do not change much over the years, such as “Woodlike crops” and “Fruits and nuts”. Other crop types have a higher additional value to the land area, such as “Grassland” and “Corn”.

5. Results

This section describes the results for both water quality forecasting and reservoir operation.

5.1 Water quality forecasting

The results for water quality forecasting are divided into three types of analysis. The effects of including the climatic processes are described for each water quality parameter, together with the weighted optimal feature set. The subsection ends with summarizing the optimal performing models for each water quality parameter and forecast horizon, that were used to forecast the conjunctive use of surface and groundwater resources. The results of all feature sets and forecast horizons and all optimal performing models are included in Appendix 2A and 2B.

5.1.1 Nitrate

The Linear Regression algorithm gives the best performance for predicting nitrate on all forecast horizons, of which forecast horizon 1 is shown in Figure 23A and 23B.

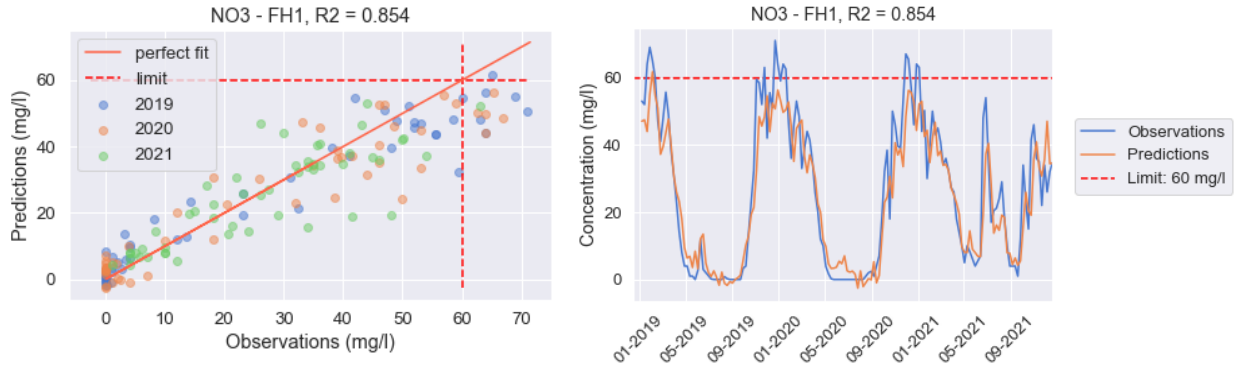


Figure 23: Nitrate – forecast horizon 1. A) Scatter plot showing predictions and observations for all tested years. B) Time series showing predictions and observations. Model: Linear Regression with Calendar + Precipitation feature set.

The model has an overall performance of R^2 = 0.854, which is considered good model performance. The model overestimates the nitrate concentration for low concentrations, while it rather underestimates medium to peak concentration. The models for the other forecast horizons show similar results, but with lower performance. Figure 24 shows the effect of including and excluding different features on the model performance for predicting 1 week, 2 weeks and 3 weeks ahead (resp. FH1 – FH3). The weighted optimal feature set includes calendar features, precipitation, and discharge (see Appendix 2C).

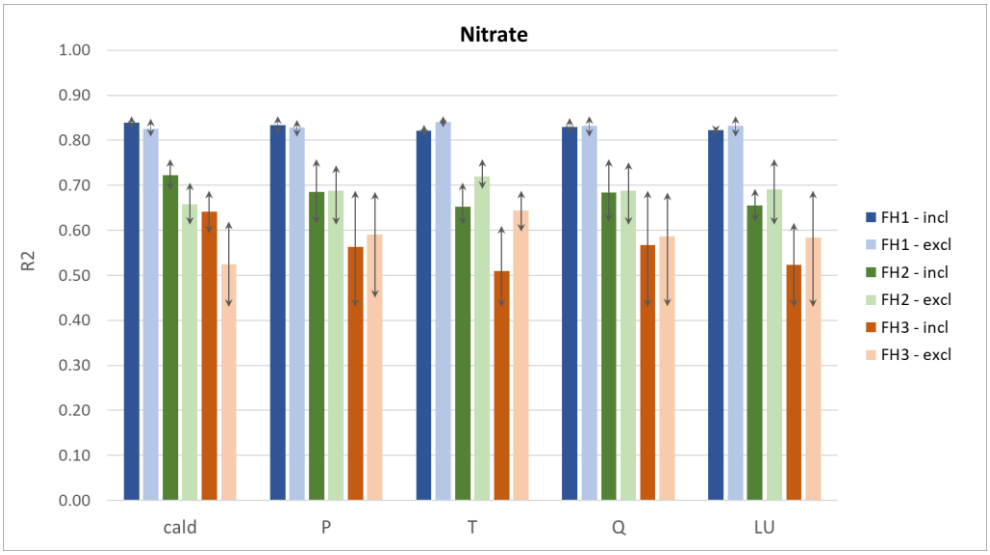


Figure 24: Effect of including calendar features (cald), precipitation (P), temperature (T), discharge (Q), and land use (LU) on model performance for nitrate (Linear Regression). Error bars show the minimum and maximum value for including or excluding the feature for each forecast horizon.

Calendar features

Including calendar features improves model performance for all forecast horizons, indicating that seasonal effects are important in the prediction of nitrate concentration. The minor improvement for forecast horizon 1 indicates that calendar features are not so important for forecasting one week ahead. Most improvement can be found for predictions further ahead in time.

Precipitation

Including precipitation does not improve model performance, except for a minor improvement for forecast horizon 1. Although the average model performance decreases when precipitation is included for forecast horizon 2 and 3, the best model performance was found for models that include precipitation. The large error bars indicate that the combination of precipitation with other features is important, but the individual effect of precipitation is low. This indicates that indirect processes have more predictive value than direct precipitation-related processes.

Temperature

Including temperature decreases model performance for all forecast horizons. The error bars do not overlap, which means that including temperature always has a negative effect on the model performance, and that the combination with other features does not change this. The most deteriorating effect on the model performance was found for the longest forecast horizon.

Discharge

The effect of discharge on the model performance is comparable to precipitation. Including discharge features has a minor negative influence on the model performance for all forecast horizons. The error bars are large, indicating that the individual effect of discharge is not important, but its combination with other features can cause different results. For FH2 and FH3, the best model performance was found for models that include discharge and precipitation, indicating that dilution is an important process here.

Land use

Land use has a negative effect on the model performance in the prediction of nitrate concentration, especially for longer forecast horizons. The large error bar for excluding land use is caused by the imbalance of experiments. The set of experiments excluding land use is 16, while only 4 experiments include land use. This causes a larger spread of results for experiments excluding land use and makes it more difficult to assess the individual effect of land use. The best model performance was never found for models that include land use.

5.1.2 Phosphate

The Random Forest Regressor algorithm gives the best performance for predicting phosphate on all forecast horizons, of which forecast horizon 1 is shown in Figure 25A and 25B.

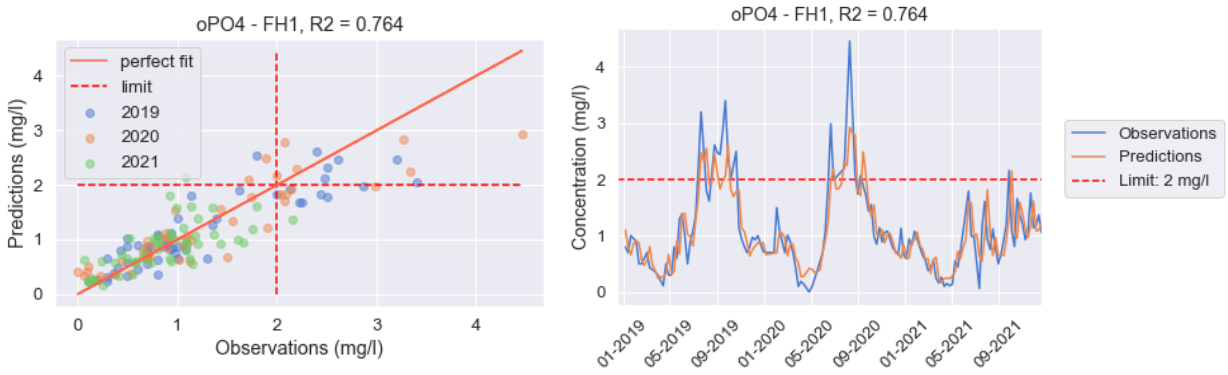


Figure 25: Phosphate – forecast horizon 1. A) Scatter plot showing predictions and observations for all tested years. B) Time series showing predictions and observations. Random Forest Regressor with Calendar + Precipitation + Temperature + Discharge + LU feature set.

The model has good overall performance of R^2 = 0.764. The model accurately predicts phosphate concentration for low to medium concentrations. Exceedance of the limit concentration is mostly predicted correctly, although the exact values of the peaks are underestimated. For longer forecast horizons, the model does not predict exceedance of the limit concentration correctly. Low concentrations are forecasted accurately for longer forecast horizons. Figure 26 shows the effect of including and excluding different features on the model performance for different forecast horizons. The weighted optimal feature set includes calendar features, precipitation, temperature, discharge, and land use (see Appendix 2C).

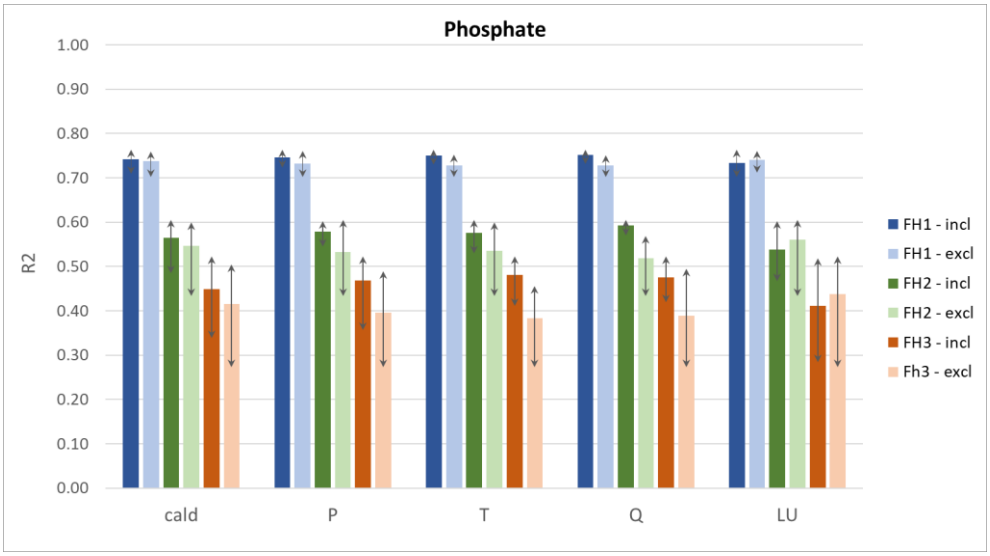


Figure 26: Effect of including calendar features (cald), precipitation (P), temperature (T), discharge (Q), and land use (LU) on model performance for phosphate (Random Forest Regressor). Error bars show the minimum and maximum value for including or excluding the feature for each forecast horizon.

Calendar features

Including calendar features generally improves model performance for all forecast horizons. The error bars are large, indicating that the combination between features is important. The largely overlapping error bars between including and excluding the feature indicate that there are other processes that are more important in improving model performance.

Precipitation

Including precipitation positively influences the model performance for all forecast horizons. For FH1 and FH2, the error bars do overlap minorly which means that precipitation is an important process in the prediction of phosphate. Forecast horizon 2 shows abnormal behavior. Including precipitation gives a better general performance with a small error bar, indicating that precipitation individually improves model performance. However, the optimal performing model does not include precipitation. For FH3, the error bars are larger which indicates that the combination of features with precipitation is more important than the individual contribution of precipitation.

Temperature

Including temperature improves model performance for all forecast horizons. The difference over forecast horizons is similar to the results found for precipitation. For larger forecast horizons, a larger improvement can be found when including the temperature feature. The small error bars strengthen the individual contribution of including temperature in the prediction of phosphate.

Discharge

Discharge improves the model performance for all forecast horizons. The error bars for FH2 are small and do not overlap with excluding the feature, so the model always improves when discharge is included. Discharge and temperature are mainly important on longer forecast horizons, while precipitation is more important for a short forecast horizon. Short-term changes in phosphate are prone to the wash-off effect of precipitation, while long-term changes are captured through dilution with temperature and discharge.

Land use

Land use negatively affects the model performance for all forecast horizons. The error bars largely overlap, so the model performance depends more on other features than on the land use features.

5.1.3 Sulfate

The Random Forest Regressor algorithm gives the overall best performance for predicting sulfate on all forecast horizons, of which forecast horizon 1 is shown in Figure 27A and 27B.

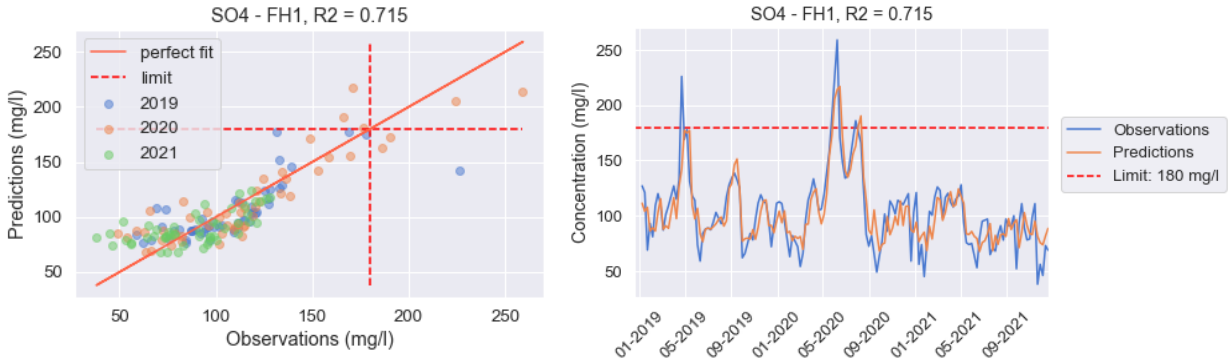


Figure 27: Sulfate – forecast horizon 1. A) Scatter plot showing predictions and observations for all tested years. B) Time series showing predictions and observations. Random Forest Regressor with Calendar + Precipitation + Discharge feature set.

The model has a moderate performance of $R^2 = 0.715$. The model captures the general behavior of the system and accurately forecasts when peaks are expected. Low concentrations are overestimated by the model while medium concentrations are underestimated. The small-scale changes in the lower concentration regime are not accurately forecasted, but also not important. For longer forecast horizons, the model overestimates sulfate concentrations rather than underestimating it (see Appendix 2B). The weighted optimal feature set includes calendar features, precipitation, and discharge (see Appendix 2C).

Figure 28 shows the effect of including and excluding different features on the model performance for predicting 1 week, 2 weeks and 3 weeks ahead (resp. FH1 – FH3).

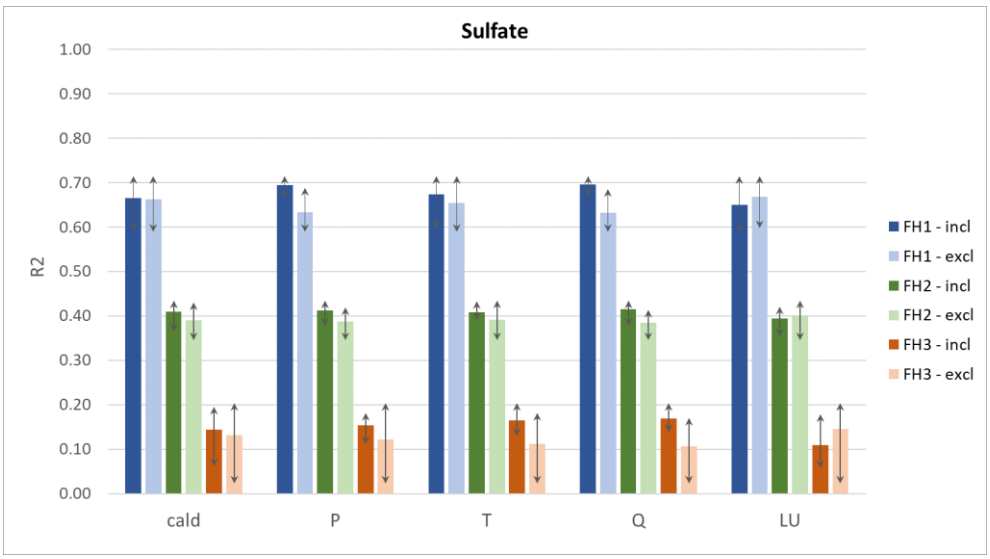


Figure 28: Effect of including calendar features (cald), precipitation (P), temperature (T), discharge (Q), and land use (LU) on model performance for sulfate (Random Forest Regressor). Error bars show the minimum and maximum value for including or excluding the feature for each forecast horizon.

Calendar features

Including calendar features generally improves model performance. For FH1, only a minor improvement can be found, with largely overlapping error bars. This shows that calendar features are not the most important parameter in predicting sulfate concentration. The optimal model performance for FH3 excludes calendar features, while on general the performance is better when calendar features are included.

Precipitation

The model performance for all forecast horizons improves when precipitation is included. The small error bars for including precipitation indicate that the individual contribution of precipitation is important. FH3 shows abnormal behaviour. The optimal model performs better when precipitation is excluded, although the overall performance is worse when precipitation is excluded.

Temperature

Including temperature improves model performance. The error bars largely overlap for FH1, indicating low importance for the first forecast horizon. Although the optimal model for both FH1 and FH3 does not include temperature, the general improved performance with small error bars indicates that temperature is an important process to include.

Discharge

Including discharge improves model performance for all forecast horizons. According to the small error bars for including the feature, the individual contribution of discharge is important. The greatest improvement for discharge can be found for forecast horizon 3. The results for including discharge are similar to the precipitation results.

Land use

Land use reduces model performance for all forecast horizons. The error bars largely overlap for all forecast horizons, so the performance depends more on other features.

5.1.4 Conductivity

The Random Forest Regressor algorithm gives the overall best performance for predicting conductivity on all forecast horizons, of which forecast horizon 1 is shown in Figure 29A and 29B.

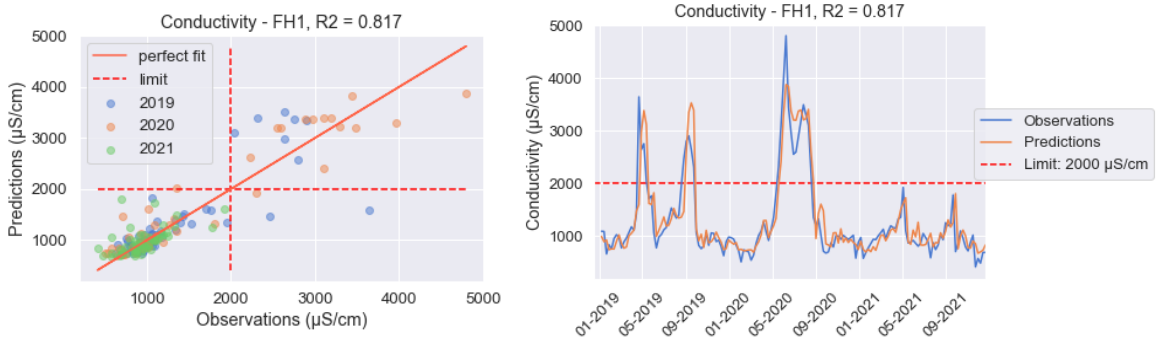


Figure 29: Conductivity – forecast horizon 1. A) Scatter plot showing predictions and observations for all tested years. B) Time series showing predictions and observations. Random Forest Regressor with Calendar + Precipitation + Temperature + Discharge + Land Use feature set.

The model has an overall good performance of R^2 = 0.817. The model accurately captures the general behavior of the system and predicts when conductivity peaks occur. Peak values are rather overestimated than underestimated. Low concentrations are more often underestimated. For longer forecast horizons, peak concentrations are more underestimated. The weighted optimal feature set includes calendar features, precipitation, temperature, discharge, and land use (see Appendix 2C).

Figure 30 shows what the effect of including and excluding different features is on the model performance for predicting 1 week, 2 weeks and 3 weeks ahead (resp. FH1 – FH3).

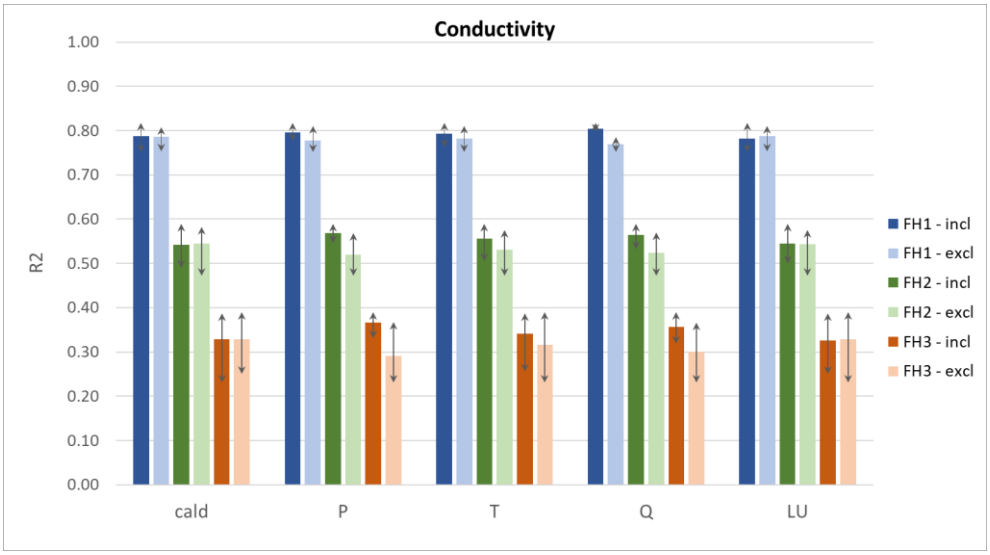


Figure 30: Effect of including calendar features (cald), precipitation (P), temperature (T), discharge (Q), and land use (LU) on model performance for conductivity (Random Forest Regressor). Error bars show the minimum and maximum value for including or excluding the feature for each forecast horizon.

Calendar features

Including calendar features generally reduces model performance for conductivity, but the large overlapping error bars indicate low feature importance. Performance is almost equal for including and excluding the calendar features. The general model performance is lower, but the optimal models do contain the calendar features for FH1 and FH2.

Precipitation

Precipitation improves model performance for conductivity. The small error bars show that the individual contribution of precipitation is important in predicting future conductivity, and that the combination with other features is not that important. Most improvement can be found for a longer forecast horizon.

Temperature

Temperature generally improves model performance but is not the most important process to include. The error bars largely overlap, indicating that the difference between including and excluding the feature is not that distinct. For FH3, the general performance is better when temperature is included, while the optimal model excludes temperature.

Discharge

Discharge positively influences the model performance. The small error bars indicate that the specific combination with other features does not change the performance much, so the individual contribution of discharge to the model performance is important. The results are comparable to the results found for precipitation, indicating that dilution is an important process here.

Land use

Land use generally reduces model performance. The results for including and excluding the feature are comparable, indicating that the feature is not of high importance in the prediction of conductivity. For FH3, the best model performance can however be found for a model that includes land use next to all other features.

5.1.5 Bentazon

The (Linear) Support Vector Machine Regressor algorithm gives the overall best performance for predicting bentazon concentration on all forecast horizons, of which forecast horizon 1 is shown in Figure 31A and 31B.

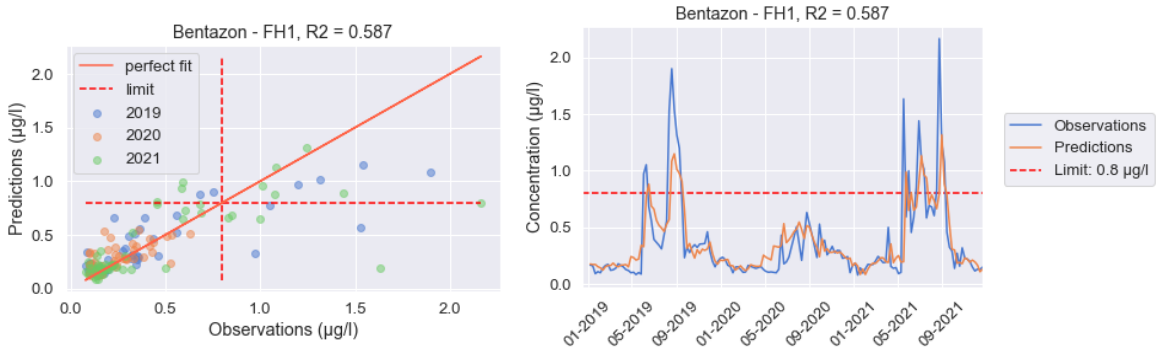


Figure 31: Bentazon – forecast horizon 1. A) Scatter plot showing predictions and observations for all tested years. B) Time series showing predictions and observations. Linear Support Vector Regressor with Calendar + Precipitation + Discharge feature set.

The model has moderate overall performance of R^2 = 0.587. The overall behavior of the system is well captured and the model predicts peaks. However, the peak concentrations are often underestimated by the model. For longer forecast horizons, the model does not correctly predict when peaks occur (see Appendix 2B). The weighted optimal feature set includes calendar features, temperature, and discharge (see Appendix 2C).

Figure 32 shows what the effect of including and excluding different features is on the model performance for predicting 1 week, 2 weeks and 3 weeks ahead (resp. FH1 – FH3). It should be noted that general model performance is low for bentazon, especially for higher forecast horizons, which complicates drawing straight conclusions about the relevant processes. The processes captured in the bentazon models are not able to explain most of the variability in concentration.

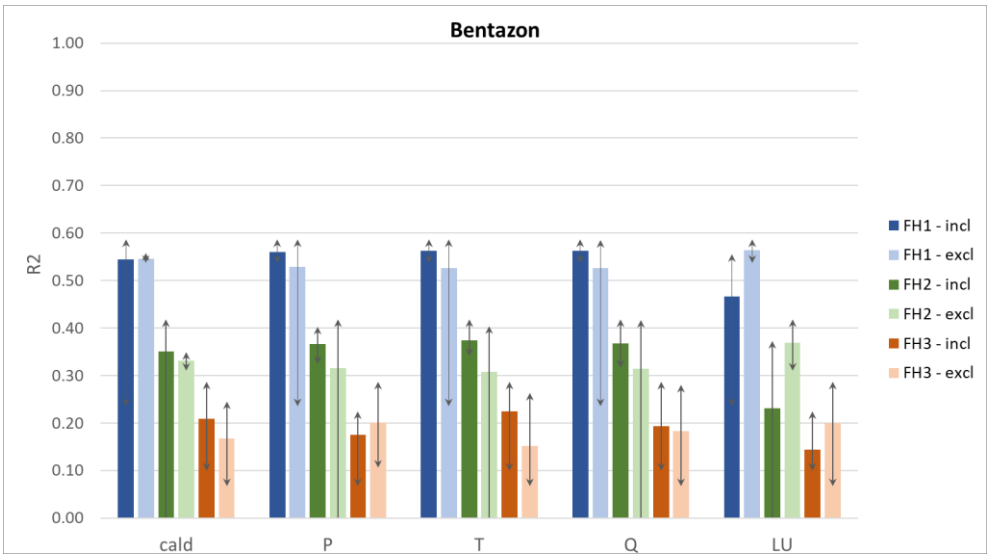


Figure 32: Effect of including calendar features (cald), precipitation (P), temperature (T), discharge (Q), and land use (LU) on model performance for bentazon (Random Forest Regressor). Error bars show the minimum and maximum value for including or excluding the feature for each forecast horizon. Large error bounds for incl cald, excl P, excl T, excl Q and incl LU are caused by negative performance for cald_LU.

Calendar features

The average model performance with excluding calendar features is higher for FH1, while the optimal model does include calendar features. Calendar features positively influence model performance for FH2 and FH3. The large negative error bound for FH1 and FH2 is caused by the outlier for experiment cald_LU.

Precipitation

For short forecast horizons, including precipitation positively influences model performance. The model performance decreases when precipitation is included for FH3. The best model performance for FH2 can however be found for a model that does not include precipitation. The error bounds for including the feature are small, such that the combination with other features is less relevant. In case of FH3, the combined features are more relevant.

Temperature

The results for the temperature features are comparable to the results for precipitation, except for FH3. Including temperature positively influences model performance for FH3. Again, the experiment cald_LU causes the lower error bound for excluding temperature. Small error bounds for FH1 and FH2 indicate high feature importance.

Discharge

The results for discharge are similar to precipitation. There is only a small difference in model performance between including and excluding discharge for FH3, with largely overlapping error bounds. The maximum performance for including or excluding discharge features lies relatively close to each other, indicating that in finding the optimal model, discharge is not an important process.

Land use

Land use generally decreases model performance for all forecast horizons. The large error bounds indicate low feature importance. Most of the variability in R^2 is caused by the other features and not by land use.

5.1.6 Synthesis

All calendar and climatic features except land use positively contribute to model performance in the prediction of phosphate, sulfate, conductivity, and bentazon (see Table 11). Model performance for nitrate can only be improved by including calendar features. The weighted optimal feature sets do not agree with the predictive value of including or excluding the individual processes.

Table 11: Predictive value of input features on the water quality parameters. +: increased model performance, -: decreased model performance. Blue cells show the weighted optimal feature set for each water quality parameter.

	Nitrate	Phosphate	Sulfate	Conductivity	Bentazon
Calendar	+	+	+	-	+
Precipitation	-	+	+	+	+
Discharge	-	+	+	+	+
Temperature	-	+	+	+	+
Land use	-	-	-	-	-

Calendar features have predictive power for all water quality compounds except for conductivity. This indicates that all water quality compounds except conductivity are predictable with their seasonal effects. For conductivity, other processes are more important and calendar features do not improve model performance. Including precipitation, discharge, or temperature, positively influences model performance for all water quality compounds except nitrate. Hydrological processes such as wash-off, dilution, and eutrophication are all important processes in the prediction of these water quality compounds. Land use negatively influences model performance for all water quality compounds. Defining the ratio of different crop types over the catchment area has no predictive powers in the models applied here.

The combination of features is important. The weighted optimal feature sets and the optimal performing models do not always agree with the findings including or excluding each climatic process individually. For nitrate, the weighted optimal feature set includes precipitation and discharge, although the average performance of including precipitation or discharge does not improve. More examples can be found in Table 11. This finding stresses the interrelatedness of the climatic processes and shows the importance of experimenting with different feature sets.

Including each climatic process influences the model performance in a similar direction (positively or negatively) for all forecast horizons. All models perform better on short forecast horizons than on longer ones, which indicates that the autoregressive behaviour of the input features is important. The influence of the climatic processes on the model performance is larger for longer forecast horizons, although the relative importance of the processes can be different.

5.1.7 Optimal performing models for reservoir operation

The optimal performing forecasting models and their performance for all water quality parameters and forecast horizons are shown in Figure 33 and Table 12. The results of all experiments (combination of algorithm, feature set, forecast horizon and water quality parameter) are attached in Appendix 2. Nitrate has the best overall model performance, followed by phosphate. Conductivity and sulfate perform well on a short forecast horizon, but less on a longer forecast horizon. Bentazon has the worst overall model performance.

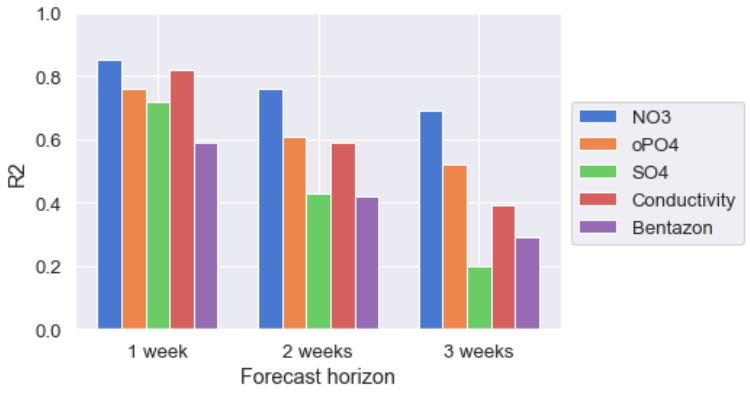


Figure 33: Model performance for all water quality parameters and all forecast horizons.

Table 12: Optimal performing models for water quality forecasting. P = Precipitation, Q = Discharge, T = Temperature, LU = Land Use. Green = good (R^2>0.75), yellow = moderate (0.25<R^2<0.75), orange = low (R^2<0.25) model performance.

Variable	Forecast horizon	Algorithm	Feature set	Model R2	Baseline R2	ΔR2
NO3	1 week	LR	Calendar + P	0.854	0.858	-0.004
	2 weeks	LR	Calendar + P + Q	0.758	0.722	0.036
	3 weeks	LR	Calendar + P + Q	0.687	0.59	0.097
oPO4	1 week	RFR	Calendar + P + T + Q + LU	0.764	0.737	0.027
	2 weeks	RFR	Calendar + T + Q	0.605	0.411	0.194
	3 weeks	RFR	Calendar + P + T + Q	0.523	0.241	0.282
SO4	1 week	RFR	Calendar + P + Q	0.715	0.489	0.226
	2 weeks	RFR	Calendar + P + Q	0.434	0.149	0.285
	3 weeks	RFR	Base + T + Q	0.203	-0.239	0.442
Conductivity	1 week	RFR	Calendar + P + T + Q + LU	0.817	0.743	0.074
	2 weeks	RFR	Calendar + P + T + Q + LU	0.589	0.475	0.114
	3 weeks	RFR	Base + P + Q	0.391	0.168	0.223
Bentazon	1 week	LSVR	Calendar + P/T + Q	0.587	0.452	0.135
	2 weeks	LSVR	Calendar + T + Q	0.418	0.066	0.352
	3 weeks	SVR	Calendar + T + Q	0.288	-0.214	0.502

All models perform better than the approximated persistence baseline model, except for the NO3 1 week forecast. The Random Forest Regressor performs best on phosphate, sulfate and conductivity, while Linear Regression and Support Vector Regression are the optimal algorithms for respectively nitrate and bentazon. The best improvement with respect to the baseline models was found for longer forecast horizons. The models in Table 12 were the models used in the adaptive forecast-based reservoir operation approach in the next subsection.

5.2 Reservoir operation

A predefined traditional production strategy and a new adaptive forecast-based production strategy were compared based on two key performance indicators under three different scenarios, to answer research question 2.

5.2.1 Developing production strategies

Two production strategies were developed and defined by a set of simple operational rules. The predefined production strategy is based on the current strategy that the drinking water company adopts. The water quality forecasting model was used to design an adaptive forecast-based production strategy.

5.2.1.1 Predefined production strategy

In the predefined production strategy of the Blankaart, the production rate depends on the number of consecutive weeks of low water quality, based on water quality measurements. If the number of consecutive weeks of low water quality exceeds a certain threshold, the production rate should be adjusted to half of the production capacity. This maximum number of consecutive weeks of low water quality depends on the time of the year. Figure 34 shows the daily average temperature and the maximum number of consecutive weeks of low water quality for each period. In summer, water quality is much more uncertain than in winter. This means that the response time in summer should be much shorter than in winter. The production rate should be halved as quickly as possible to save water in the reservoir to avert water shortages. Therefore, the maximum number of consecutive weeks follows the inverse pattern of the daily average temperature in the area.

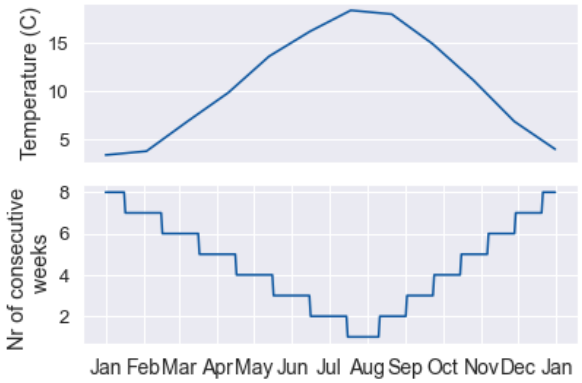


Figure 34: Daily average temperature in the area and the maximum number of consecutive weeks of low water quality.

The operational rules for the traditional production strategy are:

- If $Q_{in}(t\text{-}t_{limit}:t) = 0$ -> $Q_{prod}(t) = 0.5*Q_{prod,max}$
- If $Q_{in}(t\text{-}t_{limit}:t) > 0$ -> $Q_{prod}(t) = Q_{prod,max}$
- Constraint rules (see section 3.2.1)

5.2.1.2 Adaptive forecast-based production strategy

An adaptive forecast-based production strategy is proposed and illustrated in Figure 35, implementing the water quality forecasting models that were developed in the previous sections. The water quality forecast determines the production rate from the reservoir, but intake is still only driven by water quality measurements. If all water quality variables do not exceed the limit value, intake will be maximum. If the current water quality is insufficient, intake will be 0 m^3/day.

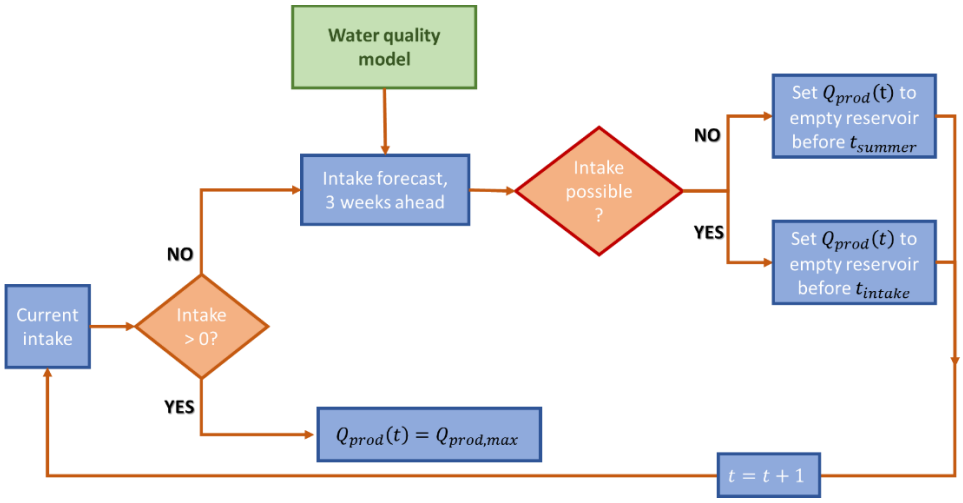


Figure 35: Adaptive forecast-based production strategy with water quality forecast.

A forecast horizon of three weeks is used. The forecast reliability is expected to be relatively good for forecasting three weeks ahead, due to autocorrelative processes (see Appendix 1B). First, the current time step is assessed. If intake at the current time step is possible, production is maximum. If intake at the current time step is not possible, the water quality forecast is used. If intake is then possible within the forecast horizon of three weeks, the reservoir should be emptied just before intake is possible. Since the intake capacity is much higher than the production capacity, the reservoir will be quickly filled when intake is possible after a dry period.

If the water quality remains low and intake is not possible in the forecasted period, the production rate should be adjusted such that the reservoir is emptied before the start of winter (1st of October). During winter, the demand is much lower and surface water availability is higher. This enables the operator to take more risk with emptying the reservoir at the end of summer. The forecast model is not a perfect forecast. During the forecasted period, the predicted intake can change and therefore the production rate can change as well.

The operational rules for the adaptive forecast-based production strategy are:

- If $Q_{in}(t) > 0$ -> $Q_{prod}(t) = Q_{prod,max}$
- If $Q_{in}(t) == 0$:
    - If intake is possible in forecast period -> *empty reservoir day before intake is possible*
        - $t_{empty} = t_{intake} - 1$
        - $V(t_{empty}) = V_{min}$
        - $Q_{out}(t) = (V(t-1)-V_{min})/dt$
    - If intake is not possible in that period -> *empty reservoir just before end of summer*
        - $t_{empty} = t_{winter}$
        - $V(t_{empty}) = V_{min}$
        - $Q_{out}(t) = (V(t-1)-V_{min})/dt$

5.2.2 Comparing production strategies

Both strategies were simulated with the reservoir model that is described in section 3.2.

5.2.2.1 Scenario 1: Multiple short drought periods

The first intake scenario is the year 2019, when multiple short drought periods occurred. Figure 36 shows the intake, volume, production, and groundwater pattern for both the traditional and the adaptive forecast-based production strategy.

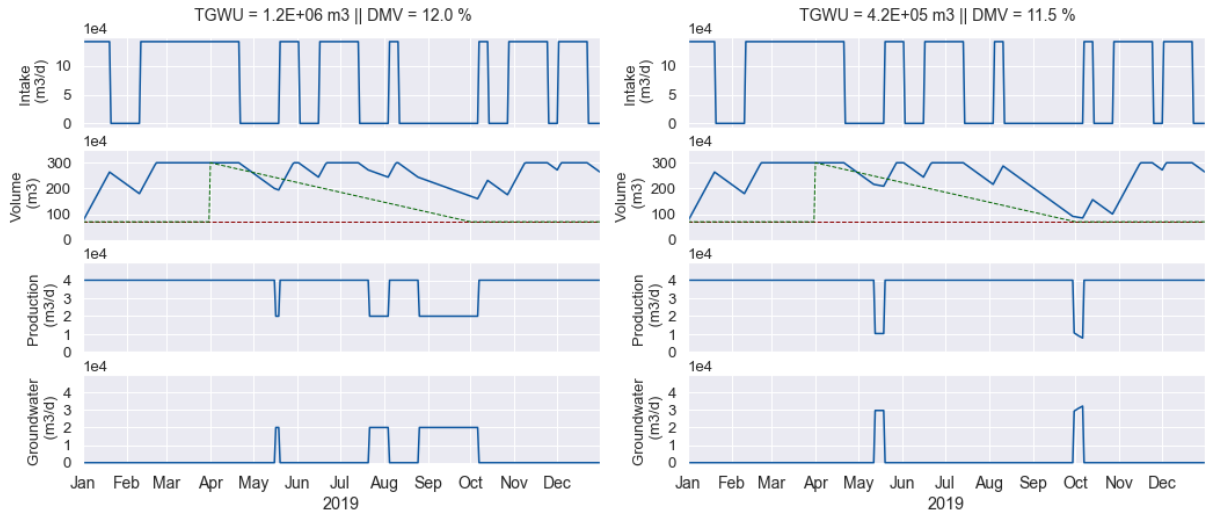


Figure 36: Traditional (left) and adaptive forecast-based (right) operation of the Blankaart reservoir, scenario 1. Red dotted line (Volume) = hard volumetric constraint. Green dotted line (Volume) = soft volumetric constraint.

Groundwater use

The traditional strategy has a total groundwater use of $1.2*10^6$ m^3, while the adaptive forecast-based strategy uses $4.2*10^5$ m^3 in a year. In the traditional strategy, the production rate is more often switched to a low rate, caused by the longer drought periods in July, August and September. Figure 34 shows that the operator under a traditional strategy should quickly change to a low production rate in summer to decrease the risk of water shortages. Using the adaptive forecast-based approach, clearly more risk can be taken. For example, during the drought period in July, no groundwater is used because of the forecasted inflow in the beginning of August. Overcoming the longer dry period by emptying the reservoir before the 1st of October clearly has the advantage of using more of the reservoir's water.

Risk of water shortages

The risk of water shortages, expressed in days of minimum volume, is similar for both years. In May, both strategies show exceedance of the soft volumetric constraint. The soft volumetric constraint is high because the risk of an empty reservoir is high, but the probability that the complete summer period is dry is actually low. The volume approaches the soft constraint for the new production strategy in August and September but does not exceed it.

5.2.2.2 Scenario 2: One long summer drought

Figure 37 shows the intake, volume, production, and groundwater pattern for the traditional and new strategy under scenario 2. This scenario resembles the year 2020 and shows one long drought period from halfway May until the start of September.

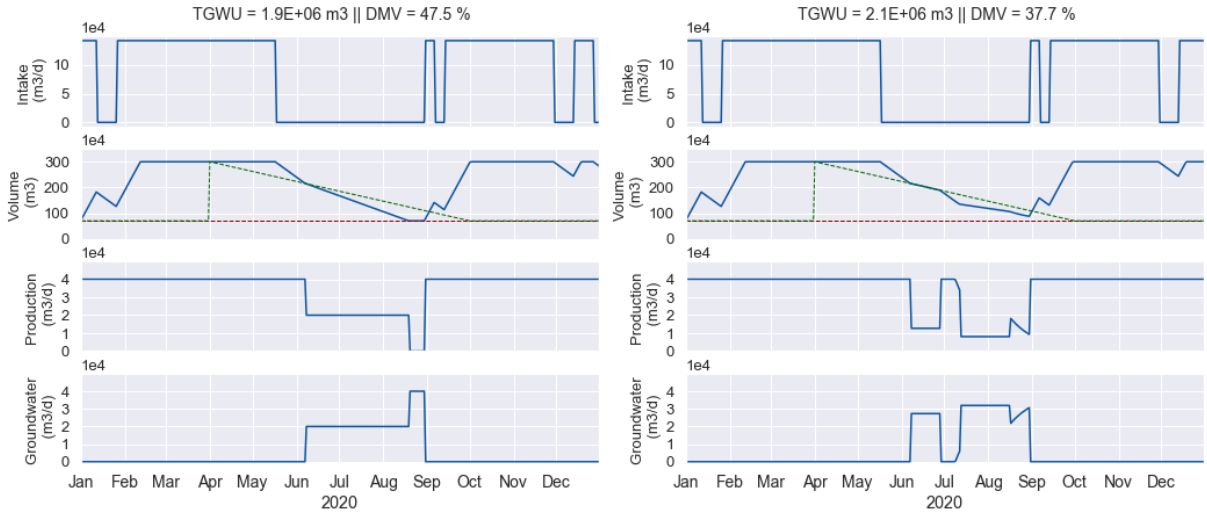


Figure 37: Traditional (left) and adaptive forecast-based (right) operation for the Blankaart reservoir, scenario 2. Red dotted line (Volume) = hard volumetric constraint. Green dotted line (Volume) = soft volumetric constraint.

Groundwater use

The total groundwater use is slightly higher for the new production strategy than for the traditional one. The new strategy adapts to the fact that intake is not possible in the forecast windows, while the traditional strategy keeps producing at half production until the volumetric constraint is reached.

Some indications of a wrong forecast can be found. Firstly, the production rate remains high at the end of May, while the intake pattern shows that intake is not possible within three weeks so the production rate should have switched to low production. Secondly, the production rate increases to maximum during the beginning of July, which indicates that the forecast must have predicted good surface water availability, while that is still not the case.

Risk of water shortages

The DMV is lower for the adaptive forecast-based strategy than for the traditional strategy. The long dry period causes the reservoir under the traditional strategy to be completely shut-off at the end of August, due to the hard volumetric constraint. This is not the case for the new strategy, where there is always a minimum production rate such that the reservoir does not have to be switched off completely.

The faulty high production rates at the end of May and the beginning of July cause less groundwater use which comes at the cost of risking water shortages, as can be seen by the volumetric constraint that quickly gets exceeded.

5.2.2.3 Scenario 3: Normal climatic conditions

Figure 38 shows the intake, volume, production, and groundwater pattern for a scenario under normal climatic conditions (2021).

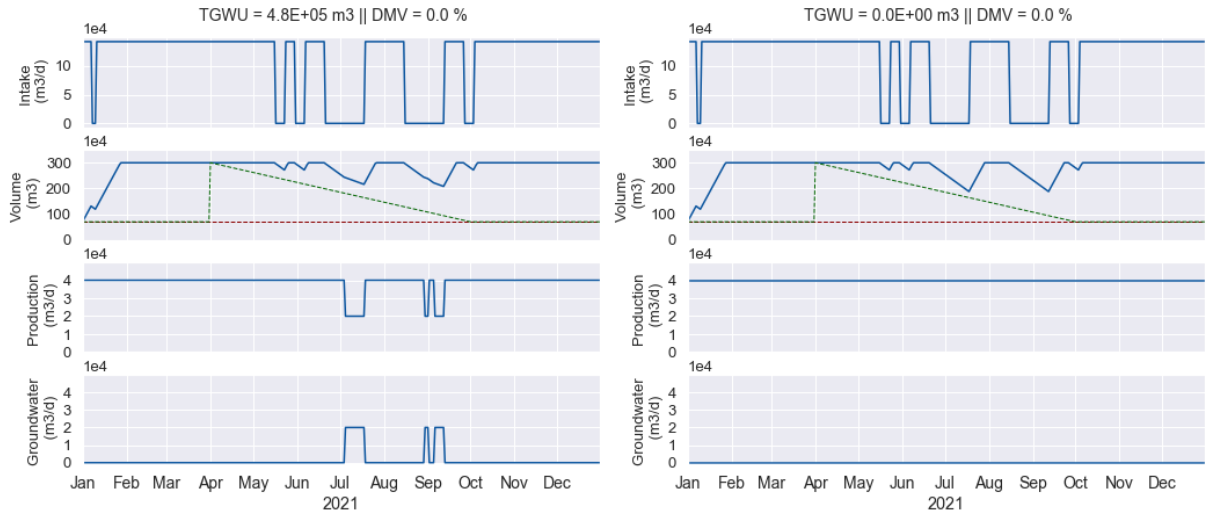


Figure 38: Traditional (left) and adaptive forecast-based (right) operation for the Blankaart reservoir, scenario 2. Red dotted line (Volume) = hard volumetric constraint. Green dotted line (Volume) = soft volumetric constraint.

Groundwater use

The traditional strategy uses more groundwater than the newly proposed strategy, but both are using a low amount of groundwater compared to the other studied years. The advantage of the forecast-based strategy can be seen in July and August, where the model correctly predicts the end of a drought period what results in a high production rate. Contrary, the drought periods cause low production rate in the traditional strategy which results in more groundwater use.

Risk of water shortages

The risk of water shortages is zero for both strategies. The volume never exceeds the soft volumetric constraint. The drought periods are short and do not happen at the riskiest moments of the year.

5.2.3 Summary

The results of both KPI's (Total Groundwater Use and Days Minimum Volume) for the years 2019-2021 are shown in Figure 39. A "perfect forecast"-based strategy is included to show what the implications of forecast performance for the adaptive forecast-based approach are. The optimal strategy should minimize both TGWU and DMV in different years.

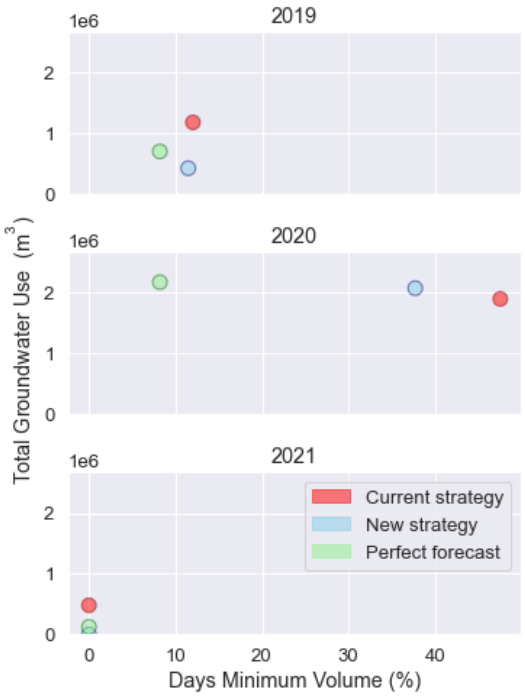


Figure 39: Summary of the key performance indicators TGWU and DMV for the traditional strategy, the adaptive forecast-based strategy and the perfect forecast-based strategy, evaluated over the years 2019-2021.

The adaptive forecast-based strategy shows advantages compared to the traditional strategy. The risk of water shortages, explained with DMV, is lower or equal to the traditional strategy, for all years. The total groundwater use can certainly be confined for years with multiple short droughts, but not for a year with one long drought. However, the strategy lowers the risk of water shortages.

The perfect forecast shows how the results of the adaptive forecast-based approach can be improved. It was expected that the perfect forecast performs better in all cases, but the results do not agree with this hypothesis. The forecast model does not always accurately predict high peaks in the water quality parameters. This means that the production rate remains high because the model predicts good water quality, although the actual water quality is low. The perfect forecast will in that case predict low water quality, leading to a lower production rate and using more groundwater. However, the perfect forecast heavily limits the risk of water shortages, which becomes obvious when considering the year 2020.

5.2.4 Uncertainty analysis

The results of the uncertainty analysis are shown in Figure 40. Next to the extreme scenarios, the results of the actual prediction and the perfect forecast were plotted.

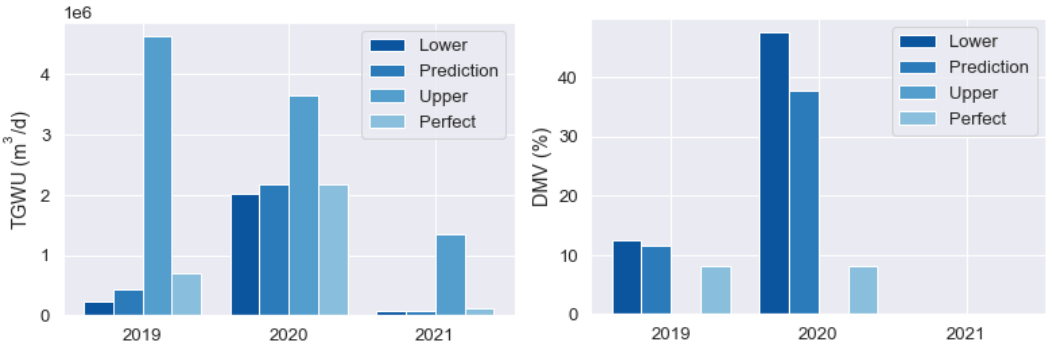


Figure 40: Uncertainty analysis for four different intake scenarios based on the model performance of the water quality forecasting model.

The lower bound forecast leads to a lower TGWU and a higher DMV compared to the other scenarios. The model always predicts low values for the water quality parameters, never exceeding their limit values. This means that the production rate stays high, but the risk of water shortages is much higher, leading to increased values for DMV. The upper bound forecast shows results contrary to the lower bound. The model more often predicts low water quality, which leads to a lower production rate and more groundwater use. A safe strategy with regard to water shortages is used, which leads to higher groundwater use.

Comparing the actual prediction to a perfect forecast model shows contradictory results. The perfect forecast does not show better results for all years. The perfect prediction performs better on DMV in 2019 and better on TGWU in 2020. The risk of water shortages is lower for the perfect prediction in 2019, but the amount of groundwater used was higher than for the actual forecast. The actual forecast predicted high water quality for most of the dry periods, which caused the production rate to remain higher. The perfect forecast, however, switched to a lower production rate and adopts a safer strategy. This leads to a lower DMV.

In 2020, the TGWU is lower for the perfect prediction than for the actual prediction. The actual prediction wrongly forecasts the end of the long dry period in 2020, which means that the reservoir is not emptied before intake can start again. The perfect prediction does predict this perfectly which means that more groundwater can be saved. However, the perfect prediction exceeds the soft constraint of the minimum volume, leading to a higher value of DMV.

5.2.5 Sensitivity analysis

A sensitivity analysis was done to find how the model constraints and initial conditions influence the performance of the adaptive forecast-based strategy.

5.2.5.1 Initial volume

Figure 41 shows how the Total Groundwater Use and the Days Minimum Volume depend on the initial volume in the reservoir.

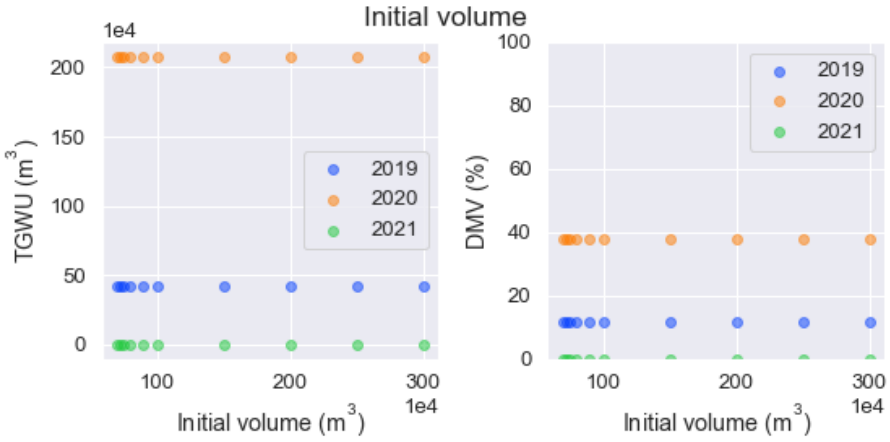


Figure 41: Sensitivity of TGWU and DMV to initial reservoir volume V_{init}.

Both TGWU and DMV are not influenced by the initial volume. The initial intake rate in the first week of each year is maximum. This ensures that the reservoir fills during the first week of the year, such that the initial volume does not have much influence on the model outcome. In practice, a lower initial volume would cause more groundwater use. It is not expected to influence the DMV because this KPI is determined only over the summer months (1 April – 1 October) while the initial volume is defined on January 1st.

5.2.5.2 Initial production rate

Figure 42 shows how TGWU and DMV depend on the initial production rate.

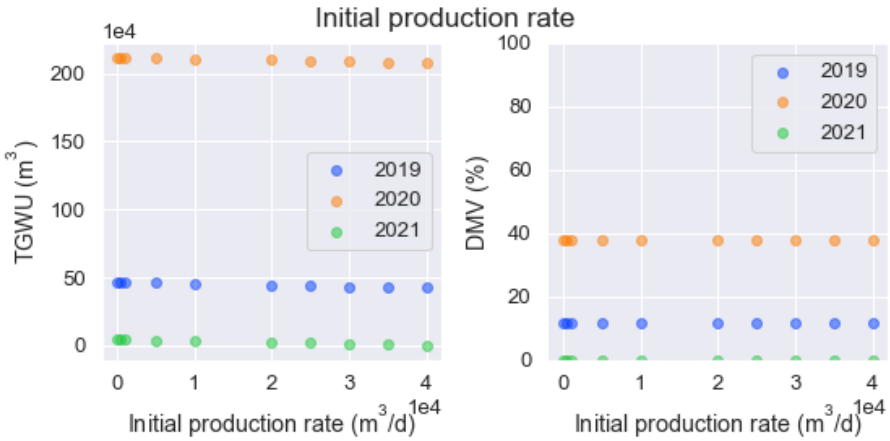


Figure 42: Sensitivity of TGWU and DMV to initial production rate Q_{init}.

The initial production rate has a minor influence on the TGWU. A lower initial production rate causes a minor increase in the TGWU. Less water can be produced from surface water which leads to a slightly higher TGWU. Similar to the initial volume, the initial production rate does not influence DMV.

5.2.5.3 Minimum reservoir volume

Figure 43 shows how the minimum reservoir volume influences the TGWU and DMV.

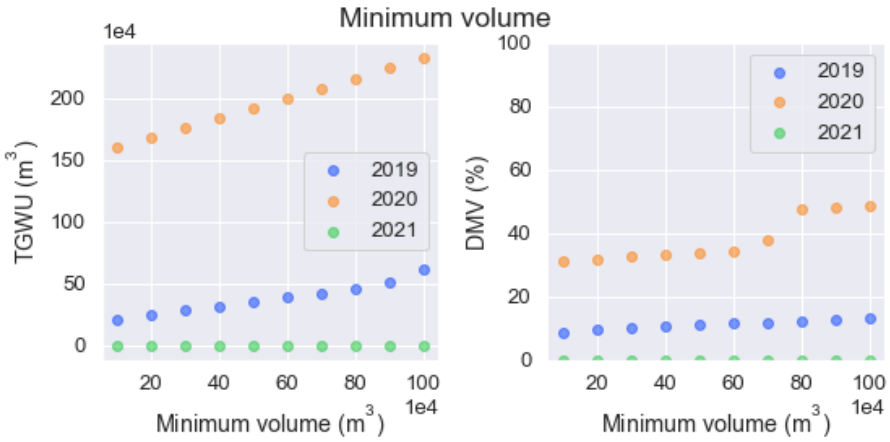


Figure 43: Sensitivity of TGWU and DMV to minimum reservoir volume V_{min}.

Both the TGWU and DMV depend on the minimum reservoir volume. 2019 and 2020 show that a higher minimum volume causes an increased TGWU, while no such dependency can be found for 2021. An increased minimum volume leads to a lower total reservoir volume and therefore a lower buffer capacity. Similarly, DMV increases with a larger minimum volume for 2020 and 2021. The soft constraint for minimum volume depends on the actual hard constraint of minimum volume (which is altered here). A higher minimum volume thus leads to a higher soft constraint and more violation of this constraint, obviously leading to a higher DMV. An abrupt change in DMV can be seen in 2020, indicating an extreme change in performance between a minimum reservoir volume of $7*10^5$ and $8*10^5$ m^3. Although the performance indicator shows very different results, the actual difference in volumetric response between the cases is not that large. In the latter case, the volume just minimally exceeds the soft constraint, resulting in a higher value for DMV. This is a shortcoming of the DMV indicator.

5.2.5.4 Maximum reservoir volume

Figure 44 shows how TGWU and DMV depend on the maximum reservoir volume.

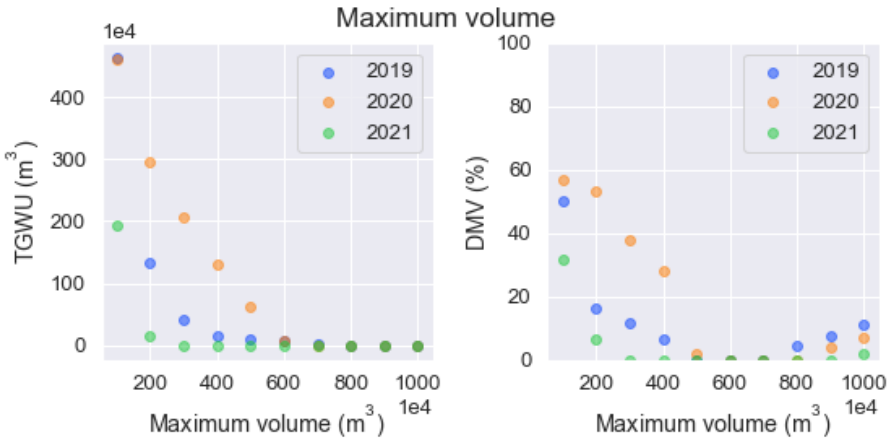


Figure 44: Sensitivity of TGWU and DMV to maximum reservoir volume V_{max}.

Increasing the reservoir volume leads to less groundwater use because of an increased buffer capacity. Doubling the reservoir volume from $3*10^6$ to $6*10^6$ m^3 would lead to (close to) zero groundwater use for all considered years. The DMV indicator highly depends on the maximum reservoir volume. The buffer capacity is increased when the maximum reservoir volume increases, which causes a lower DMV. On the other hand, the soft constraint for DMV depends on the maximum volume and increases when V_{max} increases. This causes a trade-off with an optimal value between $5*10^6$ and $7*10^6$ m^3.

5.2.5.5 Production capacity

Figure 45 shows how the production capacity influences TGWU and DMV.

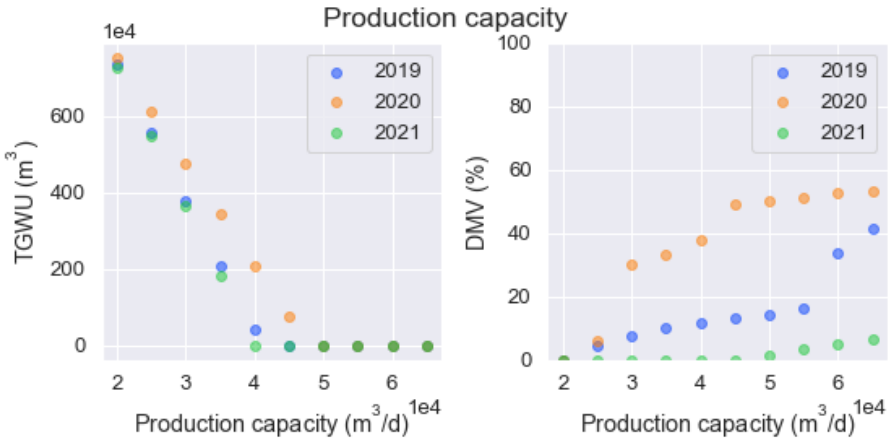


Figure 45: Sensitivity of TGWU and DMV to production capacity Q_{max}.

A higher production capacity leads to less groundwater use because of increased buffer capacity. Increasing the production capacity to $5*10^4$ m^3/day leads to zero Total Groundwater Use for all considered years. However, this comes at the cost of having a higher DMV. Similar as the performance jump in 2020 for minimum reservoir volume, an abrupt change in DMV can be seen when comparing a production capacity of $2.5*10^4$ and $3*10^4$ m^3/day. The same happens in 2019 between a production capacity of $5.5*10^4$ and $6*10^4$ m^3/day. This is the limit where the actual reservoir volume just exceeds the soft volumetric constraint in summer. The same happens for a higher production capacity in 2019.

5.2.5.6 Intake rate

Figure 46 shows how the intake rate of river water to the reservoir influences TGWU and DMV.

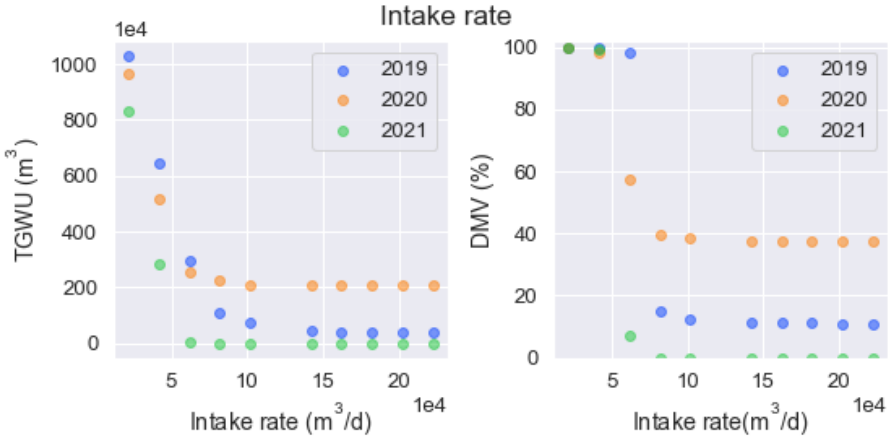


Figure 46: Sensitivity of TGWU and DMV to intake rate Q_{in}.

Lower intake rate causes more groundwater use. This effect is greater for years in which multiple short droughts occur. The reservoir cannot be filled completely during short periods of good water quality, which causes more groundwater use. An intake rate higher than $1*10^5$ m^3/day causes a minor decrease in groundwater use for 2019 but not for the other years. The volumetric constraints limit the buffer capacity of the reservoir here. The same effect holds for the risk of water shortages. The reservoir does not fill completely during short periods, which causes the reservoir volume to be lower during summer, leading to a higher risk of water shortages.

6. Discussion

This section discusses the shortcomings as well as the relevance of the performed research. The two major research steps are discussed: water quality forecasting and reservoir operation.

6.1 Water quality forecasting

The first part of this research concerned the forecasting methods for water quality prediction for the conjunctive use of surface and groundwater resources.

6.1.1 Previous work on water quality forecasting

Table 13 compares the performance of previous research on water quality forecasting using machine learning methods. Water-related problems are often characterized by noisy and poor quality data, which stresses the need to compare different combinations of algorithms and features to find the best performance in predicting the target variable (Solomatine & Ostfeld, 2008). The selection of research includes a large variety of study areas and evaluates different machine learning algorithms. The variety of algorithms shows the case-dependency of best performing algorithms. There is not one algorithm that performs best in all different situations and study areas.

Table 13: Comparison between results of the current study and previous works in literature. WT = Water Temperature, TURB = Turbidity; TDS = Total Dissolved Solids; WOY = Week Of Year; SMRI = Snow Melt Rainfall Index; EC = Electric Conductivity; SAR = Sodium Adsorption Ratio. R2 was not available for Alnahit et al. (2022).

Literature	Study area	ML algorithm	Target variable	Features	Performance
Ahmed et al. (2019)	Rawal Lake, Pakistan	Gradient boosting	Water quality index	WT, TURB, pH, TDS	$R^2 = 0.75$
		Polynomial regression	Water quality index	WT, TURB, pH	$R^2 = 0.54$
Simeonov et al. (2003)	Aliakmon River, Greece	Principal Component Analysis, Multiple Linear Regression	NO3	N.A. (PCA)	$R^2 = 0.57$
			PO4		$R^2 = 0.67$
			Conductivity		$R^2 = 0.49$
Asadollah et al. (2021)	Lam Tsuen River network, Hong Kong	Extra Tree Regression	Water quality index	EC, pH, WT, TURB, BOD, COD, NO3, NO2, PO4	$R^2 = 0.98$
				BOD, TURB, PO4	$R^2 = 0.97$
Koranga et al. (2022)	Nainital Lake, India	Random Forest	Water quality index	pH, TURB, TDS	$R^2 = 0.92$
Sattari et al. (2016)	Lighvan Chay River, Iran	Support Vector Machine	Conductivity	SAR, Na, Mg, Ca, Cl, P, SO4, HCO3	$R^2 = 0.92$
			TDS	SAR, Na, Mg, Ca, Cl, P, SO4, HCO3	$R^2 = 0.91$
Chang et al. (2015)	Dahan River, Taiwan	Artificial Neural Network	NH3-N	Q, P, industrial parameters, WQ parameters	$R^2 = 0.93$
Zaniolo et al. (2019)	Lake Como, Italy	Extreme Learning Machines	Water deficit	WOY, SMRI, T	$R^2 = 0.74$

Alnahit et al. (2022)	South Carolina, North Carolina and Georgia, USA	Random Forest	Total Nitrogen	LU, P	RMSE = 0.82
			Total Phosphorus	LU, watershed soil properties	RMSE = 0.07

No comparable research was found that uses water quality forecasting to develop an adaptive forecast-based approach for the conjunctive use of surface and groundwater resources. Other works focused either on water quality forecasting or on reservoir operation with other limitations to surface water availability. Previous research on water quality forecasting with machine learning models focuses on nowcasts rather than forecasting multiple weeks ahead with a multi-step forecasting approach.

Different target variables are used in the comparable literature. Some research focuses on the prediction of the water quality index, which is an integrated indicator of several physical and chemical factors that describe water quality (Asadollah et al., 2021). Other research focuses on the prediction of one or two water quality parameters, while the current research focuses on the prediction of five water quality parameters. Most research does not include climatic processes as features, but only depends on measured water quality features such as turbidity, pH and total dissolved solids. This often results in better model performance compared to when climatic processes are used, because all data comes from the same data source which needs less data preprocessing.

A similar study with the same case study was performed in 2019 by VITO NV (confidential source). Their analysis focuses on climate effects in modeling nitrate, phosphate, electric conductivity, and bentazon. A similar dataset was used to find the predictive power of historic values, seasonal effects, temperature, and discharge in water quality forecasting. A comparison between the VITO and the FH1 forecasting models in the current study are shown in Table 14. The VITO research focuses on a nowcast rather than a forecast.

Table 14: Results of VITO research conducted in 2019 on the case study.

	VITO research		Current research (FH1)	
	Optimal feature set	R^2	Optimal feature set	R^2
Nitrate	T + Q/P + lagged NO3	0.73	Cald + P + Q + lagged	0.85
Phosphate	T + Q/P + lagged PO4	0.56	Cald + P + T + Q + LU + lagged	0.76
Conductivity	Q + T + water levels	0.80	Cald + P + T + Q + LU	0.82
Bentazon	T	0.66	Cald + T + Q	0.59

The performance of the current research is better for nitrate, phosphate, and conductivity. The performance for bentazon here is lower compared to the VITO research. Improved performance for nitrate, phosphate and conductivity can be caused by the larger dataset that is used for training and testing the model in the current research. The training set contains more periods where limit concentrations were exceeded, which can lead to better performance. The lower performance for bentazon can be explained by the fact that the governmental regulations have changed such that the test set differs too much from the training set, an issue that was less problematic for the VITO research because it took place earlier in time.

Lagged values and rolling statistics of the water quality data were always used as important contributions in the forecasting models of the current research. They did not result in better model performance for all forecasting models in the VITO research. Calendar features always showed to improve model performance in the current research while they were not included in the VITO research. Land use was not covered in the VITO research. Underestimation of peak concentrations was found for both the current and the VITO research. Low data availability causes imbalance of peak concentrations to occur in both training and testing set, which leads to an underestimation of this event in both studies.

The VITO research focuses on finding the optimal model for predicting the water quality parameters. This is part of the current research, although this research focuses more on finding the actual predictive power of each individual climatic process and the interactions with other processes. Besides, the current research proposes a methodology to forecast several weeks ahead in time, while the VITO research only predicts the current values of water quality parameters. A nowcast with the current methodology would result in better results due to the autoregressive behaviour of the models, which leads to even more improvement compared to the VITO research.

6.1.2 Machine learning algorithms

The studied machine learning algorithms in this research were Linear Regression, Decision Trees, Support Vector Machines, K-Nearest Neighbors, and Random Forest. Although the focus of the present research is not to compare machine learning algorithms, it is interesting to see that the best performing algorithm differs per water quality compound. Random Forest Regressor performs best for phosphate, sulfate, and conductivity prediction. Random Forest Regressors perform well on noisy and nonlinear data, which works as an advantage for these water quality compounds. Linear Regression is the best performing algorithm for nitrate. This can be explained by the high autocorrelative behaviour of nitrate, which has more linear characteristics than the external features. (Linear) Support Vector Machines performed best for bentazon prediction, which indicates complex and highly nonlinear relations for this compound.

6.1.3 Climatic processes

The present research focuses specifically on using climatic variables in water quality forecasting. Chang et al. (2015), Zaniolo et al. (2019), and Alnahit et al. (2022) do include external climatic features, watershed soil properties, land use and industrial parameters. Zaniolo et al. (2019) show model improvement when calendar features are included in the prediction of water deficit. In the present research, calendar features improve model performance for almost all models and is always included in the best performing models, which indicates high seasonal behaviour for the water quality parameters.

A strong relation between precipitation and the nutrients and pesticides was expected in the present research, due to the wash-off effect (Rostami et al., 2018). Precipitation improved model performance for all water quality parameters except nitrate. Nitrate has strong autocorrelative effects, which might cause all other features to be less important. Nitrate is the only negatively correlated parameter with precipitation, which might cause the decreased predictive power. This does not agree with the findings in Alnahit et al. (2022), where precipitation was used as an important process in predicting nitrogen concentrations. All optimal models for the water quality parameters did include precipitation, which coincides with the expectation that precipitation is an important climatic process in water quality prediction.

A strong relation between nitrate and phosphate and temperature was expected, while a low importance of temperature in the prediction of conductivity, sulfate and bentazon was expected. Nevertheless, temperature improves all models, except for nitrate prediction, where it severely diminishes model performance. Temperature was an important process to include in phosphate and bentazon prediction. Enhanced nutrient cycling rates might cause more phosphate to accumulate in the water system, although this does not comply with the findings for nitrate (Fukushima et al., 2000). Enhanced crop growth during high temperatures might lead to a stronger predictive value of temperature for bentazon concentrations. Zaniolo et al. (2019) also found improved model performance when temperature is included in predictive models for water quality.

Discharge is an important process in the prediction of all water quality parameters. It quantifies the amount of water in the stream, which is directly related to concentration or conductivity. Precipitation and discharge showed similar performance results for all water quality parameters, indicating that these processes are related in their predictive value. Chang et al. (2015) also found the best model performance when discharge and precipitation are included in water quality prediction.

A strong relation between land use and the nutrients and pesticides was expected, but this relation was not found. Land use only appeared to add noise to the models and has no predictive value in the present research. This effect is related to the specific composition in which land use is included in the model. The crop type features are constant over the year and do not change on a smaller time scale. This time scale does not coincide with the time scale of how water quality changes over the year, which leads to noise. Figure 22 also shows that some crops cover only a very small land area and do not change much over the years, leading to extra noise. Alnahit et al. (2022) did find that land use has a predictive value for water quality prediction, but their research does not focus on predicting time series values. Ahearn et al. (2005) performed a similar analysis to the current research and did find predictive value for some of the land cover types. Their research includes many more sub-watersheds, which leads to improved predictive value.

Low model performance indicates that not all variability in the water quality concentration is captured in the model. Relatively low model performance can be found for sulfate and bentazon prediction. Sulfate concentration is known to be related to the amount of industrial effluent discharged on the river (Silva et al., 2002). Industrial processes were not included in the model, which explains the missing amount of variability that is not explained by the model. Bentazon is a highly diffuse source of pollution, for which it is unknown in the current model when and where it exactly originates from. Additional features such as timing of pesticide application should be included to capture more bentazon variability over the year. The governmental regulations of bentazon application in the area have also changed in the year 2018 (Vlaamse Milieumaatschappij, 2017). This means that the system in the test set (2019-2021) is not comparable to the train set (2011-2018) for bentazon, which causes low performance.

6.1.4 Model performance and feature importance

The R^2 performance metric was chosen for evaluation because it gives a general indication of the performance and most of comparable research is evaluated with this score, which enables easy comparison to literature. Tuning the models to this metric implicates focus on the overall performance. This means that performance in predicting peak concentrations can come at the cost of performance for lower concentrations, especially in imbalanced datasets. Additional performance metrices that capture peak prediction better can be used to improve the methodology. Good model performance (R^2>0.5) was found for all water quality parameters forecasting 1 week ahead, while forecasting further ahead in time leads to lower model performance. This shows the importance of autocorrelation in water quality modelling: less autocorrelation leads to lower model performance.

Evaluating the importance of climatic processes in the forecasting model is not always straightforward and easy to explain (Feng et al., 2019; Freeman et al., 2018; Guo et al., 2018). Some climatic processes were always important and their combination with other features did not cause major performance changes. In many cases, however, model performance differed with specific combinations of climatic processes. The performance difference between specific combinations of climatic processes shows the complexity of the climate system and its interconnections. Different combinations of climatic processes can lead to different information that is included in the model. Different algorithms and forecast horizons also showed different best performing feature sets.

Feng et al. (2019) used machine learning approaches to predict atmospheric pollution and also found that feature importance differs much per target variable and algorithm used. The relations between forecasted parameters and independent variables are sometimes not explainable due to the complexity of the climatic system. Including more input features did not necessarily lead to better model performance in the present research. Freeman et al. (2018) used meteorological inputs to predict air quality and found that less features often results in better performance. Time series of input features are often collinear and nonstationary, which can lead to high system complexity (Freeman et al., 2018). System complexity can be reduced when a selection of input features are removed, which often leads to better model performance. The engineered features in the current research are complex and cover a large memory of past climatic data. Improvements in the methodology

can be made to investigate the effect of system complexity by doing an extensive feature engineering process of including and excluding different sets and configurations of the climatic processes.

Algorithm performance can be improved with hyperparameter tuning. A cross validation grid search can be used to find the optimal combination of hyperparameters per model. The performance of each model is then evaluated for each combination of hyperparameters. Since hyperparameter tuning is an extensive process, it was not implemented in the current research. Especially more complicated algorithms like the Support Vector Machine could benefit from hyperparameter tuning.

6.1.5 Site-specific investigation

The inconsistent findings between different studies and study areas for water quality prediction and their important variables (see Table 13) shows the importance of site-specific investigation, as highlighted by Fukushima et al. (2000). Urban areas might have much more point source pollution while diffuse pollution occurs more in agricultural areas. Important climatic processes can therefore differ much between different areas. Agricultural areas also contain more diffuse pollutants while urban areas are easier to model because they contain more point source pollution. If the proposed methodology is to be repeated in a new study area, an extensive investigation of the study area and the expected important climatic processes is needed. Interesting anthropogenic features to add would be GIS information on urbanization rate, percentage of impervious land, industrialization, or other features that capture anthropogenic processes. Weather forecast, next to past weather conditions, would also be interesting to implement as input features.

6.1.6 Model performance influence on reservoir operation

The model performance in water quality forecasting influences model performance in reservoir operation. If good water quality is forecasted while the actual water quality is not good, the risk of water shortages becomes higher. The reservoir will continue production while the intake rate is switched down. If a period of poor water quality is predicted while the water quality is actually good, groundwater use becomes higher because the surface water is not optimally used. The performance of all forecasting models decreased for longer forecast horizons. Most forecasting models (especially on longer forecast horizons) underestimate the water quality concentrations, which leads to a higher risk of water shortages. Increasing the model performance therefore leads to a lower risk of water shortages and a higher total groundwater use, when compared to the actual forecast (see Figure 40). Better model performance improves overall performance over multiple scenarios.

6.2 Reservoir operation

This section discusses the second major part of the research: developing an optimal production strategy for the reservoir operation.

6.2.1 Previous work on adaptive forecast-based production strategies

Gavahi et al. (2019) developed an adaptive forecast-based approach in which streamflow is predicted instead of water quality as a measure for water availability. They showed the adaptive forecast-based approach has better reliability than a predefined strategy. Implementing a receding operation horizon causes a major improvement in its adaptation to changes and uncertainties. Ahmadi et al. (2015) compared a non-adaptive and an adaptive approach in reservoir operation for hydroelectric power generation. They showed that applying adaptive operational rules improves reliability and vulnerability of the system under a changing climate.

The currently proposed adaptive forecast-based strategy always improves vulnerability (risk of water shortages). In contrast to previous research, reliability (groundwater use) is not always improved in the current study. Reliability does improve in years where multiple short drought periods occur, but not when one long dry period occurs (2020). Other studies use a longer evaluation period for the retrospective analysis: nine years for Gavahi et al. (2019) and three times periods of 14 years for Ahmadi et al. (2015). Due to low data availability, it was not possible to include more years in the current research.

6.2.2 Simplifications with respect to real-world case

The reservoir model that was used in the present research was a simplification of the actual situation. Natural in- and outgoing fluxes such as the precipitation and evaporation were neglected. The effect of adding these fluxes would be minimal since their value relative to the other system parameters is small. Adding these extra fluxes would not result in different. An extra groundwater inlet and river outlet were artificially added to overcome problems with the model constraints, such as overflowing of the reservoir. Besides this, the intake of river water was always maximum or zero, with no flexibility of in-between values. This is not the case in the real-world situation, where the operator is able to flexibly set the intake rate of water to the reservoir. Complicating the reservoir is not expected to lead to different model results, because the limitations similarly influence both strategies. A simple approach is advantageous because it improves communication and understanding of the system parameters and response variables.

The modelled current strategy is a simplification of the actual current strategy. In the real-world situation, the strategy is handled with much more flexibility. In reality, a buffer zone ('Evaluate') is implemented for the limit concentrations. The evaluation zone is used to consider many more variables, such as weather forecasts and the current water quality and quantity in the reservoir, to decide about the production rate. The actual current strategy is therefore more anticipative than the modelled one and is closer to the newly proposed adaptive forecast-based production strategy. This could lead to a lower risk of water shortages and less groundwater use than the modelled situation and therefore better performance for the actual strategy. To quantify how well the modelled strategy resembles the actual strategy, the simulated years can be compared to the actual measured production data.

6.2.3 Performance assessment

Two key performance indicators were used to assess the performance of the production strategies. Reliability was expressed with the Total Groundwater Use (TGWU), while the vulnerability of the strategies was assessed with a measure for the risk of water shortages: the Days of Minimum Volume (DMV).

TGWU is the sum of all groundwater use over the year. The measure highly depends on the yearly drought pattern. In years with more drought days, more groundwater is expected to be needed to complement surface water production. Water demand in the modelling phase was always equal to the production capacity of the reservoir. All water that could not be produced from surface water through the Blankaart reservoir was replaced with groundwater. This choice was made because the production from surface water resources should always be maximum, so it does not depend on the water demand. In reality, groundwater use does depend on water demand, because higher water demand leads to higher groundwater use. Since the same scenarios were used to evaluate both production strategies, the fluctuating water demand is not expected to influence the comparison of groundwater use.

DMV depends on the soft constraint for minimum drought volume in the reservoir. This constraint only holds in summer and rigorously changes from minimum to the maximum reservoir volume between the 31st of March till the 1st of April. The soft constraint should capture the risk of water shortages. If the complete summer period is dry and we do not want to end up with an empty reservoir before the end of summer, the reservoir should have that minimum volume if the production rate is 30% of the production capacity. This is an artificial constraint and merely an approximation of the actual natural system. The method can be improved by assessing the actual probability that intake is not possible from each moment in summer until the end of summer, based on the historical data.

Because of complexity, the key performance indicators could not be expressed in terms of costs. The costs for the risk of water shortages are partly measures that the company can take to avoid a low reservoir volume, and the costs of the consequences of a water shortage. Consequences of a water shortage are low water pressure and low drinking water quality. One measure that the operators can take to avoid reaching the minimum level is flushing the reservoir. Part of the produced surface water is then used as input to the reservoir.

The reservoir can remain operational because the minimum volume is not reached. This approach is highly inefficient, because clean water is used as input to the reservoir and treated twice before supplied to the area. Other measures can be more focused on raising awareness with the end users about their water use in times of drought.

The water quality models were trained on the years 2011-2018 and tested on 2019-2021. The best performing models were selected from the tested data and used in the reservoir operation phase, for 2019-2021. The same time period was thus used for testing the water quality models and evaluating reservoir operation performance. This leads to improved performance in the evaluation of the forecast-based strategy compared to when completely new data would be used. However, it is necessary to assess the performance of the production strategy on these dry years to show what the implications of a forecast-based production strategy would be in a real-world case. Improvements can be made when more historic data on years with limited surface water availability is available.

6.2.4 Data availability

Water quality measurements were available on a weekly frequency. The data was not resampled to a daily frequency because this reduces credibility. The intake forecast therefore only changes on a weekly time step. This reduces flexibility in the response to river water quality. Increasing the data frequency of water quality measurements can therefore improve performance in limiting total groundwater use and the risk of water shortages.

6.3 Recommendations for future work

The methodology for developing water quality forecasting can be applied to other system parameters that limit surface water availability under uncertain climatic conditions. Investigation of the study area and the important climatic and system parameters is necessary to reduce complexity and focus on the necessary processes. The forecasting model can be improved by increasing the data resolution to a daily basis, which not only increases operational flexibility, but also results in better model performance. With more available data, more advanced models such as an Artificial Neural Network can be applied which is expected to lead to better forecasting results.

The sensitivity analysis showed that the initial conditions do not influence the key performance indicators much. The initial conditions are set at the start of each calendar year, while the performance is mostly evaluated during the summer months. The reservoir constraints do influence model performance. Increasing the reservoir volume to $6*10^6$ m^3 could lead to zero groundwater use when a forecast-based production strategy is implemented. The sensitivity analysis also shows that the current production capacity is a limiting factor in the total groundwater use. If the production capacity increases, the reservoir can be emptied more quickly before the end of a dry period. This increases the buffer capacity of the reservoir because more surface water can be stored and produced, which limits the use of groundwater resources.

The approximation of the traditional strategy can be improved in the reservoir operation optimization phase. Investigation and communication with the case study's operators is necessary to develop an accurate model that resembles the real-world situation. A complete set of different drought scenarios is necessary to properly evaluate different production strategies in periods with low surface water availability.

7. Conclusion

Periods with low surface water availability are expected to increase in the future in severity and frequency. The aim of this research was **to explore the role of uncertainty in climatic conditions in the conjunctive use of surface and groundwater resources for drinking water production by developing an adaptive forecast-based production strategy**. A case study of conjunctive use of surface and groundwater resources by a Belgian drinking water company was used to obtain the research goal. The present research focused on developing an adaptive forecast-based production strategy using four climatic processes in the prediction of river water quality. Implementation of the forecast-based approach has shown to improve operational performance for the conjunctive use of surface and groundwater resources. The following two research questions were answered.

1. *How can conjunctive use of surface and groundwater resources be decided on under uncertain climatic conditions?*

Conjunctive use of surface and groundwater resources can be decided on with a forecast of river water quality that limits surface water availability. Data-driven approaches are used to predict water quality parameters using the parameter's autoregressive behaviour, precipitation, temperature, discharge, and land use. Experiments with alternating machine learning algorithms, input feature sets, forecast horizons and water quality parameters resulted in different forecasting performance. Machine learning algorithms have different characteristics that are suitable to specific cases. Feature sets consisting of different configurations of precipitation, temperature, discharge, and land use resulted in different performance because of site-specificity and the interrelatedness of climatic processes. For nitrate prediction, the autocorrelation and seasonal effects are of main importance, while all other features reduce model performance. For the other water quality parameters, all external features except land use improve model performance. Land use shows no improvement to any of the models, except when it is combined with all other features. Adding climatic processes shows most effect for longer forecast horizons. With relatively simple machine learning models and open-source data, good to moderate model performance was found for all models forecasting 1 or 2 weeks ahead in time. Low model performance was found for longer forecast horizons, although climatic processes have more predictive power here.

2. *How can an adaptive forecast-based strategy be developed for the optimal conjunctive use of surface and groundwater resources under uncertain climatic conditions?*

An adaptive water quality forecast-based production strategy was developed to optimize the conjunctive use of surface and groundwater resources under uncertain climatic conditions. An adaptive approach is favored over a predefined one when climatic conditions are uncertain. The goal of the new production strategy was to meet the water demand with the trade-off of minimizing groundwater use and minimizing the risk of water shortages, under the physical constraints of the operating reservoir and surface water availability. The adaptive forecast-based approach maximizes the use of the surface water reservoir by emptying it before a period of good water quality and regulating the production rate such that production is never stopped. The adaptive forecast-based approach was compared to a predefined production strategy. The new strategy always averts the risk of water shortages, and reduces groundwater use for years with multiple short droughts. Improving the water quality forecasting model results in a lower risk of water shortages.

The current research uses relatively simple machine learning models to develop an adaptive forecast-based approach. The methodology is easy to implement when water quality data is available. All other data sources are open-source data that are publicly available. Due to the data-driven approach, it is not necessary to investigate all individual relations in the climate-water system to develop good performing predictive models. An extensive analysis of the relation between climatic processes and water quality parameters shows the predictive power of climatic processes in water quality forecasting. The forecasting models were evaluated through a real-world case study in which improved performance over multiple years of retrospective analysis was found. The adaptive forecast-based approach performs better than the predefined approach because it better averts the risk of water shortages and performs well in years with short periods of low water quality.

References

Ahearn, D. S., Sheibley, R. W., Dahlgren, R. A., Anderson, M., Johnson, J., & Tate, K. W. (2005). Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California. *Journal of Hydrology*, *313*(3–4), 234–247. https://doi.org/10.1016/j.jhydrol.2005.02.038

Ahmadi, M., Haddad, O. B., & Loáiciga, H. A. (2015). Adaptive Reservoir Operation Rules Under Climatic Change. *Water Resources Management*, *29*(4), 1247–1266. https://doi.org/10.1007/s11269-014-0871-0

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, *11*(11), 2210. https://doi.org/10.3390/w11112210

Alemu, E. T., Palmer, R. N., Polebitski, A., & Meaker, B. (2011). Decision Support System for Optimizing Reservoir Operations Using Ensemble Streamflow Predictions. *Journal of Water Resources Planning and Management*, *137*(1), 72–82. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000088

Allawi, M. F., Jaafar, O., Mohamad Hamzah, F., Koting, S. B., Mohd, N. S. B., & El-Shafie, A. (2019). Forecasting hydrological parameters for reservoir system utilizing artificial intelligent models and exploring their influence on operation performance. *Knowledge-Based Systems*, *163*, 907–926. https://doi.org/10.1016/j.knosys.2018.10.013

Alnahit, A. O., Mishra, A. K., & Khan, A. A. (2022). Stream water quality prediction using boosted regression tree and random forest models. *Stochastic Environmental Research and Risk Assessment*. https://doi.org/10.1007/s00477-021-02152-4

Asadollah, S. B. H. S., Sharafati, A., Motta, D., & Yaseen, Z. M. (2021). River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*, *9*(1), 104599. https://doi.org/10.1016/j.jece.2020.104599

Baken, S., Baetens, E., & Smolders, E. (2016). *Een beschrijving en verklaring van de zomerse fosfaatpieken in Vlaams oppervlaktewater—A description and explanation of summer phosphate peaks in Flemish surface water (in Dutch)*. https://doi.org/10.13140/RG.2.1.3524.3280

Baker, A. (2003). Land use and water quality. *Hydrological Processes*, *17*(12), 2499–2501. https://doi.org/10.1002/hyp.5140

Benítez-Gilabert, M., Alvarez-Cobelas, M., & Angeler, D. G. (2010). Effects of climatic change on stream water quality in Spain. *Climatic Change*, *103*(3–4), 339–352. https://doi.org/10.1007/s10584-009-9778-9

Bloetscher, F., Hammer, N. H., & Berry, L. (2014). How climate change will affect water utilities. *Journal - American Water Works Association*, *106*(8), 176–192. https://doi.org/10.5942/jawwa.2014.106.0112

Bolouri-Yazdeli, Y., Bozorg Haddad, O., Fallah-Mehdipour, E., & Mariño, M. A. (2014). Evaluation of Real-Time Operation Rules in Reservoir Systems Operation. *Water Resources Management*, *28*(3), 715–729. https://doi.org/10.1007/s11269-013-0510-1

Chang, F.-J., Tsai, Y.-H., Chen, P.-A., Coynel, A., & Vachaud, G. (2015). Modeling water quality in an urban river using hydrological factors – Data driven approaches. *Journal of Environmental Management*, *151*, 87–96. https://doi.org/10.1016/j.jenvman.2014.12.014

Coordinatiecommissie Integraal Waterbeleid. (2016). *Stroomgebiedbeheerplan voor de Schelde 2016-2021. Bekkenspecifiek deel IJzerbekken*. https://www.vlaanderen.be/publicaties/stroomgebiedbeheerplan-voor-de-schelde-2016-2021-bekkenspecifiek-deel-ijzerbekken

Dagli, C. H., & Miles, J. F. (1980). Determining operating policies for a water resource system. *Journal of Hydrology*, *47*(3–4), 297–306. https://doi.org/10.1016/0022-1694(80)90098-0

Das, B., Singh, A., Panda, S. N., & Yasuda, H. (2015). Optimal land and water resources allocation policies for sustainable irrigated agriculture. *Land Use Policy*, *42*, 527–537. https://doi.org/10.1016/j.landusepol.2014.09.012

De Watergroep. (2021a). *De Watergroep België*. De Watergroep. https://www.dewatergroep.be/nl-be/drinkwater

De Watergroep. (2021b). *Internal information De Watergroep*.

Delpla, I. (2009). Impacts of climate change on surface water quality in relation to drinking water production. *Environment International*, 9.

Feng, R., Zheng, H., Gao, H., Zhang, A., Huang, C., Zhang, J., Luo, K., & Fan, J. (2019). Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. *Journal of Cleaner Production*, *231*, 1005–1015. https://doi.org/10.1016/j.jclepro.2019.05.319

Freeman, B. S., Taylor, G., Gharabaghi, B., & Thé, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, *68*(8), 866–886. https://doi.org/10.1080/10962247.2018.1459956

Fukushima, T., Ozaki, N., Kaminishi, H., Harasawa, H., & Matsushige, K. (2000). Forecasting the changes in lake water quality in response to climate changes, using past relationships between meteorological conditions and water quality. *Hydrological Processes*, *14*(3), 593–604. https://doi.org/10.1002/(SICI)1099-1085(20000228)14:3<593::AID-HYP956>3.0.CO;2-O

Gavahi, K., Mousavi, S. J., & Ponnambalam, K. (2019). Adaptive forecast-based real-time optimal reservoir operations: Application to Lake Urmia. *Journal of Hydroinformatics*, *21*(5), 908–924. https://doi.org/10.2166/hydro.2019.005

Guo, Z., Zhou, K., Zhang, X., & Yang, S. (2018). A deep learning model for short-term power load and probability density forecasting. *Energy*, *160*, 1186–1200. https://doi.org/10.1016/j.energy.2018.07.090

Hashimoto, T., Stedinger, J. R., & Loucks, D. P. (1982). Reliability, resiliency, and vulnerability criteria for water resource system performance evaluation. *Water Resources Research*, *18*(1), 14–20. https://doi.org/10.1029/WR018i001p00014

Hayashi, M. (2004). Temperature-Electrical Conductivity Relation of Water for Environmental Monitoring and Geophysical Data Inversion. *Environmental Monitoring and Assessment*, *96*(1–3), 119–128. https://doi.org/10.1023/B:EMAS.0000031719.83065.68

Heylen, I. J. (n.d.). *De hydrologie van het IJzerbekken*. 6.

Kalkhoff, S. J., Hubbard, L. E., Tomer, M. D., & James, D. E. (2016). Effect of variable annual precipitation and nutrient input on nitrogen and phosphorus transport from two Midwestern agricultural watersheds. *Science of The Total Environment*, *559*, 53–62. https://doi.org/10.1016/j.scitotenv.2016.03.127

Kendall, J. (1975). Hard and soft constraints in linear programming. *Omega*, *3*(6), 709–715. https://doi.org/10.1016/0305-0483(75)90073-0

Khare, D., Jat, M. K., & Ediwahyunan. (2006). Assessment of counjunctive use planning options: A case study of Sapon irrigation command area of Indonesia. *Journal of Hydrology*, *328*(3–4), 764–777. https://doi.org/10.1016/j.jhydrol.2006.01.018

KIWA NV. (1990). *Bestrijdingsmiddelen en drinkwatervoorziening in Nederland*. https://www.ircwash.org/sites/default/files/243-90BE-7820.pdf

Koranga, M., Pant, P., Kumar, T., Pant, D., Bhatt, A. K., & Pant, R. P. (2022). Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand. *Materials Today: Proceedings*, *57*, 1706–1712. https://doi.org/10.1016/j.matpr.2021.12.334

Kundzewicz, Z. W., & Döll, P. (2009). Will groundwater ease freshwater stress under climate change? *Hydrological Sciences Journal*, *54*(4), 665–675. https://doi.org/10.1623/hysj.54.4.665

Lazzeri, F. (2020). *Machine Learning for Time Series Forecasting with Python®* (1st ed.). Wiley. https://doi.org/10.1002/9781119682394

Lepot, M., Aubin, J.-B., & Clemens, F. (2017). Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water*, *9*(10), 796. https://doi.org/10.3390/w9100796

Li, L., Xu, H., Chen, X., & Simonovic, S. P. (2010). Streamflow Forecast and Reservoir Operation Performance Assessment Under Climate Change. *Water Resources Management*, *24*(1), 83–104. https://doi.org/10.1007/s11269-009-9438-x

Lin, N. M., & Rutten, M. (2016). Optimal Operation of a Network of Multi-purpose Reservoir: A Review. *Procedia Engineering*, *154*, 1376–1384. https://doi.org/10.1016/j.proeng.2016.07.504

Maiolo, M., Mendicino, G., Pantusa, D., & Senatore, A. (2017). Optimization of Drinking Water Distribution Systems in Relation to the Effects of Climate Change. *Water*, *9*(10), 803. https://doi.org/10.3390/w9100803

Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood Prediction Using Machine Learning Models: Literature Review. *Water*, *10*(11), 1536. https://doi.org/10.3390/w10111536

Nielsen, A. (2019). *Practical time series analysis: Prediction with statistics and machine learning*. O'Reilly Media, Inc.

O'Connell, E. (2017). Towards Adaptation of Water Resource Systems to Climatic and Socio-Economic Change. *Water Resources Management*, *31*(10), 2965–2984. https://doi.org/10.1007/s11269-017-1734-2

Ozaki, N., Fukushima, T., Harasawa, H., Kojiri, T., Kawashima, K., & Ono, M. (2003). Statistical analyses on the effects of air temperature fluctuations on river water qualities. *Hydrological Processes*, *17*(14), 2837–2853. https://doi.org/10.1002/hyp.1437

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (n.d.). Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*, 6.

Pereira, J. L. J., Oliver, G. A., Francisco, M. B., Cunha, S. S., & Gomes, G. F. (2021). A Review of Multi-objective Optimization: Methods and Algorithms in Mechanical Engineering Problems. *Archives of Computational Methods in Engineering*. https://doi.org/10.1007/s11831-021-09663-x

Renwick, M. E. (2018). *Institutions and conjunctive water management among three western states* (K. W. Easter & M. E. Renwick, Eds.; 1st ed.). Routledge. https://doi.org/10.4324/9781351159289

Rosenboom, A., Karan, S., Badawi, N., Gudmundsson, L., Hansen, C. H., Nielsen, C. B., Plauborg, F., & Olsen, P. (2021). *The Danish Pesticide Leaching Assessment Programme 2019* (p. 210). http://pesticidvarsling.dk/wp-content/uploads/2021/01/The-Danish-Pesticide-Leaching-Assessment-Programme-2019-.pdf

Rostami, S., He, J., & Hassan, Q. (2018). Riverine Water Quality Response to Precipitation and Its Change. *Environments*, *5*(1), 8. https://doi.org/10.3390/environments5010008

Sarwar, A., & Eggers, H. (2006). Development of a conjunctive use model to evaluate alternative management options for surface and groundwater resources. *Hydrogeology Journal*, *14*(8), 1676–1687. https://doi.org/10.1007/s10040-006-0066-8

Sattari, M. T., Joudi, A. R., & Kusiak, A. (2016). Estimation of Water Quality Parameters With Data-Driven Model. *Journal - American Water Works Association*, *108*, E232–E239. https://doi.org/10.5942/jawwa.2016.108.0012

Scanlon, B. R., Reedy, R. C., Faunt, C. C., Pool, D., & Uhlman, K. (2016). Enhancing drought resilience with conjunctive use and managed aquifer recharge in California and Arizona. *Environmental Research Letters*, *11*(3), 035013. https://doi.org/10.1088/1748-9326/11/3/035013

Silva, A. J., Varesche, M. B., Foresti, E., & Zaiat, M. (2002). Sulphate removal from industrial wastewater using a packed-bed anaerobic reactor. *Process Biochemistry*, *37*(9), 927–935. https://doi.org/10.1016/S0032-9592(01)00297-7

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., & Kouimtzis, Th. (2003). Assessment of the surface water quality in Northern Greece. *Water Research*, *37*(17), 4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

Singh, A. (2014). Simulation–optimization modeling for conjunctive water use management. *Agricultural Water Management*, *141*, 23–29. https://doi.org/10.1016/j.agwat.2014.04.003

Singh, A., Thakur, N., & Sharma, A. (2016). *A review of supervised machine learning algorithms*. http://ieeexplore.ieee.org/servlet/opac?punumber=7589474

Snyman, J. A., & Wilke, D. N. (2018). *Practical Mathematical Optimization* (Vol. 133). Springer International Publishing. https://doi.org/10.1007/978-3-319-77586-9

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3–22. https://doi.org/10.2166/hydro.2008.015

Srubbe, I. J. (2005). *Flood management in Flanders with special focus on navigable waterways*. 17.

Tiyasha, Tung, T. M., & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, *585*, 124670. https://doi.org/10.1016/j.jhydrol.2020.124670

Tong, S. T. Y., & Chen, W. (2002). Modeling the relationship between land use and surface water quality. *Journal of Environmental Management*, *66*(4), 377–393. https://doi.org/10.1006/jema.2002.0593

Vlaamse Landmaatschappij. (2021). *Mestrapport 2021*.

Vlaamse Milieumaatschappij. (n.d.). *Geopunt Vlaanderen*. Geopunt Vlaanderen. Retrieved December 15, 2021, from https://www.geopunt.be

Vlaamse Milieumaatschappij. (2017). *Pesticiden in de waterketen 2015-2016* (p. 74).

Vlaamse Milieumaatschappij. (2021a). *Nutriënten in oppervlaktewater en grondwater in landbouwgebied*.

Vlaamse Milieumaatschappij. (2021b, December). *Fosfaat in oppervlaktewater in landbouwgebied (2002-2021)*. Vlaamse MilieuMaatschappij. https://www.vmm.be/sectoren/landbouw/nitraat-in-oppervlaktewater-in-landbouwgebied

Vlaamse Milieumaatschappij. (2022, April 9). *Emissies van nutriënten naar oppervlaktewater (2010-2020)*. Vlaamse MilieuMaatschappij. https://www.vmm.be/water/kwaliteit-waterlopen/emissies-van-nutrienten-naar-oppervlaktewater

Wu, D., Jennings, C., Terpenny, J., Gao, R. X., & Kumara, S. (2017). A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *Journal of Manufacturing Science and Engineering*, *139*(7), 071018. https://doi.org/10.1115/1.4036350

Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews*, *73*, 1104–1122. https://doi.org/10.1016/j.rser.2017.02.023

Zaniolo, M., Giuliani, M., & Castelletti, A. (2019). Data-driven modeling and control of droughts. *IFAC-PapersOnLine*, *52*(23), 54–60. https://doi.org/10.1016/j.ifacol.2019.11.009

Zhang, X. (2015). Conjunctive surface water and groundwater management under climate change. *Frontiers in Environmental Science*, *3*. https://doi.org/10.3389/fenvs.2015.00059

Zhu, D., Cheng, X., Sample, D. J., & Yazdi, M. N. (2019). The effect of temperature on sulfate release from Pearl River sediments in South China. *Science of The Total Environment*, *688*, 1112–1123. https://doi.org/10.1016/j.scitotenv.2019.06.185

Appendix 1 – Methodology: additional material

1A: Intake strategy

The intake strategy of De Watergroep is based on the concentration of several compounds. The parameters considered in this research are nitrate (NO3), (orto-)phosphate (oPO4), sulfate (SO4), conductivity and bentazon. These compounds represent nutrients, pesticides and the salination of water which are all important parameters for water quality. From all measured compounds, these five also have the longest historic record and are thus useful for machine learning applications. Table A1 shows the limit concentrations for the intake strategy for each water quality parameter.

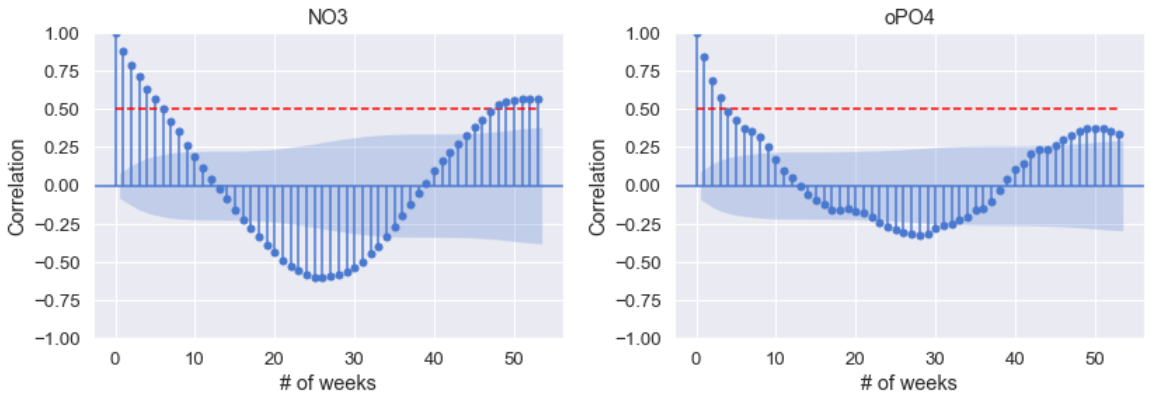
Table A1: Intake limit concentrations (De Watergroep, 2021).

	Okay	Evaluate	Alarm	
	Intake possible	*Operator should decide*	*Intake not possible*	Unit
NO3	< 40	40 – 60	> 60	mg/l
oPO4	< 1	1 – 2	> 2	mg/l
SO4	< 150	150 – 180	> 180	mg/l
Conductivity	< 1000	1000 – 2000	> 2000	µS/cm
Bentazon	< 0.3	0.3 – 0.8	> 0.8	µg/l

The water quality compounds are measured at several measurement locations along the river Yser. The most influential measurement point is located directly next to the Blankaart reservoir and therefore chosen to consider in the current research. The limit concentrations are used to support the operator to assess whether intake to the reservoir is possible or not. If the measured concentration is below the "Okay" limit, intake is possible. If the concentration exceeds the "Alarm" concentration, intake is not possible. Although this scheme assists the operators, it is not directive. The operator can take a different decision if other processes and conditions play a role at that moment.

1B: Autocorrelation of water quality parameters

Figures A1 – A5 show the autocorrelation plots of all water quality parameters. The correlations are plotted on the vertical axis and the lags on the horizontal axis. 95% confidence intervals are plotted in light blue. Correlation of R>0.5 is considered to be contributive to the predictive power of each model.



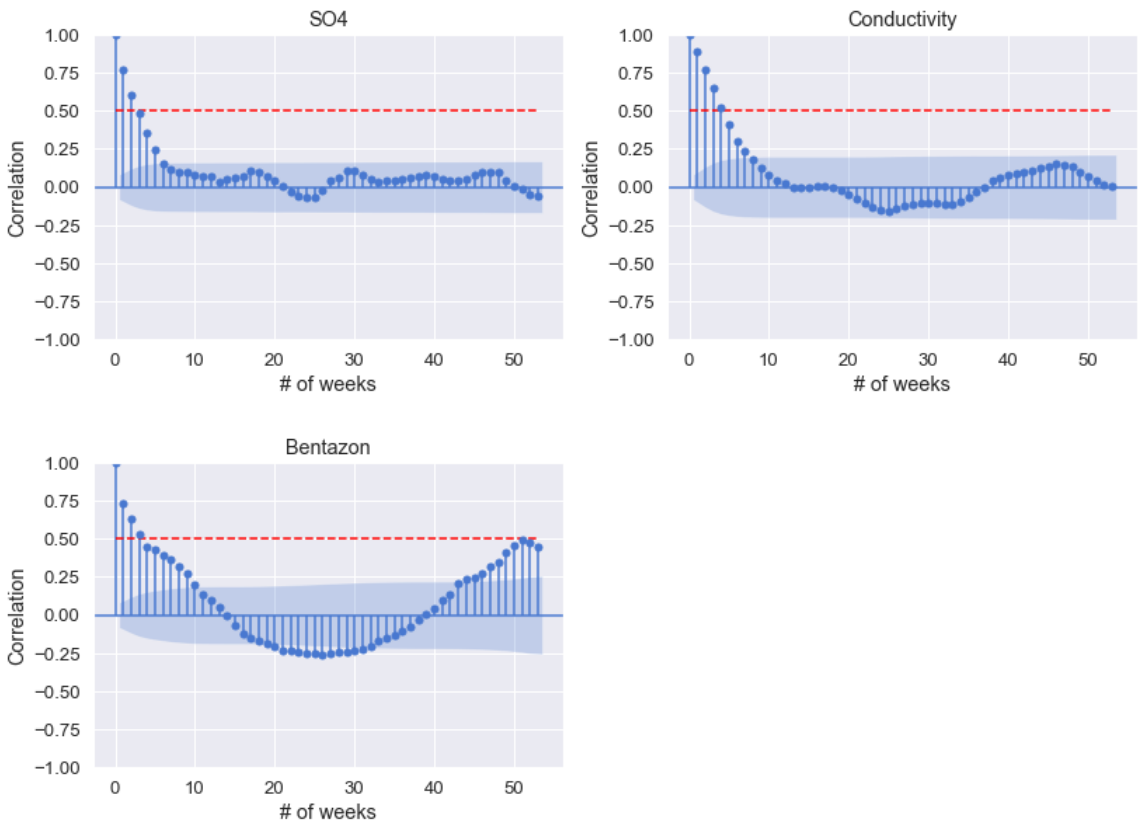


Figure A1 – A5: Autocorrelation plots with 95% confidence intervals for all water quality parameters. Red dotted line: correlation coefficient = 0.5.

1C: Rainfall spatial interpolation methods

Three spatial interpolation methods were used to approximate the mean areal precipitation in the Yser catchment with data from the seven rainfall stations. All methods are used to calculate the weights of each rainfall station to the mean areal precipitation (MAP) and are based on (Garcia et al., 2008; Shaw, 2011):

$$\mathrm{M.A.P.} = w_1{*}P_{\mathrm{lofintele}} + w_2{*}P_{\mathrm{zarren}} + w_3{*}P_{\mathrm{poperinge}} + w_4{*}P_{\mathrm{vlamertinge}} + w_5{*}P_{\mathrm{ieper}} + w_6{*}P_{\mathrm{geluwe}} + w_7{*}P_{\mathrm{roeselare}}$$

Arithmetic mean

The simplest method applied is the arithmetic mean. All rainfall stations are assigned equal means, such that they contribute equally to the mean areal precipitation. The rainfall stations Roeselare and Geluwe are both assigned a weight of 0 in this case. They are situated far outside the catchment and expected to have a much lower contribution to the MAP than the other stations.

$$w_x = 1/\#\mathrm{stations}$$

More sophisticated methods to determine the MAP are the use of Thiessen polygons and the inverse distance weighting method.

Thiessen polygons

The Yser catchment is divided into seven polygons, each describing the area of influence for each rainfall station, see Figure A6.

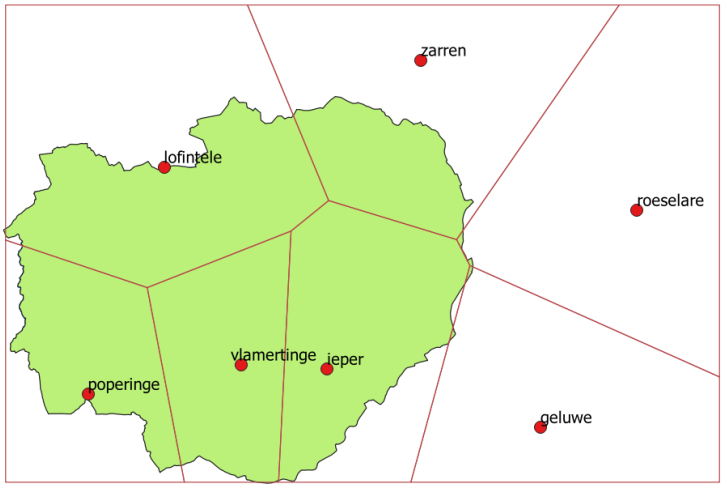


Figure A6: Thiessen polygons describing the area of influence of each rainfall station. The green area is the Yser catchment upstream of the Blankaart intake point.

The weight assigned to each rainfall station is determined by dividing the polygon area by the total catchment area. The open source QGIS software was used to calculate the catchment area and the area of each polygon.

$$w_x = \frac{A_{polygon}}{A_{catchment}}$$

Inverse distance weighting

The QGIS software was used to divide the catchment area into a number of grid cells and perform inverse distance weighting. Each rainfall station contributes to the interpolated value of a grid cell based on the inversed distance between the rainfall station and the intended grid cell. The mean areal precipitation is found by taking the average of all grid cells in the area. To find the weights of each rainfall station to the mean areal precipitation, a dataset of 24 values for each rainfall station and the mean areal precipitation based on inverse distance weighting was obtained. Multiple linear regression was then used to find the weight of each rainfall station to the mean areal precipitation.

Table A2: Weights for all rainfall stations found with each spatial interpolation method.

Rainfall station	Arithmetic mean	Thiessen	Inverse Distance Weighting
Lofintele	0.2	0.2629	0.2235
Ieper	0.2	0.2560	0.1982
Vlamertinge	0.2	0.2019	0.1929
Poperinge	0.2	0.1607	0.1920
Zarren	0.2	0.1173	0.0951
Roeselare	0	0.0007	0.0349
Geluwe	0	0.0006	0.0632

The methods with the strongest correlation between the cumulative antecedent precipitation and the current concentration were used for each water quality parameter (see Table 7).

Table A3: Strongest correlated methods for spatial interpolation of precipitation. Methods are explained in Appendix 1B.

Parameter	Method
NO3	Lo-Fintele rainfall station
oPO4	Arithmetic mean
SO4	Inverse Distance Weighting
Conductivity	Inverse Distance Weighting
Bentazon	Lo-Fintele rainfall station

1D: Feature engineering

Precipitation

Table A4: Pearson correlation coefficient and optimal lag for cumulative antecedent precipitation per water quality parameter.

Parameter	R	Optimal lag (days)
NO3	0.65	77
oPO4	-0.4	100
SO4	-0.3	46
Conductivity	-0.4	77
Bentazon	-0.3	100

Temperature

Table AA5: Pearson correlation coefficient and optimal lag for average antecedent daily maximum temperature per water quality parameter.

Parameter	R	Optimal lag (days)
NO3	-0.78	33
oPO4	0.60	86
SO4	0.12	14
Conductivity	0.48	36
Bentazon	0.46	28

1E: Days of Minimum Volume

DMV is the percentage of summer (1 April – 1 October) days that the reservoir volume exceeds the minimum drought volume ($V_{min,drought}$). The minimum drought volume is a soft constraint that describes the risk of water shortages. It defines what the reservoir volume should be on each summer day to overcome a dry summer that extends till 1 October, with a production rate of $0.3*Q_{prod,max}$. It depends on the day of the year, the production capacity and the hard constraint of the minimum volume.

The period from 1 April until 1 October has 183 days. If this period starts with a full reservoir ($V = 3*10^6$ m^3) and the reservoir should be empty ($V = 7*10^5$ m^3) at the end of the period, the production rate should be around 12,568 m^3/day, which is approximately 30% of the production capacity. The minimum production rate in summer is therefore $0.3*Q_{prod,max}$. The minimum drought volume is approximated as follows:

$$V_{min,drought} = 0.3*Q_{prod,max}*(t_{1\ October}-t_{date}) + V_{min}$$

The minimum drought volume when $Q_{prod,max}$ = $4*10^4$ m^3/day and V_{min} = $7*10^5$ m^3 is shown in Figure A7.

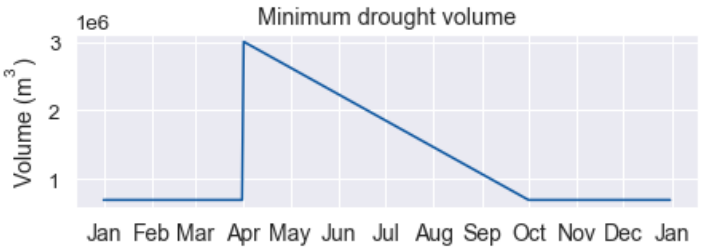


Figure A7: Minimum drought volume for $Q_{prod,max}$ *= 4*10^4 m^3/day and* V_{min} *= 7*10^5 m^3.*

DMV is defined as the percentage of summer days that the actual reservoir volume is below the minimum drought volume.

Appendix 2 – Model performance of water quality forecasting

2A: Experiments

This appendix shows the model performance of all combinations of water quality parameters, forecast horizons, feature sets and algorithms. Negative model performance was indicated with a blue color. The red/green color scales show how well each model performs relative to the other models in the forecast horizon. Green is good model performance, red resembles low model performance. Extremely negative numbers (<< -1) were excluded from calculating the average performance per feature set or algorithm since they were considered outliers.

Table A6: Performance (R^2) for nitrate – forecast horizon 1.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.829	0.387	0.772	0.698	0.745	0.523	0.659
base_P	0.848	0.423	0.794	0.738	0.770	0.583	0.693
base_T	0.807	0.611	0.757	0.650	0.738	0.529	0.682
base_Q	0.833	0.461	0.727	0.733	0.768	0.578	0.684
base_LU	0.829	0.580	0.772	0.698	0.747	0.523	0.691
base_P_T	0.815	0.776	0.763	0.678	0.780	0.587	0.733
base_P_Q	0.838	0.675	0.724	0.699	0.764	0.567	0.711
base_T_Q	0.816	0.424	0.724	0.663	0.770	0.558	0.659
base_P_T_Q	0.817	0.614	0.733	0.664	0.768	0.582	0.696
base_P_T_Q_LU	0.817	0.657	0.732	0.664	0.774	0.582	0.704
cald	0.845	0.450	0.797	0.695	0.750	0.542	0.680
cald_P	0.854	0.687	0.795	0.717	0.782	0.613	0.741
cald_T	0.830	0.392	0.785	0.664	0.737	0.545	0.659
cald_Q	0.839	0.450	0.757	0.669	0.782	0.555	0.675
cald_LU	-7.21E+25	0.658	0.583	0.517	0.751	0.374	0.577
cald_P_T	0.834	0.467	0.785	0.695	0.754	0.593	0.688
cald_P_Q	0.850	0.394	0.769	0.667	0.761	0.601	0.674
cald_T_Q	0.826	0.623	0.738	0.640	0.761	0.542	0.688
cald_P_T_Q	0.834	0.666	0.754	0.641	0.769	0.584	0.708
cald_P_T_Q_LU	-3.83E+28	0.506	0.749	0.518	0.769	0.443	0.597
average	0.831	0.545	0.750	0.665	0.762	0.550	

Table A7: Performance (R^2) for nitrate – forecast horizon 2.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.692	0.513	0.636	0.607	0.658	0.435	0.590
base_P	0.706	0.597	0.656	0.596	0.663	0.471	0.615
base_T	0.612	0.481	0.630	0.552	0.628	0.426	0.555
base_Q	0.701	0.396	0.583	0.614	0.659	0.454	0.568
base_LU	0.692	0.492	0.637	0.607	0.663	0.435	0.587
base_P_T	0.615	0.442	0.615	0.554	0.658	0.467	0.559
base_P_Q	0.697	0.413	0.589	0.553	0.654	0.440	0.558
base_T_Q	0.621	0.409	0.570	0.545	0.611	0.433	0.532
base_P_T_Q	0.619	0.344	0.568	0.526	0.631	0.457	0.524
base_P_T_Q_LU	0.619	0.435	0.568	0.526	0.639	0.457	0.541
cald	0.742	0.506	0.712	0.647	0.675	0.479	0.627
cald_P	0.752	0.408	0.708	0.657	0.672	0.529	0.621
cald_T	0.690	0.487	0.693	0.628	0.633	0.460	0.598
cald_Q	0.744	0.403	0.664	0.636	0.681	0.476	0.601
cald_LU	-3.67E+26	0.411	0.686	0.481	0.657	0.334	0.514
cald_P_T	0.697	0.506	0.675	0.647	0.635	0.500	0.610
cald_P_Q	0.758	0.495	0.670	0.614	0.673	0.508	0.620
cald_T_Q	0.695	0.374	0.647	0.592	0.640	0.458	0.568
cald_P_T_Q	0.706	0.402	0.655	0.605	0.633	0.493	0.583
cald_P_T_Q_LU	-2.50E+28	0.397	0.505	0.509	0.648	0.385	0.489
average	0.685	0.446	0.633	0.585	0.651	0.455	

Table A8: Performance (R^2) for nitrate – forecast horizon 3.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.616	0.081	0.539	0.487	0.555	0.350	0.438
base_P	0.609	0.289	0.549	0.495	0.563	0.358	0.477
base_T	0.452	0.263	0.516	0.475	0.510	0.342	0.426
base_Q	0.619	0.353	0.476	0.503	0.560	0.380	0.482
base_LU	0.616	0.061	0.538	0.487	0.558	0.350	0.435
base_P_T	0.431	0.368	0.518	0.474	0.539	0.364	0.449
base_P_Q	0.598	0.214	0.488	0.443	0.542	0.336	0.437
base_T_Q	0.450	0.149	0.459	0.483	0.509	0.340	0.398
base_P_T_Q	0.430	0.166	0.481	0.441	0.507	0.350	0.396
base_P_T_Q_LU	0.430	0.238	0.481	0.441	0.500	0.350	0.407
cald	0.684	0.529	0.641	0.642	0.625	0.422	0.590
cald_P	0.677	0.393	0.647	0.616	0.605	0.470	0.568
cald_T	0.593	0.300	0.635	0.592	0.538	0.405	0.510
cald_Q	0.685	0.383	0.607	0.595	0.633	0.421	0.554
cald_LU	-6.44E+27	0.495	0.634	0.469	0.611	0.279	0.498
cald_P_T	0.593	0.350	0.620	0.598	0.534	0.434	0.521
cald_P_Q	0.687	0.122	0.624	0.595	0.612	0.448	0.515
cald_T_Q	0.598	0.206	0.587	0.540	0.530	0.393	0.476
cald_P_T_Q	0.609	0.217	0.592	0.590	0.539	0.419	0.495
cald_P_T_Q_LU	-5.80E+29	0.166	0.607	0.490	0.531	0.348	0.428
average	0.575	0.267	0.562	0.523	0.555	0.378	

Table A9: Performance (R^2) for phosphate – forecast horizon 1.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.700	0.549	0.693	0.573	0.712	0.595	0.637
base_P	0.721	0.539	0.701	0.572	0.724	0.611	0.645
base_T	0.726	0.572	0.707	0.502	0.730	0.644	0.647
base_Q	0.709	0.635	0.695	0.612	0.744	0.633	0.672
base_LU	0.700	0.595	0.693	0.573	0.702	0.595	0.643
base_P_T	0.737	0.570	0.714	0.521	0.751	0.659	0.658
base_P_Q	0.713	0.589	0.700	0.558	0.748	0.624	0.655
base_T_Q	0.726	0.672	0.703	0.512	0.758	0.668	0.673
base_P_T_Q	0.734	0.640	0.715	0.534	0.747	0.666	0.673
base_P_T_Q_LU	0.734	0.714	0.717	0.534	0.760	0.666	0.687
cald	0.696	0.634	0.693	0.524	0.719	0.555	0.637
cald_P	0.715	0.558	0.695	0.561	0.739	0.616	0.647
cald_T	0.711	0.526	0.701	0.488	0.739	0.613	0.630
cald_Q	0.695	0.727	0.688	0.550	0.753	0.576	0.665
cald_LU	-1.32E+26	0.596	0.077	0.230	0.709	0.373	0.397
cald_P_T	0.723	0.505	0.702	0.539	0.749	0.648	0.644
cald_P_Q	0.709	0.564	0.697	0.559	0.731	0.621	0.647
cald_T_Q	0.708	0.663	0.694	0.503	0.760	0.621	0.658
cald_P_T_Q	0.720	0.656	0.701	0.553	0.751	0.654	0.672
cald_P_T_Q_LU	-4.25E+29	0.615	-1.166	0.248	0.764	0.427	0.178
average	0.715	0.606	0.576	0.512	0.740	0.603	

Table A10: Performance (R^2) for phosphate – forecast horizon 2.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.420	0.152	0.432	0.403	0.433	0.404	0.374
base_P	0.478	0.221	0.458	0.454	0.558	0.453	0.437
base_T	0.453	0.416	0.447	0.353	0.529	0.461	0.443
base_Q	0.444	0.333	0.445	0.475	0.581	0.414	0.449
base_LU	0.420	0.142	0.432	0.403	0.467	0.404	0.378
base_P_T	0.494	0.265	0.481	0.344	0.554	0.487	0.437
base_P_Q	0.470	0.480	0.460	0.433	0.570	0.456	0.478
base_T_Q	0.453	0.513	0.449	0.356	0.594	0.477	0.474
base_P_T_Q	0.500	0.507	0.485	0.356	0.588	0.486	0.487
base_P_T_Q_LU	0.500	0.529	0.483	0.356	0.599	0.486	0.492
cald	0.467	0.305	0.464	0.384	0.505	0.420	0.424
cald_P	0.513	0.266	0.483	0.426	0.569	0.477	0.456
cald_T	0.457	0.509	0.469	0.355	0.543	0.456	0.465

cald_Q	0.473	0.408	0.471	0.400	0.590	0.447	0.465
cald_LU	-1.66E+27	0.301	-0.559	0.178	0.485	0.276	0.136
cald_P_T	0.504	0.316	0.493	0.360	0.545	0.506	0.454
cald_P_Q	0.512	0.498	0.481	0.433	0.599	0.479	0.500
cald_T_Q	0.461	0.487	0.470	0.342	0.605	0.478	0.474
cald_P_T_Q	0.510	0.497	0.494	0.373	0.599	0.510	0.497
cald_P_T_Q_LU	-3.68E+27	0.438	-1.641	0.158	0.603	0.355	-0.017
average	0.472	0.379	0.310	0.367	0.556	0.447	

Table A11: Performance (R^2) for phosphate – forecast horizon 3.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.341	0.103	0.317	0.258	0.273	0.298	0.265
base_P	0.430	-0.145	0.365	0.352	0.356	0.408	0.294
base_T	0.395	-0.032	0.372	0.382	0.411	0.386	0.319
base_Q	0.364	0.176	0.332	0.333	0.448	0.340	0.332
base_LU	0.341	0.154	0.319	0.258	0.284	0.298	0.276
base_P_T	0.434	0.089	0.384	0.357	0.481	0.458	0.367
base_P_Q	0.424	0.225	0.360	0.338	0.420	0.389	0.359
base_T_Q	0.394	0.157	0.368	0.351	0.468	0.373	0.352
base_P_T_Q	0.448	0.235	0.391	0.346	0.504	0.450	0.396
base_P_T_Q_LU	0.448	0.301	0.392	0.346	0.504	0.450	0.407
cald	0.440	0.018	0.382	0.352	0.386	0.388	0.328
cald_P	0.491	0.211	0.423	0.405	0.456	0.451	0.406
cald_T	0.423	0.046	0.407	0.386	0.411	0.394	0.344
cald_Q	0.440	0.216	0.390	0.367	0.447	0.401	0.377
cald_LU	-8.08E+27	-0.014	-1.10	0.153	0.339	0.218	-0.080
cald_P_T	0.466	0.133	0.451	0.382	0.495	0.496	0.404
cald_P_Q	0.503	0.242	0.432	0.406	0.428	0.441	0.409
cald_T_Q	0.425	0.141	0.411	0.370	0.489	0.421	0.376
cald_P_T_Q	0.370	0.298	0.464	0.383	0.523	0.486	0.421
cald_P_T_Q_LU	-4.68E+27	0.363	-6.40	0.170	0.518	0.338	-1.002
average	0.424	0.146	-0.027	0.335	0.432	0.394	

Sulfate

Table A12: Performance (R^2) for sulfate – forecast horizon 1.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.595	0.102	0.383	0.560	0.607	0.180	0.404
base_P	0.610	0.358	0.436	0.502	0.676	0.188	0.462
base_T	0.608	0.206	0.338	0.383	0.598	0.125	0.376
base_Q	0.612	0.259	0.391	0.614	0.681	0.164	0.454
base_LU	0.595	0.118	0.386	0.560	0.590	0.180	0.405
base_P_T	0.606	0.230	0.395	0.449	0.685	0.144	0.418
base_P_Q	0.608	0.425	0.437	0.500	0.701	0.172	0.474
base_T_Q	0.621	0.396	0.330	0.417	0.670	0.134	0.428
base_P_T_Q	0.590	0.244	0.395	0.460	0.706	0.139	0.422
base_P_T_Q_LU	0.590	0.182	0.395	0.460	0.714	0.139	0.413
cald	0.590	0.233	0.373	0.362	0.620	0.052	0.372
cald_P	0.606	0.308	0.416	0.413	0.666	0.096	0.418
cald_T	0.593	0.208	0.357	0.326	0.610	0.053	0.358
cald_Q	0.607	0.332	0.384	0.380	0.688	0.061	0.409
cald_LU	-2.578	0.202	-0.051	-0.190	0.592	-0.005	-0.338
cald_P_T	0.587	0.319	0.409	0.404	0.674	0.091	0.414
cald_P_Q	0.595	0.330	0.419	0.431	0.715	0.092	0.430
cald_T_Q	0.604	0.444	0.361	0.383	0.673	0.071	0.423
cald_P_T_Q	0.570	0.130	0.408	0.427	0.706	0.088	0.388
cald_P_T_Q_LU	-1.14E+27	0.165	-0.096	0.030	0.703	0.029	0.167
average	0.425	0.260	0.343	0.394	0.664	0.110	

Table A13: Performance (R^2) for sulfate – forecast horizon 2.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.331	-0.009	0.163	0.273	0.346	0.059	0.194
base_P	0.334	-0.166	0.172	0.133	0.378	0.072	0.154
base_T	0.326	-0.058	0.138	0.173	0.394	0.022	0.166

base_Q	0.331	-0.052	0.145	0.277	0.377	0.050	0.188
base_LU	0.331	-0.024	0.163	0.273	0.354	0.059	0.193
base_P_T	0.329	-0.263	0.148	0.224	0.391	0.049	0.146
base_P_Q	0.323	-0.164	0.170	0.132	0.430	0.070	0.160
base_T_Q	0.323	-0.020	0.098	0.193	0.405	0.023	0.170
base_P_T_Q	0.302	-0.215	0.144	0.218	0.407	0.047	0.150
base_P_T_Q_LU	0.302	-0.226	0.145	0.218	0.420	0.047	0.151
cald	0.324	0.018	0.141	0.151	0.364	-0.002	0.166
cald_P	0.333	-0.069	0.178	0.202	0.413	0.024	0.180
cald_T	0.293	-0.116	0.153	0.084	0.409	-0.008	0.136
cald_Q	0.331	-0.088	0.132	0.146	0.419	0.003	0.157
cald_LU	-2.52E+26	-0.139	0.177	-0.319	0.392	-0.028	0.017
cald_P_T	0.288	-0.236	0.179	0.216	0.400	0.017	0.144
cald_P_Q	0.314	-0.054	0.173	0.203	0.434	0.021	0.182
cald_T_Q	0.296	0.051	0.124	0.086	0.414	-0.007	0.161
cald_P_T_Q	0.266	-0.388	0.161	0.204	0.433	0.014	0.115
cald_P_T_Q_LU	-1.26E+27	-0.143	0.182	-0.139	0.410	0.001	0.062
average	0.319	-0.118	0.154	0.147	0.400	0.027	

Table A14: Performance (R^2) for sulfate – forecast horizon 3.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.151	-0.549	0.043	-0.042	0.023	0.007	-0.061
base_P	0.154	-0.208	0.039	-0.019	0.114	0.016	0.016
base_T	0.125	-0.264	0.022	0.109	0.131	-0.025	0.016
base_Q	0.150	-0.238	0.019	0.073	0.141	-0.002	0.024
base_LU	0.151	-0.357	0.044	-0.042	0.058	0.007	-0.023
base_P_T	0.115	-0.430	0.038	0.086	0.157	-0.007	-0.007
base_P_Q	0.138	-0.346	0.039	0.016	0.165	0.013	0.004
base_T_Q	0.116	-0.236	0.015	0.117	0.203	-0.022	0.032
base_P_T_Q	0.098	-0.418	0.039	0.100	0.151	-0.011	-0.007
base_P_T_Q_LU	0.098	-0.365	0.039	0.100	0.179	-0.011	0.007
cald	0.161	-0.263	0.047	-0.017	0.089	-0.029	-0.002
cald_P	-0.090	-0.296	0.057	0.119	0.112	-0.018	-0.019
cald_T	0.115	-0.352	0.052	0.056	0.149	-0.034	-0.002
cald_Q	0.150	-0.080	0.039	-0.028	0.169	-0.032	0.036
cald_LU	-0.912	-0.231	-0.295	-0.399	0.062	-0.032	-0.301
cald_P_T	0.104	-0.493	0.061	0.140	0.171	-0.027	-0.007
cald_P_Q	0.134	-0.248	0.052	0.098	0.182	-0.020	0.033
cald_T_Q	0.118	-0.304	0.038	0.055	0.195	-0.036	0.011
cald_P_T_Q	0.091	-0.282	0.064	0.113	0.170	-0.028	0.021
cald_P_T_Q_LU	-2.14E+27	-0.288	-0.014	-0.212	0.138	-0.011	-7.74E-02
average	0.060	-0.312	0.022	0.021	0.138	-0.015	

Conductivity

Table A15: Performance (R^2) for conductivity – forecast horizon 1.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.763	0.632	-0.675	0.581	0.753	-0.101	0.326
base_P	0.771	0.634	-0.598	0.692	0.777	-0.105	0.362
base_T	0.769	0.640	0.062	0.589	0.774	-0.090	0.457
base_Q	0.768	0.639	-0.710	0.601	0.801	-0.122	0.329
base_LU	0.763	0.576	-0.675	0.581	0.761	-0.101	0.317
base_P_T	0.770	0.599	0.012	0.698	0.786	-0.088	0.463
base_P_Q	0.771	0.604	-0.626	0.685	0.808	-0.110	0.355
base_T_Q	0.769	0.571	0.053	0.605	0.805	-0.098	0.451
base_P_T_Q	0.771	0.644	0.010	0.701	0.804	-0.090	0.473
base_P_T_Q_LU	0.771	0.571	0.008	0.701	0.797	-0.090	0.460
cald	0.759	0.614	-0.610	0.583	0.764	-0.119	0.332
cald_P	0.769	0.595	-0.554	0.583	0.785	-0.113	0.344
cald_T	0.759	0.605	0.060	0.520	0.763	-0.106	0.434
cald_Q	0.763	0.608	-0.660	0.606	0.810	-0.123	0.334
cald_LU	0.405	0.575	0.040	0.195	0.752	-0.123	0.307
cald_P_T	0.758	0.607	0.007	0.545	0.781	-0.098	0.433
cald_P_Q	0.769	0.580	-0.583	0.570	0.801	-0.115	0.337
cald_T_Q	0.759	0.653	0.050	0.531	0.799	-0.105	0.448
cald_P_T_Q	0.763	0.665	0.003	0.530	0.798	-0.100	0.443
cald_P_T_Q_LU	-6.92E+28	0.598	0.119	0.307	0.817	-0.111	0.346

average	0.746	0.610	-0.263	0.570	0.787	-0.105	

Table A16: Performance (R^2) for conductivity – forecast horizon 2.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.504	0.312	-0.721	0.259	0.473	-0.110	0.119
base_P	0.493	0.250	-0.615	0.399	0.569	-0.106	0.165
base_T	0.517	0.314	0.017	0.357	0.512	-0.093	0.271
base_Q	0.508	0.327	-0.716	0.279	0.532	-0.117	0.135
base_LU	0.504	0.295	-0.721	0.259	0.507	-0.110	0.122
base_P_T	0.504	0.295	-0.041	0.434	0.569	-0.090	0.278
base_P_Q	0.499	0.331	-0.643	0.406	0.576	-0.108	0.177
base_T_Q	0.518	0.243	0.007	0.377	0.568	-0.092	0.270
base_P_T_Q	0.513	0.212	-0.049	0.437	0.564	-0.087	0.265
base_P_T_Q_LU	0.513	0.220	-0.048	0.437	0.582	-0.087	0.269
cald	0.500	0.320	-0.659	0.365	0.490	-0.122	0.149
cald_P	0.490	0.246	-0.594	0.409	0.560	-0.117	0.166
cald_T	0.501	0.331	0.018	0.306	0.506	-0.109	0.259
cald_Q	0.503	0.296	-0.672	0.355	0.539	-0.120	0.150
cald_LU	-0.388	0.247	0.004	0.076	0.500	-0.126	0.052
cald_P_T	0.481	0.223	-0.040	0.388	0.546	-0.105	0.249
cald_P_Q	0.497	0.290	-0.617	0.402	0.566	-0.117	0.170
cald_T_Q	0.503	0.287	0.001	0.301	0.566	-0.110	0.258
cald_P_T_Q	0.495	0.231	-0.045	0.373	0.565	-0.106	0.252
cald_P_T_Q_LU	-2.90E+28	0.246	0.062	0.254	0.589	-0.116	0.207
average	0.453	0.276	-0.304	0.344	0.544	-0.107	

Table A17: Performance (R^2) for conductivity – forecast horizon 3.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.296	-0.068	-0.760	0.017	0.254	-0.110	-0.062
base_P	0.277	0.101	-0.662	0.100	0.367	-0.108	0.012
base_T	0.320	-0.092	-0.031	0.123	0.287	-0.100	0.085
base_Q	0.301	0.033	-0.736	0.039	0.322	-0.112	-0.025
base_LU	0.296	0.063	-0.760	0.017	0.251	-0.110	-0.041
base_P_T	0.304	0.254	-0.093	0.183	0.361	-0.092	0.153
base_P_Q	0.280	0.288	-0.686	0.117	0.391	-0.106	0.047
base_T_Q	0.316	-0.092	-0.048	0.145	0.323	-0.099	0.091
base_P_T_Q	0.310	0.155	-0.095	0.190	0.372	-0.092	0.140
base_P_T_Q_LU	0.310	0.140	-0.095	0.190	0.363	-0.092	0.136
cald	0.315	-0.108	-0.720	0.172	0.231	-0.123	-0.039
cald_P	0.284	0.168	-0.635	0.231	0.360	-0.121	0.048
cald_T	0.328	-0.089	-0.038	0.245	0.256	-0.112	0.098
cald_Q	0.310	-0.013	-0.702	0.175	0.319	-0.121	-0.005
cald_LU	-1.65E+28	0.117	-0.023	-0.117	0.305	-0.128	0.031
cald_P_T	0.005	0.195	-0.096	0.285	0.331	-0.109	0.102
cald_P_Q	0.288	0.214	-0.653	0.236	0.361	-0.118	0.055
cald_T_Q	0.326	-0.157	-0.051	0.247	0.368	-0.111	0.103
cald_P_T_Q	0.307	0.161	-0.096	0.289	0.366	-0.109	0.153
cald_P_T_Q_LU	-2.14E+28	0.207	0.022	0.079	0.387	-0.119	0.115
average	0.286	0.074	-0.348	0.148	0.329	-0.110	

Bentazon

Table A18: Performance (R^2) for bentazon – forecast horizon 1.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.516	0.106	0.539	0.323	0.236	0.564	0.381
base_P	0.486	0.204	0.539	0.148	0.314	0.464	0.359
base_T	0.410	-1.390	0.548	0.043	0.141	0.534	0.048
base_Q	0.501	-0.855	0.538	0.328	0.251	0.560	0.221
base_LU	0.516	-0.489	0.539	0.323	0.259	0.564	0.285
base_P_T	0.388	-0.600	0.542	0.042	0.184	0.497	0.176
base_P_Q	0.484	-0.379	0.544	0.119	0.326	0.478	0.262
base_T_Q	0.407	-1.439	0.549	-0.053	0.162	0.531	0.026
base_P_T_Q	0.379	-0.521	0.558	-0.040	0.276	0.499	0.192
base_P_T_Q_LU	0.379	-0.574	0.556	-0.040	0.240	0.499	0.177
cald	0.512	-0.569	0.582	0.125	0.285	0.527	0.244
cald_P	0.471	-0.007	0.578	0.108	0.375	0.447	0.329

cald_T	0.467	-1.320	0.585	0.170	0.202	0.545	0.108
cald_Q	0.505	-0.559	0.584	0.111	0.306	0.534	0.247
cald_LU	-9.644	-0.304	0.236	-0.017	0.348	0.181	-1.533
cald_P_T	0.413	-0.443	0.577	0.186	0.316	0.501	0.258
cald_P_Q	0.459	-0.664	0.587	0.083	0.331	0.445	0.207
cald_T_Q	0.452	-0.870	0.587	0.201	0.198	0.548	0.186
cald_P_T_Q	0.385	-0.516	0.586	0.191	0.277	0.512	0.239
cald_P_T_Q_LU	-1.30E+25	-0.506	0.537	0.063	0.322	0.101	0.104
average	-0.106	-0.585	0.545	0.121	0.267	0.476	

Table A19: Performance (R^2) for bentazon – forecast horizon 2.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.247	-2.809	0.311	0.053	-0.162	0.354	-0.334
base_P	0.222	-1.041	0.325	-0.128	0.010	0.309	-0.051
base_T	0.165	-3.329	0.343	-0.093	-0.088	0.342	-0.443
base_Q	0.238	-3.156	0.317	0.077	-0.055	0.350	-0.372
base_LU	0.247	-2.702	0.311	0.053	-0.139	0.354	-0.313
base_P_T	0.135	-0.421	0.350	-0.449	-0.003	0.320	-0.011
base_P_Q	0.214	-1.063	0.325	-0.153	0.063	0.335	-0.046
base_T_Q	0.161	-1.872	0.345	-0.262	-0.126	0.332	-0.237
base_P_T_Q	0.135	-1.130	0.342	-0.493	-0.069	0.314	-0.150
base_P_T_Q_LU	0.135	-1.257	0.347	-0.493	-0.071	0.314	-0.171
cald	0.240	-3.557	0.404	-0.012	-0.138	0.370	-0.449
cald_P	0.148	-1.305	0.393	-0.012	-0.074	0.302	-0.091
cald_T	0.269	-0.503	0.417	-0.048	-0.087	0.356	0.067
cald_Q	0.236	-1.646	0.403	-0.105	-0.119	0.341	-0.148
cald_LU	-16.622	-2.431	-0.108	-0.058	-0.114	0.128	-3.201
cald_P_T	0.153	-0.701	0.402	-0.147	-0.069	0.311	-0.008
cald_P_Q	0.126	-0.972	0.403	-0.058	-0.100	0.307	-0.049
cald_T_Q	0.267	-3.121	0.418	-0.136	-0.135	0.353	-0.392
cald_P_T_Q	0.112	-0.988	0.403	-0.141	-0.062	0.322	-0.059
cald_P_T_Q_LU	-4.29E+27	-1.090	0.373	0.001	-0.099	0.016	-0.160
average	-0.738	-1.755	0.341	-0.130	-0.082	0.307	

Table A20: Performance (R^2) for bentazon – forecast horizon 3.

	LR	DT	LSVR	KNR	RFR	SVR	average
base	0.065	-0.760	0.117	-0.233	-0.173	0.106	-0.146
base_P	0.020	-2.703	0.155	-0.180	-0.158	0.068	-0.466
base_T	-0.041	-1.572	0.196	-0.241	-0.149	0.238	-0.262
base_Q	0.059	-1.789	0.124	-0.272	-0.377	0.119	-0.356
base_LU	0.065	-1.623	0.116	-0.233	-0.160	0.106	-0.288
base_P_T	-0.103	-0.864	0.210	-0.501	-0.095	0.222	-0.189
base_P_Q	0.041	-1.631	0.158	-0.299	-0.276	0.118	-0.315
base_T_Q	-0.054	-1.424	0.198	-0.378	-0.204	0.245	-0.269
base_P_T_Q	-0.079	-1.360	0.206	-0.420	-0.137	0.224	-0.261
base_P_T_Q_LU	-0.079	-0.933	0.206	-0.420	-0.180	0.224	-0.197
cald	0.033	-1.573	0.227	-0.281	-0.057	0.264	-0.231
cald_P	-0.063	-1.464	0.203	-0.194	-0.128	0.191	-0.242
cald_T	0.031	-1.688	0.222	-0.260	0.029	0.280	-0.231
cald_Q	0.036	-3.494	0.230	-0.431	-0.198	0.226	-0.605
cald_LU	-20.816	-1.778	-1.440	-0.045	-0.211	0.144	-4.024
cald_P_T	-0.068	-0.852	0.208	-0.256	-0.015	0.211	-0.129
cald_P_Q	-0.044	-3.663	0.207	-0.269	-0.167	0.182	-0.626
cald_T_Q	0.026	-1.396	0.226	-0.334	-0.057	0.288	-0.208
cald_P_T_Q	-0.037	-0.870	0.217	-0.215	-0.112	0.211	-0.134
cald_P_T_Q_LU	-1.10E+29	-1.328	0.265	0.055	-0.184	0.100	-0.218
average	-1.165	-1.638	0.113	-0.270	-0.150	0.188	

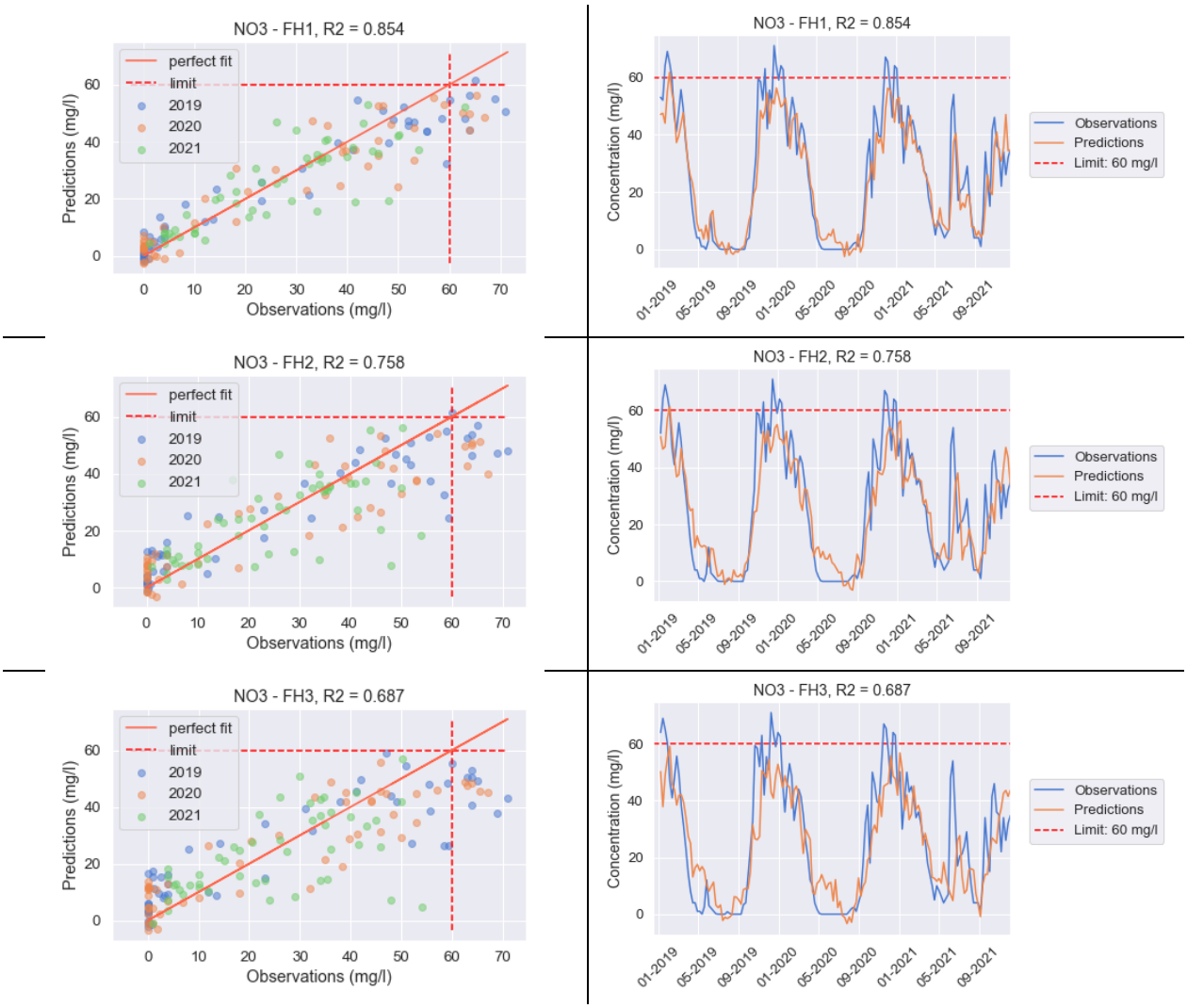
2B: Optimal performing water quality models

2B1. Nitrate

The optimal performing water quality models for nitrate prediction are shown in Table A21. The scatterplots and time series are included in Figures A8 – A13.

Table A21: Optimal performing models for nitrate prediction.

Variable	Forecast horizon	Algorithm	Feature set	Model R2	Baseline R2	ΔR2
NO3	1 week	LR	Calendar + P	0.854	0.858	-0.004
	2 weeks	LR	Calendar + P + Q	0.758	0.722	0.036
	3 weeks	LR	Calendar + P + Q	0.687	0.59	0.097



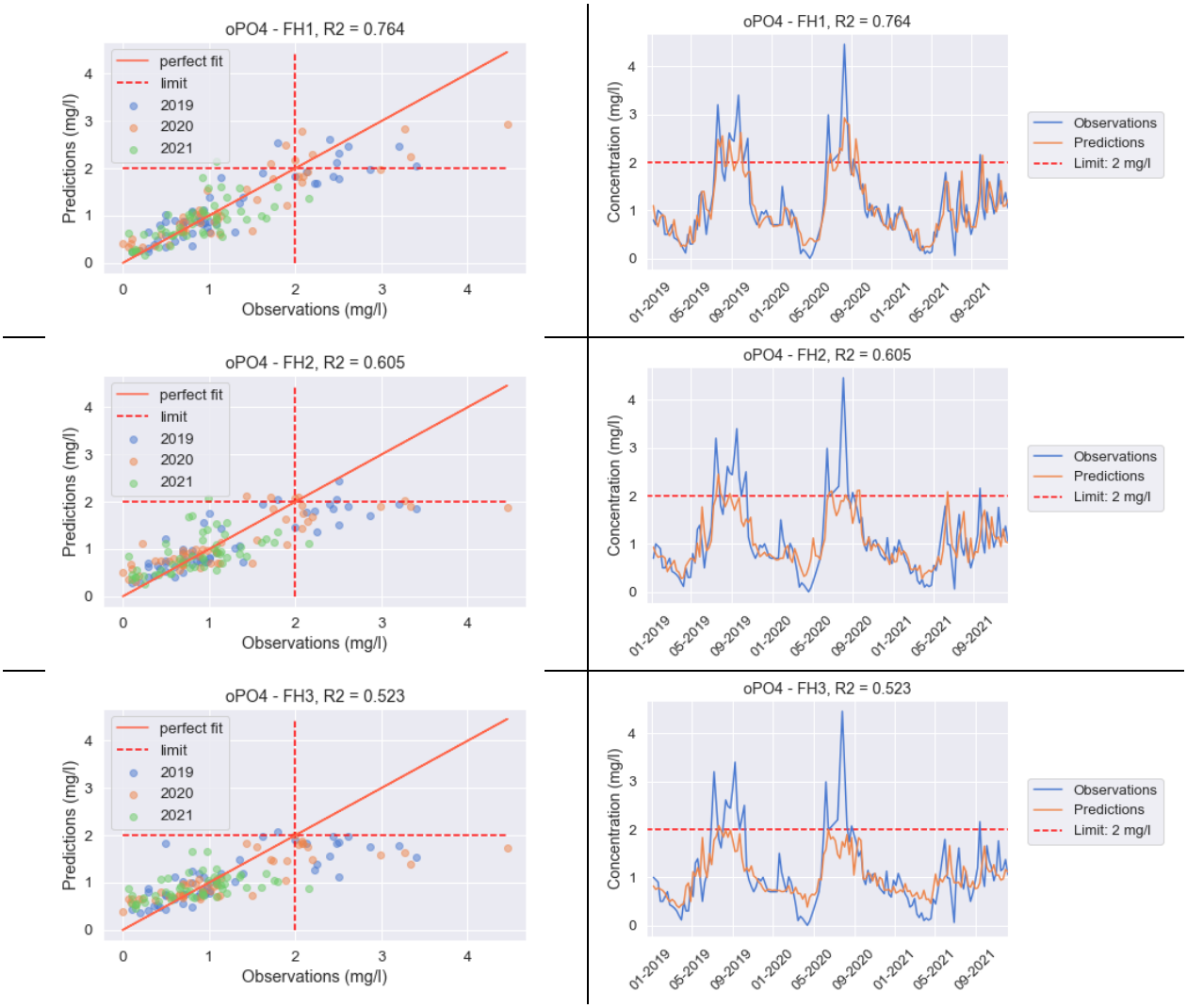
Figures A8 – A13: Scatterplots and time series showing nitrate predictions and observations for all tested years.

2B2. Phosphate

The optimal performing water quality models for phosphate prediction are shown in Table A22. The scatterplots and time series are included in Figures A14 – A19.

Table A22: Optimal performing models for nitrate prediction.

Variable	Forecast horizon	Algorithm	Feature set	Model R2	Base R2	ΔR2
oPO4	1 week	RFR	Calendar + P + T + Q + LU	0.764	0.737	0.027
	2 weeks	RFR	Calendar + T + Q	0.605	0.411	0.194
	3 weeks	RFR	Calendar + P + T + Q	0.523	0.241	0.282



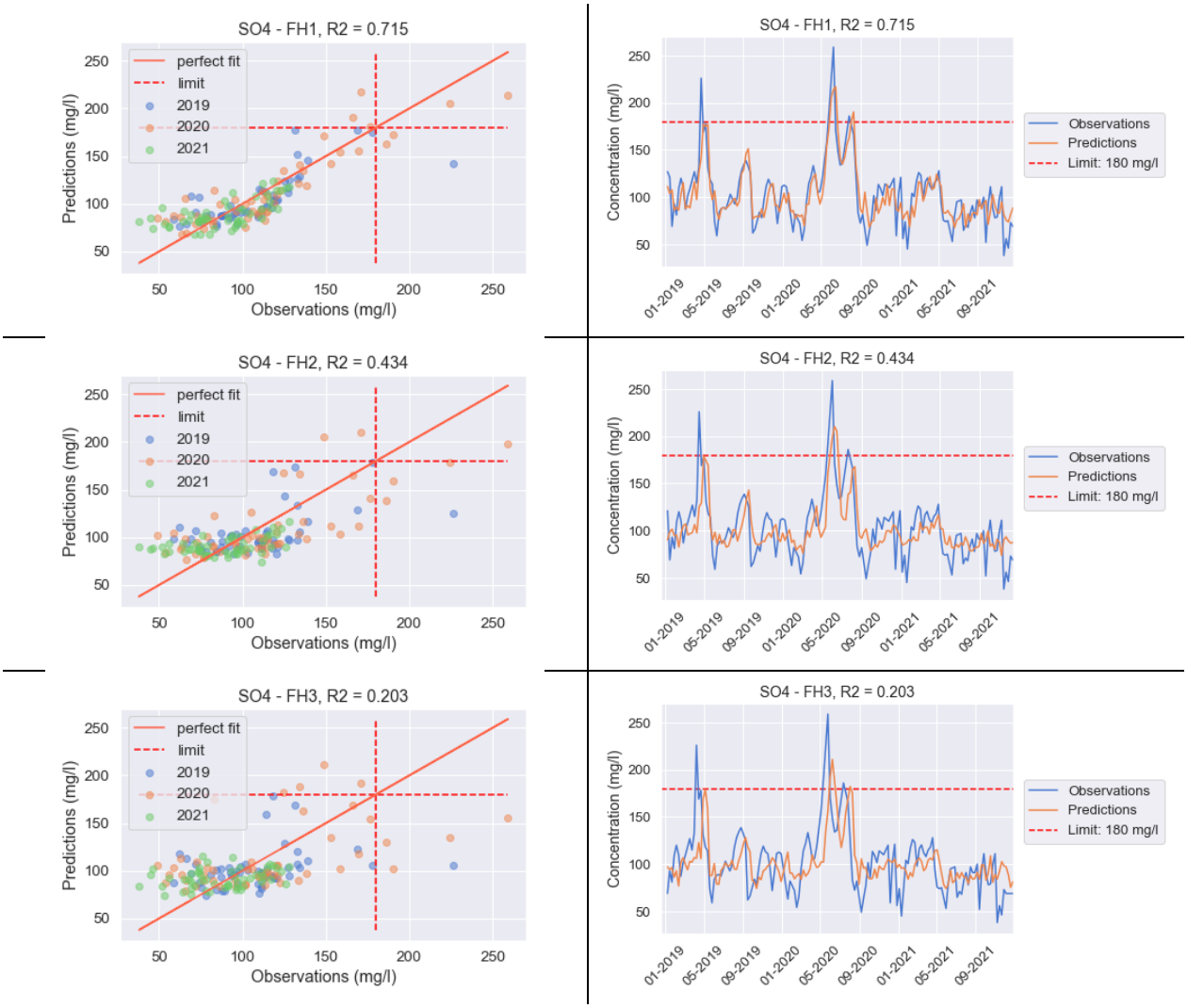
Figures A14 – A19: Scatterplots and time series showing phosphate predictions and observations for all tested years.

2B3. Sulfate

The optimal performing water quality models for sulfate prediction are shown in Table A23. The scatterplots and time series are included in Figures A20 – A25.

Table A23: Optimal performing models for sulfate prediction.

Variable	Forecast horizon	Algorithm	Feature set	Model R2	Base R2	ΔR2
SO4	1 week	RFR	Calendar + P + Q	0.715	0.489	0.226
	2 weeks	RFR	Calendar + P + Q	0.434	0.149	0.285
	3 weeks	RFR	Base + T + Q	0.203	-0.239	0.442



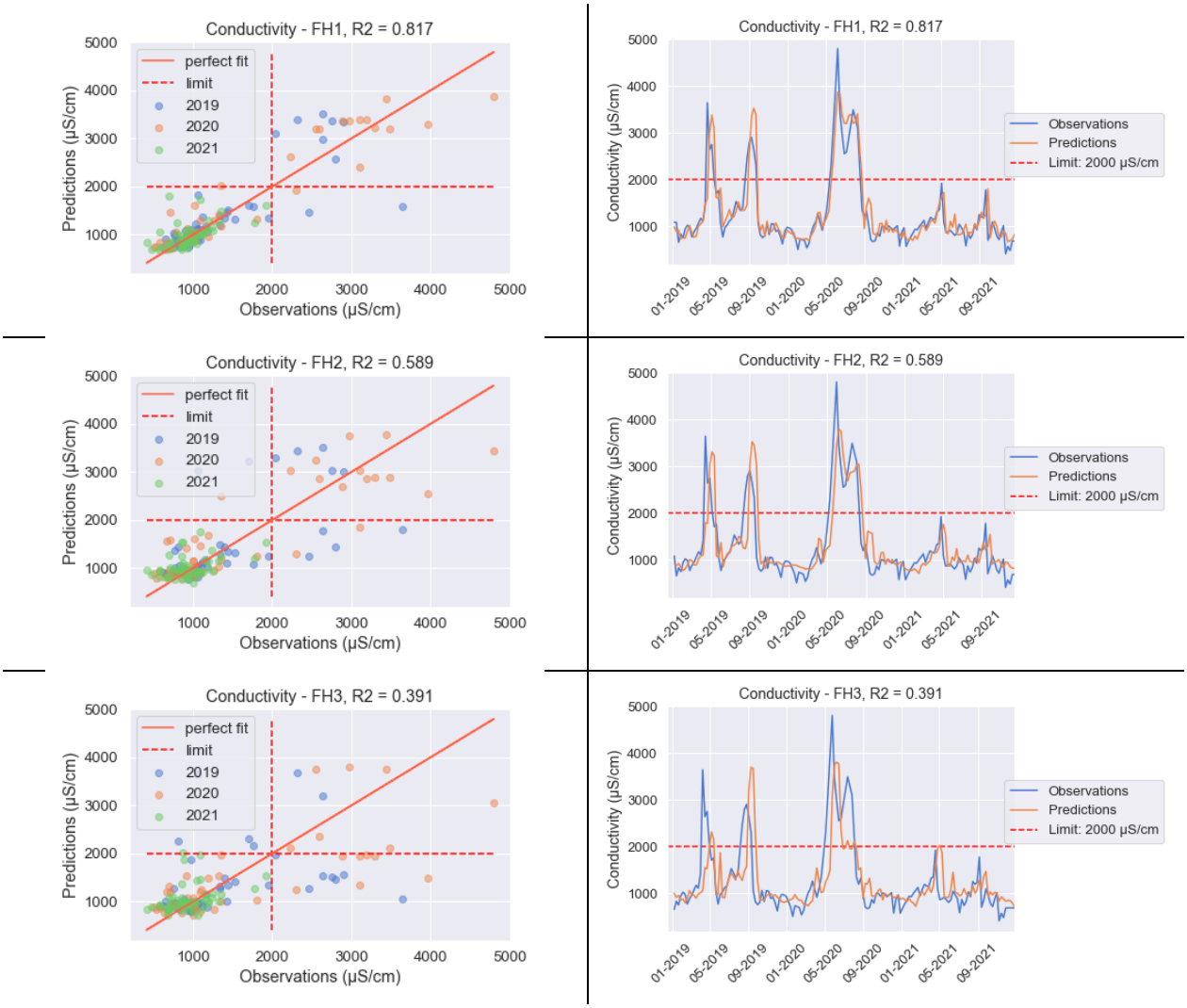
Figures A20 – A25: Scatterplots and time series showing sulfate predictions and observations for all tested years.

2B4. Conductivity

The optimal performing water quality models for conductivity prediction are shown in Table A24. The scatterplots and time series are included in Figures A26 – A31.

Table A24: Optimal performing models for conductivity prediction.

Variable	Forecast horizon	Algorithm	Feature set	Model R2	Base R2	ΔR2
Conductivity	1 week	RFR	Calendar + P + T + Q + LU	0.817	0.743	0.074
	2 weeks	RFR	Calendar + P + T + Q + LU	0.589	0.475	0.114
	3 weeks	RFR	Base + P + Q	0.391	0.168	0.223



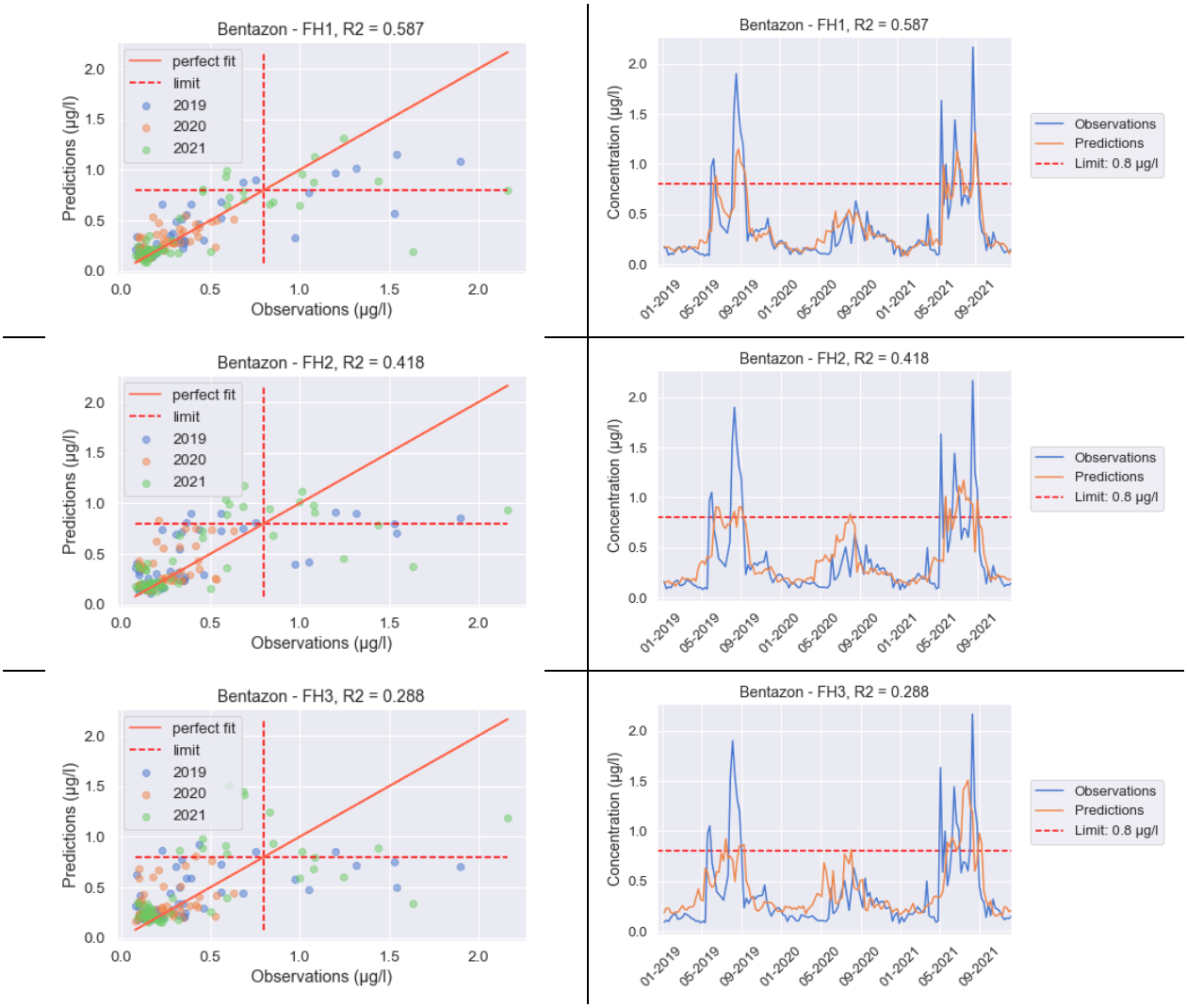
Figures A26 – A31: Scatterplots and time series showing conductivity predictions and observations for all tested years.

2B5. Bentazon

The optimal performing water quality models for bentazon prediction are shown in Table A25. The scatterplots and time series are included in Figures A32 – A37.

Table A25: Optimal performing models for bentazon prediction.

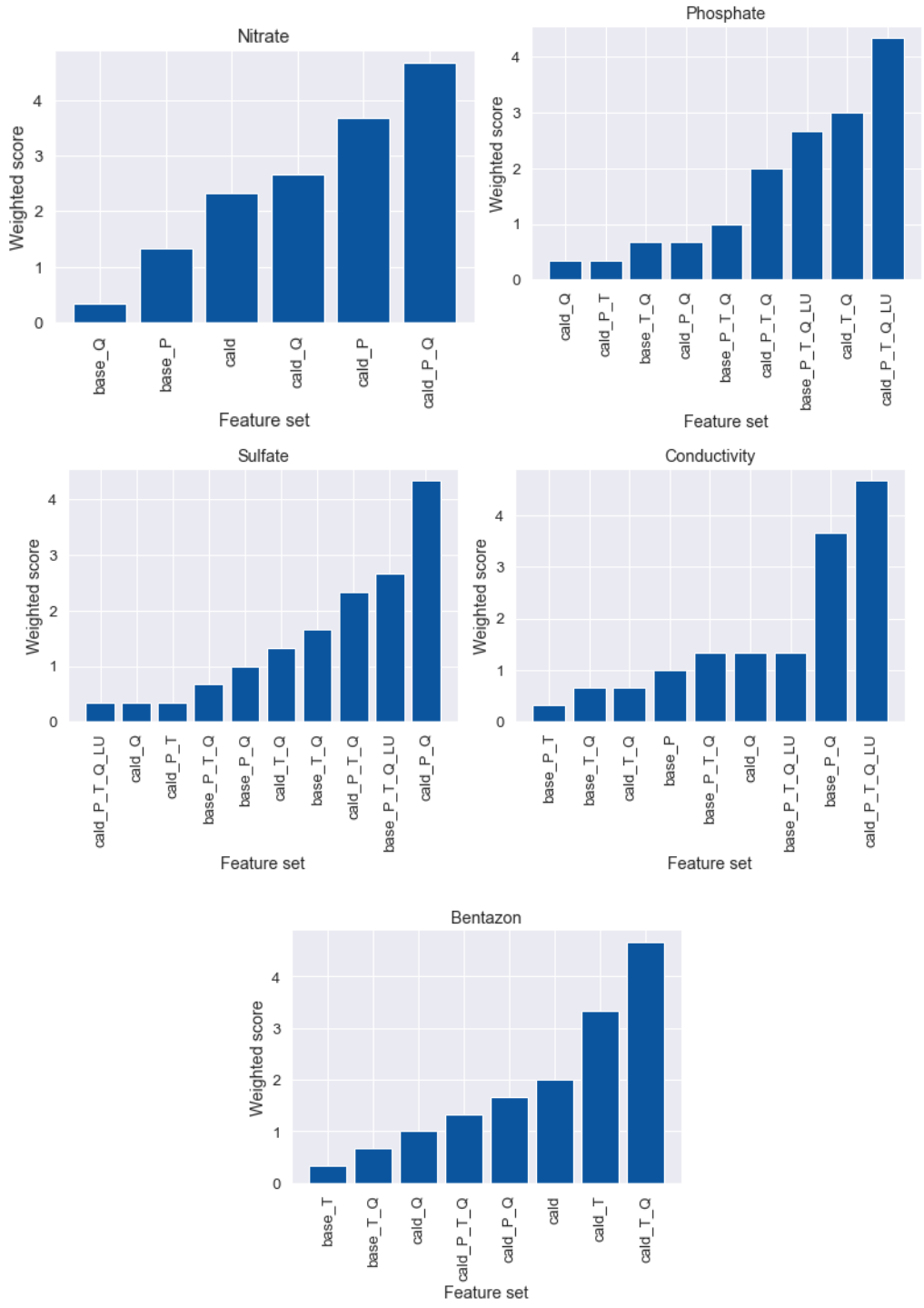
Variable	Forecast horizon	Algorithm	Feature set	Model R2	Base R2	ΔR2
Bentazon	1 week	LSVR	Calendar + P/T + Q	0.587	0.452	0.135
	2 weeks	LSVR	Calendar + T + Q	0.418	0.066	0.352
	3 weeks	SVR	Calendar + T + Q	0.288	-0.214	0.502



Figures A32 – A37: Scatterplots and time series showing bentazon predictions and observations for all tested years.

2C: Weighted optimal feature sets

The feature sets are weighted over all forecast horizons to find a general feature set per water quality parameter for all forecast horizons. The results are shown in Figures A38 – A42.



Figures A38 – A42: Weighted optimal feature set per water quality parameter over all forecast horizons.