Application of Pastas for real-time groundwater level forecasting in Northern Limburg

MSc Internship

waterschap limburg

WAGENINGEN UNIVERSITY & RESEARCH

Nelen & Schuurmans

Karl Schütt

11-10-2019

Application of Pastas for real-time groundwater level forecasting in Northern Limburg

MSc Internship

Author:
Karl Schütt
Registration number: 950322756110

Supervised by:
Thomas Berends (Nelen & Schuurmans)
Evert Wielsma (Nelen & Schuurmans)
Jac Peerboom (Waterschap Limburg)

Examiner:
Victor Bense (Wageningen University and Research)

Nelen & Schuurmans
Zakkendragershof 34-44
3511 AE Utrecht

www.nelen-schuurmans.nl

Projectgegevens

Datum : 11-10-2019

Preface

This report is the result of my MSc internship at Nelen & Schuurmans which I did in collaboration with Waterboard Limburg. Fulfilling this internship is the final part of completing my Master's degree at Wageningen University and Research. During the internship I worked mainly at the office of Nelen & Schuurmans, but I also worked at the headquarters of Waterboard Limburg. It was therefore an experience by which I became more acquainted with both fields.

There are several persons I want to thank. First, I would like to thank Thomas Berends and Evert Wielsma, my supervisors at Nelen & Schuurmans, for their guidance and advice during my internship. Besides I would like to thank Jac Peerboom (Waterboard Limburg) for his feedback and providing the opportunity to work at the waterboard. Also, I would like to thank the other involved people at Waterboard Limburg for their support, feedback and input during the discussions: Nicole Mulders, Peter Hulst, Karlijn Kessels, Arnoud Soetens and Jurriaan Cok.

Abstract

Waterboard Limburg and Nelen & Schuurmans are working on an operational decision supportive system (BOS-OMAR), which can be used for optimizing the management of surface water and groundwater. Accurate real-time forecasts of local groundwater levels at monitoring wells could be useful in order to verify the performance of area covering groundwater level predictions of BOS-OMAR. It was explored during this study whether Pastas could be suitable for this application. Pastas is an open source Python package for processing, simulating and analyzing hydrological timeseries. Hydrological timeseries are simulated using transfer-noise models, which convert timeseries of stresses to the simulated variable using response functions. A script was developed for running real-time groundwater level forecasts with Pastas. Six groundwater level timeseries from monitoring locations within Northern Limburg were selected for testing. First, it was verified how well the groundwater level timeseries could be simulated with Pastas using precipitation, potential evaporation, the surface water level and discharge of the Meuse as stresses. Secondly, hindcasts of the groundwater levels were performed with the developed script using calibration timespans of varying lengths. It was concluded that at most locations the groundwater dynamics could be explained well using the precipitation excess (precipitation minus potential evaporation) as stress. However, some sub-yearly patterns could not be explained. This is most likely caused by uncertainty in input data, exclusion of local stresses and limitations of the models in accounting for the non-linear nature of groundwater response to meteorological forcing. The influence of the length of historical data used for calibration on hindcasting performance turned out to be location specific.

Contents

1 Introduction

1.1 Introduction

It is expected that droughts will occur more frequently in the Netherlands during summertime (van den Hurk, et al., 2014; van der Linden, et al., 2019). The occurrence of droughts has a negative impact on agriculture and nature (Stokkers, et al., 2018; Vos & Kuiters, 2007). A precipitation deficit can lead to limited water availability in the subsurface, which limits the growth of agricultural crops and other vegetation species. On the long term, meteorological droughts cause hydrological droughts because the groundwater tables will drop when recharge remains limited (Tallaksen & van Lanen, 2004). The higher located sand covered areas in the Netherlands are vulnerable, since the recharge of the groundwater system over there mainly depends on the local precipitation excess.

Early insights in the development of droughts are important for taking measures to diminish future impacts as much as possible. Typical measures involve temporary interventions in local and regional water systems, such as the supply of surface water from elsewhere, retaining water as much as possible and stimulating spare use of the available ground and surface water resources. With the aid of meteorological and hydrological models, the impact of future droughts can be assessed.

Waterboard Limburg and Nelen & Schuurmans are working on an operational decision supportive instrument (BOS-OMAR) (Graaf, et al., 2017), which can be used for optimizing the management of surface water and groundwater. BOS-OMAR generates short-term area covering forecasts of the groundwater levels in the former management area of Waterboard Peel en Maasvallei. Besides, surface water levels and fluxes are also forecasted. Forecasts are generated two times every day. At the moment the model is still under development, and it is ought to generate forecasts for a period of one to four months ahead every two weeks.

Additional insights in the future dynamics of groundwater levels at monitoring locations would be valuable, because these could be helpful to assess the accuracy of area covering groundwater level forecasts from BOS-OMAR. In order to obtain such insights, a separate modelling tool could be developed with Pastas (Collenteur, et al., 2019) for independently generating real-time forecasts of groundwater levels at monitoring wells. Pastas is an open source Python package which can be used for processing, simulating and analyzing groundwater level timeseries. Timeseries are simulated with the aid of transfer-noise models that calculate groundwater levels based on only explanatory timeseries. The advantage of such models is that there is no information required about physical properties of the subsurface, and that only timeseries of the precipitation surplus and optionally other explanatory timeseries have to be used as input (Knotters & Bierkens, 1999a). Transfer-noise models are often used for simulating and analyzing timeseries, but it has also been stated that these models can be applied in the field of forecasting (Keith, et al., 1994).

Pastas can easily be integrated in larger applications because it is written in Python, and therefore it is interesting to explore the possibilities. However, no specific functionality for providing real-time forecasts of timeseries has been implemented in Pastas yet. Challenges of performing real-time groundwater level forecasts with Pastas are describing the groundwater dynamics as good as possible by explanatory timeseries and keeping the timeseries models actualized as much as possible. Since groundwater level dynamics are often not stationary in time (Ritzema, et al., 2012), it would be interesting to explore which influence the length of the selected historical calibration data has on the performance in predicting groundwater levels in the future.

1.2 Goal

The goal of this study is evaluating whether Pastas is suitable for generating real-time forecasts of groundwater levels at monitoring wells in the northern part of the management area of Waterboard Limburg. Firstly, a script for generating real-time groundwater level forecasts with Pastas was written in Python. Afterwards, timeseries models were created for simulating groundwater levels at six monitoring sites within the study area. After having set up the models, they were tested for real-time application. It was explored how well these models performed and which influence varying the length of the calibration dataset had on the performance of the forecasts.

1.3 Research questions

The main research question of this study is: *Up to which extent is Pastas suitable for generating real-time groundwater level forecasts at monitoring wells within Northern Limburg?*

Besides, the following sub questions were composed:

- Which stresses explain the variance of groundwater dynamics locally, and up to how far can the groundwater dynamics be explained?
- How well can real-time groundwater level forecasts be performed with Pastas, and what influence does the length of the calibration period have on the performance?

1.4 Results

The results of this study are:

- A concept of a functionality for Pastas for performing real-time groundwater level forecasts at monitoring wells
- This report in which the method and outcomes of the study are described

1.5 Reading guide

In chapter 2, the selected data and project area are discussed. The methodology is explained in chapter 3. The results are included in chapter 4 and the results are discussed in chapter 5. The conclusion and recommendations are given in chapter 6.

2 Study area and data

In this chapter, the study area will be described. Afterwards, the selected data will be discussed

2.1 Study area and geohydrology

The study area involves the former management area of Waterboard Peel en Maasvallei (see figure 1). The subsurface in the area largely consists out of sandy deposits (Doppert, et al., 1975). The shallowest part of the subsurface mainly involves the formation of Boxtel, which can be characterized by moderately permeable eolian sandy deposits which have a total thickness varying from 2 to 20 meters locally (Massop, et al., 2005). The more permeable formations of Kreftenheye, Veghel, Sterksel and Kiezeloöliet are located directly below and are considered as the phreatic aquifers in the area (van Rooijen, 1989). These formations have been deposited by the Meuse, and contain coarse sand and gravel. The Miocene sands (formation of Breda) are poorly permeable and can be considered as the hydrological base of the groundwater system. Alongside the river Meuse, Holocene clayey or loamy fluvial deposits are found which belong to the Betuwe formation.

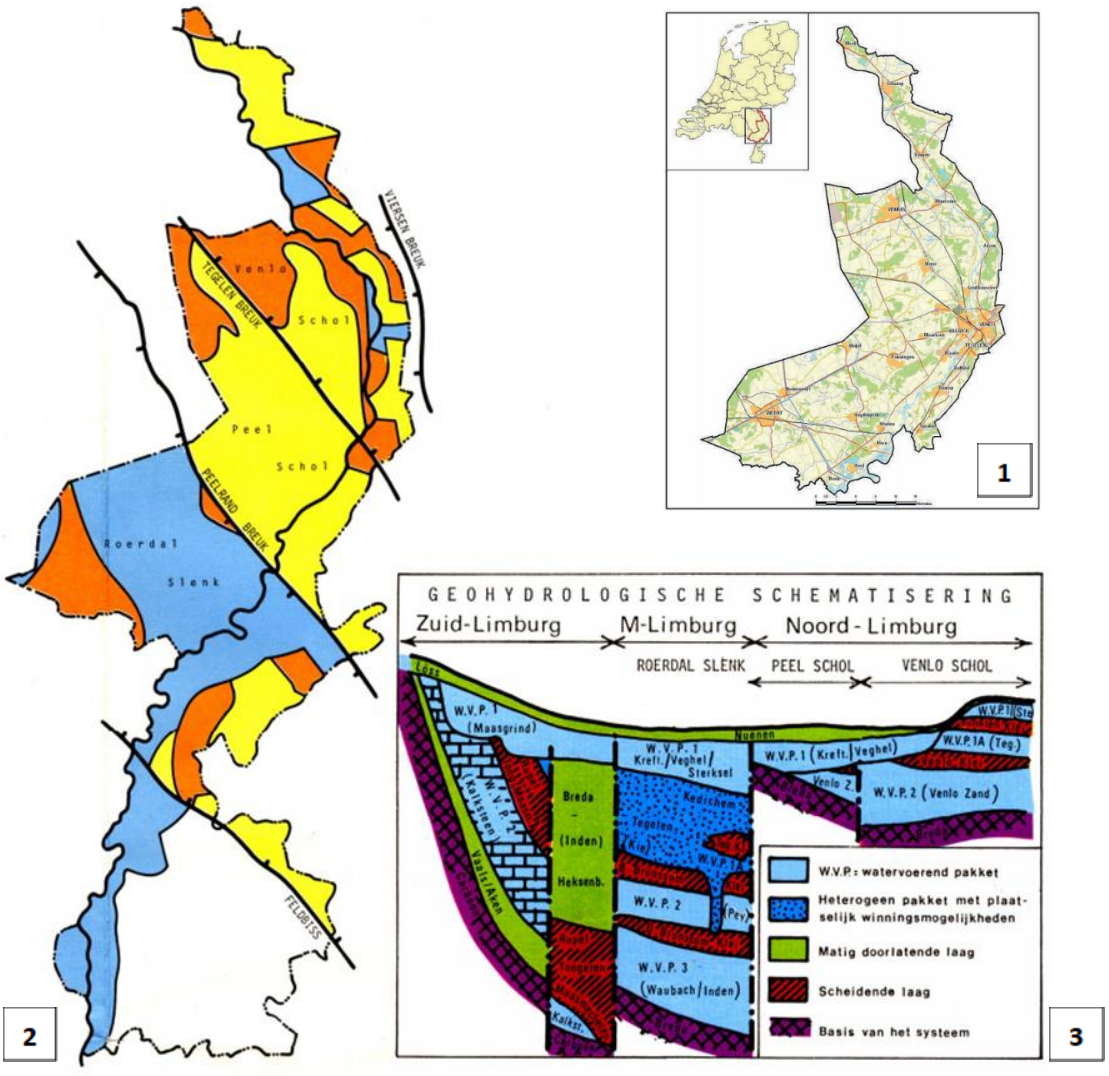


Figure 1: Study area and geohydrological structure (Massop, et al., 2005; Ogink, 2011). 1: Study area. Southern Limburg and a part of Middle Limburg are not included., 2: Faults in the province of Limburg. The Peelrand fault and Tegelen fault cross the study area. 3: Visualization of the geohydrology of the province of Limburg. Aquifers are marked in blue; aquitards are marked in brown and moderately permeable layers are marked in green. The hydrological base is marked in purple. The left part (Southern Limburg) does not belong to the study area.

Some faults that are part of the Roerdal graben system cross the area, of which the Peelrandbreuk is the most well-known (figure 1). The faults have an impact on groundwater flow since they contain fine sediments that are poorly permeable. Seepage occurs locally at the higher located side of the faults (van Balen, 2009). Also, human interactions with the subsurface and groundwater system influence the groundwater level regime locally, such as drainage and groundwater abstractions for example. Further, groundwater tables nearby the Meuse, channels and streams are locally influenced by the respective surface water levels and discharges.

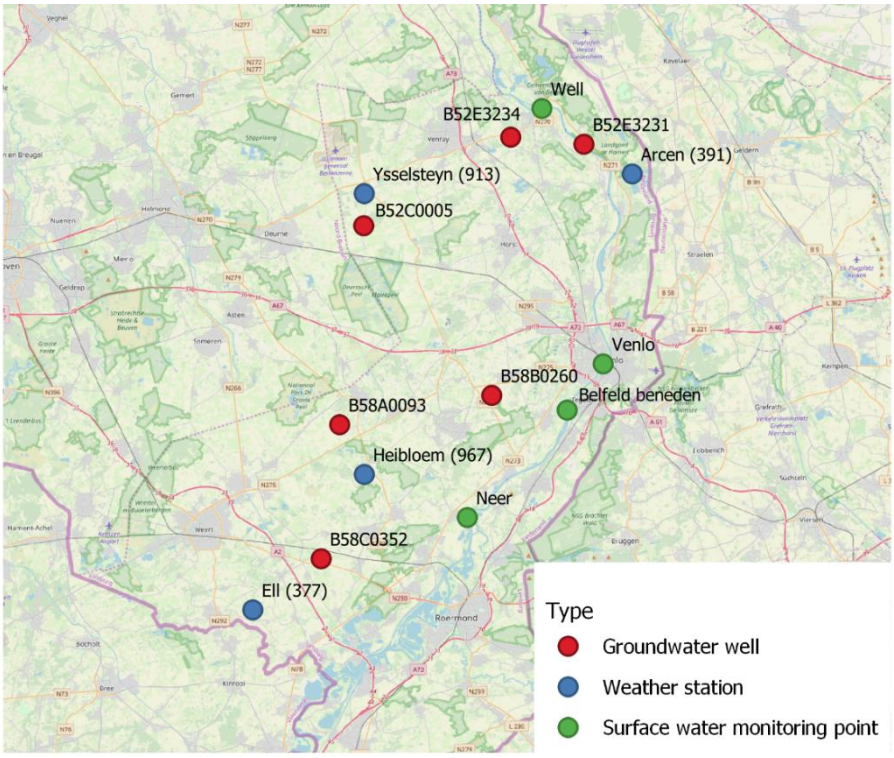


Figure 2: Overview of the selected groundwater wells and the monitoring locations of potential explanatory stresses.

WELL	SELECTION OF (POTENTIAL) EXPLANATORY STRESSES			
	Precipitation	*Reference evaporation*	*Water level Meuse*	*Discharge Meuse*
B58C0352	Ell	Ell	Neer	Venlo
B58A0093	Heibloem	Ell	Neer	Venlo
B52C0005	Ysselsteyn	Arcen	Well	Venlo
B58B0260	Heibloem	Arcen	Belfeld	Venlo
B52E3231	Arcen	Arcen	Well	Venlo
B52E3234	Arcen	Arcen	Well	Venlo

Table 1: Selection of potential stresses for explaining groundwater level dynamics. For each groundwater well, timeseries of potential explanatory stresses were obtained from the most nearby located monitoring location with available data. It should be noted that there was no reference evaporation data accessible at the weather stations at Ell and Heibloem. Further, measurements of the discharge of the Meuse were only available at Venlo.

2.2 Data

Six groundwater level timeseries were selected for this study (figure 2, table 1). The locations of measurement are equally distributed over the former management area of waterboard Peel en Maasvallei and all have a different distance towards the river Meuse. For simulating the groundwater level timeseries, potential explanatory timeseries of stresses were selected. Explanatory meteorological timeseries for each groundwater monitoring location where gathered from the most nearby located weather stations. Also, additional explanatory surface water timeseries of the Meuse were selected from the most nearby located monitoring points. Timeseries of other local stresses, such as local surface water levels of streams and groundwater abstractions, have not been selected since they were mostly not complete or scarcely available. All the selected timeseries have a daily observation frequency.

WELL	SURFACE LEVEL (M W.R.T. NAP)	FILTER NUMBER	TOP OF FILTER (M W.R.T. SURFACE LEVEL)	BOTTOM OF FILTER (M W.R.T. SURFACE LEVEL)	MONITORING PERIOD
B58C0352	28,51	1	-1,80	-3,8	2001 - 2017
B58A0093	30,83	2	-30	-32	2001 - 2017
B52C0005	29,53	1	-24,60	-44,91	2012 - 2019
B58B0260	34,60	2	-6,16	-9,16	2012 - 2019
B52E3231	13,74	1	-2.91	-3.91	2008 - 2017
B52E3234	19,75	1	-2.39	-3.39	2008 - 2018

Table 2: Information of the groundwater monitoring wells

2.2.1 Groundwater timeseries

The selected groundwater level timeseries where obtained from Dinoloket. The monitoring locations of the selected timeseries are shown in figure 2. Besides, an overview of the most important information of the monitoring wells is added in table 2. A more extensive overview of the local structure of the subsurface at the monitoring locations is added in appendix I.

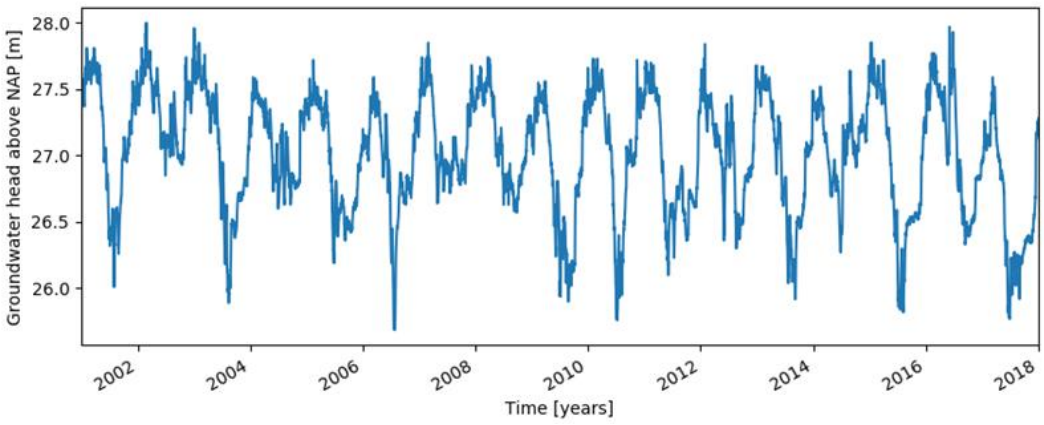


Figure 3: Timeseries of the hydraulic head at monitoring well B58C0352

B58C0352

Monitoring location B58C0352 is located southward to Leveroy. The observed timeseries involve the average groundwater head between a depth of 1,8 and 3,8 meters below surface level within the moderately permeable formation of Boxtel. Because of the shallow depth, the hydraulic head is considered to be sensitive to human interactions in the subsurface. The local land coverage type consists of grassland. Measurements of the hydraulic head have been taken on a daily frequency between 2001 and 2017.

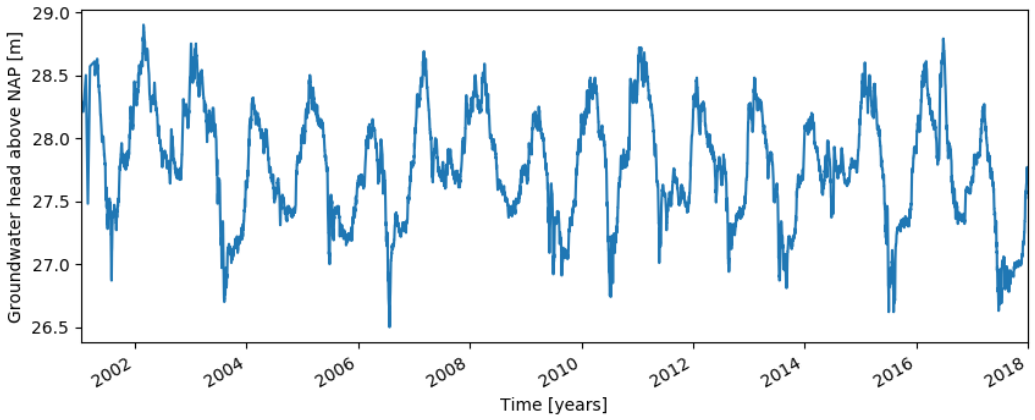


Figure 4: Timeseries of the hydraulic head at monitoring well B58A0093

B58A0093

Monitoring location B58A0093 is located southeastward to Meijel. The timeseries involves measurements of the average hydraulic head between 30 and 32 meters below surface level, within the local more permeable formation of Beegden. At a depth of approximately 6 to 10 meters below surface level, a local layer of loam can be found. The other part of the subsurface between the surface level and filter bottom mainly consists of sand layers varying from fine to coarse sand. The local land coverage type involves agricultural crops (maize). Measurements at daily base have been taken from 2001 till 2017.

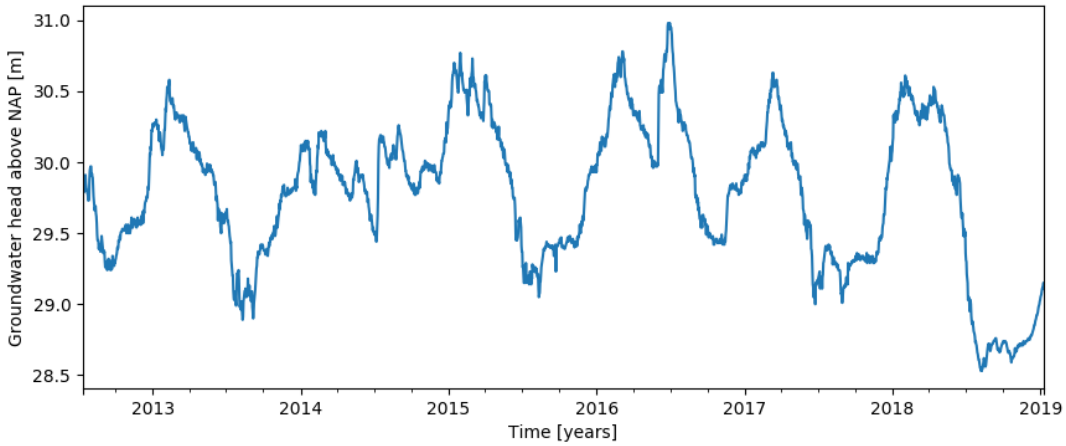


Figure 5: Timeseries of the hydraulic head at monitoring well B58C0005

B52C0005

Monitoring location B52C0005 is located in western direction of Ysselsteyn. The timeseries that where obtained from this location involve measurements of the average hydraulic head within the Kiezelooliet formation between 24,60 and 44,91 meters with respect to surface level. The subsurface mainly consist of sandy layers, but at a depth of approximately 20 up to 25 meters a local clay lens can be found, which lays on top of the aquifer in which the hydraulic head is measured. Further the land coverage type involves deciduous forest. Measurements on daily basis started from 2012 at this location.

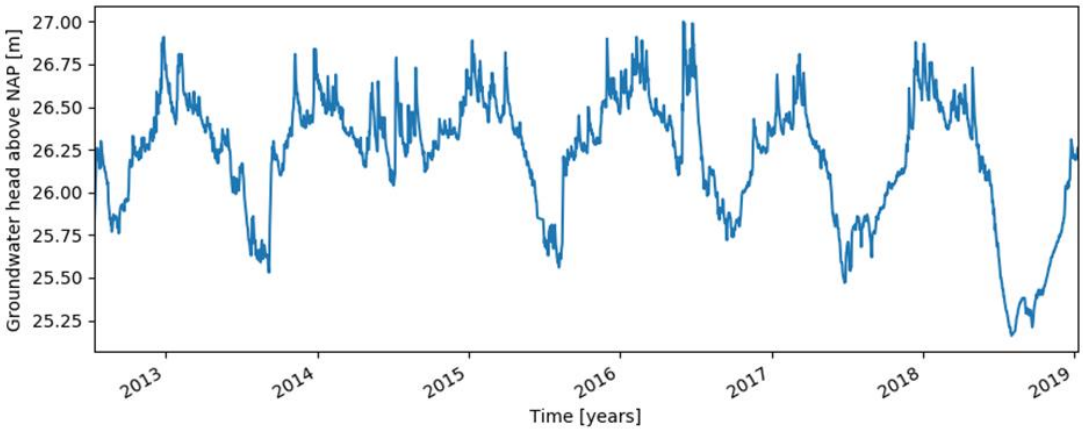


Figure 6: Timeseries of the hydraulic head at monitoring well B58B0260

B58B0260

Monitoring well B58B0260 is located south-eastwards to Maasheeze. The average hydraulic head between a depth of -6,16 to -9,16 meters below surface level is measured (within the formation of Beegden). The subsurface consist of moderately fine sand, and the land coverage type involves grassland. Measurements on daily basis have been taken between 2012 and 2019.

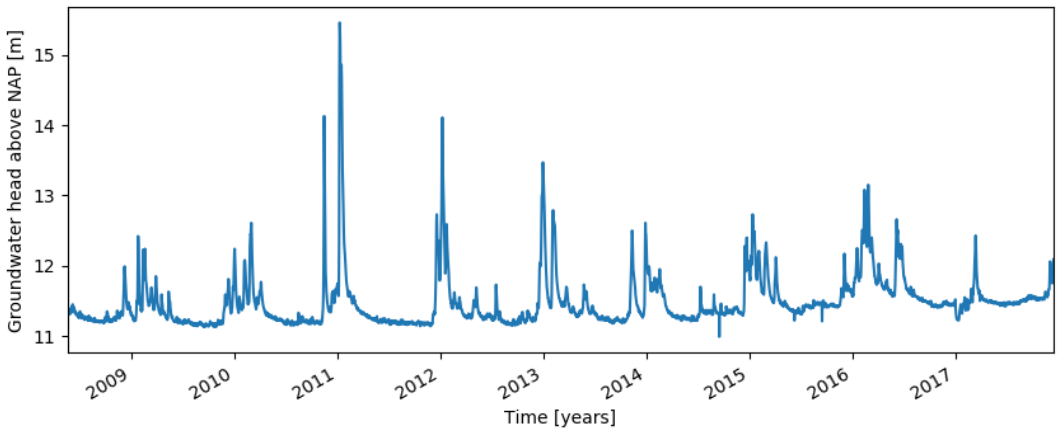


Figure 7: Timeseries of the hydraulic head at monitoring well B52E3231

B52E3231

Monitoring location B52E3231 is located near Wellerlooi nearby the river Meuse (within a distance of 50 meters). The timeseries involve the average measured hydraulic head between a depth of 2,91 to 3,91 meters below surface level in a locally confined aquifer. From the graph it can be seen that the water level in the Meuse has a considerable impact on the hydraulic head (see figure 10 for comparison). Above the sand layer in which the head is measured, a clay layer of a thickness of 2 meters is found (see appendix I). The local land coverage type concerns grassland. Measurements on daily basis have been taken between 2008 and 2017.

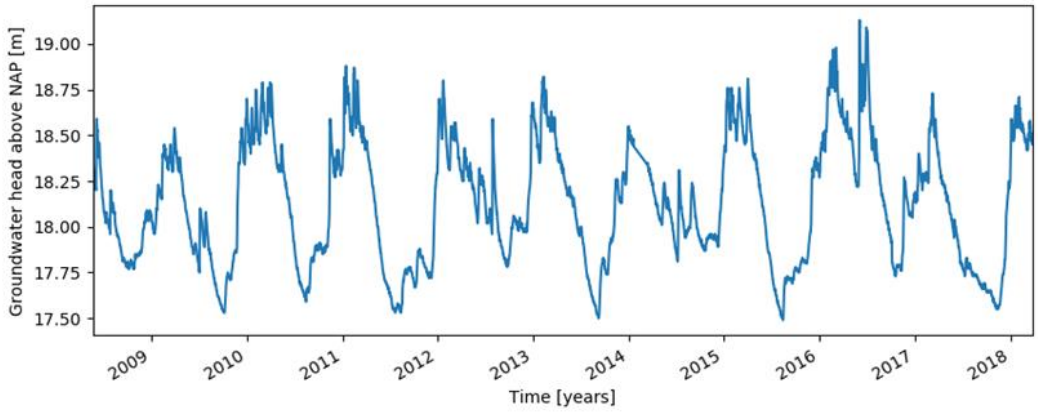


Figure 8: Timeseries of the hydraulic head at monitoring well B52E3234

B52E3234

Monitoring well B52E3234 is located eastwards from Wanssum, at a distance of 2.5 kilometers from the river Meuse. The measured head involves the average hydraulic head at a depth between 2.39 and 3.39 meters below surface level. The local subsurface consists of fine to moderately fine sands from the formation of Boxtel and Beegden, and the local land coverage type concerns deciduous forest. Measurements at a daily base have been taken place between 2008 and 2018. It should be noted that there was a gap in the measurements between January and March 2014

2.2.2 Meteorological data

The collected meteorological data has been collected from the KNMI database, and involves timeseries of precipitation and reference evaporation from four weather stations (Heibloem, Ysselsteyn, Ell and Arcen). No reference evaporation data was available at the weather stations at Ysselsteyn and Heibloem.

Reference evaporation involves the evaporation of a reference crop (hypothetic field of grassland). To obtain a more accurate estimation of the evaporation locally the potential evaporation was calculated, which corresponds with the evaporation of a certain land coverage type under optimal circumstances (Bot, 2011). This was done by multiplying the reference evaporation with crop factors. The crop factors vary for different land coverage types, and vary also for each month. An overview of the used crop factors is included in table 3.

LAND COVERAGE TYPE	LOCATIONS	CROP FACTORS											
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
GRASS	B58C0352 B58B0260 B52E3231	0.9	0.9	0.9	1	1	1	1	0.95	0.9	0.9	0.9	0.9
MAIZE	B58A0093	0.9	0.9	0.6	0.4	0.7	1.05	1.25	1.20	1.20	0.9	0.9	0.9
DECIDUOUS FOREST	B52C0005 B52E3234	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8

Table 3: Crop factors for calculating the potential evaporation at the groundwater monitoring locations (Bot, 2011). The local land coverage types where determined with the LGN7 map from Wageningen University & Research

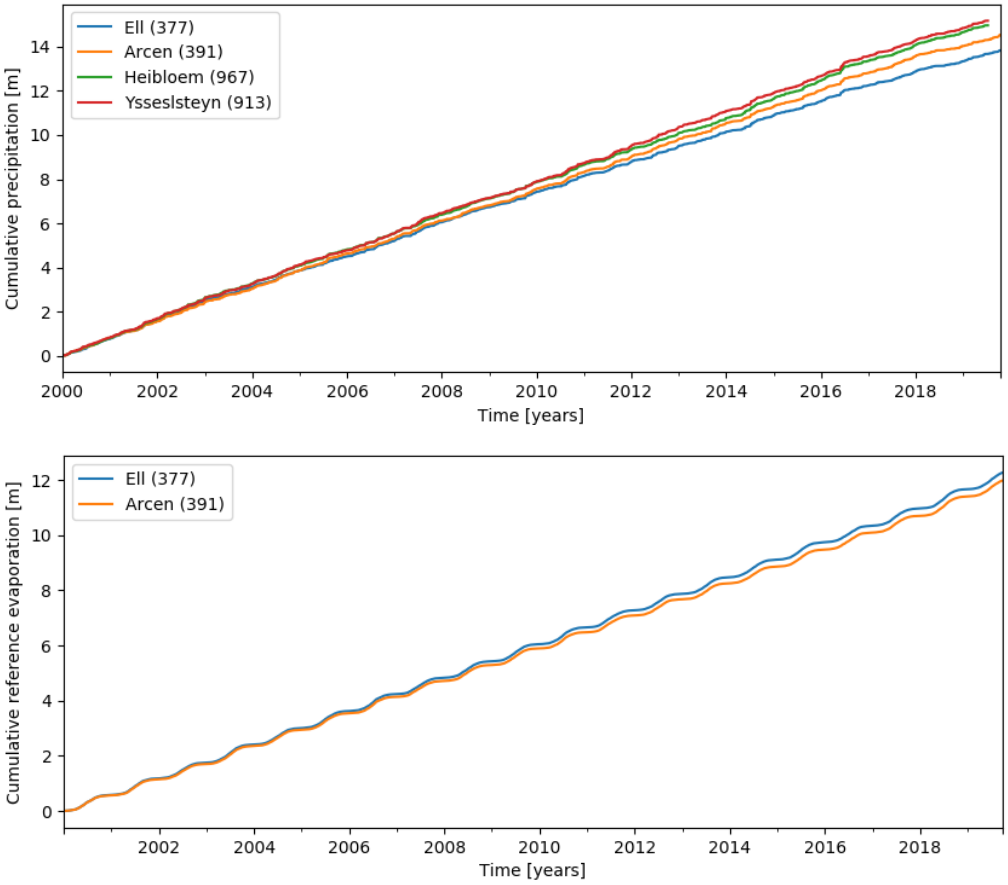


Figure 9: Cumulative precipitation and reference evaporation timeseries from the weather stations in the study area.

2.2.3 *Surface water level timeseries*

Additionally, surface water timeseries were selected as potential explanatory stresses. The timeseries where obtained from the database of Rijkswaterstaat, and they can be divided in water level timeseries and discharge timeseries (figure 10). At three locations, surface water level timeseries where taken (Well, Neer and Belfleld beneden). The water level of the Meuse is regulated by wears. From the graph of the water level timeseries at Well can be concluded that the surface water level has undergone a systematic change between 2014 and 2016. Also, in the winter of 2016-2017 an accident happened: a cargo ship collided with the weir downstream resulting in a temporary drop of the water level locally. Discharge timeseries were only available at the monitoring location in Venlo.

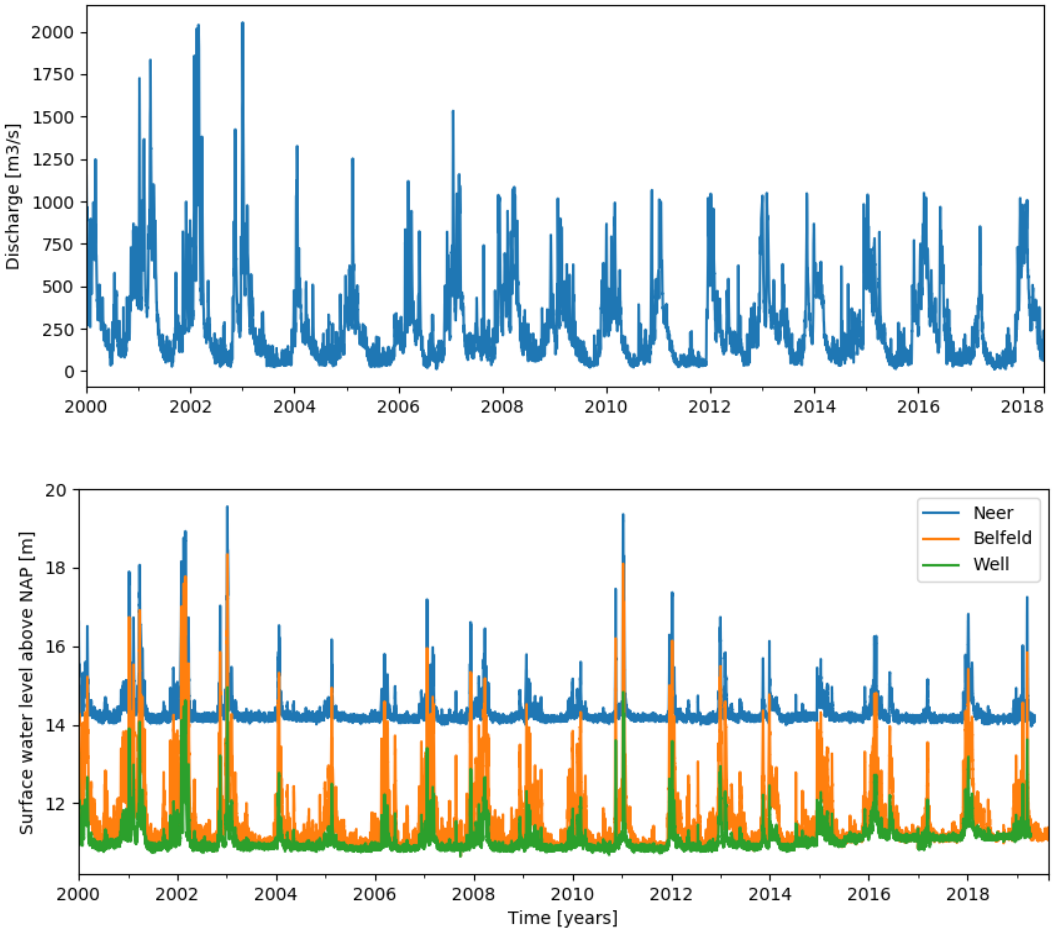


Figure 10: Collected surface water timeseries. Above: discharge timeseries of the Meuse at Venlo. Below: water level timeseries of the Meuse at Well, Neer and Belfeld.

3 Method

In this chapter, a description of the methodology is given. First, the model structure of Pastas is discussed. Afterwards, the created functionality for real-time simulations is explained. At the end, the methodology of the model study is explained.

3.1 Pastas

Pastas (Collenteur, et al., 2019) is a Python package that has been developed for analyzing and simulating hydrological timeseries. The model structure of Pastas is schematized in figure 11. A Pastas model consists of a transfer-noise model which contains multiple components that can be added manually: one or multiple stress models, a noise model, a constant, a transformation and additional features. Commonly, a Pastas model contains at least one stress model. Usually, the model is solved first because the parameter values are not known in advance. When a model is solved, timeseries with observations of the groundwater level (oseries) are used for calibration. Parameter values can also be set manually.

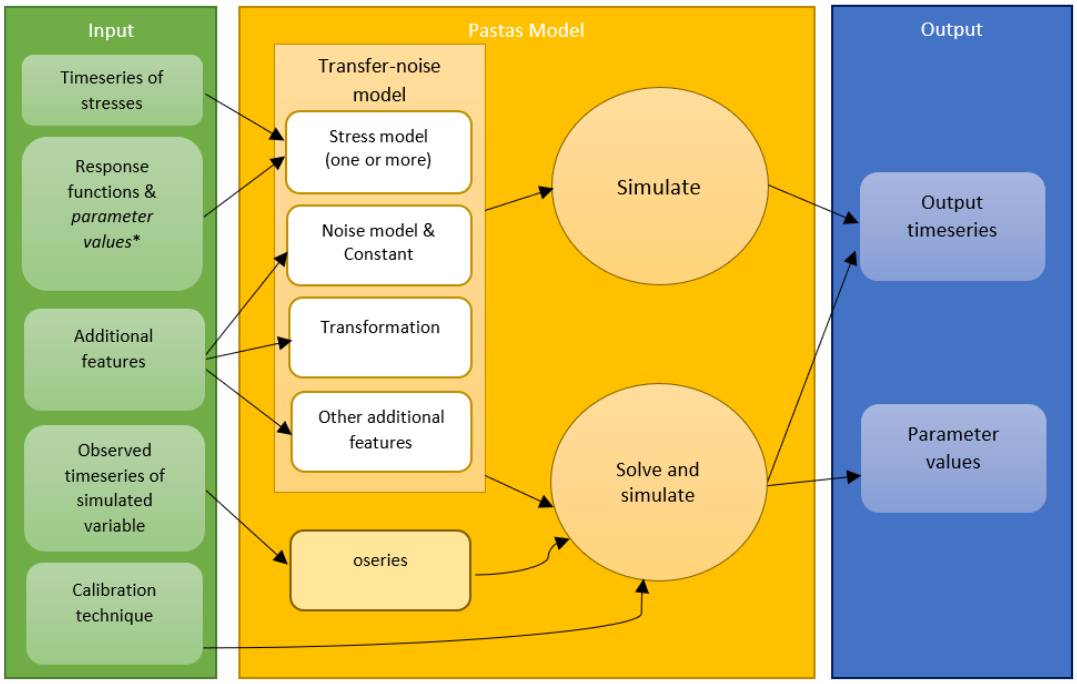


Figure 11: Schematization of PASTAS. Groundwater level timeseries are simulated with a transfer-noise model, which consists of one or multiple stress models, a noise model, a constant and optionally a transformation and additional features. The model is parameterized by calibrating the transfer-noise model on the observational timeseries of the groundwater level. It is also possible to simulate groundwater levels by setting the parameter values manually.

3.1.1 Input data

A Pastas model requires input data for model calibration and simulation. The input data involves timeseries of one or multiple stresses which govern the behavior of the hydrological variable that is simulated. Observed timeseries of the simulated variable (oseries) are also required as input in order to solve the model and set the parameter values.

3.1.2 *Stress model*

For each stress that is included in a Pastas model, a stress model is created. A stress model transforms input timeseries of a stress into a contribution to the simulation. Multiple stress models can be added to a Pastas model, and the output that is individually generated by the stress models is added to obtain the output signal (figure 12). Besides, other elements which contribute to the simulation can optionally be included in a Pastas model (see paragraph 3.3.2, 3.3.3 and 3.3.4).

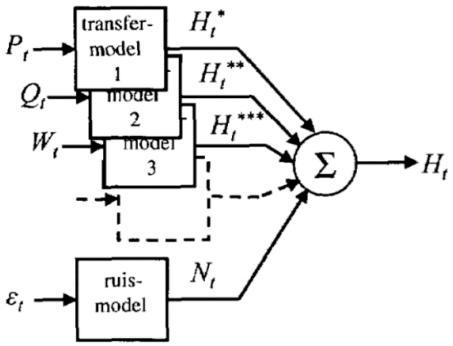


Figure 12: Groundwater timeseries simulation using a transfer noise model (Knotters & Bierkens, 1999a). *Different stresses are handled by different stress models, and the output is summed to obtain the predicted groundwater level. Also, other components such as a noise model can be added.*

For transforming a stress signal into a contribution to the simulation, response functions are used (Collenteur, et al., 2017). These functions contain parameters which have to be calibrated, and the functions generate characteristic curves which represent the step response. After having obtained the step response curve, the derivative is determined which represents the block response curve (figure 13). The block response curve describes the influence of one unity stress at a certain timestep on the state of the simulated variable in the period afterwards. A stress signal at a certain timestep has a temporary influence on the simulated variable in the future, and the period of influence is referred to as the response time (Knotters & Bierkens, 1999b). During this study, the response time is defined as the amount of days in which 95% of the step response has been occurred. The block response curve is used to calculate the temporary reaction of the simulated variable to the stress signals at all timesteps individually. Al these temporary reactions are convolved by means of a fast Fourier transform to obtain the output signal.

A couple of different response functions is available in Pastas at the moment. The four-parameter response function is versatile and can be used for multiple purposes. It can be used for simulating groundwater response to meteorological forcing and other types of forcing. Because it is versatile, it was used for simulating the response of the groundwater level to meteorological forcing and influences of surface water levels and discharges as well during this study. The function is given below:

$$s(t) = A \frac{1}{\int_0^\infty \tau^{n-1} \cdot e^{-\tau/a - b/\tau} d_\tau} \int_0^t \tau^{n-1} \cdot e^{-\tau/a - b/\tau} d_\tau$$

It contains four parameters should be calibrated: A, a, b and n. All the parameters have a specific influence on the shape, which is visualized in figure 13.

3.1.3 *Noise model & Constant*

In order to distinguish noise from real observations, a noise model is commonly added to the simulation. With a noise model, the dynamics of the groundwater head without noise can be assessed. A Pastas noise model is able to detect noise that is completely random, but correlated noise can also be detected.

A constant is also commonly included in the simulation: it is basically a constant value that is added to the simulation. This can be minimum of the observed series for example.

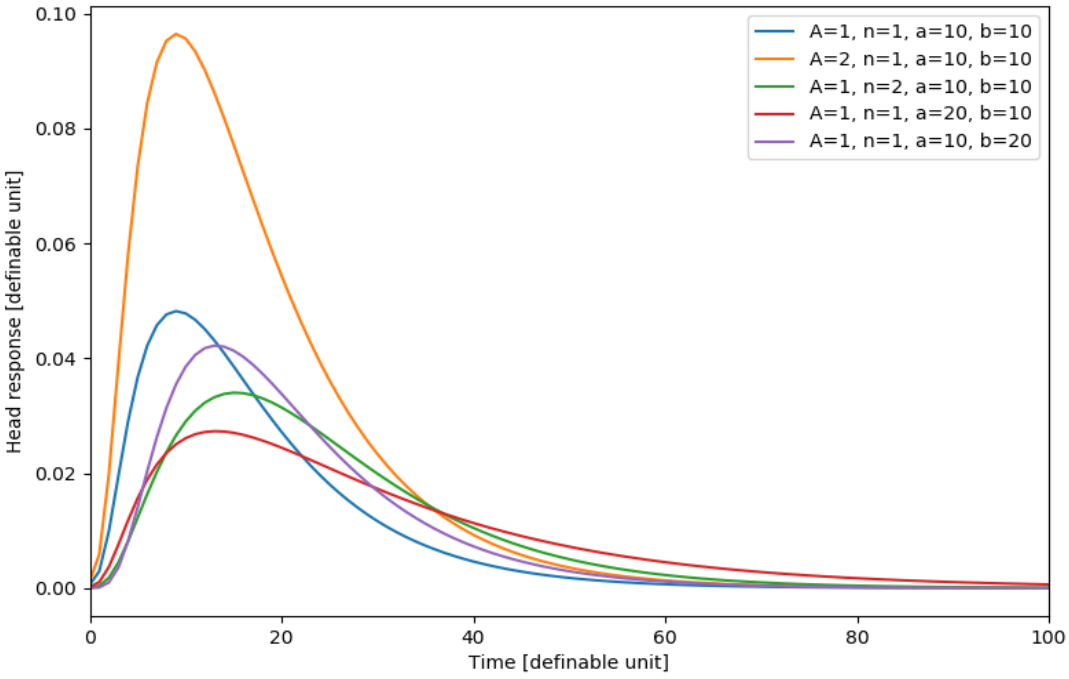


Figure 13: Four Parameter block response curves generated by varying the parameter settings.

3.1.4 *Transformation*

It is also possible to perform a transformation on the simulated output. With a transformation, nonlinear behavior in groundwater dynamics can be simulated for specific cases, involving the slower rise of a groundwater head when it exceeds a drainage level or when surface level is reached. When a transformation is applied, two new parameters can be implemented depending on the situation. One parameter represents a level above which the behavior of the dynamics of the groundwater head is changed, and the other parameter represents the factor of which simulated values above the threshold are multiplied with.

3.1.5 *Additional features*

Several additional features can be added to the simulation: A linear trend, a step trend and a factor. A linear trend can be used for modelling gradual changes in the groundwater regime and a step trend can be used for simulating sudden changes in the groundwater regime. A factor can be implemented to multiply a specific stress with.

3.1.6 *Model solver*

Pastas contains a solver module which can be used for calibrating models on the observed timeseries of the variable that should be simulated. In the module, different solvers are available. The optimization objective is minimizing the residual series, of which the sum of squares is the minimization objective. The least squares solver is the default solver within Pastas. This solver calibrates the model using the least squares function from the Scipy Python package. The default method is a Trust Region Reflective algorithm, which is particularly suitable for large sparse problems with bounds.

3.2 Real-time simulation algorithm

Because there was no specific application available yet for performing real-time forecasts with Pastas, a real-time simulation class was developed in Python during this study. The real-time simulation class can be used to govern a Pastas model. It has the ability to update Pastas models with recent data. Also, a functionality is included for performing forecasts. An overview of the class and its functionalities and variables is given in table 4.

Real time simulation project class

Functions	Description
Init	Initializes the real time simulation project. As input, a PASTAS model and a name for the project should be given.
Update	By calling this function the model is updated. Three input arguments are required: - A dataframe containing newly observed timeseries of the stresses - A dataframe containing newly observed timeseries of the output variable - The length of the calibration period The new data is appended to the historical data of the model, and the historical dataset is cropped to the specified length. Afterwards the model is calibrated with the updated dataset to obtain the new parameter values.
Forecast	By calling this function, a forecast of the simulated variable is performed. As input, a dataframe should be handed in that contains timeseries representing future values of the stresses. These timeseries could involve forecasts, but they could also involve observed data. The function commands the Pastas model to perform a simulation with the given input data and the most recently determined parameter values. A data frame with predictions of the output variable is returned afterwards. The length of the forecast is inferred by the length of the input data frame
Variables	**Description**
Name	Contains the name of the model
Model	Contains the PASTAS model
Stress model names	Vector with the names of all the stress models that are included in the PASTAS model
Calibration window size	Size of the calibration period
Parameters	Data frame containing the parameters of the PASTAS model and their values
Stress model types	Vector that contains the types of the stress models that are included in the PASTAS model

Table 4: Overview of the functionalities of the developped script for performing real-time forecasts with Pastas models. It concerns a class in which a Pastas model can be integrated. The class can be used for updating the model and for performing forecasts.

3.3 Model study

The model study was performed in two phases. In the first phase, it was explored which stress timeseries contribute to the groundwater dynamics locally. In the second phase, real-time hindcasts of the groundwater levels where performed using the developed script (section 3.2).

3.3.1 Data selection and performance assessment

All selected groundwater timeseries (chapter 2) were split up in two parts. The first part involves historical data, which was used during the first phase of this study. The historical data was also used in the second phase of this study in the calibration process for the real-time hindcasts. The remaining part of the timeseries involves the last year of available data and was used for validation of hindcasting performance during the second phase. All timeseries of the stresses were split up in the same way. Only for the historical data of stresses, a year of data in advance of the start of the historical groundwater timeseries was selected for model warming up.

For judging model performance in simulating and forecasting groundwater levels, two performance measures where used: the root mean squared error (RMSE) and the mean average error (MAE). RMSE measures the goodness of fit relevant to high errors and MAE measures the goodness of fit while judging all errors equally (Wang, et al., 2018). Additionally, during the setup of the models in the first phase r-squared (R^2) was used to measure the extent up to which the variability of the predicted variable was explained by the model and the Akaike information criteria (AIC) was used to judge model performance relative to model complexity. With AIC it is possible to verify the contribution of adding extra parameters to the model relative to the increase in model performance (Oberfell, et al., 2019). The ultimate values for all performance criteria would be: R2=1, RMSE=0, MAE=0 and AIC = 0. The formulas of the different performance criteria are given below.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(h_i^o - h_i^p)^2}{\sum_{i=1}^{n}(h_i^o - \bar{h}^o)^2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(h_i^p - h^o)^2}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|h_i^p - h^o\right|$$

$$\mathrm{AIC} = (2k + 1) + N\log(S/N)$$

In the formulas for calculating R2, RMSE and MAE, h_i^o and h_i^p are the observed and predicted groundwater levels at the ith timestep, $\bar{h}^o$ is the mean observed groundwater level and *n* is the number of observations. In the formula for calculating AIC, k is the number of parameters, *N* is the number of data points and *S* is the sum of squared residuals (Wang, et al., 2018; Oberfell, et al., 2019).

3.3.2 *Phase 1: explaining the groundwater level dynamics locally*

In order to determine the relevance of adding different stresses as input for simulating the groundwater levels locally, different model setups were tested. An overview is included in table 5. The performance of different model setups in predicting the groundwater levels was verified in three rounds:

- Round1: During the first round, the stresses that were expected to mainly govern the groundwater dynamics at each location where used as input and the performance in simulating the groundwater level was determined afterwards. Also, the performance of models using different ways of implementing these stresses was explored.
- Round 2: In round 2, the best performing model setup of round 1 was combined with additional stresses.
- Round 3: In round 3, the best performing setup of the first two rounds was taken and combined with a transform and a linear trend respectively.

At the end, the best performing model setups out of all three rounds were selected for each location to perform hindcasts with during the next phase of this study.

Round	Setup	B58C0352	B58A0093	B52C0005	B58B0260	B52E3231	B52E3234
1	*1*	P & ET_{pot}	P & ET_{pot}	P & ET_{pot}	P & ET_{pot}	h_{Meuse}	P & ET_{pot}
	2	R	R	R	R	Q_{Meuse}	R
	3	-	-	-	-	h_{Meuse} & Q_{Meuse}	-
2	1	Best of round 1 & h_{Meuse}	Best of round 1 & h_{Meuse}	Best of round 1 & h_{Meuse}	Best of round 1 & h_{Meuse}	Best of round 1 & P & ET_{pot}	Best of round 1 & h_{Meuse}
	2	Best of round 1 & Q_{Meuse}	Best of round 1 & Q_{Meuse}	Best of round 1 & Q_{Meuse}	Best of round 1 & Q_{Meuse}	Best of round 1 & R	Best of round 1 & Q_{Meuse}
3	1	Best out of first two rounds & transform	Best out of first two rounds & transform	Best out of first two rounds & transform	Best out of first two rounds & transform	Best out of first two rounds & transform	Best out of first two rounds & transform
	2	Best out of first two rounds & linear trend	Best out of first two rounds & linear trend	Best out of first two rounds & linear trend	Best out of first two rounds & linear trend	Best out of first two rounds & linear trend	Best out of first two rounds & linear trend

Table 5: Overview of the different model setups that were tested. The different stress models included in each model setup are indicated with their stresses: P is the precipitation, ETpot is the potential evaporation, R is the precipitation excess (P- ETpot), hMeuse is the water level of the Meuse and QMeuse is the discharge of the Meuse. The tests were performed in three rounds. In round 2 and 3, additional stress models were added to the best model setup in the previous rounds.

3.3.3 *Phase 2: performing hindcasts*

After having assessed which stresses mainly contributed to the groundwater level dynamics locally, the resulting models were tested for real-time application. This was done by performing hindcasts: groundwater levels in the year with the most recent observations were forecasted using parameter values obtained from calibration on historical data. The setup of the hindcasts is schematized in figure 14 and can be explained as follows:

- The initial models were calibrated with the historical datasets that were used in the first phase of the model study. With the resulting parameter settings, forecasts of the groundwater level were performed for periods of 30 and 120 days ahead. Observational data of the stresses from the dataset of the most recent year of observations (section 3.3.1) were used as input for the forecasts.
- Subsequently, the models were updated with two weeks of observational data and were re-calibrated with the updated dataset. New forecasts were performed afterwards.
- This process was repeated 14 times: until the end of the datasets with the year of most recent observations was reached for the 120 days-ahead forecasts.

The performance of the hindcasts was analyzed afterwards with timeseries of the observed groundwater level. The influence of varying the length of the historical calibration dataset on the performance of groundwater level forecasts was explored by running multiple hindcast scenarios for each location in which calibration periods with different lengths were used. A schematic overview is included in table 6.

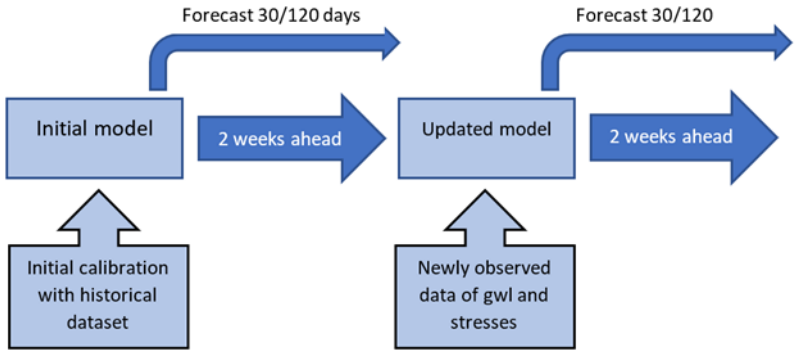


Figure 14: Schematization of the real-time groundwater level simulations. Every two weeks, forecasts of the groundwater level were generated for a period of 30 and 120 days ahead based on future data of the stresses. The model was updated afterwards with new observational series and recalibrated.

Scenario	Length of calibration period					
	B58C0352	**B58A0093**	**B52C0005**	**B58B0260**	**B52E3231**	**B52E3234**
1	Entire dataset (15 years)	Entire dataset (15 years)	Entire dataset (6 years)	Entire dataset (6 years)	Entire dataset (8 years)	Entire dataset (9 years)
2	10 years	10 years	3 years	3 years	3 years	3 years
3	5 years	5 years	2 years	2 years	2 years	2 years
4	3 years	3 years	1 year	1 year	1 year	1 year
5	2 years	2 years	0.5 year	0.5 year	0.5 year	0.5 year
6	1 year	1 year	-	-	-	-
7	0.5 year	0.5 year	-	-	-	-

Table 6: Overview of the real-time hindcast scenarios for each location

4 Results

In this chapter, the results of the model study are discussed separately for each location. First, the procedure in selecting the best model is discussed (phase 1). Afterwards, the performance of the best model resulting from the three rounds of the selection procedure is discussed (phase 1). At the end, the performance of the hindcasts with the best model is discussed (phase 2). The observations will be discussed further in Chapter 5.

4.1 B58C0352

4.1.1 Model Setup

The dynamics of the groundwater level at location B58C0352 can mainly be explained by the precipitation excess (see appendix II). Adding the water level and discharge of the Meuse as stresses did not contribute to an improvement in predictive performance, indicating that the Meuse has no influence on the fluctuations of the local groundwater level. Adding a linear trend and transformation to the simulation did also not improve the performance, indicating the absence of influences of drainage or a linear trend on the groundwater level dynamics.

The resulting predictions of the Pastas model that simulated the groundwater level based on the precipitation excess are shown in figure 15. The model performs well in simulating the annual groundwater level fluctuations (R^2 = 0.91, RMSE = 0.13m, MAE = 0.1m). However, in winter and summer time some peaks and troughs in the groundwater level are missed. During summertime it is likely that groundwater abstractions for agricultural use purposes occur in the proximity, resulting in an additional lowering of the groundwater level. The response time of the system was approximately around one year.

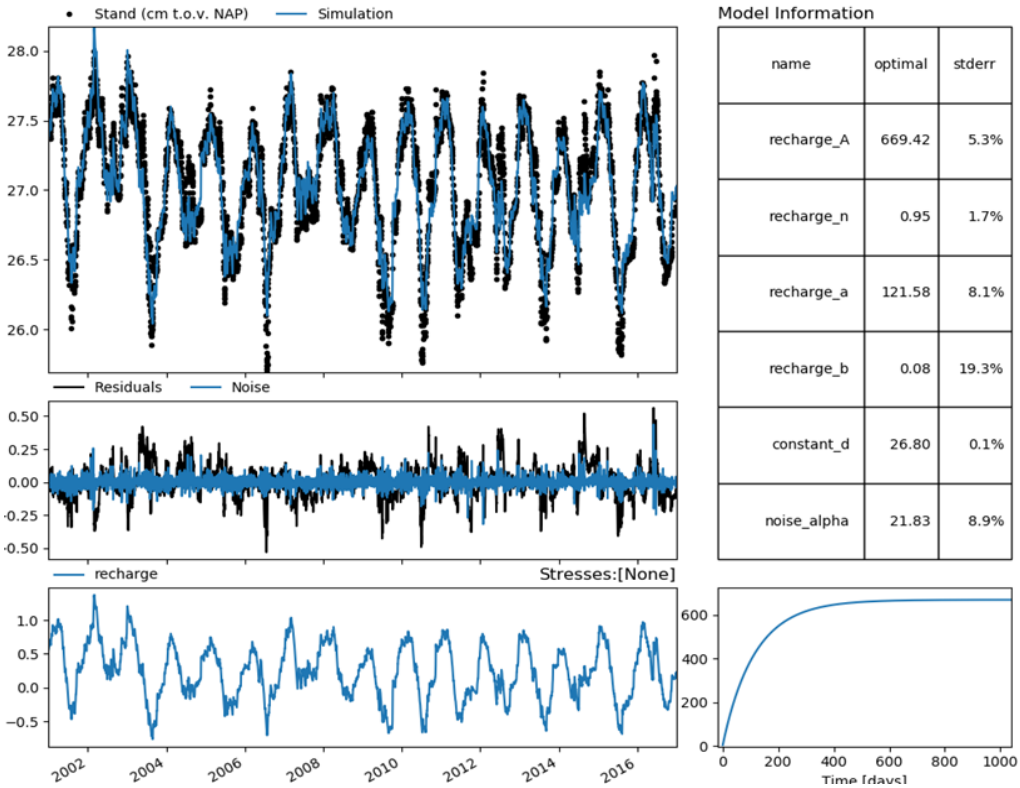


Figure 15: Simulations of the groundwater level at location B58C0352 by implementing the precipitation excess timeseries as stress.

4.1.2 *Hindcasts*

The performance of the hindcasts is shown in figure 16. The hindcasted groundwater level deviated averagely between 0 and 0.25 meters with respect to the observed groundwater level (RMSE and MAE). Firstly, the average deviations for the 120-day hindcasts seem to be higher than the average deviations of the 30 days-ahead hindcasts. Since the last hindcasts were performed in August 2017, only the 120 days-ahead hindcasts cover the last three months of that year. During that period, the hindcasts started to deviate more from the observed groundwater level. This caused the 120 days-ahead hindcasts to have a larger average deviation than the 30 days-ahead hindcasts.

For the 30 days-ahead hindcasts, no clear pattern was found in the relation between the length of the calibration period and the performance of the hindcasts. For the 120 days-ahead hindcasts, it seems that the overall performance increased a little when shorter calibration periods down to two years were used. When using shorter calibration periods, the hindcasted groundwater level during the last part of 2017 was lower and therefore biases with the observed groundwater level also decreased (see Appendix III). This observation will be discussed in chapter 5. Further, it was observed that groundwater level hindcasts using a calibration period of one year were less accurate and precise in general: they were visually also more unstable (see Appendix III). This will also be discussed in Chapter 5.

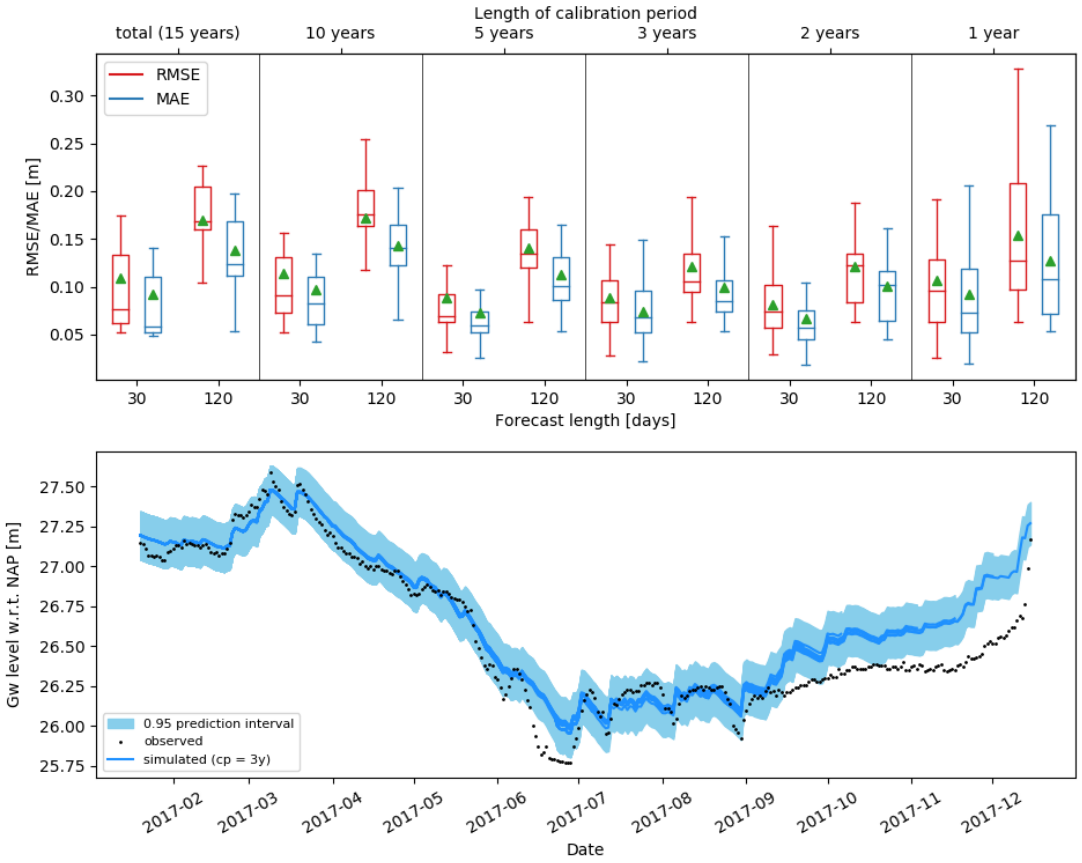


Figure 16: Above: performance of 30 and 120 days-ahead hindcasts for location B58C0352 using varying calibration period lengths. The boxplots represent the spread in RMSE and MAE for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B58C0352 using a calibration period of three years. Every two weeks, the model was recalibrated and a new hindcast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.

In figure 16, the 120 days-ahead groundwater level hindcasts using a calibration period of three years are displayed as well. In the spring of 2016, the observed groundwater level is mostly within the prediction interval. However, during summertime and autumn there are some larger deviations occurred.

4.2 B58A0093

4.2.1 Model setup

The performance of the different model setups at monitoring location B58A0093 were similar to the results at location B58C0352 (appendix II). The local groundwater level could be explained well by implementing the precipitation excess as a stress model. The predictions of the resulting Pastas model are shown in figure 17. The model also was able to simulate the observed groundwater level quite well (R^2 = 0.87, RMSE = 0.15m, MAE = 0.12m). Similar to the simulations of the groundwater level at location B58C0352, peaks and troughs in the groundwater level are missed sometimes. During summertime it is likely that groundwater abstractions for agricultural use occur in the proximity, resulting in an extra lowering of the groundwater level. The response time of the system seems to be approximately one year.

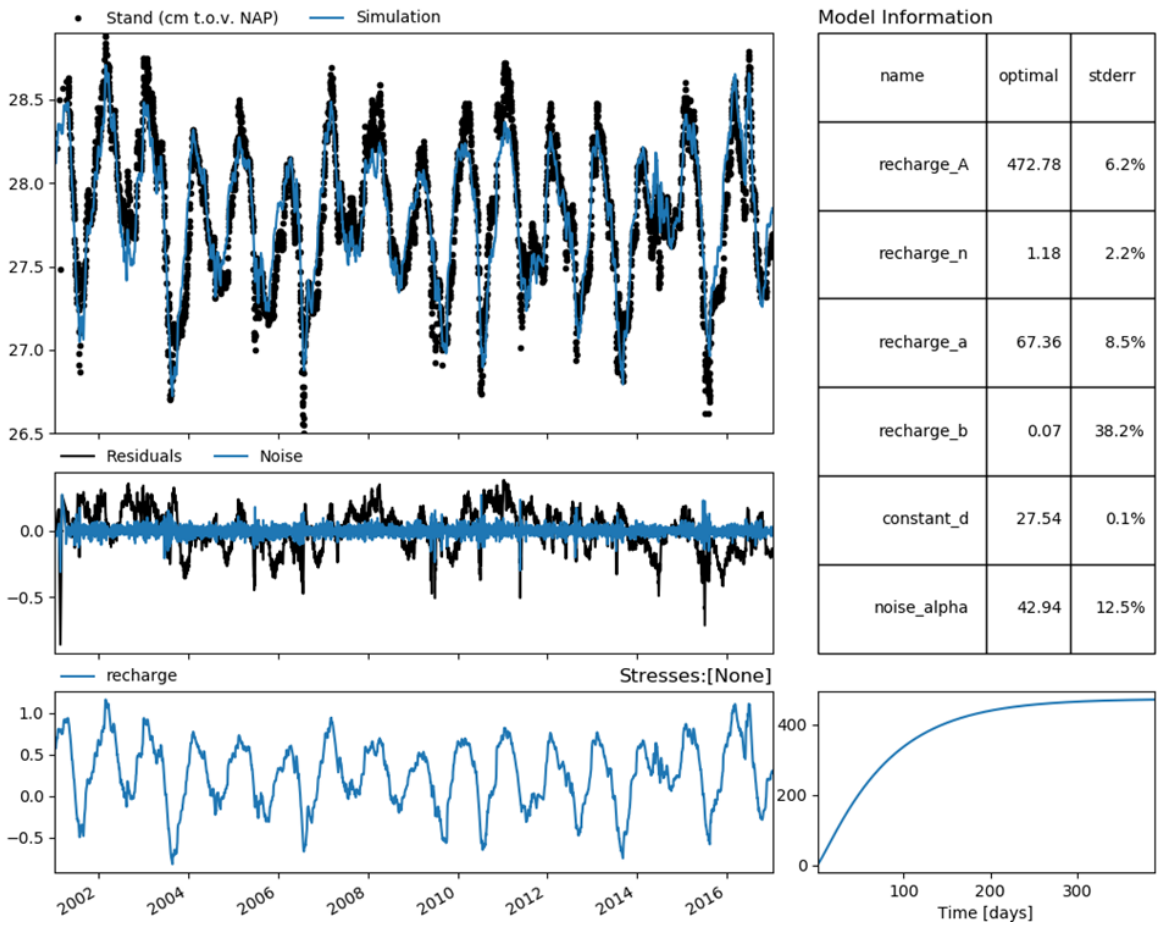


Figure 17: Simulations of the groundwater level using the precipitation excess timeseries as input with one single stress model for location B58A0093.

4.2.2 Hindcasts

The performance of groundwater level hindcasts using different calibration period lengths is shown in figure 18. The RMSE and MAE seems to range in between 0.3 and 0 meter. Similar to the results at location

B58C0352, it seems that the average deviations of the 120-days ahead hindcasts are larger. This is also because only the 120 days ahead hindcasts overlap the last three months of 2017, in which the deviation between the observed and hindcasted groundwater level was relatively large.

Also, it seems that the average performance of the hindcasted groundwater level increases when shorter calibration periods are used (especially for the 120 days-ahead hindcasts): the hindcasted groundwater level during especially the last part of 2017 decreased when the length of the calibration period decreased in general. Hindcasts in which a calibration period of one year is used seem to be the most unstable visually (Appendix III).

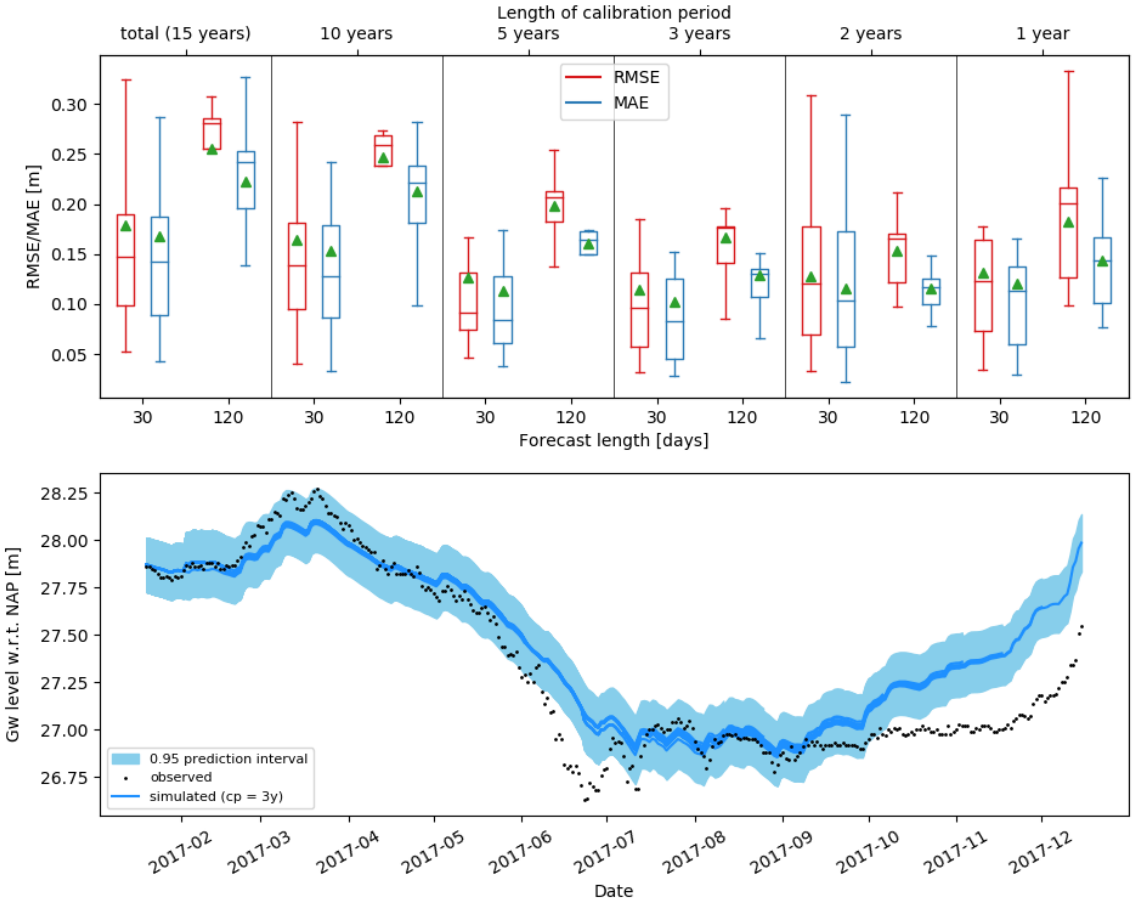


Figure 18: Above: performance of 30 and 120 days-ahead hindcasts using varying calibration period lengths for location B58A0093. The boxplots represent the spread in RMSE and MAE for the hindcasts for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B58A0093 using a calibration period of three years. Every two weeks, the model was recalibrated and a new hindcast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.

In figure 18, the 120 days-ahead groundwater level hindcasts with a calibration period of 3 years are also included. The hindcasting performance seems to be quite similar to the hindcasting performance at location B58C0352. Observed groundwater levels in the start of 2017 are within the prediction interval, but during summertime and autumn there are some larger deviations. The deviation during summertime is possibly caused by groundwater abstractions. The deviation during autumn will be discussed further in Chapter 5.

4.3 B52C0005

4.3.1 Model setup

At location B52C0005, the performance of the different model setups in describing the groundwater dynamics was quite similar (appendix II). Although the model setup in which precipitation and potential evaporation were implemented as separate stress models performed a little bit better, it should be noted that the model complexity did not contribute sufficiently to model performance compared with the model in which the precipitation excess was implemented as a single stress model. Therefore, it was chosen to use the model setup with the precipitation excess as only as best performing model in the first round. In the second round, it was remarkable that the inclusion of the water level of the Meuse did not lead to increased model performance and that the inclusion of the discharge of the Meuse did lead to increased model performance. When taking the AIC values into account, the increase in model performance was not significant compared to the increased model complexity. Therefore, it was decided to exclude timeseries of the Meuse as explanatory variable. The model with only the precipitation excess as input was selected as best model after all. Groundwater predictions with the resulting model seemed to be quite accurate (R^2 = 0.86, RMSE = 0.16, MAE = 0.13). The response time of the systems seems to be around two years.

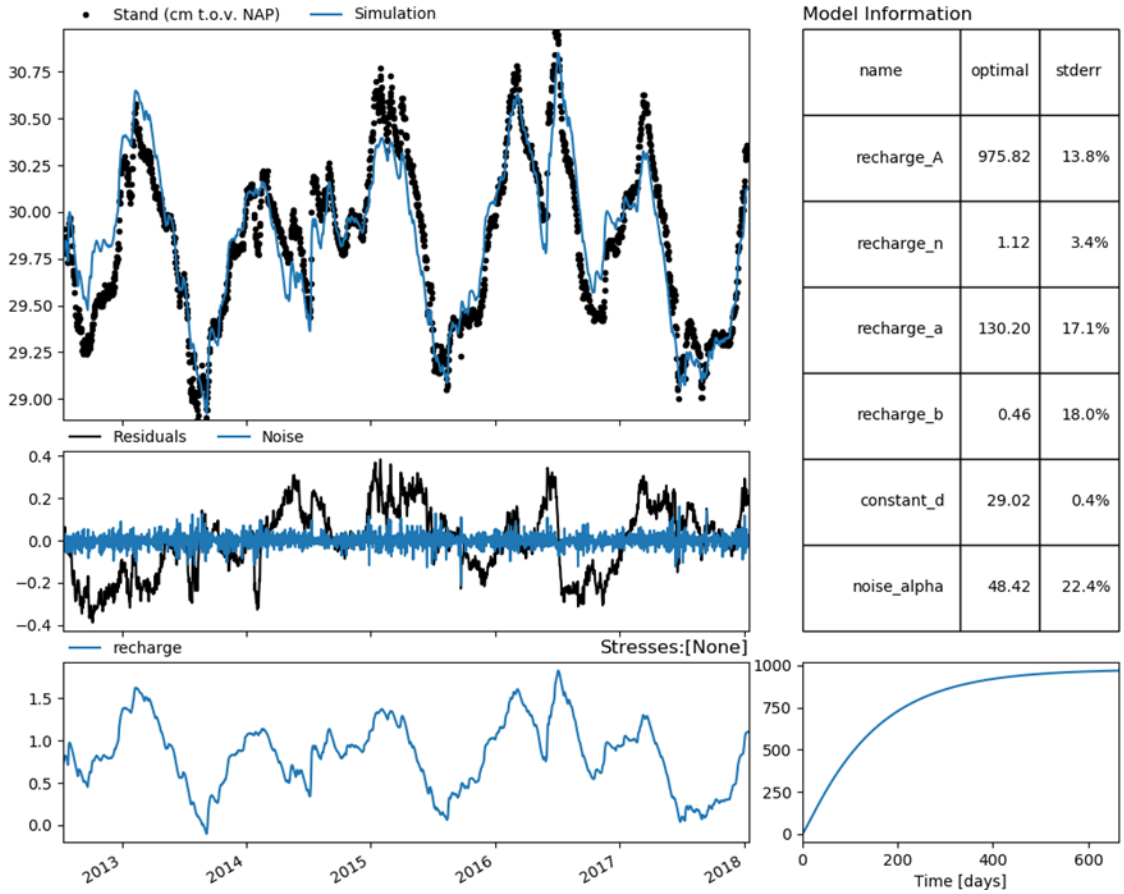


Figure 19: Simulations of the groundwater level for location B52C0005 using the precipitation excess timeseries as input with one single stress model.

4.3.2 *Hindcasts*

There was no large difference in the performance of the hindcasts generated with calibration periods of several lengths (figure 20). The 30 days-ahead hindcasts seemed to be most accurate when a calibration period of one year was used, but in contrary the 120 days-ahead hindcasts with the same calibration period were the worst. The plot of the different hindcasting scenarios in appendix III shows that on the long term (120 days) the hindcasts are more unstable when using a calibration period of one year: especially during summertime the groundwater level was not captured well. When longer calibration periods were used, the 120 days-ahead hindcasts seemed to be more stable.

In figure 20, the 120-days ahead groundwater level hindcasts during which the entire set of historical data was used for calibration are also shown. It seems that the lowering of the groundwater level during summer is captured quite well. This is remarkable, since the observed groundwater level in the summer of 2018 was very low compared to the previous years. However, especially during the spring of 2018 and December 2018, the observations deviate relatively a lot with the hindcasted groundwater level (see chapter 5).

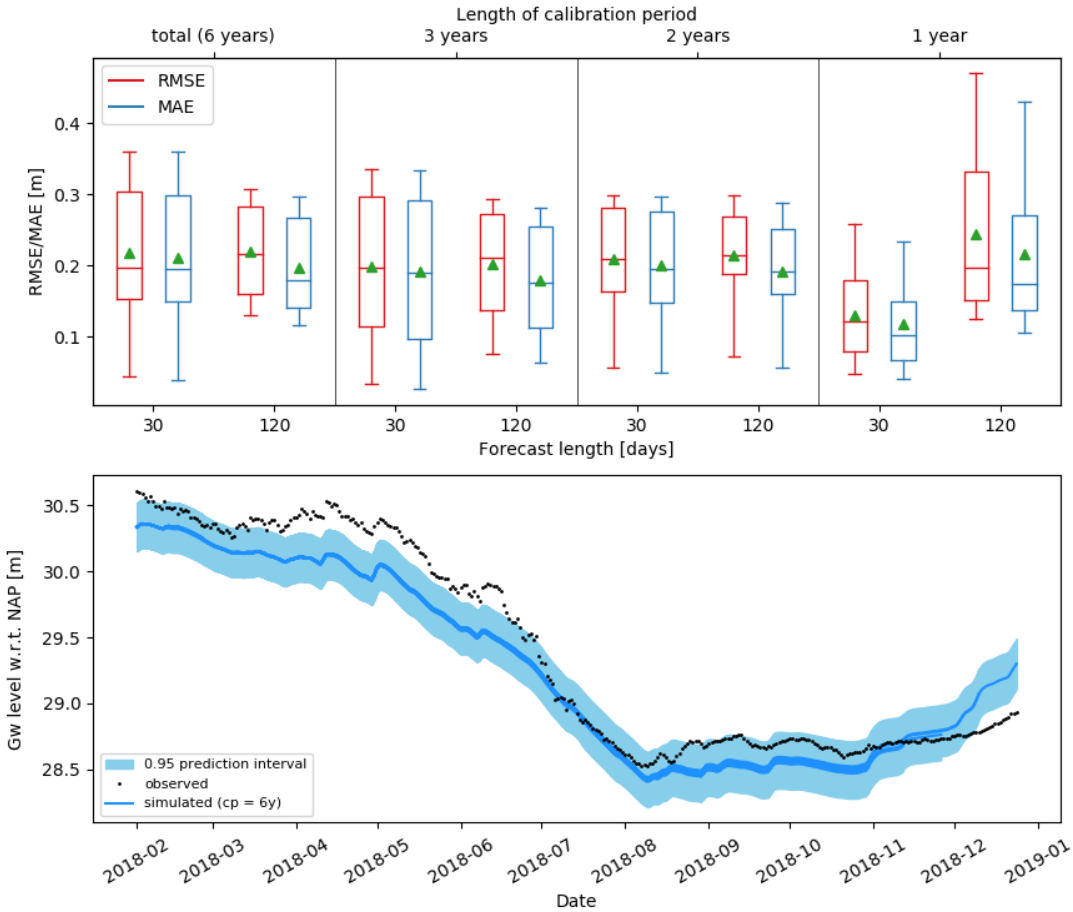


Figure 20: *Above: performance of 30 and 120-days ahead hindcasts using varying calibration period lengths for location B52C0005. The boxplots represent the spread in RMSE and MAE for the hindcasts for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B52C0005 using the whole set of historical data for calibration. Every two weeks, the model was recalibrated and a new hindcast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.*

4.4 B58B0260

4.4.1 Model setup

For location B58B0260, the precipitation excess seemed also to be most suitable for explaining the local groundwater dynamics. The resulting model performed relatively well and had a relatively low bias (Appendix II). Adding the other stresses and input variables to the model did not lead to a significant increase in predictive performance. The simulations of the groundwater level using the precipitation excess as input are shown in figure 21. During wintertime and summertime, some peak values do not seem to be explained very well. Further it is visible that the response time is roughly around a year. The model performed quite well however (R^2 = 0.9, RMSE = 0.1, MAE = 0.08).

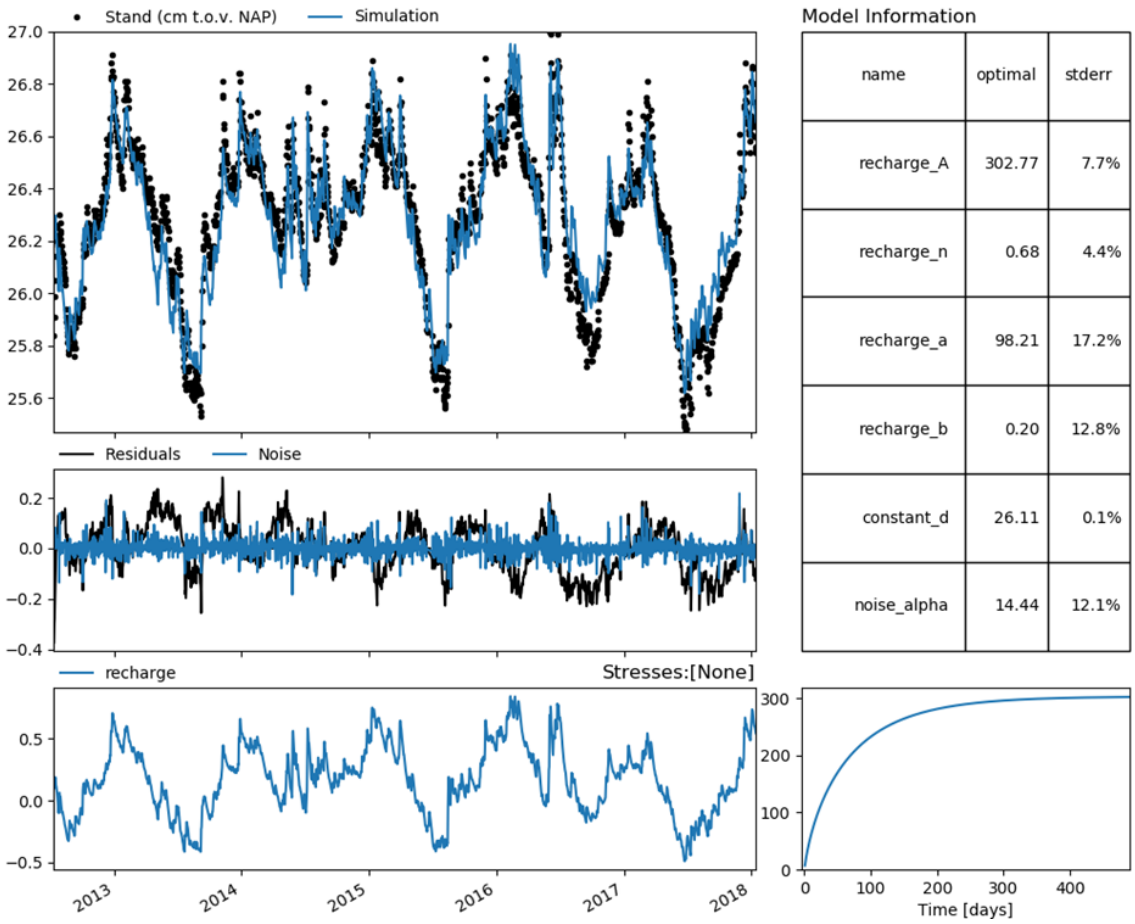


Figure 21: Simulations of the groundwater level for location B58B0260 using the precipitation excess timeseries as input with one single stress model.

4.4.2 Hindcasts

The boxplots of the performance of the hindcasts (figure 22) show that for especially the 120 days-ahead hindcasts the performance increased when the length of the calibration period decreased down to two years. Similar to the results at location B58C0352 and B58A0093, the deviations of the 120-day ahead hindcasts were larger: the last hindcasts were performed in September 2019 and only the 120-day hindcast

cover the last part of 2018, in which biases between the hindcasts and observations were relatively high (appendix III). The hindcasts using a calibration period of one year were very unstable (appendix III). In figure 22, the 120-days ahead hindcasts using a calibration period of two years are also plotted. It seems that the observed groundwater level is hindcasted quite well, although sometimes there were some larger deviations. It seems that there might be some patterns in groundwater response to recharge which cannot be explained by the model, causing the sub-annual deviations (see chapter 5). Especially during summertime and autumn, the forecasted groundwater level seemed to respond more strongly to precipitation than the observed groundwater level. Further, the peaks in the groundwater level during the spring of 2018 were underestimated.

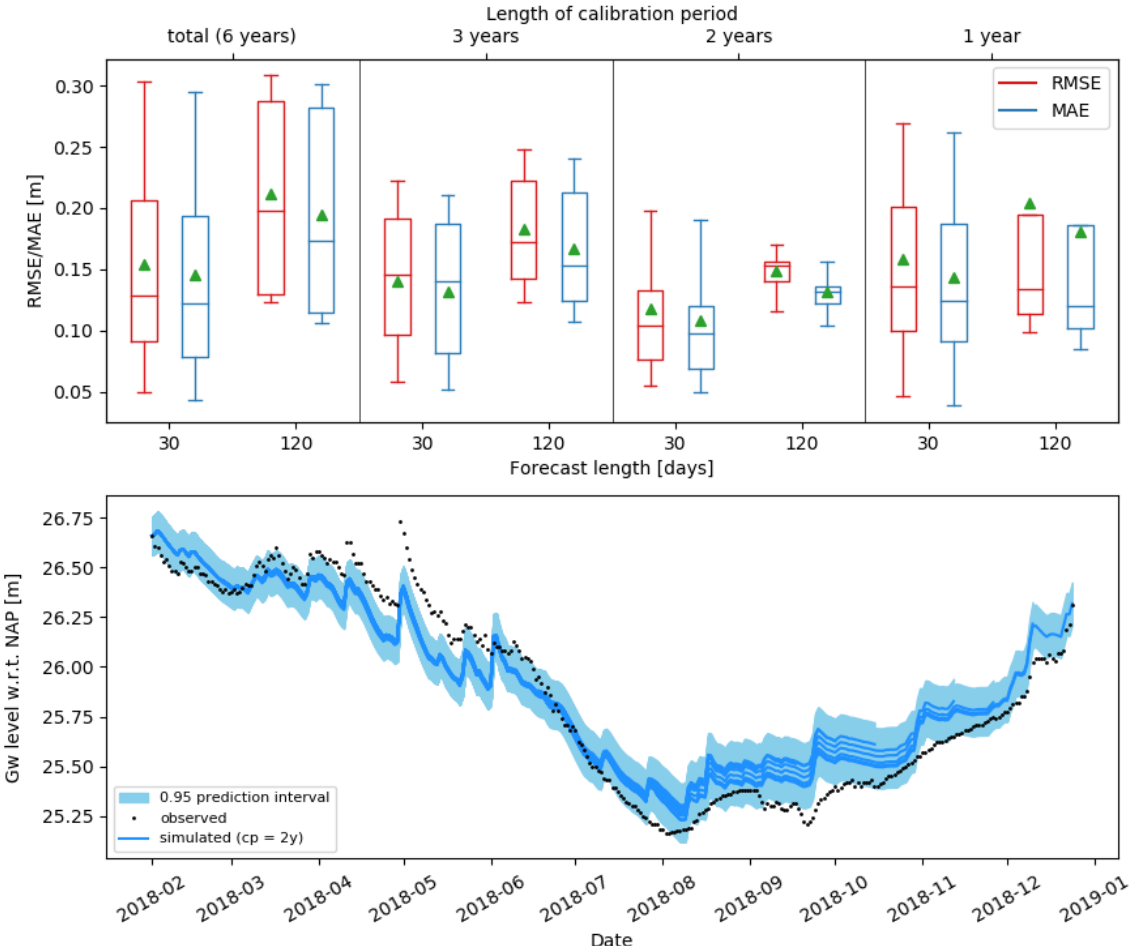


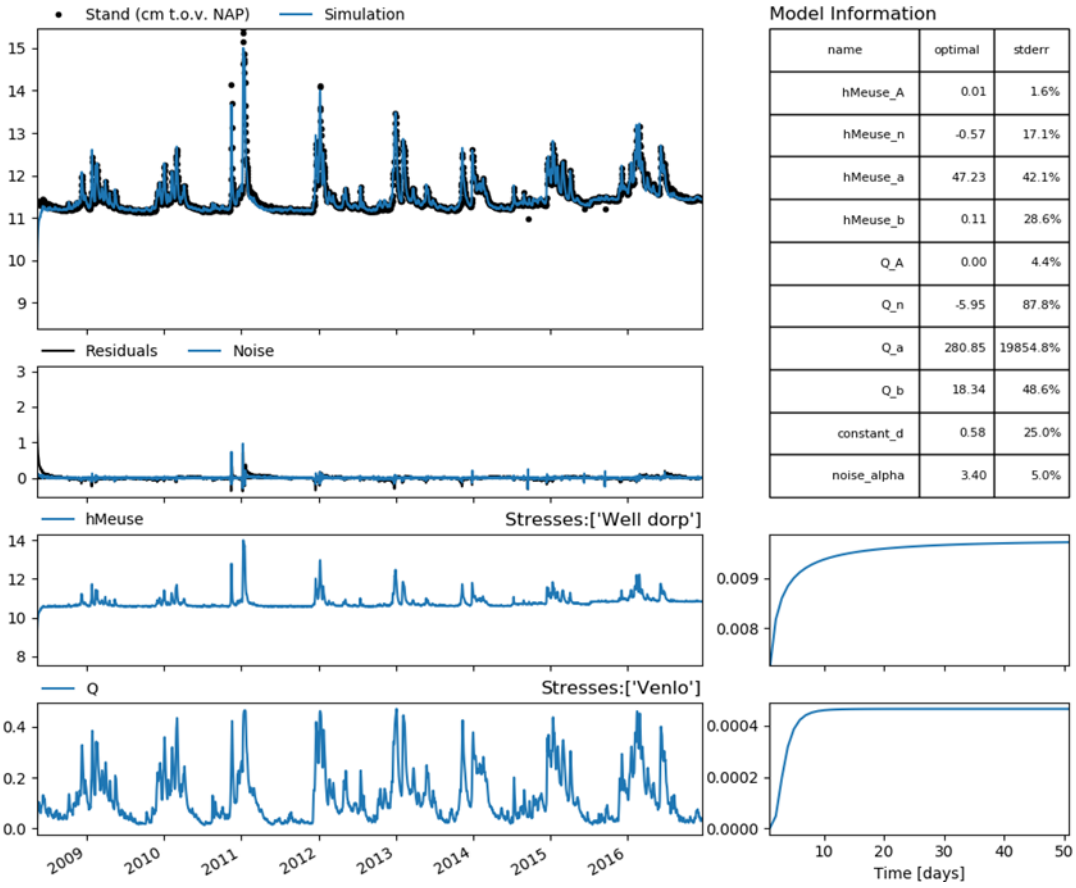
Figure 22: Above: performance of 30 and 120 days-ahead hindcasts using varying calibration period lengths for location B58B0260. The boxplots represent the spread in RMSE and MAE for the hindcasts for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B58B0260 using a calibration period of two years. Every two weeks, the model was recalibrated and a new hindcast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.

4.5 B52E3231

4.5.1 Model setup

For location B52E3231, the surface water level timeseries were used as primary input variables to build the model on, since the groundwater level timeseries at this location visually seemed to be similar to those timeseries. It was concluded that the groundwater level timeseries could not only be explained by the surface water level timeseries however when looked at the performance (appendix II). The model that used the water level and discharge of the Meuse as input variables seemed to be the best option, looked at the performance.

The simulation using the water level and discharge of the Meuse as stresses is also shown in figure 23. It seems that the groundwater level is simulated quite well (R^2 = 0.95, RMSE = 0.09m, MAE = 0.05m). Also, it seems that the structural change in the regime of the surface water level of the Meuse during the simulation period has also been captured well (since it was also visible in the groundwater level timeseries).



name	optimal	stderr
hMeuse_A	0.01	1.6%
hMeuse_n	-0.57	17.1%
hMeuse_a	47.23	42.1%
hMeuse_b	0.11	28.6%
Q_A	0.00	4.4%
Q_n	-5.95	87.8%
Q_a	280.85	19854.8%
Q_b	18.34	48.6%
constant_d	0.58	25.0%
noise_alpha	3.40	5.0%

Figure 23: Simulations of the groundwater level at location B52E3231 using the surface water level and discharge of the Meuse as input with one single stress model.

4.5.2 *Hindcasts*

The performance of the hindcasts was quite well in general. The RMSE and MAE were both in between 0.1m and zero meters for all the scenarios. The performance decreased a little bit when the calibration period decreased for as well the 30 days-ahead as the 120 days ahead hindcasts. The performance of longer calibration periods was slightly better. Especially the water level predictions with calibration periods of one and two years seemed to be more unstable (appendix III). This could be caused because the structural change in the surface water regime was not fully covered. In figure 24, the 120-days ahead hindcasts are plotted. It can be seen that the hindcasts seem to be stable and accurate, indicating that the groundwater level dynamics are captured well by the model. The observed groundwater level is almost entirely is located within the prediction interval.

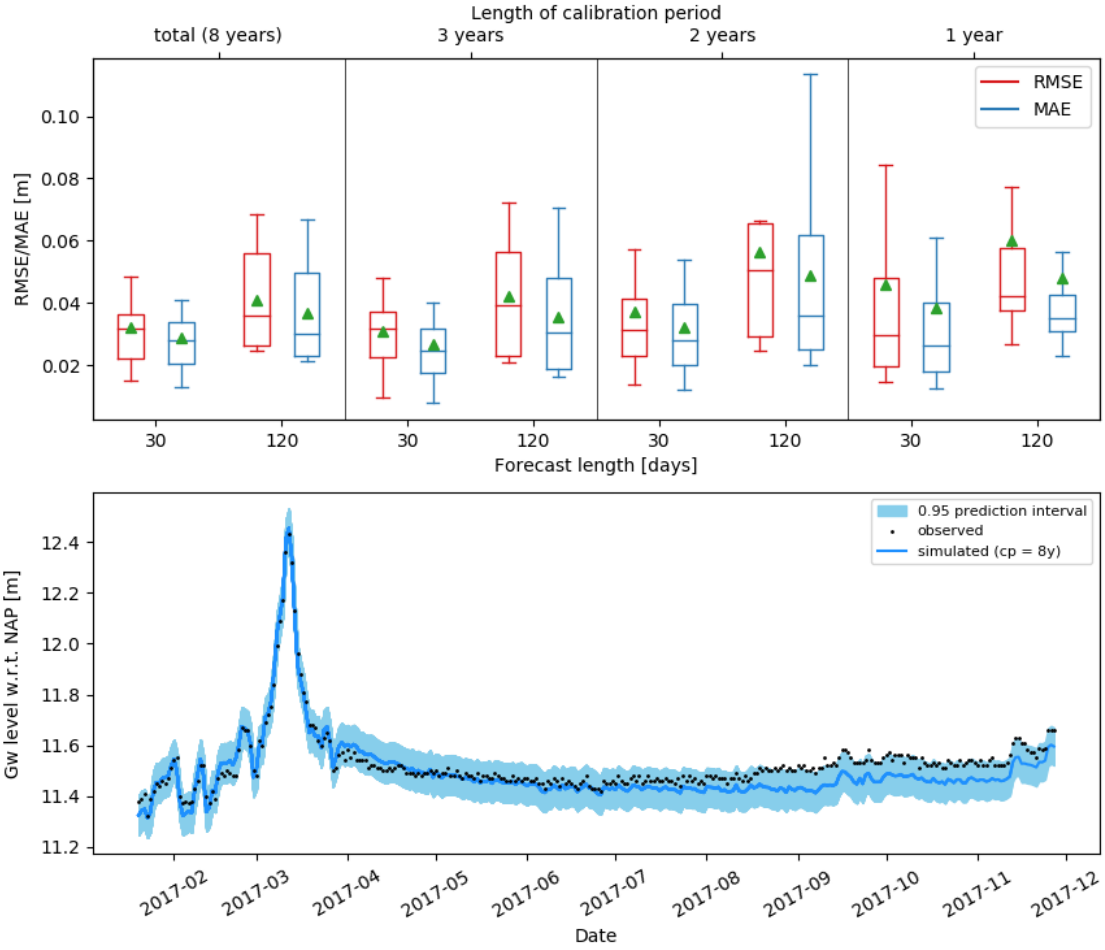


Figure 24: Above: performance of 30 and 120 days-ahead hindcasts using varying calibration period lengths for location B52E3231. The boxplots represent the spread in RMSE and MAE for the hindcasts for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B52E3231 using the total set of historical data for calibration. Every two weeks, the model was recalibrated and a new hindcast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.

4.6 B52E3234

4.6.1 Model setup

For location B52E3234, the potential evaporation and precipitation timeseries were used as primary input variables (Appendix II). Similarly, to the other locations for which the meteorological forcing was the most important stress for explaining the groundwater dynamics, the precipitation excess was taken as a single stress model. The local groundwater level was simulated quite well (R^2 = 0.92, RMSE = 0.09m, MAE = 0.07m). Only two patterns in the groundwater level could not be explained very well: the temporary lowering of the groundwater level in the winter of 2014-2015 and the peak of the groundwater level in the summer of 2016.

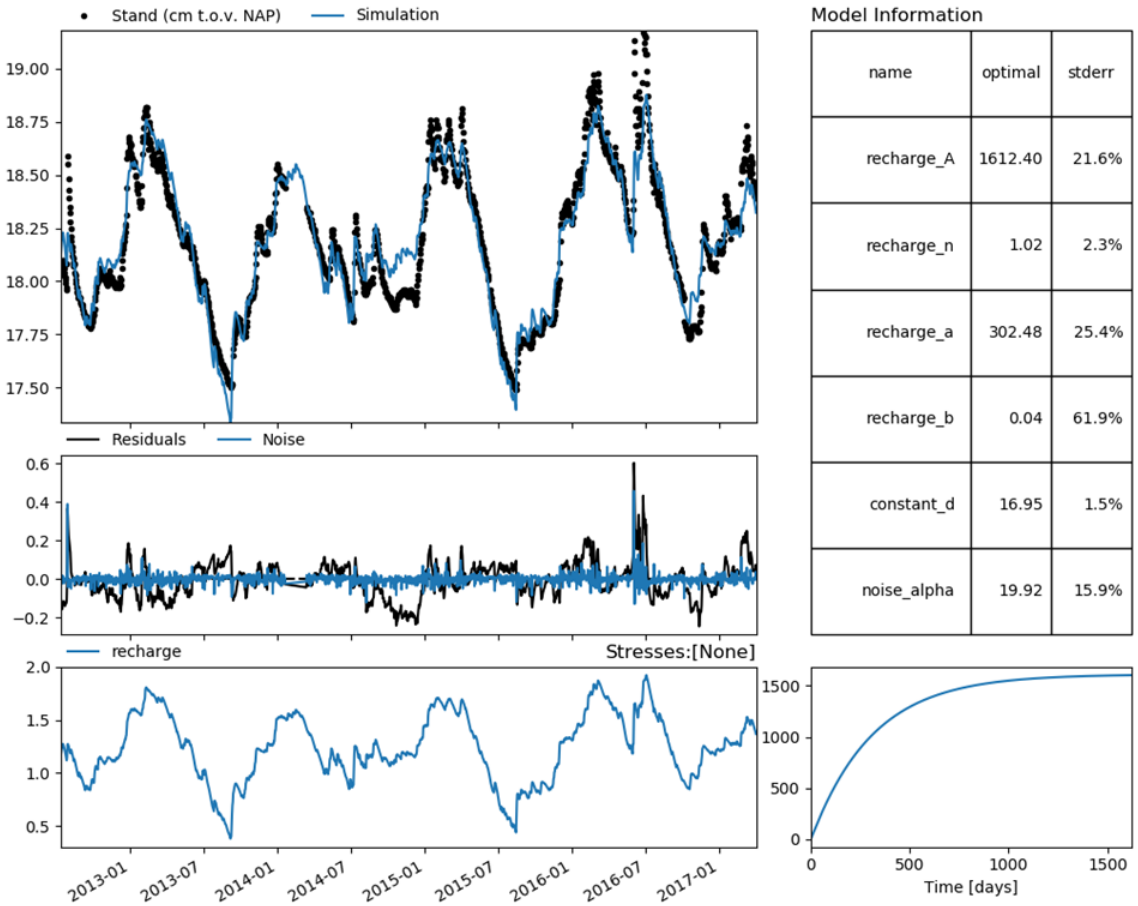


Figure 25: Simulations of the groundwater level using the precipitation excess timeseries as input with one single stress model for location B52E3234.

4.6.2 Hindcasts

From figure 26, it can be concluded that there was no large difference between the performance in hindcasts of the groundwater level using calibration periods of a length varying between the total length of the historical dataset and two years (appendix III). The biases varied mainly between 0.3 and 0 m. Only the performance of the groundwater level hindcasts with one year as calibration period were worse.

In figure 26, the 120 days-ahead groundwater level hindcasts using the total set of historical data for calibration are also plotted. The groundwater levels at the end of the hindcasting period were hindcasted most accurately. Especially during summertime and autumn there were relatively large deviations with the

observed groundwater level. In the middle of the summer, the forecasted drop of the groundwater level seemed to be lower than the drop in the observed groundwater level. In Autumn, the hindcasted rise in the groundwater level seemed to be excessive compared with the observed groundwater level. Since the observed groundwater level exceeded the 0.95 prediction interval during both events, it seems that these patterns cannot explained well by the groundwater dynamics in the past. This indicates that other processes or variables have an influence on the groundwater level besides the precipitation excess.

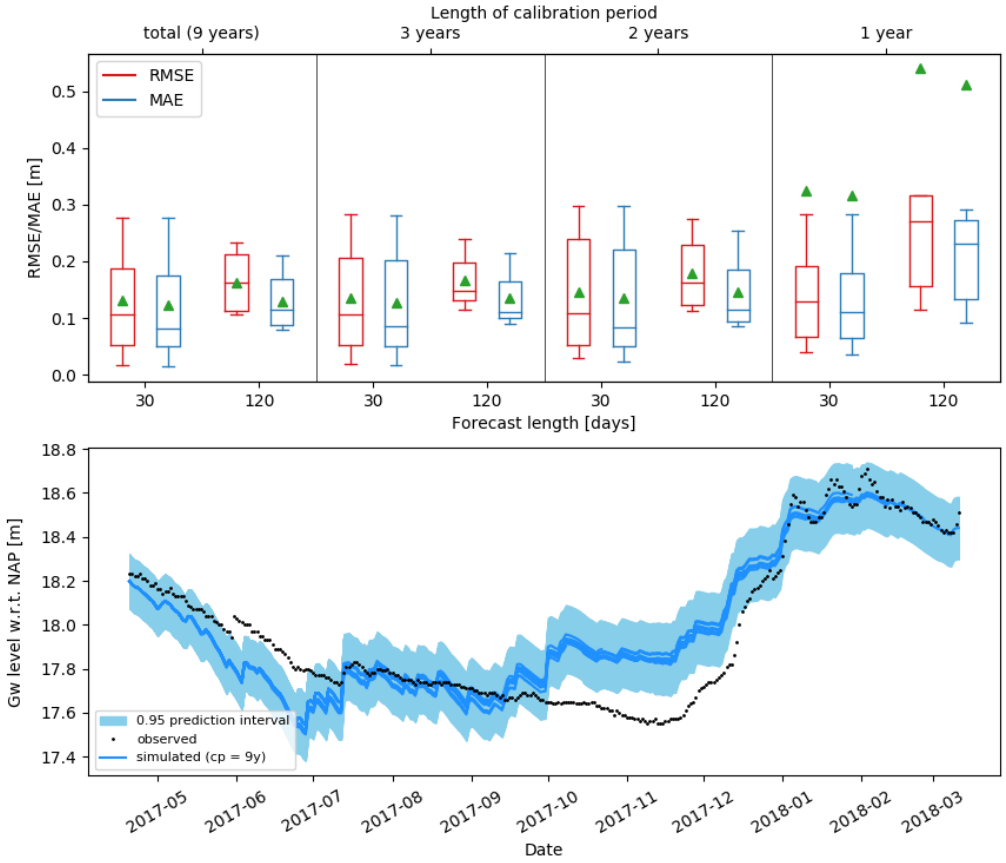


Figure 26: Above: performance of 30 and 120-days ahead hindcasts using varying calibration period lengths for location B52E3234. The boxplots represent the spread in RMSE and MAE for the hindcasts for each scenario, and the average performance is marked with a green triangle. Outliers are not displayed. Below: Continuous hindcasts of the groundwater level 120 days ahead in 2017 for location B52E3234 using a calibration period of two years. Every two weeks, the model was recalibrated and a new forecast was performed and plotted. The last hindcast was performed four months before the end of the observed timeseries. The 0.95 prediction interval, calculated with the standard deviation of the residuals during the calibration period, has also been plotted.

5 Discussion

5.1 Model performance in general

5.1.1 Selection of explanatory variables and stress models

For the majority of the locations, the precipitation excess was most suitable for describing the groundwater level dynamics. The surface water timeseries of the Meuse seemed to be irrelevant as explanatory variable at most locations, with exception of location B52E3231. Location B52E3231, at which the groundwater level was observed within a confined aquifer, is located within a very close distance to the Meuse (±50 meters), and it was therefore expected that the timeseries of the Meuse have a large influence on the local groundwater dynamics. The other locations are located at higher altitude within the landscape, and it is therefore likely that the precipitation excess mainly influences the groundwater level dynamics over there.

Further, implementing rain and evaporation as separate stresses in separate stress models did not yield better model performance than implementing the precipitation excess as stress in a single stress model. This indicated that an increase model complexity did not yield better predictions for the precipitation excess dependent groundwater level timeseries.

5.1.2 Difficulties in explaining groundwater level dynamics

The multi annual fluctuations in groundwater levels could in almost all cases be described quite well with the selected explanatory variables as input. However, some specific sub-yearly patterns were not simulated well. This observation was done in the model setup phase as well as in the hindcasting phase of this study. There are several possible explanations for the limitations of the models in describing groundwater level dynamics at sub-yearly timescale. A general overview of likely causes of limitation in model performance for all locations is also included in appendix IV.

Exclusion of stresses:

The groundwater fluctuations at most of the locations cannot only be described by the local precipitation excess. There are several other processes that have an influence on the dynamics of the observed hydraulic heads, such as human interventions in the hydrological system like for example drainage and groundwater abstractions (Ritzema, et al., 2012).

At all the locations, groundwater abstraction wells for agricultural use are located within a distance of one kilometer, and could therefore have an influence on the groundwater level dynamics during summertime. Especially at location B58C0352 and B58A0093 some specific unexplained structural troughs in observed groundwater head during the growing season (in June and July) seem to be possible groundwater abstractions. At the other locations it is visually hard to detect potential groundwater abstractions from the timeseries visually, but it should be noted that agricultural abstractions in the proximity may possibly have an influence on the groundwater dynamics.

Drainage did probably not cause differences between simulations and observations at sub-yearly scale, especially since including a transformation or a linear trend in all the models did not increase the performance.

Limitations in simulating non-linear response to with a single stress model:

Transfer noise-models in which precipitation and evaporation are used as stresses assume the groundwater level to respond linearly (Petesron & Western, 2013; Zaadnoordijk, et al., 2018). In reality, groundwater level response to meteorological forcing is non-linear. Large rainfall events will produce recharge but a lot of precipitation might be lost to interception and runoff, which are functions of antecedent conditions such as

soil moisture. Smaller precipitation events might produce relatively little recharge because of interception and transpiration by vegetation, or might even refill the unsaturated zone not far enough to produce recharge. During very dry periods, the groundwater might also be transpired by vegetation. At all the locations at which the precipitation excess (precipitation minus potential evaporation) was used as the main explanatory variable, this limitation could have caused the groundwater level predictions to be inaccurate. At locations B58C0352, B58A0093 and B52E3234 the hindcasted groundwater level rose remarkably faster than the observed groundwater level during the autumn of 2017. At locations B52C0005 and B58B0260, the hindcasted groundwater level also seemed to respond more strongly to the precipitation excess in summer/autumn than the observed groundwater level. At all these locations this might have happened because the model does not account for non linear response caused by antecedent conditions in the unsaturated zone for example.

Uncertainties in forcing data:

It should also be noted that for the locations at which the precipitation excess was used as explanatory variable, the representativity of used precipitation timeseries is uncertain. For all the locations data from the most nearby located weather stations was selected, but it is not known up to how far this data is representative for the local precipitation at the groundwater level monitoring wells.

Besides, the potential evaporation was used during this study. The potential evaporation concerns evaporation under optimal conditions, but differs from the actual evaporation.

5.2 Influence of varying the length of the calibration period

When the calibration period was shorter than two years, the performance in simulations was worse in general. This can be explained well theoretically. In order to calibrate the transfer noise model, historical data with the length of at least one times the response time of the system is required to set up a successful model (Knotters & Bierkens, 1999b). The estimated response times for the models in which the precipitation excess was used as stress was mostly in between one and two years. Only by capturing the response time of the groundwater level to the stress entirely in the calibration dataset, the behavior of the groundwater level can be captured accurately enough by the model for generating stable and relatively accurate forecasts.

Further, it was also observed that performance for especially the 120 days-ahead hindcasts seemed to be optimal for relatively short calibration periods in general (2-3 years) at three of the six locations. In all cases, the hindcasts seemed to be more accurate because the hindcasted groundwater levels in summer/autumn were lower and therefore the average deviations with the observed groundwater level were lower as well. It might be premature to state that forecasts using only the more recent historical data are better, because not all dynamics could be fully explained by the models (section 5.1.2). Using short calibration periods can also cause the model to be overfitted more easily on patterns that actually cannot be explained by the used model: the uncertainty of capturing the true process increases when the length of the calibration period is shortened (Zaadnoordijk, et al., 2018). Therefore, no concrete statement can be made about the influence of the length of the calibration period on performance except that this length should at least cover the response time of the system in order to generate stable hindcasts.

6 Conclusion and recommendations

6.1 Conclusion

During this study, Pastas models were set up for performing groundwater level hindcasts at six different locations within Northern Limburg. The selected timeseries involve measurements of the hydraulic head at varying depths. It can be concluded that Pastas can be applied for performing real-time forecasts of groundwater levels in the future. However, some difficulties might occur in capturing the groundwater dynamics at sub-yearly timescale. If Pastas will be applied for real-time purposes these difficulties should be accounted for or addressed.

For all the locations, it was explored first which potential explanatory stresses were required for explaining the groundwater level dynamics and up to how far the groundwater dynamics could be explained. Most of the groundwater levels could be simulated well by implementing only the local precipitation excess as stress. In one case, the groundwater dynamics were explained well by the discharge and water level of the river Meuse. Although the yearly behavior of the groundwater levels was captured quite accurately by the models, the models had difficulties in explaining some specific patterns in the timeseries at sub-yearly timescale. There are several explanations. Firstly, other factors that influenced the groundwater level in reality were not included such as agricultural abstractions for example. Secondly, the input data of stresses might not be completely location representative. Further, non-linear response of groundwater levels to meteorological forcing is not taken into account when simulating groundwater levels with the precipitation excess (precipitation minus potential evaporation) as stress.

After having set up the models, hindcasts of the groundwater levels were performed for 30 and 120 days ahead using calibration periods of varying lengths. Firstly, it was concluded for all the locations that calibration periods of one year or shorter should be omitted due to moderate performance. This corresponds with the statements in literature that at least the response time of the system should be covered in a calibration dataset in order to generate stable groundwater level predictions. Further, for some of the locations it was concluded that for especially the 120 days-ahead groundwater level predictions relative short calibration periods (two or three years) yielded the best performing forecasts. Conditions in the recent past could be more representative for the present, but it should be noted that models with a shorter calibration timespan are more likely to be overfitted on patterns which cannot directly be explained by the model.

6.2 Recommendations

Several recommendations are done:

- Setting up Pastas models for simulating more groundwater levels within the management area of Waterboard Limburg
- The developed functionality for real-time forecasting with Pastas can be integrated in FEWS and used to forecast future groundwater levels based on forecasts or historical records of precipitation or other input stresses to support decisions of water managers in the future. These forecasts can be used for comparing area covering groundwater levels of BOS-OMAR.
- In order to improve forecasting performance, several actions can be taken:
    - Firstly, it can be explored whether using an unsaturated zone model could aid in improving forecasts. In order to account for non-linear reaction of the groundwater level to meteorological forcing, an unsaturated zone model could be used for calculating the

groundwater recharge. The calculated groundwater recharge could be handed over to Pastas as stress to predict the groundwater levels.

- Besides an uncertainty band could be implemented for structural annual deviations caused by stresses of which the data cannot be obtained (groundwater abstractions for example)

7 References

Bot, B. (2011). *Grondwaterzakboekje.* Rotterdam: Bot Raadgevend Ingenieur.

Collenteur, R., Bakker, M., Caljé, R., & Schaars, F. (2019). Pastas: open-source software for time series analysis in hydrology. doi:Zenodo. http://doi.org/10.5281/zenodo.1465866

Doppert, J., Ruegg, G., van Staalduinen, C., Zagwijn, W., & Zandstra, J. (1975). *Toelichting bij geologische overzichtskaarten van Nederland.* Haarlem: Rijks Geologische Dienst.

Graaf, M., Kessels, K., & Overbeek, C. (2017). Waterbeheer in droge en natte tijden met BOS-OMAR. *H2O*.

Keith, W., Hipel, A., & McLeod, I. (1994). Time Series Modelling of Water Resources and Environmental Systems. *45*.

Knotters, M., & Bierkens, M. (1999a). Tijdreeksmodellen voor de grondwaterstand. *Stromingen, 5*(3), 35-50.

Knotters, M., & Bierkens, M. (1999b). Hoe lang moet je de grondwaterstand meten om iets over de dynamiek te weten? *Stromingen, 5*(4).

Massop, H., Gaast, J., & Kiestra, E. (2005). *De doorlatendheid van de bodem voor infiltratiedoeleinden. Een gebiedsdekkende inventarisatie voor waterschap Peel en Maasvallei.* Wageningen: Alterra.

Obergfell, C., Bakker, M., & Maas, K. (2019). Identification and Explanation of a Change in the Groundwater Regime using Time Series Analysis. Delft: Delt University of Technology.

Ogink, H. (2011). *Validatie NHI waterschap Peel en Maasvallei.* STOWA.

Peterson, T., & Western, A. (2013). Nonlinear time-series modeling of unconfined groundwater head. *Water Resources Research*, 8330-8355.

Ritzema, H., Heuvelink, G., Heinen, M., Bogaart, P., van der Bolt, F., Hack-ten Broeke, M., . . . Vroom, H. (2012). *Meten en interpreteren van grondwaterstanden. Analyse van methodieken en nauwkeurigheid.* Wageningen: Alterra.

Stokkers, R., Prins, H., Jager, J., & van Asseldonk, M. (2018). *Effecten droogte en hitte op inkomens land- en tuinbouw.* Wageningen: Wageningen Economic Research.

Tallaksen, L., & van Lanen, H. (2004). *Hydrological drought: Processes and estimation methods for streamflow and groundwater.* Elsevier.

van Balen, R. (2009). Peelrandbreuk en maashorst. *Grondboor & Hamer*.

van den Hurk, B., Klein Tank, A., Attema, J., Bakker, A., Beersma, J., Bessembinder, J., . . . van Zadelhoff, G. (2014). *KNMI (2014): KNMI'14: Climate Change scenarios for the 21st Century – A Netherlands perspective.* De Bilt: KNMI.

van der Linden, E., Haarsma, R., & van der Schrier, G. (2019). Impact of climate model resolution on soil moisture projections in central-western Europe. *Hydrology and Earth System Sciences*, 191-206.

van Rooijen, P. (1989). Grondwater in Limburg. *Grondboor & Hamer*, pp. 377-386.

Vos, C., & Kuiters, L. (2007). *Effecten van klimaatverandering op de natuur.* Wageningen: Alterra.

Wang, X., Tailian, L., Zheng, X., Peng, H., Xin, J., & Zhang, B. (2018). Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Applied Water Science, 8*(125).

Zaadnoordijk, W., Bus, S., Lourens, A., & Berendrecht, W. (2018). Automated Time Series Modeling for Piezometers in the National Database of the Netherlands. *Groundwater*.

I. Subsurface structure at monitoring sites

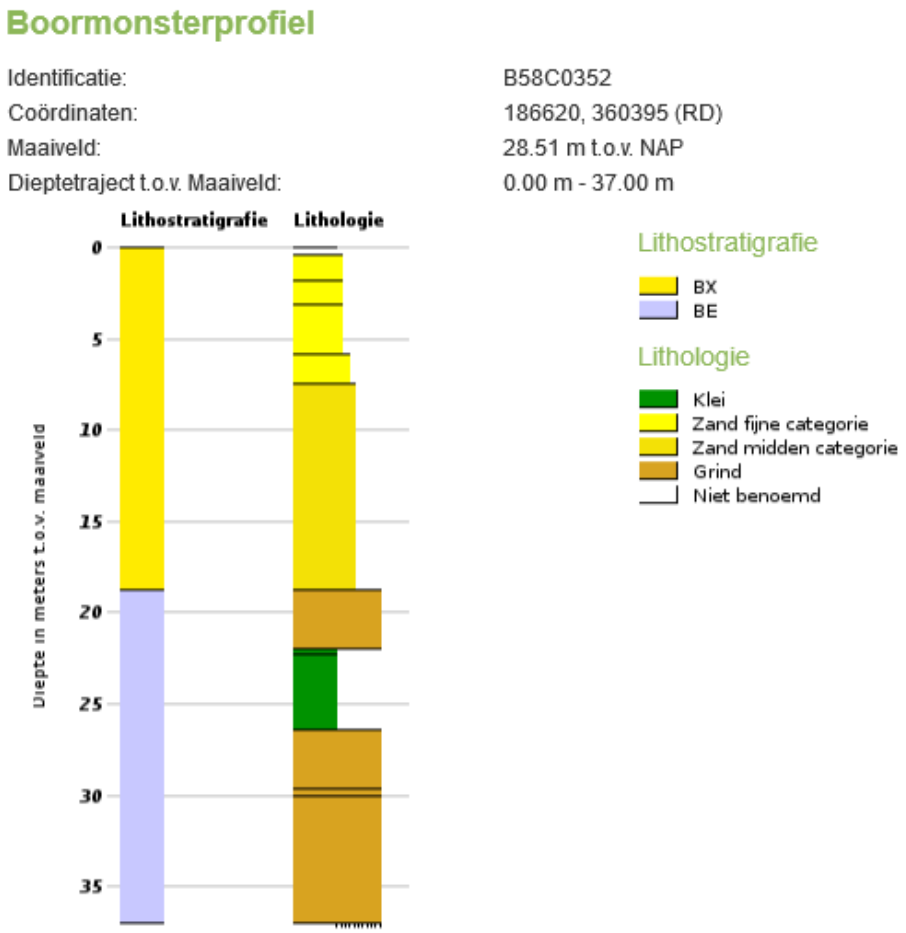


Figure 27: Profile of the subsurface at monitoring site B58C0352 from 0 m to 37 m below surface level. The upper part (0 m to ± 18 m below surface level) consists of fine sands which belong to the formation of Boxtel. The lower part (down to 35 meters below surface level) consists of deposits from the formation of Beegden (gravel and clay). At this location, the hydraulic head was measured within the formation of Boxtel at a depth between 1,8 and 3,8 meters below surface level.

Boormonsterprofiel

Identificatie:	B58A0093
Coördinaten:	188070, 370900 (RD)
Maaiveld:	30.83 m t.o.v. NAP
Dieptetraject t.o.v. Maaiveld:	0.00 m - 35.30 m

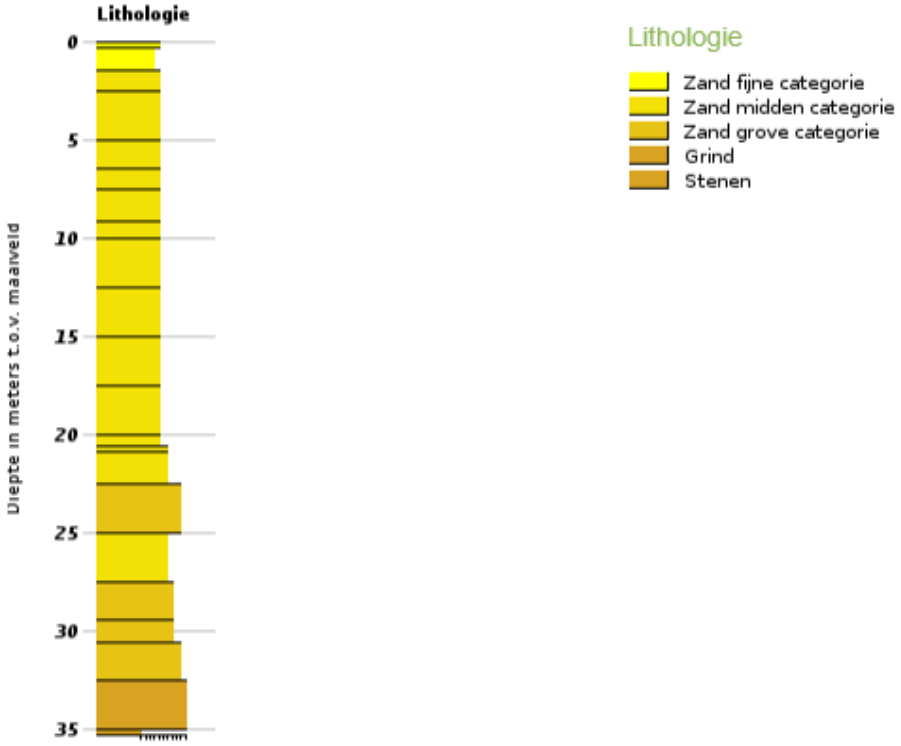


Figure 28: Profile of the subsurface at monitoring site B58A0093 from 0 m to 35 m below surface level. The upper part (0 m to ± 23 m below surface level) consists of fine sands which belong to the formation of Boxtel. The lower part (23 to 35 meters below surface level) consists of deposits from the formation of Beegden (gravel and fine to coarse sand). At this location, the hydraulic head was measured within the formation of Beegden at a depth between 30 and 32 meters below surface level.

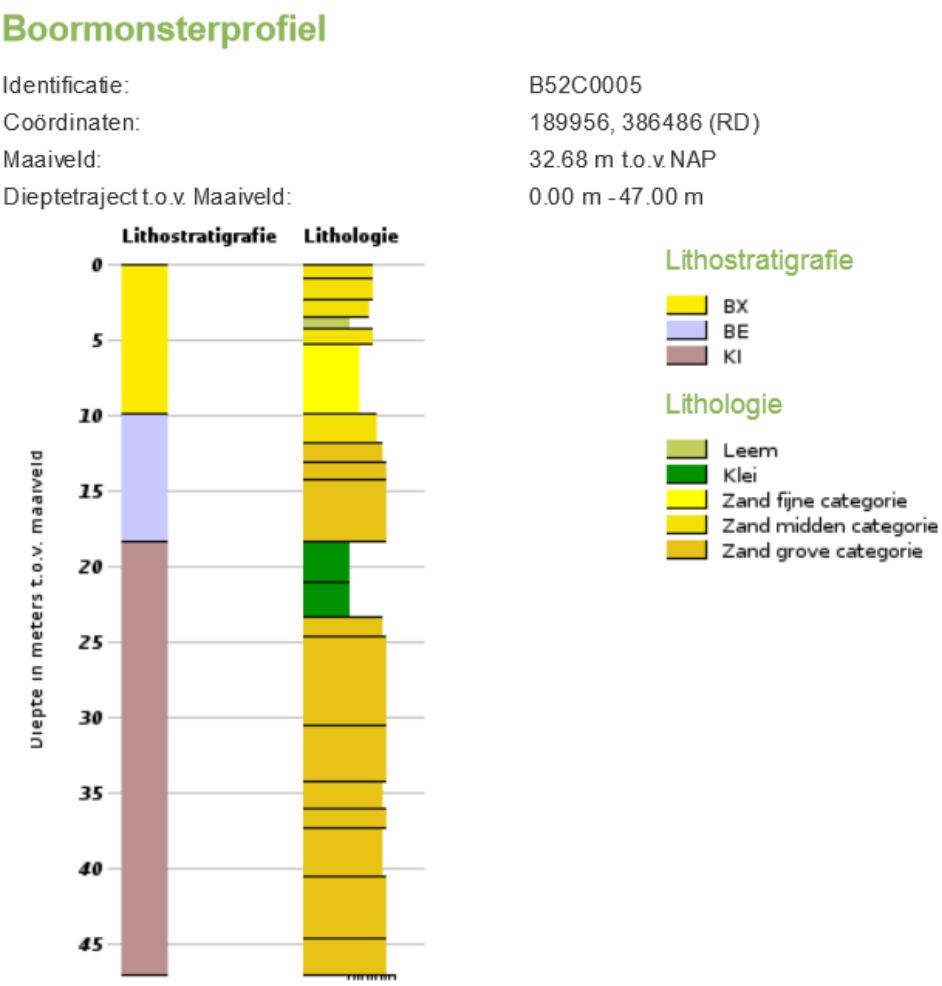


Figure 29: Profile of the subsurface at monitoring site B52C0005 from 0 m to 47 m below surface level. The upper part (0 m to ± 10 m below surface level) consists of fine sands which belong to the formation of Boxtel. The middle part (±10 to ±18 meters below surface level) consists of deposits from the formation of Beegden (fine to coarse sand). The lower part consists of deposits belonging to the Kiezelooliet formation (clay and coarse sands). At this location, the hydraulic head was measured within the formation of Kiezelooliet under locally confined conditions at a depth between 24,6 and 44,91 meters below surface level.

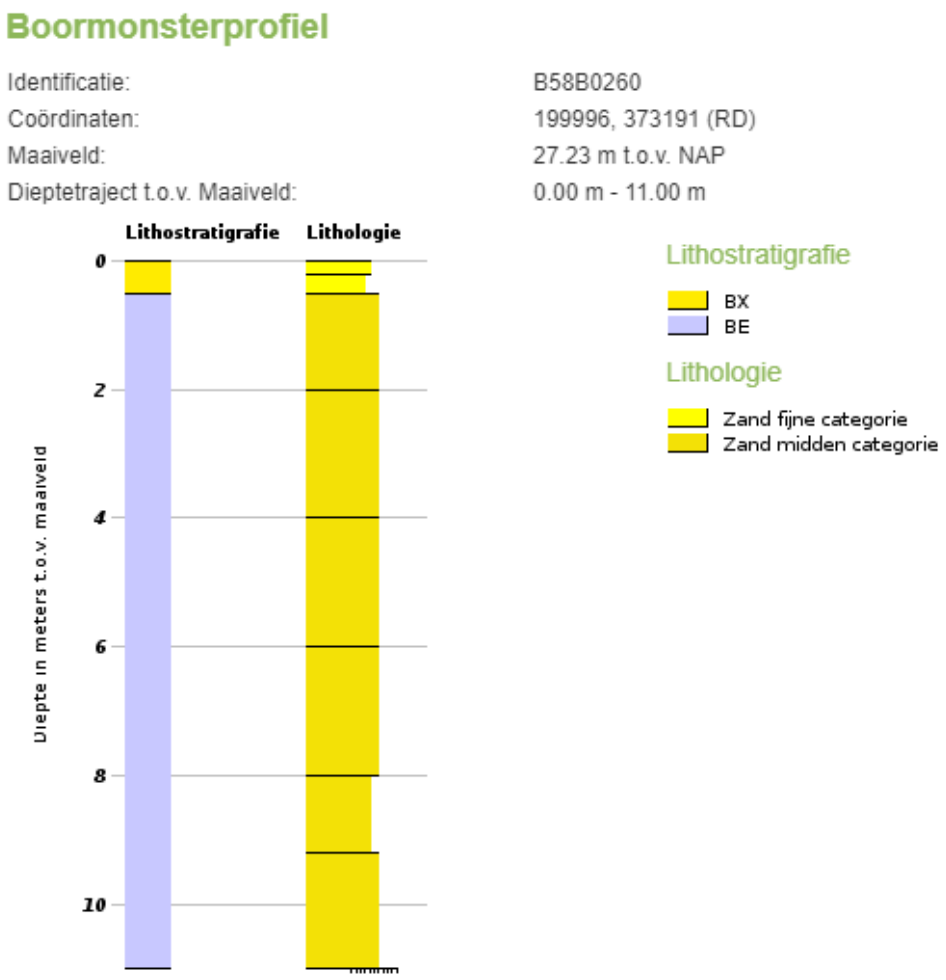


Figure 30: Profile of the subsurface at monitoring site B58B0260 from 0 m to 13 m below surface level. The upper part (0 m to ± 0.5 m below surface level) consists of fine sands which belong to the formation of Boxtel. The lower part (down to 13 meters below surface level) consists of deposits from the formation of Beegden (moderate fine sand). At this location, the hydraulic head was measured within the formation of Beegden at a depth between 6,16 and 9,16 meters below surface level.

Boormonsterprofiel

Identificatie:	B52E3231
Coördinaten:	207216, 392869 (RD)
Maaiveld:	13.66 m t.o.v. NAP
Dieptetraject t.o.v. Maaiveld:	0.00 m - 4.00 m

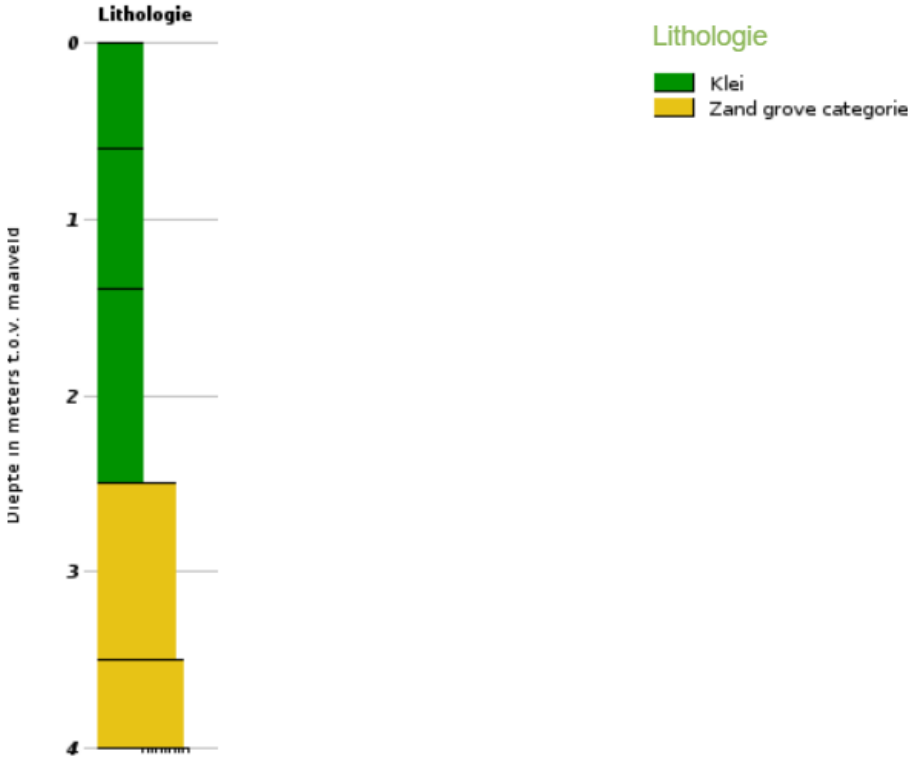


Figure 31: Profile of the subsurface at monitoring site B52E3231 from 0 m to 4 m below surface level. The deposits within the profile belong to the formation of Beegden (clay and coarse sand). At this location, the hydraulic head was measured under locally confined conditions at a depth between 2,91 and 3,91 meters below surface level.

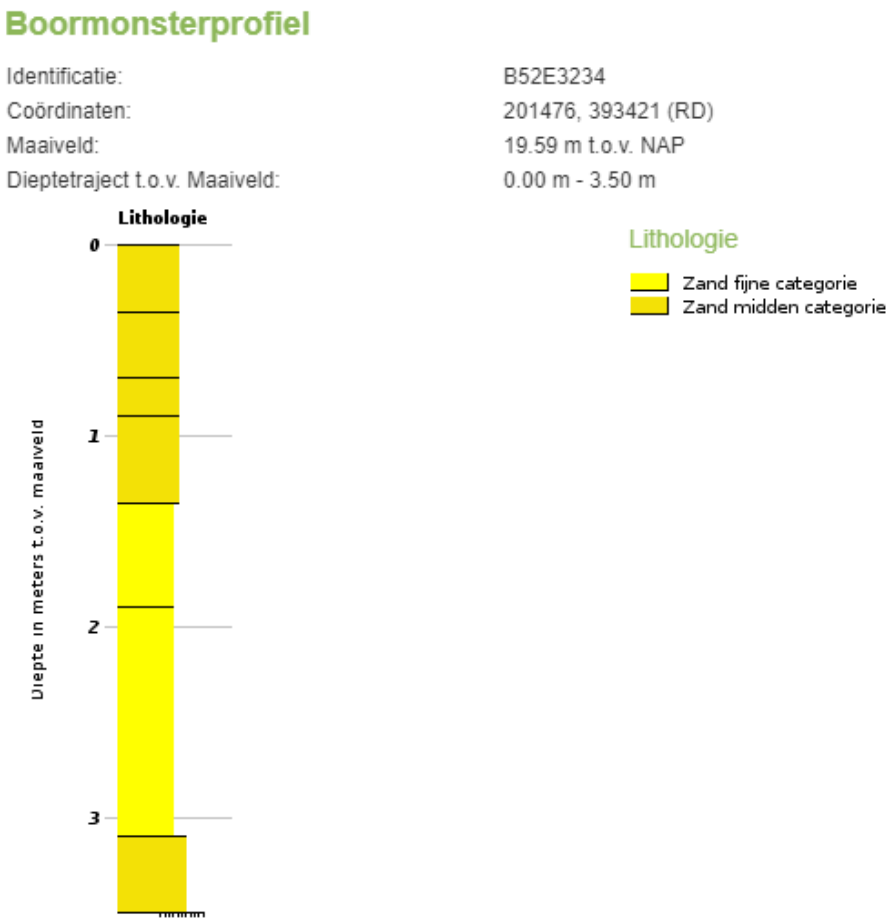
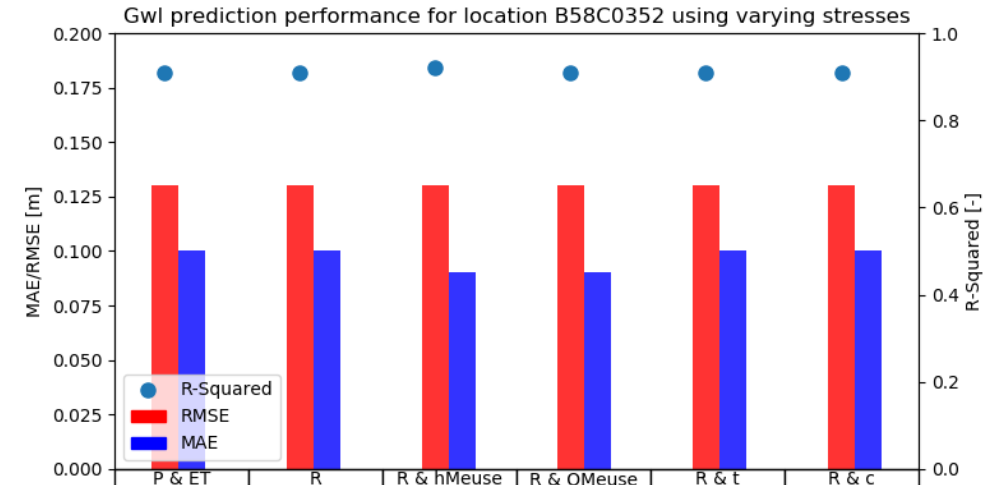


Figure 32: Profile of the subsurface at monitoring site B52E3234 from 0 m to 3,5 m below surface level. The deposits within the profile belong to the formation of Boxtel (fine to moderate fine sand). At this location, the hydraulic head was measured at a depth between 2,39 and 3,39 meters below surface level.

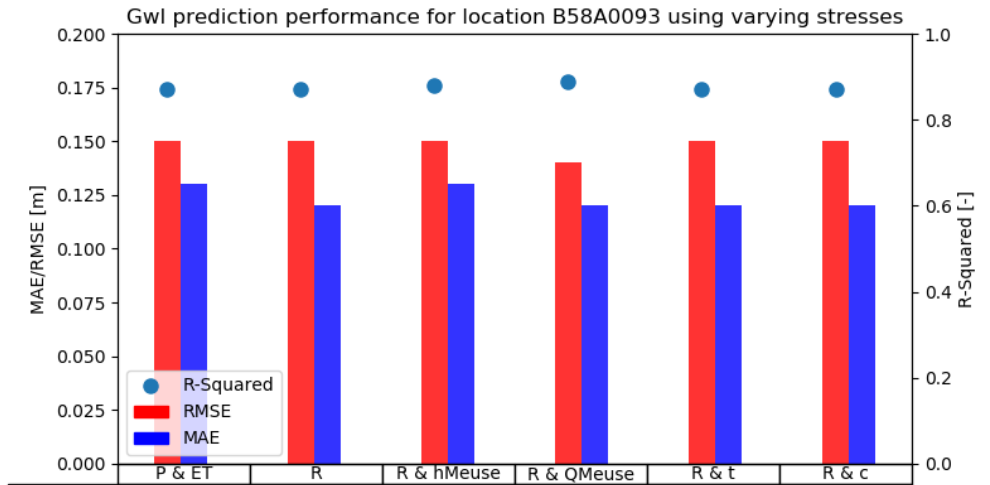
II. Model performance for varying compositions

In this appendix, the performance of models with varying compositions in simulating the groundwater level is included for each monitoring location. In the different model setups, several stresses were implemented: precipitation (P), potential evapotranspiration (ETpot), the precipitation excess (R), the water level of the Meuse (hMeuse) and the discharge of the Meuse (QMeuse). Recharge was calculated by subtracting the potential evapotranspiration from the precipitation. Also, additional elements were implemented: a transformation (t) and a constant (c).



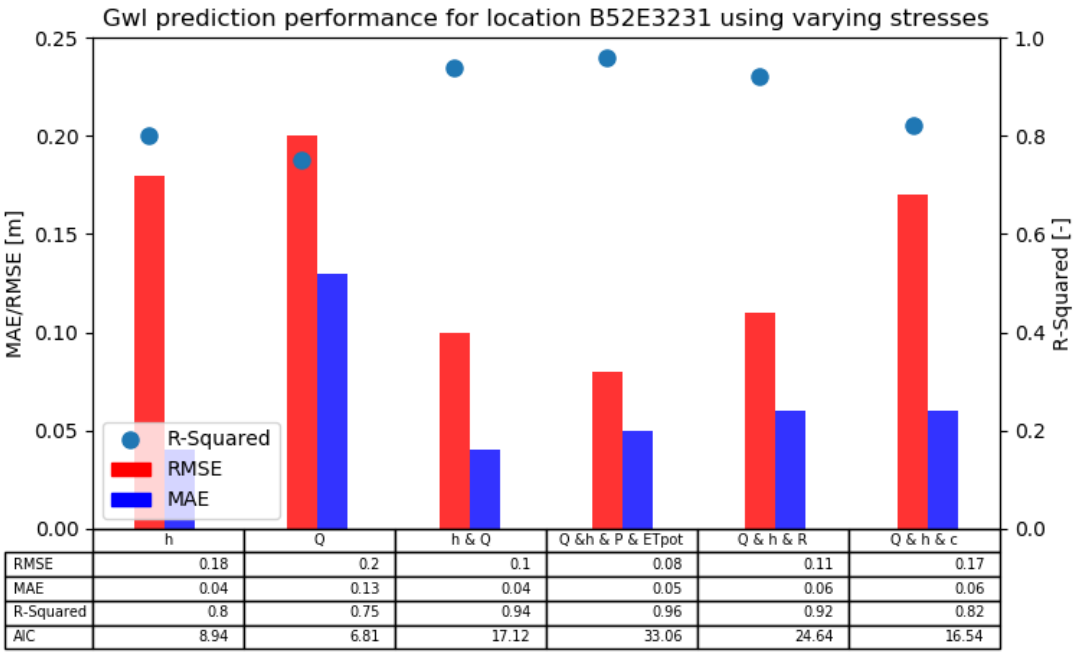
	P & ET	R	R & hMeuse	R & QMeuse	R & t	R & c
RMSE	0.13	0.13	0.13	0.13	0.13	0.13
MAE	0.1	0.1	0.09	0.09	0.1	0.1
R-Squared	0.91	0.91	0.92	0.91	0.91	0.91
AIC	15.9	7.9	17.9	17.9	7.9	11.9

Figure 34: Performance of models with varying compositions in simulating groundwater levels at location B58C0352. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.



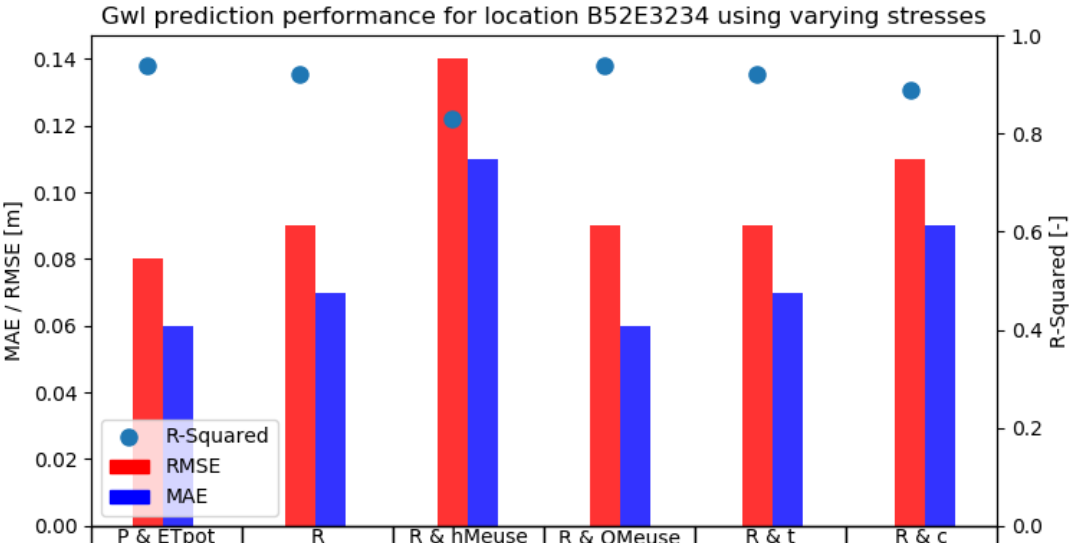
	P & ET	R	R & hMeuse	R & QMeuse	R & t	R & c
RMSE	0.15	0.15	0.15	0.14	0.15	0.15
MAE	0.13	0.12	0.13	0.12	0.12	0.12
R-Squared	0.87	0.87	0.88	0.89	0.87	0.87
AIC	16.33	8.33	16.33	16.33	8.33	12.33

Figure 35: Performance of models with varying stresses in simulating groundwater levels at location B58C0093. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.



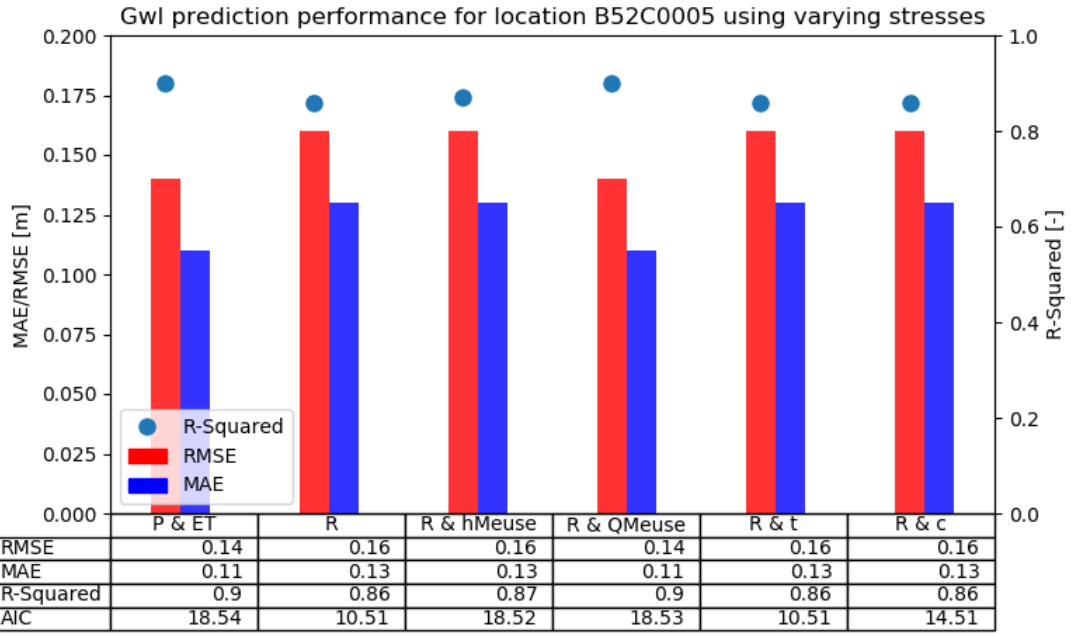
	h	Q	h & Q	Q &h & P & ETpot	Q & h & R	Q & h & c
RMSE	0.18	0.2	0.1	0.08	0.11	0.17
MAE	0.04	0.13	0.04	0.05	0.06	0.06
R-Squared	0.8	0.75	0.94	0.96	0.92	0.82
AIC	8.94	6.81	17.12	33.06	24.64	16.54

Figure 36: Performance of models with varying stresses in simulating groundwater levels at location B52E3231. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.



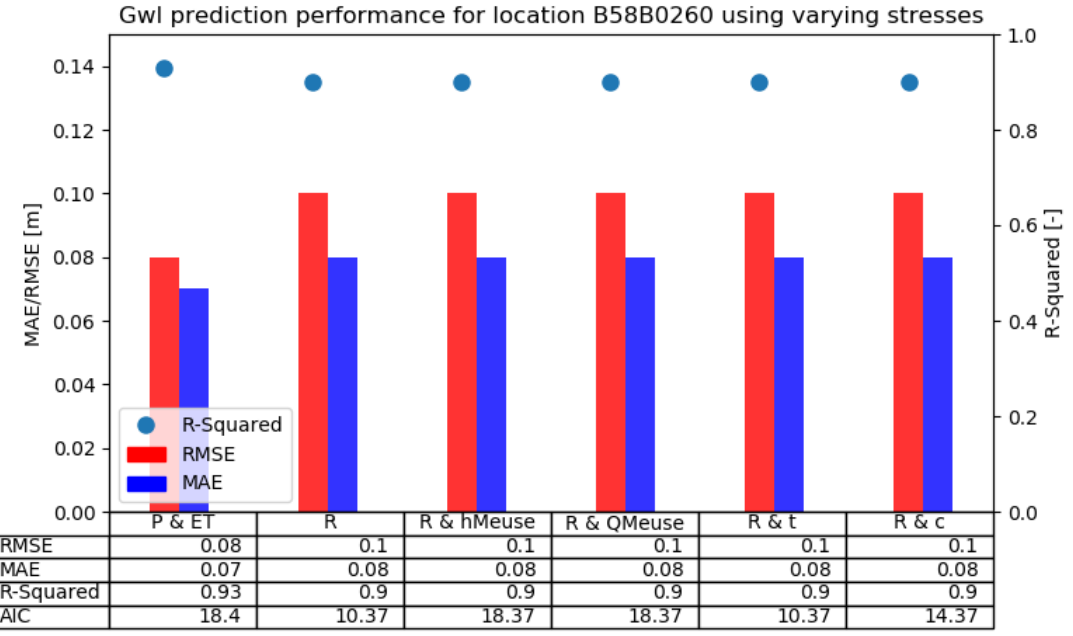
	P & ETpot	R	R & hMeuse	R & QMeuse	R & t	R & c
RMSE	0.08	0.09	0.14	0.09	0.09	0.11
MAE	0.06	0.07	0.11	0.06	0.07	0.09
R-Squared	0.94	0.92	0.83	0.94	0.92	0.89
AIC	19.38	11.33	19.3	19.36	11.33	15.31

Figure 37: Performance of models with varying stresses in simulating groundwater levels at location B52E3234. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.



	P & ET	R	R & hMeuse	R & QMeuse	R & t	R & c
RMSE	0.14	0.16	0.16	0.14	0.16	0.16
MAE	0.11	0.13	0.13	0.11	0.13	0.13
R-Squared	0.9	0.86	0.87	0.9	0.86	0.86
AIC	18.54	10.51	18.52	18.53	10.51	14.51

Figure 33: Performance of models with varying stresses in simulating groundwater levels at location B52C0005. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.



	P & ET	R	R & hMeuse	R & QMeuse	R & t	R & c
RMSE	0.08	0.1	0.1	0.1	0.1	0.1
MAE	0.07	0.08	0.08	0.08	0.08	0.08
R-Squared	0.93	0.9	0.9	0.9	0.9	0.9
AIC	18.4	10.37	18.37	18.37	10.37	14.37

Figure 39: Performance of models with varying stresses in simulating groundwater levels at location B58B0260. For each scenario the included stresses / additional elements are mentioned. If more than one stress is mentioned, they were implemented as separate stress models. Performance is expressed with MAE, RMSE and R-squared.

III. Plots of hindcasts

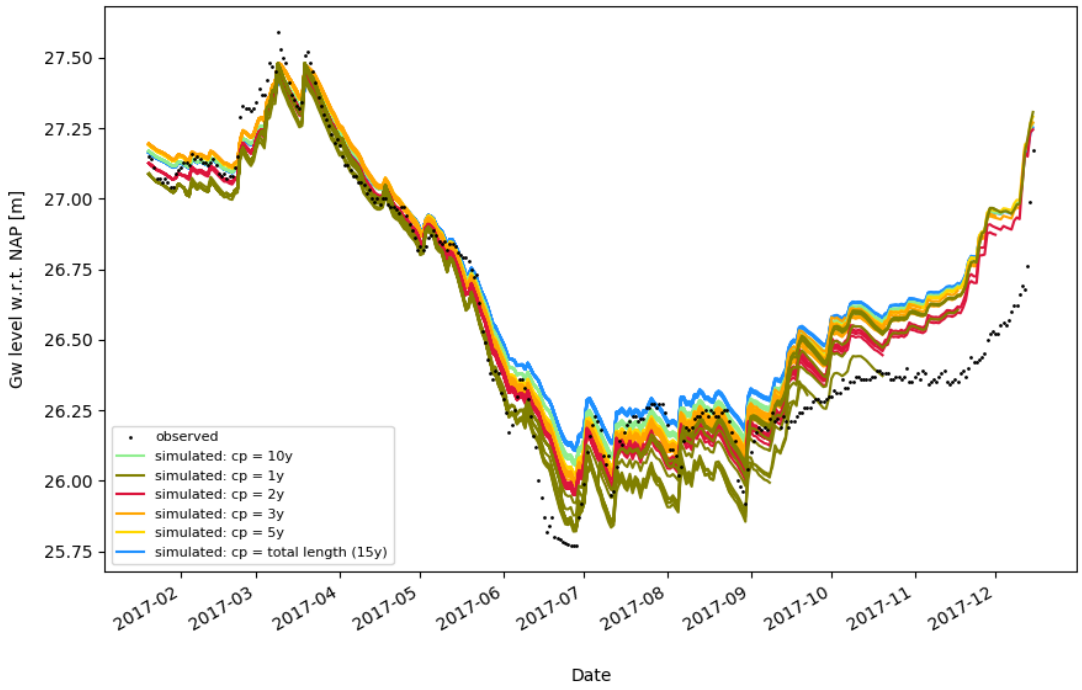


Figure 40: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location B58C0352 using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 September 2017.

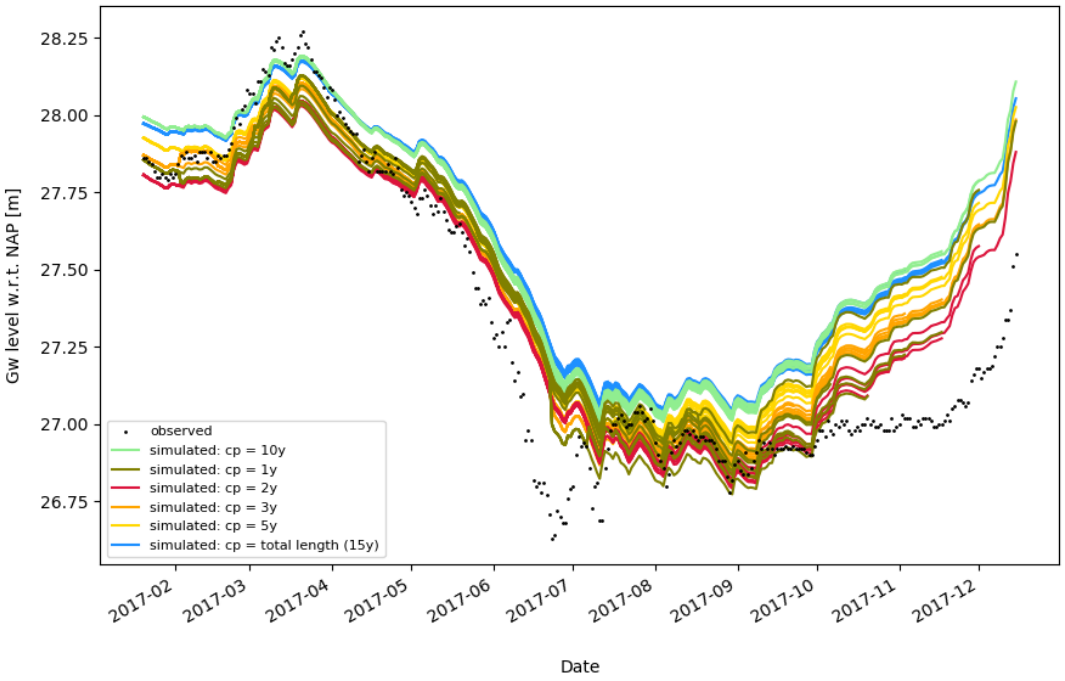


Figure 41: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location B58A0093 using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 September 2017.

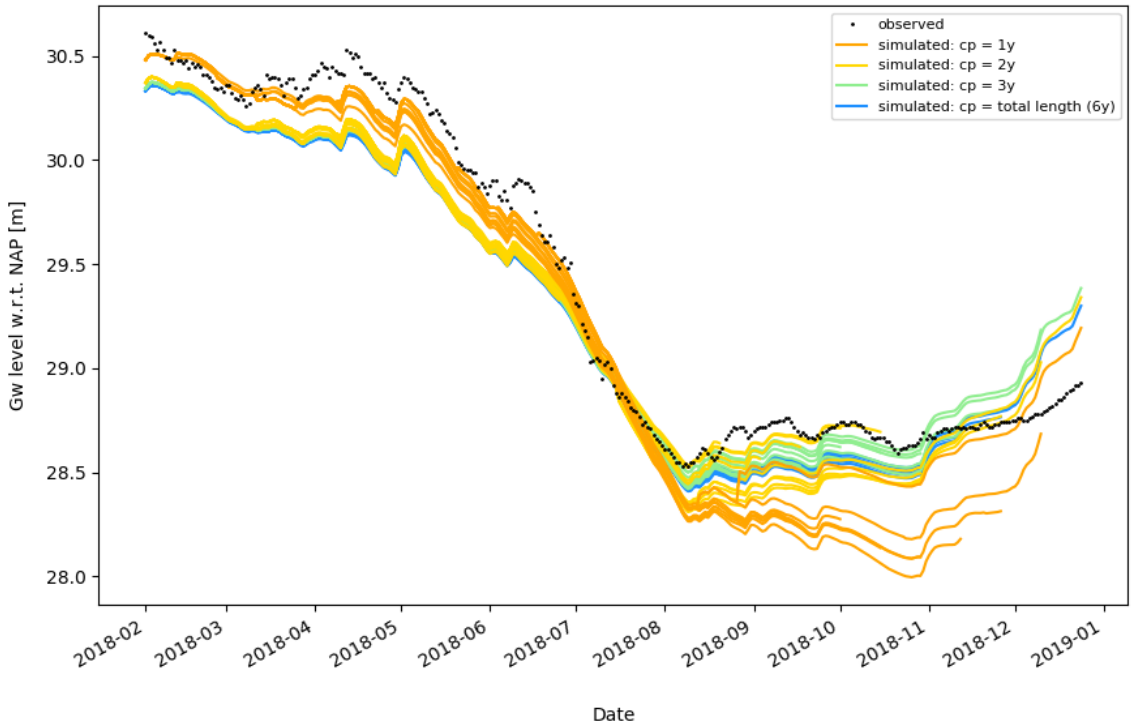


Figure 42: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location B52C0005 using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 September 2018.

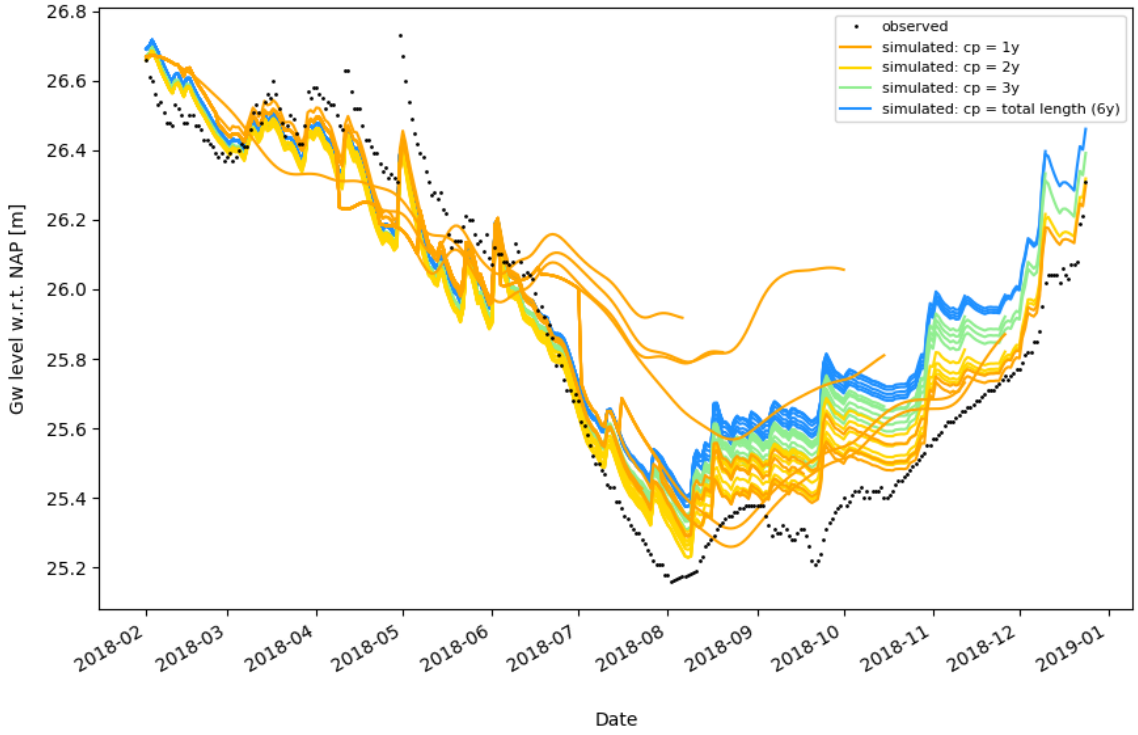


Figure 43: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location B58B0260 using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 September 2018.

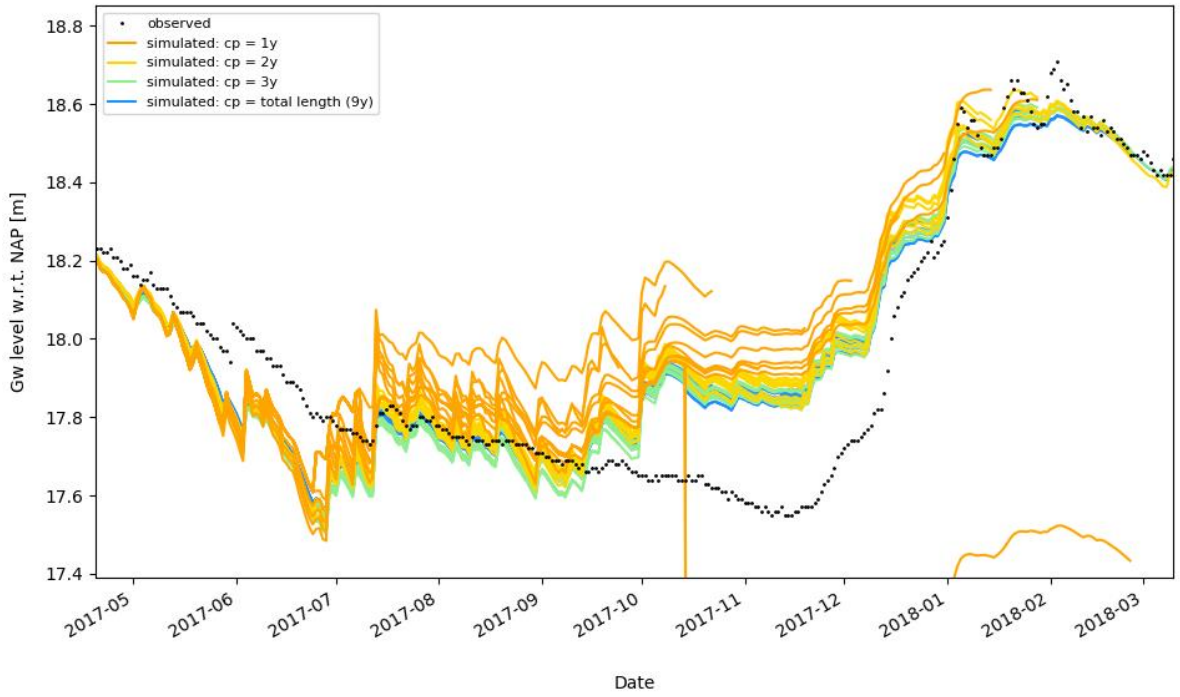


Figure 44: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location B52E3234 using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 November 2017.

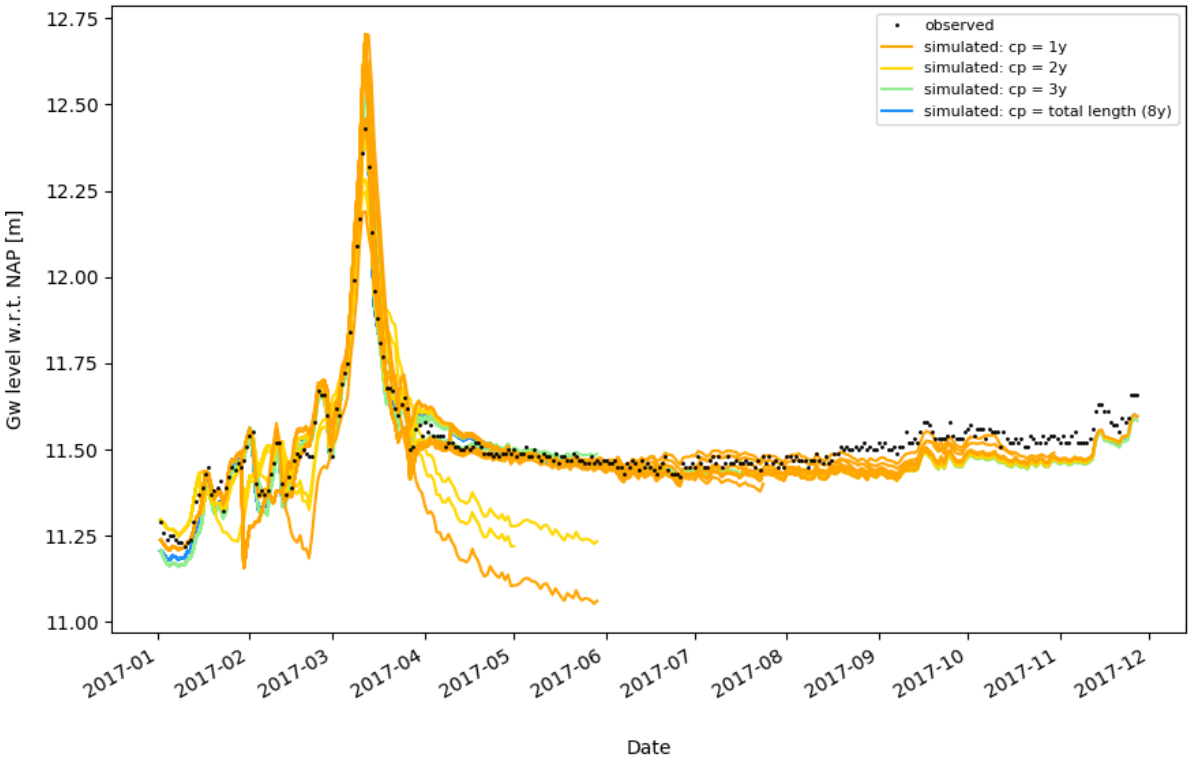


Figure 45: Plots of the real-time 120 days ahead hindcasts of the groundwater level at location *B52E3231* using calibration periods of varying lengths. The length of the calibration period is indicated with cp. Hindcasts were performed every two weeks and all hindcasts are plotted for each scenario. The last hindcasts for each scenario were performed on 1 November 2017.

IV. Overview possible causes limitations in groundwater level predictions

POSSIBLE CAUSES OF LIMITATIONS IN GROUNDWATER LEVEL PREDICTIONS	LOCATIONS	EXPLANATION
EXCLUSION OF STRESSES:		
(AGRICULTURAL) GROUNDWATER ABSTRACTIONS	B58A0093, B58C0352	Some specific troughs in the groundwater timeseries during summertime (June/July) seem to be possibly caused by groundwater abstractions.
UNCERTAINTY IN INPUT DATA:		
PRECIPITATION	B58C0352, B58A0093, B52C0005, B58B0260, B52E3234	Precipitation data were obtained from the most nearby located weather stations. Locally, the real precipitation might differ
EVAPORATION	B58C0352, B58A0093, B52C0005, B58B0260, B52E3234	The potential evaporation was used for each location to calculate the precipitation excess. The potential evaporation involves the evaporation under optimal circumstances, and does not equal the actual evaporation.
LIMITATIONS IN SIMULATING NON-LINEAR RESPONSE TO STRESSES:		
PRECIPITATION EXCESS	B58C0352, B58A0093, B52C0005, B58B0260, B52E3234	At the mentioned locations, the groundwater level was simulated using the precipitation excess (precipitation minus potential evaporation) as stress. The reaction of the groundwater level to meteorological forcing is non-linear, and is not captured by the stress models with the precipitation excess as input: single stress models are constrained to simulating linear response

WELL	SURFACE LEVEL (M W.R.T. NAP)	FILTER NUMBER	TOP OF FILTER (M W.R.T. SURFACE LEVEL)	BOTTOM OF FILTER (M W.R.T. SURFACE LEVEL)	MONITORING PERIOD
B58C0352	28,51	1	-1,80	-3,8	2001 - 2017
B58A0093	30,83	2	-30	-32	2001 - 2017
B52C0005	29,53	1	-24,60	-44,91	2012 - 2019
B58B0260	34,60	2	-6,16	-9,16	2012 - 2019
B52E3231	13,74	1	-2.91	-3.91	2008 - 2017
B52E3234	19,75	1	-2.39	-3.39	2008 - 2018