Improving flow forecast skill by assimilating groundwater observations in WALRUS

A case study for the Hupsel Brook catchment

R.J. Lubben

December 2020

MSc thesis

Hydrology and Quantitative Water Management Group
Wageningen University

MSc Thesis
Hydrology and Quantitative Water Management Group
Wageningen University

Improving flow forecast skill by assimilating groundwater observations in WALRUS

A case study for the Hupsel Brook catchment

Author:
R.J. Lubben

Supervisors:
Dr.Ir Claudia Brauer
Prof.Dr.Ir Albrecht Weerts

December 6, 2020

Abstract

Flow forecasts derived with hydrological models are used more frequently to decide if water has to be retained or discharged based on expected flow conditions. Former studies show that data assimilation using discharge or surface water level can be successful and can improve forecast skill. However, updating the soil moisture state has proven to be more difficult due to limited available data and large spatial and temporal variability. Therefore, this study aimed to use already widely available groundwater level observations to update storage deficit and groundwater states. For the analysis WALRUS is combined with OpenDA (an open source tool for data assimilation) embedded in Delft-FEWS (an operational data management platform used by the water authority). This configuration is used to assimilate discharge and groundwater level observations, and derived storage deficit, in WALRUS for the Hupsel Brook catchment. The results show that using groundwater in addition to discharge for data assimilation does not increase forecast skill substantially. However, this can be related to the initial high skill for discharge assimilation which limits further improvement. Assimilation with only groundwater observations does show a lower but more constant skill over lead time. However, as the analysis period was relatively short (1 year), future research is needed to assess and quantify skill increase using data of other lowland catchments.

Contents

List of figures

List of tables

1 | Introduction

1.1 General introduction

Climate change is expected to affect global hydrological conditions. For Western Europe, summer precipitation is expected to decrease while the intensity of events will increase (Christensen and Christensen, 2004). Hydrological droughts will become more apparent due to increasing evaporation and longer periods without precipitation. This is expected to decrease river flow and water quality and will affect water availability for agriculture, nature, industry and drinking water production (van Vliet et al., 2013). During winter an increase of precipitation by $3.5-7\%$ per degree warming is expected due to higher air temperatures (van den Hurk et al., 2007). This will increase river discharge and the risk of flooding (Dankers and Feyen, 2008). Floods often lead to property damage and loss of agricultural yield. Together with climate change induced sea level rise and land subsidence, lowland areas are becoming more vulnerable to flooding (van den Hurk et al., 2007).

The impacts of extreme hydrological events such as floods and droughts can be estimated by using hydrological models. Flow forecasts made with these models can be used to decide if water has to be retained or discharged based on expected flow conditions. However, prediction uncertainty becomes larger with longer forecasts because of errors in the initial conditions, observations, simulation and forcing. Data assimilation can reduce these discrepancies between observations and model outputs by frequently updating the model state(s) with new observations (Liu et al., 2012). Figure 1.1 shows the process of state updating by comparing the observed state value with the value of the calculated model state. The updated value is determined by certainty of the model state and the observed value. After the state update, a forecast can be made using input data for the model. When a new observation becomes available, the model state can be corrected again before starting the next prediction.

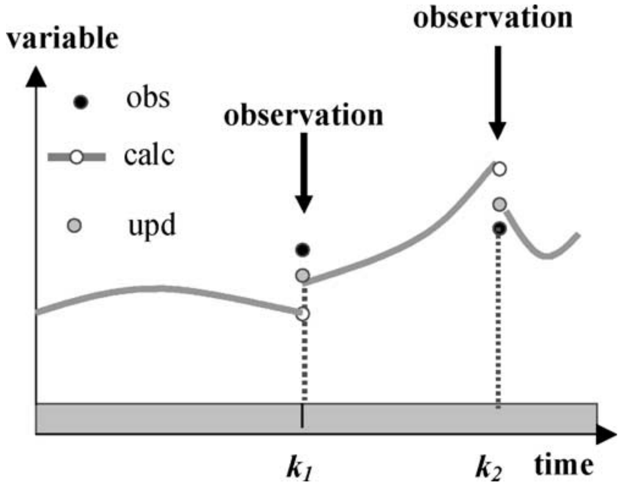


Figure 1.1: Schematic diagram that shows the principle of state updating. When observations become available (black point), the model state (white point) is corrected and an updated value is calculated (gray point). Adapted from Aubert et al. (2003).

1.2 Applications of data assimilation

Different methods exist for updating the model states and addressing uncertainty in observations and model outputs.The Ensemble Kalman Filter (EnKF) is most often used in hydrology (Valk, 2019; Rakovec et al., 2012). This filter is an improved version of the Kalman Filter which uses a probability density function for the observation and corresponding model state using the mean and standard deviation (Kalman, 1960). This method is computationally expensive for models with a larger number of parameters. Therefore, the Ensemble Kalman Filter uses Monte-Carlo sampling to create a smaller set of ensembles of which the mean and standard deviation can be used as uncertainty estimate (Reichle, 2008). This filter is more robust in operational mode than the Particle Filter due to the lower sensitivity to model uncertainty and model misspecification (Weerts and Serafy, 2006). In addition, it is already implemented in the OpenDA software that can be used for data assimilation (Weerts and van Osnabrugge, 2020). For a more detailed description and derivation of the Ensemble Kalman Filter see Weerts and Serafy (2006).

Two state variables that are often used for data assimilation to improve flow forecasting are the surface water level or related discharge and soil moisture content (Rakovec et al., 2012). The potential of discharge and surface water level assimilation is shown by former studies. Rakovec et al. (2012) successfully applied data assimilation using discharge in the Ourthe catchment. Weerts and Serafy (2006) studied the effect of using different filters with discharge assimilation. Valk (2019) and Sun et al. (2020) both showed that assimilation of surface water level for the Regge catchment can be successful and that surface water state updating improved forecasting skill substantially by providing accurate forecasts up to five days. McMillan et al. (2013) used discharge assimilation in seven catchments in New Zealand using an operational flow forecasting system. The forecasts with assimilation showed significantly higher skill compared to simulations without assimilation.

Soil moisture has been applied previously with varying results. Aubert et al. (2003) used probed soil moisture data for improving flow forecasting for the Seine river and showed that this improves forecasting during normal flow conditions and flood events. In addition, Montzka et al. (2011) applied soil moisture assimilation with remotely sensed data and this proved to be successful. According to Komma et al. (2008) soil moisture is highly dependent on soil type and using streamflow assimilation for predicting soil moisture can be more effective than remote sensing. Chen et al. (2011) states that the increase of forecasting performance using soil moisture observations depends on model structure and the degree of coupling between the shallow and deeper soil moisture layers. Lee et al. (2011) used soil moisture assimilation in the Eldon basin, a 795 km^2 headwater catchment located in an agricultural area of the US. During this study soil moisture assimilation did not increase forecasting performance for the Eldon river significantly. McMillan et al. (2013) concluded that prediction performance would most likely increase when soil moisture content is not used for assimilation. Although soil moisture can be used effectively for state updating, the needed data is often not available at the needed update frequency. In addition, soil moisture data could yield large spatial variability under dry conditions due to decoupling of shallow and deeper soil moisture layers.

1.3 Groundwater and data assimilation

Groundwater is important for flow forecasting because it influences runoff by regulating the storage capacity in the soil (Brauer et al., 2011). Despite the importance of groundwater levels for closely linked groundwater and surface water systems, studies of both discharge and groundwater assimilation are scarce (He et al., 2019). Zhang et al. (2016) showed with a MIKE-SHE model of the Ahlergaarde catchment that groundwater assimilation significantly improves forecasting performance. Compared to soil moisture, groundwater level has a lower, but still significant, spatial variability on catchment scale and is expected to resemble the regional water storage more closely (Rakovec et al., 2012). Groundwater levels are already measured in the Netherlands for modelling and monitoring purposes. Despite the large availability, these data are currently not used for flow forecasting.

The implementation of groundwater assimilation in the Netherlands is previously explored by Valk (2019) and Ogilvie (2016). Valk (2019) used WALRUS, a rainfall-runoff model specifically designed for lowland catchments, in combination with Delft-FEWS, a modelling framework for flow forecasting, in the Regge catchment. The results of this study show that there is high correlation between observed and modelled groundwater levels for the Regge catchment. This shows opportunity for applying groundwater assimilation using this model configuration. Soil moisture deficit, quickflow reservoir level and surface water level were also identified as most important state variables for data assimilation in WALRUS. Ogilvie (2016) showed that the relation between groundwater observations and modelled soil moisture deficit can be used for state updating in WALRUS and recommends to use a filter that takes model uncertainty into account.

Using groundwater assimilation for improving flow forecasting shows potential. However, it is not applied in the Netherlands despite the abundance of available groundwater data. Water authorities, such as Water Authority Rijn en IJssel, explore the potential of data assimilation for flow forecasting. However, a common methodology and potential of groundwater assimilation are unknown. Therefore, this study aims to identify the effects of using different observation locations, update intervals and preprocessing methods on forecasting skill. For this study the Hupsel Brook catchment in the management area of Water Authority Rijn en IJssel is used.

1.4 Research questions

This study aims to identify the methodology and potential of groundwater assimilation for flow forecasting in lowland catchments using WALRUS in combination with Delft-FEWS and OpenDA. The main research question is: **How does groundwater assimilation affect forecast skill with a rainfall-runoff model?** To find an answer to this research question, the following sub questions will be answered:

- What is the relation between groundwater level, precipitation and surface water level?

- How do groundwater observations have to be preprocessed for state updating?
- How does groundwater assimilation affect model performance with lead time?
- What is the relation between forecast skill and used observations for assimilation?

The next chapter shows the field site and used data for the analysis. The third chapter describes the methodology consisting of: the model descriptions, statistical methods, model runs and efficiency estimators. The results are shown in chapter 4. The fifth and sixth chapter contain the discussion and conclusion with final recommendations. Appendix A contains a list with used abbreviations.

2 | Field site and data

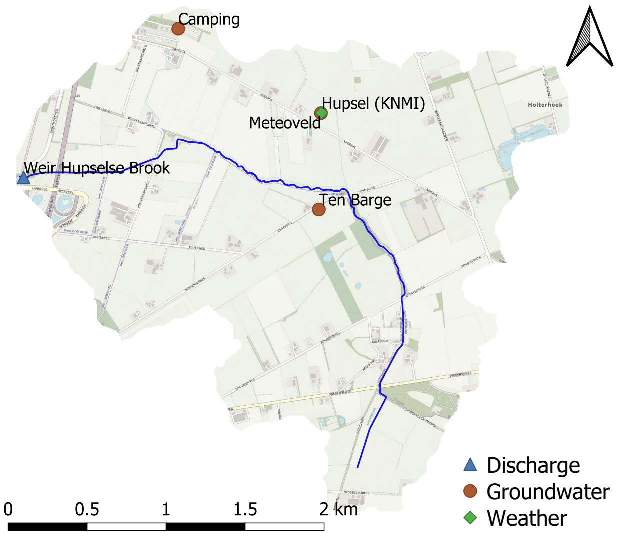


Figure 2.1: Locations of discharge, groundwater and meteorological observations used for the analysis

2.1 Catchment description

The Hupsel Brook has a small catchment (6.5 km^2) with a mainly agricultural land use. This is one of the most well-studied catchments in the Netherlands, so processes and characteristics are well known. The soil consists of an impermeable clay layer below a loamy sand layer which results in single aquifer drainage towards the brook (Brauer et al., 2018). Together with an average catchment slope of 0.8%, this results in a relatively fast response time (Brauer et al., 2011). More detailed information about the Hupsel Brook catchment is available by Brauer et al. (2018).

2.2 Used data

Data between the January 2012 to May 2017 is collected. Figure 2.1 shows the used measurement locations. For the observed discharge the flume is used of which the relation between surface water level and discharge is known. The precipitation and potential evapotranspiration is measured at KNMI station Hupsel. These data are available at KNMI (2020). Three piezometers for the observation of groundwater levels are used: Camping (Ca), Meteoveld (Mv) and Ten Barge (Tb). The measurements of used period are conducted by Wageningen University and are freely accessible via the data portal of Water Authority Rijn en IJssel (Rijn and IJssel, 2020). All variable data, except potential evapotranspiration, has a temporal resolution of one hour. The potential evapotranspiration is estimated as daily sum using daily temperature and radiation data and the Makkink method. The total daily evaporation disaggregated to hourly evaporation. This is done by dividing the radiation of each our over the total daily radiation and multiplying this fraction with the total daily evaporation as done by Brauer et al. (2014a). The years 2013 and 2015 are selected for simulations based on the data gaps, WALRUS model performance and groundwater level fluctuations over each year (Figure 2.2). For the groundwater locations few data are missing during spring for 2013 and 2015 (Appendix B). Data of the observed discharge, precipitation and evapotranspiration for these years is complete. The mean and standard deviation for each observation are given in Table 2.1.

Table 2.1: Mean and standard deviation (Sd) of used variables for 2013 and 2015.

	2013		2015	
Variable	Mean	Sd	Mean	Sd
Q [mm h^{-1}]	0.031	0.047	0.046	0.063
P [mm h^{-1}]	0.081	0.468	0.098	0.553
ET [mm h^{-1}]	0.066	0.115	0.069	0.117
Ca [mm]	-1182	365	-1044	446
Mv [mm]	-1231	376	-1073	345
Tb [mm]	-497	220	-310	257

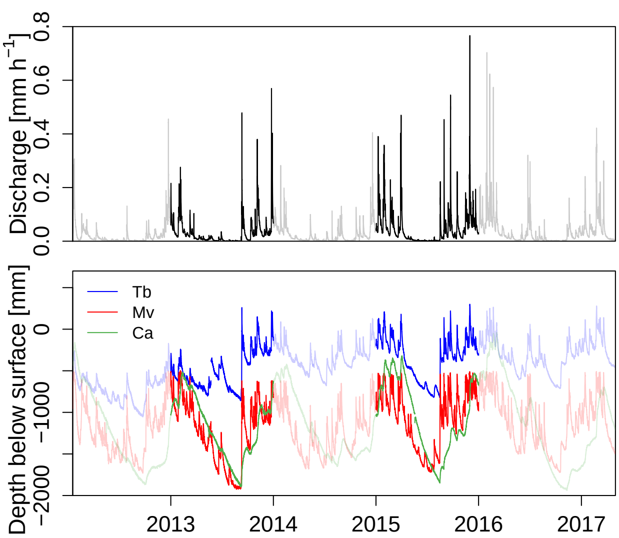


Figure 2.2: Observed discharge and groundwater levels with the years that are used for the analysis in bright colours.

3 | Methodology

The first section of this chapter gives a description of the used models and forecasting framework. After this the observation analysis and the observation model are discussed. This is followed by the conducted experiments and used settings of FEWS. Finally the methodology to determine forecast skill over lead time is explained.

3.1 Models and framework

First the WALRUS model that is used as rainfall-runoff model to simulate discharge is described. This is followed by the description of the DELFT-FEWS modelling framework used for creating forecasts with WALRUS. Finally the OpenDA tool is discussed which is used to perform state updating with hydrological models.

3.1.1 WALRUS

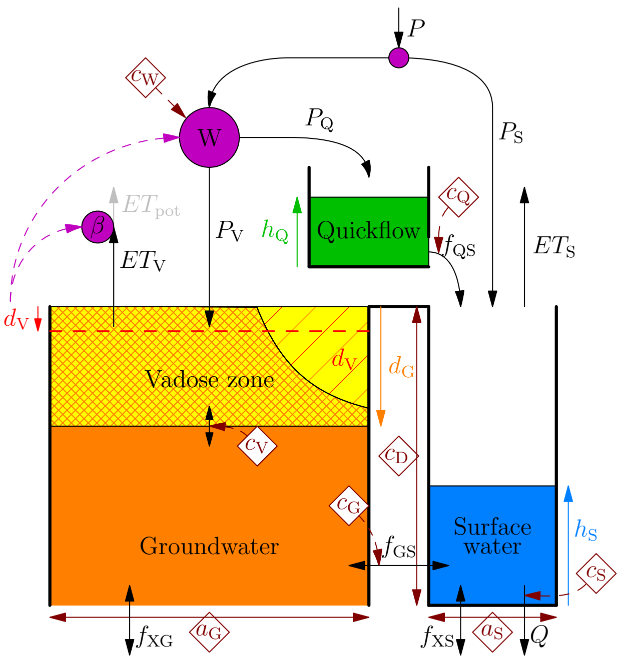


Figure 3.1: Schematic diagram with the model components and input variables of WALRUS. Adapted from Brauer et al. (2014a).

The Wageningen Lowland Runoff Simulator (WALRUS) is a rainfall-runoff model specifically designed for lowland catchments (Brauer et al., 2014a). The model accounts for four important lowland processes: groundwater-unsaturated zone coupling, wetness dependent flow routes, groundwater-surface water feedbacks and seepage and surface water supply. WALRUS is a water balance model that consists of three reservoirs: a soil reservoir, a quickflow reservoir and a surface water reservoir (Figure 3.1).

The soil reservoir consists of the groundwater zone and vadose zone. The vadose zone extends from the surface down to the groundwater level and is characterized by soil moisture deficit (d_V) which affects the evapotranspiration reduction and the wetness index. The fluxes in and out of this zone are infiltration and exfiltration as result of precipitation and evapotranspiration. The groundwater zone is characterized by the groundwater level with respect to surface (d_G). The groundwater level responds to soil moisture deficit variation, which can be caused by precipitation, evapotranspiration, interaction with surface water or seepage. The quickflow reservoir accounts for fast flow through drain pipes, animal burrows and soil cracks. The amount of precipitation that becomes quickflow (overland runoff) is determined by the wetness index (W) of the vadose zone.

The surface water state is characterized by the surface water level (h_S) state. The corresponding discharge (Q) is calculated via a Q-h_S relation. The head of the surface water is affected by: precipitation, evaporation, discharge, extraction or supply, quickflow and groundwater exfiltration into the surface water (Brauer et al., 2014a).

WALRUS is originally coded and used in R. However, a C++ version of WALRUS is implemented in the SOBEK modelling suite, through which it can be coupled to a SOBEK 1D-flow model (Deltares, 2018). In this way open channel flow from SOBEK and rainfall-runoff from WALRUS can be combined. SOBEK is a modelling suite used to study on hydrodynamics, rainfall runoff and real time control. However, for this study SOBEK is only used as interface for WALRUS in FEWS.

3.1.2 DELFT-FEWS

DELFT-FEWS (from here on referred to as FEWS) is a modelling framework that provides a platform in which operational forecasting systems can be constructed (Werner et al., 2013). FEWS is the abbreviation of Flood Early Warning System. The framework has no modelling capabilities and solely relies on external models. The predictions from FEWS can be used as guidance for decision makers to issue warnings. It is used by several water authorities, including Water Authority Rijn en IJssel, for discharge forecasting. The system links models to input data and allows data assimilation with observations by third party software. It also contains data processing functions such as data validation, interpolation, aggregation and disaggregation. The communication between FEWS and external models is done by an XML based interface that consists of time series data, parameters, states, meta-information and run diagnostics (Werner et al., 2013). The combination of WALRUS rainfall-runoff and SOBEK 1D channel flow is already implemented in FEWS and also used for data assimilation in former studies (Izeboud, 2017; Valk, 2019; Sun et al., 2020).

3.1.3 OpenDA

OpenDA is open source software that facilitates data assimilation in external models (Weerts and van Osnabrugge, 2020). It consists of three building blocks: the algorithm or method, modeller and observer. These building blocks can be configured to the needs of the user by XML configuration files. OpenDA contains several data assimilation methods such as the Particle Filter, several Ensemble Kalman filter types and the Ensemble Square Root Filter. The software needs input variables and observations and is already incorporated in FEWS (Weerts and van Osnabrugge, 2020).

3.2 Analysis

The analysis is split into four parts linked to the research questions given in the introduction. At first the relation between the different observation types is determined using correlation and cross-correlation. This is necessary to see if these relations affect forecasting performance and to determine the input for the observation model. In the next part the observation model is created to convert point observations to catchment scale variables that can be used as WALRUS input. This model is calibrated and used to prepare the observations for data assimilation by using the validation outputs. The third step involves the preparation of the model configuration and data, creating forecasts with the FEWS configurations and determining the forecasting efficiency over lead time. Finally the forecasts made by assimilating different groundwater observations and the observation model output are compared. The methodology is further explained in this paragraph and summarized in Figure 3.2 (next page).

3.2.1 Observation analysis

The input variables for the Hupsel Brook catchment are compared to determine the relation between groundwater level, precipitation and discharge. The catchment response time is estimated by using the cross-correlation between precipitation and discharge. The correlation between groundwater level and discharge is estimated for comparison with forecast efficiency when using assimilation with the given state. The degree of correlation can possibly be used to select the most effective groundwater measurement location for updating the groundwater state for assimilation. Further, the lag between the groundwater level and discharge is estimated with cross-correlation. This shows the delay between discharge and groundwater fluctuations in time which can be an important indicator for the selection of the optimal observation for assimilation.

3.2.2 Observation model

One of the main difficulties in using observed groundwater levels for updating states in WALRUS is that these observations are point measurements while WALRUS calculates a groundwater level at the catchment scale. In addition, the groundwater level in WALRUS represents the slow seasonal variation while the fast response of the groundwater table to rainfall events is represented in the quickflow reservoir. This means that the groundwater level observations have to be translated to a catchment scale average. For this a multiple linear regression model, referred to as observation model, is created. This model uses the modelled state by WALRUS as dependent variable and the observed groundwater levels as explanatory variables (left section of Figure 3.2 on the next page).

$$y_\mathrm{i} = \beta_0 + \beta_1 x_\mathrm{i1} + \beta_2 x_\mathrm{i2} + \beta_\mathrm{n} x_\mathrm{in} \tag{3.1}$$

The storage deficit (d_V) state output from WALRUS is used as dependent variable y_i in equation 3.1. The explanatory variables, x_in in the formula are the observed groundwater levels which are fitted with β_n to the modelled storage deficit state. Storage deficit can be updated with observed groundwater levels due to the direct relation between groundwater and the storage deficit of the vadose zone in WALRUS (see Table 1 in Brauer et al., 2014a). Ogilvie (2016) used the same method for updating the storage deficit state from WALRUS successfully. The observation model is fitted on data of

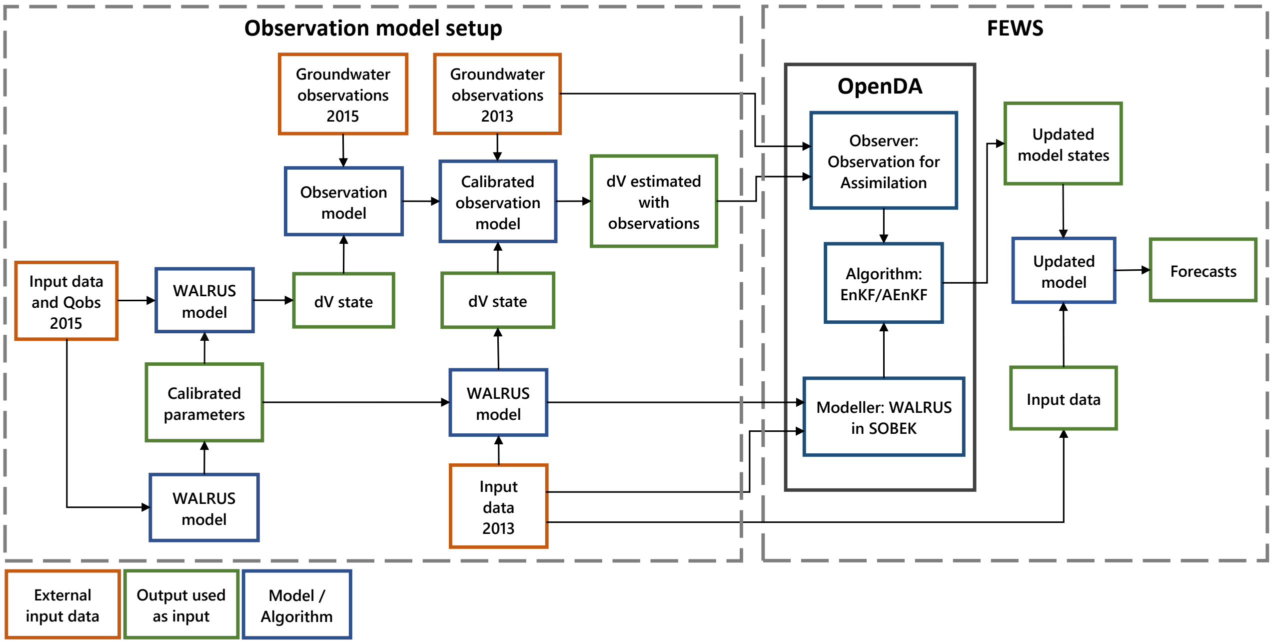


Figure 3.2: Flowchart that summarizes the analysis sections of the observation model and FEWS configuration. The flowchart starts on the left with input data and observed discharge for 2015 and continues to the right. The input data of 2013 is only used to estimate the correlation between the observation model output and WALRUS output and for the forecast with FEWS.

2015 as calibration. After this the observation model is validated with data of 2013. The output of the validation year is used for updating the storage deficit (d_V) state of WALRUS in FEWS. Missing groundwater data are interpolated using linear interpolation (Appendix B).

3.2.3 FEWS-OpenDA configuration

The existing FEWS-OpenDA configuration of Water Authority Rijn en IJssel is used to predict discharge for the Hupsel Brook catchment. A single channel branch is made in SOBEK to which the WALRUS rainfall-runoff model discharges. The SOBEK model is only used as interface for WALRUS in FEWS. The parameter values for the WALRUS models in R, used for the observation model, and SOBEK, used for assimilation and forecasting, are the same (Table 4.1 in section 4.1) and calibrated for 2015.

The right part of Figure 3.2 shows the forecasting process by FEWS with OpenDA. The imported data consist of: precipitation, evapotranspiration and discharge for WALRUS. For state updating the groundwater observations from the Camping (Ca), Meteoveld (Mv) and Ten Barge (Tb) locations are imported. The storage deficit output from the observation model is also imported into FEWS. These observations are used by the observer in OpenDA. The forecasting interval is one day with a forecast duration of 5 days. This is done in sequence for every day of 2013. Uncertainty in the model output is created by perturbing precipitation which is given a standard deviation with a factor of 0.25. This is done by selecting 16 precipitation values from a probability density function using Monte Carlo simulation and running the model with these values. The resulting ensemble output for a given model state is compared with the observed state variable using the Ensemble Kalman Filter Algorithm (EnKF). If using the probability density function of the observation improves the model output the observed variable will be used to update model states (Weerts and Serafy, 2006). For the experiments the EnKF and Asynchronous Ensemble Kalman Filter (AEnKF) are used. The Asynchronous filter makes it possible to update the model with observations by a given interval larger than the temporal resolution of the observed variable. This reduces the computational time substantially due to the lower number of updates.

Table 3.1: Simulations made with the FEWS-OpenDA configuration. For each simulation the name, assimilated state(s), used standard deviation(s) (Sd) and update interval (Int) of the configuration are described.

Name	State 1	Sd 1	State 2	Sd 2	Int
Q	Q	2.5%	-	-	1h
dV	dV	2.5%	-	-	1h
QdV	Q	2.5%	dV	2.5%	1h
Mv	Mv	25mm	-	-	1h
Ca	Ca	25mm	-	-	1h
Q 6h	Q	2.5%	-	-	6h
QdV 6h	Q	2.5%	dV	2.5%	6h
Q 12h	Q	2.5%	-	-	12h
QdV 12h	Q	2.5%	dV	2.5%	12h
Sd5	Q	2.5%	dV	5mm	1h
Sd15	Q	2.5%	dV	15mm	1h
Sd25	Q	2.5%	dV	25mm	1h
Sd32	Q	2.5%	dV	32mm	1h

3.2.4 FEWS-OpenDA simulations

To assess the forecasting performance increase when using groundwater observations and the observation model outputs for data assimilation several simulations are carried out (Table 3.1). First the WALRUS outputs between FEWS and the observation model are compared. After this, the first five simulations as given in Table 3.1 where made to estimate the effect of using different states for assimilation.

For the first run only discharge is assimilated. The second simulation used the storage deficit state from the observation model to update the storage deficit state (d_V) of WALRUS. The third simulation is a combination of storage deficit and discharge assimilation to explore the potential of combining these observed variables for data assimilation. In addition, two model runs with the observed groundwater level at the Camping and Meteoveld locations are used to see if there is a relation between assimilation efficiency and the used groundwater location (Table 3.1). Finally, an Asynchronous Ensemble Kalman Filter with 6 and 12-hour intervals is used to see if lower updating frequencies increase forecasting efficiency when assimilating both discharge and storage deficit (middle section in Table 3.1).

The effect of uncertainty in the observation on forecast skill is studied by using different values for the standard deviation. The standard deviation of the observed discharge is not changed and is set to 2.5% of the observed value. The forecast efficiency over lead time is determined as well as the effect on an event of 5 days starting on the 10th of September 2013. The precipitation amount for this event was exceptionally large (Brauer et al., 2016) and WALRUS estimated the discharge peak poorly with open loop simulations. Several simulation runs are made to explore the effect of different standard deviations in more detail (last section of Table 3.1). The QdV model run is also used for this analysis. All simulations start at the 25th of January 2012 and end at the first of January 2014.

3.3 Estimation of forecast skill

The Continuous Ranked Probability Score (CRPS) and Receiver Operator Characteristics curve (ROC) are used to determine forecast skill. The effect of different standard deviations on forecast skill is also visually estimated by comparing the forecast discharge and storage deficit with the observations.

3.3.1 CRPS over lead time

The Continuous Ranked Probability Score (CRPS) is a verification tool used to determine the skill of ensemble forecasts. The CRPS is a continuous version of the Ranked Probability Score (RPS) which means that the CRPS is not proportional to the number of classes by which the forecast missed the observation (Hersbach, 2000). For deterministic forecasts the CRPS is equal to the Mean Absolute Error (MAE) and ranges from 0 (perfect forecast) to 1 (no relation) (Hersbach, 2000). The CRPS uses a heaviside function for the observed value and compares the forecast ensemble with this function. For a more detailed description see Hersbach (2000).

The CRPS score for a given lead time is determined with the "crps" function from the "Verification" R-package. This function uses the mean and standard deviation of the forecast ensembles and the observed discharge for a given lead time as input. However, the overall CRPS could not be calculated due to missing CRPS values for some instances. This was resolved by manually recalculating the average CRPS without the use of missing values. A more detailed description of the CRPS calculation in R is given in Appendix C.

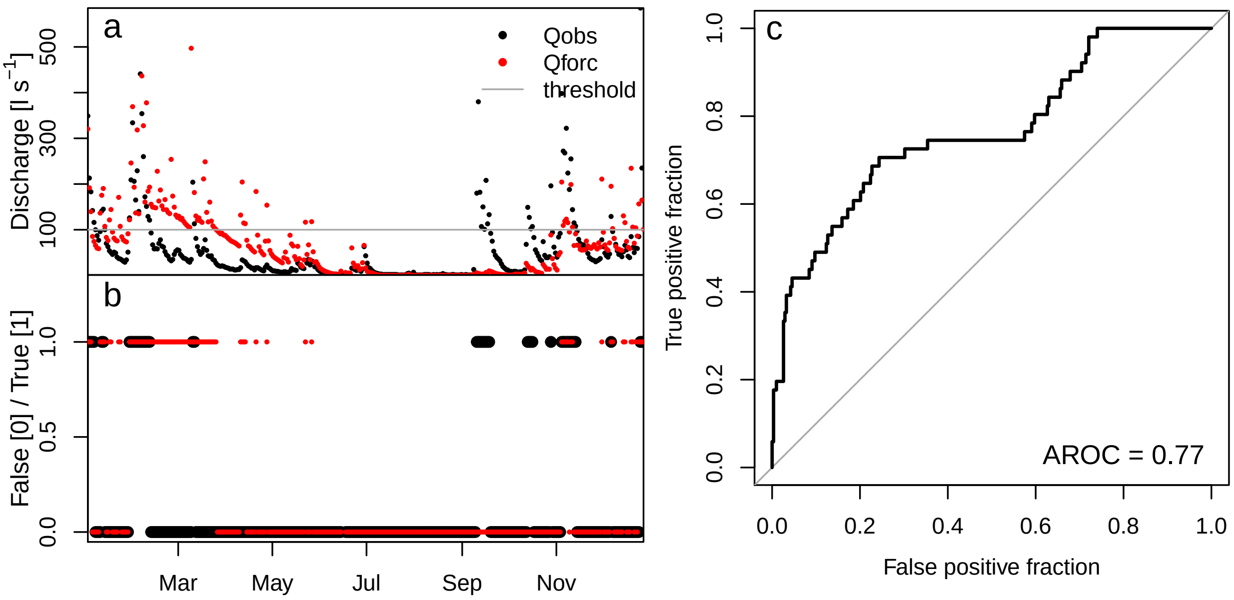


Figure 3.3: Methodology for creating an ROC curve: a) determining the observed (black) and forecast (red) discharge for a given lead time above the threshold level (gray line), b) Indicate threshold exceedance with 1 and non-exceedane with 0 and c) create an ROC curve with the exceedance probability and the ensemble means of forecast discharge and calculate the AROC.

3.3.2 AROC over lead time

ROC (Receiver Operator Characteristic) curves are performance indicators that are widely used to measure the skill of dichotomous (binary) forecasts. These curves are used to compare the hit rate (HR) and false alarm rate (FAR) for different probability thresholds (Alfieri et al., 2012). The closer the curve is to the upper left corner of the graph, the higher the accuracy of the forecast (Figure 3.3c). The overall performance of the ensemble forecast can be determined by calculating the area under the ROC curve, which summarizes the system skill for all the probability thresholds (Alfieri et al., 2013). For a perfect forecast the $AROC = 1$. If the curve is close to the $x{=}y$ line it is considered a random forecast and not able to predict an event ($AROC = 0.5$) (Alfieri et al., 2012). An AROC of 0.7 is assumed to be the limit for a useful prediction system Buizza et al. (1999). For this study the mean of the ensembles is used to calculate the AROC for lead times up to five days.

For the calculation of the AROC for a given lead time the observed discharge has to be evaluated over a given threshold to determine the occurrence of a flood (Figure 3.3a). All observed hourly discharges above the threshold are indicated with 1 and discharges below the threshold are indicated with 0 (Figure 3.3b). The R-package "pROC" uses the dichotomous and forecast discharge to create an ROC curve by determining the HR and FAR and calculates the AROC as can be seen in Figure 3.3c. This method is used to calculate the AROC for all lead times. The mean of the AROC over all lead times is also calculated for easier comparison. A more detailed description of the AROC calculation in R is given in Appendix C.

4 | Results

The first section contains the calibration of the WALRUS model which is used to make forecasts and to calibrate the observation model. This is followed by the relation between input variables. Section 4.3 describes the calibration and validation of the observation model. The next section contains an analysis of the updated model performance when updating different states. Section 4.5 shows the results of using different states, update intervals and standard deviations for the storage deficit state. Finally, the forecast efficiency between runs is compared by using the CRPS (Continuous Ranked Probability Score) and AROC (Area below the Receiver Operator Characteristics curve) over lead time.

4.1 Calibration of WALRUS model

The WALRUS parameters found with calibration are given in table 4.1. The c_W, c_V, c_G and c_Q parameter values are optimized by using the Levenberg-Marquardt optimization algorithm included in the R package "minpack.lm". The model is calibrated with data from the year 2015 and validated with data from 2013 by using the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970). The WALRUS model output is given in Appendix D. The skill is further evaluated by using the Kling-Gupta efficiency (KGE, Gupta et al., 2009). The two efficiency scores use different fundamentals to estimate model efficiency and decomposition of the KGE allows to better understand model performance (Knoben et al., 2019). The starting parameter values for calibration are from Brauer et al. (2014a). For the initial groundwater level a fraction of 0.8 is used for the G_{frac} parameter. The soil type (cal_H) and Q-h_S relation where already available for this catchment. The Q-h_S relation for WALRUS in SOBEK is added as table with a discharge value for every millimetre water depth between 0 and 1500 millimetres.

Table 4.1: This table lists the used WALRUS parameter values for the Hupsel Brook catchment. The values for c_W, c_V, c_G and c_Q are found with auto-calibration.

c_W	c_V	c_G	c_Q	a_S	c_D	st
356	0.21	5e6	3.3	0.01	1500	cal_H

Table 4.2: This table shows the NSE and decomposed KGE values for the WALRUS model calibration (cal) and validation (val) for the Hupsel Brook catchment.

Indicator	cal (2015)	val (2013)
NSE	0.78	0.72
KGE	0.79	0.85
KGE (r)	0.89	0.86
KGE (β)	0.85	0.97
KGE (α)	0.91	0.99

The NSE and KGE scores are given in Table 4.2. The KGE is decomposed into the Pearson correlation (r), bias (β) and relative variability (α) components (Gupta et al., 2009). For the calibration run the NSE and KGE scores are similar. The NSE score for the calibration period (0.78) is assumed to be good. The model has a lower, but still acceptable, NSE score for the validation run (0.72) (Ritter and Muñoz-Carpena, 2013). However, the KGE score for the validation run is higher and decomposition shows that the correlation (r) is the only component at which the validation run scored lower. Therefore, the validation run is assumed to be sufficient.

4.2 Relation between input variables

This section describes the relation between the observed precipitation, discharge and groundwater levels. The relation between precipitation and discharge is used to estimate the response time. The correlation and cross-correlation between discharge and groundwater level is also estimated. Data between January 2012 and May 2017 is used for this part of the analysis.

4.2.1 Response time

For the cross-correlation function, missing values are removed. The response time is estimated to be 12 hours based on the lag times with the highest correlation (Appendix E). However, it was found that the response time depends on the momentary catchment conditions and varies over the year, which was also found by Brauer et al. (2018). Also, the overall correlation is low (maximum: $R^2 = 0.21$). Therefore, the response time is also verified by single events (Appendix E). The estimated response time with this method was 5 to 12 hours dependent on discharge conditions before the event.

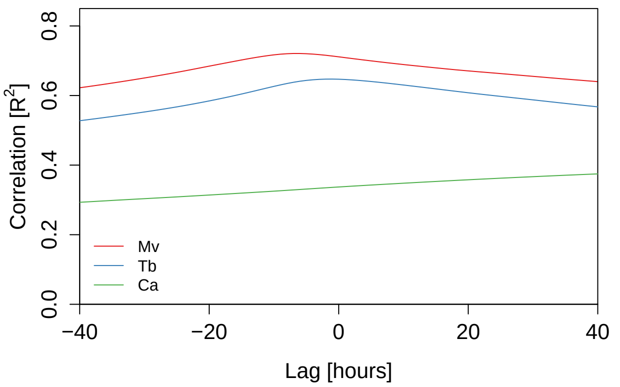


Figure 4.1: Cross-correlation between groundwater level and discharge for Camping (Ca), Meteoveld (Mv) and Ten Barge (Tb). Note: the maximum cross-correlation for Ca is not visible in this figure.

4.2.2 Discharge and groundwater

The relation between groundwater level and discharge is estimated with correlation. The relation with discharge is the strongest for the Meteoveld location ($R^2 = 0.71$). The Ten Barge (Tb) piezometer also shows a relatively high correlation of $R^2 = 0.65$. The Camping (Ca) location yields the lowest correlation with only $R^2 = 0.34$. Figure 4.1 shows the cross-correlation between discharge and groundwater level. The locations of Meteoveld (Mv) and Ten Barge (Tb) locations have a lag time of -7 and 0 hours. The Camping (Ca) has a far larger and positive lag time between 500 and 1500 hours. This indicates that fluctuations in groundwater level at Ca are delayed compared to fluctuations in discharge. The negative lag time for Mv indicates that the discharge follows groundwater level in time. The positive lag time at Ca indicates that the groundwater level follows discharge. The lag time of 0 hours for Tb can be caused by the close proximity of the Hupsel Brook and temporary inundations by the brook.

4.3 Observation model

This section shows the calibration and validation of the observation model. This observation model is used to calculate a catchment scale storage deficit value used for updating WALRUS. The observation model used to estimate storage deficit (d_V) from WALRUS with observed groundwater levels uses all three groundwater measurement locations.

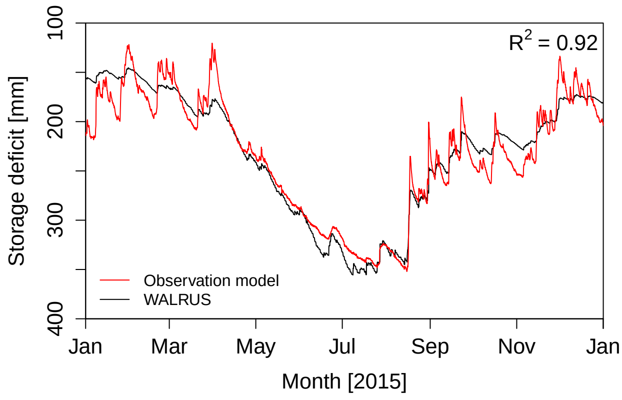


Figure 4.2: Fitted observation model for the calibration year (2015) with simulated storage deficit of WALRUS.

4.3.1 Calibration

Two data inputs are needed to calibrate teh observation model. As dependent variable the storage deficit state (d_V) with a one hour resolution for 2015 from the WALRUS model is used. All three observation locations are significant when used in the linear regression model. Therefore, all three observation locations are used. Figure 4.2 shows the fitted observation model and simulated d_V by WALRUS. The correlation between the storage deficit output from WALRUS and the observation model is high ($R^2 = 0.92$). Other methods of increasing the observation model accuracy by reducing hysteresis and variability where not satisfactory (Appendix F). Other regression model shapes did not increase the relation between the modelled storage deficit and observed groundwater level and a linear model showed to be sufficient (Figure 4.3). Valk (2019) also showed that quadratic and logarithmic regression models do not increase the fit substantially. An observation model with the WALRUS

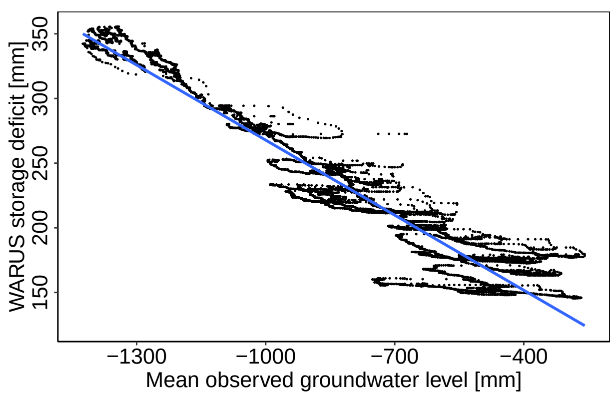


Figure 4.3: Fitted linear regression model (blue line) for the relation between modelled storage deficit by WALRUS and observed groundwater level for 2015.

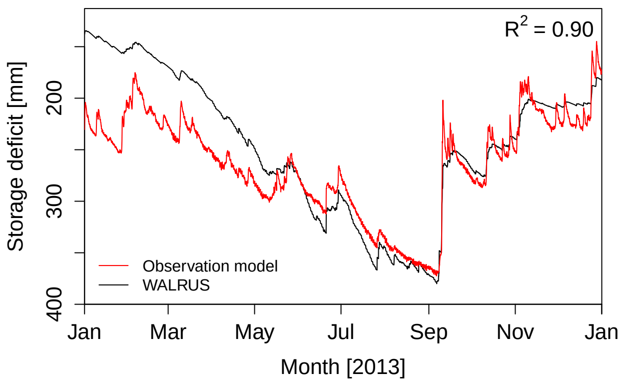


Figure 4.4: Calculated storage deficit by the observation model and WALRUS for 2013.

groundwater state (d_G) as dependent variable was also created. However, d_G is a dependent variable in WALRUS which can limit assimilation effectiveness. Therefore, the observation model for d_G is not used for the analysis. The calibration and validation of this observation model are given in Appendix G.

4.3.2 Validation

To be able to update the storage deficit state (d_V) of WALRUS before making a new prediction, the storage deficit values for state updating need to be estimated from groundwater level observations. This is done with the observation model that uses the measured groundwater levels to estimate storage deficit as used by WALRUS. The observation model was able to estimate the storage deficit state from WALRUS with a correlation of $R^2 = 0.90$ (see Figure 4.4). However, there is a deviation of on average 50 millimetres storage deficit between the outputs of WALRUS and the observation model during the first months.

4.4 FEWS state updating

This section shows the effect of assimilation on the output of the updated model. Also, the use of observations for assimilation is studied by comparing the modelled d_V state of different runs. The updated models are used to make the forecasts described in the next section.

4.4.1 Updated model output

The output from the updated models is compared with the model without assimilation using the NSE and KGE to see if updated models perform better than the non-updated (open loop) model (Table 4.3). The updated

Table 4.3: The NSE and KGE scores of the updated discharge from the WALRUS model used to make predictions. The model output without update (Open Loop Simulation; Ols) is given for reference.

Run	Q	dV	QdV	Mv	Ca	Ols
NSE	0.95	0.69	0.95	0.83	-0.04	0.80
KGE	0.92	0.70	0.90	0.80	0.41	0.74
KGE r	0.98	0.90	0.98	0.92	0.68	0.91
KGE β	0.93	0.80	0.91	0.85	1.35	0.81
KGE α	0.96	1.20	0.97	0.89	1.35	0.85

models with discharge and discharge with storage deficit perform the best and both have a NSE of 0.95. This is an increase when compared to the open loop model without assimilation. Overall, the models updated with only storage deficit ($NSE = 0.69$) or groundwater level from Ca ($NSE = -0.04$) perform worse than the open loop run. Both models have a larger relative variability (α) and Ca also has a larger bias (β) when compared to the open loop simulation which is also indicated by a negative NSE (McCuen et al., 2006).

The simulated hydrographs from the updated models for the event starting at the 10th of September are shown in Figure 4.5. This shows that updating storage deficit (d_V) results in an overestimation of the discharge peak with more than 100%. Assimilation of only the Meteoveld (Mv) or Camping (Ca) observations leads to an underestimation of discharge. Overall, assimilation of only discharge (Q) and discharge with storage deficit (QdV) from the observation model seem to perform equally well.

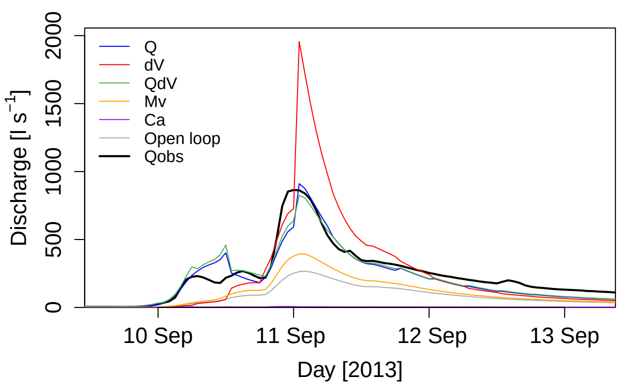


Figure 4.5: Updated discharge for September 2013 from models with assimilation of only discharge (Q), only storage deficit (dV), both discharge and storage deficit (QdV) and groundwater levels of Meteoveld (Mv) and Camping (Ca; purple line at bottom of graph). With observed (Qobs) and open loop discharge for reference.

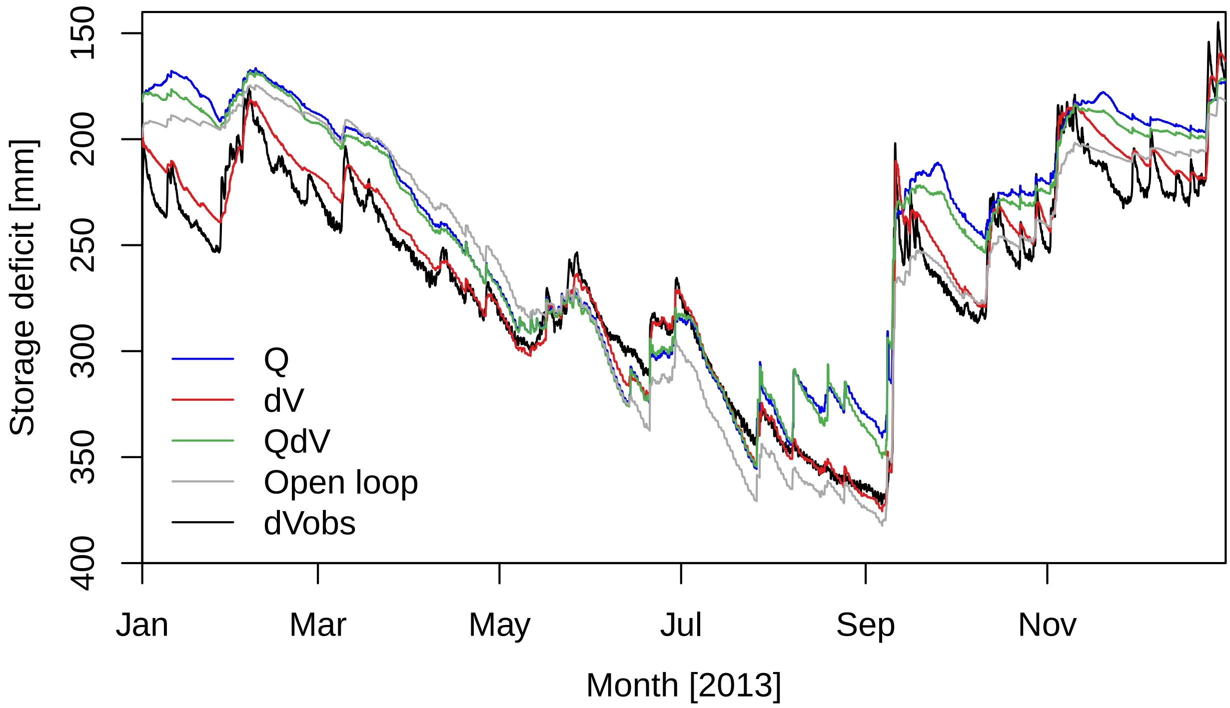


Figure 4.6: Updated storage deficit state of WALRUS for assimilation of only discharge (Q), only storage deficit (dV) and both discharge and storage deficit (QdV). The storage deficit from the observation model (dVobs) and open loop simulation (Open loop) are given for reference.

4.4.2 Updated storage deficit

To see if the storage deficit observations are used for state updating the simulated storage deficit states from different runs are compared. In Figure 4.6 can be seen that the state with only discharge updated (Q) and with both discharge and storage deficit updated (QdV) are similar. The state with only storage deficit updated followed the observed storage deficit more closely than the other two simulations. However, simulated storage deficit of the run for assimilation of Q with dV (QdV) is mostly between the Q and dV simulations. This shows that the storage deficit observations are used when both updating the storage deficit and discharge, but not with large effects on the simulated state. Similar performance of QdV and Q can be caused by the already high performance of Q resulting in dV not being used for updating. In addition, the uncertainty in the dV observation is given as 2.5% of the observation (same as discharge). This means that a larger deficit results in a larger assumed error in the observation, because a fraction of the observed is used as standard deviation. For example, a storage deficit of 350 millimetre will have a standard deviation of $350 * 0.025 = 8.75$ millimetres. The effect on forecast skill is examined further in section 4.5.4.

4.5 Forecast comparison

In this paragraph the forecasting efficiency of different observations, 6-hour and 12-hour interval for state updating and the effect of different standard deviations for the observed variables are shown. All forecasts are compared by using the AROC and CRPS over lead time. For this the AROC threshold levels of 200, 100 and 10 l s^{-1} are used. The results for the CRPS and AROC for 100 l s^{-1} are shown in Figure 4.8. The figures with AROC scores for 10 l s^{-1} and 200 l s^{-1} can be found in Appendix H.

4.5.1 AROC thresholds

The alarm levels are relatively low because only a few larger discharge peaks are available for 2013. The alarm level of 200 l s^{-1} is exceeded for 5.9% of the time. This resulted in an highly variable output for AROC over lead time due to the low number of discharges above the threshold. A lower threshold level of 100 l s^{-1} is exceeded by 16% of the observed discharge values. This is enough data to estimate the AROC over lead time. The third threshold of 10 l s^{-1} was exceeded by 69% of the observations and is also used for calculating the AROC.

4.5.2 Forecast skill for different observations

Figures 4.8a and b show the CRPS and AROC scores for the five assimilation runs with only discharge, only storage deficit, discharge and storage deficit and groundwater levels of Camping and Meteoveld. All predictions have an AROC score of more than 0.7 during the whole prediction period. However, there is a 24-hour periodicity visible which is likely a result of the used FEWS settings. By making forecasts every 24 hours and calculating the CRPS and AROC for every hour of the forecast, only parts of discharge peaks are captured. The CRPS and AROC for Ca show the lowest skill of all runs. Overall, the assimilation run with both discharge and storage deficit updated, show slightly higher CRPS and AROC scores over longer lead times compared to the run with only discharge assimilation (Figure 4.8a and b). However, this may also be an effect related to randomness (further explained in chapter 5). The Mv and dV runs show both a lower initial skill compared to the Q and QdV runs. However, they seem to be more constant over longer lead times. With a threshold level of 10 l s^{-1} (Appendix H Figure a) these runs show the highest skill, which might indicate that updating only storage deficit or groundwater level is useful for low discharge conditions.

4.5.3 Reduced updating frequency

Assimilation of only discharge and discharge with storage deficit is compared for different update intervals. Figures 4.8c and d show the effect of an updating interval of 6 hours. Overall, both the CRPS and AROC show the same pattern and all forecasts are relatively similar. The QdV with an update interval of 6 hours seems to perform the least followed by the Q with an interval of 1 hour according to the CRPS.

With a 12-hour interval the forecast for which only discharge is assimilated has the lowest performance (Figures 4.8e and f) which may be related to the response time of the catchment. This difference is likely a result of making a forecast 12 hours after the update just before the model receives a new observation. The QdV run with a 1-hour interval has a slightly higher skill compared to the 12-hour interval for both the CRPS and AROC (100 l s^{-1}). However, for a lower threshold of 10 l s^{-1} the QdV with a 12-hour interval has a better score than the 1-hour interval simulation (QdV 1h). This shows that there is no direct relation between the updating interval and model performance for this period.

4.5.4 Observation uncertainty

From the updated storage deficit and discharge was found that the standard deviation is an important parameter when using the Ensemble Kalman Filter. Khaki et al. (2017) also found that changing the observation uncertainty can alter the results substantially. Therefore, the effect of different standard deviations for the dV observations used for the QdV run is examined. This is done by using the CRPS and AROC for the whole period (Figure 4.8 g and h) and for an precipitation event in September 2013 (Figure 4.7). The standard deviation as 2.5% of the observation shows the highest skill for the CRPS and AROC (100 l s^{-1}). However, with the lower AROC threshold level of 10 l s^{-1} a standard deviation of 5 millimetres for the storage deficit shows more skill.

For the precipitation event in September (Figure 4.7) the forecast with a standard deviation of 15 millimetres for the observation uncertainty estimates the maximum observed peak height the closest. Standard deviations of 25 millimetre, 32 millimetre and as 2.5% overestimate the largest peak while a deviation of 5 millimetre leads to an underestimation. All forecasts have perform better compared to the run without assimilation. The figure also does show that forecast peak height is influenced by storage deficit. Overall, all forecasts with assimilation did perform better than the open loop run.

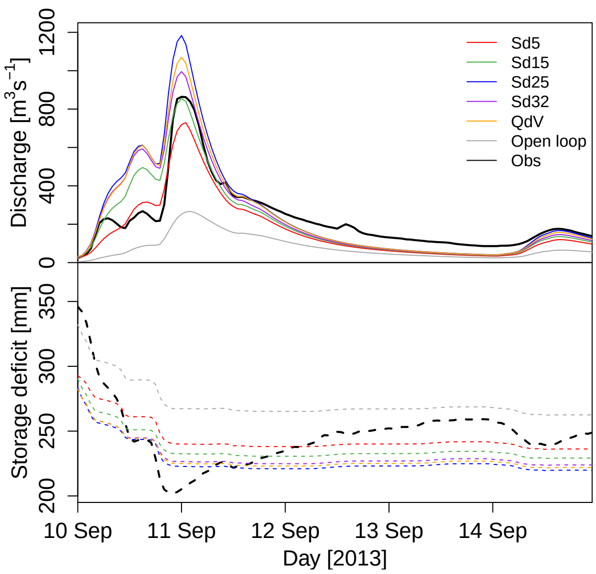


Figure 4.7: Forecast discharge and storage deficit state for different standard deviations for an precipitation event in September 2013. The discharge from the open loop simulation and the observed discharge are given for reference.

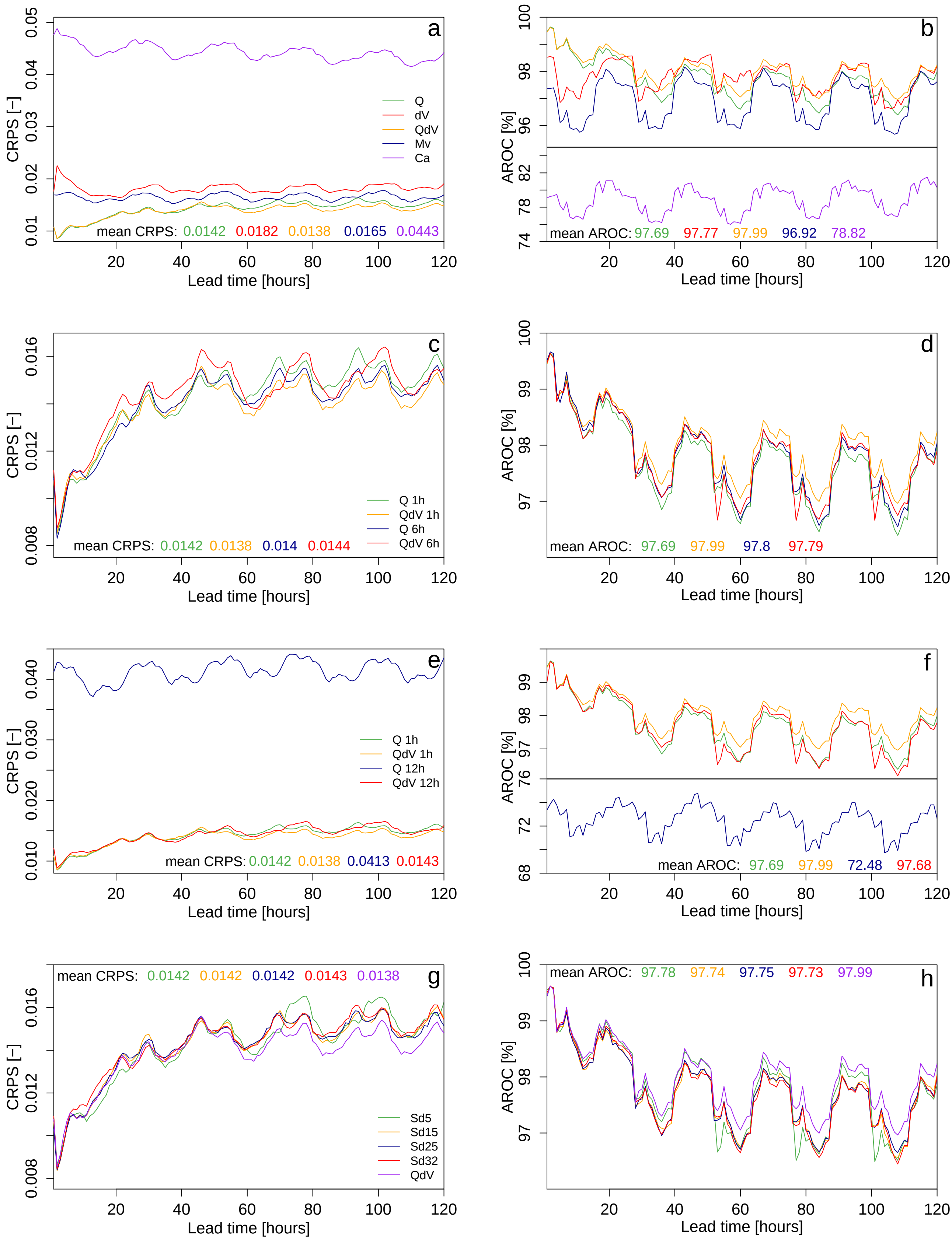


Figure 4.8: CRPS and AROC (100 l s^{-1} threshold) over lead time for the first five runs (a and b), assimilation of Q and Q with dV with a 1 and 6 hour interval for state updating (c and d) and assimilation of Q and Q with dV with a 1 and 12 hour interval for state updating (e and f). Figures g and h show the effect of using different standard deviations for the storage deficit observation when updating both discharge and storage deficit. Note: panels b and f are multipanel plots with different y-axis scales to increase readability.

5 | Discussion

5.1 Groundwater-discharge relation and assimilation efficiency

The relation between groundwater level and discharge largely determines the skill of the forecast when using the groundwater observations for state updating. Groundwater level from the Camping location has a low correlation and large positive lag time with discharge and the run with assimilation of Ca shows the lowest skill of all forecasts. In contrast, the Meteoveld location has the highest correlation and a negative lag time of only 7 hours with discharge and has the highest skill from all forecasts which do not use discharge for assimilation (Figure 4.8a and b). The similar performance of the Mv (groundwater level) and dV (storage deficit) observations is caused by the use of Mv for the observation model. The response time for this catchment is between 5 and 12 hours and may be related to the skill increase by using groundwater assimilation. Valk (2019) studied groundwater and discharge assimilation for the Regge catchment using different filters and found a skill increase when using both states compared to only discharge. This catchment has a similar response time as the Hupsel Brook (7 to 10 hours; Heuvelink et al., 2020). This indicates that response time may not be a suitable characteristic to determine assimilation efficiency. Therefore, it is recommended to estimate forecast skill with groundwater assimilation for catchments with other characteristics, such as soil type, catchment slope and a larger base flow.

5.2 Relation between observations

Response time is estimated with cross-correlation between precipitation and discharge and individual rainfall events. Both methods yielded different results and response time varied depending on catchment wetness which is according to Brauer et al. (2018) the main predictor for discharge response. The cross-correlation analysis showed a response time of 12 hours where the individual events showed a values ranging between 5 and 12 hours. These results differ from Brauer et al. (2011) who also found a response time for this catchment by analysing an extreme event. They found that discharge started to increase after 7 hours and that the peak was reached at 23 hours after the event. The difference in results is likely caused by the used methodology, since Brauer et al. (2011) studied one extreme event in summer and this study focused on the response time over more than 5 years and different wetness conditions.

In addition, forecasts made with the observation model output had lower skill than forecasts with assimilation of Meteoveld time series data. This shows that using one observation to explain a catchment scale average state can be sufficient. However, for larger catchments the observation model can be useful to estimate a catchment averaged state from observed groundwater levels, that still contains a portion of the observed variability.

5.3 Observation model

The observation model is used to convert point observations to a catchment average. Positive lag times (Camping) can be of limited use for making predictions because of the delayed response of groundwater. However, this location is still used for the observation model because, it has the same significance as the Meteoveld and Ten Barge observations. This shows that a weak relation between groundwater level and discharge can still be of value for the observation model.

The difference of 50 millimetre at the start of the prediction for 2013 is caused by the used simulation period and the effects of freezing. During January, February and March of 2013, 41 days with a daily average temperature below zero degrees Celsius are recorded at the Hupsel station (KNMI, 2020). During these days a total of 32.7 millimetres of precipitation was recorded with non-freezing days in between. This can lead to a smaller storage deficit in WALRUS because the snow module was not used and thus precipitation was always assumed to be rain. In contrast, the observed groundwater level can remain lower due to a reduced infiltration capacity by frozen soil. However, this does not explain the difference in storage deficit over a longer period. This is caused by the used simulation period for WALRUS in R for 2015 used to calibrate the observation model. Appendix I shows the observation model result when using a longer spin-up time. This increases the correlation of the observation model further to $R^2 = 0.96$ for both the d_G and d_V state. The root mean square error (RMSE) between WALRUS storage deficit and the observation model for 2013 decreases from 36.87 to 26.35 millimetres when using a simulation period with one year spin-up time.

Two additional observation models were made to estimate both the groundwater level (d_G) and quickflow reservoir level (h_Q) state and from observed groundwater levels. However, the d_G observation model was not used because updating the storage deficit (d_V) state was expected to yield better results, because the groundwater level in WALRUS is calculated with the storage deficit from the vadose zone as input (Brauer et al., 2014a). However, OpenDA updates all states of the model instead of only states for which observations are available. This is different from the method used by Ogilvie (2016) who found that the updated discharge state quickly returned to pre-update levels because the other model states where not updated. Therefore, updating the WALRUS groundwater level with OpenDA can still be useful for assimilation. An attempt was made to use fast groundwater fluctuations for updating the quickflow reservoir of WALRUS. However, this observation model yielded no significant correlation for the validation year due to the zero inflated data from the quickflow reservoir. A more detailed description can be found in Appendix J.

The mean and standard deviation (32.49 millimetre) of the error in storage deficit between the observation model and the WALRUS state were expected to be used as the uncertainty in the observation for FEWS. The error is considered Gaussian and needed a bias correction of -17.77 millimetres. Appendix K shows a histogram with the difference. However, this standard deviation made the observation too uncertain and the observations were not used to update the model due to the model uncertainty being smaller. Therefore, the standard deviation is reduced and the effect of varying the standard deviation is explored.

5.4 FEWS configuration

The WALRUS models for the forecast in FEWS and for the observation model are compared to be sure that they give the same results. However, after making simulations there was a discrepancy between both models at the beginning of 2013. This difference was caused by a reduced the spin-up time of the FEWS simulation. Also, a one-hour shift between both model outputs was found. This can be related to the way WALRUS in R stores the simulated discharge. A graph with the difference is given in appendix L. This error is consistent for all runs which makes them still comparable. However, the score of the open loop simulation is different when using a longer spin-up time for the configuration. This results in a higher NSE of the WALRUS model in FEWS ($NSE = 0.8$) than the R-version used for the observation model ($NSE = 0.72$), see Table 4.3.

The importance of updating groundwater levels can be more related to low flow scenarios. However, due to perturbation of precipitation, the uncertainty of model output becomes very low during periods without precipitation. Therefore, the groundwater is not updated during these situations and the value of groundwater observations for predicting low flows could not be quantified properly. Perturbing other non-zero state(s) instead of precipitation, like potential evapotranspiration or the groundwater state itself, can keep the model uncertainty larger and more constant when no precipitation is available.

The precipitation event starting on the 10th of September 2013 was used to estimate the effect of different observations for assimilation on updated discharge. Sun et al. (2020) studied the updated model output for the Regge over the first 8 days of August in 2006. Their results also show that the open loop (non-updated) model underestimates peak discharge compared to the observation and that the EnKF simulated discharge resembles observed discharge more closely. This shows that discharge assimilation for WALRUS is especially useful for the correction of model states, during a fast (several days) decrease in storage deficit.

5.5 CRPS and AROC

To evaluate forecast performance the area under the Receiver Operator Characteristic curve (AROC) and Continuous Ranked Probability Score (CRPS) over lead time are calculated. The resulting figures are used to compare different states, intervals and standard deviations for assimilation. However, Monte Carlo sampling of precipitation in FEWS, used to estimate uncertainty in model outputs, adds randomness to the results. Therefore, small differences in forecast skill cannot be directly linked to a the changed configuration input.

The AROC and CRPS in Figure 4.8 show a 24-hour periodicity. However, forecast skill is expected to decreases towards longer lead times. Therefore, the used function was rebuilt but no related error in the post-processing steps could be found. So, the periodicity is most likely caused by the calculation over lead time in hours for forecasts made every 24 hours. Sun et al. (2020) used an forecast interval of 2 days with discharge assimilation for the Regge catchment. Their calculated MAE and CRPS with lead time show an 48-hour periodicity resulting in two peaks of skill increase between 1 and 120 hours lead time for the EnKF filter. Therefore, the periodicity in skill over lead time can be related to the used forecast interval. This periodicity can be avoided by using the same temporal scale for skill over lead time and the forecast interval.

The AROC for a threshold of 200 l s^{-1} showed a strong temporary decrease in skill that can be related to the low number of discharge peaks in one year. Therefore, lower AROC thresholds are used which makes the comparison between forecasts still possible. The number of discharge values for the AROC is further reduced by the 24-hour interval discussed before. Therefore, a longer simulation period and forecast interval of 1 hour is needed to estimate the AROC for higher thresholds which are more realistic for flood warning. In addition, another forecast skill estimator such as the CRPS can be used which uses all available forecast data.

5.6 Recommendations

WALRUS is a rainfall-runoff model that fills the gap between parametric and distributed models. It is capable of modelling multiple states including surface water level and related discharge, groundwater level and storage deficit. However, hydrological models calibrated on discharge can show high efficiency while the other model states are vastly different from the observations. This depends on model structure and the degree in which physical processes are incorporated. This has been studied for WALRUS by Brauer et al. (2014b) and it was found that WALRUS estimates other states such as groundwater level and unsaturated zone processes well. WALRUS can also be calibrated on groundwater levels. Therefore, updating the WALRUS groundwater and storage deficit states is possible. However, results may differ with other hydrological models because of differences in model structure and the representation of physical processes.

The forecast skill when only using discharge for assimilation is already high. This means that skill increase is limited when also using groundwater level observations. This can be caused by the fast response time of this catchment. It is possible that the effect of groundwater assimilation can be more substantial in catchments that have a slower response time and are more dependent on groundwater (larger baseflow). Valk (2019) found that a combination of storage deficit and groundwater level performed the best when updating two states for the Regge catchment (1015 km^2). Therefore, the small increase in forecast skill for the Hupsel catchment can be significant and show that groundwater assimilation does increase forecast skill.

Initially, the Groenlose Slinge catchment in the management area of Water Authority Rijn en IJssel was also chosen to estimate assimilation effectivity. For this 194 km^2 large catchment a response time between 17 and 41 hours was found. Also, other catchment characteristics such as catchment slope, soil type and size may affect assimilation performance. Unfortunately, building the FEWS configuration for the Groenlose Slinge was not possible due to time constraints. However, the variable relations and observation model results are added in Appendix M. This combined with the results of previous studies (Valk, 2019; Ogilvie, 2016) shows that this method can be used to convert point measurements into a catchment average.

6 | Conclusion

In this study the effect of using groundwater observations for data assimilation on flow forecast skill is investigated. Forecasts that use different observations, update intervals and observation uncertainties for assimilation are compared. Results show that assimilating groundwater in addition to assimilating discharge only results in a small increase in forecast skill. A possible explanation for this can be the fast response time and high temporal variation in groundwater level of this catchment. In addition, the already high forecast skill with only discharge assimilation likely limits skill increase. Forecast skill remains more constant over lead time when using only groundwater observations compared to discharge. Combining groundwater and discharge gives a higher overall skill that is more constant over lead time. This is confirmed by former studies were also a significant skill increase was found when assimilating groundwater in addition to assimilating discharge. However, the skill increase of groundwater assimilation for the Hupsel Brook catchment is uncertain because the small increase in forecast skill can be a result of Monte Carlo simulations. Therefore, it is recommended to estimate the potential of groundwater assimilation for other lowland catchments with different characteristics and a longer simulation period (more than one year).

Groundwater level observations with the strongest correlation with discharge yielded the highest forecast skill when used for assimilation. In addition, an observation model was established to convert the measurements into a catchment average state that can be used for updating the model. Using the output from this model for assimilation yielded similar results as the groundwater time series that has the highest correlation with discharge. This can be related to the small size of the used catchment and indicates that only using the groundwater observations with the strongest correlation with discharge could be sufficient for smaller catchments. In this case, the correlation and cross-correlation between discharge and groundwater level can be used as predictor for forecast skill. However, these results show that the observation model is capable of representing the model states used for assimilation with observed groundwater levels which can be useful for (larger) catchments with more spatial variability.

Acknowledgements

First I want to thank my supervisors Claudia Brauer and Albrecht Weerts for their supervision, support with WALRUS and FEWS and scripts, supplying literature and data and pointing out possible research directions. I also want to thank Eoin Burke from Water Authority Rijn en IJssel for the FEWS-OpenDA configuration, explanation of the configuration and help with setting up the configuration for the Hupsel Brook catchment. I would like to thank Water Authority Rijn and IJssel for sharing measurement data. I also want to thank the students that attended the thesis ring for their constructive feedback on sections of this report. Finally, I want to thank my parents for their support and a place to work when I could not work on campus due to the COVID-19 restrictions.

Bibliography

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., mar 2013. GloFAS – global ensemble streamflow forecasting and flood early warning. Hydrology and Earth System Sciences 17 (3), 1161–1175.

Alfieri, L., Thielen, J., Pappenberger, F., mar 2012. Ensemble hydro-meteorological simulation for flash flood early detection in southern switzerland. Journal of Hydrology 424-425, 143–153.

Aubert, D., Loumagne, C., Oudin, L., sep 2003. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model. Journal of Hydrology 280 (1-4), 145–161.

Brauer, C., van der Velde, Y., Teuling, A., Uijlenhoet, R., 2018. The hupsel brook catchment: Insights from five decades of lowland observations. Vadose Zone Journal 17 (1), 180056.

Brauer, C. C., Overeem, A., Leijnse, H., Uijlenhoet, R., jun 2016. The effect of differences between rainfall measurement techniques on groundwater and discharge simulations in a lowland catchment. Hydrological Processes 30 (21), 3885–3900.

Brauer, C. C., Teuling, A. J., Overeem, A., van der Velde, Y., Hazenberg, P., Warmerdam, P. M. M., Uijlenhoet, R., jun 2011. Anatomy of extraordinary rainfall and flash flood in a dutch lowland catchment. Hydrology and Earth System Sciences 15 (6), 1991–2005.

Brauer, C. C., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., oct 2014a. The wageningen lowland runoff simulator (WALRUS): a lumped rainfall–runoff model for catchments with shallow groundwater. Geoscientific Model Development 7 (5), 2313–2332.

Brauer, C. C., Torfs, P. J. J. F., Teuling, A. J., Uijlenhoet, R., oct 2014b. The wageningen lowland runoff simulator (WALRUS): application to the hupsel brook catchment and the cabauw polder. Hydrology and Earth System Sciences 18 (10), 4007–4028.

Buizza, R., Hollingworth, A., Lalaurette, F., Ghelli, A., 1999. Probabilistic predictions of precipitation using the ecmwf ensemble prediction system. Wea. Forecasting 14, 168–189.

Chen, F., Crow, W. T., Starks, P. J., Moriasi, D. N., apr 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. Advances in Water Resources 34 (4), 526–536.

Christensen, O., Christensen, J., dec 2004. Intensification of extreme european summer precipitation in a warmer climate. Global and Planetary Change 44 (1-4), 107–117.

Dankers, R., Feyen, L., oct 2008. Climate change impact on flood hazard in europe: An assessment based on high-resolution climate simulations. Journal of Geophysical Research 113 (D19).

Deltares, Apr. 2018. SOBEK, User Manual. Available at url: www.deltares.nl/en/software/sobek. Deltares, Bousinesqweg 1, Delft, version revision 55205 Edition. URL `www.deltares.nl`

Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., oct 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377 (1-2), 80–91.

He, X., Lucatero, D., Ridler, M.-E., Madsen, H., Kidmose, J., Hole, Ø., Petersen, C., Zheng, C., Refsgaard, J. C., may 2019. Real-time simulation of surface water and groundwater with data assimilation. Advances in Water Resources 127, 13–25.

Hersbach, H., oct 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15 (5), 559–570.

Heuvelink, D., Berenguer, M., Brauer, C. C., Uijlenhoet, R., mar 2020. Hydrological application of radar rainfall nowcasting in the netherlands. Environment International 136, 105431.

Izeboud, P., 2017. Reducing water level prediction uncertainty in a delfland polder using data assimilation. Master's thesis, TU Delft.

Kalman, R. E., mar 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82 (1), 35–45.

Khaki, M., Ait-El-Fquih, B., Hoteit, I., Forootan, E., Awange, J., Kuhn, M., dec 2017. A two-update ensemble kalman filter for land hydrological data assimilation with an uncertain constraint. Journal of Hydrology 555, 447–462.

KNMI, May 2020. Knmi daggegevens van het weer in nederland. url: https://www.knmi.nl/nederland-nu/klimatologie/daggegevens (acessed on may 20, 2020). website.

Knoben, W. J. M., Freer, J. E., Woods, R. A., oct 2019. Technical note: Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. Hydrology and Earth System Sciences 23 (10), 4323–4331.

Komma, J., Blöschl, G., Reszler, C., aug 2008. Soil moisture updating by ensemble kalman filtering in real-time flood forecasting. Journal of Hydrology 357 (3-4), 228–242.

Lee, H., Seo, D.-J., Koren, V., dec 2011. Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states. Advances in Water Resources 34 (12), 1597–1615.

Liu, Y., Weerts, A. H., Clark, M., Franssen, H.-J. H., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., Restrepo, P., oct 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences 16 (10), 3863–3887.

McCuen, R. H., Knight, Z., Cutter, A. G., nov 2006. Evaluation of the nash–sutcliffe efficiency index. Journal of Hydrologic Engineering 11 (6), 597–602.

McMillan, H. K., Hreinsson, E. ., Clark, M. P., Singh, S. K., Zammit, C., Uddstrom, M. J., jan 2013. Operational hydrological data assimilation with the recursive ensemble kalman filter. Hydrology and Earth System Sciences 17 (1), 21–38.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J. H., Canty, M., Vereecken, H., mar 2011. Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. Journal of Hydrology 399 (3-4), 410–421.

Nash, J., Sutcliffe, J., apr 1970. Riverflow forecasting through conceptual models part i-a discussion of principles. J. Hydrol. 10.3 (3), 282–290.

Ogilvie, R., 2016. Application of data assimilation to the walrus model in the reusel catchment. Master's thesis, Wageningen University and Research.

Rakovec, O., Weerts, A. H., Hazenberg, P., Torfs, P. J. J. F., Uijlenhoet, R., sep 2012. State updating of a distributed hydrological model with ensemble kalman filtering: effects of updating frequency and observation network density on forecast accuracy. Hydrology and Earth System Sciences 16 (9), 3435–3449.

Reichle, R. H., nov 2008. Data assimilation methods in the earth sciences. Advances in Water Resources 31 (11), 1411–1418.

Rijn, IJssel, May 2020. Waterdata meetdata van kaart. url: Waterdata.wrij.nl. acessed on 18-05-2020.

Ritter, A., Muñoz-Carpena, R., feb 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. Journal of Hydrology 480, 33–45.

Sun, Y., Bao, W., Valk, K., Brauer, C. C., Sumihar, J., Weerts, A. H., 2020. Improving flow forecasting skill of lowland hydrological models using ensemble kalman filter and unscented kalman filter. AGU journal.

Valk, K., Mar. 2019. State updating for improving flood forecast accuracy in lowland catchments with walrus. Master's thesis, Wageningen University and Research.

van den Hurk, B., Tank, A., Lenderink, G., van Ulden, A., van Oldenborgh, G., Katsman, C., van den Brink, H., Keller, F., Bessembinder, J., Burgers, G., Komen, G., Hazeleger, W., Drijfhout, S., aug 2007. New climate change scenarios for the netherlands. Water Science and Technology 56 (4), 27–33.

van Vliet, M. T., Franssen, W. H., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., Kabat, P., apr 2013. Global river discharge and water temperature under climate change. Global Environmental Change 23 (2), 450–464.

Verschaeren, P. J., jan 2015. Het scheiden van langzame en snelle regen-afvoerprocessen in grondwaterdynamiek (the separation of quick and slow rainfall-runoff processes in groundwater dynamics). Bsc Thesis.

Weerts, A., van Osnabrugge, B., 2020. OpenDA for data assimilation within a Delft-FEWS system. Deltares.

Weerts, A. H., Serafy, G. Y. H. E., sep 2006. Particle filtering and ensemble kalman filtering for state updating with hydrological conceptual rainfall-runoff models. Water Resources Research 42 (9).

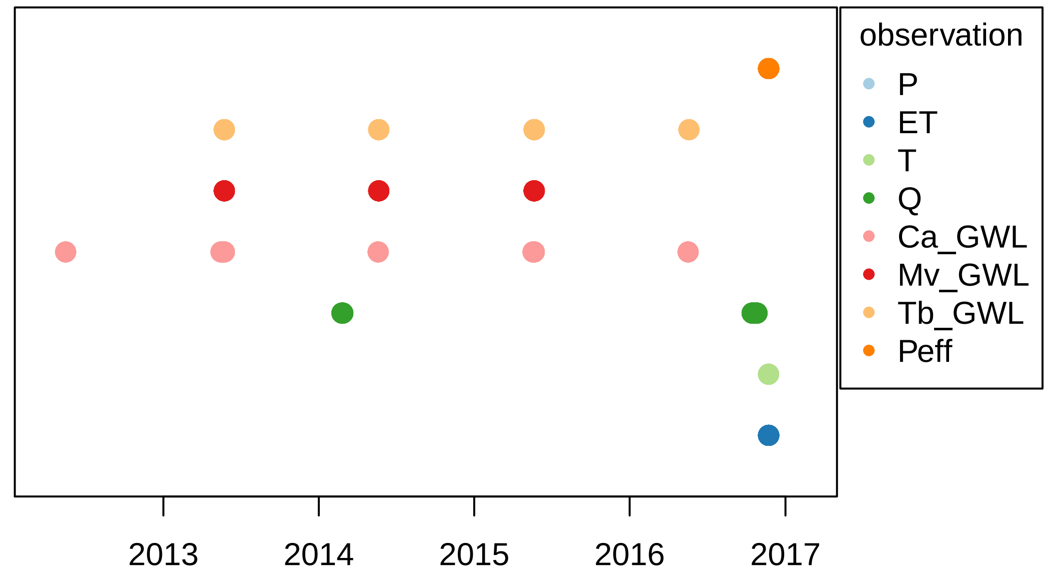
Werner, M., Schellekens, J., Gijsbers, P., van Dijk, M., van den Akker, O., Heynert, K., feb 2013. The delft-FEWS flow forecasting system. Environmental Modelling & Software 40, 65–77.

Zhang, D., Madsen, H., Ridler, M. E., Kidmose, J., Jensen, K. H., Refsgaard, J. C., oct 2016. Multivariate hydrological data assimilation of soil moisture and groundwater head. Hydrology and Earth System Sciences 20 (10), 4341–4357.

A | Used abbreviations

AEnKF	Asynchronous Ensemble Kalman Filter
AROC	Area under the Receiver Operator Characteristic curve
Ca	Camping, name of groundwater level measurement location
CRPS	Continuous Ranked Probability Score
EnKF	Ensemble Kalman Filter
(Delft-)FEWS	Flood Early Warning System
KNMI	Royal Netherlands Meteorological Institute
MAE	Mean absolute error
Mv	Meteoveld, name of groundwater level measurement location
RMSE	Root Mean Square Error
Tb	Ten Barge, name of groundwater level measurement location
WALRUS	WAgeningen Lowland RUnoff Simulator

B | Missing observation data



Data gaps for meteorological, discharge and groundwater data between January 2012 and May 2017 (including effective precipitation). Note: the precipitation data does not contain gaps.

C | Calculation of CRPS and AROC in R

This appendix gives a more detailed description of the calculation of the CRPS and AROC scores from forecast and observation data. The forecasts from FEWS are converted from NetCDF to CSV format by using the "Dataset" function from the Python package "netCDF4". These CSV files with the five day forecasts are then loaded in R. The following sections give a description of the steps needed for the calculation.

CRPS

This describes the methodology used to calculate the CRPS scores from the FEWS forecast outputs and observed discharge for a given lead time.

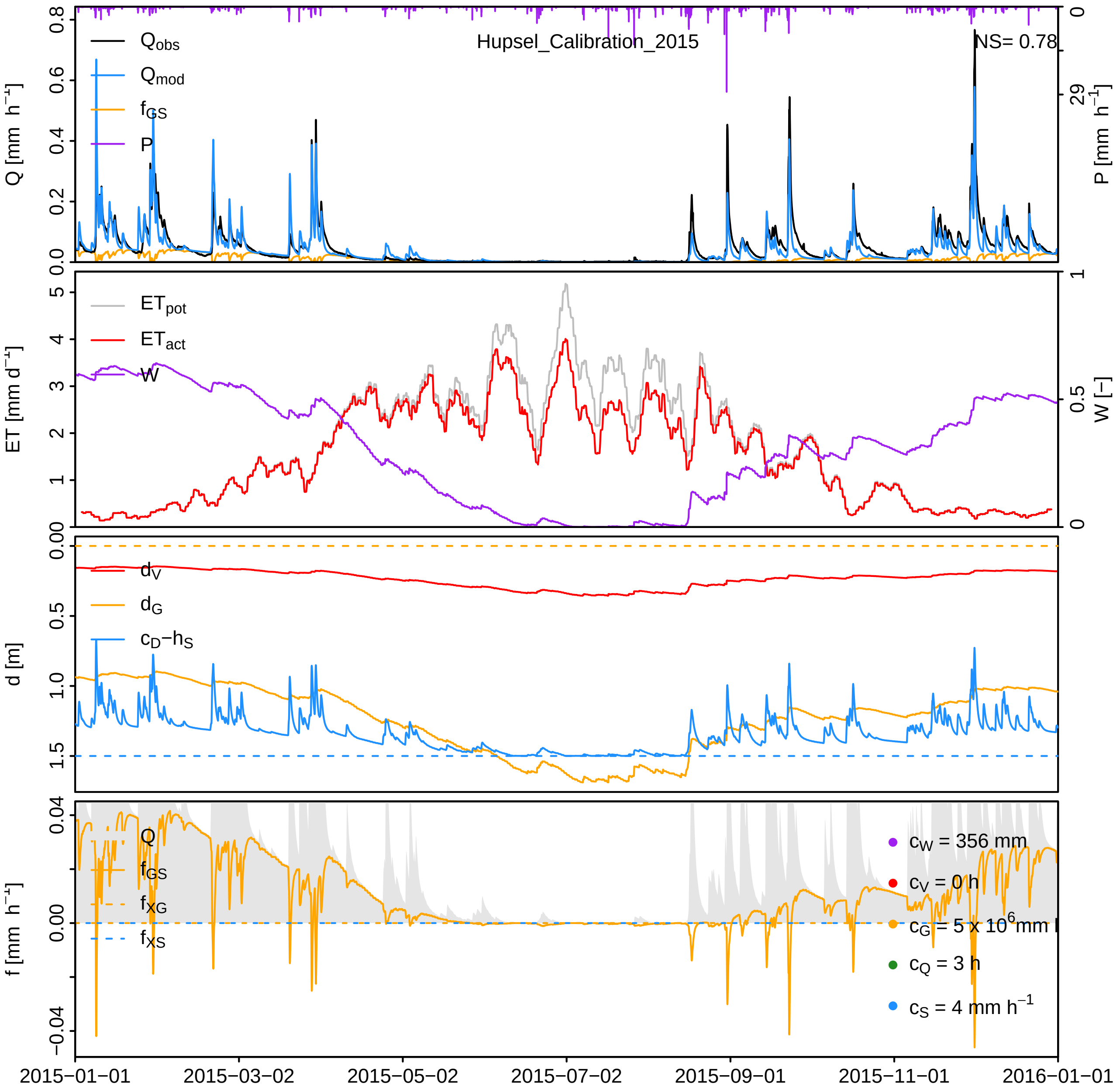
1. Determine mean and standard deviation of the ensembles for each forecasted hour for each forecast.
2. Store the means and standard deviations in two separate matrices with as rows the forecasted hour and as columns the means or standard deviations of the five day forecast. This results in two matrices with with 120 rows (5 day forecast with 1 hour interval) and 361 columns (forecasts made in one year).
3. Create a similar matrix for the observations by selecting 5 days of hourly data starting at the beginning of the forecast and placing this in a column. The next column is filled with similar data but shifted 24 hours (forecasting interval) to line up with the forecasting data.
4. Select the observed and forecasted data for a given lead time with the row index and store the observed discharge in a vector and the forecasted data in a matrix with two vectors containing the mean and standard deviations.
5. Use the vector with observed discharges and the matrix with the mean and standard deviation for the "crps" function of the "Verification" R-package to calculate the CRPS.

AROC

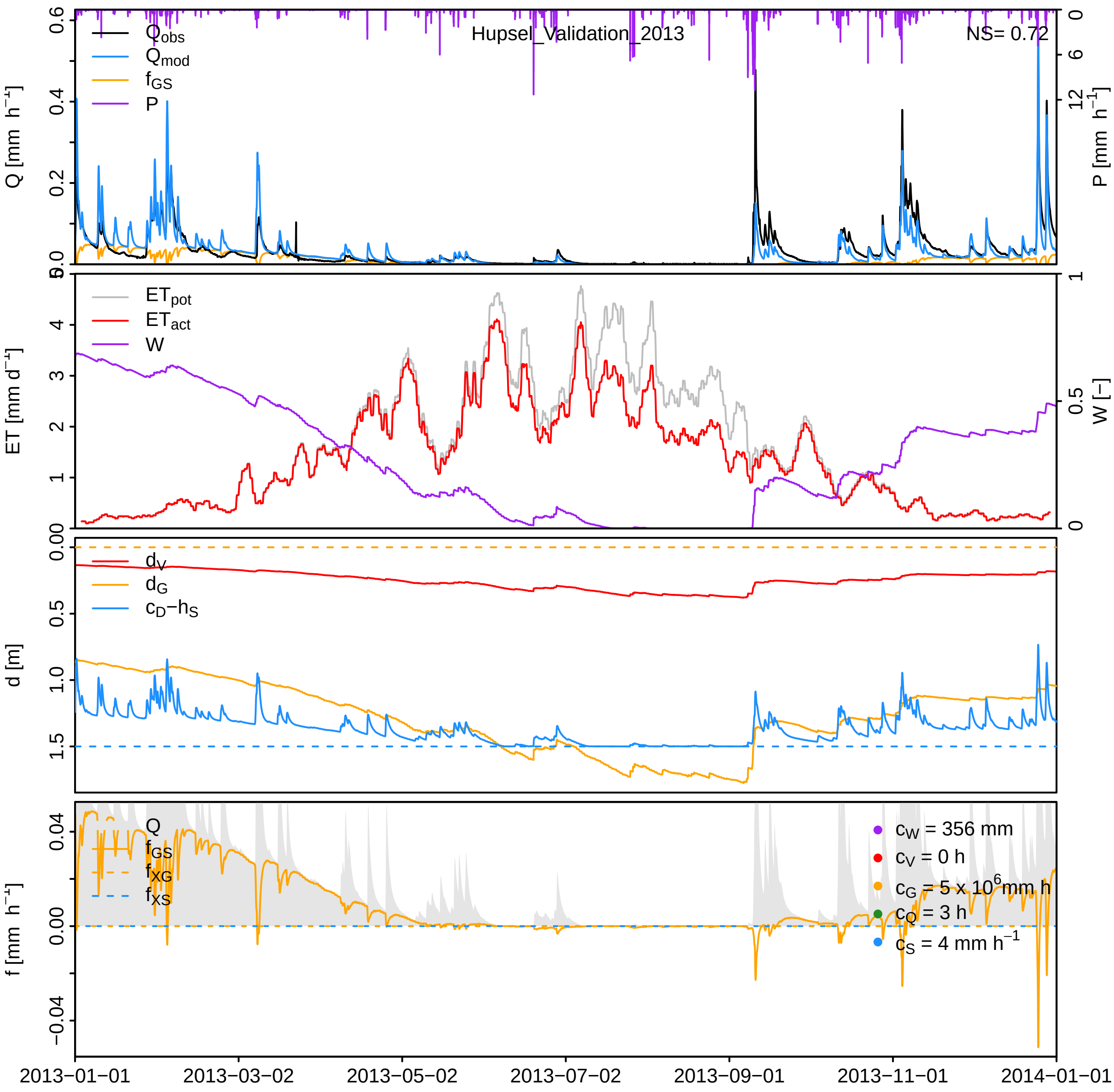
This section describes the methodology used to calculate the AROC scores from the FEWS forecast outputs and observed discharge for a given lead time.

1. Determine mean of the ensembles for each forecasted hour for each forecast.
2. Store the means in a matrix with as rows the forecasted hour and as columns the means or standard deviations of the five day forecast. This results in a matrix with 120 rows (5 day forecast with 1 hour interval) and 361 columns (forecasts made in one year).
3. Create a similar matrix for the observations by selecting 5 days of hourly data starting at the beginning of the forecast and placing this in a column. The next column is filled with similar data but shifted 24 hours (forecasting interval) to line up with the forecasting data.
4. Select the observed and forecasted data for a given lead time with the row index and store the observed discharge and the ensemble means of the forecasted discharge in two vectors.
5. Use the two vectors with observed discharges and the vector with the ensemble means for the "ROC" function of the "pROC" R-package to calculate the AROC with a graph or with the "AUC" if only the area under the curve is needed.

D | WALRUS model Calibration and Validation



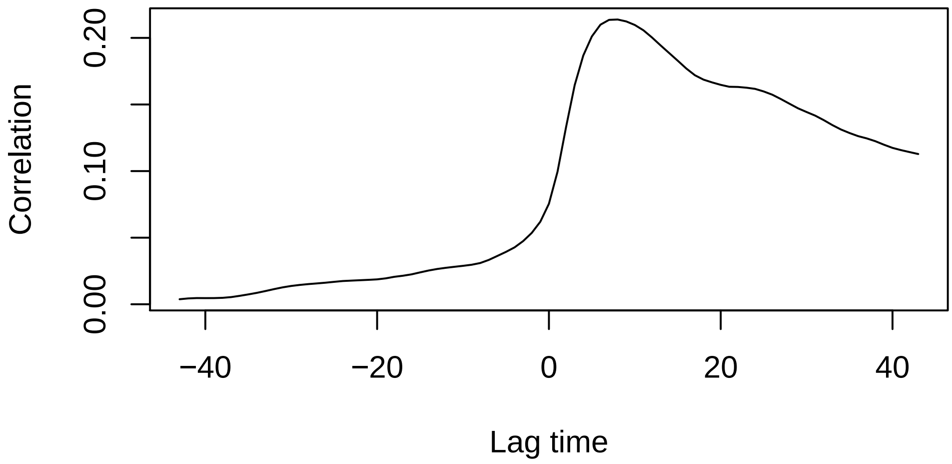
WALRUS model calibration output for 2015 for the Hupsel brook. The c_W, c_V, C_G and c_Q parameter values are optimized with the Levenberg-Marquardt algorithm. The resulting d_V, d_G and h_S states are used to calibrate the observation model.



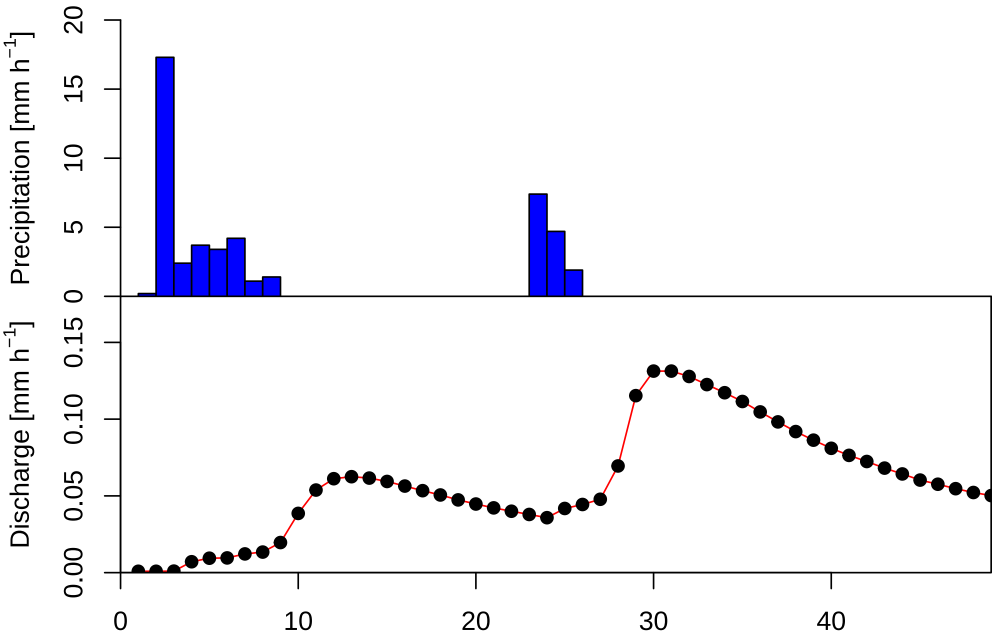
WALRUS model validation output for 2013 for the Hupsel brook catchment. The d_V state output is used to estimate the observation model efficiency.

E | Estimation of response time

The response time is estimated with cross-correlation between discharge and precipitation and with multiple events by comparing the delay between precipitation and discharge peaks.



Cross-correlation between discharge and precipitation. The lag at which maximal correlation occurs is assumed to be the response time in hours.



One example of a visual comparison of the delay between discharge and precipitation between the 28th and 30th of July 2013.

F | Improving the observation model

The accuracy of the observation model can possibly be increased by decreasing the hysteresis and by reducing the variability of the observation. This has been tried by using a weighted moving average function to decrease variability, using a rolling average function and by shifting the groundwater observations with the lag time at which correlation is at maximum. This analysis is carried out for the period between November 2013 and April 2014.

Linear model with rolling average of observations

This method uses the rolling average function "rollmean" in R to calculate the rolling average over a given interval. The rolling averages of the Meteoveld and Ten Barge observations are calculated with a 1 hour resolution and used as explanatory variables in a linear model that has storage deficit from WALRUS as dependent variable. The output from the linear model had a correlation of $R^2 = 0.90$ over these 5 months with a rolling average window of 250 hours backwards in time (forwards not possible for forecast). This was an increase when compared to the normal observation model ($R^2 = 0.72$). However, this method also removes a lot of variability which could be needed for assimilation. Also, it introduces extra parameters for the averaging window which may be affected by catchment conditions that vary over time (wetness).

Comparison of moving average and lagged moving average

For the given period the moving average and lagged moving weighted average from different locations is compared with the WALRUS output. The correlations for this analysis are given in the table below. The lag of Ca is not used because the groundwater level reacts slower than discharge. The correlation was the highest with the moving weighted average of the lagged observations from Meteoveld. However, this methodology is not used for the final observation model for several reasons. The use of the MWA function introduces two extra parameters to the observation model (averaging window and weight) which is not preferred when taking in account the limited increase of accuracy. Also, the lag time varies over the year due to different degrees of wetness and therefore has to be estimated again for each new simulation period.

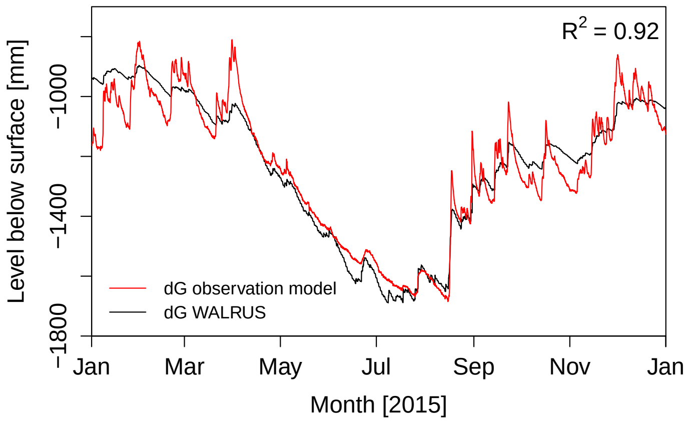
Conclusion

Due to the extra parameters needed for these methodologies and the limited increase in correlation between observed and modeled groundwater these methods are not used for the final observation model. The final model uses only the observed groundwater levels without any modifications for the linear model. However, These results show that is may be possible to increase the correlation between modeled and observed groundwater by using the moving weighted average function rolling average (also found by Verschaeren (2015)) and by shifting data with the lag time when maximal correlation occurs.

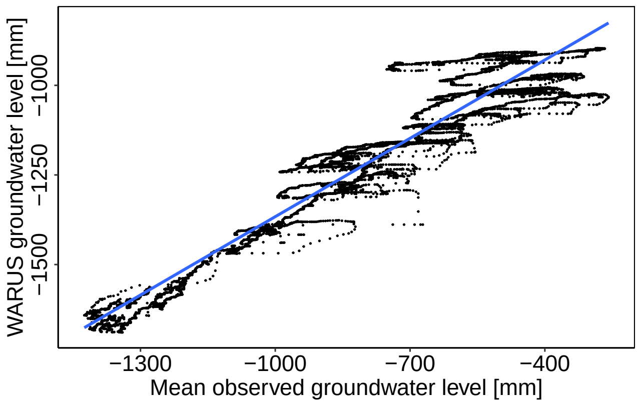
A table with the correlation between WALRUS groundwater level and observed groundwater level.

Observation	R^2
WALRUS	1.00
Ca	0.47
Tb	0.67
Mv	0.82
MWA Tb	0.74
MWA Mv	0.87
MWA Tb lagged	0.78
MWA Mv lagged	0.90

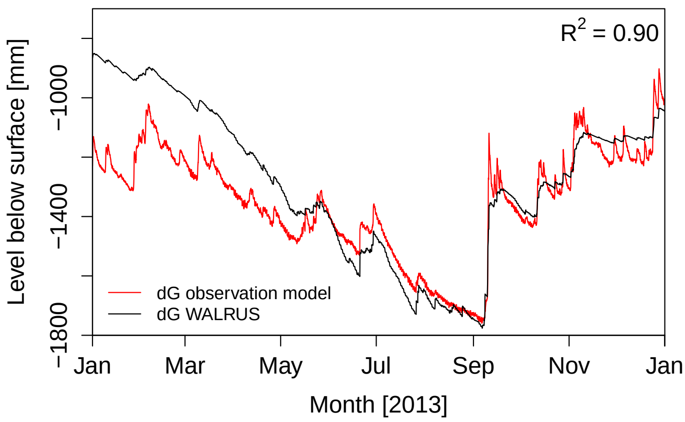
G | Observation model for groundwater state



Fitted observation model for the calibration year (2015) with simulated groundwater level by WALRUS.

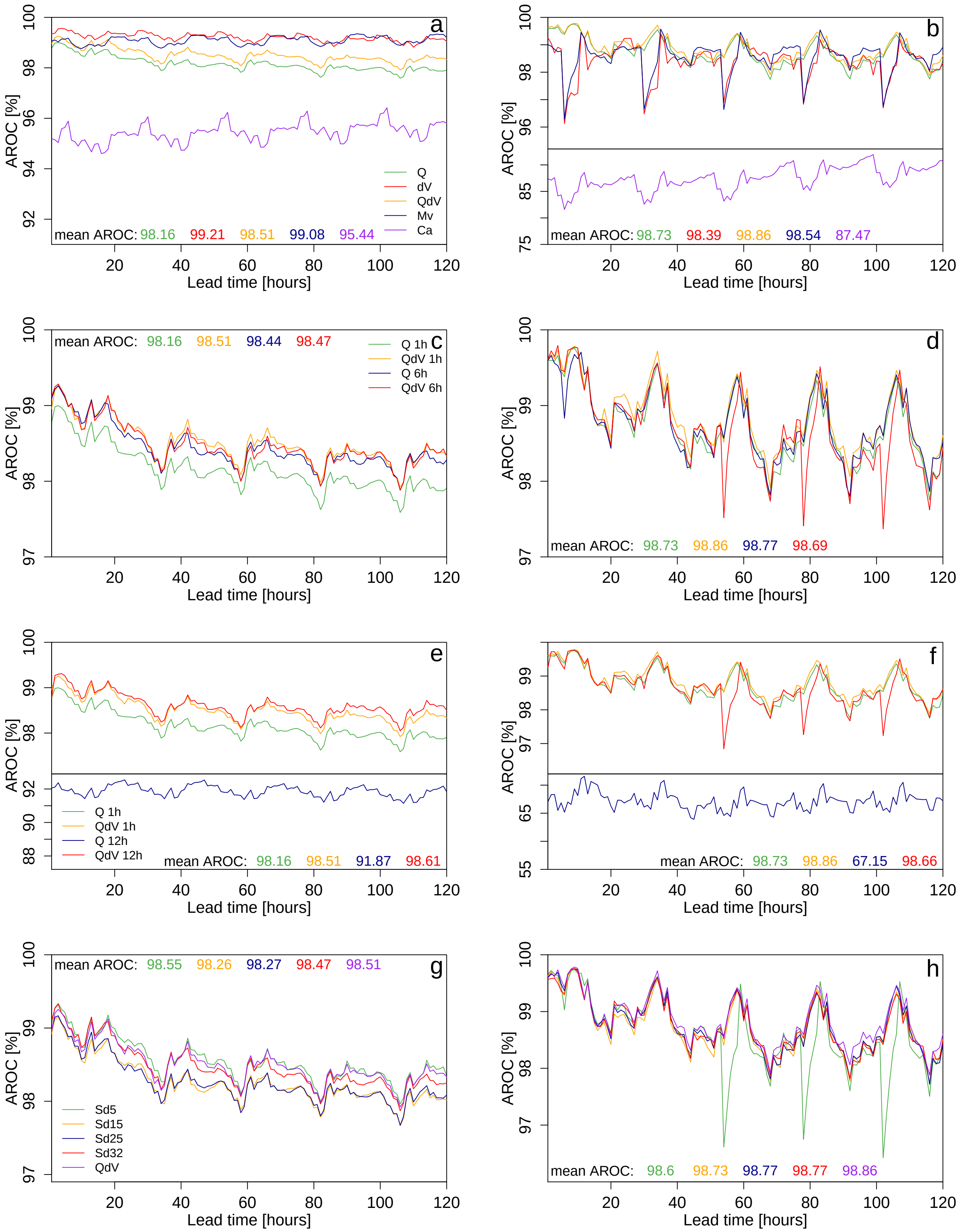


Calibration of observation model used to estimate the d_G state with observed groundwater level. The model is calibrated with WALRUS groundwater level output (d_G) for 2015.



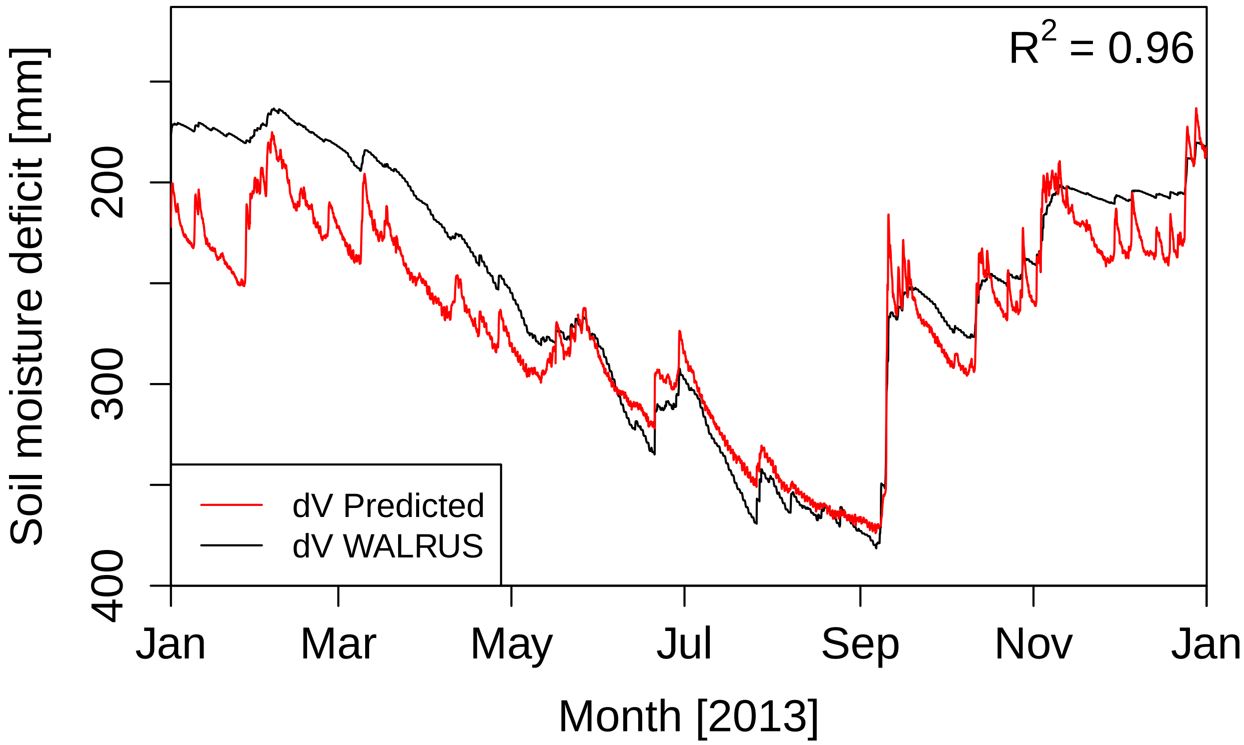
Predicted groundwater state for WALRUS by observation model for 2013.

H AROC for 10 and 200 l s^{-1}



AROC over lead time for an alarm level of 10 (first column) and 200 l s^{-1} (second column) for the first five runs (a and b), assimilation of Q and Q with dV with a 1 and 6 hour interval for state updating (c and d) and assimilation of Q and Q with dV with a 1 and 12 hour interval for state updating (e and f). The effect of changing the standard deviation for d_V is given in sub-figures g and h.

I | Observation model with longer spin-up time



Observation model for d_V with a longer WALRUS run reducing the effect of the initial states.

J | Observation model for quickflow reservoir

At first was attempted to separate slow and fast processes in groundwater into seasonality and faster fluctuations. For this the relation between groundwater level and the quickflow reservoir level was estimated for the year 2013 for the Hupsel Brook catchment. However, the correlation was low ($R^2 = 0.14$) due to the large amount of zero's in the quickflow reservoir level data. When adding the effective precipitation (P-ETpot) to the linear model the correlation increased to ($R^2 = 0.30$. It was found that the level of the quickflow reservoir is not directly linked to groundwater levels but rather to the change in groundwater level.

several methods have been tried to estimate this relation:

1. Using a Poisson model
2. Using the change in groundwater level
3. Using a function that separates the fast and slow groundwater fluctuations

These methodologies will be explained in more detail in this section.

Poisson model

A poisson model was fitted to account for the zero's in the quickflow reservoir data. A Poisson regression model can deal with zero inflated data by assuming that these are generated by a different non-related process. However, this is not the case with reservoir levels which are always dependent on the same processes. Also, the Poisson regression is mainly used for count data without an internal relation. This is not the case for a reservoir because the reservoir state is dependent on the previous state and is subjected to hysteresis. When using the Poisson model the correlation increased to 0.55 when using: Ca, Mv, Tb, Peff ($P - PET$) and P.

Groundwater level change

The change in groundwater level is estimated by making a function that subtracts the groundwater level at time $n+1$ over the groundwater level of time n. This change in groundwater level was used to estimate the relation with quickflow. However, this resulted in poor correlation ($R^2 = 0.06$).

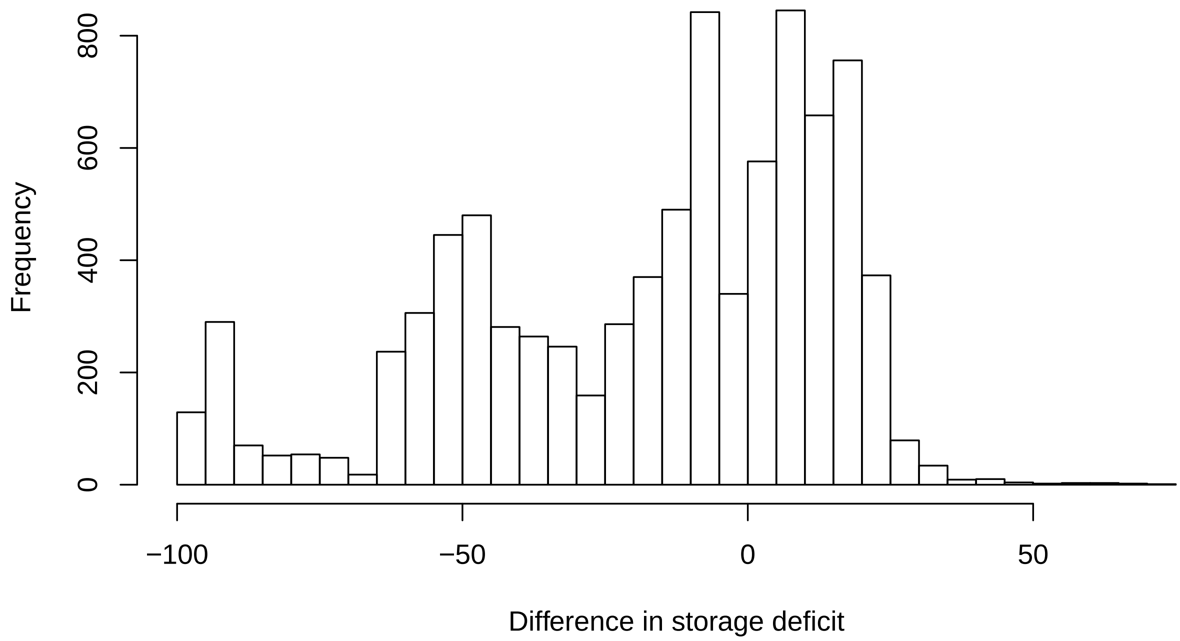
Function for separation of variability

Finally a function was created that uses the observation with the highest correlation with discharge and subtracts the moving average from that observation to reside with only the fast fluctuations. This function needs the observed groundwater data, the modelled hQ for comparison and the averaging time and weight for the weighted average. The positive fluctuations (rise in groundwater level) are used and fitted using the correlation and RMSE by adjusting the weight and averaging time of the moving average function. This resulted in a relatively good fit comparable to the poisson function ($R^2 = 0.52$, RMSE $= 0.72$). However, there are important downsides to this method:

1. It fully describes h_Q by the observed groundwater level, however WALRUS also uses runoff from precipitation for this based on the wetness parameter.
2. The units do not work out because the estimated h_Q state value with the function is two orders of magnitude larger.
3. Extra parameter values are needed which have to be estimated for other simulation periods which is not desired for prediction purposes.
4. All data is used to create the fit which means that it cannot be used for predictions unless this relation is valid for more simulation periods and the correlation is high with the calibration dataset.
5. Making a function to estimate h_Q is not possible because this variable is dependent on the W variable in WALRUS which is dependent on other model variables.

For these reasons the fast fluctuations in groundwater levels are not used for state updating of the hQ reservoir.

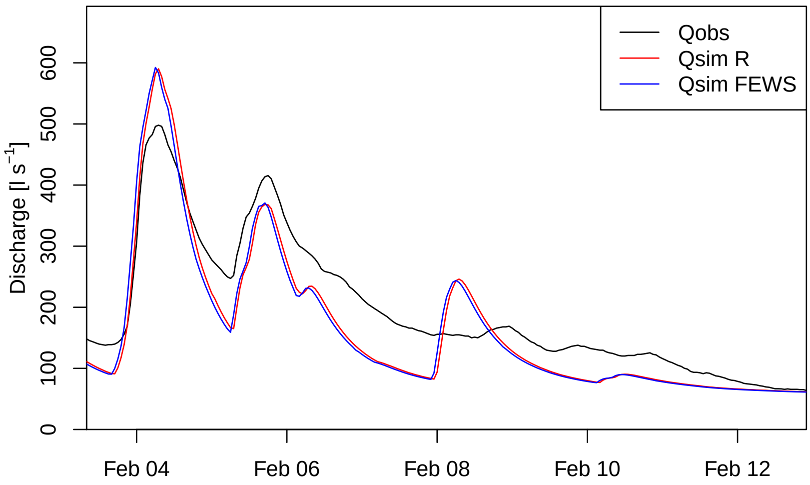
K | Storage deficit difference observation model and WALRUS



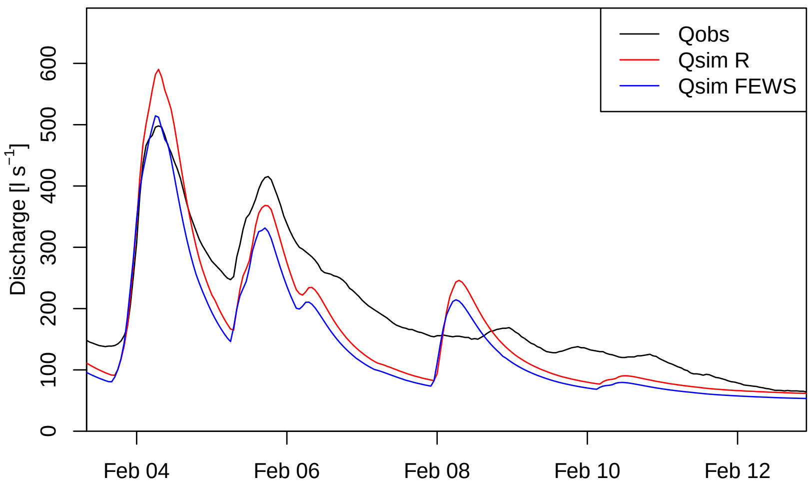
Difference between simulated storage deficit and estimated storage deficit by the observation model for 2013. The data in this figure is not bias corrected.

L | Difference in WALRUS discharge between FEWS and R

The difference between R and SOBEK discharge output is caused by the short spin-up time used for the SOBEK simulation. For earlier simulations a 2 month spin-up period was used for WALRUS in SOBEK which yielded almost the same result as the output from WALRUS in R (Figure **??**). It was found that the small difference (1 hour shifted) can be caused by the way WALRUS in R stores the output. However, the decreased spin-up time for the model runs used in this study resulted in the large difference as seen in Figure **??**. Unfortunately, no time was available to redo these simulations.



This figure shows the modeled discharge of WALRUS from SOBEK and WALRUS from R with the same model parameters for an event in September 2013 with a 2 month spin-up time.



This figure shows the modeled discharge of WALRUS from SOBEK and WALRUS from R with the same model parameters for an event in February 2013.

M | Relations and observation model Groenlose Slinge

Used data

The groudwater observations are selected based on their proximity to the Groenlose Slinge river and data availability. The used data is measured between 28 October 2018 and 20 October 2020. Five groundwater locations are used: Beltrum Kempersweg (Bk), Mulliersweg Winterswijk (Mw), Walienseweg Huppel (Wh), Oude Schooldijk Stelkampsveld (Os) and Wiechersweg Stelkampsveld (Sw). Discharge of the Hagenbeekbrug location near Borculo is used. For precipitation an average precipitation was calculated using Voronoi or Thiessen polygons because radar data from the KNMI was temporary unavailable. For this the following stations were used: Harreveld (13.6% of catchment area), Hengelo (GLD;5%), Hupsel (29.5%), Ratum-Henxel (39.2%) and VredenKA (Germany; 12.8%). The potential evapotraspiration is estimated from the Hupsel station similar to the analysis for the Hupsel Brook catchment. The discharge from the water treatment plant of Winterswijk is not used because it did not yield a significant increase in model efficiency.

Variable relations

The response time of the Groenlose Slinge catchment is estimated to be 21 hours based on lag time from cross-correlation between precipitation and discharge.Also, the delay between certain precipitation events and resulting discharge peaks are evaluated. This showed a variable response time of 17 to 41 hours. The correlation between discharge and the groundwater observations ranges between 0.78 - 0.84 for the five selected groundwater observations. The cross-correlation lag times, between discharge and groundwater level, for the locations are: 18 hours (Sw), 2 hours (Mw), 0 hours (Kb) and -4 hours (Os and Wh). This shows that most groundwater locations have a delayed response compared to discharge.

WALRUS model

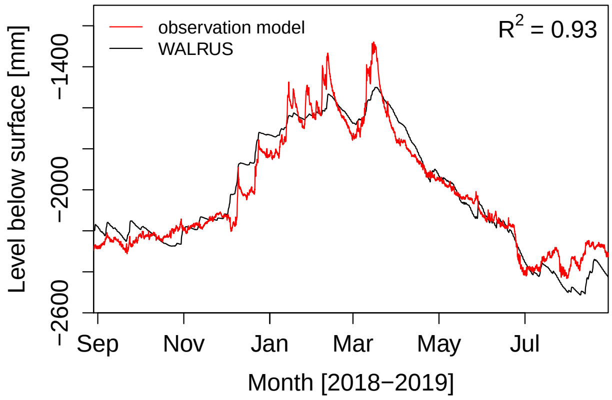
The parameters for the WALRUS model are given in the table below. The G_{frac} calibration parameter was set to 0.9 for auto-calibration. The model is calibrated by using one year of data from October 2018 to October 2019. The NSE value for simulated discharge over the calibration period was 0.94. The model is validated with observed discharge from October 2019 to October 2020 which yielded a NSE score of 0.78. A catchment area of 188 km^2 is used with the default $Q - h_S$ relation.

This table shows the WALRUS parameter values for the Groenlose Slinge catchment.

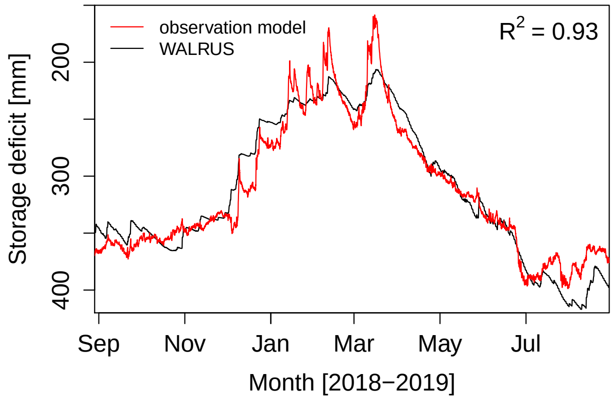
Par	cW	cV	cG	cQ	cS	cD	aS	st
Val	366	1	1.25e6	30	1.6	2300	0.01	loamy sand

Observation model

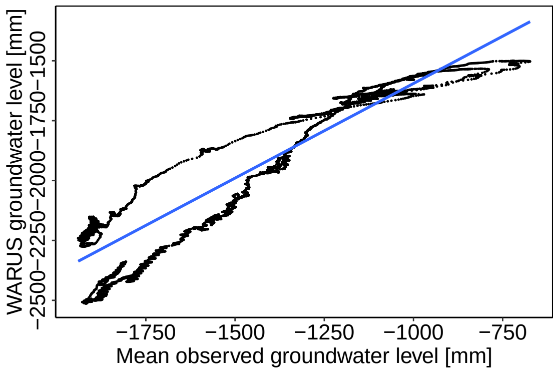
All 5 groundwater locations are used for the observation model. The correlation for the fitted model for 2018-2019 was $R^2 = 0.93$. For the validation the predicted storage deficit and groundwater both had a correlation of $R^2 = 0.90$ with the simulated WALRUS states. Due to the effect of hysteresis two clear differences in groundwater level are visible. Therefore, it can be preferable to separate the summer and winter discharge and use two separate observation models to get a better approximation of the catchment average groundwater level and storage deficit. The calibrated and validated observation models for the Groenlose Slinge are shown on the next page.



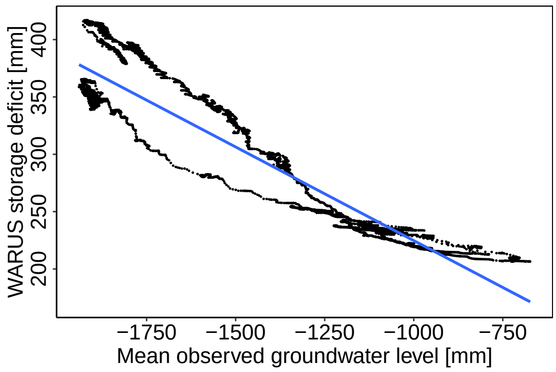
Calibrated groundwater level observation model for the year (2018-2019) calibrated on the simulated groundwater level (d_G) of WALRUS.



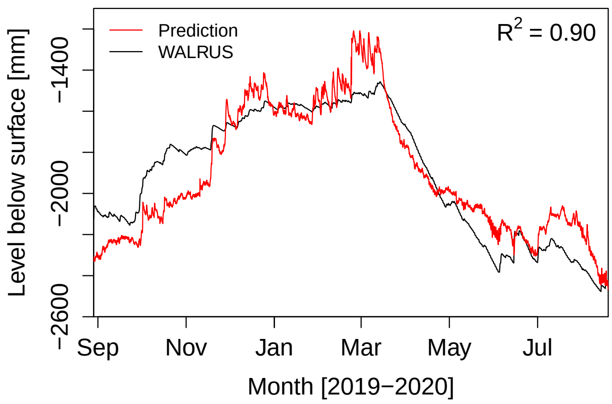
Calibrated storage deficit observation model for the year (2018-2019) calibrated on the simulated storage deficit (d_V) of WALRUS.



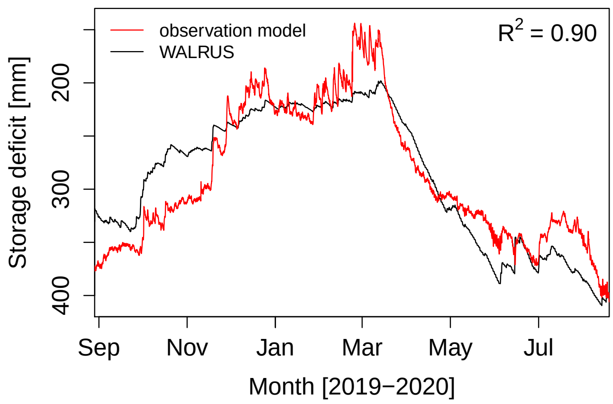
Fitted regression model as observation model used to estimate the d_G state from observed groundwater level. The model is calibrated with WALRUS groundwater level output (d_G) for 2018-2019.



Fitted regression model as observation model used to estimate the d_V state from observed groundwater levels. The model is calibrated with WALRUS storage deficit output (d_V) for 2018-2019.



Validation of observation model for updating the groundwater state (d_G) of WALRUS for 2019-2020.



Validation of observation model for updating the storage deficit state (d_V) of WALRUS for 2019-2020.