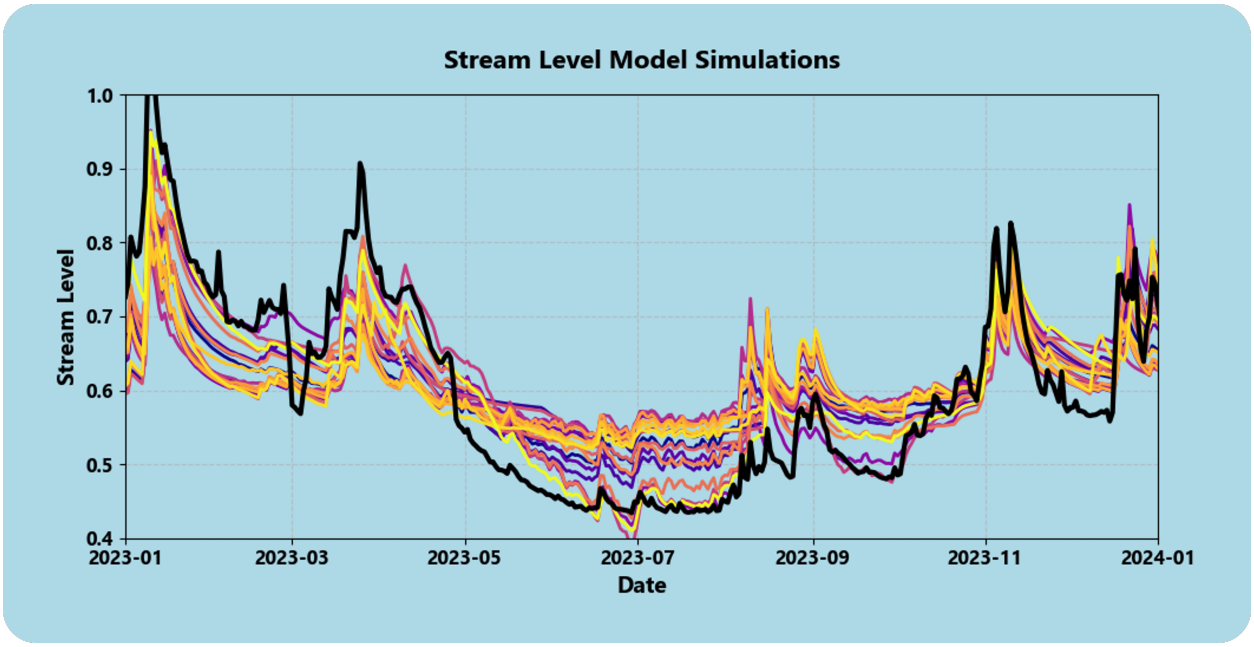


Quantifying Groundwater-Surface Water Interactions with Transfer Function Models

Assessment of a data-driven method applied to a case study of a tunnelling project in Kolmården, Sweden

Master's thesis in Infrastructure and Environmental Engineering

JOAQUIM ALTIMIRAS GRANEL

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Master's thesis 2024

Quantifying Groundwater-Surface Water Interactions with Transfer Function Models

Assessment of a data-driven method applied to a case study of a tunnelling project in Kolmården, Sweden

JOAQUIM ALTIMIRAS GRANEL

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Engineering Geology
Chalmers University of Technology
Gothenburg, Sweden 2024

Quantifying groundwater-surface water interactions with transfer function models
Assessment of a data-driven method applied to a case study of a tunnelling project in Kolmården, Sweden
JOAQUIM ALTIMIRAS GRANEL

Supervisors: Ezra Haaf, Department of Geology and Geotechnics
Johanna Merisalu, Department of Geology and Geotechnics
Andreas Berg, COWI AB
Examiner: Lars Rosén, Department of Geology and Geotechnics

Master's Thesis 2024
Department of Geology and Geotechnics
Division of Architecture and Civil Engineering
Engineering Geology
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Example of modelling results from one of the stream level models, with simulations using different methods for potential evapotranspiration.

Typeset in LaTeX
Gothenburg, Sweden 2024

Quantifying groundwater-surface water interactions with transfer function models
Assessment of a data-driven method applied to a case study of a tunnelling project in Kolmården, Sweden

JOAQUIM ALTIMIRAS GRANEL
Department of Geology and Geotechnics
Chalmers University of Technology

Abstract

This project aims to develop and evaluate a data driven method for analysing groundwater-surface water interactions. These interactions are analysed by using time-series analysis and transfer function modelling to create models for groundwater head, stream flow and stream level. The method is applied to a case study: a tunnelling project in Kolmården, Sweden, where a recipient stream in a nature reserve risks being affected by tunnel-induced groundwater drawdown due to leakage. The method consists of two modelling stages. In the first stage groundwater head models (GWMs) are developed. In the second stage, the GWMs are used to extend existing groundwater head time series and develop three types of models: models that estimate baseflow (BFM), stream total flow (TFM), and stream level (SLM). Additionally, the results are simulated using 18 different methods for estimation of potential evapotranspiration, to see how different methods impact results.

Results from the first modelling stage indicate higher model performance for models representing groundwater level in soil, compared to rock, and for models using longer calibration periods across all four seasons. Results from the second modelling stage present very poor results for the stream baseflow models (BFMs), which is probably related to the baseflow separation procedure's inability to account for specific local conditions. Total stream flow models (TFMs) and stream level models (SLMs) perform much better, with R^2-values around 0.6 in validation. However, all models show signs of overfitting, with R^2-values above 0.8 in calibration. The contribution from groundwater head to the variations seems to be higher for the SLMs.

The developed method constitutes a simplified and data driven approach for analysing groundwater-surface water interactions, making it possible to study, e.g., how tunnel-induced groundwater drawdown might affect a watercourse. Further development is recommended, e.g., using additional model evaluation metrics and simulating different leakage rate scenarios with pump test data. The method shows promise for allowing effective modelling of the dynamic response of groundwater-fed surface water to hydroclimatic variables, and can be seen as an additional tool for representing hydrogeological systems.

Keywords: hydrogeological modelling, groundwater-surface water interactions, time series analysis, transfer-function models, impulse response models, Pastas, Python

Acknowledgements

The final results of this project are a product of the continuous expertise, discussions and support of the supervisors of this project. Therefore, I would like to express my gratitude, especially to Ezra and Johanna at the Division of Geology and geotechnics at Chalmers, not only for the guidance, but also for introducing me to the field of hydrogeology and interesting discussions. I had the privilege of working in a great and supportive work environment thanks to the colleagues at COWI, who allowed me to join them at the office. For that I am very grateful, especially towards Andreas for enabling this collaboration, and always answering my questions. Finally, a special thanks to the people close to me for all the support during this spring.

Joaquim Altimiras Granel, Gothenburg, June 2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

BFM	Stream Baseflow Model
GWM	Groundwater Model
masl.	Meters above sea level
PET	Potential Evapotranspiration
SLM	Stream Level Model
SMHI	Swedish Meteorological and Hydrological Institute
TFM	Stream Total Flow Model
TFN	Transfer Function Noise (Modelling)

Contents

List of Figures

List of Tables

1
Introduction

While hydrology is the broader study of all the water on earth, the field of hydrogeology relates to all the water under the surface of the Earth and relationship of its processes with geologic materials (Fetter & Kraemer, 2022). Understanding the complex three-dimensional hydrogeological system at a site can be challenging, as the available information is limited to measurements at a few points (Bakker & Post, 2022). Creating models to describe these systems and understand the behaviour of the groundwater is therefore a common approach in the field of hydrogeology.

In projects that require construction underground, it is important to understand how the groundwater will be affected. Constructions under the groundwater level are exposed to the risk of groundwater inflow, which can cause groundwater drawdown (Sundell et al., 2019). Tunnelling-induced drawdown is a well-known issue and can have significant impacts on the environment and activities at the surface. The properties of the hydrogeological system including the distribution and connectivity of aquifers, as well as groundwater dependent ecosystems, will determine the magnitude of negative consequences induced by leakage into an underground construction, e.g. a tunnel (Yoo, 2016). For tunnels excavated in rock, it is difficult to predict how the water moves due to natural rock formations like dykes and fractures (Zarei et al., 2011). The extent of the groundwater inflow can be mitigated by using grouting techniques, which consists of injecting cement and chemical compounds in cavities along the tunnel walls to prevent groundwater from flowing in (Swanson & Babcock, 2023). Groundwater head can be measured with pressure transducers, commonly called divers, which are placed inside groundwater wells under the water table and measure absolute pressure exerted by the water column and the atmosphere.

This type of groundwater drawdown is interesting when studying interactions between groundwater and surface water bodies. The areas in the subsurface where the groundwater and the surface water meet are called hyporheic zones and is where there is an exchange of water (Fetter & Kraemer, 2022). Water can either move from the groundwater into the stream, creating a gaining stream, or from the stream into the groundwater, creating a losing stream (Kalbus et al., 2006). The characteristics of this water exchange are related to hydraulic head, and can vary over time due to, e.g., seasonal precipitation patterns. The interactions are complex as other factors, like geological conditions, have a significant impact on the fluxes (Sophocleous, 2002). By understanding this interaction at a site, it is easier to predict how changes in the groundwater head might impact the surface water, which can be used to protect the ecological values of surface water courses.

Another challenge in these types of projects is to accurately estimate the natural hydrological processes, e.g. recharge, and how the hydrogeological conditions will be affected by changes imposed by construction (Collenteur, Bakker, Klammler, & Birk, 2021). A common tool for handling these challenges is to develop models that are a simplified representation of the complex real system (Bakker & Post, 2022). Groundwater models are often used to quantify groundwater head at different points at a site (Anderson et al., 2015). Two categories of such mathematical models can be identified, process-based models and data-driven models. Process-based models are based on physical principles and processes to estimate the flow of groundwater. They are solved numerically, meaning that an unknown variable is approximated using methods that solve partial differential equations like finite-element (FE) or finite-difference (FD) (Anderson et al., 2015). In hydrogeology, these models can be used to represent the whole hydrogeological system in an area, but they usually require a significant amount of data and conceptual understanding of the site, and are therefore time-consuming to develop (Haaf et al., 2023). Data-driven models use empirical or statistical equations and existing data-series of known variables to calculate values of an unknown variable. They can be considered simpler since they can be solved analytically, i.e., an exact value for the unknown dependent variable is obtained. The choice of which model type to use depends on the type of problem at hand. This project aims to apply a data-driven approach using transfer function (TF) modelling and time series analysis, i.e., variations over time, to describe the interactions between groundwater and surface water.

1.1 Aim

The aim of this project is to develop a method constituting a data-driven approach for describing the relationship between groundwater levels and discharge in surface water recipients. The approach will be based on time series analysis and, specifically, transfer function models. It will be developed and demonstrated using data obtained from an existing investigation program for a tunnelling project in the Getå nature reserve in Kolmården, Sweden.

1.2 Specific Objectives

• Gather, process, and interpret field data to ensure adequate data quality for modelling.

• Develop local models for simulation of groundwater head at the points where pressure transducers measuring groundwater head are installed at the site, by using climate stresses such as precipitation and potential evapotranspiration.

• Analyse how different methods for estimating potential evapotranspiration affect the results.

• Develop models describing the relationship between groundwater head and flow in the stream, using the developed groundwater head models as input stress in addition to climate stresses.

• Evaluate model performance across calibration and validation periods by comparing simulation results and observed values, and optimise the models by adjusting parameters iteratively.

• Identify sources of uncertainties to understand limitations of the method and draw conclusions regarding significant model parameters based on the model performance.

• Ensure that the developed method constitutes a streamlined workflow that is comprehensible, applicable and that improves conceptualisation of the hydrological system.

1.3 Project Limitations

The project will focus on developing and optimising models to understand the relationship between surface water and groundwater and does not aim to specifically quantify the inflow rate to the planned tunnel. Additionally, it does not aim to evaluate or propose measures to mitigate the inflow of groundwater, or treatment or use of the water that enters the tunnel. These issues might be relevant to study in future research and are briefly discussed in Chapter 5.

2
Case Study Description

The case study used to evaluate this data-driven modelling approach for groundwater-surface water interactions is a tunnelling project planned in Kolmården, outside the city of Norrköping in Sweden. The tunnel will be approximately 8 km long, and is part of a larger infrastructure project, the East Link, which consists of a new 160 km long double-track railway between Stockholm and Linköping (Trafikverket, 2024).

2.1 Project Background

In 2022, the Swedish government decided that new main railway lines were to be built to connect the three largest cities, Stockholm, Gothenburg and Malmö (Trafikverket, 2023a). One of the main parts of this development is the 160 km railway in the project called Ostlänken, which connects the cities of Stockholm and Linköping (Trafikverket, 2023b). In the area of Kolmården, near the city Norrköping, an 8 km long tunnel will be excavated approximately 100 m underground (Trafikverket, 2022). The tunnel passes through a ravine of a few hundred meters at Getå, which is in a nature reserve and has a stream, Getåbäcken, that leads to the recipient, the bay Bråviken. The Swedish National Transportation Agency, Trafikverket, is the project owner and due to the specific conditions of the site, there is a risk that groundwater could flow into the tunnel and cause groundwater drawdown, which could affect the recipient stream at the Getå nature reserve (Trafikverket, 2023c). The stream is considered to be groundwater-fed and has a high ecological value for the nature reserve and should therefore be protected.

2.2 Case Study Area

The area is mainly covered by nature with a main highway, the E4, passing through south of the planned tunnel route. There are also several minor roads with houses alongside. The terrain in the area is quite hilly and has pronounced topographical gradients towards Bråviken in the south. The Getå stream comes down from the lakes in the north and passes the area in a pronounced ravine, meaning that higher elevations can be found on both sides of the ravine. The stream then continues down in a southeastern direction until it reaches the recipient. In the middle of the area, north of Algutsbo, another stream called Rödmossebäcken merges with Getåbäcken and continues downstream. The stream is mostly 2-3 meters wide and has a mean flow of 90 l/s at the point leading out to Bråviken (SMHI, 2020).

The main hydrogeological concern in this area is that a fracture zone in the bedrock passes through in the same direction as the Getå stream, which increases the risk of negatively impacting the stream because of tunnel-induced groundwater drawdown (Trafikverket, 2023c). An overview of the area is presented in Figure 2.1.

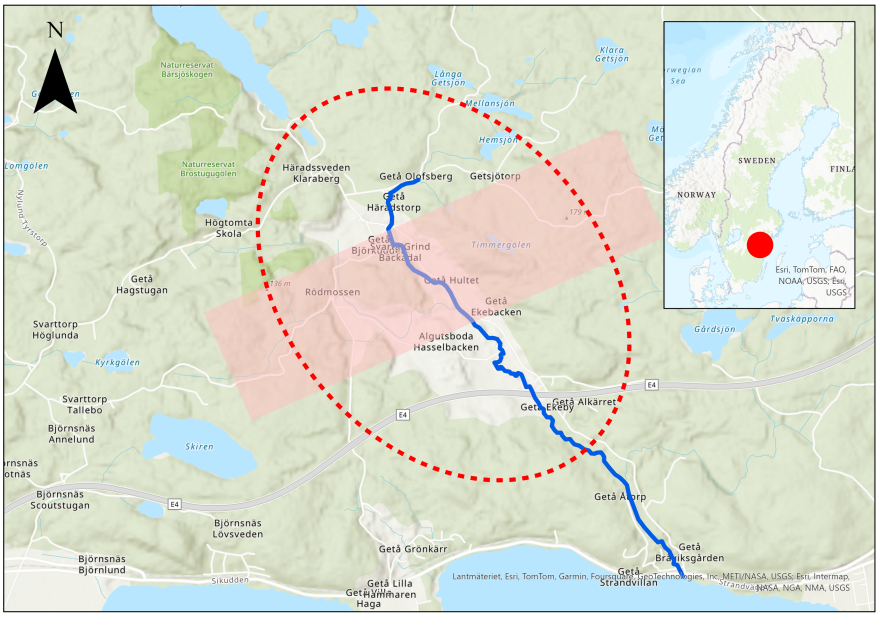


Figure 2.1: Overview of the case study area. The red dotted figure highlights the main area of interest with the Getå stream, the blue line, passing through the middle. The pink highlighted rectangle shows approximately where the tunnel passes the ravine.

2.3 Geological Setting

The bedrock in the area is between 2000 to 1650 million years old and was formed during the Svecokarelian orogeny (SGU, 2024). The Getå ravine was developed over time following the underlying fault zone in the bedrock. Soil type maps from the Geological Survey of Sweden (SGU), show that the area consists of mostly bedrock, postglacial sand and gravel at the surface (SGU, 2023a). An overview of the geology at the site is presented in Figure 2.2. In the soil type map the ravine, which is where the stream passes, is clearly distinguished in the yellow area representing postglacial silt. This geological setting is expected since the retreating ice sheet left fine-grained sediments in the ravine. Underneath the silt and clay, there is glaciofluvial sediment with higher hydraulic conductivity that can be considered an aquifer. The modelled soil depth is up to 10-20 m along the ravine, but the topmost impermeable layer might be very thin at some parts (SGU, 2023b). Previous investigations in the area suggest that at some parts the stream has eroded the top layer of low-conductivity soils, down to the glaciofluvial sediment (Trafikverket, 2023c). This enables a connection between the stream and the groundwater.

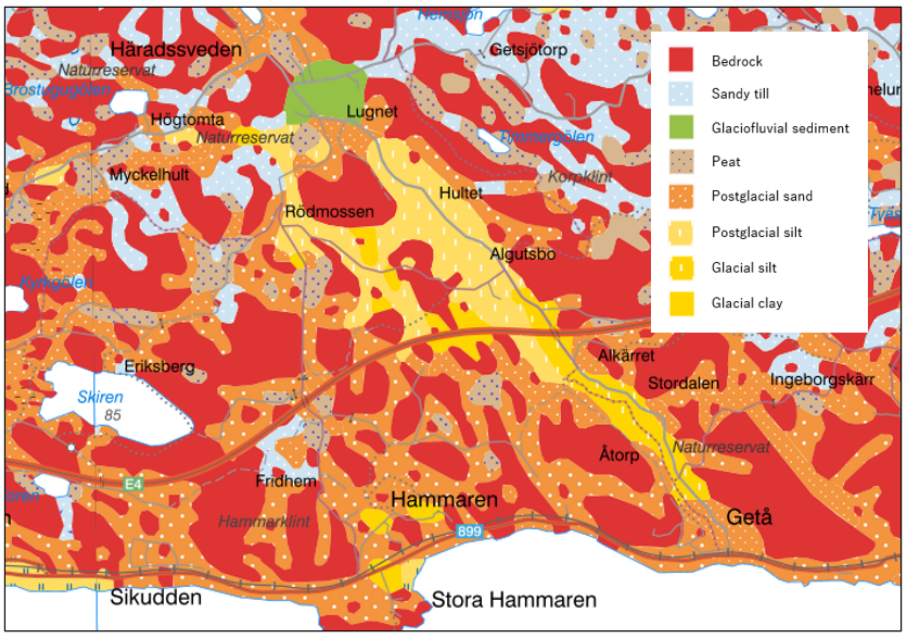


Figure 2.2: Map of soil types in the case study area from SGU (SGU, 2023a).

An additional hydrogeological concern within the project is the fact that the ravine in Getå has developed in a fracture zone, meaning that the bedrock is also permeable and can be considered an aquifer along the ravine (Länsstyrelsen, 2014). To visualise the geological setting and the challenges in the project, a conceptual cross-section of the stratigraphy in the ravine is presented in Figure 2.3. With this geological setting of several permeable layers, there is a potential risk that leakage to the tunnel will cause significant groundwater drawdown, which would impact the stream.

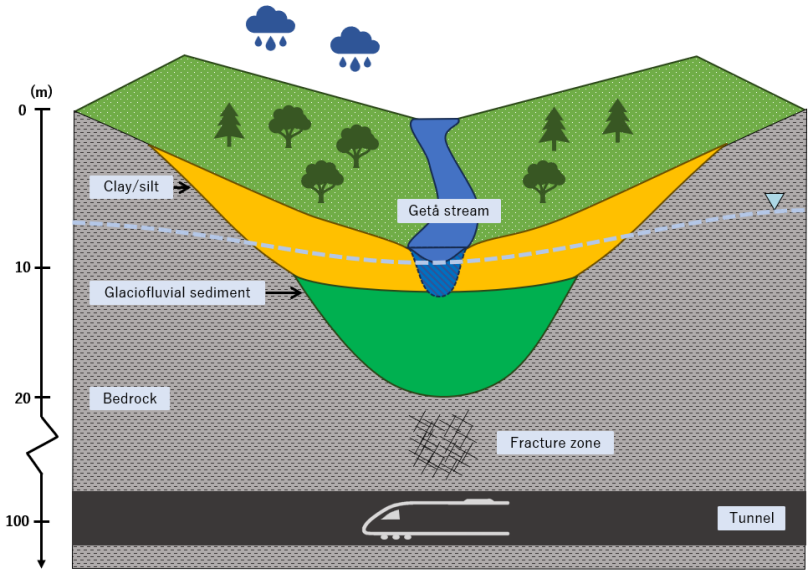


Figure 2.3: Conceptualisation of the stratigraphy in the ravine as a cross-section. The stream seems to have eroded down to the glaciofluvial sediment at some places along the stream.

2.4 Meteorological Data

The Swedish Meteorological and Hydrological Institute (SMHI) has one weather station that is very close to the area, approximately at 2 km from Getå, called Kolmården-Strömsfors A. From this station, precipitation, wind, humidity, and temperature data could be collected. In the city of Norrköping, at approximately 13 km, there is another weather station called Norrköping-SMHI collecting more types of data, which was used for data collection of sun hours and shortwave radiation. These data sets from SMHI are open to the public and were used in the models developed in this study.

2.5 Available Data

Since the project is in an investigation phase and has been for several years at the time of writing, more than 200 groundwater wells have been installed along the planned railway. In the catchment area in Getå, 26 groundwater wells have been installed. Since December 2022, researchers at Chalmers University of Technology have been collecting high-resolution (every 3 h) groundwater head data with pressure transducers at nine of the groundwater wells, see detailed information in Table 2.1. The barometric pressure is measured in well 17S1KB15 along diver EB938, but above the water table. An additional transducer, labelled DF155, measures stream level at one point, and Trafikverket measures stream flow at another point called MP4 Algutsbo. The locations for measurements of groundwater head, stream level and stream flow are presented in Figure 2.4.

Table 2.1: Details about groundwater head measurements. Well labels with corresponding diver labels and well types.

Well label	**Diver label**	**Well type**
22S108GU	DF150	Soil
17HB103	DF164	Rock
17HB102	DF165	Rock
22S107GU	EB829	Soil
22S119GU	EC402	Soil
17HB104	EG534	Rock
16S113GU	X3469	Soil
17S1KB15	EB938 + DE842 (barometric)	Rock
15S113GU	EG551	Soil

Figure 2.4: Map with locations of points for measurements of groundwater head, stream level and stream flow. The blue line represents the Getå stream.

3
Methods

3.1 Overview

The project is planned to consist of two modelling stages. The first stage consists of developing several local groundwater head models (GWMs), capable of simulating groundwater head changes. The second modelling stage consists of using the models from the first stage to extend the time series for groundwater head. They are then used as inputs to the three model types in the second stage: stream total flow models (TFMs), stream baseflow models (BFMs), and stream level models (SLMs). All models include measured precipitation data and modelled potential evapotranspiration data to represent recharge. The modelling approach is to use transfer function modelling, time series analysis, and the Python package Pastas as the main tool (Collenteur et al., 2019). A simplified version of the modelling workflow is presented in Figure 3.1.

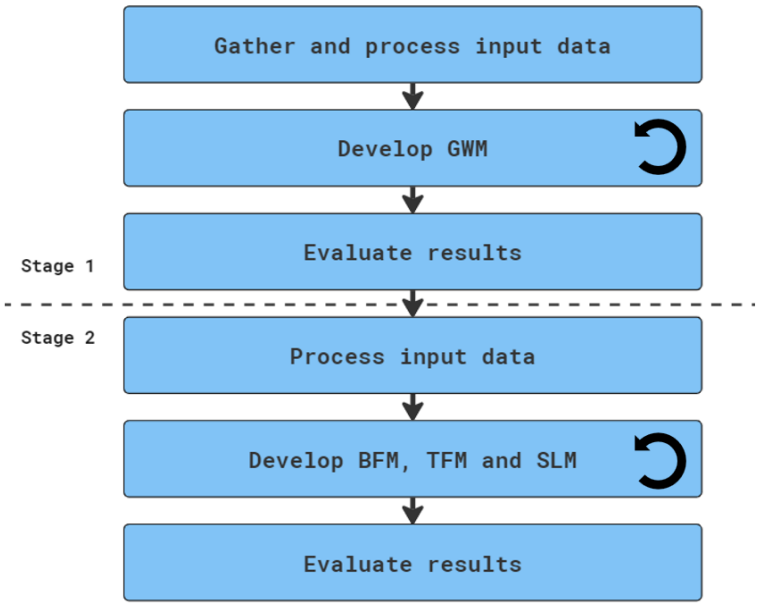


Figure 3.1: Presentation of the modelling workflow. The arrows inside the model development highlight that the process is iterative.

Recharge is the water added to the zone of saturation of the hydrogeologic system and can come from different processes (European Environment Agency, 2001). In the models of this project, recharge is represented through measured precipitation

and potential evapotranspiration (PET), which is modelled. Temperature data was also added separately to be able to account for snow. Measured data for groundwater head, stream total flow, stream baseflow and stream level is also needed for calibrating the models. Figure 3.2 presents a flowchart of the modelling approach and data sets needed for each stage and model.

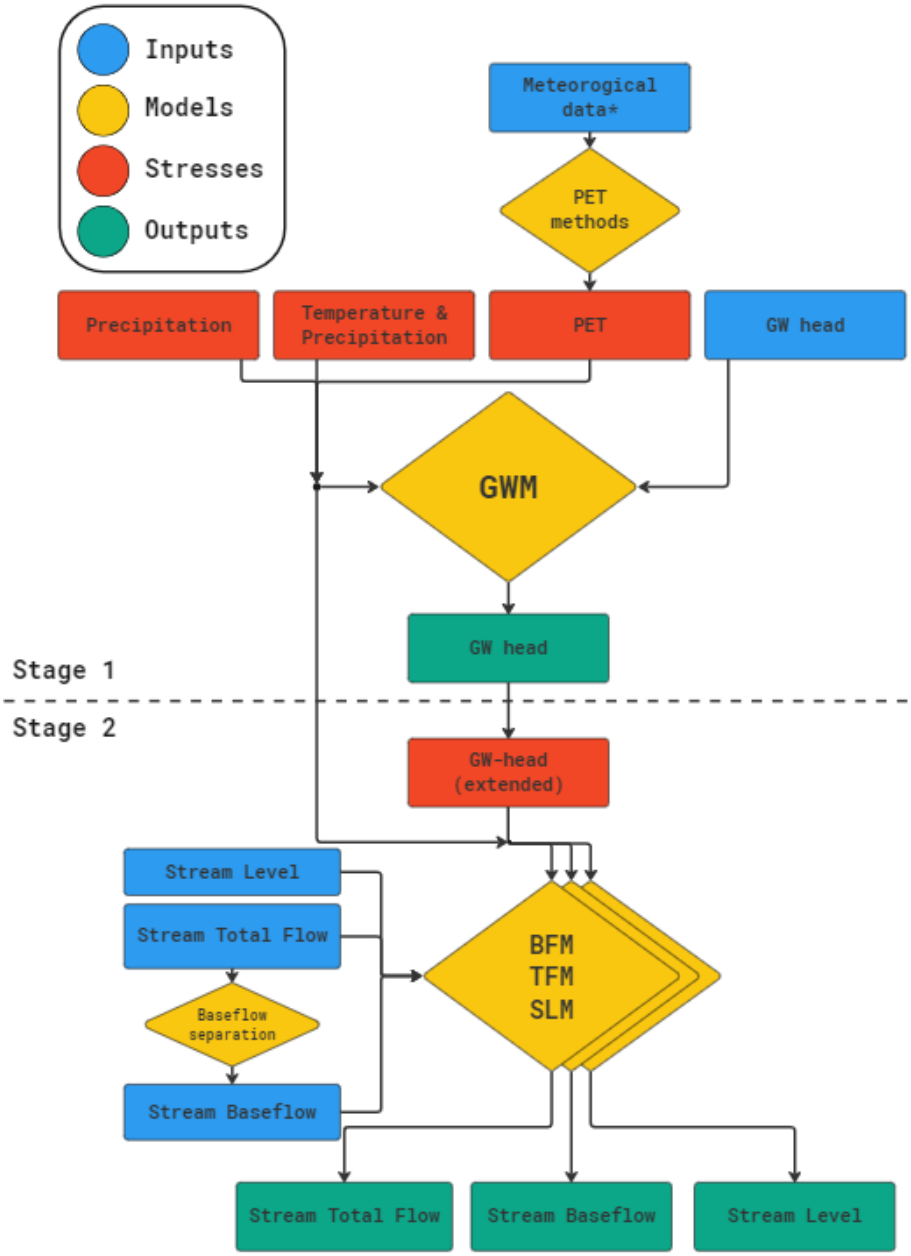


Figure 3.2: Flowchart of the modelling approach and data sets for each stage. The meteorological data consists of the parameters needed for estimation of potential evapotranspiration, see Section 3.3.2.

3.2 Transfer Function Noise Modelling and Python

The method developed in this study uses time series and the concept of transfer function noise models for analysing groundwater-surface water interactions. In practice, this was done by coding with Python across several steps, from processing data to creating models and visualising results. Some minor data processing was done in Microsoft Excel. Several specialised Python packages were used in different stages of the project: Pastas, PyEt and Baseflow. All scripts and coding details can be found on GitHub, see information in Appendix A.

3.2.1 Transfer Function Noise Modelling

Transfer function noise models combine one or more input series and translate these into a single output series using a transfer function that describes the relationship between the inputs and the outputs (Hipel & McLeod, 1994). When trying to model a real complex system, the transfer function will not represent the system perfectly, which leads to disturbances in the results, commonly called noise. The use of TFN-models is a common approach in hydrology where single outputs are caused by several inputs. In groundwater modelling, Von Asmuth et al. (2002), introduced a new type of TFN-model based on predefined impulse-response (IR) functions. Since the introduction of this type of model, it has become increasingly used in the field of hydrogeology (Collenteur et al., 2019).

As mentioned, TFN-models model a single unknown variable using several known variables as input. This is done by using data sets for the inputs and creating a model that can produce values that correspond to the observed target variable. In hydrogeology that could mean using measured precipitation as an input and try to model groundwater head as a target output. By having measured data for groundwater head, the model can be calibrated. The model structure of a TFN-model can be expressed as in Equation 3.1 (Collenteur et al., 2019).

$$h(t) = \sum_{m=1}^{M} h_m(t) + d + r(t) \tag{3.1}$$

- $h(t)$ is the observed groundwater head.
- $h_m(t)$ is the head contribution from the stress m.
- M is the total number of stresses.
- d is the base elevation.
- $r(t)$ is the residual term.

Each stress, m, contributes to the head and can be computed through convolution of a stress time series and an impulse response function, see Equation 3.2 (Collenteur et al., 2019).

$$h_m(t) = \int_{-\infty}^{t} S_m(\tau)\theta_m(t-\tau)d\tau \tag{3.2}$$

- S_m is the time series of stress m.
- θ_m is the impulse response function for stress m.

There are different impulse response functions that can be used for different purposes, e.g. Gamma, Exponential, and Hantush (Collenteur et al., 2019). The response functions describe how the dependent variable will respond to the applied stress and can have different levels of complexity depending on the amount of parameters it uses. The residuals of the model can be modelled with a noise model to better understand how the developed model performs.

3.2.2 Python Package Pastas

Pastas is an open-source Python package used specifically for analysing groundwater time series with TFN-modelling, and has two purposes according to the developers: "*provide a scientific framework to develop and test new methods, and to provide a reliable ready-to-use software tool for groundwater practitioners*" (Collenteur, Bakker, Calje, & Schaars, 2021). Collenteur et al. (2019) explain the inner workings and functionality of the Pastas package in detail. Pastas uses TFN-modelling and IR-functions to simulate output series.

Pastas uses two types of time series to create models, dependent and independent time series (Collenteur et al., 2019). The dependent time series are the targeted series that will be modelled, e.g., groundwater head, and are added as an *oseries*. Independent time series are used as inputs to model the dependent time series and are often referred to as stresses, e.g., precipitation and potential evapotranspiration. Pastas has an object-oriented design and three main classes, *Model*, *StressModel* and *RechargeModel*. Once the time series have been collected and read, a *Model* can be created that stores the dependent time series as an *oseries*. The stresses are then applied to the model through stress models, of which there are different types.

In this study three types of stress models were used: *StressModel*, *RechargeModel*, and *TarsoModel* (Collenteur, Bakker, Calje, & Schaars, 2021). The *StressModel* is the simplest one and uses a single stress time series and a response function to compute the contribution of that stress to head fluctuations. The *RechargeModel* and *TarsoModel* combine specifically precipitation and potential evapotranspiration data to represent recharge as a stress. Both stress model types allow the opportunity to choose a specific model for recharge (Collenteur, Bakker, Calje, & Schaars, 2021). For all models developed in this study, the *FlexModel* was used as recharge model since it offers the possibility to consider presence of snow by adding a temperature time series. The *TarsoModel* (Threshold Autoregressive Self-Exciting Open-loop) is a model that breaks down the data into different regimes based on threshold values (Knotters & De Gooijer, 1999). A piecewise linear function is created for each regime, describing the relationship and modelling groundwater head. This allows for a non-linear analysis where different conditions can be applied to different regimes.

If several stress models were added to a model, the contributions of each stress model added to a model was obtained using the Pastas function *get_contribution*.

3.3 Data Sets and Data Processing

As seen in Figure 3.2, several data sets were used to develop the models. All data is related to the case study and was collected continuously in time series. Before using the data sets in the models, they had to be processed which included several steps: unit adjustment, temporal resolution adjustment, quality control, filtering and data imputation. The approach for data processing was to keep the original raw data in Excel files (.xlsx), and then apply the processing steps in Python scripts. Using time series with daily values is recommended for this type of modelling (Collenteur, Bakker, Calje, & Schaars, 2021).

3.3.1 Meteorological Data

Data sets for precipitation, temperature and meteorological data needed for PET were collected from SMHI's weather stations. The meteorological data needed for estimating potential evapotranspiration are described further in Section 3.3.2. All data was resampled to daily values, precipitation in total [mm/day], and temperature in mean [°C/day].

3.3.2 Estimation of Potential Evapotranspiration

Potential evapotranspiration (PET) is a term used to describe the water loss in a natural system due to surface evaporation and transpiration from vegetation, which often do not need to be separated in hydrogeology (Fetter & Kraemer, 2022). It is one of the main parameters when studying the hydrological characteristics of a site and has historically been very important in agriculture. However, measuring PET directly is difficult and expensive, and therefore several methods for estimating PET have been developed over the years (Vremec et al., 2024).

To obtain PET time series, the Python package PyEt was used. PyEt is an open-source Python package that uses meteorological data to estimate daily PET (Vremec et al., 2023). At the time of writing the package supports PET estimations based on 18 different methods. The parameters needed for estimating PET with all methods are, mean, minimum and maximum day temperature [°C], mean day wind speed [m/s], incoming solar radiation [MJ m^{-2} d^{-1}], site elevation [masl.], site latitude, and mean, minimum and maximum relative humidity [%]. All the methods in PyEt are presented in Table 3.1. There are three types of PET methods, that use different parameters: methods based on radiation, methods based on temperature, and methods based on a combination of both (Vremec et al., 2023). Most methods have been developed focusing on agriculture.

Table 3.1: All methods for estimating PET in the PyEt Python package.

Method Name	Model Type	Reference
Penman	Combination	(Penman, 1947)
FAO-56	Combination	(Allen et al., 1998)
Priestley-Taylor	Combination	(Priestley & Taylor, 1972)
Kimberly-Penman	Combination	(Wright, 1982)
Thom-Oliver	Combination	(Thom & Oliver, 1977)
Blaney-Criddle	Temperature	(Blaney & Criddle, 1950)
Hamon	Temperature	(Hamon, 1963)
Romanenko	Temperature	(Romanenko, 1961)
Linacre	Temperature	(Linacre, 1977)
Haude	Temperature	(Haude, 1955)
Turc	Radiation	(Turc, 1961)
Jensen-Haise	Radiation	(Jensen & Haise, 1963)
McGuinness-Bordne	Radiation	(McGuinness & Bordne, 1972)
Hargreaves	Radiation	(Hargreaves & Samani, 1983)
FAO-24	Radiation	(Jensen et al., 1990)
Abtew	Radiation	(Abtew, 1996)
Makkink	Radiation	(Makkink, 1957)
Oudin	Radiation	(Oudin et al., 2005)

3.3.3 Groundwater Head Data

Groundwater head data was needed for the models in the first modelling stage for calibration. The groundwater head was collected at nine locations with pressure transducers that were placed in groundwater wells measuring absolute pressure, see Figure 2.4. An additional transducer was placed above the water column in another groundwater well to measure the barometric pressure. The transducers yield the pressure in cm H_2O. By subtracting the barometric pressure from the absolute pressure in each time step, the remaining value is the pressure exerted by the water column above the diver level. This procedure is done for barometric compensation and gives a comprehensible value since the transducers measure the pressure in cm H_2O. This could then be transformed to meters above sea level (masl.) by subtracting the diver depth in the well from the reference level. By then adding the water column height, the values have been transformed from absolute pressure in cm H_2O to masl. Figure 3.3 illustrates conceptually how this was done.

3.3.4 Stream Level Data

Stream level data was needed for the SLMs in the second modelling stage to use it as a time series for calibration. The data was obtained from an additional pressure transducer placed in a groundwater well located in the Getå stream. The stream level is calculated with the same principle as the groundwater head data, using the barometric data for compensation. The groundwater well with this data is located close to well 22S119GU, see Figure 2.4.

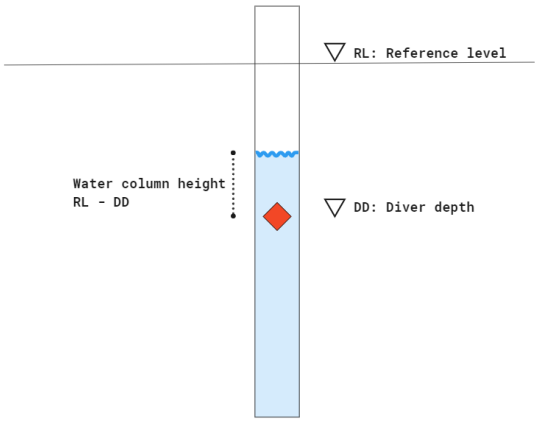


Figure 3.3: Visualisation of conversion from pressure to masl.

3.3.5 Stream Flow Data

Stream flow data was needed for the TFMs in the second modelling stage to use it as a time series for calibration. The measurements were collected continuously at a point called MP4 Algutsbo by Trafikverket, which is located downstream of the divers, see Figure 2.4.

3.3.6 Estimation of Baseflow

Baseflow data was needed in the calibration of the BFMs, in the second modelling stage. Baseflow is defined as the part of the total flow in a stream that originates from groundwater and depends on the specific conditions of a watercourse (Fetter & Kraemer, 2022). It has been shown that baseflow in Sweden, even during precipitation events, can be as high as 95%, meaning that almost all flow originates from groundwater in that case (Rodhe, 1987). Estimating how much of the total flow in a stream is baseflow is complicated, and there are several methods that can be used to estimate it. The Python package Baseflow uses 12 different methods for baseflow separation and chooses the best result based on the metric Kling-Gupta efficiency (KGE) (Xie et al., 2020). The best performing model is then exported and used for calibration in the BFMs.

3.3.7 All Processed Input Data

All these data sets were used either as independent stresses in the models or as dependent series for calibration. All the processed input data sets are presented in Table 3.2 as daily values.

Table 3.2: All data sets after processing, resampled to daily values. The asterisk indicates that the data is modelled instead of measured.

Data Set	Unit	Start of Series	End of Series
Temperature	°C (mean)	1995-12-17	2024-04-09
Precipitation	mm (total)	1995-12-14	2024-04-09
Potential Evapotranspiration*	mm (total)	1995-01-01	2024-04-09
Groundwater Head	masl. (mean)	2022-12-07	2024-03-19
Stream Level	masl. (mean)	2022-12-07	2024-03-19
Stream Flow	m^3/s (mean)	2022-12-01	2024-04-04
Stream Baseflow*	m^3/s (mean)	2022-12-01	2024-04-04

3.4 Modelling Stage 1: Groundwater Head Models

The first modelling stage consisted of creating models for groundwater head (GWM) for each diver location. This was done using the Python package *Pastas*. Four sets of input data were needed for each GWM: precipitation, temperature and potential evapotranspiration for stresses, and measured groundwater head for calibration. In total, nine GWMs could be created since there were nine divers measuring groundwater head, and one measuring barometric pressure. However, three of these were later discarded, see Section 4.1.

Each model was created through an iterative process of testing different stress models and changing model parameters. The optimisation aimed at finding a model variation that both performed well continuously over the calibration and validation periods, and that represented the real system as accurately as possible. When a final model variation was chosen, simulations were done for each PET method across the calibration and validation period. The four best PET methods were then highlighted, and the best performing model was used for the next modelling stage.

3.5 Modelling Stage 2: Stream Flow, Stream Baseflow, and Stream Level Models

The aim of the second modelling stage was to model stream total flow, stream baseflow and stream level, using extended groundwater head data from the first modelling stage as stresses. Three types of models were created: baseflow models (BFMs), total flow models (TFMs), and stream level models (SLMs). Each model type used a different dependent series for calibration, i.e., baseflow for the BFMs, total flow for the TFMs, and stream level for the SLMs, see Table 3.2.

Two types of stresses were added to all these models: one *RechargeModel* representing recharge with precipitation and PET data, and one *StressModel* with extended groundwater head data from one of the GWMs from the first stage. Since there

were six complete GWMs, six models of each type could be created, each using a different groundwater head series as a stress. An example is the model labelled TFM-17HB102, presented in Section 4.2.2, which modelled the total stream flow by using a total flow series for calibration, and a groundwater head series from well 17HB102 as a stress. The modelling development process was iterative and exploratory, to identify a general baseline model that was not too complex and that produced reasonable results consistently. In practice, this means that the models were run in several iterations trying out different combinations of stress models, response functions and parameters, like turning on and off presence of snow and groundwater uptake.

3.5.1 Extending Groundwater Series From GWM

When a stress is added to a Pastas model, a default warm-up period of 10 years is automatically implemented. This means that any stress time series that is added in a stress model must have data for the last 10 years. By using the GWMs from the previous modelling stage, the existing time series for groundwater head could be extended by simulations, since the GWMs only used precipitation and PET. This was done for each GWM. An extra quality control step was added at this point by controlling the simulated levels with manually measured values for nearby wells. Since there are many groundwater wells in the area that have been part of control programs over the years, there were time series with monthly values for some of them. By standardising the results, a visual quality control could be done to check model fits.

3.6 Model Evaluation

To evaluate the model results, the metric coefficient of determination, R^2, was used. R^2 is a metric that ranges from 0 to 1 and that describes how much of the variation in one series, e.g., measured groundwater head, is explained by another series, e.g., modelled groundwater head (Britannica, 2024). Using R^2 on its own has some limitations as it does not show if a model is biased, with values continuously too high or too low, or overfitted, i.e., too adjusted to the calibration data (Frost, 2024). It is good practice to also evaluate the residual plots when using R^2. In this study R^2 was used as a general metric to compare models, and its use is discussed further in Section 5.3. The choice of the best model variations was mainly based on the results in validation, since that describes the real performance of the model with data that it has not been calibrated against.

4
Results

This chapter presents the results for the first and second modelling stage. For each modelling stage, data, model development, and specifications of the final models are presented. By simulating the models iteratively, using different PET methods, a more in-depth analysis of the impact of PET data was performed.

4.1 Modelling Stage 1: GWM

Of the nine time series from divers, six were considered to be good enough, based on data quality and continuity, for creating a groundwater head model (GWM). The data sets for divers EB938 (17S1KB15), EG551 (15S113GU) and X3469 (16S113GU) were discarded during the model development process, mainly due to lack of consistent data and overall short measuring periods. The lack of data made it difficult to properly calibrate and validate the models.

For each model, measured groundwater head was used as a time series for calibration in Pastas. Time series for precipitation, PET and temperature were added as stresses to the models. Table 4.1 and Figures 4.1 and 4.2 present the time series of these inputs. Of the six measured groundwater head series, four have a length of 15 months, and two have a length of 8 months. To maintain consistency across the different models, the ratio 75-25% was chosen to divide the periods into calibration and validation. The main reason for this choice of ratio was that a calibration period of one year was desired, to capture the variations over one whole year. Detailed model results from the calibration period are presented in Appendix B.

Table 4.1: Precipitation, temperature and PET data used for the GWMs.

Time series	**Source**	**Start of series**	**End of series**
Precipitation	Measured (SMHI)	1995-12-14	2024-04-09
Temperature	Measured (SMHI)	1995-12-17	2024-04-09
PET	Estimated (PyEt)	1996-02-01	2024-04-09

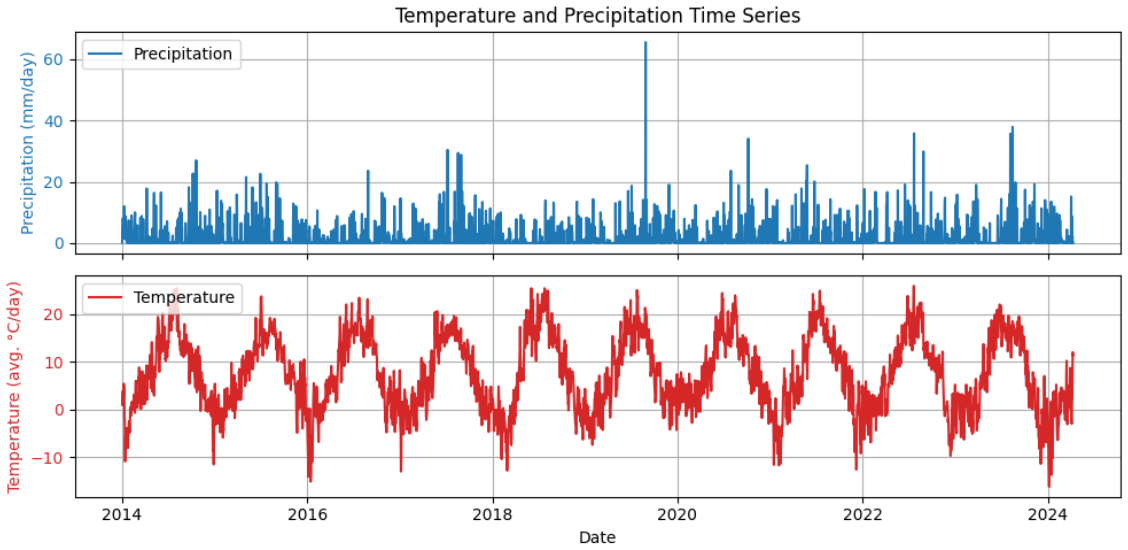


Figure 4.1: Time series for precipitation and temperature used in all GWMs. 2014-2024.

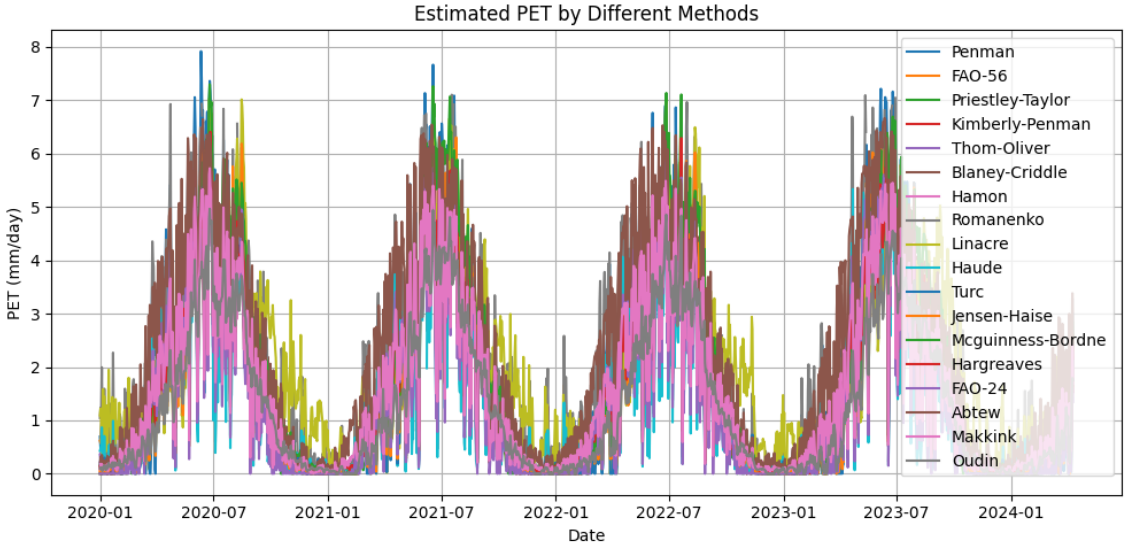


Figure 4.2: Time series for PET estimated with 18 different methods in PyEt, 2020-2024.

4.1.1 GWM-22S108GU

Well 22S108GU was driven in soil approximately 130 m from the Getå stream in the middle of the area, with an elevation of 76 masl. The model contains one stress model, a *RechargeModel* used to represent the recharge in the system with precipitation and PET data. The *RechargeModel* uses a *FlexModel* with a *Gamma* response function and accounts for groundwater uptake and presence of snow. The calibration period for the model is eleven months, from 2022-12-22 to 2023-11-22, and the validation period is four months, from 2023-11-22 to 2024-03-19. The results for the different PET methods are presented in Figure 4.3.

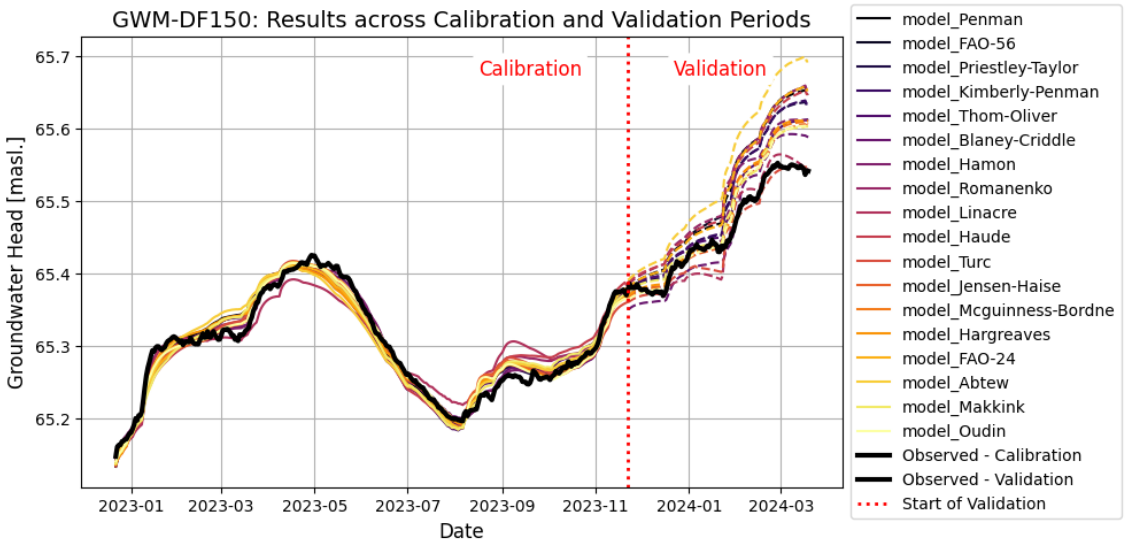


Figure 4.3: Results for GWM-22S108GU across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-22S108GU during calibration and validation are presented in Table 4.2. Turc's PET method was chosen as the best method for the next modelling stage.

Table 4.2: Top four results for GWM-22S108GU based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
Blaney-Criddle	0.99	0.81
Linacre	0.89	0.89
Turc	0.95	0.94
Jensen-Haise	0.97	0.79

4.1.2 GWM-17HB103

Well 17HB103 was driven in rock approximately 2 km east of the Getå stream, close to well 17HB104, with an elevation of 140 masl. The model contains one stress model, a *TarsoModel* used to represent the recharge in the system with precipitation and PET data. The *TarsoModel* uses a *FlexModel* with an *Exponential* response function and accounts for groundwater uptake and presence of snow. The calibration period for the model is eleven months, from 2022-12-22 to 2023-11-22, and the validation period is four months, from 2023-11-22 to 2024-03-19. The results for the different PET methods are presented in Figure 4.4.

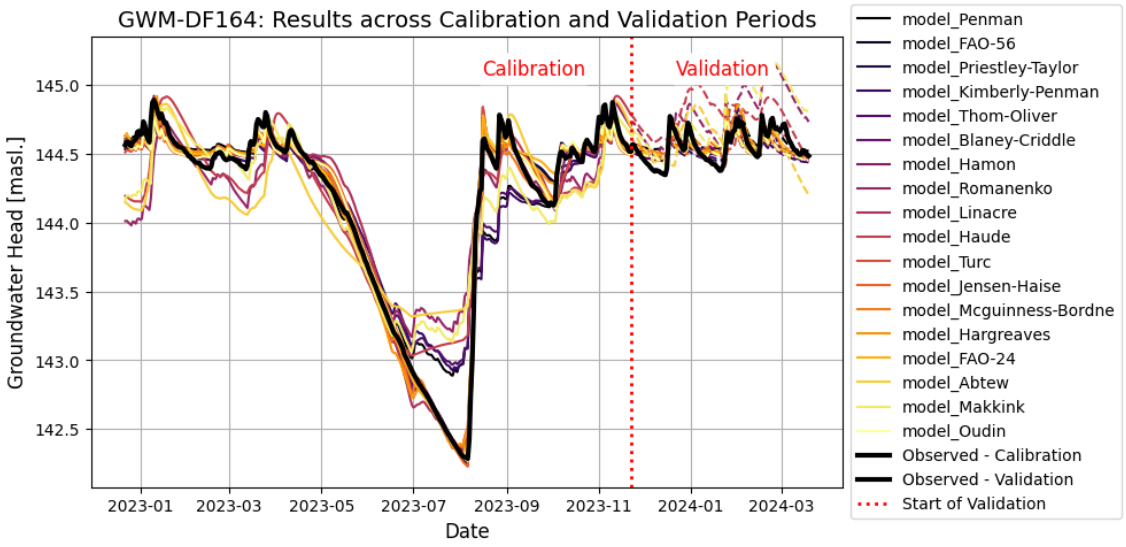


Figure 4.4: Results for GWM-17HB103 across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-17HB103 during calibration and validation are presented in Table 4.3. Turc's PET method was chosen as the best method for the next modelling stage.

Table 4.3: Top four results for GWM-17HB103 based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
Priestley-Taylor	0.99	0.52
Blaney-Criddle	0.99	0.54
Turc	0.99	0.59
Hargreaves	0.99	0.49

4.1.3 GWM-17HB102

Well 17HB102 was driven in rock approximately 2 km west of the Getå stream with an elevation of 103 masl. The model contains one stress model, a *TarsoModel* used to represent the recharge in the system with precipitation and PET data. The *TarsoModel* uses a *FlexModel* with an *Exponential* response function and accounts for groundwater uptake and presence of snow. The calibration period for the model is eleven months, from 2022-12-22 to 2023-11-22, and the validation period is four months, from 2023-11-22 to 2024-03-19. The results for the different PET methods are presented in Figure 4.5.

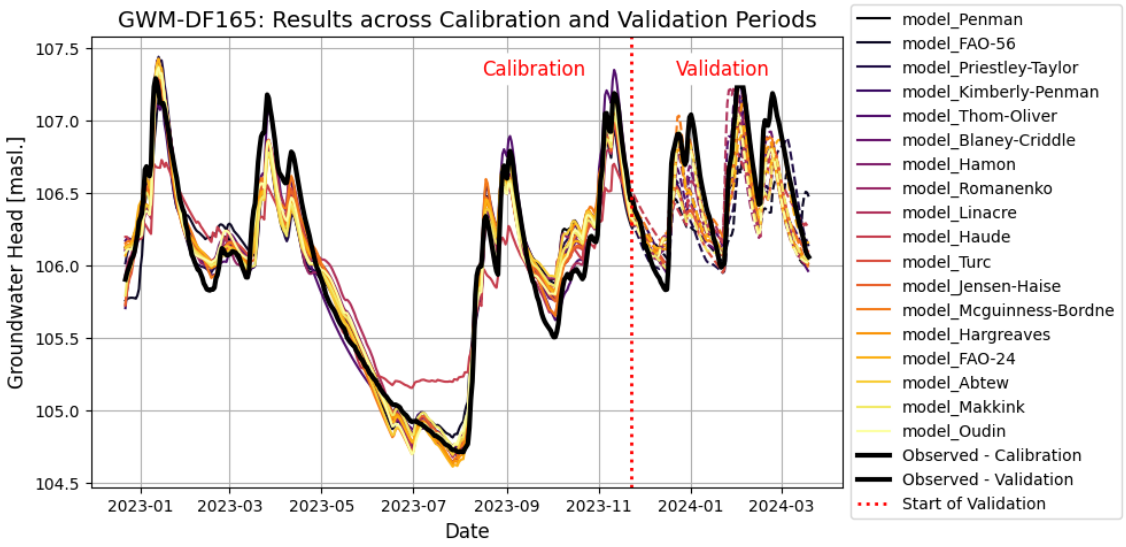


Figure 4.5: Results for GWM-17HB102 across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-17HB102 during calibration and validation are presented in Table 4.4. McGuinness-Bordne's PET method was chosen as the best method for the next modelling stage.

Table 4.4: Top four results for GWM-17HB102 based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
Thom-Oliver	0.97	0.67
Blaney-Criddle	0.95	0.68
McGuinness-Bordne	0.95	0.73
FAO-24	0.94	0.68

4.1.4 GWM-22S107GU

Well 22S107GU was driven in soil approximately 130 m from the Getå stream close to well 22S108GU, with an elevation of 76 masl. The model contains one stress model, a *RechargeModel* used to represent the recharge in the system with precipitation and PET data. The *RechargeModel* uses a *FlexModel* with a *Gamma* response function and accounts for groundwater uptake and presence of snow. The calibration period for the model is eleven months, from 2022-12-22 to 2023-11-22, and the validation period is four months, from 2023-11-22 to 2024-03-19. The results for the different PET methods are presented in Figure 4.6.

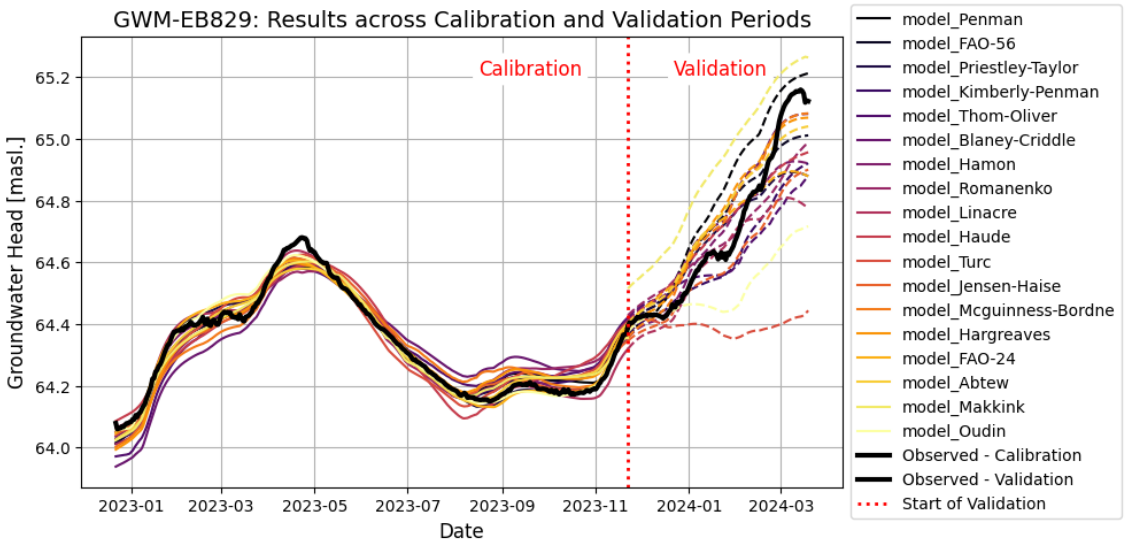


Figure 4.6: Results for GWM-22S107GU across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-22S107GU during calibration and validation are presented in Table 4.5. McGuinness-Bordne's PET method was chosen as the best method for the next modelling stage.

Table 4.5: Top four results for GWM-22S107GU based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
FAO-56	0.96	0.88
McGuinness-Bordne	0.92	0.92
Hargreaves	0.94	0.88
Abtew	0.96	0.87

4.1.5 GWM-22S119GU

Well 22S119GU was driven in soil approximately 10 meters from the Getå stream, close to the flow measurement point MP4 Algutsbo, with an elevation of 59 masl. The model contains one stress model, a *RechargeModel* used to represent the recharge in the system with precipitation and PET data. The *RechargeModel* uses a *FlexModel* with a *Gamma* response function and accounts for groundwater uptake and presence of snow. The calibration period is shorter than for previous models and is six months, from 2023-07-13 to 2024-01-15. The validation period is two months, from 2024-01-15 to 2024-03-19. The results for the different PET methods are presented in Figure 4.7.

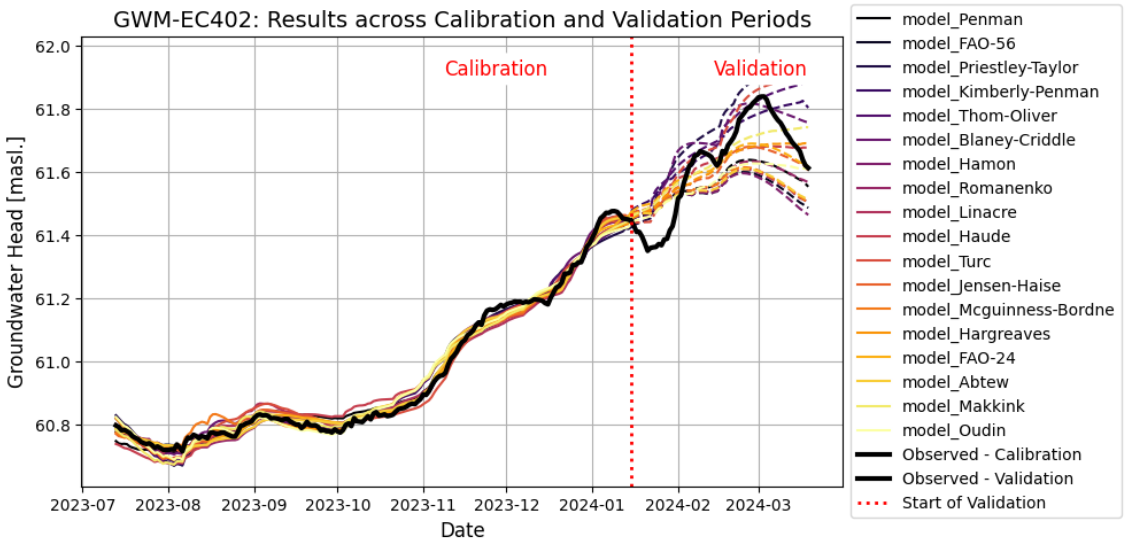


Figure 4.7: Results for GWM-22S119GU across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-22S119GU during calibration and validation are presented in Table 4.6. Makkink's PET method was chosen as the best method for the next modelling stage.

Table 4.6: Top four results for GWM-22S119GU based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
Blaney-Criddle	0.99	0.65
Jensen-Haise	0.99	0.64
Hargreaves	0.99	0.67
Makkink	0.99	0.71

4.1.6 GWM-17HB104

Well 17HB104 was driven in rock approximately 2 km east of the Getå stream with an elevation of 139 masl. The model contains one stress model, a *TarsoModel* used to represent the recharge in the system with precipitation and PET data. The *TarsoModel* uses a *FlexModel* with an *Exponential* response function and accounts for groundwater uptake and presence of snow. The calibration period is shorter also for this model and is six months, from 2023-07-13 to 2024-01-15. The validation period is two months, from 2024-01-15 to 2024-03-19. The results for the different PET methods are presented in Figure 4.8.

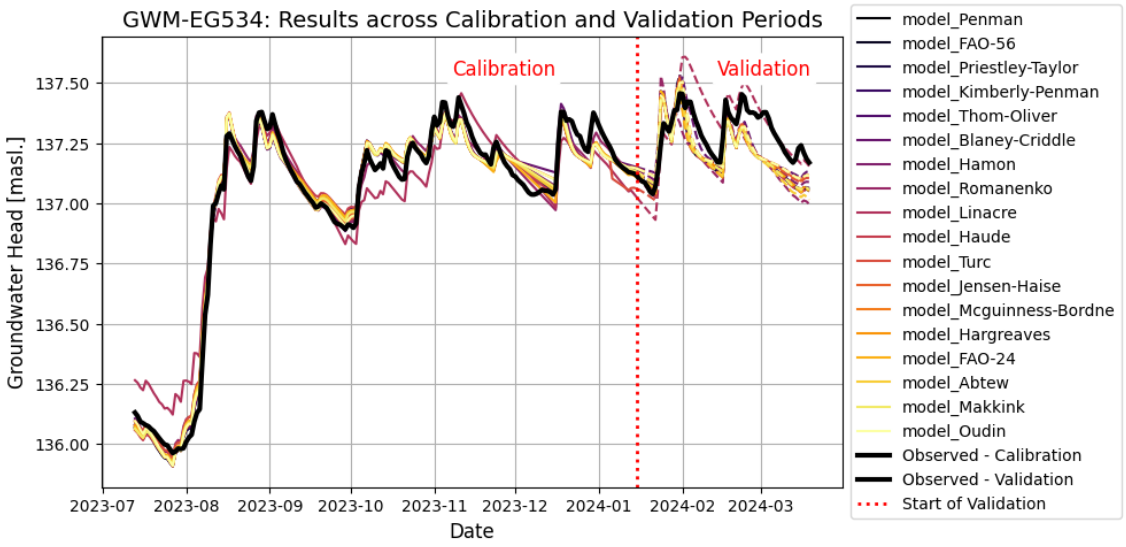


Figure 4.8: Results for GWM-17HB104 across the calibration and validation periods for all the PET methods.

The PET methods that produce the best results for GWM-17HB104 during calibration and validation are presented in Table 4.7. Blaney-Criddle's PET method was chosen as the best method for the next modelling stage.

Table 4.7: Top four results for GWM-17HB104 based on R^2 across the calibration and validation periods. The best model is highlighted in green.

PET method	R^2 Calibration	R^2 Validation
Blaney-Criddle	0.98	0.30
Linacre	0.94	0.23
McGuinness-Bordne	0.98	0.21
Oudin	0.98	0.21

4.1.7 GWM Summary

The best GWMs were used to extend the groundwater head time series in the next modelling stage. Two model results are significantly better than the others, GWM-22S108GU and GWM-22S107GU, which were based on data sets measured close to the Getå stream. Table 4.8 presents a summary of the best performing GWMs.

Table 4.8: Calibration and validation results of chosen models for each GWM that was used further in the second modelling stage. The asterisk highlights the models with shorter time series.

GWM	**Well type**	**PET method**	**R^2 Cal.**	**R^2 Val.**
GWM-22S108GU	Soil	Turc	0.95	0.94
GWM-17HB103	Rock	Turc	0.99	0.59
GWM-17HB102	Rock	McGuinness-Bordne	0.95	0.73
GWM-22S107GU	Soil	McGuinness-Bordne	0.92	0.92
GWM-22S119GU*	Soil	Makkink	0.99	0.71
GWM-17HB104*	Rock	Blaney-Criddle	0.98	0.30

4.2 Modelling Stage 2: TFM, BFM, and SLM

The six best performing GWMs were used to hindcast the existing groundwater head series back to 2012-01-01. Two stresses were then added to the models: recharge, with a *RechargeModel* with precipitation and PET data, and groundwater head with a *StressModel* and a *Gamma* response function. In this stage three types of Pastas-models were developed, Total Flow Models (TFM), Baseflow Models (BFM), and Stream Level Models (SLM). Six models of each type could be created, each using a different groundwater head series as a stress. All models represent the system at the flow measurement point called MP4 Algutsbo, see Figure 2.4. All BFMs and TFMs presented in this section had a calibration period of twelve months, from 2022-12-01 to 2023-12-01, and a validation period of four months, from 2023-12-01 to 2024-04-01. The SLMs were a few days shorter due to limited data, starting on 2022-12-07 and ending on 2024-03-19. The choice of the best model variations was based mainly on the performance in validation, given that the performance in calibration was good in relation to the other model variations. The specific contribution of each stress model is also presented for the best model variation of each model type.

4.2.1 Baseflow Models (BFM)

To calibrate a model that simulates the baseflow in the stream, a time series for the baseflow is needed. After applying baseflow separation with different methods, the EWMA-method proved to be the best performing one based on Kling-Gupta efficiency. The resulting baseflow series is presented in Figure 4.9.

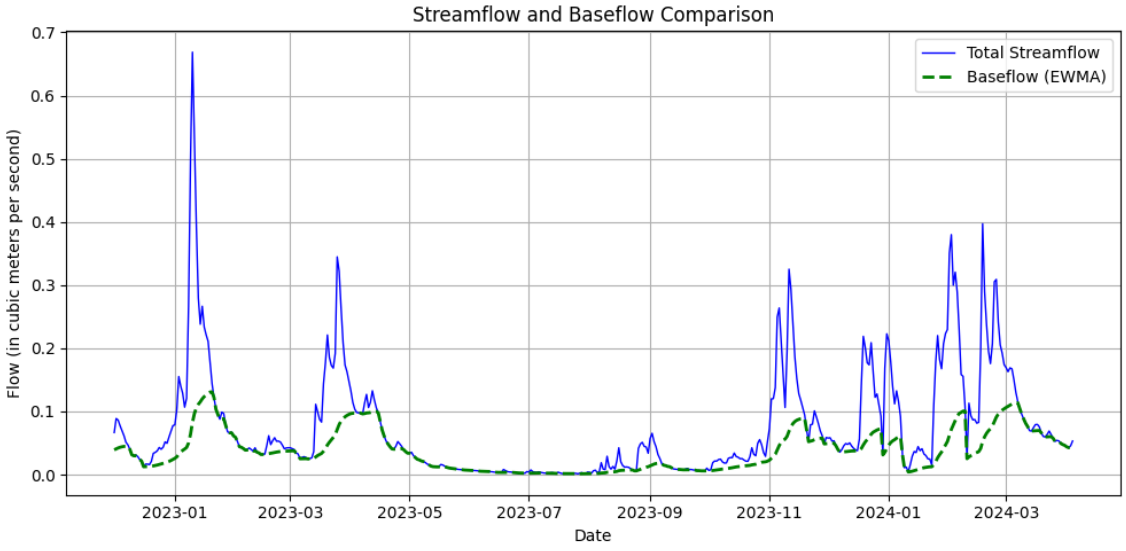


Figure 4.9: Total flow and baseflow series for measuring point MP4 Algutsbo after baseflow separation with the best performing method, EWMA.

With the obtained baseflow time series, Pastas models for simulating baseflow, BFMs, could be developed. The BFMs included two stress models. The first one is a *RechargeModel* with a *Gamma* response function, accounting for groundwater

uptake and presence of snow, used to represent recharge. The second one is a *Stress-Model* with a hindcasted groundwater head series and a *Gamma* response function. Each model was tested with the 18 different PET methods, and the three best model results were selected for evaluation. The results for the six baseflow models, with the three best PET methods for each, are presented in Table 4.9.

The best-performing BFM is BFM-22S119GU with Turc's PET method, and the contribution for each of the two stress models, in the calibration period, is presented in Figure 4.10. Running the model with only a *RechargeModel* results in R^2=0.93. With both stress models added there is a slight improvement, with R^2=0.96.

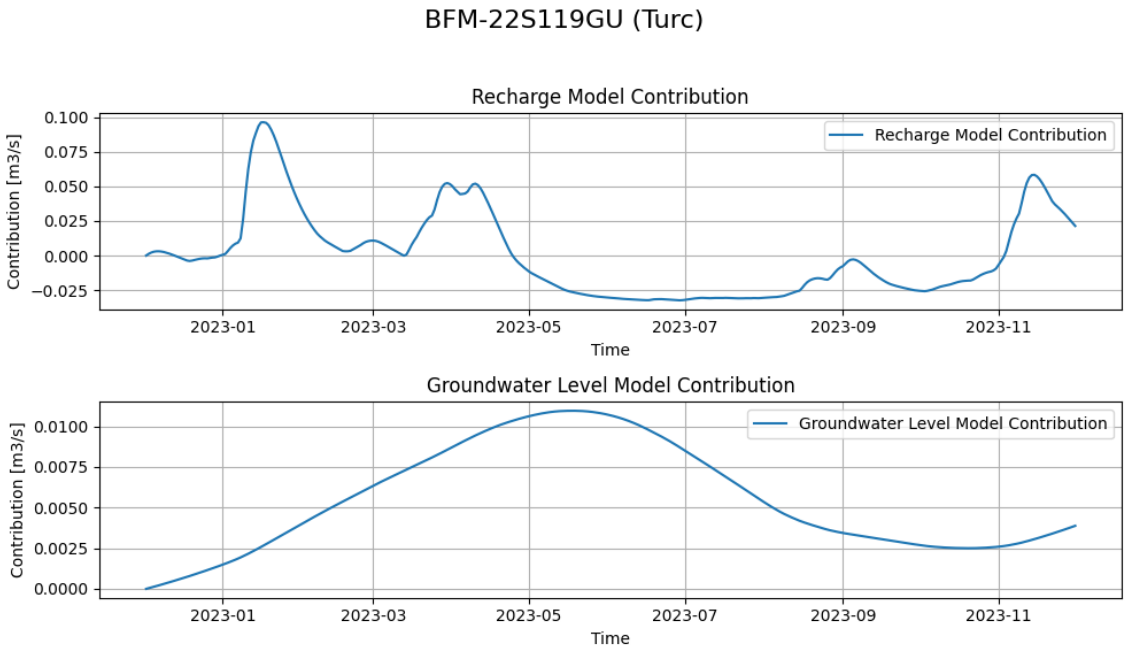


Figure 4.10: Contributions of the two stress models for BFM-22S119GU with Turc's PET method. The data has been adjusted with baseline adjustment, meaning that the initial value is set as a reference value and the absolute variation is presented.

Table 4.9: Top three results for each of the six baseflow models. Best models highlighted in green. Mean result for the best model of each GWM is shown in the last row.

Baseflow Model Results

GWM	PET method	R^2 Calibration	R^2 Validation
22S108GU	Hamon	0.94	-0.40
	Turc	0.94	-0.06
	Jensen-Haise	0.89	-0.39
17HB103	Blaney-Criddle	0.93	-0.53
	Hamon	0.91	-0.48
	Jensen-Haise	0.92	-0.38
17HB102	Penman	0.82	-0.26
	Hamon	0.93	-0.61
	Oudin	0.96	-0.50
22S107GU	Hamon	0.92	-0.67
	Haude	0.87	-0.68
	Turc	0.94	0.03
22S119GU	Turc	0.96	0.62
	Abtew	0.83	0.05
	Oudin	0.93	-0.40
17HB104	Hamon	0.89	-0.40
	Jensen-Haise	0.91	-0.33
	Oudin	0.91	-0.48
Mean (of best models)		**0.90**	**-0.06**

4.2.2 Total Flow Models (TFM)

For the TFMs, the aim was to develop models capable of simulating the total flow in the stream, therefore the total measured stream flow, i.e., the blue series in Figure 4.9, was used as a time series for calibration. These models also included two stress models, a *RechargeModel* with a *Gamma* response function and accounting for groundwater uptake and presence of snow, and a normal *StressModel* with a hindcasted groundwater head series and a *Gamma* response function. Like the BFMs, each TFM was tested with the 18 different PET methods, and the three best model results were selected for evaluation. The results for the six total flow models, with the three best PET methods for each, are presented in Table 4.10.

The best-performing TFM is TFM-22S107GU with Priestley-Taylor's PET method, and the contribution for each of the two stress models, in the calibration period, is presented in Figure 4.11. Running the model with only a *RechargeModel* results in R^2=0.85. With both stress models added there is a slight improvement, with R^2=0.88.

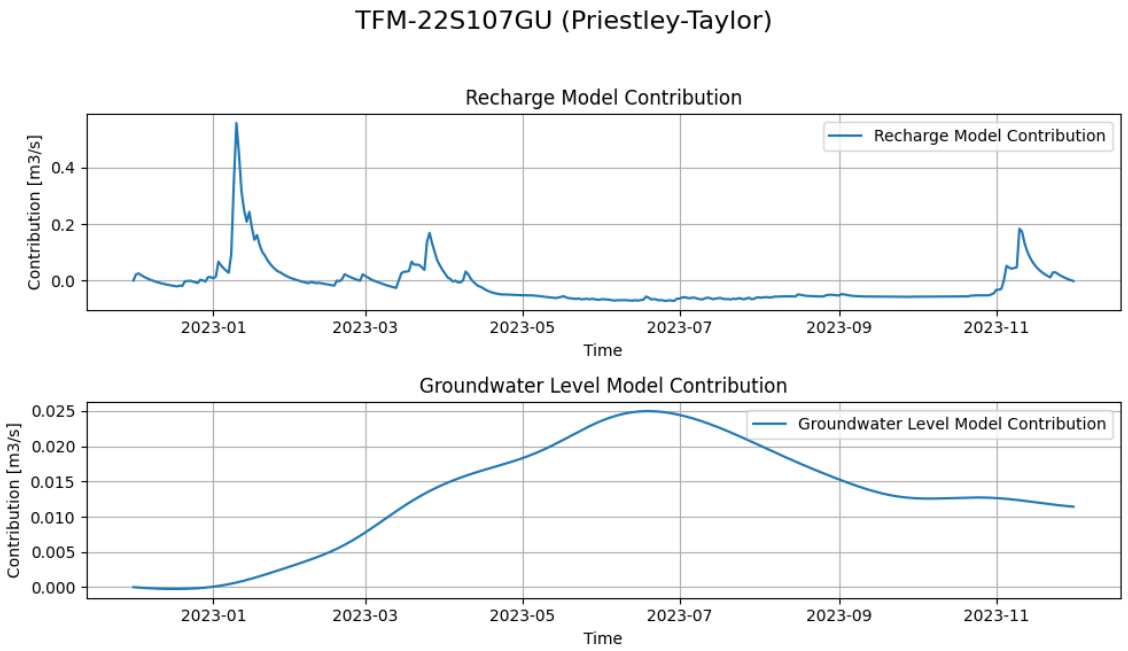


Figure 4.11: Contributions of the two stress models for TFM-22S107GU with Priestley-Taylor's PET method. The data has been adjusted with baseline adjustment, meaning that the initial value is set as a reference value and the absolute variation is presented.

Table 4.10: Top three results for each of the six total flow models. Best models highlighted in green. Mean result for the best model of each GWM is shown in the last row.

Total Flow Model Results

GWM	PET method	R^2 Calibration	R^2 Validation
22S108GU	Turc	0.90	0.64
	Jensen-Haise	0.94	0.45
	Hargreaves	0.86	0.48
17HB103	FAO-56	0.85	0.49
	Blaney-Criddle	0.90	0.46
	Hargreaves	0.88	0.47
17HB102	Turc	0.89	0.52
	Jensen-Haise	0.90	0.63
	FAO-24	0.89	0.61
22S107GU	Priestley-Taylor	0.88	0.66
	Turc	0.92	0.47
	Abtew	0.85	0.52
22S119GU	Linacre	0.88	0.43
	Jensen-Haise	0.89	0.49
	Oudin	0.92	0.40
17HB104	Penman	0.91	0.36
	Turc	0.92	0.38
	Oudin	0.93	0.50
Mean (of best models)		**0.89**	**0.57**

4.2.3 Stream Level Models (SLM)

For the stream level models (SLMs), the time series from diver DF155, placed in the Getå stream, was used after barometric compensation. The data is presented in Figure 4.12.

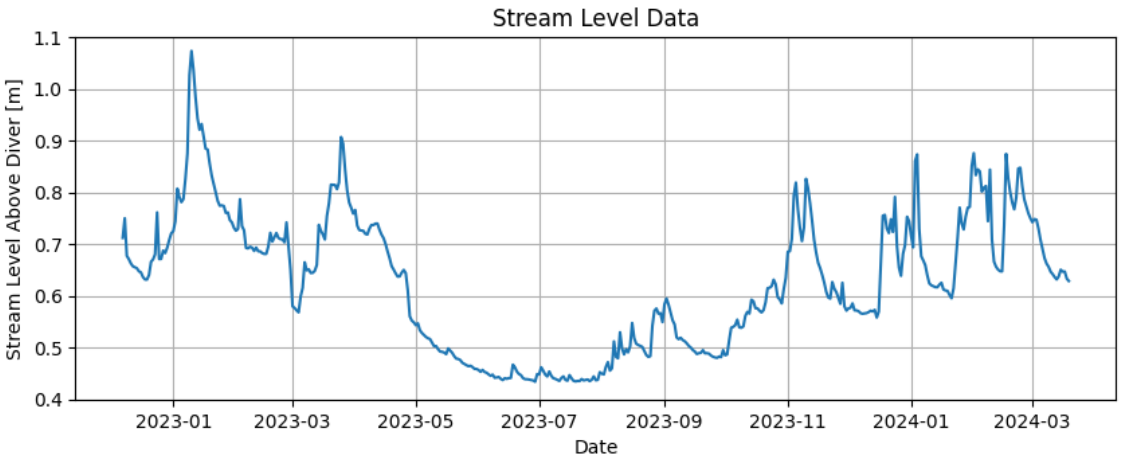


Figure 4.12: Stream level data above the pressure transducer used for calibrating the SLMs.

The SLMs included two stress models like the BFMs and TFMs, a *RechargeModel* with a *Gamma* response function and accounting for groundwater uptake and presence of snow, and a normal *StressModel* with a hindcasted groundwater head series and a *Gamma* response function. Each SLM was tested with the 18 different PET methods, and the three best model results were selected for evaluation. The results for the six stream level models, with the three best PET methods for each, are presented in Table 4.11.

The best-performing SLM is SLM-22S107GU with Blaney-Criddle's PET method, and the contribution for each of the two stress models, in the calibration period, is presented in Figure 4.13. Running the model with only a *RechargeModel* results in R^2=0.71. With both stress models added there is a considerable improvement, with R^2=0.87.

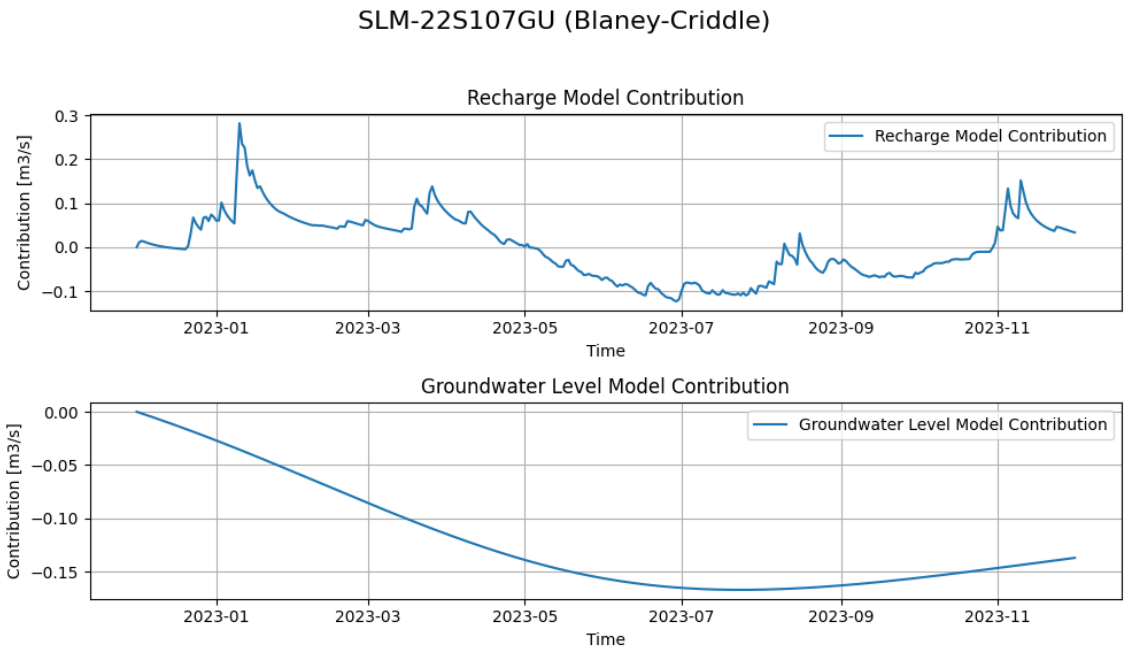


Figure 4.13: Contributions of the two stress models for SLM-22S107GU with Blaney-Criddle's PET method. The data has been adjusted with baseline adjustment, meaning that the initial value is set as a reference value and the absolute variation is presented.

Table 4.11: Top three results for each of the six stream level models. Best models highlighted in green. Mean result for the best model of each GWM is shown in the last row.

Stream Level Model Results

GWM	PET method	R^2 Calibration	R^2 Validation
22S108GU	Blaney-Criddle	0.88	0.52
	Hargreaves	0.85	0.54
	Abtew	0.81	0.46
17HB103	Priestley-Taylor	0.84	0.58
	Jensen-Haise	0.80	0.65
	Hargreaves	0.84	0.55
17HB102	Penman	0.71	0.62
	Hargreaves	0.85	0.52
	FAO-24	0.72	0.59
22S107GU	Blaney-Criddle	0.87	0.66
	Jensen-Haise	0.72	0.58
	McGuinness-Bordne	0.88	0.42
22S119GU	Penman	0.70	0.48
	Kimberly-Penman	0.74	0.50
	Abtew	0.71	0.28
17HB104	Blaney-Criddle	0.81	0.65
	Turc	0.79	0.56
	FAO-24	0.80	0.56
Mean (of best models)		**0.80**	**0.60**

5
Discussion

In this chapter the results from both modelling stages are analysed and discussed. The method is evaluated further by analysing uncertainties in the models and limitations in the modelling approach. Lastly, further development and application of the method is discussed.

5.1 Model Evaluation: GWM

The GWMs in modelling stage 1 perform quite well in estimating the groundwater head, with all values higher than R^2=0.90 in the calibration period. In the validation period, however, some models perform worse with values around R^2=0.70 and below, see Table 4.8. Two factors that influence the results and that contribute to this variation can be identified. The first one is the type of well from which the time series was collected. During model development, the models created from time series originating from wells driven in rock proved to be considerably harder to model. Using a *TarsoModel* for these wells continuously produced better results, being better at simulating the threshold effect behaviour that those time series displayed. The time series for those wells often displayed drastic peaks and troughs at certain levels, compared to soil well time series that had smoother changes. This could be related to the fracture network in the rock, its connectivity with other aquifers and surface water bodies, as well as the fractures intersecting with the well. Therefore, using that stress model for recharge was preferred in models based on wells in rock. However, even with this optimisation procedure of finding the best stress model for the data, it can be seen in Table 4.8 that the GWMs for 17HB103, 17HB102 and 17HB104 performed worse as a group compared to 22S108GU, 22S107GU and 22S119GU that used data from wells in soil. Both model variations use the same input data, precipitation and PET, so it can therefore be concluded that with this simplified approach, it is easier to model groundwater head variation for a well in soil, than a well in rock.

The second factor that can be identified is the length of the time series. It is intuitively logical that the more data a model has for calibration, the better it will be at representing the real system. This can be observed in the results for the GWMs for 22S119GU and 17HB104 in Figure 4.8, which were calibrated during summer and autumn, and validated during winter. The seasonal variations can be significant in hydrogeological contexts and should be considered when choosing calibration periods. Using a calibration period of one full year allows the model to adapt to seasonal patterns, making it better at predicting future variations.

Regarding the choice of PET, it is difficult to draw specific conclusions about which methods are best to use. The conclusion that could be drawn from the results in the first stage is that the methods Blaney-Criddle, McGuinness-Bordne and Hargreaves seem to produce good results for this application.

5.2 Model evaluation: BFM, TFM, and SLM

The performance of the second stage models varies significantly. The results from the first modelling stage are important to consider when evaluating the results from the second stage, since the models from the first stage were used to extend the existing groundwater head series.

Results for the baseflow models in Table 4.9 are good in the calibration period with R^2 values above 0.85, but very poor in the validation period with most values being even negative. A possible explanation for this is the baseflow separation procedure performed before creating the models. The different methods used in the Baseflow package yield very different results, and do not consider specific local conditions. The methods are simple and only use one parameter, the size of the catchment area. Using modelled baseflow as a calibration time series therefore means that the BFMs are calibrated against a potentially poor data set. Therefore, using baseflow data that has been modelled through baseflow separation theory does not seem to be a good approach for this type of application due to the inability to account for local conditions.

The total flow models perform significantly better than the baseflow models in the validation periods, with the best models producing results higher than R^2=0.60. The performance in the calibration period was good, with values higher than R^2=0.85. A main difference between the BFMs and TFMs is that they should capture different system behaviours. The BFMs should capture less drastic changes in the flows, as it should represent slower processes. The TFMs, on the other hand, should capture drastic flow rate changes impacted by precipitation events. This could mean that different PET methods could work better for each model type. For the BFMs, methods like Hamon, Oudin, Turc and Jensen-Haise, repeatedly perform better than other methods, but still the overall results for the BFMs are not satisfactory. For the TFMs, Turc and Jensen-Haise perform well repeatedly, but Hamon does not appear at all in the top results. Most methods used for estimating PET were developed for agricultural purposes, and the methods use different parameters. What can be concluded within the scope of this thesis is that in this type of application, the different methods have a significant impact on the result. To be able to draw more specific conclusions about which methods to use in the context of hydrogeology and for certain local conditions, it is recommended that these methods are evaluated further in future research. The best-performing methods in this study could be compared more in detail, to identify trends by looking at, e.g., maximum and minimum values and time lag. Additionally, they could be tested in different models, representing different site conditions, to see the resilience and consistency of the methods.

The stream level models performed slightly worse than the BFMs and TFMs in the calibration period, but still produced good results overall with R^2 values above 0.80. In the validation period, however, the SLMs performed as good as the TFMs with R^2 values around 0.60. This means that the developed models capture most of the variations in stream level and stream total flow, indicating that it is possible to use this approach to describe the impact of groundwater drawdown on the stream. However, there does seem to be some room for improvement since there is a discrepancy between the calibration and validation results for the TFMs and SLMs, indicating some degree of overfitting.

From the graphs showing the contributions of each stress model, Figures 4.10, 4.11 and 4.13, it can be concluded that the stress models related to groundwater head have a minor impact on the daily variations in stream flow and baseflow. For stream level, however, it seems like the groundwater head stress model has a larger contribution. The low contribution of the groundwater head stress model might be due to the recharge stress model representing the system's response to precipitation and PET, where groundwater head is already included. The added groundwater head stress model might be adding some specific information, hence the minor increase in model performance. One interesting aspect is that groundwater head seems to have a positive contribution during summer for the BFMs and TFMs, but a negative contribution for the SLMs. Unfortunately, a deeper analysis of these trends could not be done within the scope of this thesis due to time constraints.

This type of data-driven modelling using impulse-response functions has been done in previous research by Long (2015), where a similar model based on time series was developed for simulating stream flow. In the study, it was concluded that it is possible to accurately simulate stream flow using this approach, and the benefits of using data-driven modelling in hydrogeological contexts are highlighted further. However, the study also highlights that in complex systems, like in karst, more complex models might be required. Another study by Lu et al. (2022) also explored groundwater-surface water interactions using two steps and focusing on specifically baseflow. The baseflow separation techniques applied in the study are more advanced than the ones used in this thesis, as they combined several methods and were able to adapt the baseflow separation more to local conditions. The study concludes that the simulated groundwater inflow captures some of the variations in the baseflow, but not the peaks, and that the differences are related to delayed water release from other storages. More detailed techniques for baseflow separation therefore seem to improve the model performance.

It is worth noting that the used metric, R^2, considers all variations, including the peaks, which can be quite significant in the stream flow series. Other metrics might need to be used to continue the optimisation process and finding the best model variations, which is discussed further in the following Section 5.3.

5.3 Uncertainties and Method Limitations

Models are simplified representations of a complex reality, and always include some degree of uncertainties. This is especially true when modelling in several steps, where uncertainties can accumulate and have a significant impact on the results. In the method that has been developed in this study, this is particularly relevant in the second stage model results. All models related to 17HB104 and 17HB103 should be evaluated carefully, as the results for these two GWMs produced R^2 values below 0.60 in the first stage.

Another challenge in modelling is overfitting, which can occur during model development when the model is allowed too much flexibility in its parameters and fits the calibration excessively (IBM, 2024). This results in a model that yields good results initially in calibration, but that performs worse when trying to predict values in validation. This was observed continuously during the iterative model development process. Some of the model parameters in Pastas produce good calibration results but poor validation results. It is therefore very important to understand the parameters in the models and validate during the model development. This was clearly observed in the BFMs, where most calibration results were above R^2=0.90, but most validation results were negative.

Another important factor to consider is the choice of metric for model evaluation. In this method the coefficient of determination, R^2, was chosen as it offers the opportunity to quickly compare different model variations. However, using R^2 as a single metric for model evaluation can be misleading as it does not show bias or overfitting, and it is therefore better to combine R^2 with other metrics (Ford, 2015). If the data for the case study was to be analysed further, other and more specific measures could be used used for evaluating results. This would make it possible to draw more specific conclusions. One of these measures is $MAE_{0.2}$, which is the mean absolute error for the lowest 20 % of values (Ahmed, 2023). The most important aspect for the case study is how the stream might be affected during the most vulnerable periods of the year, which will occur during the summer months when the water levels are at their lowest. Using this metric would improve the model optimisation process, by identifying which model variation captures this behaviour best. Another suggestion is evaluating model results using Spearman's rank correlation coefficient, which is non-parametric and uses monotonic functions to describe correlation between two variables, meaning that the variations can be studied instead of absolute values (Laerd Statistics, 2018).

5.4 Further Development and Application

The method that has been developed within the scope of this study is an initial data-driven method that can be used to model stream flow rate and stream level. For the specific case study, the next step could be to simulate different leakage rates in the tunnel, to see how the groundwater and the stream is affected by leakage to the

tunnel. Doing this, and hence getting a better understanding of the groundwater-surface water interactions of the site, would allow for a clearer communication in the permit process. In terms of modelling, this could mean performing a pump test in the area and use the results as calibration data in the models. An additional stress model could then be added to the models to simulate the water loss due to leakage, since pumping has the same effect on groundwater level as leakage by lowering it. The impact on the stream could then be simulated.

Another possible improvement for the method is to use longer time series by continuing the measurements on the site. With longer calibration periods the models would become better at representing the hydrogeological behaviour at the site. Combining this with more metrics for model evaluation, as mentioned in Section 5.3, would improve the optimisation process. A part of this is understanding which PET methods are most suited for different conditions. In all results in both modelling stages, using different PET methods for recharge produces very different modelling results.

An important aspect to study further is the applicability of the developed method, i.e., to identify in which type of projects the method could be used. The developed data-driven approach is simpler in several aspects than process-based models that require more specific details of the site. The aim of the data-driven approach is to develop a more general conceptual understanding of the behaviour of the groundwater at the site. It is easier to implement and quicker to obtain results, but it is not possible to, e.g., see in detail how a specific point or small area will be affected by a change in the system, which is possible in a process-based numerical model. The choice of method therefore depends on several factors, e.g., what is being studied, what level of detail is needed, and available resources. It might be possible that different modelling approaches could be used within the same project, but at different stages. In large infrastructure projects, like the one used as a case study here, groundwater levels are often monitored over longer periods of time, which is the only thing needed in the proposed method alongside basic meteorological data. However, data resolution needs to be considered, as this approach works better with daily values. Monitoring programs in early investigation programs might only collect groundwater head data once a week or once a month. This type of data-driven modelling approach does not aim at replacing existing modelling, but should rather be seen as an additional tool for representing natural, complex hydrogeological systems.

6
Conclusions

The data-driven method developed in this study consists of two modelling stages. Results from the first stage show that with the limited data-approach used, the method is better for modelling groundwater head changes in soil, compared to in rock. The length of the calibration period also proved to have a significant impact on the results, with longer calibration periods performing better.

In the second modelling stage, it was seen that modelling the baseflow yielded poor results, with only a few models producing positive R^2-values in validation. This is assumed to be related to the baseflow separation which introduces additional uncertainties to the modelling. The TFMs and SLMs produced better results, with mean results in validation of R^2=0.57 and R^2=0.60. The TFMs and SLMs both showed signs of overfitting with higher calibration results at 0.89 and 0.80. Overall, this means that the models capture most of the variations in stream level and stream flow, and describe the groundwater-surface water interactions fairly well.

The analysis of PET methods shows that the different methods can have a significant impact on the results. For the GWMs, Blaney-Criddle, McGuinness-Bordne and Hargreaves were best, while Turc and Jensen-Haise were better for the BFMs, TFMs and SLMs. Further analysis is recommended to better understand which methods work best under different conditions.

Noteworthy limitations in the method that can be considered in further work are the metrics used for model evaluation. R^2 could be used with additional metrics like $MAE_{0.2}$ and residual plots that would improve the model performance and understanding. Another improvement for further work is using longer time series for longer calibration periods, to give the models more calibration data and see model performance across more seasons.

The developed method constitutes a simplified and data-driven approach for analysing groundwater-surface water interactions. The approach works better for simulating total stream flow and stream level, compared to stream baseflow. Some improvements, like using additional metrics and pump test data, could be made to increase method performance and reliability. The data needed for this type of approach is relatively easy to obtain, making it an additional relevant tool for representing and understanding hydrogeological systems. Overall, with the proposed further development, the method has good potential to be used for describing the groundwater-surface water interactions of the case study and other similar projects.

References

Abtew, W. (1996). Evapotranspiration measurements and modeling for three wetland systems in south Florida. *Journal of the American Water Resources Association (JAWRA)*, *32*(3), 465–473. https://doi.org/10.1111/J.1752-1688.1996.TB04044.X

Ahmed, M. W. (2023). Understanding Mean Absolute Error (MAE) in Regression: A Practical Guide. https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration* (tech. rep.). Food and Agriculture Organization (FAO).

Anderson, M. P., Woessner, W. W., & Hunt, R. J. (2015). *Applied Modelling Groundwater Simulation of Flow and Advective Transport*. Elsevier.

Bakker, M., & Post, V. (2022). *Analytical Groundwater Modeling; Theory and Applications using Python*. CRC Press.

Blaney, H. F., & Criddle, W. D. (1950). Determining water requirements in irrigated areas from climatological and irrigation data.

Britannica. (2024). Coefficient of determination. https://www.britannica.com/science/coefficient-of-determination

Collenteur, R. A., Bakker, M., Calje, R., & Schaars, F. (2021). *Pastas Documentation Release 0.17.0* (tech. rep.).

Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., & Schaars, F. (2019). Pastas: Open Source Software for the Analysis of Groundwater Time Series. *Groundwater*, *57*(6), 877–885. https://doi.org/10.1111/GWAT.12925

Collenteur, R. A., Bakker, M., Klammler, G., & Birk, S. (2021). Estimation of groundwater recharge from groundwater levels using nonlinear transfer function noise models and comparison to lysimeter data. *Hydrol. Earth Syst. Sci*, *25*, 2931–2949. https://doi.org/10.5194/hess-25-2931-2021

European Environment Agency. (2001). groundwater recharge. https://www.eea.europa.eu/help/glossary/eea-glossary/groundwater-recharge

Fetter, C. W., & Kraemer, D. (2022). *Applied Hydrogeology* (5th). Waveland Press.

Ford, C. (2015). Is R-squared Useless? | UVA Library. https://library.virginia.edu/data/articles/is-r-squared-useless

Frost, J. (2024). How To Interpret R-squared in Regression Analysis. https://statisticsbyjim.com/regression/interpret-r-squared-regression/

Haaf, E., Giese, M., Reimann, T., & Barthel, R. (2023). Data-Driven Estimation of Groundwater Level Time-Series at Unmonitored Sites Using Comparative

Regional Analysis. *Water Resources Research*, *59*(7). https://doi.org/10.1029/2022WR033470

Hamon, W. R. (1963). Estimating Potential Evapotranspiration. *American Society of Civil Engineers Transactions*, *128*(3415).

Hargreaves, G. H., & Samani, Z. A. (1983). Estimating Potential Evapotranspiration. *Journal of Irrigation and Drainage Engineering*, *109*(3), 341–343. https://doi.org/10.1061/(ASCE)0733-9437(1983)109:3(341)

Haude, W. (1955). Determination of evapotranspiration by an approach as simple as possible. *Mitt Dt Wetterdienst*.

Hipel, K. W., & McLeod, A. I. (1994). Chapter 17 Constructing Transfer Function-Noise Models. *Developments in Water Science*, *45*(100), 573–616. https://doi.org/10.1016/S0167-5648(08)70679-8

IBM. (2024). What is Overfitting? | IBM. https://www.ibm.com/topics/overfitting

Jensen, M. E., Burman, R. D., & Allen, R. G. (1990). Evapotranspiration and irrigation water requirements: a manual. *ASCE manuals and reports on engineering practice (USA). no. 70.*, 332.

Jensen, M. E., & Haise, H. R. (1963). *Estimating evapotranspiration from solar radiation* (tech. rep. No. IR4).

Kalbus, E., Reinstorf, F., & Schirmer, M. (2006). Hydrology and Earth System Sciences Measuring methods for groundwater-surface water interactions: a review. *Hydrol. Earth Syst. Sci*, *10*, 873–887. www.hydrol-earth-syst-sci.net/10/873/2006/

Knotters, M., & De Gooijer, J. G. (1999). TARSO modeling of water table depths. *Water Resources Research*, *35*(3), 695–705. https://doi.org/10.1029/1998WR900049

Laerd Statistics. (2018). Spearman's Rank-Order Correlation. https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

Länsstyrelsen. (2014). *Skötselplan för naturreservatet Getåravinen* (tech. rep.).

Linacre, E. T. (1977). A simple formula for estimating evaporation rates in various climates, using temperature data alone. *Agricultural Meteorology*, *18*(6), 409–424. https://doi.org/10.1016/0002-1571(77)90007-3

Long, A. J. (2015). RRAWFLOW: Rainfall-response aquifer and Watershed Flow Model (v1.15). *Geoscientific Model Development*, *8*(3), 865–880. https://doi.org/10.5194/GMD-8-865-2015

Lu, M., Rogiers, B., Beerten, K., Gedeon, M., & Huysmans, M. (2022). Exploring river-aquifer interactions and hydrological system response using baseflow separation, impulse response modeling, and time series analysis in three temperate lowland catchments. *Hydrology and Earth System Sciences*, *26*(13), 3629–3649. https://doi.org/10.5194/hess-26-3629-2022

Makkink, G. F. (1957). Testing the Penman formula by means of lysimeters. *Journal of the Institution of Water Engineers*.

McGuinness, J. L., & Bordne, E. F. (1972). A comparison of lysimeter derived potential evapotranspiration with computed values.

Oudin, L., Michel, C., & Anctil, F. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? *Journal of Hydrology*, *303*(1-4), 275–289. https://doi.org/10.1016/J.JHYDROL.2004.08.025

Penman, H. L. (1947). Natural evaporation from open water, bare soil and grass. *158*, 285. https://royalsocietypublishing.org/

Priestley, C. H. B., & Taylor, R. J. (1972). *On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters* (tech. rep.).

Rodhe, A. (1987). *The Origin of Streamwater Traced by Oxygen-18* [Doctoral dissertation].

Romanenko, V. A. (1961). Computation of the autumn soil moisture using a universal relationship for a large area. *Proc. of Ukrainian Hydrometeorological Research Institute*.

SGU. (2023a). Kartvisaren Jordarter 1:25 000-1:100 000. https://www.sgu.se/produkter-och-tjanster/kartor/kartvisaren/jordkartvisare/jordarter-125-000-1100-000/

SGU. (2023b). Kartvisaren Jorddjup. https://www.sgu.se/produkter-och-tjanster/kartor/kartvisaren/jordkartvisare/jorddjup/

SGU. (2024). Sveriges berggrund. https://www.sgu.se/om-geologi/berg/sveriges-berggrund/

SMHI. (2020). Om tjänsten Modelldata per område. https://www.smhi.se/data/hydrologi/vattenwebb/ladda-ner-modelldata-per-omrade-1.116164

Sophocleous, M. (2002). Interactions between groundwater and surface water: The state of the science. *Hydrogeology Journal*, *10*(1), 52–67. https://doi.org/10.1007/S10040-001-0170-8/METRICS

Sundell, J., Haaf, E., Tornborg, J., & Rosén, L. (2019). Comprehensive risk assessment of groundwater drawdown induced subsidence. *Stochastic Environmental Research and Risk Assessment*, *33*(2), 427–449. https://doi.org/10.1007/S00477-018-01647-X/FIGURES/16

Swanson, J., & Babcock, B. N. (2023). Beneath the Surface: An Overview of Grouting Applications for Successful Tunnel Grouting — TBM: Tunnel Business Magazine. https://tunnelingonline.com/beneath-the-surface-an-overview-of-grouting-applications-for-successful-tunnel-grouting/

Thom, A. S., & Oliver, H. R. (1977). On Penman's equation for estimating regional evaporation. *Quarterly Journal of the Royal Meteorological Society*, *103*(436), 345–357. https://doi.org/10.1002/QJ.49710343610

Trafikverket. (2022). *Program Ostlänken - Affärsplan* (tech. rep.).

Trafikverket. (2023a). Nya stambanor mellan Stockholm, Göteborg och Malmö. https://www.trafikverket.se/vara-projekt/projekt-som-stracker-sig-over-flera-lan/nya-stambanor-mellan-stockholm-goteborg-och-malmo/

Trafikverket. (2023b). Ostlänken, en ny dubbelspårig järnväg. https://www.trafikverket.se/vara-projekt/projekt-som-stracker-sig-over-flera-lan/ostlanken-en-ny-dubbelsparig-jarnvag/

Trafikverket. (2023c). *PM miljökvalitetsnormer för vatten Ostlänken delen Stavsjö-Loddby* (tech. rep.).

Trafikverket. (2024). The East Link, a new doubletrack railway. https://bransch.trafikverket.se/en/startpage/projects/Railway-construction-projects/ostlanken-east-link-project/

Turc, L. (1961). Water requirements assessment of irrigation, potential evapotranspiration: Simplified and updated climatic formula. *Annales Agronomiques*, *12*, 13–49.

Vremec, M., Collenteur, R., & Birk, S. (2024). PyEt v1.3.1: a Python package for the estimation of potential evapotranspiration. https://gmd.copernicus.org/preprints/gmd-2024-63/

Vremec, M., Collenteur, R. A., & Birk, S. (2023). HESSD - Technical note: Improved handling of potential evapotranspiration in hydrological studies with PyEt. https://hess.copernicus.org/preprints/hess-2022-417/

Wright, J. L. (1982). New Evapotranspiration Crop Coefficients. *Journal of the Irrigation and Drainage Division*, *108*(1), 57–74. https://doi.org/10.1061/JRCEA4.0001372

Xie, J., Liu, X., Wang, K., Yang, T., Liang, K., & Liu, C. (2020). Evaluation of typical methods for baseflow separation in the contiguous United States. *Journal of Hydrology*, *583*, 124628. https://doi.org/10.1016/J.JHYDROL.2020.124628

Yoo, C. (2016). Ground settlement during tunneling in groundwater drawdown environment – Influencing factors. *Underground Space*, *1*(1), 20–29. https://doi.org/10.1016/J.UNDSP.2016.07.002

Zarei, H. R., Uromeihy, A., & Sharifzadeh, M. (2011). Evaluation of high local groundwater inflow to a rock tunnel by characterization of geological features. *Tunnelling and Underground Space Technology*, *26*(2), 364–373. https://doi.org/10.1016/J.TUST.2010.11.007

A

Appendix A - Programming with Python: All scripts

All the coding was done using Python version 3.11.8 in Jupyter Notebooks in the Visual Studio Code software. The artificial intelligence tool ChatGPT model 4.0 was used continuously during the iterative modelling process, mainly for troubleshooting code. All the scripts used in the project are published on the following GitHub repository: **https://github.com/J-AltiGran/hydro_thesis.git**

The input data is not open-source and could not be added to the GitHub repository. The file structure in the repository is presented in Figure A.1.

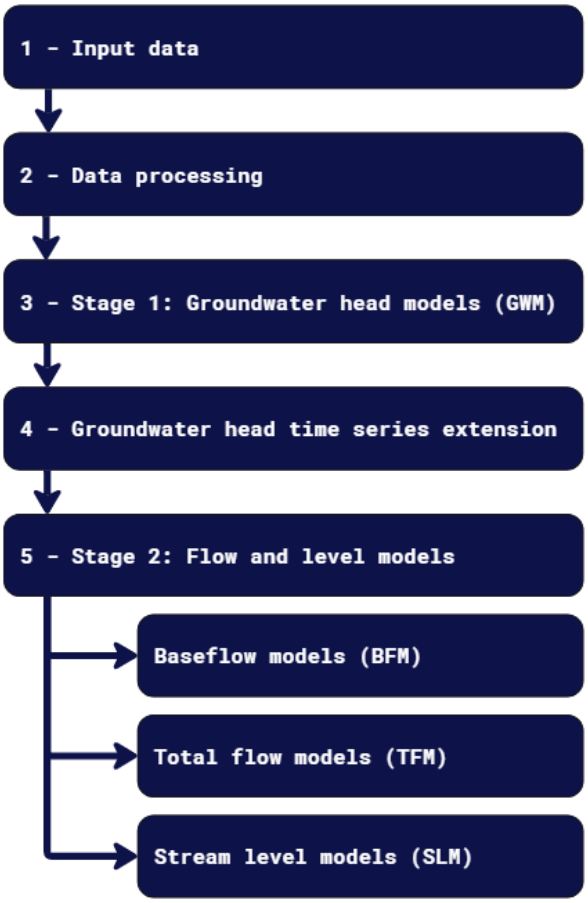


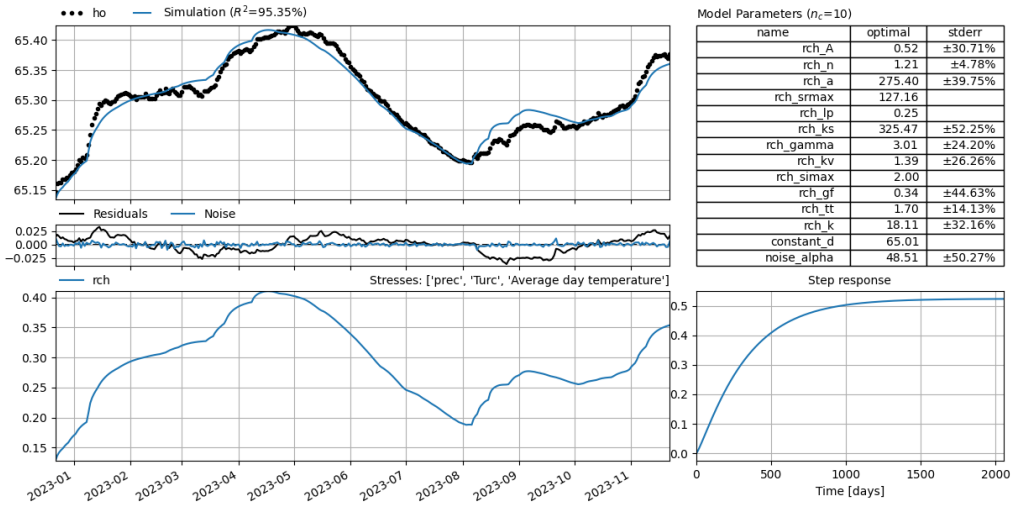
Figure A.1: File structure of the project in the GitHub repository.

B

Appendix B - Calibration results GWMs

This appendix presents the calibration results for the groundwater head models (GWMs) in the first modelling stage. The best model variation for each GWM is presented. As mentioned in Appendix A, all scripts for the project are published on GitHub.

B.1 GWM-22S108GU



name	optimal	stderr
rch_A	0.52	±30.71%
rch_n	1.21	±4.78%
rch_a	275.40	±39.75%
rch_srmax	127.16	
rch_lp	0.25	
rch_ks	325.47	±52.25%
rch_gamma	3.01	±24.20%
rch_kv	1.39	±26.26%
rch_simax	2.00	
rch_gf	0.34	±44.63%
rch_tt	1.70	±14.13%
rch_k	18.11	±32.16%
constant_d	65.01	
noise_alpha	48.51	±50.27%

Figure B.1: Calibration results for GWM-22S108GU.

B.2 GWM-17HB103

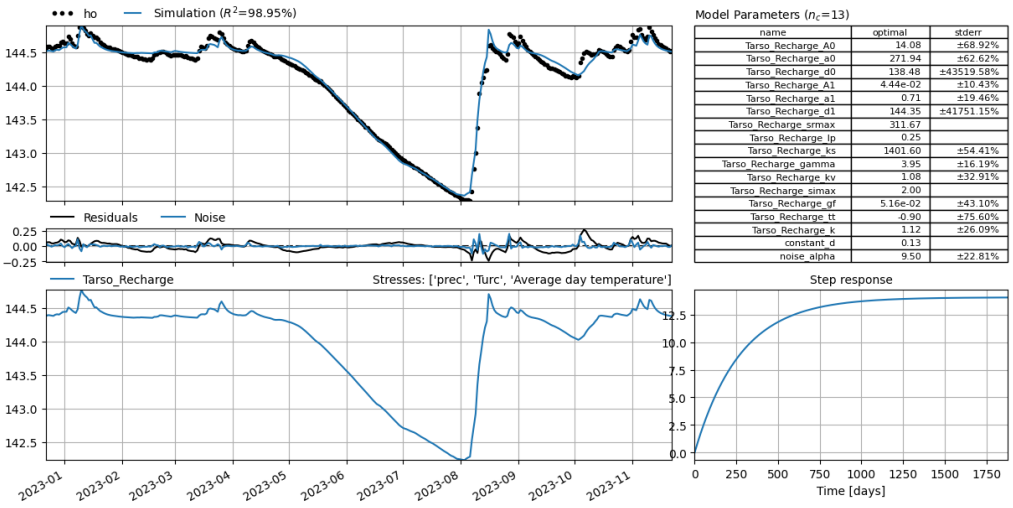


Figure B.2: Calibration results for GWM-17HB103.

B.3 GWM-17HB102

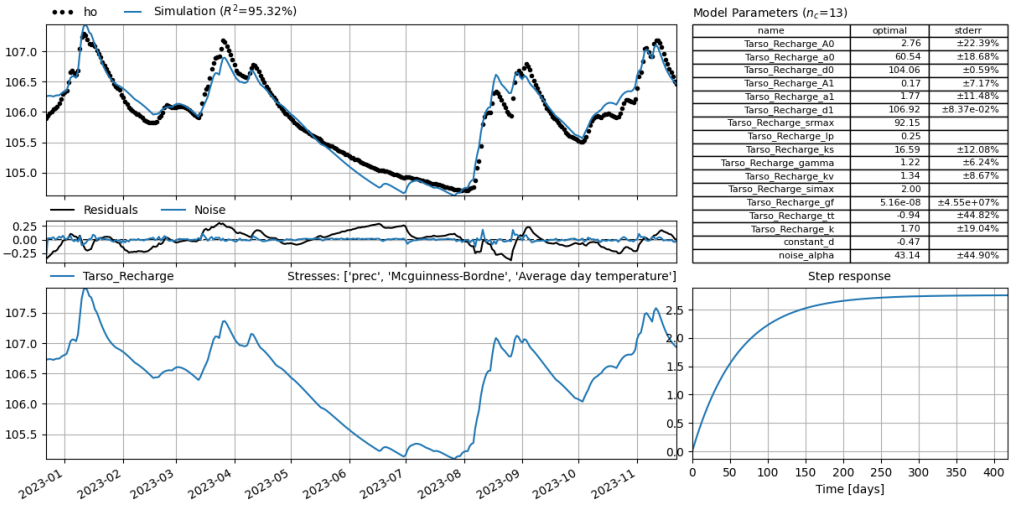


Figure B.3: Calibration results for GWM-17HB102.

B.4 GWM-22S107GU

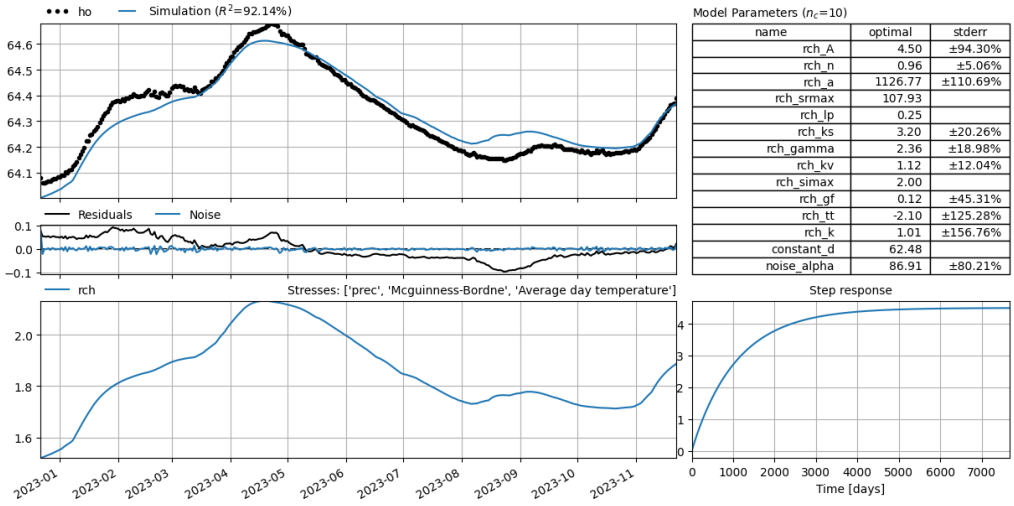


Figure B.4: Calibration results for GWM-22S107GU.

B.5 GWM-22S119GU

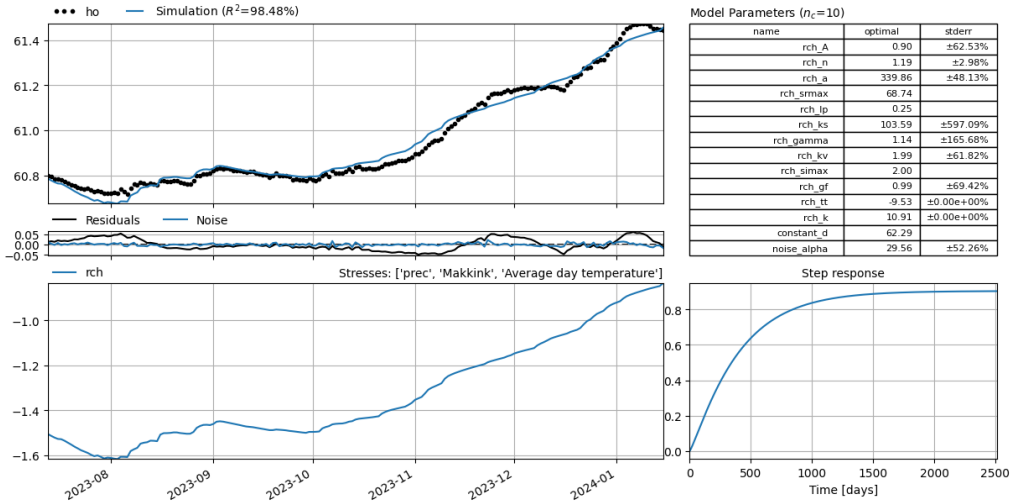


Figure B.5: Calibration results for GWM-22S119GU.

B.6 GWM-17HB104

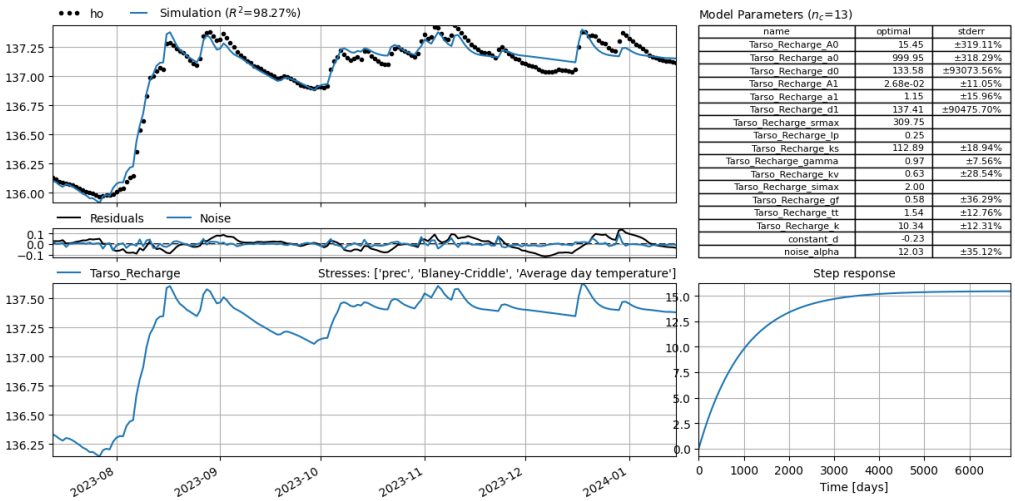


Figure B.6: Calibration results for GWM-17HB104.

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY