HISTORICAL METEOROLOGICAL TIME SERIES IN TRANSFER FUNCTION-NOISE MODELLING: HOW EXTREME ARE CURRENT GROUNDWATER LEVELS?

MSc Thesis

Civil Engineering and Management

Author

Wout Schutten

Date

28-02-2024

COLOPHON

TITLE
Historical meteorological time series in transfer function-noise modelling: how extreme are current groundwater levels?

AUTHOR
W.A. Schutten

EMAIL
w.a.schutten@student.utwente.nl
woutschutten@gmail.com

VERSION
Final report

DATE
28-02-2024

INSTITUTION
University of Twente – Enschede
Civil Engineering and Management
Faculty of Engineering Technology

INTERNAL SUPERVISORS
Dr. Ir. H.J. Hogeboom
Dr. Ir. D.C.M Augustijn

UNIVERSITY OF TWENTE.

COMPANY
HKV lijn in water – Amersfoort
Water en klimaat

EXTERNAL SUPERVISORS
Dr. Ir. M. Pezij
U.N. Jungermann MSc

HKV LIJN IN WATER

PREFACE

I am proud to present my graduation research to complete my master's degree in Civil Engineering and Management at the University of Twente. From September 2023 to February 2024, I studied the extremity of current groundwater droughts using long-term groundwater simulations.

Throughout my MSc journey, my passion for hydrology and water management steered my academic pursuits. There was just one important part missing: groundwater. Its hidden nature, combined with the recent dry periods in the region in the Netherlands where I live, sparked a passion to understand this crucial resource and the growing relevance of groundwater droughts.

Conducting my research at HKV provided me valuable insights into the groundwater system. Furthermore, working with an unfamiliar model broadened my horizons, addressing aspects not covered in my coursework. I also learnt a lot about communicating, writing, and managing a research report. Therefore, I would like to thank Michiel Pezij and Nicole Jungermann from HKV for their guidance during this process. Their expertise in groundwater modelling and climate series really helped me. In addition, the feedback and critical questions stimulated me to grow intellectually and made sure I got the most out of this study. The new office in Amersfoort was also very comfortable, so I always enjoyed going there.

I would also like to thank Denie Augustijn and Rick Hogeboom from the University of Twente for their guidance. Especially their support in the process and their critical view of the approach ensured a smooth research process. Special thanks to Bart Strijker of HKV for his insights into groundwater statistics, which gave me food for thought.

Finally, I would like to thank my friends, family and girlfriend for their support and encouragement that kept me motivated throughout this journey. Graduating has been a rewarding experience, despite the occasional challenges. My research has convinced me that there is still considerable work to be done for a robust groundwater system in the Netherlands. I am therefore looking forward to putting my knowledge into practice.

I hope you enjoy reading this thesis.

Wout Schutten

Amersfoort, February 2024

SUMMARY

Groundwater is an important part of the water availability in the Netherlands and serves as a necessary source for agriculture, industry, drinking water and nature. To protect the groundwater availability, it is crucial to monitor groundwater levels as this provides insight into current groundwater conditions. The dry summer of 2018 demonstrated that sufficient groundwater availability cannot be guaranteed every year, in large parts of the Netherlands. Among other things, a changing climate, increased groundwater withdrawals and changes in the water system have made the Netherlands vulnerable to groundwater droughts. Therefore, it is important to have insight in the long-term behaviour of groundwater levels, to make optimal and justified decisions on groundwater policy and the implementation of structural measures. However, the limited availability and applicability of long-term groundwater observations hinders the understanding of long-term groundwater fluctuations, the classification of extreme dry conditions, and the frequency of extreme drought events.

This study presents long-term groundwater levels for four locations in the Netherlands. For the first time, a data-driven time series model is combined with historical meteorological data to simulate groundwater levels for the period 1910-2022. The time series analysis is conducted using a transfer function-noise model, in which fluctuations of groundwater levels are modelled by precipitation and evaporation data. The simulated groundwater levels are historical projections based on the current climate and groundwater system, which allows for different drought statistics to be derived and compared with groundwater observations. The findings of this study indicate that groundwater observations over the last eight years do not provide a comprehensive representation of long-term groundwater levels. Long-term groundwater simulations offer improved insights, particularly in estimating return periods of droughts. Furthermore, the choice of drought statistic plays a crucial role the characterisation of drought intensity, duration, and frequency.

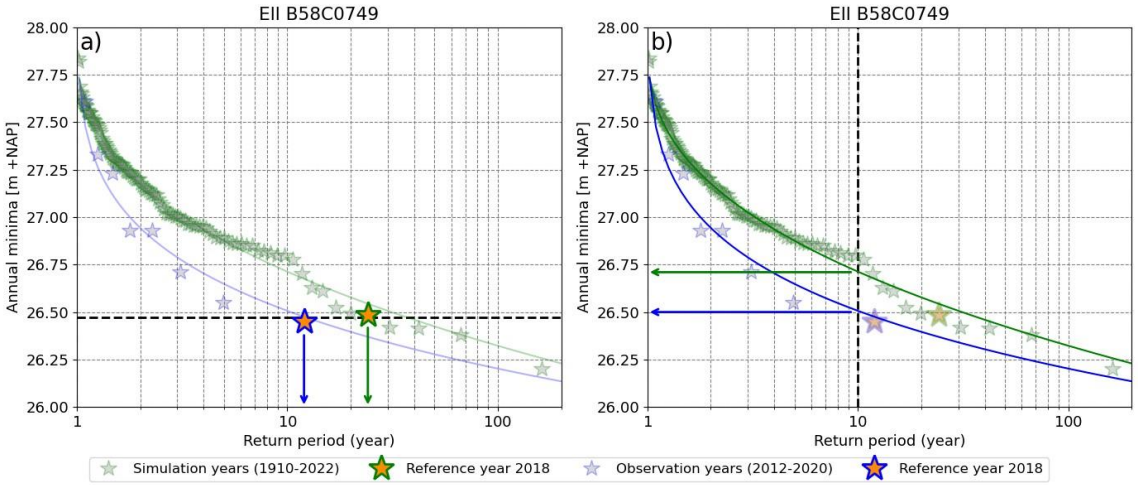


Figure 1: Differences in return periods between groundwater observations (2012-2020) and long-term simulations.

The figure above shows an example of the drought characterisation using annual minimum groundwater levels. The drought of 2018 occurs less frequently than expected based on observations (Figure 1a). Groundwater levels like those observed in the 2018 drought appear every 12 years based on observations. However, long-term simulations suggest a lower frequency, with similar levels occurring only once every 25 years. Moreover, if a measure is justified for droughts that occur on average every 10 years (T10), the intervention would be taken too late or not at all, based on observations (Figure 1b). Therefore, the use of long-term simulations with historical meteorological data gives an improved insight in return periods for current droughts.

CONTENTS

1 INTRODUCTION

The Netherlands has a substantial groundwater reservoir, which serves various crucial purposes, including supply for drinking water, industrial, and agricultural practices. Additionally, groundwater determines the baseflow of streams, water management in urban areas and is necessary for plant growth in nature (Geologische Dienst Nederland, 2023). Phreatic groundwater levels increase when the amount of precipitation exceeds evaporation rates. This establishes a seasonal pattern characterised by higher groundwater levels during the winter and lower groundwater levels during the summer, due to increased evaporation. Phreatic groundwater also drains to surface water and deeper layers. Imbalances in groundwater levels, whether too high or too low, lead to various challenges and issues.

Hydrological droughts can occur as a result of prolonged periods of precipitation deficit, leading to a decline in groundwater levels due to insufficient groundwater recharge combined with increased abstraction, posing a significant threat to water availability. The recent droughts in 2018 and 2022 have provided a lot of attention regarding hydrological droughts due to severe damage to a wide range of sectors (Brakkee et al., 2021). In 2018, declining water reserves inflicted substantial damage, mainly to agriculture, but also to navigation, drinking water and industry, with estimated economic losses ranging from 450 to 2080 million euros (Van Hussen et al., 2019). Moreover, the increase in climate variability is expected to lead to even drier years in the future (Philip et al., 2020). In addition, a growing population requires changes in land use and an increased water demand. All these factors put pressure on the groundwater availability. Well-informed groundwater management and insight in the extremes of droughts is needed to overcome these challenges.

The existing design and use of the water system has left areas vulnerable to droughts because groundwater levels decline faster in times of low precipitation. Drainage through pipes, ditches and streams, groundwater withdrawals for drinking water and agricultural irrigation play an important part. In a country that has traditionally prioritised discharging precipitation, the drought in 2018 was perceived as a turning point by many.

To move towards a climate-resilient groundwater system, it is necessary to have a proper understanding of the hydrological state of the water system (van den Eertwegh et al., 2021). Effectively mitigating the effects of droughts depends on gaining insight into (extremely) low groundwater levels. This is because groundwater is the most persistent water storage in the landscape and the last to respond when a meteorological drought propagates through the hydrological system (Van Loon, 2013).

The analysis of groundwater conditions is based on groundwater measurements but can be improved by using groundwater models, which are valuable tools to describe groundwater level fluctuations. Among these models, time series modelling is commonly used. The time series model used in this study is a transfer function-noise model (TFN). TFN modelling uses regression to establish a relationship between various stress factors, such as precipitation and evaporation, and observed groundwater levels. This relationship can be used to extend or interpolate observed time series of groundwater levels. The use of precipitation and evaporation data is constrained by the length of observed meteorological time series. To extend observed time series of groundwater levels, it becomes necessary to use climate models or synthetic meteorological data. In this study, a detrended historical meteorological series is used, developed by HKV (Pezij & Lugt, 2023).

This study is the first to use a time series model with historical meteorological data to provide long-term groundwater levels for four monitoring wells in the Netherlands.

1.1 PROBLEM CONTEXT

Monitoring the status of groundwater resources is a crucial component of drought management. Droughts are assessed as deviations relative to 'normal' conditions. Groundwater measurements are the primary source to determine this reference. Therefore, the amount of groundwater measurements determines the description of current groundwater conditions, i.e. to determine when it is dry. Therefore, it is crucial to have information about long-term groundwater levels, which requires a minimum 30-year measurement period (van den Eertwegh et al., 2021). To identify extreme droughts, even longer series would be required. For instance, extreme value statistics necessitate a minimum sample size of 50 years (McCluskey et al., 2021). However, due to limited quality and amount of available groundwater measurements, describing groundwater droughts remains challenging (Brakkee et al., 2021).

In practice, reliable long-term groundwater measurements (over a period of 50 years) are hardly available in the Netherlands (Ritzema et al., 2012; van den Eertwegh et al., 2021). 110 groundwater monitoring wells have measurements for more than 50 years, while only 6 wells have measurements exceeding 80 years (Verhagen & Avis, 2021). Furthermore, the locations of these measurements are unevenly distributed across the Netherlands, such that there are large parts of the country where there is insufficient coverage with time series longer than 50 years. Most importantly, these historical groundwater measurements are not representative of the current situation due to the changes in the climate and water system, so that only measurements from recent years can be used to describe groundwater droughts. This results in a limited understanding of groundwater droughts, and thus lack in describing:

- Long-term (> 100-year) fluctuations of groundwater levels.
- Classification of extreme dry conditions.
- Frequency of groundwater droughts.

As a result, it is currently difficult to assess the severity and occurrence of groundwater levels during summer months, when the groundwater levels drop due to precipitation deficits. This lack of insight in describing extreme groundwater levels hinders the implementation of appropriate measures such as irrigation restrictions.

Additionally, the use of meteorological data in time series modelling shows potential for simulating long-term groundwater level time series (El Mezouary et al., 2020; Vonk, 2021). However, it remains uncertain whether long-term simulations contribute to the understanding of extreme groundwater levels and improve the characterisation of groundwater droughts. Also used for the first time are historical detrended meteorological data developed by HKV (Pezij & Lugt, 2023). However, it is unsure whether this dataset is appropriate to use to simulate long-term groundwater levels.

1.2 RESEARCH OBJECTIVE AND QUESTIONS

The objective of this study is:

To improve the understanding of the severity and occurrence of groundwater droughts by simulating long-term groundwater levels for the period 1910-2022 that represent the current climate and groundwater system.

The research objective can be translated into the following main research question:

> **Main research question:** To what extent can groundwater measurements describe the severity and occurrence of groundwater droughts?

And the following sub-questions:

> **RQ1**: How do different drought statistics describe relevant properties of groundwater droughts?
>
> **RQ2**: To what extent do the climatological properties of the historical meteorological time series match the current climate?
>
> **RQ3:** How can you use historical meteorological time series to simulate long-term groundwater heads?
>
> **RQ4:** How do drought properties differ when drought statistics are derived from long-term groundwater levels?

1.3 READING GUIDE

The theoretical background of this study is provided in the second chapter. The third chapter presents the research steps, study area and data. In this, the first research question is already answered by providing a methodology to describe groundwater droughts. Chapter 4 shows the simulation of long-term groundwater series. In chapter 5, drought statistics of observed and simulated groundwater levels are derived and compared. This is followed by a discussion of the validity and interpretation of the results in Chapter 6. Chapter 7 contains the main conclusions and chapter 8 makes recommendations for further research and practical use.

2 THEORETICAL BACKGROUND

The theoretical background is elaborated in the following paragraphs. First, background information is given about groundwater and the factors influencing groundwater levels. Next, the different stages of droughts are presented. In the fourth paragraph, various statistical methods are described. Lastly, the transfer function-noise model is explained.

2.1 GROUNDWATER WITHIN THE HYDROLOGICAL CYCLE

The hydrological cycle of the Earth consists of various processes (Figure 2). Groundwater is a major storage within the hydrological cycle. The cycle starts with evaporation of water. Evaporation occurs from soil and surface water. Vegetation also transpires, since water vapour escapes from stomata when plants photosynthesise. Therefore, the sum of this process is referred to as evapotranspiration. Evaporated water comes back as precipitation. Precipitation either flows over the surface towards lakes and rivers (surface runoff) or infiltrates into the soil. This process of infiltration saturates the soil by filling up space between soil particles. The part of the ground which is not fully saturated is known as the unsaturated or vadose zone and the water in the pores soil moisture. Water in parts of the soil that are fully saturated is called groundwater. Infiltrated water can percolate further towards deeper groundwater storages (aquifers).

The phreatic groundwater table marks the boundary between the saturated and unsaturated zones of the soil. Depending on the difference between groundwater table and surface water level groundwater can drain to surface water. However, this movement is relatively slow (Hoekstra, 2018), the average residence time of deep groundwater (100 to 200 years) is significantly longer than that of soil moisture (1 to 2 months), lakes (50 to 100 years), and rivers (2 to 6 months) (Pidwirny, 2006). This makes groundwater act as a buffer, being replenished in times of rainfall surpluses and being depleted gradually over time. The deep groundwater flows so slow, that it is often referred to as a finite resource, since extractions easily surpass replenishing rates (Hoekstra, 2018).

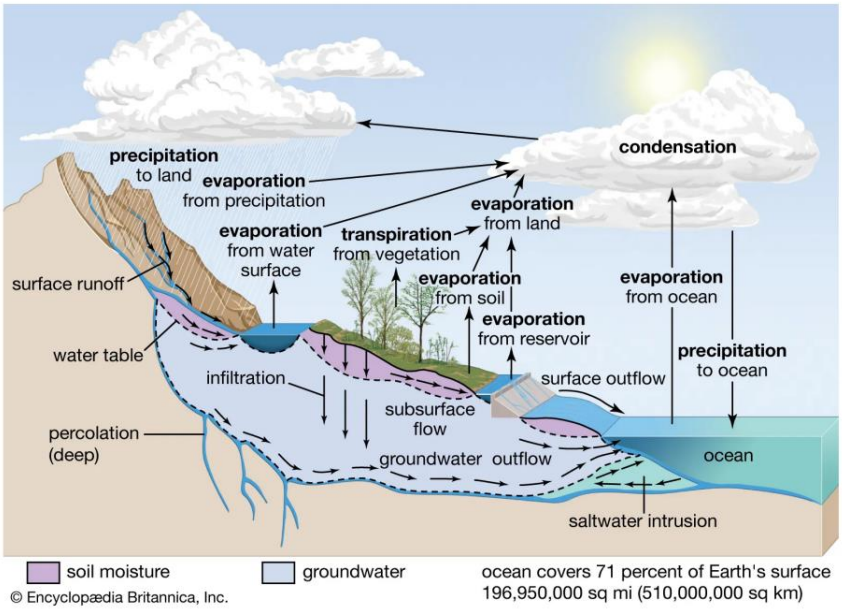


Figure 2: Hydrological cycle (The editors of Encyclopaedia Britannica, 2023).

2.2 GROUNDWATER MONITORING

Groundwater measurements are the principal source of information about groundwater levels. Four common methods exist to measure the phreatic groundwater level: monitoring wells, piezometers, open boreholes and field estimates, of which the first two are the most reliable (Ritzema et al., 2012). Groundwater monitoring wells are shallow monitoring wells that measure a head that deviates little from the phreatic groundwater level. Piezometers are monitoring wells that measure heads in deeper soil layers (Bouma et al., 2012). Field estimates are estimates of physical groundwater levels and mainly a tool to determine the location of open boreholes and not suitable for validation purposes (Ritzema et al., 2012).

Groundwater levels are dynamic and constantly vary due to short and long-term changes in meteorological conditions, climate, groundwater withdrawals and land use (Alley & Taylor, 2002; Ritzema et al., 2012). Therefore, it is important to measure groundwater levels at multiple locations. The location of groundwater measurements is key to reliable observations. Groundwater levels can vary spatially due to soil types and elevation differences, nearby surface water (Figure 3), groundwater withdrawals, presence of hard surfaces and drainage. Therefore, monitoring wells should be placed in a safe and protected location where the groundwater levels are representative for the surrounding area (Bouma et al., 2012).

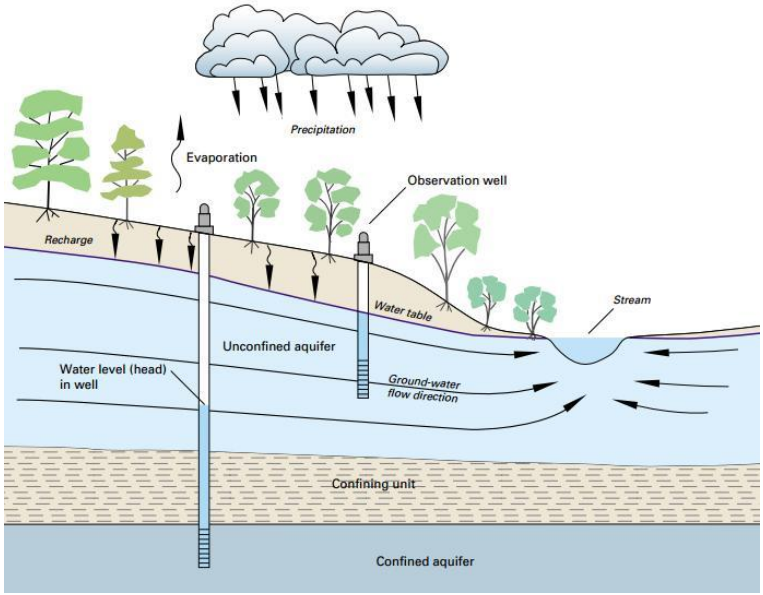


Figure 3: Subsurface groundwater flow, observation well and some hydrological elements (Alley & Taylor, 2002).

2.3 CLASSIFICATION OF DROUGHTS

Droughts are recurrent features of the climate that results from a shortfall in precipitation or increase in evapotranspiration over an extended period of time (European Drought Observatory, 2023), and leads to an unusual and temporary deficit in the water availability of water storages. A deficit occurs when the demand for groundwater is greater than the availability of groundwater. Generally, droughts can be classified in several categories, depending on their impact and prevailing effect on the hydrological system.

Meteorological droughts refer to an extensive period of abnormal precipitation deficit, in relation to the long-term average conditions for a region. The KNMI uses precipitation deficit or the Standardised Precipitation Index (SPI) to indicate meteorological droughts (KNMI, 2021b). The SPI is a statistical indicator comparing the actual precipitation with the long-term precipitation distribution for the same period at that location.

When a meteorological drought continues for longer periods and leads to a soil moisture deficit, the result is a soil moisture drought. In this situation, the drought limits the water availability for natural vegetation and crops. The European Drought Observatory uses the Combined Drought Indicator (CDI) to identify areas affected by soil moisture droughts. This indicator is computed by combining the anomalies of precipitation, soil moisture and satellite-measured plant growth (European Drought Observatory, 2023).

A hydrological drought is related to the effects on surface or sub-surface water supply, such as below-normal reservoir and lake levels, groundwater levels and declining river discharge.

The last type of drought is socio-economic drought, which occurs when the shortfall in water availability results in a disruption of the water resources system, such that the supply does not meet the economic demand. It can also refer to the ecological or health-related impacts of droughts.

Droughts are known to propagate over time (Schumacher et al., 2022; Van Loon, 2013), since a prolonged precipitation deficiency generates less input to the hydrological system.

2.4 STATISTICAL METHODS

Statistical analysis involves the characterisation of groundwater level occurrences. These characteristics encompass central tendencies (mean, median, mode), dispersion (standard deviation, variance), and the assessment of extreme event probabilities. Statistical methods serve the purpose of describing and making inferences about groundwater level behaviour. No single optimised method exists to derive groundwater statistics. Therefore, within the theoretical background of this study, various statistical methods are described and used to characterise groundwater droughts. The methods include percentiles, average low groundwater level (GLG), Standardised Groundwater Index (SGI) and extreme value statistics (EVA). The following sections provide a description of the aforementioned statistics.

2.4.1 Percentiles

A percentile is a statistical measure that indicates the percentage of data points below a specific value. For example, the 50th percentile (also known as the median) represents the middle point of the dataset, where 50% of the data points are below this value. The value of the p^{th} percentile can be determined using the following:

$$P = \frac{p}{100} * N \tag{1}$$

Where P is the position of the sorted groundwater levels that belongs to percentile p using N observations. In this method, current groundwater levels are compared with other groundwater levels over a specified period, making use of percentiles. Interpreting this comparison offers an assessment of whether the measured groundwater levels are dry, moderate, or wet, depending on their percentile rank, giving insight into the drought conditions of current groundwater levels. Percentiles are currently used in some databases of provinces and regional water authorities in the Netherlands (such as Province of Brabant and Waterschap Rijn en IJssel) and in the database of the Geologische Dienst Nederland to display groundwater levels (Zaadnoordijk et al., 2019).

2.4.2 Average low groundwater Level (GLG)

Statistics can also be used to describe the average annual fluctuations of groundwater levels, often referred to as grondwater dynamics or groundwater regime. The annual groundwater fluctuations can be characterised by using certain parameters, such as the average low (GLG) and average high (GHG) groundwater levels (van der Gaast et al., 2010) (Figure 4). Alongside the average low and high groundwater levels, parameters also exist for the Average Spring Groundwater level (GVG) and Average Groundwater level (GG). The collective term for all these parameters is the GXG.

Figure 4: Describing the annual fluctuations of groundwater using the parameters GLG and GHG.

The most useful parameter for droughts is the average low groundwater Level (GLG) since it can be interpreted as the average summer groundwater level and is often used as a general indicator of low groundwater levels. The GLG is therefore not indicating an extremely dry situation (Zaadnoordijk et al., 2019). The GLG can be defined as the average of the three lowest annual groundwater levels, averaged over a period of at least eight years:

$$GLG = \frac{\sum_1^n LG3}{n} \quad (2)$$

Where $LG3$ is the average of the three lowest groundwater levels in a hydrological year [m +NAP], and n is the number of years whose $LG3$ has been determined which should be at least eight hydrological years during which no hydrological interventions have taken place. The GLG is used in combination with the GHG in the Dutch method “grondwatertrappen” (de Gruijter et al., 2004; Hoogland et al., 2014; Knotters et al., 2018; van Kekem et al., 2005), involving various stages of combinations of average low and average high groundwater levels to provide a nationwide description of the groundwater dynamics (Knotters et al., 2018). The GLG statistic is also used to indicate low groundwater levels in the database of the Geologische Dienst Nederland (Zaadnoordijk et al., 2019).

2.4.3 Standardised Groundwater Index (SGI)

The Standardised Groundwater Index (SGI) standardises the fluctuations of groundwater time series to characterise groundwater droughts (Bloomfield & Marchant, 2013). Its primary purpose is to standardise groundwater data by transforming groundwater levels into a standardised index. The method is similar to the Standardised Precipitation Index, an index used to evaluate precipitation deficits (Devesa, 2023; KNMI, 2021b).

The calculation of the index values is based on nonparametric transformation, instead of distribution fitting (Bloomfield & Marchant, 2013). This means that the measured groundwater heads are transformed to a normal distribution, and minimal assumptions are made about the underlying distribution of the data. For each calendar month with *n* observations, probability values are uniformly spaced over interval (1/2*n*) to (1 – 1/2*n*) (Brakkee et al., 2021). The corresponding SGI values are found by applying a normal inverse cumulative distribution function to these values (Marchant & Bloomfield, 2018). The resulting SGI values are assigned to the groundwater levels by their rank from low to high (Brakkee et al., 2021), which can be formulated as:

$$SGI = F^{-1}(p) \tag{3}$$

Where F^{-1} is the inverse of the cumulative function of a normal distribution and p is the probability between 0 and 1. So basically, for every month, uniformly spaced probability values are assigned to groundwater levels, and these are transformed to a standardised index value using an inverse of the cumulative function of a normal distribution. SGI values can be understood as standardised variations from the historical mean. Positive SGI values indicate groundwater levels above average, while negative values signify groundwater levels below the historical average.

The SGI can be determined for different time scales, reflecting different types of droughts (Guo et al., 2021). The SGI is often determined for periods of 90 days (3 months; SGI-3), 30 days (1 month; SGI-1) or 10 days (SGI-0). Therefore, the value of the index is an average of the preceding 90, 30 or 10 days (Bloomfield & Marchant, 2013; van den Eertwegh et al., 2021). The *droogteportaal* uses different time scales of the SGI to show the actual drought conditions for monitoring wells in the Netherlands (van den Eertwegh et al., 2021)[1].

2.4.4 Extreme value analysis (EVA)

Extreme value analysis involves fitting an extreme value distribution to the tail of a probability distribution. EVA deals with the characterisation of the tails of a distribution, where extreme observations occur, such as extreme low groundwater levels. Extreme value statistics aims to predict probabilities for rare events using the probability of exceedance. Generally, two main approaches for extreme value statistics exist, which are Block Maxima (BM) and Peak Over Threshold (POT). When a block of one year is selected, annual minima (or maxima) are obtained. When applied to groundwater measurements, the Annual Minima (AM) technique consists of sampling the minimum groundwater level for every calendar or hydrological year. This is a simple method to determine the general trend of the lowest groundwater levels on a yearly basis (Škarpich et al., 2016).

An extreme value distribution can be fitted with the selection of (annual) minima values. The selection of the most suitable distribution is based on the Akaike Information Criterion (AIC) value, which measures the fit of a distribution in relation to other distributions. For Block Maxima extreme values, the best fit is often a generalized extreme value (GEV) distribution function:

$$GEV(\mu, \sigma, \zeta) = e^{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}} \tag{4}$$

Where μ is the location parameter [-], σ is the scale parameter [-] and ξ is the shape parameter [-]. If $\xi = 0$, the distribution is known as a Gumbel distribution. Extreme value analysis is a commonly used method for other hydrological events such as precipitation, streamflow, discharge, and lake levels (Dawley et al., 2019; Katz et al., 2002).

[1] www.droogteportaal.nl

2.5 TRANSFER FUNCTION-NOISE MODELLING

Transfer function-noise modelling is a form of time series analysis that estimates output series using one or multiple input series, based on impulse response functions and a noise model. A relationship is sought between observed groundwater levels and input stresses, like precipitation and evaporation, in the form of impulse response functions. The goal of the modelling is to define the impulse response functions using observational data and use those in combination with other input stress data to simulate groundwater levels. Many studies have shown the effectiveness of using transfer function-noise modelling in groundwater studies (Collenteur et al., 2021; El Mezouary et al., 2020; Mohanasundaram et al., 2017; Pezij et al., 2020; Ratering, 2023; Von Asmuth et al., 2002; Vonk, 2021; Zaadnoordijk et al., 2019). This data-driven approach is commonly simpler and faster than developing a numerical groundwater model (M. Bakker & Schaars, 2019) and does not need prior assumptions on model structure (Peterson & Western, 2014). Furthermore, the response of a water system to input stresses is included, which increases our understanding of the effects of an input stress (Pezij et al., 2020). Lastly, the stochastic system dynamics are modelled using a noise model (Von Asmuth et al., 2002). Commonly used tools for time series analysis are Menyanthes (von Asmuth et al., 2012), Hydrosight (Peterson & Western, 2014) and Pastas (Collenteur et al., 2019).

Pastas will be used in this study and is an open-source python package developed by TU Delft, Artesia and TU Graz (Collenteur et al., 2019). The objective of Pastas is to provide a scientific framework to develop and test new methods and to provide a tool for groundwater research. Pastas implements transfer function-noise models to simulate time series of groundwater heads using Predefined Impulse Response Functions in Continuous Time (PIRFICT) (Von Asmuth et al., 2002), in which time series are modelled as follows:

$$h(t) = \sum_{m=1}^{M} h_m(t) + d + r(t) \tag{5}$$

In which $h(t)$ is the observed groundwater head [m], $h_m(t)$ is the contribution of input stress m (impulse response function), d is the base level of the model, and $r(t)$ is the model residual (Collenteur et al., 2019). This formula describes the impact of different stresses to observed groundwater heads, comprising of a base level, response functions of different input stresses and residuals. The base level d is a constant parameter without physical meaning which is used to correct the model. Sometimes the parameter is considered the base drainage, the equilibrium level of groundwater when all input stresses are zero. However, this interpretation is not always correct. The base level can be determined using an average of the residuals after a simulation, or to include the parameter in the calibration (Collenteur et al., 2019).

The fluctuations in groundwater heads can be caused by a variety of factors. The effect of an input stress can differ depending on the model structure. Therefore, the choice of model structure has impact on the results. There are two types of input stresses, measured and ambient influences. Ambient influences have no available time series which can be used to determine the effect on the groundwater heads, making it more difficult to model the response of the groundwater level for the input stress. These include changing land use or interventions in the water system. These are typically modelled by including a trend which corrects the model for this ambient influence. Measured input stresses are influences with available time series to determine the effect on the groundwater head. Most common examples are precipitation, evapotranspiration, groundwater extractions, surface water levels. The contributions of these influences can be determined using impulse response functions.

2.5.1 Impulse response functions

The contribution of input stresses to the groundwater heads can be determined using impulse response functions ($h_m(t)$ from Eq. 5):

$$h_m(t) = \int_{-\infty}^{t} S_m(\tau)\vartheta_m(t-\tau)d\tau \quad (6)$$

In which S_m is the time series of a certain input stress m and ϑ_m is the impulse response function of m (Von Asmuth et al., 2002). The impulse response function describes the behaviour of groundwater head following an impulse of a certain input stress. Many different impulse response functions exist, depending on the type of input stress and on the study area. A simple impulse response function is the exponential function (Von Asmuth et al., 2021), in which only two parameters have to be fitted to the observed groundwater heads:

$$\vartheta(t) = Ae^{-\frac{t}{a}} \quad (7)$$

A gamma impulse response function has three parameters, and thus can describe the effect of groundwater infiltration better than an exponential function:

$$\vartheta(t) = At^{n-1}e^{-\frac{t}{a}} \quad (8)$$

Other impulse response functions are FourParam, Hantush, Polder, Double Exponential, Edelman and the Hantush Well Model (Collenteur et al., 2019).

The input stresses of precipitation and evaporation can also be combined into one recharge stress model, instead of modelling two separate stresses. Since groundwater heads do not behave the same after an equal impulse of either evaporation or precipitation, there are various approaches to model the recharge stress. These different approaches increase the accuracy because of their ability to describe the recharge flux. One simple method is linear recharge (Obergfell et al., 2019). There are also non-linear recharge stress models like TARSO (Von Asmuth et al., 2021), Berendrecht, Peterson (Peterson & Western, 2014) and FlexModel (Collenteur et al., 2021). The modelling in this study included testing different stress models, however just modelling recharge as precipitation minus evaporation proved to yield the best results (Table 20, Appendix E: Use of different stress models).

2.5.2 Noise model

Another important part of the time series modelling is the noise model. A noise model is used to reduce the model residuals and simulate the stochastic system behaviour. The residual is the difference between the observed and simulated groundwater heads, meaning the groundwater which could not be simulated by responses to the input stresses (***r*** in equation 4). Often, the residuals from the model are correlated, resulting in the residual being dependent on the residual of the previous day. Since this effect can attenuate for several days, the model may simulate excessive groundwater heads. Therefore, correlated noise is implemented using the following formula for exponential decay (Collenteur et al., 2019):

$$n_c(t) = e^{-\frac{1}{\alpha}} \cdot n_c(t-1) + n(t) \quad (9)$$

In which n_c is the correlated noise, α is the noise decay parameter and n is the uncorrelated (white) noise, result of a random process. The created noise series is added to the observed groundwater head (Collenteur et al., 2019). Without noise model it is impossible to make statistical predictions and analyse uncertainty of parameters (Von Asmuth et al., 2021).

3 METHODOLOGY

3.1 RESEARCH STEPS

The research steps are visualised in Figure 5:

- In the first research question, an overview is given of statistical methods that describe groundwater drought conditions. The results of the first research question are already included in this chapter.
- To use the historical meteorological data for long-term simulations, in the second research question, it must be validated whether it is representative of the current climate.
- In the third research question, long-term groundwater levels are simulated using the validated historical meteorological data and time series analysis. Therefore, a transfer function-noise model is developed using groundwater measurements and observed meteorological data. Then, the fitted model is used to simulate long-term groundwater levels using the historical meteorological data. The long-term groundwater levels are also validated.
- In the last research question, the methodology from the first research question is used to derive drought statistics of groundwater observations and simulations which are compared.

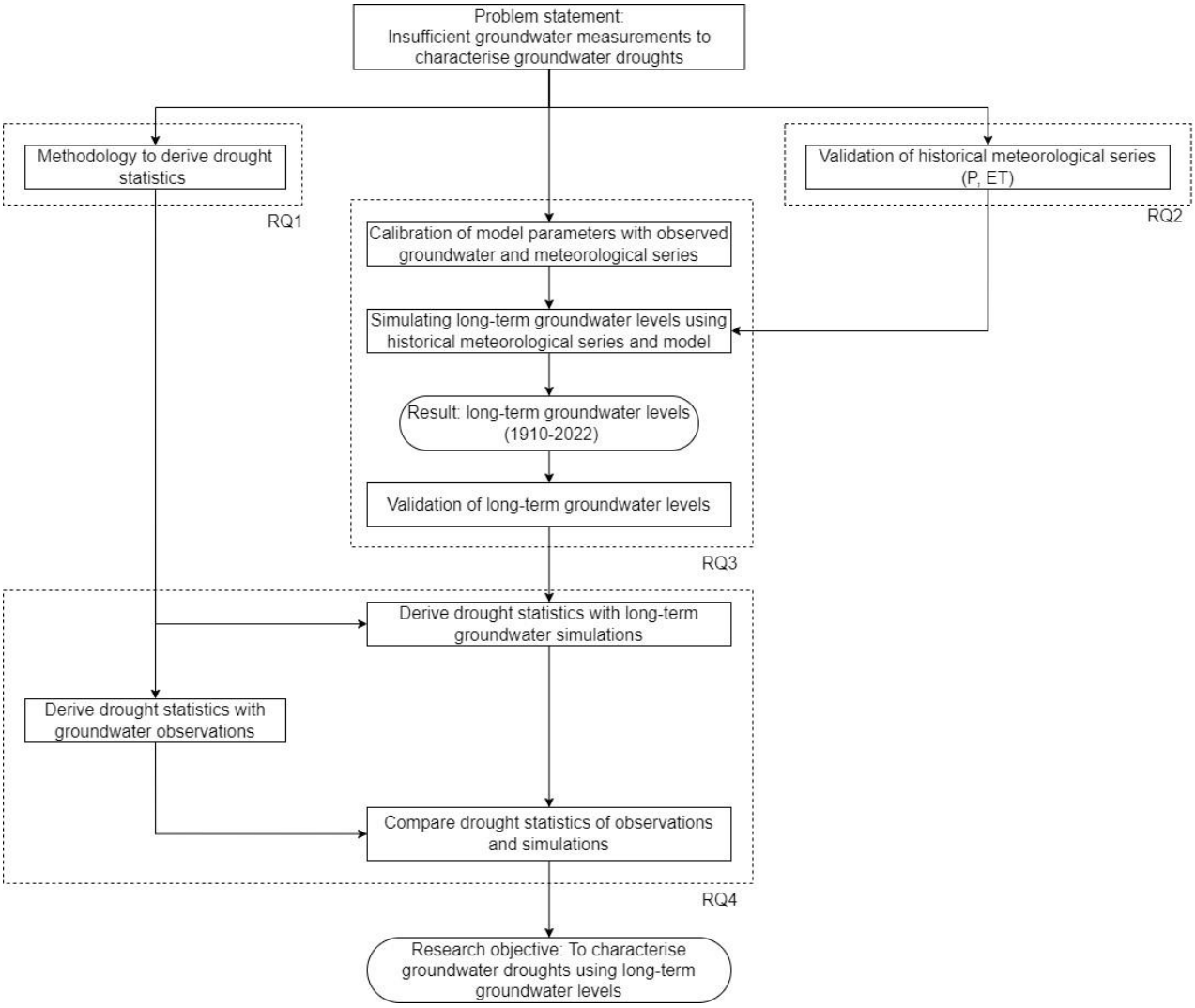


Figure 5: Flow diagram of research steps

3.2 RQ1 – METHODOLOGY TO DERIVE DROUGHT STATISTICS

The objective of the first research question is to explore statistical methods to describe groundwater drought conditions. First, relevant aspects of groundwater droughts are described. This is followed by the elaboration of the research question, as the results of this research question provide a methodology for the rest of the study. This contains the elaboration of the four methods from the theoretical background (section 2.4), describing how they characterise drought conditions.

3.2.1 Characterising drought properties in groundwater levels

The important aspects of droughts are identified to determine the ability of statistics to characterise droughts in groundwater levels. The aim is to provide insight into the duration, intensity and frequency of droughts, as these determine the impact (Petersen-Perlman et al., 2022). The extremity of a groundwater drought can therefore be assessed in terms of its intensity, its duration and its frequency:

- Intensity refers to the severity of the groundwater deficit during the period of the drought.
- The duration of a drought is the length of time during which the groundwater level remains below a defined threshold.
- Frequency describes how often a given level of drought occurs. This can be expressed in terms of return periods or probability of exceedance. The return period is the average time interval between the recurrence of a particular drought event. For example, a return period of 10 years means that, on average, similar drought conditions are expected to occur once every 10 years. The probability of exceedance is the probability that a drought of a given intensity and duration will be exceeded in a given year. For example, a probability of 0.10 [-] means that there is a 10% chance of a drought of that intensity is exceeded each year. This results in an average of one event every 10 years.

The frequency of a drought event can be determined for a certain drought intensity, such as the annual minimum groundwater level. The frequency of drought events is estimated using their plotting position (Benard & Bos-Levenbach, 1954), which is commonly used in hydrology:

$$F_i = \frac{i - 0.3}{N + 0.4} \tag{10}$$

In which F_i is the plotting position for a specific drought intensity, which is the position on a frequency curve. i denotes the rank number of the drought intensity, sorted from high to low, where the lowest values have a rank of 1. N is the total number of data values. The probability of exceedance is the survival function of F_i. The return period T is the inverse of the exceedance probability (Eq. 11). The return period indicates, on average, how often a drought event of a specific intensity and duration can be expected to occur:

$$T = \frac{1}{1 - F_i} \tag{11}$$

3.2.2 Percentiles

Percentiles are determined by first sorting a dataset in ascending order. Percentiles indicate the percentage of data points that fall below a certain value. Since groundwater levels fluctuate throughout the year, percentiles are generally determined per month, by resampling the data based on monthly averages, acknowledging that the groundwater level are typically higher in winter and lower in summer. The calculated percentile values are set to the first of the month. Ranges between two percentile values are indicated as coloured bands. This results in a regime curve that shows the fluctuations of the groundwater, using percentile ranges (Figure 6).

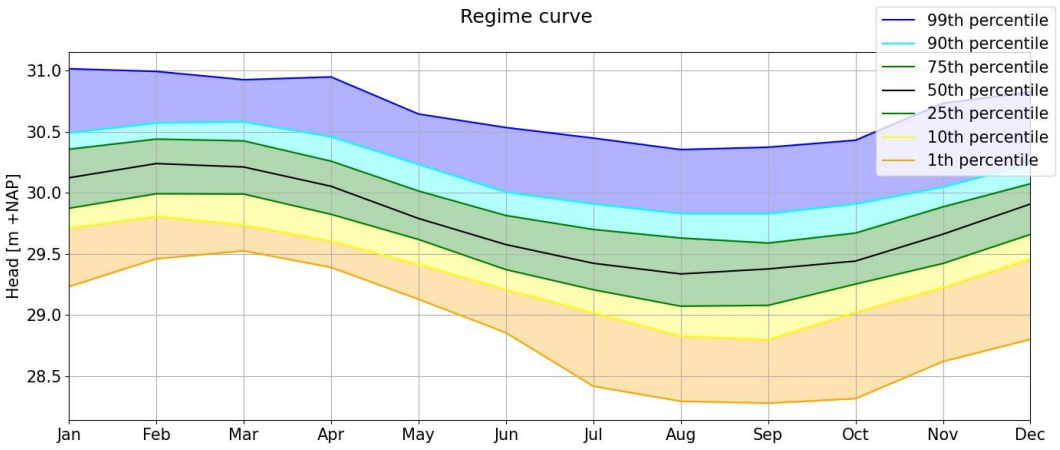


Figure 6: Example of a regime curve. The black line is the 50th percentile, the median. The coloured lines could represent the pth percentile, therefore the percentage of the groundwater levels that are below this line. The range between percentiles are classified to drought conditions.

The classification of drought conditions to percentile values are shown in Table 1. The classification of the extremely dry groundwater levels is based on the current application of percentile methods for groundwater. Some references use the 2.5th percentile (Knotters et al., 2013), 5th percentile (Finke et al., 2001), but most sources use the 10th percentile (Averink, 2013; Zaadnoordijk et al., 2019). The 1st and 99th percentiles are the minimum and maximum occurring groundwater levels.

Table 1: Classification of percentiles.

Classification	**Percentile bands**
Extremely dry	1st to 10th percentile
Dry	10th to 25th percentile
Moderate	25th to 75th percentile
Wet	75th to 90th percentile
Extremely Wet	90th to 99th percentile

By interpreting the regime curve and percentile bands, valuable insights can be gained into the intensity, duration, and frequency of drought events. The percentile bands, along with their corresponding classifications, indicate the intensity of drought. The duration of drought corresponds to the period during which groundwater resides within a particular percentile band. The frequency of drought events depends on the total amount of measurement years and the particular percentile band of the drought. Since percentiles indicate the percentage of groundwater levels that exceed the specified percentile.

3.2.3 Average low groundwater Level (GLG)

The average low groundwater level (GLG) is the average of the 3 lowest groundwater levels per year and can be considered the average summer condition (as elaborated in 2.4.2). Alongside the GLG, the average high (GHG), average spring (GVG) and average groundwater level (GG) will be determined. Current groundwater levels can be plotted alongside a line of the GLG value to identify drought conditions (Figure 7). Next, the intensity is determined by measuring the difference between the lowest value of a drought event and the GLG line. In other words, it quantifies how much below the average summer conditions (represented by a GLG line) the groundwater levels have dropped during a specific drought event. Duration refers to the time period during which the groundwater levels remain consecutively below the GLG line. It measures how long the drought event persists with groundwater conditions falling below the established threshold. Frequency is calculated by

determining the return periods for drought events using the Bernard en Bos-Levenbach plotting positions (section 3.2.1).

Figure 7: Example of characterising droughts. The horizontal line is the GLG, L is the duration of the drought. I is the intensity of the drought, the groundwater levels below GLG. The frequency of the drought is determined at the lowest point.

3.2.4 Standardised Groundwater Index (SGI)

The Standardised Groundwater Index (SGI) is determined by transforming groundwater levels to a normal distribution (as elaborated in 2.4.3). Therefore, SGI values can be understood as standardised variations from the historical mean. Positive SGI values indicate groundwater levels above the historical average, while negative values indicate groundwater levels below the historical average. The SGI value reflects the degree of this deviation and is used to classify drought conditions (Table 2). It is chosen to determine the SGI over a period of 3 months, which means that for any given day, the SGI is the average of the last 90 days. The SGI-3 reflects the characteristics of groundwater droughts (Guo et al., 2021), and makes it clearer to characterise droughts (Figure 52, Appendix C), due to the reduced short-term fluctuations.

As a result, the SGI offers valuable insights into the intensity and duration of droughts by examining how far SGI values drop below zero (Table 2) and the duration for which they remain beneath zero (Hsin-Fu Yeh & Chang, 2019). The frequency of drought events can be assessed by estimating the return periods of annual minimum SGI values.

Table 2: Index values representing groundwater drought conditions (Bloomfield & Marchant, 2013).

Condition	Criterion
No drought	SGI > 0
Mild drought	0 > SGI > -1
Moderate drought	-1 ≥ SGI > -1.5
Severe drought	-1.5 ≥ SGI > -2
Extreme drought	-2 ≥ SGI

3.2.5 Annual minima extreme value analysis (EVA)

The analysis of extreme values is based on annual minima groundwater levels. The annual minimum extreme value analysis only takes into account the lowest groundwater level in one year. The analysis of extreme values provides insight into the probability and intensity of extreme occurrences over time. It does not inherently provide information about the duration of drought events, as a single value for each year is chosen. The intensity consists of the minimum groundwater level. The frequency of drought events is determined using the return periods for the annual minima.

Furthermore, an extreme value distribution is fitted to the annual minima values to extrapolate the frequency curve to a return period of 100 years (as elaborated in 2.4.4).

3.2.6 Overview drought statistics

Table 3 gives an overview of the different drought statistics which provide insight into the extremity of groundwater drought events. This includes the characterisation of droughts according to the properties intensity, duration, and frequency. The consecutive duration of a certain intensity is noted as [m, m + n_m] which means that the intensity holds for period (months or day-of-year) m to m + n_m.

Table 3: Overview statistical methods that describe droughts using intensity, duration, and frequency.

Statistical Methods	**Intensity**	**Duration**	**Frequency**
Percentiles	P	P for [m, m + n_m]	T(P_m)
Average low groundwater Level (GLG)	Min(H) < GLG	H < GLG for [m, m + n_m]	T(min(H) < GLG)
Standardised groundwater index (SGI)	Min(SGI) < 0	SGI < 0 for [m, m + n_m]	T(min(SGI) < 0)
Annual minima extreme value analysis (EVA)	Min(H)	-	T(min(H))

3.3 RQ2 – VALIDATION OF HISTORICAL METEOROLOGICAL DATA

The objective of the second research question is to validate whether the historical meteorological time series represents the current climate. To gain insight into the drought statistics of the current climate, it is important to have long-term data that is representative of the current climate. The validation is required to ensure that it is valid to use the historical data in the modelling part of this study. The historical meteorological dataset from HKV (Pezij & Lugt, 2023) contains nationwide daily precipitation and reference crop evaporation for the period 1910-2022 projected to the current climate, in contrast to other meteorological datasets (Appendix E: Additional historical meteorological series). The dataset was developed using daily measurements of precipitation and reference crop evaporation of various KNMI weather stations. The precipitation and evaporation series were detrended per weather station to filter out long-term changes in the meteorological series, by applying a trend factor per season to each weather station. Afterwards, the data was spatially interpolated to cover a nationwide 1x1 km raster resolution, using different interpolation techniques for the precipitation and evaporation series. The interpolation of precipitation is based on spatial correlation between station observations, which are described by a variogram. There is also corrected for stations with no precipitation, using the assumption that the percentage of grid cells without daily precipitation should be equal to the percentage of stations with no daily precipitation. For evaporation data, observations are interpolated using a smooth plane. This is possible because evaporation varies gradually in space, unlike precipitation which varies widely in space.

The validation of the historical meteorological series from HKV (Pezij & Lugt, 2023) involves comparing the climatological properties of the historical precipitation and evaporation data with observed precipitation and evaporation. A period of 30 years (1991 to 2020) is assumed representative for the current climate because climate information encompasses long-term weather patterns, typically spanning 30 years (World Meteorological Organization, 2017). Furthermore, a comprehensive climate description includes not just averages but also variations from those averages (extremes) and the likelihood of these variations occurring (KNMI, 2021a), which can be described using probability density functions. Therefore, to address the second research question, the approach involves comparing probability density functions derived from historical precipitation and evaporation data to observed precipitation and evaporation at the locations of the KNMI stations.

Probability density curves are created for both precipitation and evaporation of both observed and historical data by sorting the data and calculating the probability for each value. The probability densities are accumulated to obtain a cumulative density function (CDF) for a clearer comparison. These curves can be visually compared for disparities in the probability density. Another visual inspection will be performed using quantile-quantile (Q-Q) plots, in which the quantiles of the CDF of historical meteorological data are compared to the quantiles of observed meteorological data (Das & Umamahesh, 2018; De Valk & Wijnant, 2019). If the quantiles are equal, thus having the same distribution function, the points in the Q-Q plot will fall along a straight line.

To perform a quantitative analysis, a gamma distribution is fitted to data, identified as the best fit for climate variables such as precipitation and evaporation (Dawley et al., 2019; Gupta et al., 2019; Martinez-Villalobos & David Neelin, 2019). The differences in gamma distributions are tested using the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov (K-S) test is a non-parametric hypothesis test for comparing the probability distributions of two samples. Its primary application lies in determining whether two samples are drawn from identical distributions. The null hypothesis is that the historical climate and observed climate have the same distribution. Since it is impossible to prove that the climate variables have the same distribution, the distributions are considered the same if the test fails to reject the null hypothesis. The K-S test reports the maximum difference between two cumulative distributions and calculates a p-value. The p-value [-] describes how likely it is that the data could have occurred under the null hypothesis. The smaller the p-value, the more likely it is that the null hypothesis is rejected. If the p-value is below 0.05 [-], the hypothesis is rejected (with 95% confidence), indicating samples are from different distributions. If the p-value is above 0.05 [-], the hypothesis cannot be rejected. Assuming the sample size is large enough, it is plausible to conclude that both climates have the same distribution.

If the historical meteorological time series fails to match the current climate, a bias correction is necessary. A widely recognized bias-correction method for climate variables is quantile mapping (Gupta et al., 2019; Navarro-Racines et al., 2015; Qian & Chang, 2021). Quantile mapping involves substituting the historical precipitation or evaporation with observed ones at the same cumulative density function (CDF) value within the utilised distribution for every CDF value (see Figure 8). Quantile mapping outperforms simpler methods such as shifting or scaling due to its effectiveness and minimal computational requirements (Gupta et al., 2019) and the ability to take into account the underlying distribution (Qian & Chang, 2021).

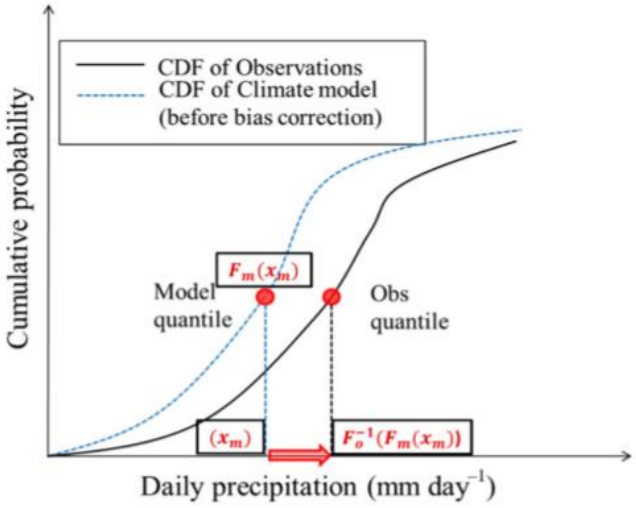


Figure 8: Quantile mapping bias correction (Gupta et al., 2019); Quantile mapping involves substituting the historical values with observed ones at the same cumulative density function (CDF) within the utilised distribution for every CDF value.

3.4 RQ3 – SIMULATION OF GROUNDWATER LEVELS

The objective of the third research question is to simulate long-term groundwater levels for the period 1910-2022. The long-term groundwater levels are simulated using the validated historical meteorological series (section 3.3) and a trained transfer function-noise model. Therefore, the first step is to fit the parameters of the model. This procedure includes a calibration and validation. The performances of the calibration and validation are evaluated using two goodness of fit metrics. After the long-term groundwater levels are simulated, these simulations are validated.

3.4.1 Calibration and validation of the model parameters

The transfer function-noise model is developed using eight years of observed groundwater levels, precipitation and reference crop evaporation. This time period is chosen such that the model represents the current response of the groundwater system (Zaadnoordijk et al., 2019), because of alterations in the water system are constantly taken place in the Netherlands. As an example, the relocation of monitoring wells may cause inhomogeneous results (Bouma et al., 2012; Ritzema et al., 2012). The calibration period is six years (2014-2020), and the validation period is two years (2012-2014). These periods are chosen since it results in enough data for calibration with some dry and moderate years. In the calibration period, the model parameters are fitted using the observed groundwater levels, precipitation and evaporation data. The validation shows how well the model simulates groundwater for a different period than the calibration period.

3.4.2 Model performance

The performances of the models are evaluated using two goodness of fit metrics. The explained variance percentage (EVP) and the root mean square error (RMSE), which are commonly used goodness of fit metric for groundwater level models (von Asmuth et al., 2012). These metrics indicate the performance of the model by comparing model simulations to actual observations. The difference between simulated and observed groundwater levels on a given day is the residual.

The EVP [%] describes the percentage of variation in the groundwater levels that is explained by precipitation and potential evapotranspiration:

$$EVP = \frac{{\sigma_h}^2 - {\sigma_r}^2}{{\sigma_h}^2} * 100\% \qquad (12)$$

Where σ_h^2 [m^2] is the variance of the groundwater observations and σ_r^2 [m^2] is the variance of the residuals. The returned value is bounded between 0% and 100%, higher percentages indicate more explained variance and therefore a better fit. The RMSE [m] measures the average difference between simulated and observed values and is calculated by the square root of the mean of the residuals:

$$RMSE = \sqrt{\frac{\sum r^2}{n}} \tag{13}$$

Where n [-] is the amount of residuals r [m]. A lower root mean square error implies smaller residuals and therefore a better fit. The goodness of fit criteria for the model performance is set to a minimum EVP of 70% and minimising the RMSE. Furthermore, the RMSE should be similar for the calibration and validation period. These criteria are often set to groundwater level models (Pezij et al., 2020; van Engelenburg et al., 2020; Zaadnoordijk et al., 2019). The performance of the model is determined for all groundwater levels in a year and specifically for summer groundwater levels, to emphasise the model performance for simulating low groundwater levels.

3.4.3 Validation of long-term groundwater levels

To validate how representative the long-term simulated groundwater levels are of current groundwater conditions, the simulated groundwater levels are compared with groundwater observations within the analysis (2012-2020). The goodness of fit metrics is used for this purpose. Furthermore, to give an indication of the quality of the simulated groundwater levels before the analysis period, the simulated groundwater levels are compared with observations before the analysis period (~2006-2012) and groundwater levels in the vicinity of the location (~1950 - 2012). These includes measurements within 2 km of the monitoring locations. As the groundwater levels are projections and not actual groundwater levels, goodness of fit metrics and a direct comparison cannot be used for this purpose. Therefore, a visual comparison using dot plots is conducted between these measurements and simulated groundwater levels, to indicate to what extent the simulated values are relatable to measurements, even though they are not supposed to be the same.

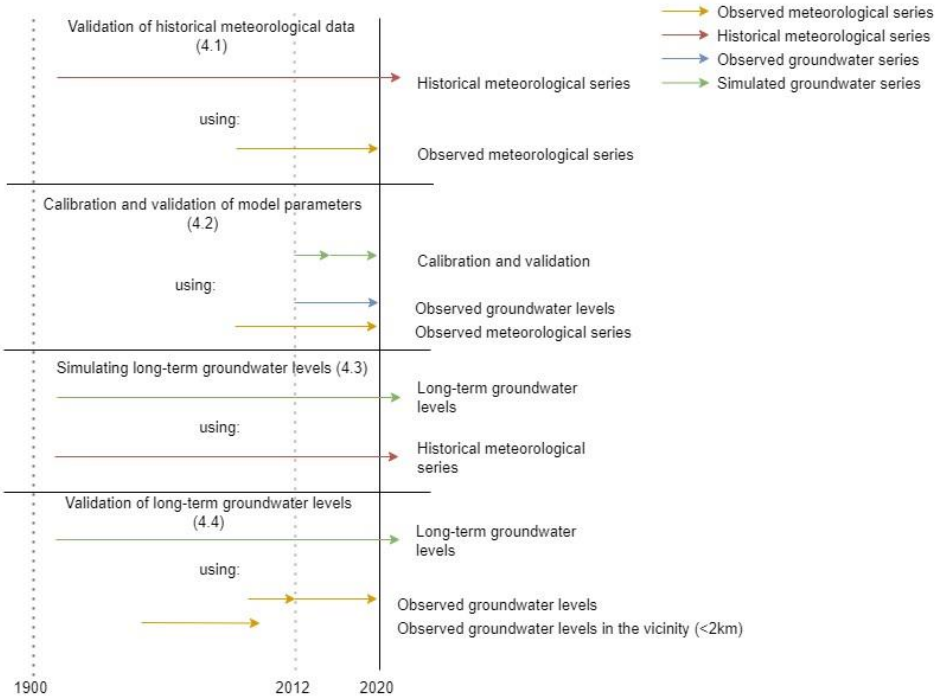


Figure 9: Different time periods for the simulation of long-term groundwater levels.

3.5 RQ4 – COMPARING DROUGHT STATISTICS OF OBSERVED AND SIMULATED GROUNDWATER LEVELS

The objective of the fourth research question is to derive drought statistics from groundwater observations and long-term simulations. Furthermore, the drought statistics of the observations and simulations are compared to assess whether the groundwater measurements are representative of long-term groundwater levels.

In the first part, drought statistics are derived using the approach defined in the first research question (Section 3.2). Therefore, for each location, each drought statistic is derived from groundwater observations (2012-2020) and long-term simulations (1910-2022). Frequency lines are plotted, showing the return period with associated drought intensity. Scatterplots are also produced showing the drought duration with associated drought intensity.

The second part involves the general comparison of observations and simulations. Therefore, the characterisation of the 2018 drought is highlighted for one location. In addition, the drought characteristics of 2018 are compared with those of 1976 and 1921, both record years in terms of precipitation deficit. The 1976 deficit was more intense, while the 1921 deficit was more prolonged. This makes it possible to determine whether the drought events are characterised differently, resulting in differences in the driest year.

3.6 STUDY AREA AND DATA

Long-term groundwater levels are simulated for locations in the management area of regional water authority Waterschap Limburg (dots in Figure 10). These locations are monitoring wells with relatively few external influences on groundwater flow (such as seepage to surface water or major groundwater withdrawals). Furthermore, Northern Limburg has sandy soils which have vulnerability to groundwater droughts (Brakkee et al., 2021; van den Eertwegh et al., 2021). Lastly, the locations have differences in land use, soil type and distance to surface water, providing a more generalised research. This is used to demonstrate whether a valid time series model can be set up for all these locations, since these are important characteristics that have impact on groundwater levels. The region has a temperate climate with a precipitation surplus of 100-350 mm/y (Brakkee et al., 2021), characterised as low-topography area above sea level (Figure 29, Appendix A).

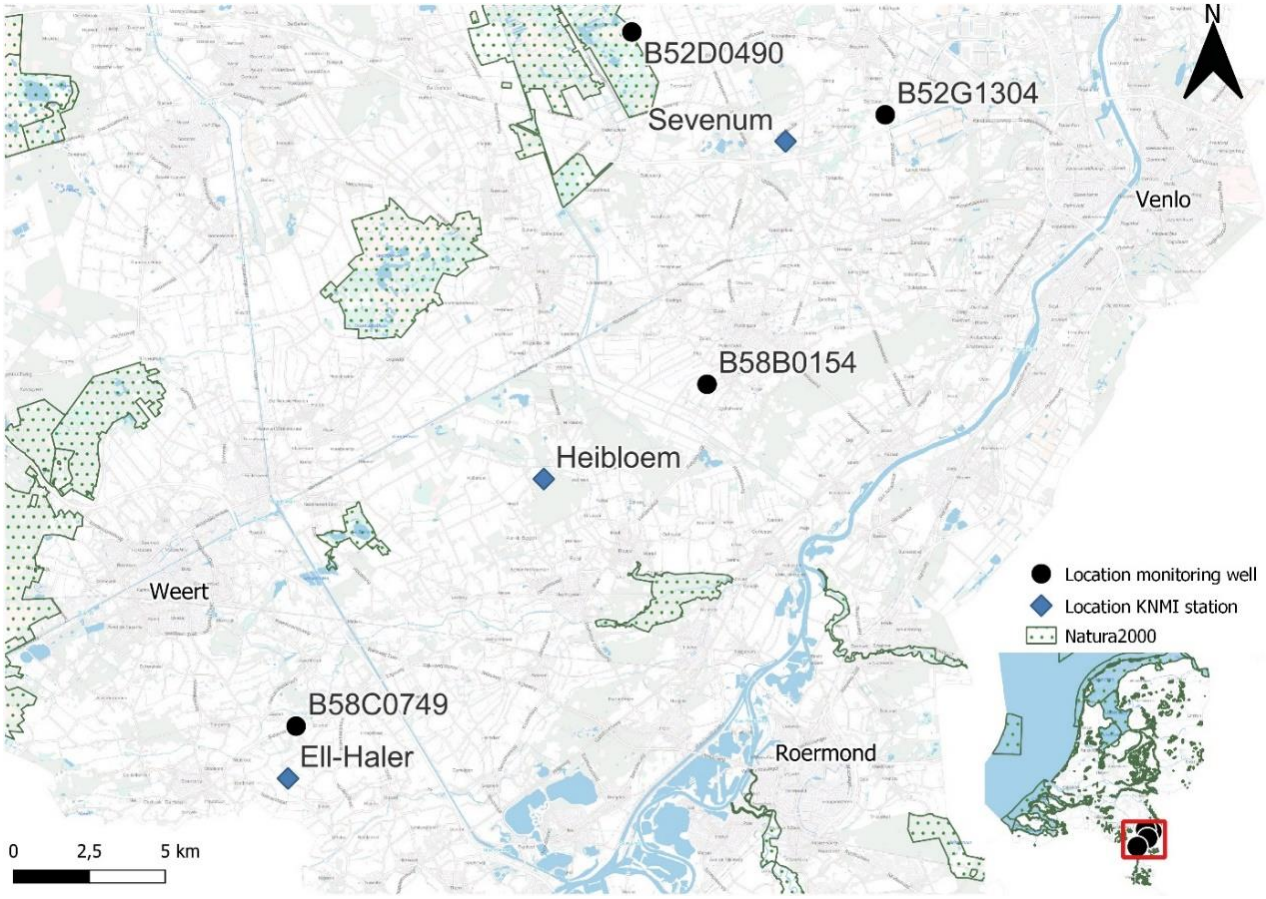


Figure 10: The four monitoring wells (dots) and the KNMI meteorological stations (diamonds) in northern Limburg (KNMI, 2023a, 2023b). The names given to the monitoring wells are Ell, Heibloem, Sevenum and Mariapeel (Table 4).

3.6.1 Study area

For each monitoring well, the land use, soil type and distance to surface water are analysed, and shown in Table 4. The given name of the locations are related to towns nearby the monitoring well. Land use is extracted from the Landelijk Grondgebruiksbestand Nederland database (LGN), which is a land cover dataset that relies on a combination of geospatial data sources and satellite data (Hazeu et al., 2023). Soil types are derived from the soil map of Basisregistratie ondergrond (BRO), which delineates the soil composition of the Netherlands down to a depth of 1.2 meters at a scale of 1:50.000 (show in Figure 31, Appendix A). The soil map SGM is used, which is a model that is based on the interpretation of field work data and is often used in hydrological studies (Programma Basisregistratie Ondergrond, 2023). Along with the type of soil, the designation according to the soil classification system in the Netherlands (H. de Bakker & Schelling, 1989) is given, which is founded on geological conditions.

The distance to nearby surface water is approximated using satellite imagery. In order to minimise the subsurface groundwater flow to surface water, the minimal distance should be 100 m from a river or canal, 25-50 m from streams and 10-25 m from a ditch (Bouma et al., 2012). Possible groundwater withdrawals by farmers (1Limburg, 2023) are shown in Figure 30, Appendix A. However, since location Ell, Sevenum and Mariapeel are located within buffer zones for desiccated nature reserves, groundwater can only be withdrawn by permit holders and is therefore minimised. Lastly, there are no groundwater protection areas (Waterschap Limburg, 2019) or drinking water withdrawals in the area (Figure 30, Appendix A). Detailed description of the locations are shown in Appendix A: Characteristics of locations.

Table 4: Characteristics of site groundwater monitoring wells.

Location name	**Code monitoring well**	**Soil** (Programma Basisregistratie Ondergrond, 2023)	**Land use** (Hazeu et al., 2023)	**Distance to surface water**
Ell	B58C0749	Loamy fine sand; Beekeerd	Grassland for nature	50 m (Vliet)
Heibloem	B58B0154	Loamy fine sand; Gooreerd	Agricultural	150 m (Egchelbeek)
Sevenum	B52G1304	Loamy fine sand; Veldpodzol	Agricultural and grassland for nature	110 m (Groote Molenbeek)
Mariapeel	B52D0490	Soft loamy fine sand; Veldpodzol	Nature reserve	30 m (Peelkanaal)

3.6.2 Data

The time-series analysis of groundwater data requires meteorological and groundwater measurements. The used data in this study are groundwater measurements, observed meteorological observations and a historical meteorological dataset, as summarised in Table 5. The locations of the measurements are shown in Figure 10. Details about the measurements are mentioned in the paragraphs below.

Groundwater measurements from the four monitoring wells (B58C0749, B58B0154, B52G1304, and B52D0490) are used, spanning from December 2, 2012 to December 2, 2020 (DINOloket, 2020). Measurements were taken every other week or daily. The properties of the observed groundwater measurements are shown in Table 13 of Appendix B.

Meteorological observations consist of daily precipitation sum and reference crop evaporation according to Makkink, which were obtained from nearby KNMI weather station Ell-Haler. Additional daily precipitation sum was obtained from KNMI precipitation stations (Heibloem, and Sevenum). Therefore, covering different time periods (1999-2023 for Ell-Haler and 1991-2020 for Heibloem and Sevenum). The reference crop evaporation data from station Ell was used for all four groundwater monitoring locations (Figure 49). Additional information about the data is given in Appendix B.

In addition to the meteorological observations obtained from KNMI, a historical meteorological dataset is used (as described in section 3.3). Referred to as historical meteorological data throughout the remainder of the report. The historical meteorological data is extracted at the location of the KNMI stations, such that it is consistent with the observed meteorological data (as described in Figure 49, Appendix B). The historical meteorological contains also daily precipitation sum and Makkink reference crop evaporation.

Table 5: The type, temporal resolution, and time period of the data.

Name	**Type**	**Temporal resolution**	**Time period**	**Reference**
Groundwater observations	Point	Every two weeks (Ell) or daily (others)	2012-2020	(DINOloket, 2020)
Daily precipitation sum	Point	Daily	1999-2023 (Ell) or 1991-2020 (others)	(KNMI, 2023a, 2023b)
Reference crop evaporation	Point	Daily	1999-2023	(KNMI, 2023b)
Historical meteorological data	Raster (1x1 km)	Daily	1910-2022	(Pezij & Lugt, 2023)

4 SIMULATING LONG-TERM GROUNDWATER SERIES

This chapter describes the simulation of long-term groundwater levels using historical meteorological data. This includes the results of the validation of the historical meteorological data using cumulative density functions (CDFs), quantile-quantile (QQ) plots, and the Kolmogorov-Smirnov (KS) test (section 4.1). The transfer function-noise model is calibrated using observed groundwater levels, precipitation and evaporation, for the period 2012-2020 (section 4.2). The performance of the calibration is assessed using the goodness of fit metrics RMSE and EVP. In section 4.3, long-term groundwater levels are simulated for the period 1910-2022 using the calibrated model and historical meteorological data. The validity of the simulation results is assessed using the goodness of fit metrics RMSE and EVP (section 4.4). Simulated groundwater levels are also visually compared to observations from groundwater monitoring stations nearby with longer time series.

4.1 VALIDATION OF HISTORICAL METEOROLOGICAL SERIES

The first step in the process of modelling long-term groundwater levels is to validate the historical meteorological time series. The goal of the model is to simulate long-term groundwater levels for the current situation. Therefore, validation is essential to ensure that the historical meteorological data represents the current climate, and therefore valid to use as input data for the simulation.

Initially, the validation shows that the historical precipitation at precipitation stations Heibloem and Sevenum do not represent the current climate (Appendix D: Validation of historical meteorological data before correction). The cumulative density functions for the historical meteorological data and current observation look very similar (Figure 59). However, the QQ-plots (Figure 60) and Kolmogorov-Smirnov test (Table 18) show that there is indeed a significant difference for precipitation at Heibloem and Sevenum. Specifically, high precipitation events are underestimated in the historical data compared to observed precipitation. However, for KNMI station Ell, the historical precipitation and reference crop evaporation are representative for the current climate. This is because this weather station was used in the development of the dataset, while the precipitation at Sevenum and Heibloem were spatially interpolated.

As a result, the precipitation for location Sevenum and Heibloem were bias corrected using observed precipitation following the quantile mapping approach. This means that the CDF of the long-term climate series is adjusted to match the CDF of the 30-year observed climate data, resulting in historical precipitation and evaporation that represents the current climate (Figure 11).

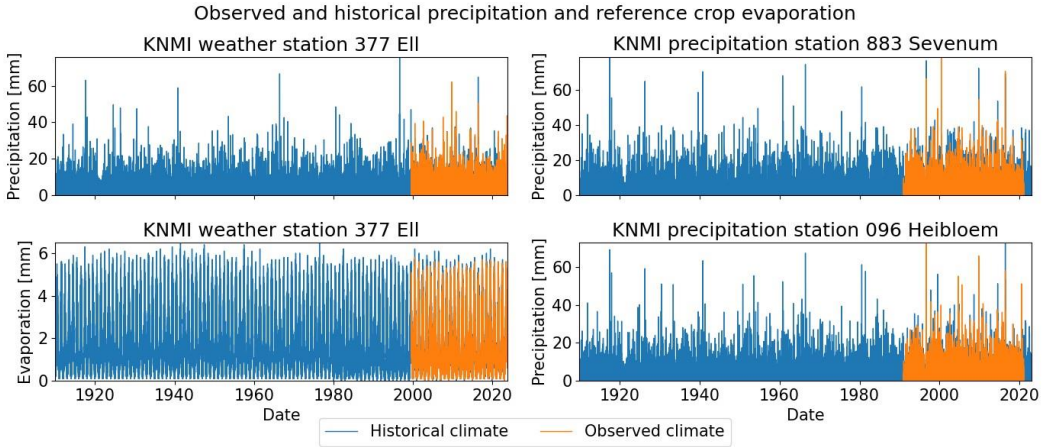


Figure 11: Observed and historical precipitation and reference crop evaporation (whereof Sevenum and Heibloem are corrected for current climate using the observations). The historical climate is validated to represent the current climate.

Figure 12 displays QQ-plots, comparing the quantiles the historical precipitation and evaporation with those of the current observed climate. The red line represents the theoretical scenario where the distributions of both climate series are identical. Based on a visual inspection of the QQ-plots, the cumulative density function of both meteorological series looks very similar, indicating that the climate of the historical series represents the current climate. The precipitation in the historical series is underestimated slightly in the most extreme quantiles. Conversely, the historical climate series tends to slightly overestimate evaporation in these extreme quantiles. However, this has minimal impact since the most extreme precipitation will not infiltrate into the ground if the infiltration capacity is reached, unless there is no surface runoff.

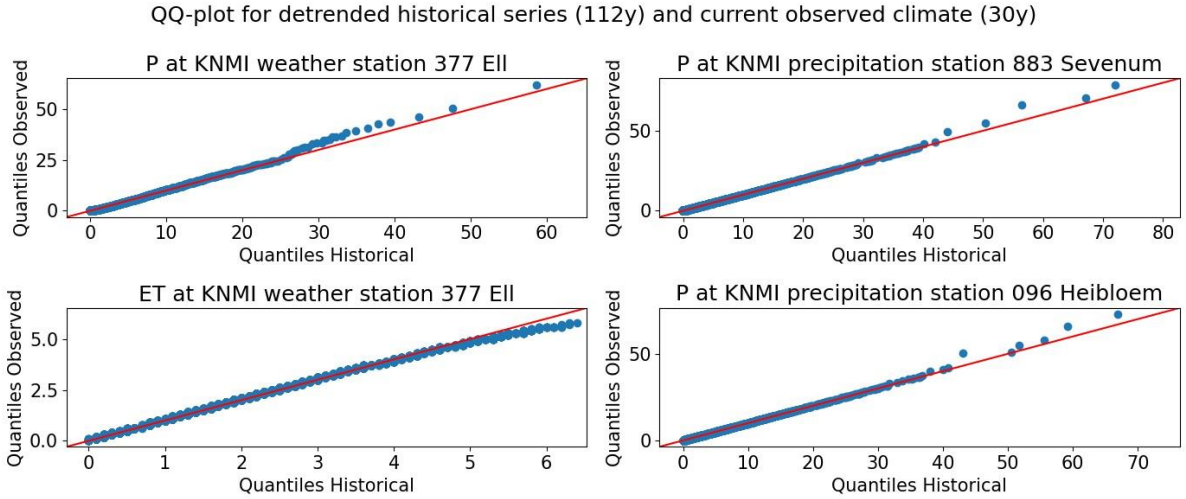


Figure 12: QQ-plots: quantiles of historical precipitation and refence crop evaporation are compared to quantiles of observed precipitation and refence crop evaporation. If the distributions are equal, the points will fall along a straight line.

Table 6 shows the results of the KS-test for the corrected historical meteorological data. In the table, this test proves that the corrected historical data represents the current climate since the p-values for the meteorological series are above the chosen criteria. Therefore, validating with 95% confidence that the corrected historical meteorological data is representative of the current climate. The K-S statistic is the maximum difference in cumulative density function value, which are also relatively small values.

Table 6: K-S statistics of corrected historical and observed climate.

Test	Ell		Heibloem	Sevenum	**Criterion**
	P	ET	P	P	
KS [mm]	0.05	0.03	0.02	0.02	
P-value [-]	0.13	0.99	0.97	0.95	**> 0.05**

4.2 CALIBRATION AND VALIDATION OF THE MODEL PARAMETERS

Figure 13 shows the time series of the calibration and validation, alongside the observed groundwater levels. Moreover, the calibration and validation goodness of fit metrics meet the criteria (Table 7) meaning that all models can accurately simulate groundwater levels. The model performances for validation of location Heibloem are lower, which is probably caused by the underestimation of groundwater levels in the year 2014.

In Appendix E (Table 19), the model performance is shown for only summer months, to give an indication of the performance of the model to simulate low groundwater years. This results in slightly lower goodness of fit values, but still comfortably above the defined criteria.

Table 7: Model performance using goodness of fit metrics EVP and RMSE.

Model Performance	**Ell**	**Heibloem**	**Sevenum**	**Mariapeel**	**Criteria**
Calibration (2014-2020)					
EVP [%]	94.9	88.6	90.3	92.6	> 70%
RMSE [m]	0.12	0.16	0.16	0.14	
Validation (2012-2014)					
EVP [%]	96.1	76.6	90.3	89.5	> 70%
RMSE [m]	0.09	0.30	0.16	0.11	

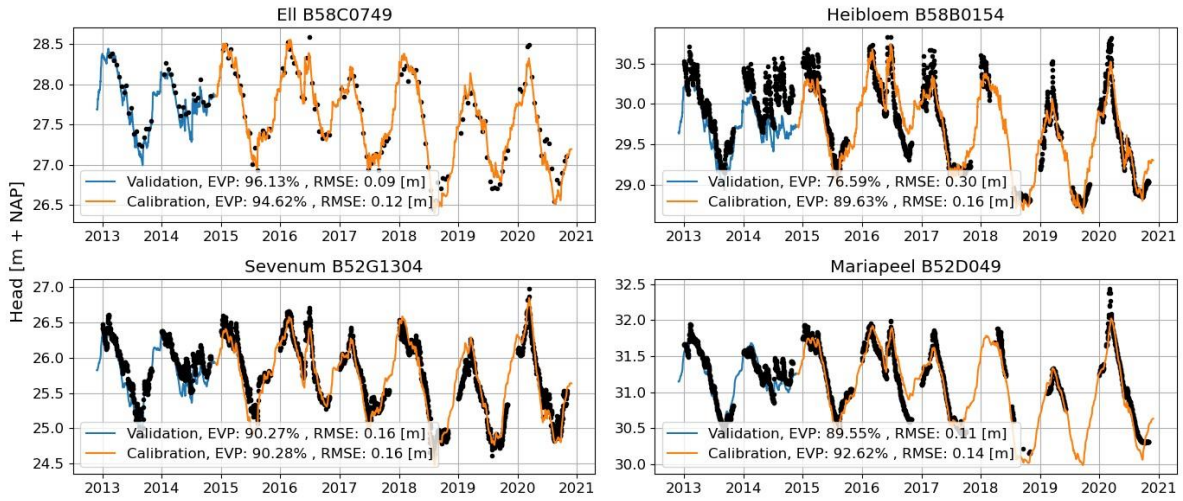


Figure 13: Calibration and validation of the time series model, together with goodness of fit metrics EVP and RMSE.

4.3 SIMULATING LONG-TERM GROUNDWATER LEVELS WITH HISTORICAL METEOROLOGICAL SERIES

The calibrated time series model can be combined with historical precipitation and evaporation data to simulate long-term groundwater levels. The obtained simulated long-term groundwater levels are shown in Figure 14. The figure shows the annual groundwater levels for the period 1910-2022, displayed throughout the year. The reference year 2018 and record year 1976 are highlighted. As can be seen, the groundwater levels in the year 1976 are very low, and lower compared to 2018.

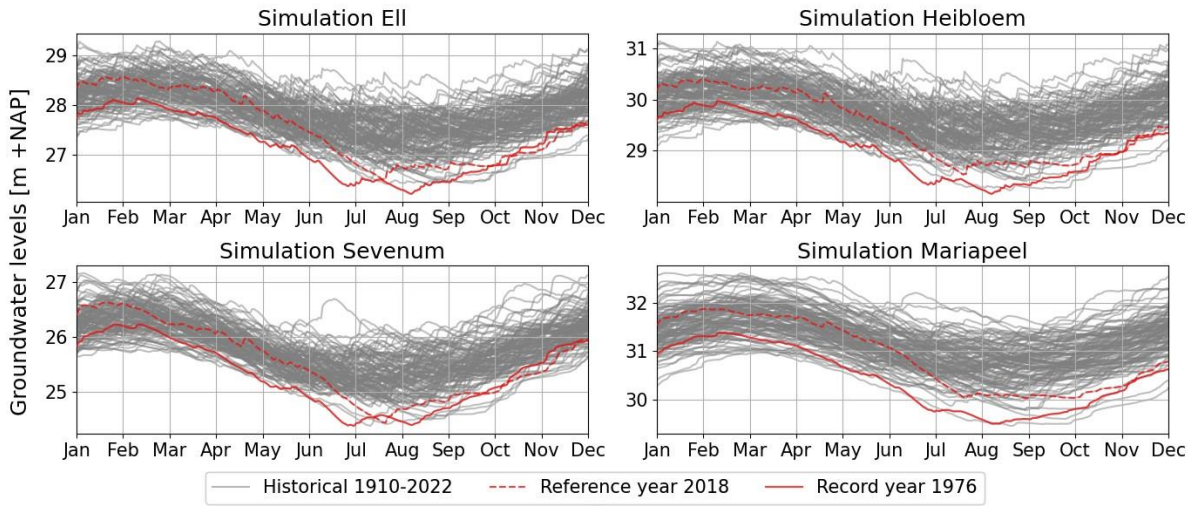


Figure 14: Long-term groundwater levels based on simulations with the calibrated model and historical meteorological data, displayed throughout the year. The dashed red line is the reference year 2018, which can be compared to observed groundwater levels. The solid red line is record-year 1976, which has the largest precipitation deficit.

4.4 VALIDATION OF LONG-TERM GROUNDWATER LEVELS

This paragraph describes the validity of the long-term groundwater simulations with historical meteorological time series. For the analysis period (2012-2020), the simulated groundwater levels are assessed using the goodness of fit metric RMSE and EVP (Figure 15). The simulations show accurate results, with EVP values around 90% (Table 8). Furthermore, simulating similar low groundwater levels compared to observations such as summer 2018.

Table 8: Validation of long-term groundwater simulations in analysis period.

	Ell	Heibloem	Sevenum	Mariapeel
EVP [%]	93.0	85.3	86.3	87.6
RMSE [m]	0.23	0.18	0.18	0.20

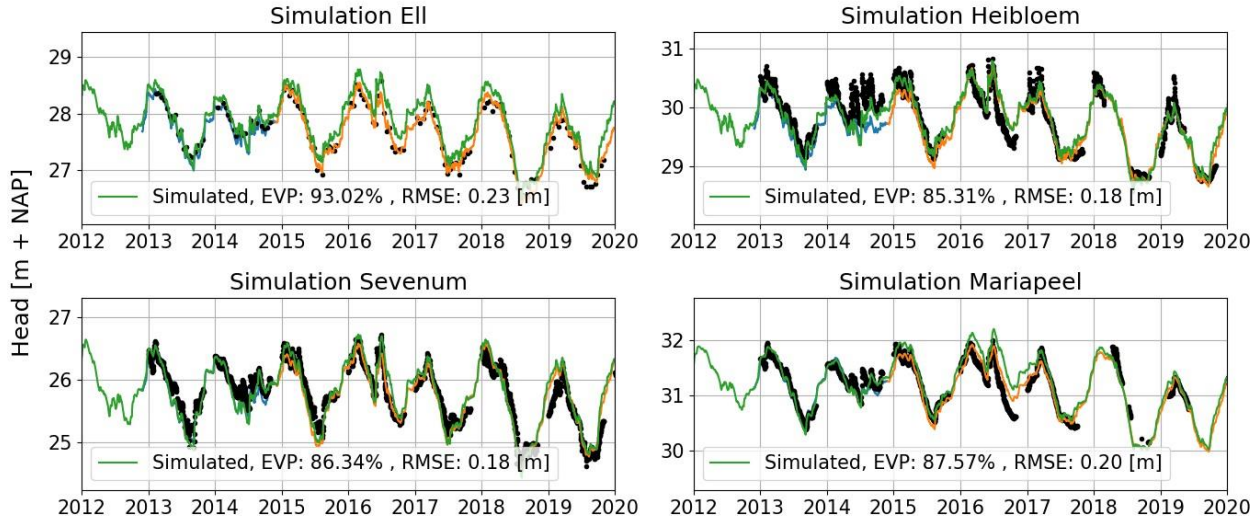


Figure 15: Validation of long-term simulated groundwater levels using measurements in the analysis period. Performance indicators are the goodness of fit metrics RMSE and EVP. The figure also includes the calibration and validation of the model (section 4.2).

To validate the long-term groundwater simulations outside the analysis period, simulated groundwater levels are visually compared to measurements before the analysis period and to observations of groundwater monitoring stations in the vicinity with a larger data availability (Figure 62, Appendix D). Since the long-term groundwater simulations are projections rather than actual events, the validity cannot be performed quantitively. However, the groundwater simulations can be visually compared to measurement before the analysis period (Figure 16) and in the vicinity within 2 km of the monitoring location (Figure 17).

These figures show that the observed groundwater measurements are mostly different to the long-term simulations in earlier years. However, this is expected since the climate and groundwater system has changed. What can be seen is that the comparison follows the same trend, indicating that dry years in the measurements are also occurring the long-term groundwater simulations.

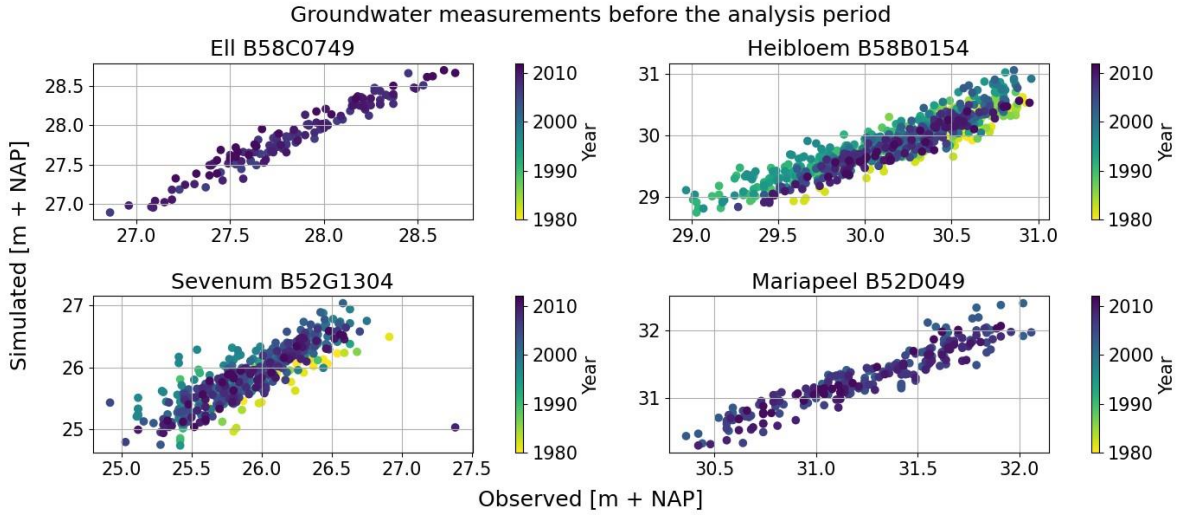


Figure 16: Validation of long-term simulated groundwater levels using measurements before the analysis period.

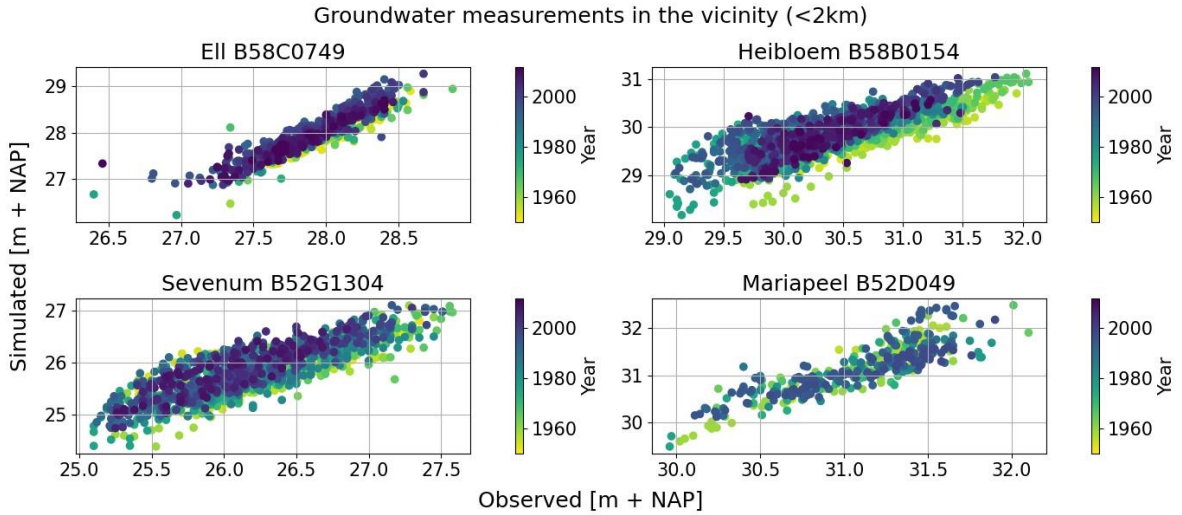


Figure 17: Validation of long-term simulated groundwater levels using measurements in the vicinity with larger data availability.

5 COMPARING DROUGHT STATISTICS OF OBSERVED AND SIMULATED GROUNDWATER LEVELS

Various drought statistics are derived from the groundwater measurements and long-term simulations. In the following paragraphs, the methodology of section 3.2 is used to derive drought statistics of groundwater observations and simulations. In Appendix E, the groundwater measurements from the monitoring wells (over the period of 2012-2020) are presented (Figure 65). In section 4.3, the time series of the groundwater simulations (over the period of 1910-2022) were presented.

5.1 PERCENTILES

Figure 18 shows the regime curve with percentile bands for the four locations. The left panels present the percentiles based on observations, while the right panels present percentiles based on simulations. The regime curve provides a visual representation of the annual groundwater dynamics, with coloured bands serving to indicate the extent of groundwater fluctuations throughout the year.

The green percentile band, representing moderate conditions, encompasses 50% of the data points. The light blue and yellow percentile bands, indicative of wet and dry conditions, respectively, each capture 15% of the data points. Lastly, the orange and dark blue percentile bands, reflecting very dry and very wet conditions, each account for 10% of the data points.

A monthly resampling yields a nuanced perspective. Moderate conditions exhibit similarity between observations and simulations. For example, the average difference in median for location Ell is 0.2m (Table 15, Appendix C). Furthermore, the green percentile bands for moderate conditions are visually similar. The average observed groundwater levels are relatively well representative for average long-term conditions. A notable difference is evident in the wider extreme percentile bands, as the simulations involve a more prolonged analysis period. Figure 18 shows that the extreme dry and wet conditions based on observations are not representative when compared to the long-term simulated conditions, since the percentile bands have different ranges.

Consequently, percentiles based on observations indicate that the conditions in reference year 2018 transitions from wet to very dry. Long-term groundwater levels indicate that the beginning of 2018 was moderate, later turning extremely dry. Furthermore, in the observations, 2018 stands out as the most extreme year, while the long-term simulations suggest otherwise.

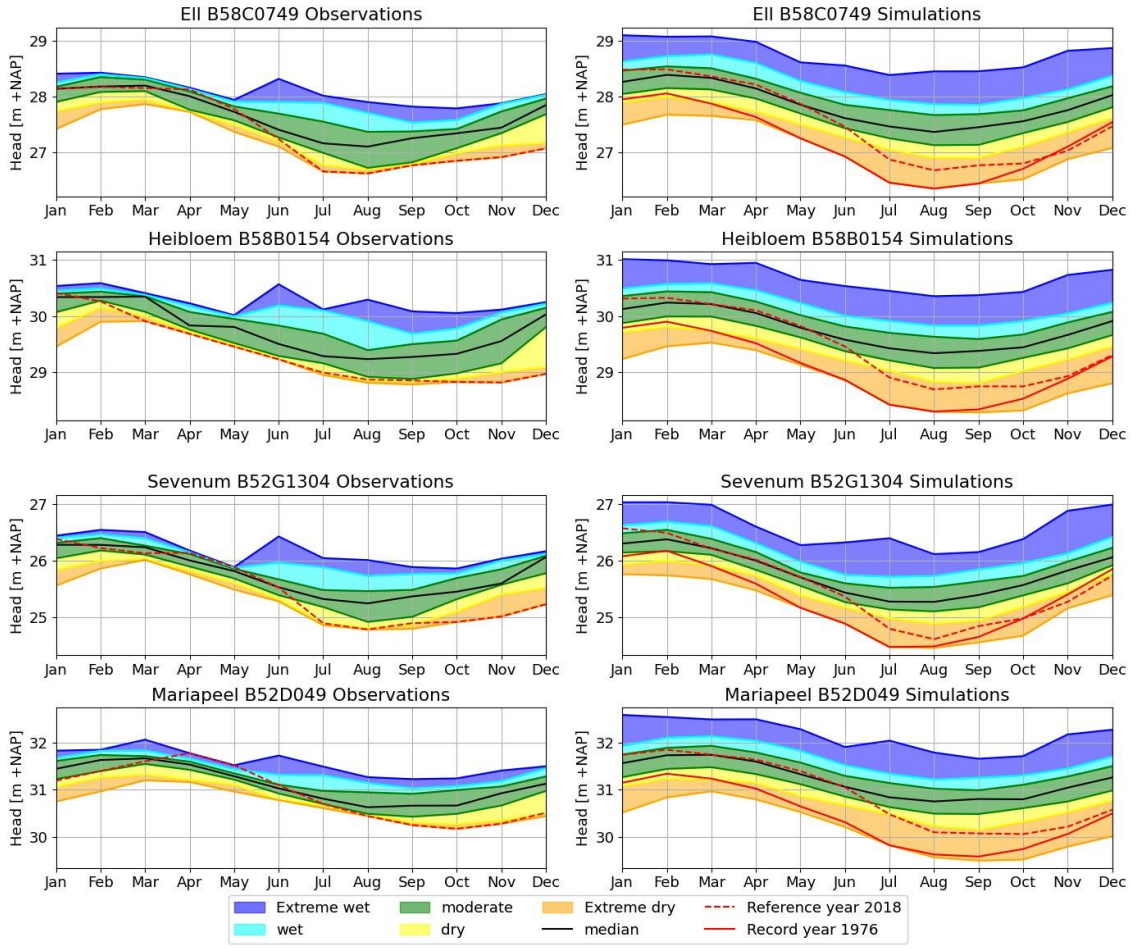


Figure 18: Regime curve with percentile bands; Extreme wet conditions defined with 90th percentile, wet condition is the 75th - 90th percentile band; 75th to 25th percentile bands indicate moderate conditions. 10th percentile demarcates dry and extreme dry conditions; resampled to monthly values.

5.2 AVERAGE LOW GROUNDWATER LEVEL (GLG)

The average high groundwater level (GHG), average spring groundwater level (GVG), average groundwater level (GG) and the average low groundwater level (GLG) are determined for the observations and simulations. Table 9 shows that most of the average groundwater statistics based on observations are higher than compared to simulations. The groundwater levels in the observation years are below-average compared to the long-term simulations. The average low groundwater level in location Ell is the lowest, indicating that the groundwater levels in the summer are highest compared to the other locations.

Table 9: Average high (GHG), average spring (GVG), average (GG) and average low (GLG) groundwater levels based on groundwater observations and model simulations for all monitoring locations.

Statistic [m below surface level]	Ell		Heibloem		Sevenum		Mariapeel	
	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated
GHG	0.87	0.64	0.92	0.98	1.28	1.15	1.09	0.98
GVG	0.98	0.83	1.12	1.16	1.46	1.44	1.22	1.14
GG	1.5	1.25	1.62	1.55	1.88	1.8	1.66	1.61
GLG	2.06	1.82	2.14	2.07	2.46	2.43	2.17	2.17

Figure 11 shows the GLG in relation to the groundwater observations and model simulations. The figures show multiple throughs below the GLG line. Observations show that the groundwater levels in 2018 drop lowest below this line while simulations show that there are multiple years with even lower groundwater levels, indicating more intense dryer periods.

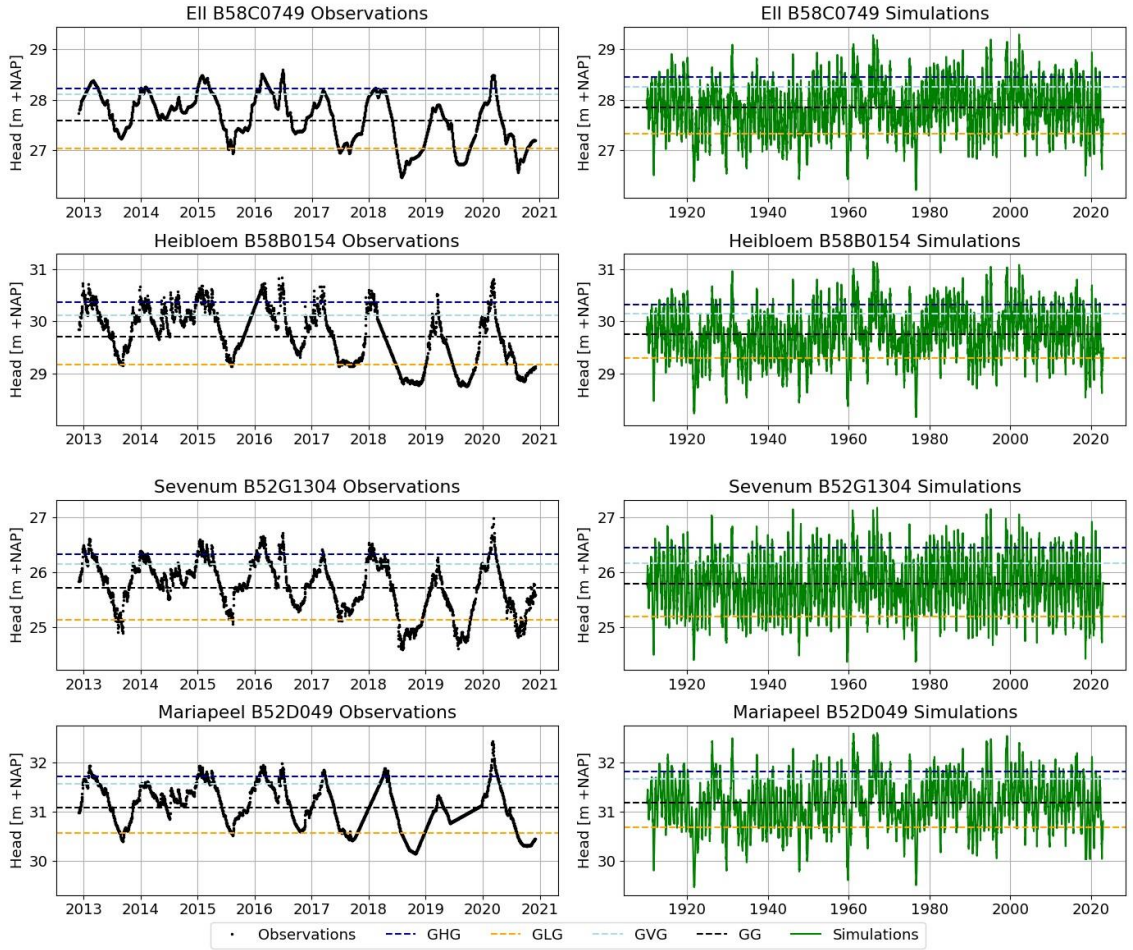


Figure 19: Average high (GHG), average spring (GVG), average (GG) and average low (GLG) groundwater levels based on groundwater observations and model simulations for all monitoring locations.

The selection of drought events, peaks with groundwater levels below the GLG, are shown in Appendix B (Figure 50). Using the plotting positions, the return periods of the drought events can be determined (Figure 20). As seen in the figure, the return periods for observations years do not exceed 1 in 10 years. The yellow stars represent the return period for reference year 2018, shown for both observations and simulations, while the red stars represent the return period for record year 1976.

Comparing the statistics, using only measurements tends to underestimate the severity of the drought (lower return period), except for Mariapeel. Furthermore, difference exist between locations, for example for a return period of ten years (T10). For location Ell and Sevenum, the groundwater level for T10 is lower using long-term simulations, indicating less head below GLG. This means that the years in the observation period indicate a drier T10 groundwater level compared to long-term simulation. For location Heibloem and Mariapeel is this the other way around.

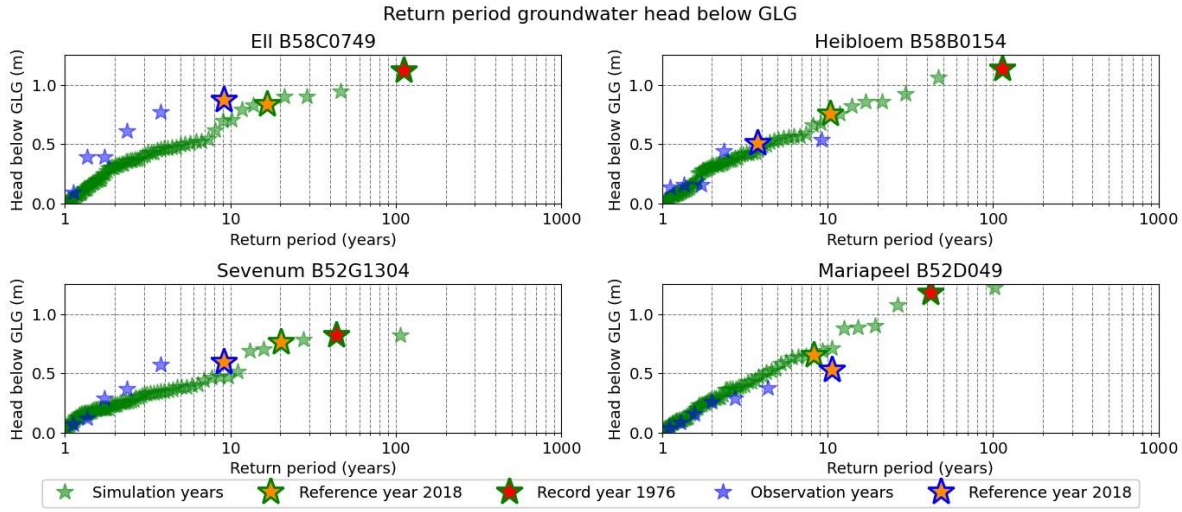


Figure 20: Return periods for groundwater levels below GLG.

Figure 21 shows the intensity (head below GLG) and the duration of drought events for simulations (green outline) and observation (blue outline). The figure shows that the drought of 1976 is more intense but is shorter compared to the drought of 1921, except for Mariapeel. Furthermore, the most prolonged drought is nine months below the GLG, which is quite more extreme compared to the year 2018.

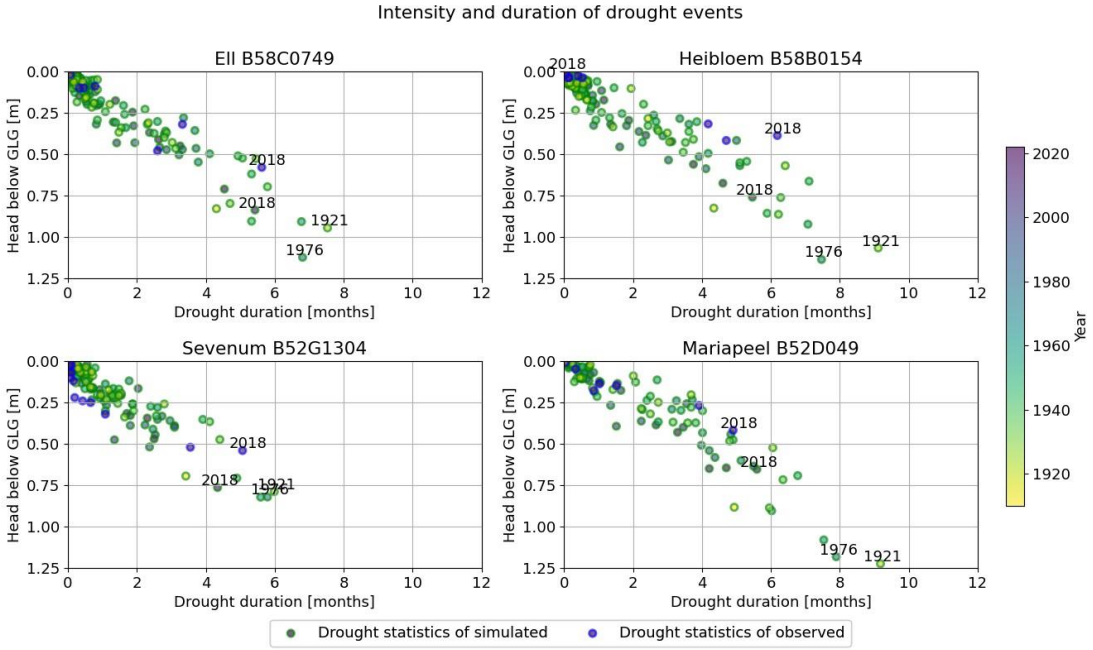


Figure 21: Drought as duration of groundwater levels below GLG.

5.3 STANDARDISED GROUNDWATER INDEX (SGI)

Figure 22 illustrates the Standardised Groundwater Index (SGI) for all monitoring stations, accumulated over a rolling window of three months (SGI-3). Positive SGI values indicate wet conditions, while negative values indicate dry periods. The left-side panels are based on groundwater observations, whereas the right-side panels are derived from groundwater simulations. Compared to time series of groundwater levels, the SGI does not contain a strong seasonal component. This means that the indication of droughts is not constrained by the rise in groundwater levels in the winter, which also increases drought duration. The SGI exhibits a consistent pattern across all monitoring stations, for observations and simulations. Furthermore, the pattern seen in the observation period is also visible in the simulation period for those years.

In the observation period, 2019 stands out as the most severe drought. However, the long-term simulations reveal that the 2018 drought appears less extreme. There have been several years in the past with even lower SGI values, indicating periods of more intense droughts (Figure 22).

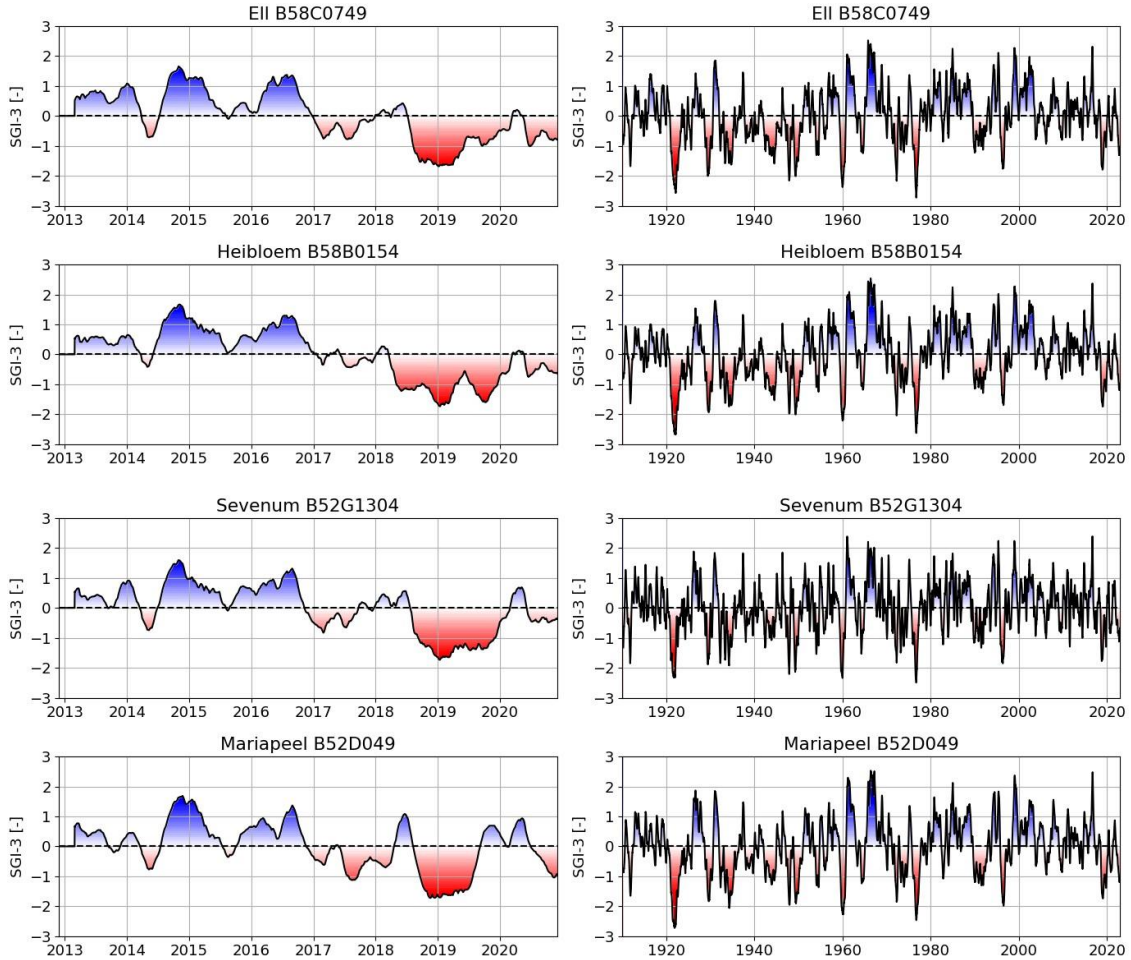


Figure 22: Standardised Groundwater Index for all monitoring stations; SGI values above zero (blue) indicate wet conditions, SGI values below zero (red) indicates dry conditions; figures on the left are derived from groundwater observations, figures on the right are derived from groundwater simulations.

Figure 23 presents a comparison return periods for annual SGI values below zero derived from observed and simulated groundwater levels. As evident in the figure, return periods for observations do not exceed 12 years, emphasizing the limited time series for drought analysis.

A comparison of the SGI for the 2018 drought indicates that using only measurements tends to overestimate the severity of the drought (lower return time for observations, implying more frequent occurrences), except for Mariapeel. This highlights the limitations of relying solely on measurements for accurate drought characterisation. Interestingly, the SGI values for the 2018 drought are consistent between short-term (observed data) and long-term (simulation data) statistics. This suggests that the intensity of simulated SGI peaks closely matches that of the observed drought event.

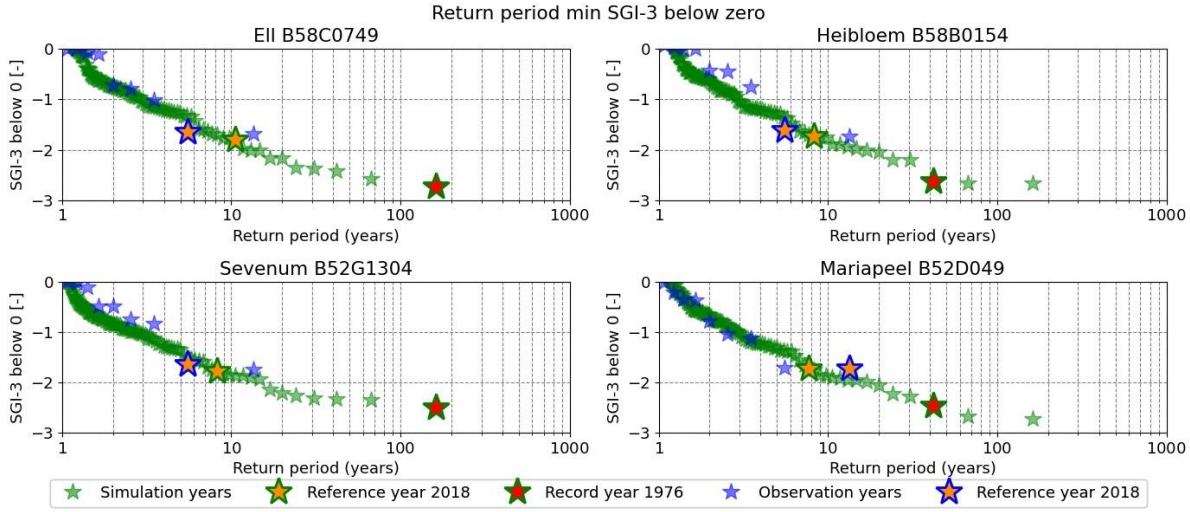


Figure 23: Return periods of minimum SGI-3 of drought events.

Figure 24 shows the intensity (SGI-3) and the duration of drought events for simulations (green outline) and observation (blue outline). The figure shows that the drought of 1921 is not always the most prolonged. However, those other droughts with long durations are a lot less intense. Furthermore, the most prolonged drought is around fifty months of SGI below zero, which is quite more extreme compared to the year 2018.

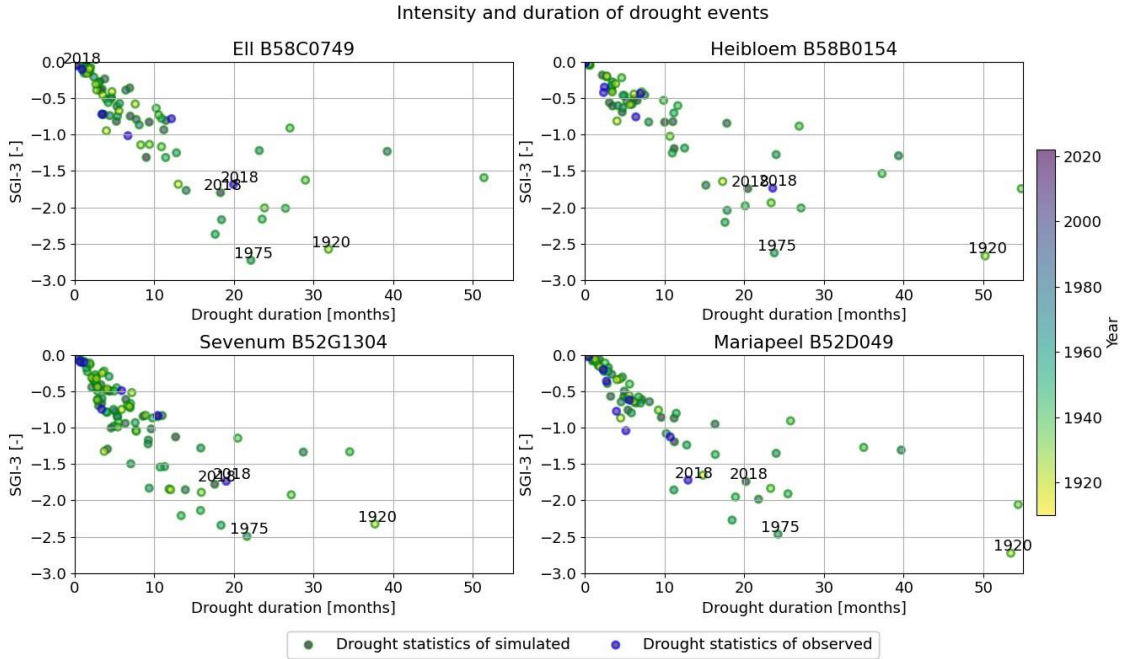


Figure 24: Drought as duration of minimum SGI-3.

5.4 ANNUAL MINIMA EXTREME VALUE ANALYSIS (EVA)

Last, groundwater droughts are classified by selecting annual minimum groundwater levels (Figure 25). The figure shows the annual minimum groundwater levels for the observation period on the left panels, indicating 2018 as the most extreme drought. The right panels show the annual minimum groundwater levels for the simulated period, indicating 1976 as the most intense drought. What also can be seen in the figure is that the seasonal fluctuations are clearly present. Furthermore, that the low groundwater levels in the summer sustain for multiple years.

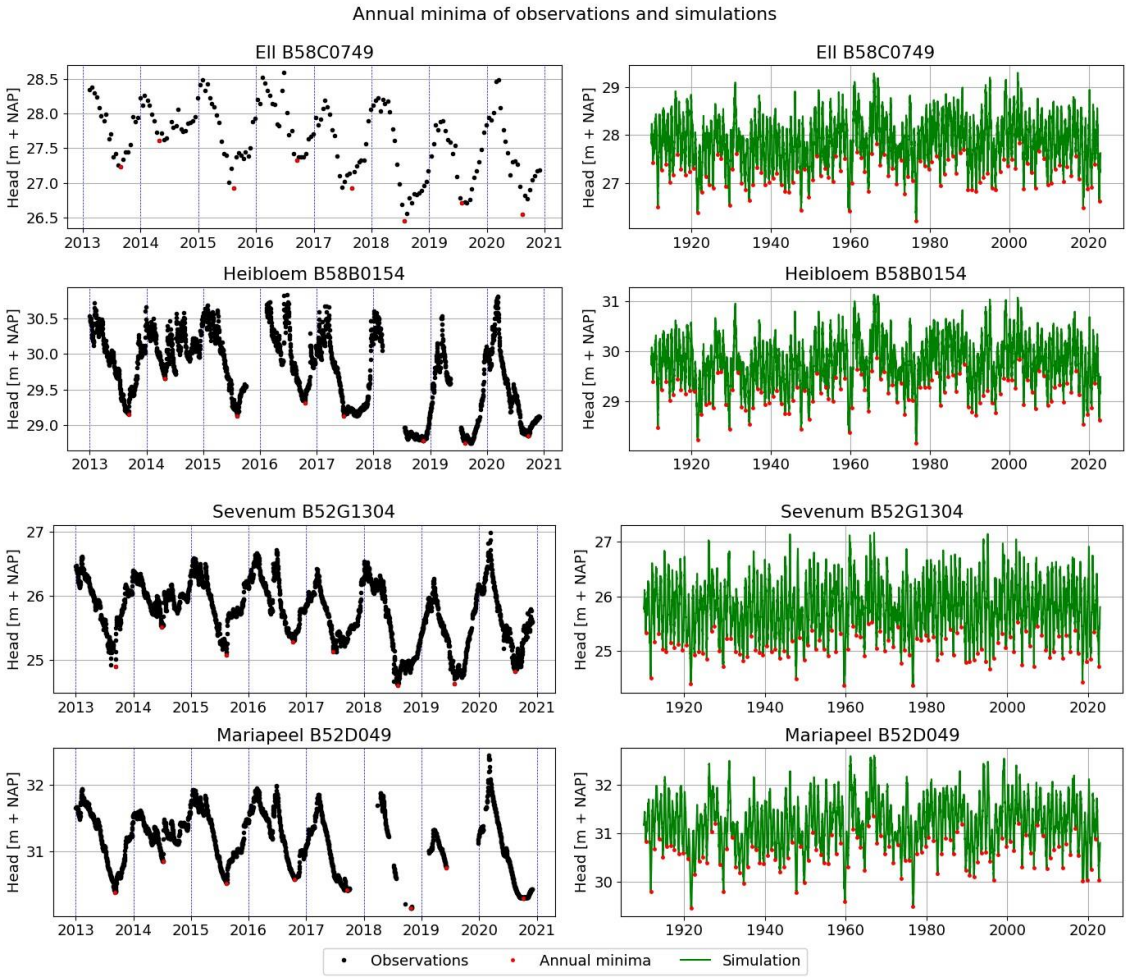


Figure 25: Annual minima for observations and simulations.

Figure 26 shows the return periods of the annual minimum groundwater heads along with the fitted extreme value distributions for observed and simulated groundwater levels. The most extreme return period for the minimum groundwater head in the observation series is 12 years. Additionally, the extreme value distributions allow for extrapolation of annual minimum groundwater levels for return periods up to 100 years. However, this is not accurate for the observations since a minimum of 30 years is needed (McCluskey et al., 2021).

When comparing the return period of 2018 between simulations and observations, the return period based on observations is a factor 2 smaller for all locations (except Mariapeel). This means that the observations give an incorrect representation of long-term frequency of droughts. Furthermore, the frequency lines of Ell and Sevenum are lower than the frequency lines of the simulations. This results in observations estimating lower groundwater levels for the same return period.

Additionally, basing criteria on return periods of observations could lead to delayed measures, resulting in the actual return period being surpassed by the time action is taken. For Heibloem and Mariapeel, this is the other way around. Most importantly, the return periods of the simulations show a clear break in trend around the T15 situation. Prior to this break, the groundwater level exhibits a gradual decrease as the return time increases, and then experiences a sudden, rapid decline until it eventually levels off.

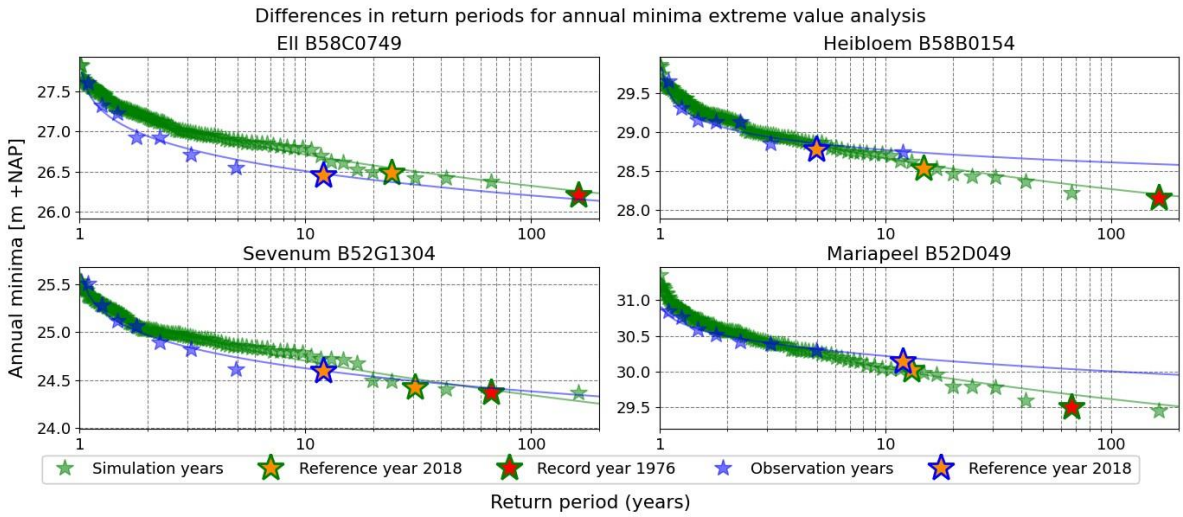


Figure 26: Return periods for annual minima with best fitting extreme value distribution.

5.5 DIFFERENCE BETWEEN DROUGHT STATISTICS OF SIMULATIONS AND OBSERVATIONS

Table 10 shows the differences between the drought statistics for the characterisation of the 2018 drought at the Ell site. Intensity and duration are similar for all drought statistics between observations and simulations. The frequency shows a significant difference. This difference applies to all percentiles, GLG and EVA. The frequency based on simulations gives a return period that is twice that of the observations.

Table 10: Difference in drought statistics for the characterisation of reference year 2018 at location Ell.

2018	**Intensity**		**Duration**		**Frequency (return period)**	
	Observations	Simulations	Observations	Simulations	Observations	Simulations
Percentiles	Minimum Extreme dry	5th percentile Extreme dry	6 months	6 months	T8	T21
GLG	0.87 m below GLG	0.84 m below GLG	5.6 months	5.4 months	T9	T17
SGI	-1.68 Severe drought	-1.79 Severe drought	20 months	18 months	T13	T11
EVA	26.5 m + NAP	26.5 m + NAP	-	-	T12	T24

Table 11 compares the drought statistics and their characterisation of the year 1976 and 1921 which can be compared to the simulation for 2018 in Table 10. The drought statistics show that although 1976 was more intense, the 1921 drought lasted longer, this emphasises that not only return periods are important.

Table 11: Drought statistics for simulated record year 1976 and long drought year 1921 at location Ell.

1976	Intensity	Duration	Frequency	1921 Intensity	Duration	Frequency
Percentiles	Minimum Extreme dry	10 months	T110	Minimum Extreme dry	11 months	T110
GLG	1.12 m below GLG	6.8 months	T112	0.95 m below GLG	7.5 months	T46
SGI	-2.72 Extreme drought	22 months	T162	-2.58 Extreme drought	32 months	T67
EVA	26.2 m + NAP	-	T161	26.4 m + NAP	-	T67

Table 12 displays the most extreme dry years based on return periods for all drought statistics and locations. This results in differences in the most extreme dry year per location and per drought statistic. However, percentiles and GLG indicate the same years as most extreme.

Table 12: Ranking of driest year for different statistical methods based on drought frequency.

	Ell		**Heibloem**		**Sevenum**		**Mariapeel**	
	Obs	Sim	Obs	Sim	Obs	Sim	Obs	Sim
Percentiles	2018	1976	2019	1976	2018	1959	2018	1921
GLG	2018	1976	2019	1976	2018	1959	2018	1921
SGI	2019	1976	2019	1921	2018	1921	2019	1976
EVA	2018	1976	2019	1976	2018	1921	2018	1959

6 DISCUSSION

This chapter consists of a discussion about the results of this study. The discussion is divided into the validity of the simulation of long-term groundwater series, interpretation of results and limitations in methodology.

6.1 VALIDITY OF SIMULATIONS

Several validations have been carried out in this study to ensure that the long-term simulated groundwater levels represent the current climate and groundwater system:

- The Kolmogorov-Smirnov test shows that the historical meteorological data are representative of the current climate, and therefore valid as input to the time series model;
- The goodness of fit metrics Explained Variance Percentage and Root Mean Square Error of the calibration of the model parameters meet the set criteria for a valid time series model;
- The calibrated model parameters show similar performance in goodness of fit metrics when validated for another time period;
- The long-term simulated groundwater levels display a similar goodness of fit performance compared to groundwater observations within the analysis period (2012-2020);
- The long-term simulated groundwater levels are comparable with historical measurements before the analysis period (1950-2012) at the same location and in the vicinity (<2 km).

Thus, the time series model effectively simulates long-term groundwater levels for the selected locations, suggesting that the fluctuations in groundwater levels can be explained by precipitation and evaporation.

The calibrated model parameters are validated over a limited two-year period, providing a good indication of the model performance. However, it is important to note that not a lot of extreme groundwater levels are included in this timeframe, which may lead to incorrect validation of extremely low groundwater levels. Furthermore, the accuracy of the goodness of fit metrics EVP and RMSE are compromised for location Ell due to the limited number of measurements. This makes it easier to achieve a higher performance indication, even though it does not necessarily imply a better fit (Collenteur, 2021). Additionally, the RMSE values for the validation and calibration should be approximately equal. This holds true for all locations except Heibloem, where the RMSE for validation is double that of the calibration (Table 7).

Validating model simulations prior to the analysis period (Figure 16, Figure 17) is challenging because the simulated groundwater levels are projections rather than observed events. These simulations represent historical groundwater levels under the current climate and water system. In particular, significant changes in the water system around 1950 make a direct comparison of groundwater levels impossible. Lastly, some groundwater levels are briefly simulated above surface level. This is probably because the model is calibrated in a relatively dry period (2014-2020) such that periods with high precipitation are modelled less accurately.

6.2 INTERPRETATION OF RESULTS

The characterisation of groundwater droughts using drought statistics from groundwater observations and long-term simulations shows significant differences (Figure 27). The methodology described in this study allows the direct translation of groundwater levels into return periods and vice versa. This means that it is possible to accurately quantify the severity of past droughts under

current conditions and makes a significant contribution to the understanding of groundwater droughts. This means that it is possible to get better estimates of how often droughts of different severity might occur by using return periods for drought analysis (Figure 27a). The drought of 2018 occurs less frequently than expected based on observations. Furthermore, the corresponding groundwater level that is expected for a given return period (Figure 27b). For example, if a drought occurring on average once in 10 years (T10) would justify a ban on irrigation, the action would be taken too late or not at all if it were based on observations.

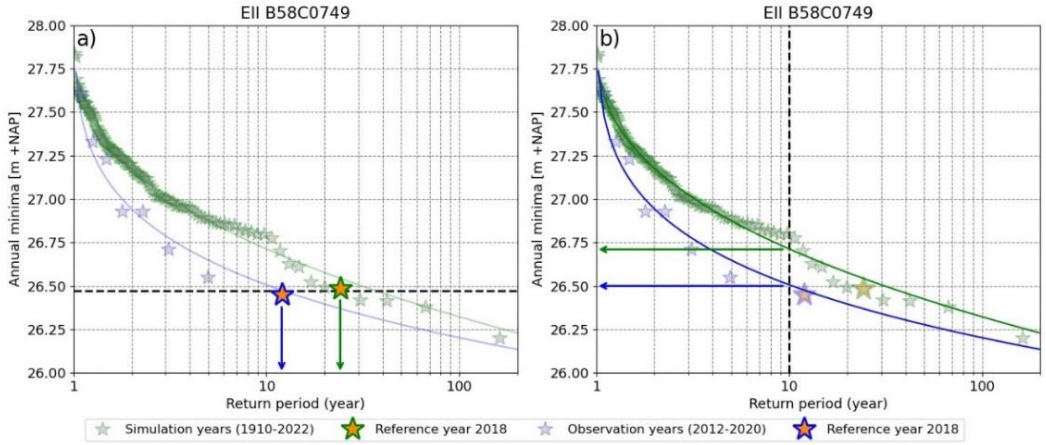


Figure 27: Implications to groundwater analysis

Furthermore, there are other things that stand out about the results. At location Sevenum, the results show notable discrepancies in comparison to other locations (Figure 28). The frequency lines exhibit a quicker flattening and reaching the minimum groundwater level sooner, around the T20 situation. This difference is particularly notable in terms of the minimum groundwater level in the driest year, which is approximately 0.3 meters higher at Sevenum.

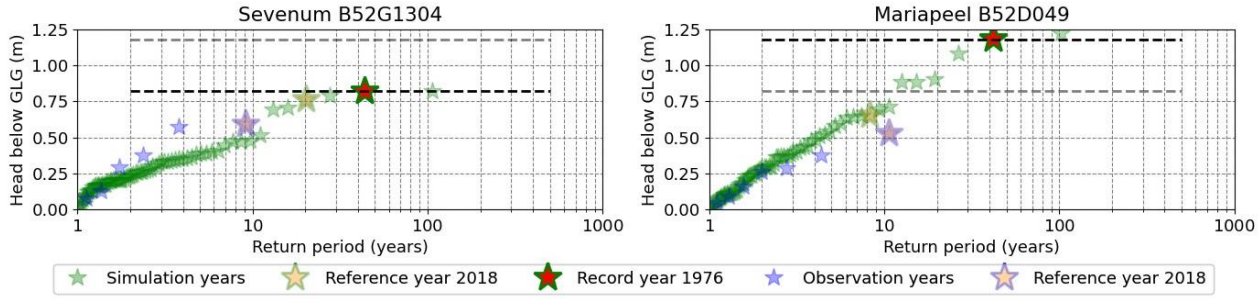


Figure 28: Droughts at location Sevenum are less intense and frequency line flattens out.

There are several possible reasons for this discrepancy, such as incorrect model inputs. However, the use of consistent evaporation data across locations and similar precipitation patterns suggests otherwise (Figure 12). Another consideration is the model calibration. It can be seen that the calibration for Sevenum leads to overfitting, given the relatively higher density of measurements in the summer of 2018 compared to the other locations (Figure 61, Appendix E). This results in those groundwater levels being overrepresented in the calibration. Overfitting means that the model fits the training data exceptionally well but struggles to simulate new unseen data. Therefore, overfitting can lead to poor performances and lacking robustness. Another explanation of the differences in location Sevenum lies in differences in soil physics. The structure of the subsurface in this study area is largely determined by fractures (Figure 32, Appendix A). These fractures originate from shallow impermeable layers in the subsurface, making the subsoil very complex (Benninga et al., 2018). Although these fractures do not occur exclusively at location Sevenum, they may occur at higher layers and therefore affect shallow subsurface flows.

The results also reveal a trend break in the frequency lines of the long-term simulated groundwater levels around the T20 return time (Figure 20, Figure 23 and Figure 26). Prior to this break, the groundwater level exhibits a gradual decrease as the return time increases, and then experiences a sudden, rapid decline until it eventually levels off. This pattern is seen at all sites, although some locations exhibit it more prominently. Possible explanations for this phenomenon include the impact of groundwater withdrawals or subsurface flows, particularly relevant at very low groundwater levels. Alternatively, it may be attributed to incorrect estimation of return times. The estimation of return times relies on the assumption that all included events are independent of each other. Due to the long residence time of the groundwater system, this assumption is not met when taking annual groundwater levels, impacting the statistical analysis. Therefore, the estimation of return period is biased by the selection of included drought events (Figure 55, Appendix C). Furthermore, higher return periods are inherently less accurate, since these have only occurred once in the simulations. Therefore, with 112 years of simulations, accurate return periods can be estimated up to approximately a return period of T25, as these events would have occurred in the simulations more than four times on average.

6.3 LIMITATIONS IN METHODOLOGY

While the methodology in this study provides valuable insights into long-term groundwater levels, it is important to acknowledge several limitations.

- This study utilises a framework based on intensity, duration, and frequency to characterise droughts. However, additional variables also contribute to the other classification of droughts, such as groundwater demand, land use and drought pattern. Consequently, there can be diverse definitions and interpretations of droughts.
- The accuracy of groundwater measurements is ±2 cm (Bouma et al., 2012; Ritzema et al., 2012), implying that the groundwater measurements potentially deviate from the actual groundwater levels. Furthermore, there can be measurements errors in the observations and although some errors are removed, some measurements errors may still be in the data.
- Although groundwater measuring methods in the Netherlands follow standardisation in terms of tube type, placement, measurement methods and data storage, there is room for interpretation by the monitoring manager regarding aspects such as location, depth, filter length, measurement frequency and soil type (Ritzema et al., 2012).
- Regarding meteorological data from the KNMI, both data from an automatic weather station (Ell) and manual precipitation stations (Heibloem, Sevenum) were used. Automatic weather stations can measure 5 - 8% less precipitation annually than manual stations, corresponding to approximately 0.3 mm per wet day (Brandsma, 2014).
- The limitations of the time series model are closely tied to the available data. Because the model is trained for six years, care must be taken to avoid overfitting, as the results for location Sevenum show. Additionally, the time series model uses potential evaporation, assuming that there is always sufficient moisture for plants to evaporate. However, during dry weather, when the groundwater levels decrease, plants dry out. Consequently, actual evaporation in dry periods might be lower than the model estimates. Therefore, the simulated groundwater levels are probably slightly lower than the actual groundwater levels.
- The model also assumes reference crop evaporation, considering a grass field (KNMI, 2018), which may lead to variations in actual evaporation for other crops and plants.
- Finally, the simulation of groundwater levels relies solely on precipitation and evaporation, not taking into account extractions and surface water. If this situation changes in the future, the results will no longer be representative of the location.

7 CONCLUSION

The main objective of this research was to characterise groundwater droughts using long-term groundwater levels. The main research question was:

> **Main research question:** To what extent can groundwater measurements describe the severity and occurrence of groundwater droughts?

Therefore, the main conclusion of this study is that groundwater measurements do not provide a comprehensive representation of long-term groundwater levels. Long-term groundwater simulations offer more accurate insights, particularly in estimating return periods of droughts, due to the larger statistical substantiation. This results in different return periods for the current droughts, e.g. the drought of 2018. It also results in different groundwater levels when return periods are used to justify actions, e.g. at T10. Therefore, the use of long-term simulations with historical meteorological data provides a more accurate characterisation of groundwater droughts, which can be used to justify measures and more precisely determine the extremity of currents droughts.

> **RQ1**: How do different drought statistics describe relevant properties of groundwater droughts?

Drought statistics have different characterisations of groundwater droughts using intensity, duration, and frequency of low groundwater levels:

- The **percentiles** method assesses drought intensity by the percentile of the monthly resampled groundwater level and classifies drought conditions based on percentile bands. It evaluates the duration of the drought by the consecutive time that groundwater levels remain within the percentile band and determines the frequency of the drought by multiplying the percentile by the number of measurements.
- The **average low groundwater (GLG)** method measures the intensity of drought by the distance below the GLG line. It determines the duration of drought by the consecutive time the groundwater level remains below the average low groundwater level and estimates the frequency of drought by identifying the return period based on the plot positions of peaks below the GLG line.
- The **Standardised Groundwater Index (SGI)** statistic describes the extent of drought using a standardised index. SGI quantifies the intensity of the drought by the value of the index if the index value drops below zero. The duration of drought is determined by the consecutive time the index value remains below zero, and the frequency of drought is estimating by identifying the return period based on the plot positions of peaks with SGI values below zero.
- The **annual minimum extreme value analysis (EVA)** measures the intensity of drought in terms of annual minimum groundwater levels. EVA cannot provide a direct interpretation of drought duration. However, EVA does estimate the frequency of droughts by identifying the return periods based on the plot positions of the minimum groundwater levels per year.

RQ2: To what extent do the climatological properties of the historical meteorological time series match the current climate?

The climatological properties of historical meteorological data were validated. The Kolmogorov-Smirnov test revealed that the historical precipitation at location Heibloem and Sevenum did not match the current climate. Therefore, the cumulative density functions of these precipitation series were bias-corrected. As a result, the historical meteorological data does match the current climate and can be used to simulate long-term groundwater levels.

RQ3: How can you use historical meteorological time series to simulate long-term groundwater heads?

To simulate long-term groundwater levels, a transfer function-noise model was developed using eight years of groundwater, precipitation, and evaporation measurements. Of these, six years were used to calibrate the model parameters and two years to validate the model parameters. The calibration and validation results met the goodness of fit criteria for both year-round and summer-only groundwater levels. So, a valid time series model can be established for the study area locations using precipitation and evaporation data.

The time series model was used with the validated historical meteorological series to simulate long-term groundwater levels. This resulted in groundwater levels for the period 1910-2022 for the current climate and groundwater system. These long-term simulated groundwater levels are validated using observations from the monitoring wells within the analysis period (2012-2020), which show a good goodness of fit. Furthermore, the long-term groundwater levels are compared to measurements outside the analysis period and observations in the vicinity with larger data availability. The peaks in the long-term simulated groundwater levels match the locations of the peaks in the observations.

RQ4: How do drought properties differ when drought statistics are derived from long-term groundwater levels?

Drought statistics have different interpretations of intensity, duration, and frequency of droughts, causing differences in identifying the most extreme dry year per location. Furthermore, drought statistics derived from long-term simulated groundwater levels show differences compared to drought statistics derived from observations. While the drought intensity and duration exhibit minimal differences, the frequency of droughts exhibits a significant difference due to the longer time series of the simulation. The frequency of droughts can be determined more precisely through long-term simulations. As an example, the return period for annual minima extreme value analysis of the groundwater drought of 2018 in location Ell is 1 in 24 years for long-term simulations versus 1 in 12 years for groundwater observations. Additionally, the analysis of drought statistics reveals that, despite the greater intensity of the 1976 drought, the 1921 drought endured for a more prolonged period. This underscores the significance of also considering duration when evaluating extremes in groundwater levels.

8 RECOMMENDATIONS

This research contributed to the understanding of long-term fluctuations of groundwater levels, the classification of groundwater droughts and their characterisation using intensity, duration and frequency. Recommendations are given for practical use and further research.

8.1 PRACTICAL USE

Examples for application. The methodology described in this study can be used to develop policy based on return periods. This study showed that it is possible to simulate long-term groundwater levels and determine their return periods. This allows criteria such as irrigation bans or the location of structural measures to be determined for each area. This can give a water authorities guidance when interventions are needed during dry summers. In addition, insurance companies could use this to determine whether pay outs are justified for drought claims, by examining the severity of certain groundwater droughts.

Importance of long-term simulations. The results in this study show the importance of extended time series for a proper analysis of drought frequency. Simulations are crucial, as measurements alone do not represent long-term groundwater conditions. The accurate characterisation of droughts gives potential for water authorities to establish policy criteria based on this. I would recommend informing water authorities about the importance of long-term groundwater levels and the use of groundwater simulations to gain a more comprehensive understanding of groundwater levels. Furthermore, the results in this study offer a more comprehensive perspective compared to the 'historical' groundwater levels presented in the databases from *droogteportaal* and *grondwaterstanden in beeld*, in which groundwater levels are modelled for only eight years.

Validity of historical meteorological data. The historical meteorological time series used in this study were recently developed by HKV. The findings in this research showed that for KNMI precipitation stations excluded from the generation of this dataset, the climate is not representative of the current climate. I recommend including this as a discussion point in that study or considering including more KNMI stations in a future version of this historical meteorological series. In Appendix E, additional historical meteorological data is described such as RACMO and GRADE. I would not advise to use those datasets, since they are more complex to use and involve a smaller dataset. These datasets would also need to be bias corrected, since they are developed for much larger areas. Furthermore, the historical meteorological dataset from HKV simulates properly.

Characterisation of droughts. The results in this study show that different drought statistics have different characterisations of groundwater droughts using intensity, duration and frequency. Regardless of the choice in method, the analysis of drought statistics reveals that, despite the greater intensity of the 1976 drought, the 1921 drought lasted longer. This underlines the importance of taking duration into account when evaluating extremes in groundwater levels. Therefore, my suggestion is to consider extreme droughts not solely on intensity and its frequency but also on duration. I would also recommend using percentiles and GLG for descriptive functions, e.g. in a database or to highlight average years. I think SGI and EVA are more useful for analysing droughts. Of these, I would prefer to use SGI to characterise droughts. As this index is standardised, there is no seasonal variation in its time series. In addition, the duration of the drought is included in this index, especially when longer time scales are used (SGI-1 or SGI-3).

Continue using time series modelling. HKV is actively engaged in several studies that use time series analysis to model groundwater. My own experience in this research agrees that time series analysis offers a highly effective, efficient, and straightforward approach to groundwater modelling. Additionally, the used transfer function-noise model is continuously updated and developed as well as its online documentation. I would therefore recommend continuing to use time series analysis for groundwater modelling.

8.2 FURTHER RESEARCH

Uncertainty analysis of historical meteorological data. This study did not include a sensitivity analysis for the use of meteorological data. As a result, it remains uncertain to what extent the model is sensitive to errors or uncertainties in the meteorological input series. To evaluate the sensitivity of meteorological data and take weather uncertainty into account, a possible approach involves reshuffling the historical meteorological data. By bootstrap resampling the historical meteorological data and feeding it into the model, it is possible to provide insights into the sensitivity of the historical meteorological data. To add to that, the approach can be extended by shuffling the historical meteorological data multiple times (e.g. >100 times), effectively creating a Monte Carlo simulation. This is easily feasible due to the high processing speed of the model. By aggregating the outcomes of these simulations, a potentially more probabilistic understanding of long-term groundwater levels can be obtained, thereby incorporating an uncertainty margin into the results. Therefore, I would recommend testing the sensitivity of the meteorological data to understand how uncertainties in the weather affect the simulations.

Sensitivity analysis of the calibration. In the case of location Sevenum, there is evidence that the model is too overfitted, indicating that the time series model for this location lacks the necessary flexibility to accurately simulate a wider range of groundwater levels. This issue may be overlooked due to the limited duration of the validation period, but extending it leads to a shorter calibration period. Sensitivity analysis can help identify overfitting by, for example, reducing the number of calibration points during dry summers or using extremely low and/or high precipitation as test inputs to evaluate their influence on simulated groundwater levels. Subsequently, a suitable solution, such as a more precise selection of calibration points, can be implemented to reduce the overfitting. Therefore, I suggest testing the sensitivity of the calibration such that overfitting can be prevented.

Use the KNMI'23 climate scenarios to simulate future groundwater levels. This study has proven that valid groundwater projections can be simulated using a time series model and other meteorological data. The KNMI climate scenarios are already being translated into future river discharges. As this dataset also includes precipitation and evaporation, this could also be done for groundwater levels, using the methodology from this study. This can provide insight into future groundwater levels for the current groundwater system.

Expanding the number of locations. This study uses four groundwater monitoring wells in Northern Limburg, which is classified as high sandy soils. I would recommend researching to what extent the results of this study holds for other geographical areas. I would expect it to be difficult to model using precipitation and evaporation alone, so this would require sufficient data on e.g. river discharge or polder water level. The generalisability of the results might be limited by location-specific factors like polder areas, varying subsoil characteristics, and diverse land use practices. Nevertheless, the interpretation regarding the differences between characterising droughts using measurements and simulations might hold true across these contexts.

Take into account dependency of multi-year droughts in statistics. As highlighted in this study, extreme droughts often extend across multiple years, influencing groundwater levels even during periods of relatively higher rainfall. For example, the drought in 2018 led to lower groundwater levels in the beginning of 2019, resulting in a very dry summer despite the increased precipitation that year. This implies that groundwater levels exhibit both seasonal fluctuations (summer-winter) and fluctuations between dry and wet periods over multiple years. This pattern also occurs in long-term groundwater simulations in this study. The estimation of return periods in this study assumes independent events, a condition not met using annual values due to the long response time of groundwater. The sensitivity analysis showed that the return period below T20 is therefore sensitive to the sampling strategy. A possible solution to reduce the dependency of years in annual statistics is to correct for the trend in fluctuations between dry and wet periods, retaining only seasonal fluctuations. However, this results in less insight in the groundwater dynamics. Alternatively, the variation between summer and winter groundwater levels can be analysed separately for each year, excluding any previous dry or wet years. This requires the use of a Bayesian statistic, where the frequency of a summer groundwater level is derived, given the previous winter groundwater level. My suggestion is to further explore the dependency of multiple years in groundwater statistics.

REFERENCES

1Limburg. (2023, June 18). *Grondwateronttrekking lijkt niemand te boeien*. https://www.1limburg.nl/nieuws/2224008/grondwateronttrekking-lijkt-niemand-te-boeien

AHN. (2022). *Actueel Hoogtebestand Nederland*. https://service.pdok.nl/rws/ahn/atom/index.xml

Alley, W. M., & Taylor, C. J. (2002). *Ground-Water-Level Monitoring and the Importance of Long-Term Water-Level Data*. https://www.researchgate.net/publication/255948122

Averink, J. (2013). *Methoden voor bepalen hoogste en laagste grondwaterstanden* [Master Thesis, Universiteit Twente]. https://essay.utwente.nl/64285/1/Averink_Arjan.pdf

Bakker, M., & Schaars, F. (2019). Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need Another Model. *Groundwater*, *57*(6), 826–833. https://doi.org/10.1111/gwat.12927

Bartholomeus, R. P., Voortman, B. R., & Witte, J.-P. M. (2011). *In search of the actual groundwater recharge*. www.kwrwater.nl

Benard, A., & Bos-Levenbach, E. C. (1954). *Het uitzetten van waarnemingen op waarschijnlijkheidspapier*. https://doi.org/10.1111/j.1467-9574.1953.tb00821.x

Benninga, H. J. F., Carranza, C. D. U., Pezij, M., Van Santen, P., Van Der Ploeg, M. J., Augustijn, D. C. M., & Van Der Velde, R. (2018). The Raam regional soil moisture monitoring network in the Netherlands. *Earth System Science Data*, *10*(1), 61–79. https://doi.org/10.5194/essd-10-61-2018

Bloomfield, J. P., & Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrology and Earth System Sciences*, *17*(12), 4769–4787. https://doi.org/10.5194/hess-17-4769-2013

Bouma, J., Maasbommel, M., & Schuurman, I. (2012). handboek meten van grondwaterstanden in peilbuizen. *STOWA*.

Brakkee, E., Van Huijgevoort, M., & Bartholomeus, R. P. (2021). Spatiotemporal development of the 2018-2019 groundwater drought in the Netherlands: a data-based approach. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-2021-64

Brandsma, T. (2014). *Comparison of automatic and manual precipitation networks in the Netherlands*.

Collenteur, R. A. (2021). How Good Is Your Model Fit? Weighted Goodness-of-Fit Metrics for Irregular Time Series. *Groundwater*, *59*(4), 474–478. https://doi.org/10.1111/gwat.13111

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. *Hydrology and Earth System Sciences*, *25*(5), 2931–2949. https://doi.org/10.5194/hess-25-2931-2021

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, *123*(17), 9391–9409. https://doi.org/10.1029/2017JD028200

Das, J., & Umamahesh, N. V. (2018). Assessment of uncertainty in estimating future flood return levels under climate change. *Natural Hazards*, *93*(1), 109–124. https://doi.org/10.1007/s11069-018-3291-2

Dawley, S., Zhang, Y., Liu, X., Jiang, P., Tick, G. R., Sun, H. G., Zheng, C., & Chen, L. (2019). Statistical analysis of extreme events in precipitation, stream discharge, and groundwater head fluctuation: Distribution, memory, and correlation. *Water (Switzerland)*, *11*(4). https://doi.org/10.3390/w11040707

de Bakker, H., & Schelling, J. (1989). *Systeem van bodemclassificatie voor Nederland*.

de Gruijter, J. J., van der Horst, J. B. F., Heuvelink, G. B. M., Knotters, M., & Hoogland, T. (2004). *Grondwater opnieuw op de kaart*.

De Valk, C., & Wijnant, I. L. (2019). *Uncertainty analysis of climatological parameters of the Dutch Offshore Wind Atlas (DOWA)*. https://www.dutchoffshorewindatlas.nl/about-the-atlas

Devesa, K. (2023). *Drought indicators in the East of the Netherlands*.

DINOloket. (2020, November 28). *Ondergrondgegevens*. TNO. https://www.dinoloket.nl/ondergrondgegevens

Droogers, P. (2009). *Verbetering bepaling actuele verdamping voor het strategisch waterbeheer*. https://edepot.wur.nl/7183

El Mezouary, L., El Mansouri, B., & El Bouhaddioui, M. (2020, March 11). Groundwater Forecasting using a Numerical Flow Model Coupled with Machine Learning Model for Synthetic Time Series. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3399205.3399230

European Drought Observatory. (2023, February 17). *Drought and Drought Observation*. https://edo-jrc-ec-europa-eu.ezproxy2.utwente.nl/edov2/html/1001.html

Finke, P. A., Bierkens, M. F. P., Zeeman, W. P. C., Schouten, G., Runhaar, J., van der Molen, P., van der Meer, W., de Gruijter, J. J., & van Bakel, P. J. T. (2001). *Beter werken met 'Waternood'; Een proeftoepassing in het herinrichtingsgebied 'De Leijen'*. https://www.researchgate.net/publication/40155498

Geologische Dienst Nederland. (2023). *Grondwaterstanden in beeld*. Grondwatertools. https://www.grondwatertools.nl/gwsinbeeld/showloc?wellid=B42H0043&screenid=001

Guo, M., Yue, W., Wang, T., Zheng, N., & Wu, L. (2021). Assessing the use of standardized groundwater index for quantifying groundwater drought over the conterminous US. *Journal of Hydrology*, *598*, 126227. https://doi.org/10.1016/J.JHYDROL.2021.126227

Gupta, R., Bhattarai, R., & Mishra, A. (2019). Development of climate data bias corrector (CDBC) tool and its application over the agro-ecological zones of India. *Water (Switzerland)*, *11*(5). https://doi.org/10.3390/w11051102

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950-2006. *Journal of Geophysical Research Atmospheres*, *113*(20). https://doi.org/10.1029/2008JD010201

Hazeu, G., Schuiling, R., Thomas, D., Vittek, M., Storm, M., & Bulens, D. J. (2023). Landelijk Grondgebruiksbestand Nederland 2021 (LGN2021). *Wageningen Environmental Research*.

Hegnauer, M., Beersma, J. J., van den Boogaard, H. F. P., Buishand, T. A., & Passchier, R. H. (2014). *Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins-Final report of GRADE 2.0*.

Hoekstra, A. Y. (2018). *Lecture Notes Water* (Vol. 885). Faculty of Engineering Technology.

Hoogland, T., Knotters, M., Pleijter, M., & Walvoort, D. J. J. (2014). *Actualisatie van de grondwatertrappenkaart van holoceen Nederland*. www.BISNederland.wur.nl

Hsin-Fu Yeh, H. F., & Chang, C. F. (2019). Using Standardized Groundwater Index and Standardized Precipitation Index to Assess Drought Characteristics of the Kaoping River Basin, Taiwan. *Water Resources*, *46*(5), 670–678. https://doi.org/10.1134/S0097807819050105

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, *25*(8–12), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

KNMI. (2018). *Referentie-gewasverdamping*. https://cdn.knmi.nl/system/ckeditor_assets/attachments/62/ref_gewasverdamping.pdf

KNMI. (2021a, January 5). *Achtergrondinformatie berekening klimaatnormalen 1991- 2020*. https://www.knmi.nl/kennis-en-datacentrum/achtergrond/achtergrondinformatie-berekening-klimaatnormalen-1991-2020

KNMI. (2021b, June 28). *Achtergrondinformatie neerslagindexen SPI en SPEI*. https://www.knmi.nl/kennis-en-datacentrum/achtergrond/achtergrondinformatie-neerslagindex-spi#:~:text=Het%20KNMI%20rekent%20de%20SPI,ontwikkeling%20van%20de%20SPI%20toegevoegd.

KNMI. (2023a). *Dagwaarden neerslagstations*. https://www.knmi.nl/nederland-nu/klimatologie/monv/reeksen

KNMI. (2023b). *Dagwaarnemingen weerstations*. https://daggegevens.knmi.nl/klimatologie/daggegevens

Knotters, M., Hoogland, T., & Brus, D. (2013). Validatie van grondwater-standskaarten met behulp van de Landelijke Steekproef Kaarteenheden. *Stromingen*, *19*(3 & 4), 35–47. https://edepot.wur.nl/315231

Knotters, M., Walvoort, D., Brouwer, F., Stuyt, L., & Okx, J. (2018). Landsdekkende, actuele informatie over grondwatertrappen digitaal beschikbaar. *Wagningen Environmental Research*.

Lenderink, G., van den Hurk, B., van Meijgaard, E., van Ulden, A., & Cuijpers, H. (2003). Simulation of present-day climate in RACMO2: first results and model developments. *KNMI*.

Martinez-Villalobos, C., & David Neelin, J. (2019). Why Do Precipitation Intensities Tend to Follow Gamma Distributions? *Journal of the Atmospheric Sciences*, *76*(11), 3611–3631. https://doi.org/10.1175/JAS-D-18

McCluskey, C. J., Guers, M. J., & Conlon, S. C. (2021). Minimum sample size for extreme value statistics of flow-induced response. *Marine Structures*, *79*, 103048. https://doi.org/10.1016/J.MARSTRUC.2021.103048

Mohanasundaram, S., Narasimhan, B., & Suresh Kumar, G. (2017). Transfer function noise modelling of groundwater level fluctuation using threshold rainfall-based binary-weighted parameter estimation approach. *Hydrological Sciences Journal*, *62*(1), 36–49. https://doi.org/10.1080/02626667.2016.1171325

Navarro-Racines, Tarapues-Montenegro, And Ramírez-Villegas, J. E., & Ramírez-Villegas. (2015). *BIAS-CORRECTION IN THE CCAFS-CLIMATE PORTAL: A description of methodologies*. http://www.ccafs-climate.org/data_bias_correction/

NHI. (2023). Documentatie datasets. *STOWA*. https://nhi.nu/documents/46/Documentatie_datasets.pdf

Obergfell, C., Bakker, M., & Maas, K. (2019). Estimation of Average Diffuse Aquifer Recharge Using Time Series Modeling of Groundwater Heads. *Water Resources Research*, *55*(3), 2194–2210. https://doi.org/10.1029/2018WR024235

Petersen-Perlman, J. D., Aguilar-Barajas, I., & Megdal, S. B. (2022). Drought and groundwater management: Interconnections, challenges, and policyresponses. In *Current Opinion in Environmental Science and Health* (Vol. 28). Elsevier B.V. https://doi.org/10.1016/j.coesh.2022.100364

Peterson, T. J., & Western, A. W. (2014). Nonlinear time-series modeling of unconfined groundwater head. *Water Resources Research*, *50*(10), 8330–8355. https://doi.org/10.1002/2013WR014800

Pezij, M., Augustijn, D. C. M., Hendriks, D. M. D., & Hulscher, S. J. M. H. (2020). Applying transfer function-noise modelling to characterize soil moisture dynamics: a data-driven approach using remote sensing data. *Environmental Modelling & Software*, *131*, 104756. https://doi.org/10.1016/J.ENVSOFT.2020.104756

Pezij, M., & Lugt, D. (2023). *Droogtestatistiek; Meteo-onderzoek ten behoeve van het waterbeheer: Deelrapport 3*. https://www.stowa.nl/publicaties/droogtestatistiek-meteo-onderzoek-ten-behoeve-van-het-waterbeheer-deelrapport-3

Philip, S. Y., Kew, S. F., Van Der Wiel, K., Wanders, N., Jan Van Oldenborgh, G., & Philip, S. Y. (2020). Regional differentiation in climate change induced drought trends in the Netherlands. *Environmental Research Letters*, *15*(9). https://doi.org/10.1088/1748-9326/ab97ca

Pidwirny, M. (2006). The Hydrologic Cycle. *Fundamentals of Physical Geography, 2nd Edition*. http://www.physicalgeography.net/fundamentals/8b.html

Programma Basisregistratie Ondergrond. (2023). *Bodemkaart (SGM)*. https://basisregistratieondergrond.nl/inhoud-bro/registratieobjecten/modellen/bodemkaart-sgm/

Qian, W., & Chang, H. H. (2021). Projecting health impacts of future temperature: A comparison of quantile-mapping bias-correction methods. *International Journal of Environmental Research and Public Health*, *18*(4), 1–12. https://doi.org/10.3390/ijerph18041992

Ratering, P. (2023). *Effects of human landscape interventions on groundwater drought* [Master Thesis, University of Twente]. https://landschapoverijssel.nl/twente/doorbraak

Ritzema, H., Heuvelink, G., Heinen, M., Bogaart, P., van der Bolt, F., Hack-ten Broeke, M., Hoogland, T., Knotters, M., Massop, H., & Vroon, H. (2012). Meten en interpreteren van grondwaterstanden. In *Alterra-rapport 2345*.

Robinson, D. A., Nemes, A., Reinsch, S., Radbourne, A., Bentley, L., & Keith, A. M. (2022). Global meta-analysis of soil hydraulic properties on the same soils with differing land use. *Science of The Total Environment*, *852*, 158506. https://doi.org/10.1016/J.SCITOTENV.2022.158506

Schumacher, D. L., Keune, J., Dirmeyer, P., & Miralles, D. G. (2022). Drought self-propagation in drylands due to land–atmosphere feedbacks. *Nature Geoscience*, *15*(4), 262–268. https://doi.org/10.1038/s41561-022-00912-7

Škarpich, V., Horáček, M., Galia, T., Kapustová, V., & Šala, V. (2016). The effects of river patterns on riparian vegetation: A comparison of anabranching and single-thread incised channels. *Moravian Geographical Reports*, *24*(3), 24–31. https://doi.org/10.1515/mgr-2016-0014

The editors of Encyclopaedia Britannica. (2023). Water cycle. In Rafferty John P. (Ed.), *Encyclopedia Britannica*. https://www.britannica.com/science/water-cycle

van den Eertwegh, G., de Louw, P., Witte, J.-P., van Huijgevoort, M., Bartholomeus, R., van Deijl, D., van Dam, J., Hunink, J., America, I., Pouwels, J., Hoefsloot, P., & de Wit, J. (2021). *Droogte Zandgronden Nederland*.

van der Gaast, J. W. J., Vroon, H. R. J., & Massop, H. Th. L. (2010). *Grondwaterregime op basis van karteerbare kenmerken*. https://edepot.wur.nl/163486

van Engelenburg, J., de Jonge, M., Rijpkema, S., van Slobbe, E., & Bense, V. (2020). Hydrogeological evaluation of managed aquifer recharge in a glacial moraine complex using long-term groundwater data analysis. *Hydrogeology Journal*, *28*(5), 1787–1807. https://doi.org/10.1007/s10040-020-02145-7

Van Hussen, K., Van De Velde, I., Läkamp, R., & Van Der Kooij, S. (2019). *Economische schade door droogte in 2018*.

van Kekem, A. J., Hoogland, T., & van der Horst, J. B. F. (2005). *Uitspoelingsgevoelige gronden op de kaart*.

Van Loon, A. F. (2013). *On the propagation of drought How climate and catchment characteristics influence hydrological drought development and recovery*.

Van Meijgaard, E., Van Ulft, L. H., Van De Berg, W. J., Bosveld, F. C., Van Den Hurk, B. J. J. M., Lenderink, G., & Siebesma, A. P. (2008). *The KNMI regional atmospheric climate model RACMO version 2.1*.

Van Voorst, L., & Van Den Brink, H. (2022). *Improving the GRADE weather generator by using synthetic datasets from RACMO and SEAS5*.

Verhagen, F., & Avis, L. (2021). Lessen uit lange grondwaterreeksen. *Stromingen*, *27*(2).

Von Asmuth, J., Baggelaar, P., Bakker, M., Brakenhoff, D., Collenteur, R. A., Ebbens, O., Mondeel, H., Klop, S., & Schaars, F. (2021). *Handleiding Tijdreeksanalyse*. STOWA. www.stowa.nl

Von Asmuth, J., Bierkens, M., & Maas, K. (2002). Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research*, *38*(12), 23-1-23–12. https://doi.org/10.1029/2001wr001136

von Asmuth, J., Maas, K., Knotters, M., Bierkens, M. F. P., Bakker, M., Olsthoorn, T. N., Cirkel, D. G., Leunk, I., Schaars, F., & von Asmuth, D. C. (2012). Software for hydrogeologic time series

analysis, interfacing data with physical insight. *Environmental Modelling and Software*, *38*, 178–190. https://doi.org/10.1016/j.envsoft.2012.06.003

Vonk, M. (2021). *Performance of nonlinear time series models to simulate synthetic ground-water table time series from an unsaturated zone model*.

Waterschap Limburg. (2019). *Kaarten en meetgegevens*. https://www.waterschaplimburg.nl/uwbuurt/kaarten-meetgegevens/

World Meteorological Organization. (2017). *WMO Guidelines on the Calculation of Climate Normals*. https://library.wmo.int/viewer/55797?medianame=1203_en_#page=1&viewer=picture&o=bookmark&n=0&q=

Zaadnoordijk, W. J., Bus, S. A. R., Lourens, A., & Berendrecht, W. L. (2019). Automated Time Series Modeling for Piezometers in the National Database of the Netherlands. *Groundwater*, *57*(6), 834–843. https://doi.org/10.1111/GWAT.12819

APPENDIX A: ADDITIONAL MAPS OF STUDY AREA

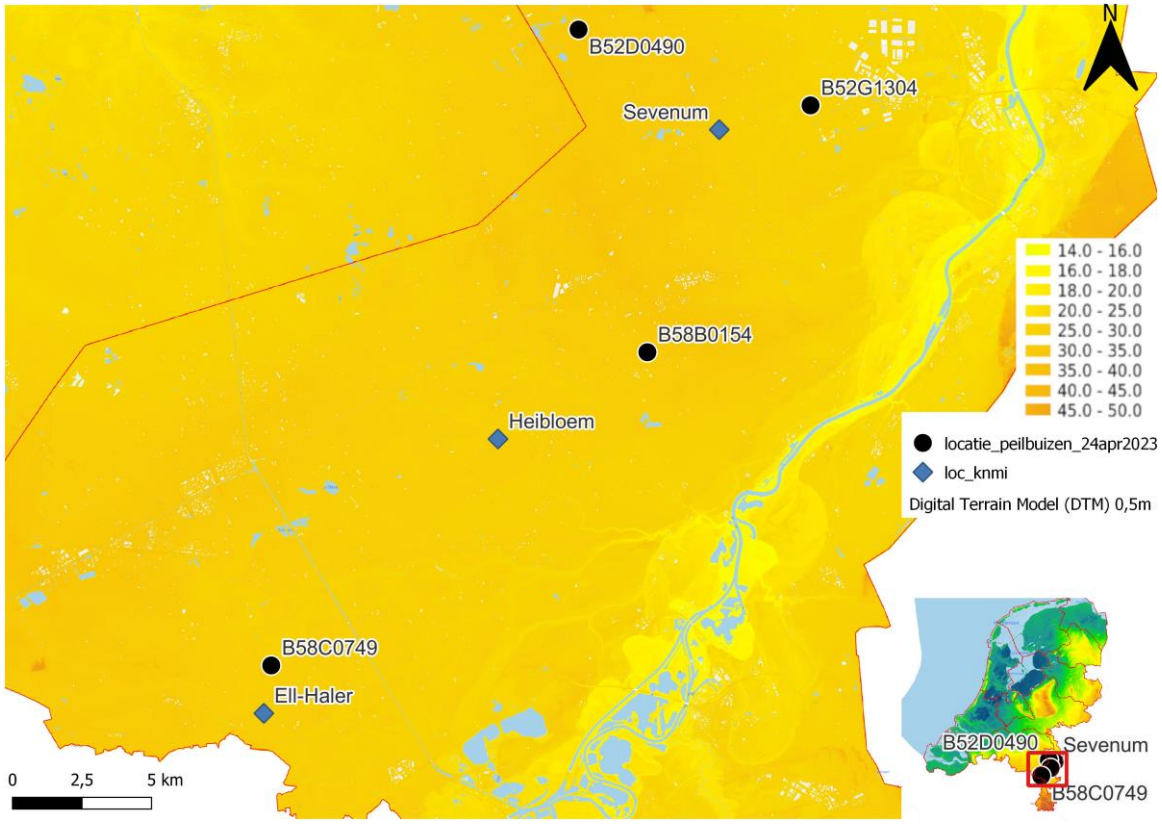


Figure 29: Ground elevation from Digital Terrain Model AHN3 (AHN, 2022).

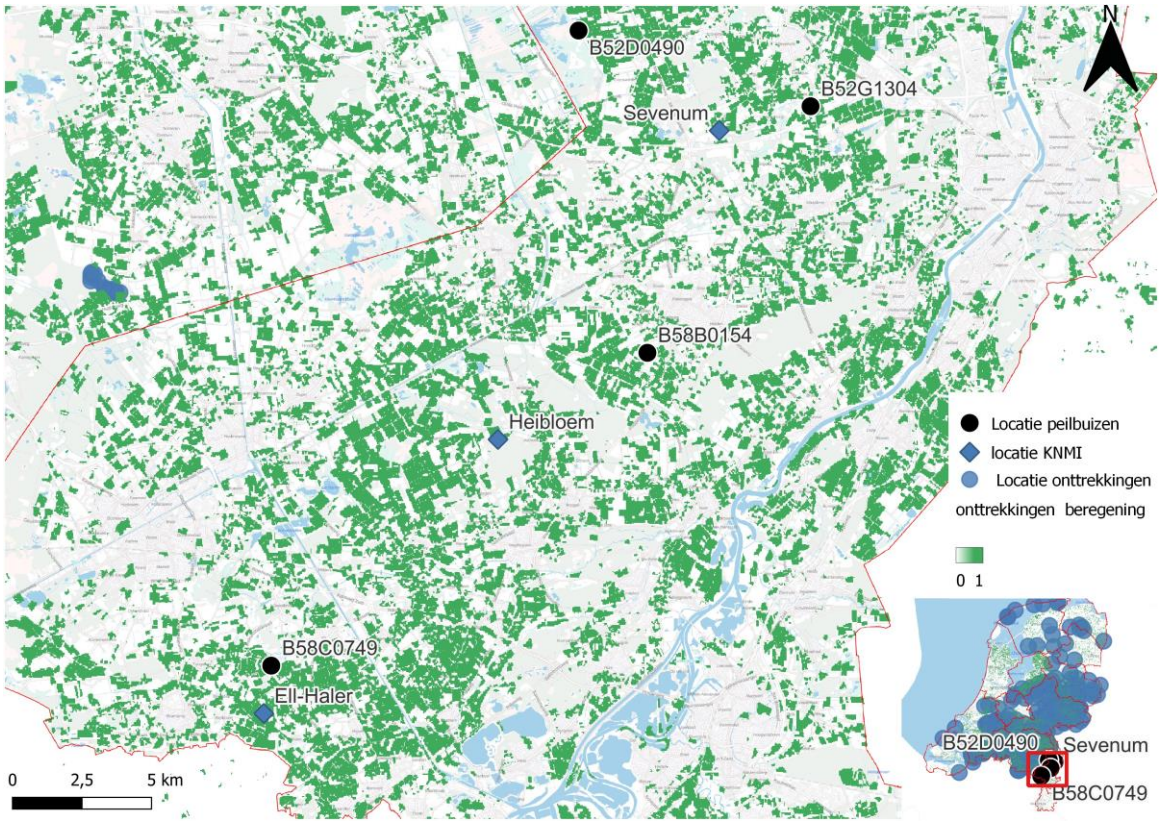


Figure 30: Drinking water withdrawals (blue patches) and possible irrigation extractions (green areas) (NHI, 2023).

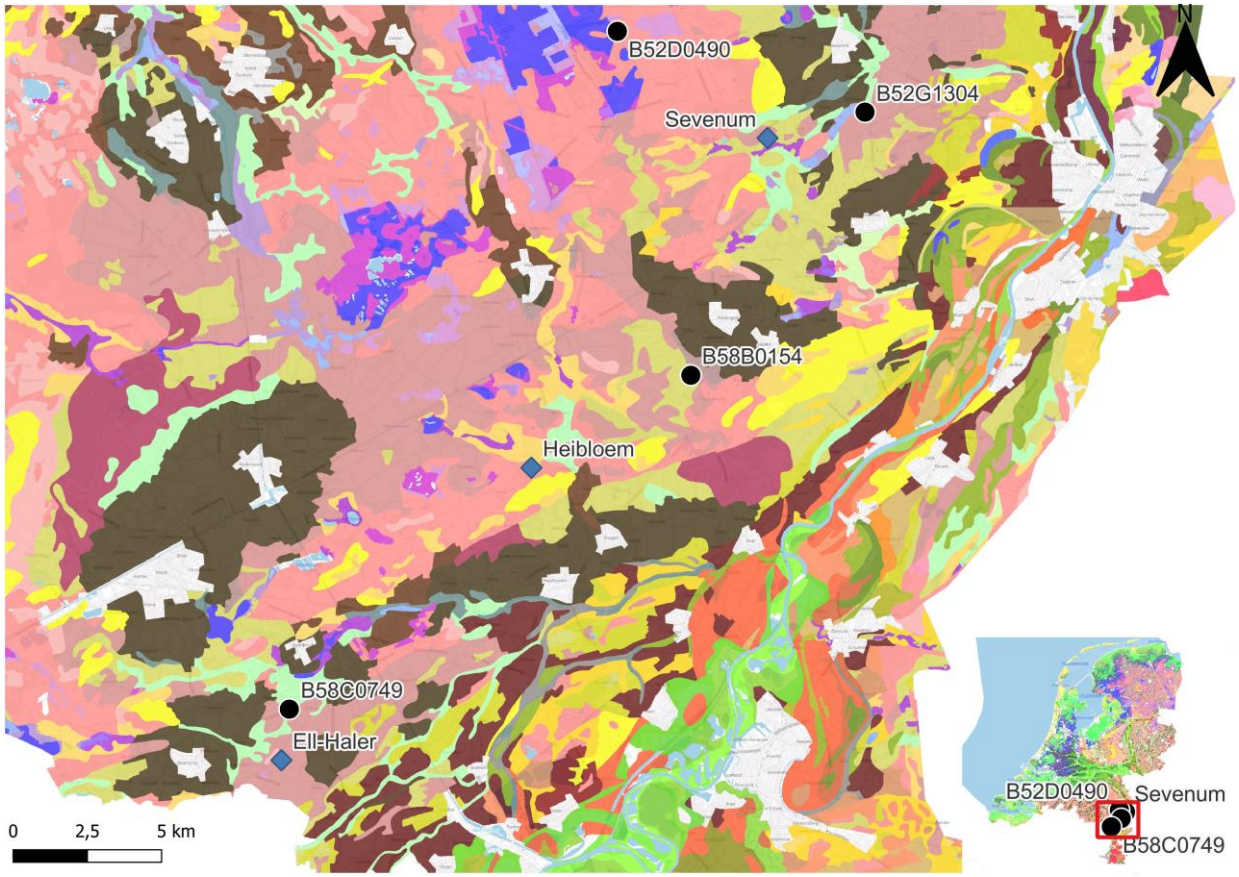


Figure 31: Soil surfaces BRO Soil Map (SGM) (Programma Basisregistratie Ondergrond, 2023).

Figure 32: Subsurface soil fractures (NHI, 2023).

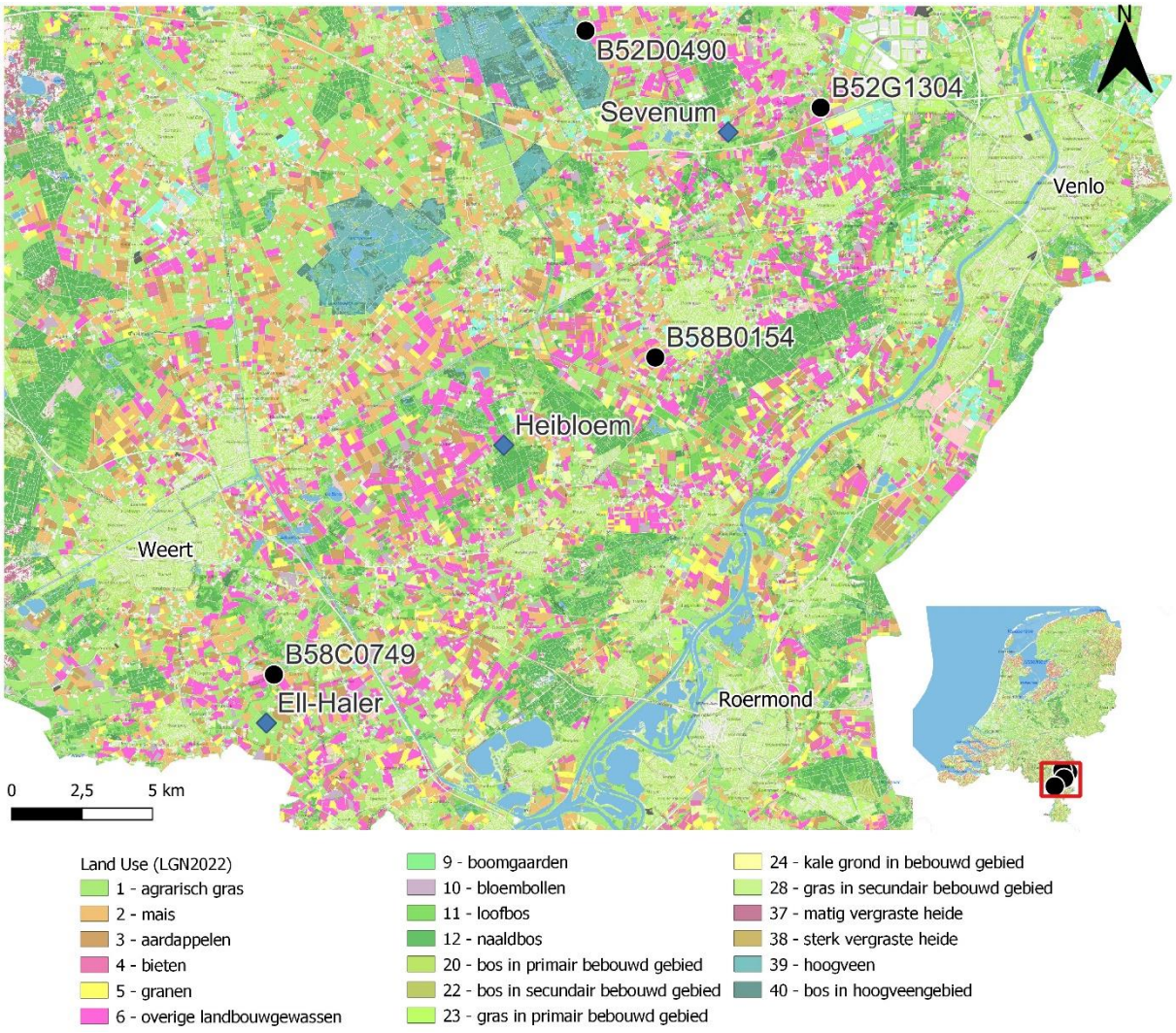


Figure 33: Land use of study area according to LGN2022 (Hazeu et al., 2023).

CHARACTERISTICS OF LOCATIONS

Ell

The land use surrounding monitoring well Ell predominantly consists of managed agrarian grasslands, with a deciduous forest (Heijkersbroek) in close proximity. The soil is classified as Beekeerdgrond, which typically comprises of sandy soils found in stream valleys. The nearest stream is Vliet at approximately 50 meters from the monitoring well.

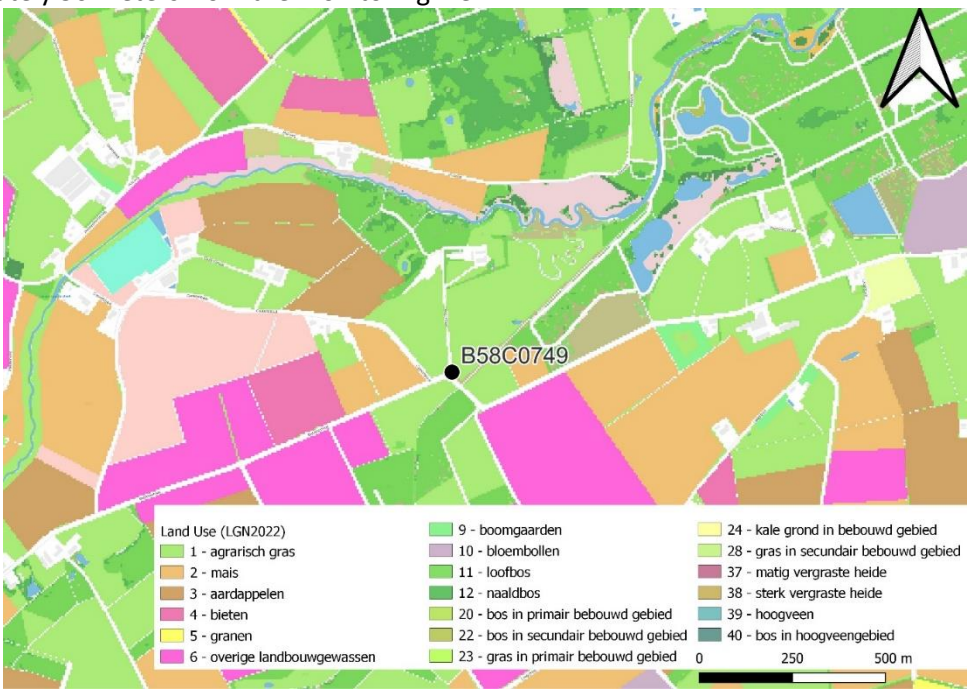


Figure 34: Land use zoomed in on location Ell.

Heibloem

The land use surrounding the Heibloem monitoring well is primarily dedicated to agriculture, featuring crops such as potatoes, corn and sugar beets. In Heibloem, the soil is classified as Gooreerdgrond, characterized by calcareous sandy soils with a topsoil primarily composed of loamy fine sand. Nearest surface water is Egchelbeek at approximately 150 meters from the monitoring well.

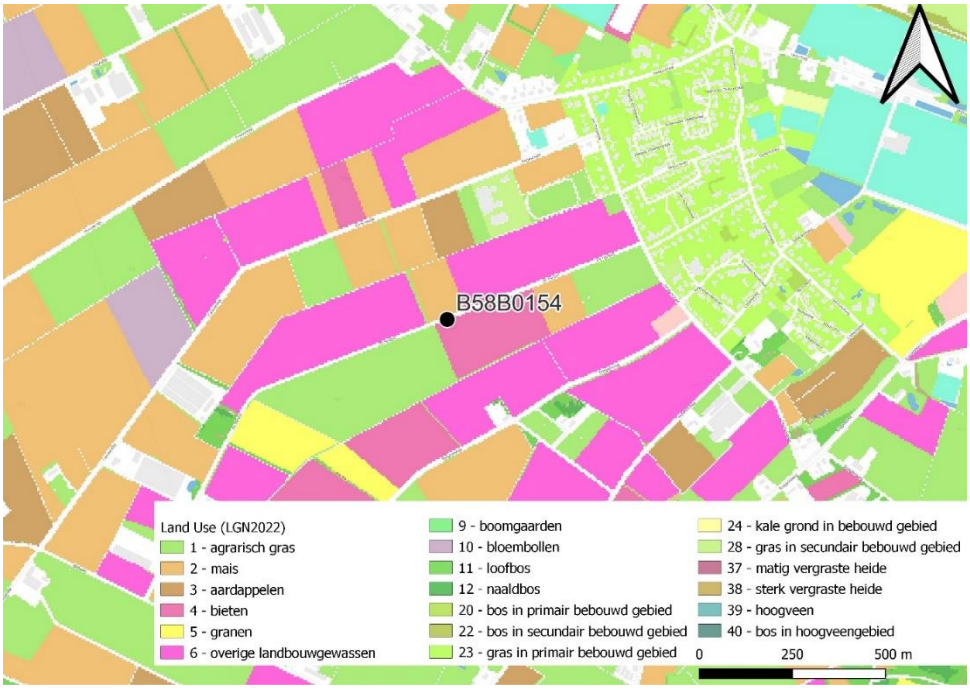


Figure 35: Land use zoomed in on location Heibloem.

Sevenum

The Sevenum monitoring well is located on the border of agricultural areas and nature-managed grasslands with adjacent deciduous forest (Elsbeemden). Similarly, in Sevenum, the soil is identified as Veldpodzolgrond, which are humic podzolic soils mainly consisting of loamy fine sand. The nearest surface water is the Groote Molenbeek at approximately 110 meters from the monitoring well.

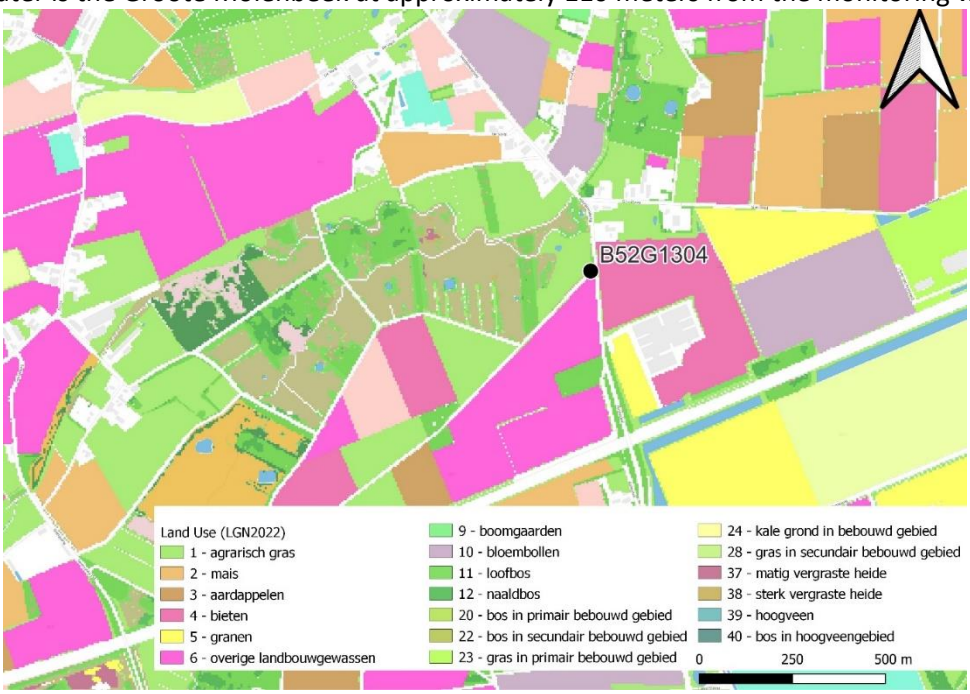


Figure 36: Land use zoomed in on location Sevenum.

Mariapeel

Finally, the Mariapeel monitoring well is situated at the edge of a Natura-2000 area (Mariapeel) and adjacent to a stretch of deciduous forest, with agricultural land nearby. Mariapeel also exhibits characteristics of veldpodzolgronden, with a soft loamy fine sand topsoil. The adjacent natural area contains peat with humuspodzol and marshy podzolic soils. The nearest surface water is the Peelkanaal at approximately 30 meters from the monitoring well. This is less than the defined criteria, such that it might impact groundwater monitoring (Bouma et al., 2012).

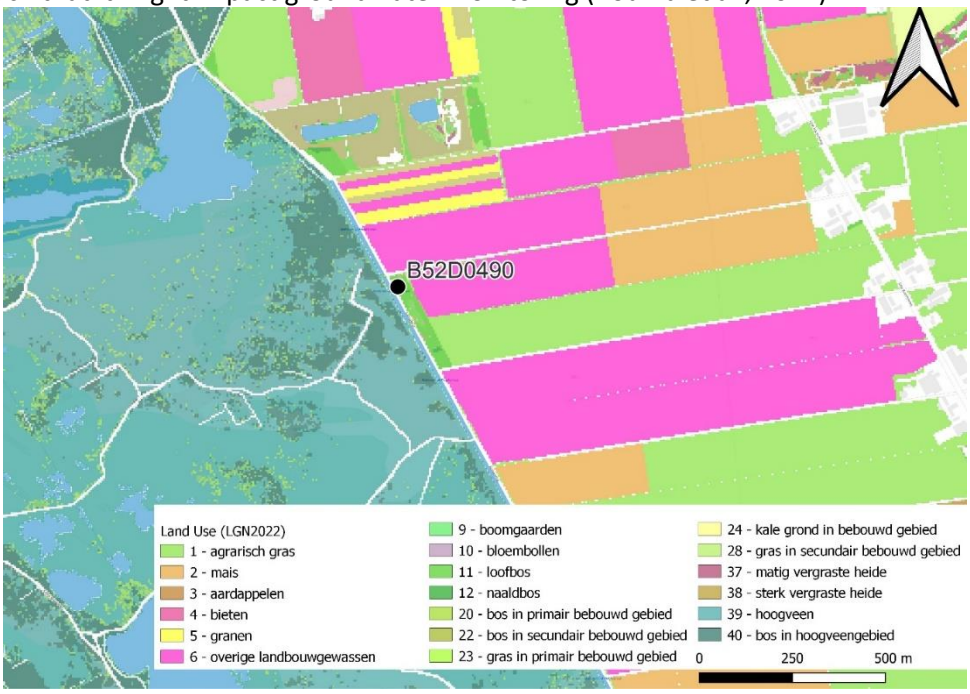


Figure 37: Land use zoomed in on location Mariapeel.

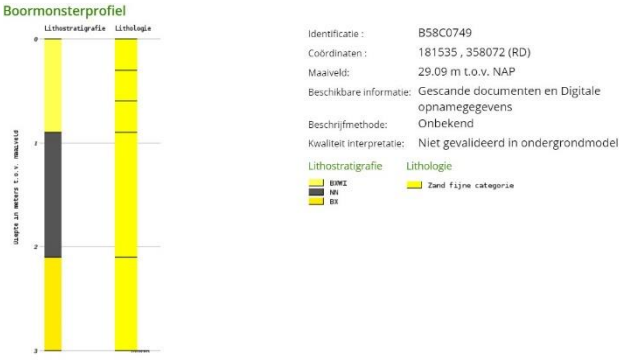


Figure 38: Profile bore sample location Ell.

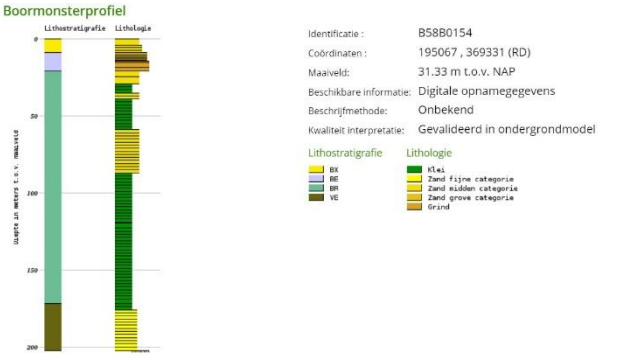


Figure 39: Profile bore sample location Heibloem.

Figure 40: Profile bore sample location Sevenum.

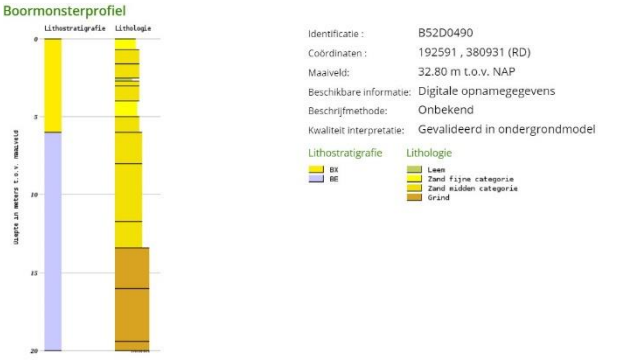


Figure 41: Profile bore sample location Mariapeel.

APPENDIX B: INPUT DATA

INPUT DATA RQ1: METHODOLOGY TO DERIVE STATISTICS OF MEASUREMENTS

Some measurements expose unrealistic changes to the groundwater head within one observation time step. These measurements are marked as measurements errors since they cause unreliability in further analysis. Therefore, these measurements are removed from the dataset. Whether a measurement is an error is based on visual inspection of the measurement series and comparing outliers with other locations and measured precipitation. The found measurement errors and a few examples are given in the appendix. The gaps due to the removed errors are not interpolated, since the measurements series already contains data gaps which are too large to interpolate.

The processed measurements of the observation wells are displayed throughout the year in Figure 42. In the figure, the annual regime of the groundwater heads is presented. Furthermore, it is easier to highlight peak events in relation to the other years. A high peak is shown in June 2016 for all locations, due to a large rainfall event. Furthermore, the groundwater heads in the years 2018, 2019 and 2020 are low, especially in the months July – September. It is also worth noting that groundwater levels in 2019 already start below average, due to low groundwater levels in late 2018.

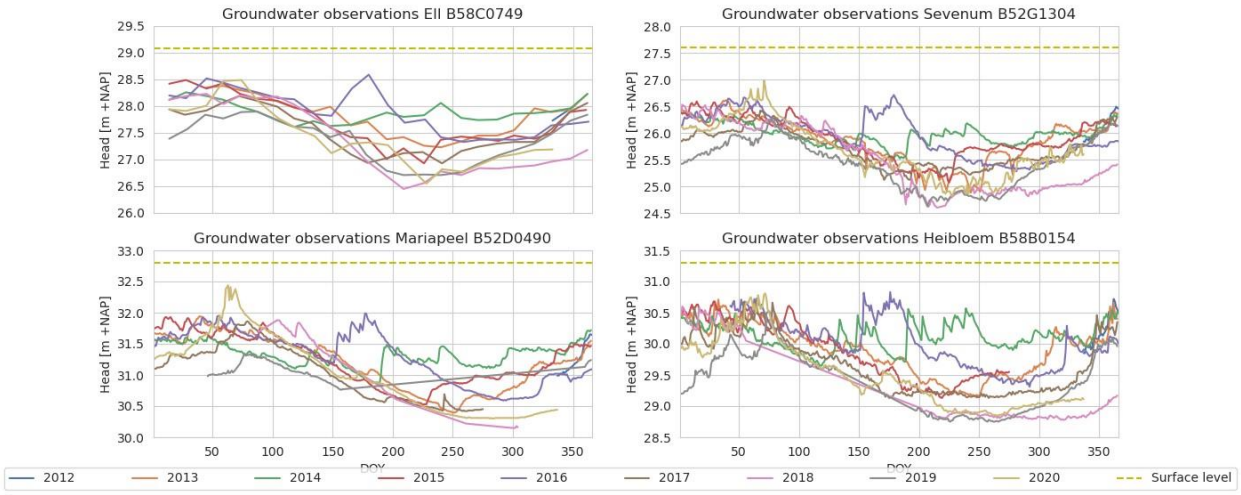


Figure 42: Groundwater observations per Day-Of-Year.

Table 13: Properties of groundwater measurements for analysis period.

Properties measurement series	monitoring well Ell	monitoring well Heibloem	monitoring well Sevenum	monitoring well Mariapeel
Start date measurement period	28-11-2012	2-12-2012	2-12-2012	2-12-2012
End date measurement period	28-11-2020	2-12-2020	2-12-2020	2-12-2020
Number of measurements	191	2574	2863	2275
Average (cm + NAP)	2763	2974	2575	3117
standard deviation (cm)	49.2	52.9	45.5	43.7
Minimum (cm + NAP)	2645	2875	2460	3015
10-percentile (cm + NAP)	2693	2891	2507	3055
50-percentile (cm + NAP)	2769	2982	2582	3123
90-percentile (cm + NAP)	2823	3039	2630	3170
Maximum (cm + NAP)	2859	3083	2698	3256
Groundwater dynamics (m)	1.3	1.48	1.23	1.15

Table 14: Dates of measurement errors based on precipitation data and visual comparison with other locations.

Measurement errors:	
Heibloem	**Sevenum**
28-10-2013	4-6-2013
4-11-2013	5-6-2013
27-1-2014	23-5-2017
15-2-2014	2-6-2017
28-5-2014	11-7-2017
26-8-2014	22-4-2020
28-12-2014	24-4-2020
9-1-2015	9-5-2020
10-1-2015	10-5-2020
14-1-2015	18-5-2020
27-2-2015	19-5-2020
30-3-2015	**Mariapeel**
31-3-2015	6-1-2015
23-2-2016	17-2-2020
5-6-2016	18-2-2020
9-3-2017	**Ell**
30-10-2018	28-12-2012
11-2-2019	28-6-2014

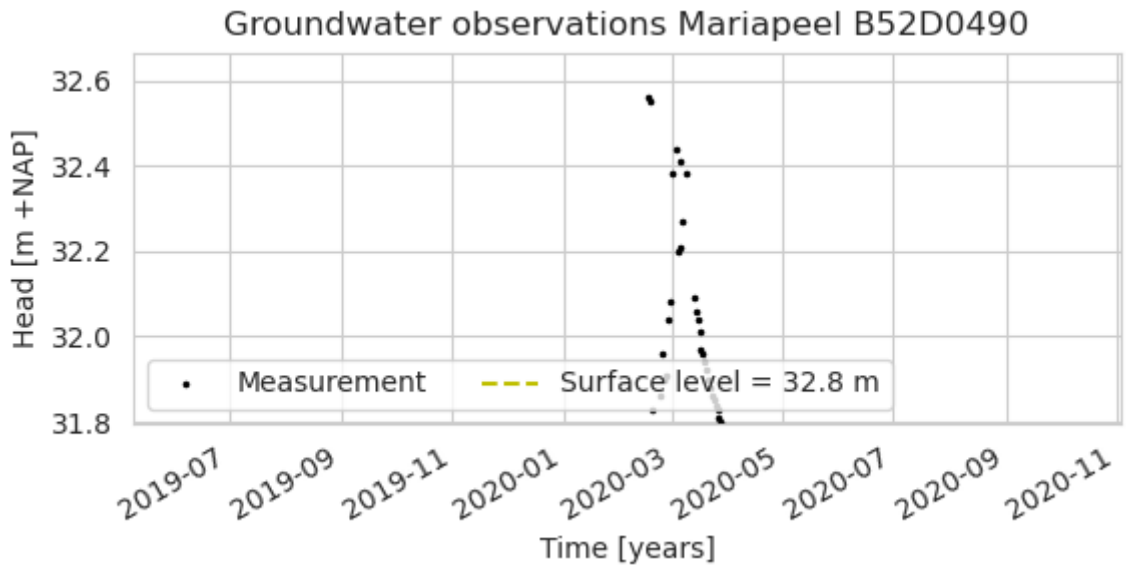


Figure 43: Example of measurement error for location Mariapeel.

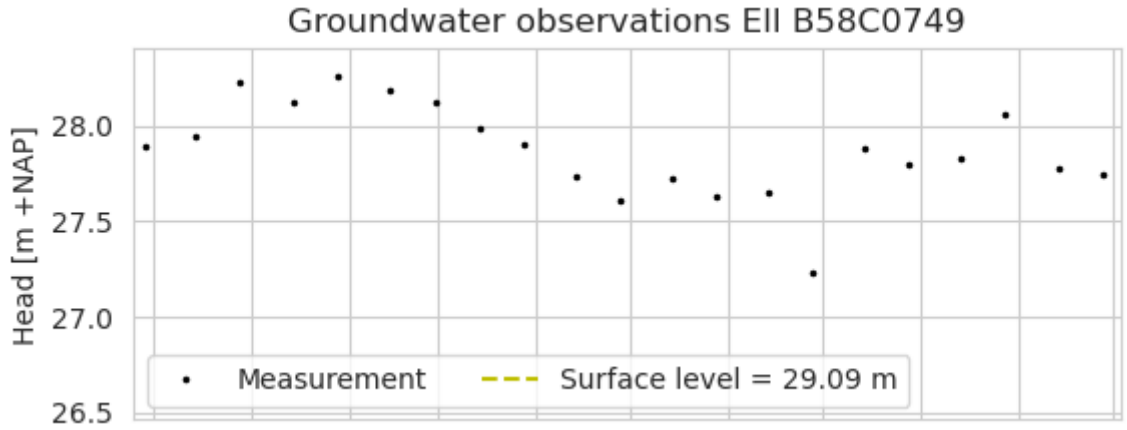


Figure 44: Example of measurement error for location Ell.

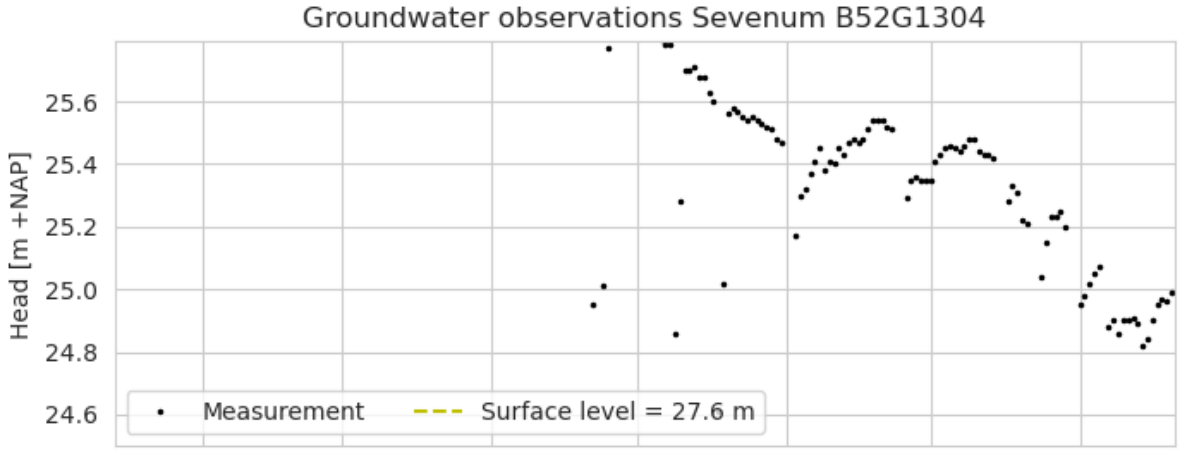


Figure 45: Example of measurement error for location Sevenum.

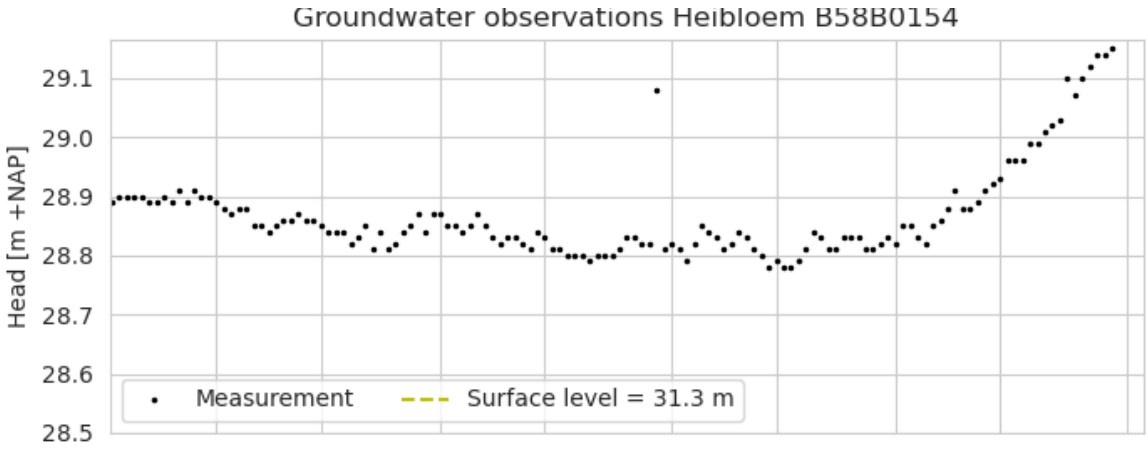


Figure 46: Example of measurement error for location Heibloem.

INPUT DATA RQ2: VALIDATION OF HISTORICAL METEOROLOGICAL SERIES

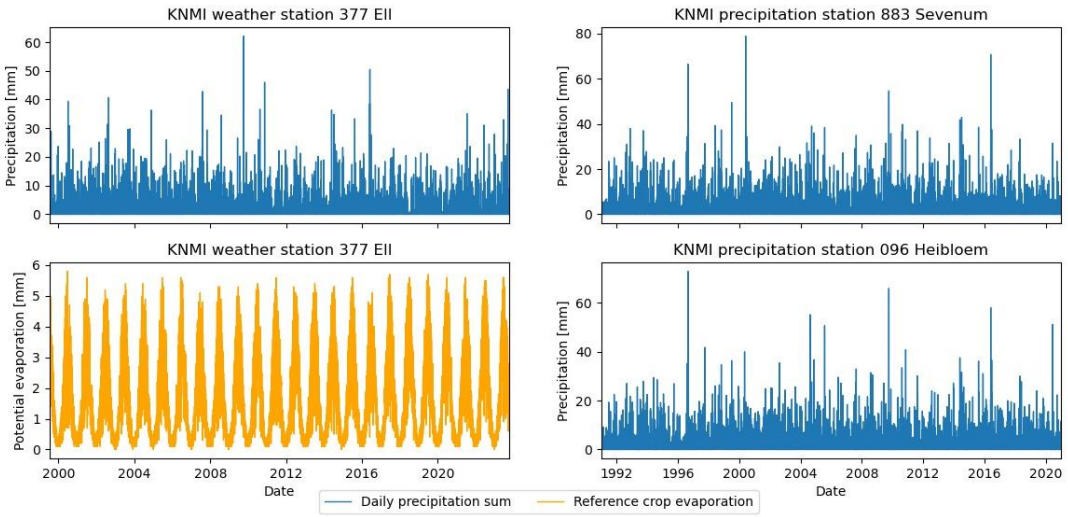


Figure 47: Observed precipitation (Ell, Heibloem, Sevenum) and evaporation (Ell) for the KNMI stations for approximately 30 years (1991-2020 for the precipitation stations and 1999-2023 for weather station Ell due to lack of data).

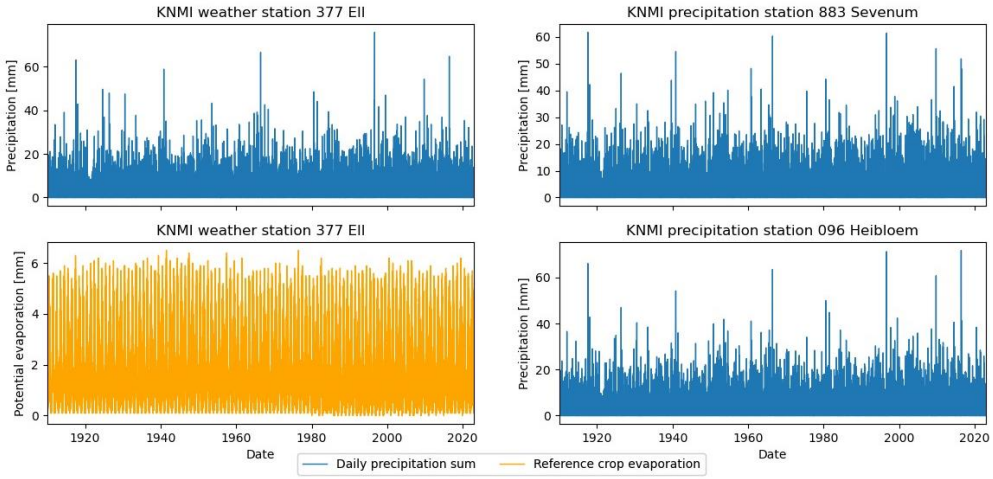


Figure 48: Historical precipitation and evaporation data extracted for the locations of the KNMI stations for the period of 1910-2022.

METEOROLOGICAL OBSERVATIONS

The KNMI is the primary source for meteorological data, utilizing automated weather stations and manual precipitation stations. Precipitation is measured using rain gauges, with 1 millimeter corresponding to 1 liter of water per square meter. Evaporation not only depends on meteorological quantities such as solar radiation, wind and temperature, but also on non-meteorological conditions such as soil moisture, crop growth and ground cover. Given this, actual evapotranspiration cannot be predicted (KNMI, 2018). Therefore, KNMI calculates reference crop evaporation from proxy meteorological data. In summer, wind tends to play a less significant role in influencing evaporation making it possible to accurately estimate evaporation based solely on the temperature and solar radiation. The KNMI developed a calculation method based on these parameters, known as the Makkink method. Reference crop evaporation is used to estimate the potential evapotranspiration of a crop or vegetation (Bartholomeus et al., 2011), and is defined as the evapotranspiration from a dry grass surface optimally supplied by water (Droogers, 2009).

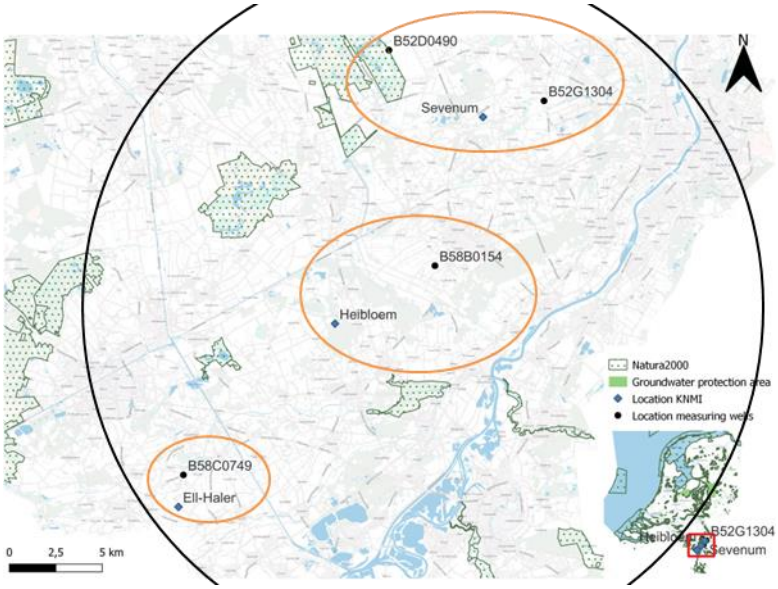


Figure 49: Locations of the input data; Historical meteorological data at locations of measuring wells; Observed precipitation data in KNMI station Ell, Heibloem and Sevenum and evaporation data at KNMI station Ell. Circles denote the use of which KNMI data for which measuring well (Orange = precipitation, black = evaporation).

APPENDIX C: ADDITIONAL RQ1: METHODOLOGY TO DERIVE STATISTICS

PERCENTILES

Table 15: Median (50th percentile) value for observations and simulations for location Ell.

Month	Observation [m+ NAP]	Simulation [m + NAP]	Difference [m+ NAP]
1	28.14	28.26	-0.12
2	28.18	28.39	-0.21
3	28.20	28.33	-0.14
4	28.00	28.14	-0.14
5	27.72	27.85	-0.13
6	27.40	27.61	-0.21
7	27.16	27.46	-0.30
8	27.10	27.36	-0.26
9	27.25	27.45	-0.20
10	27.34	27.55	-0.22
11	27.44	27.75	-0.31
12	27.84	28.02	-0.18
		sum	-2.42
		average	-0.20

AVERAGE LOW GROUNDWATER LEVEL

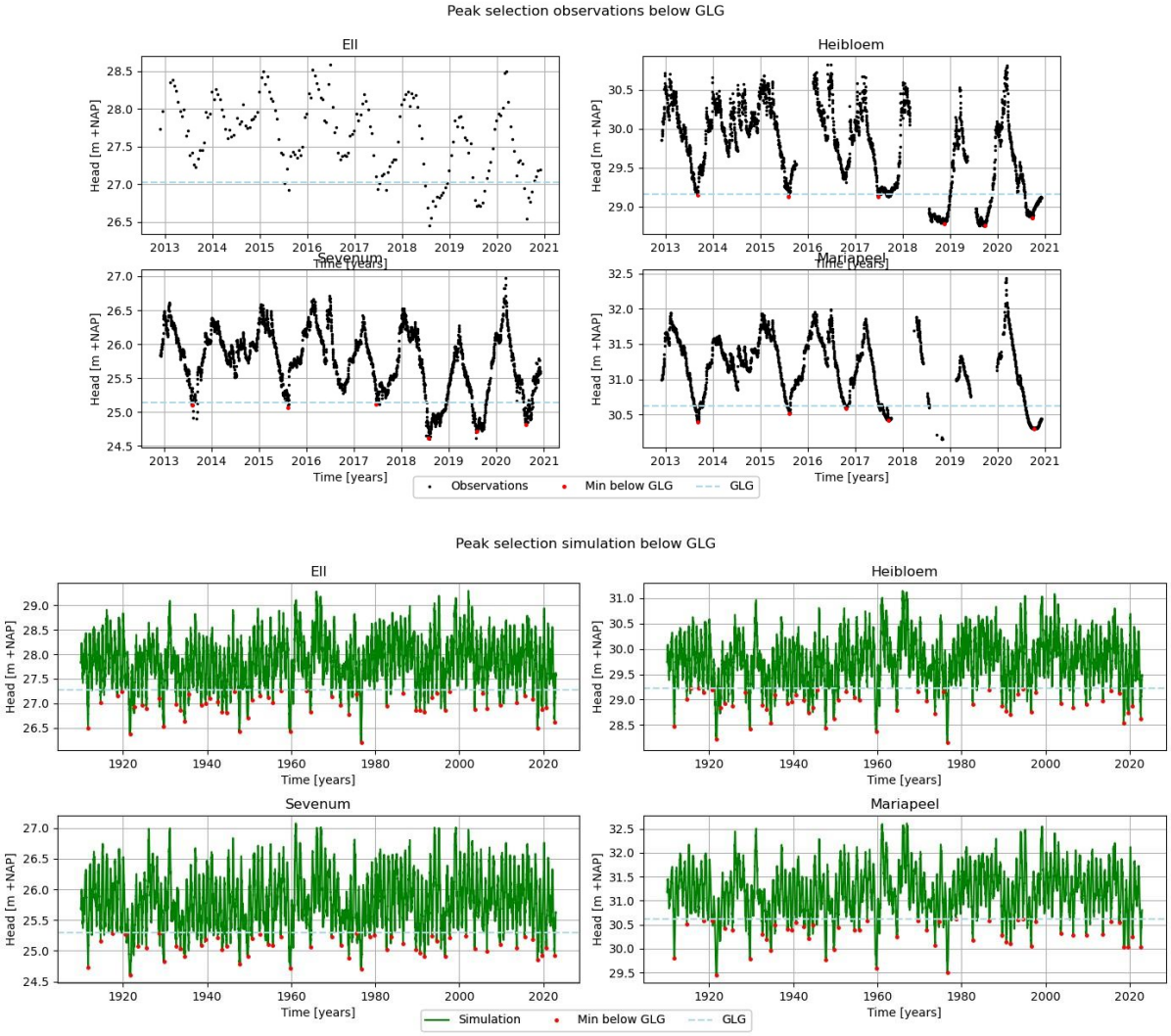


Figure 50: Selection of drought events with minimum groundwater head below GLG.

Table 16: GxG of observed and simulated groundwater levels for all locations.

Statistic [m below surface level]	Ell		Heibloem		Sevenum		Mariapeel	
	Observed	Simulated	Observed	Simulated	Observed	Simulated	Observed	Simulated
GHG	0.87	0.64	0.92	0.98	1.28	1.15	1.09	0.98
GVG	0.98	0.83	1.12	1.16	1.46	1.44	1.22	1.14
GG	1.5	1.25	1.62	1.55	1.88	1.8	1.66	1.61
GLG	2.06	1.82	2.14	2.07	2.46	2.43	2.17	2.17

STANDARDISED GROUNDWATER INDEX

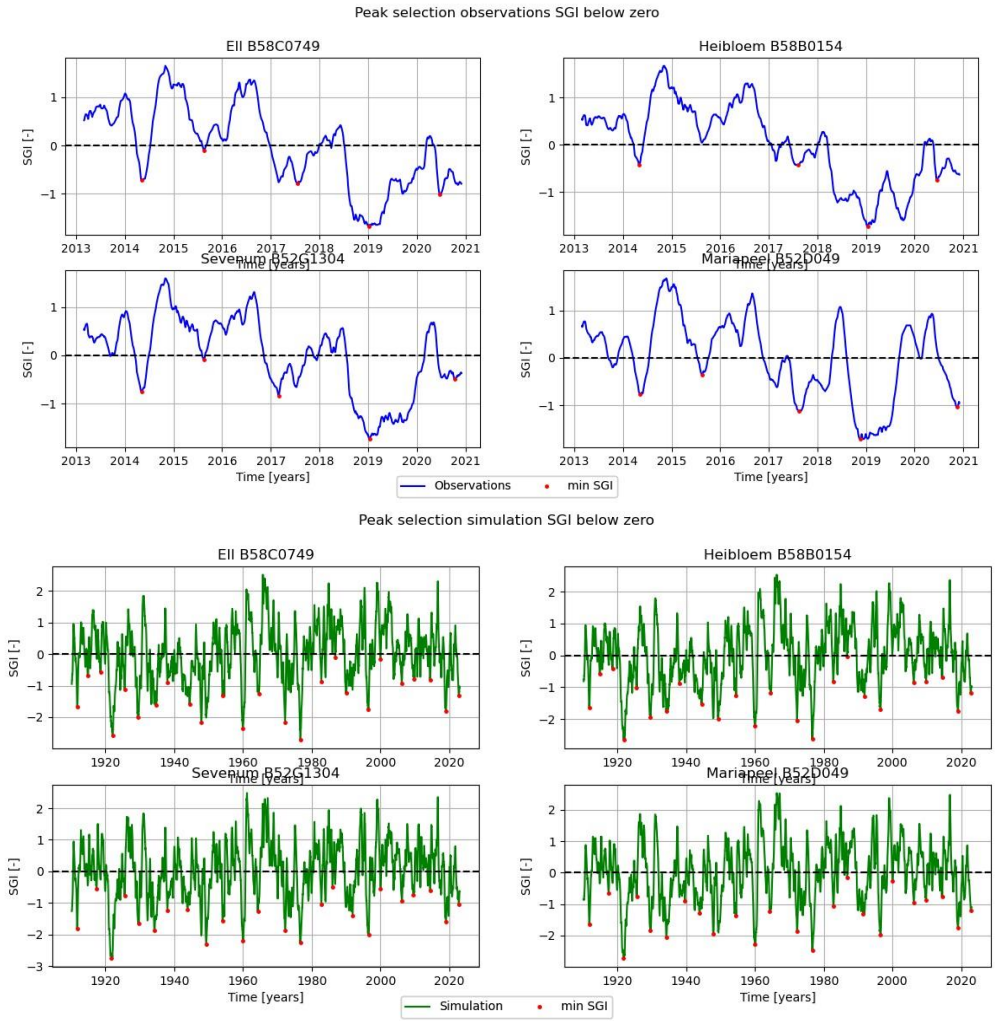


Figure 51: Peak selection for SGI below zero for observations and simulations.

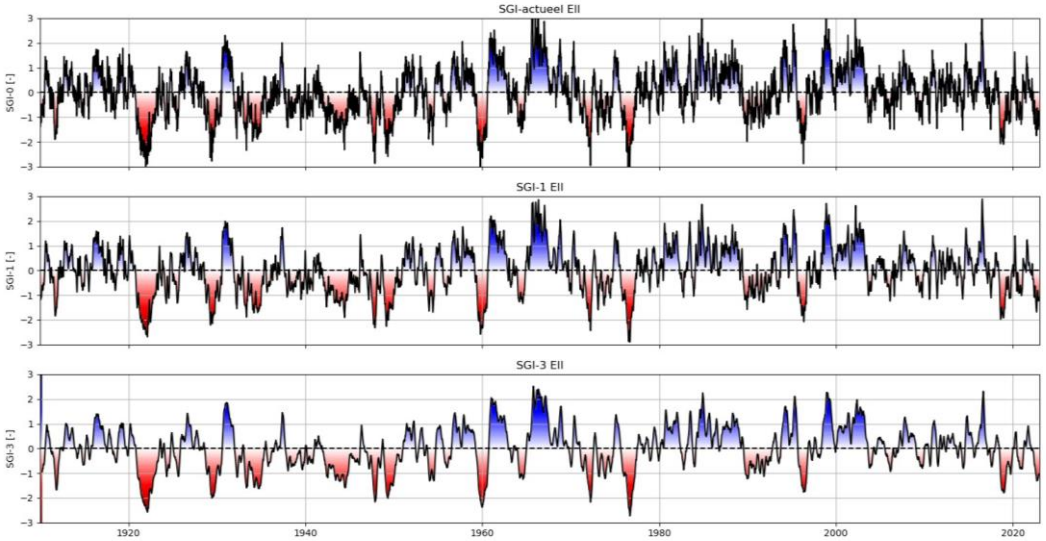


Figure 52: Difference between SGI-0, SGI-1 and SGI-3 for monitoring location Ell.

ANNUAL MINIMA EXTREME VALUE ANALYSIS

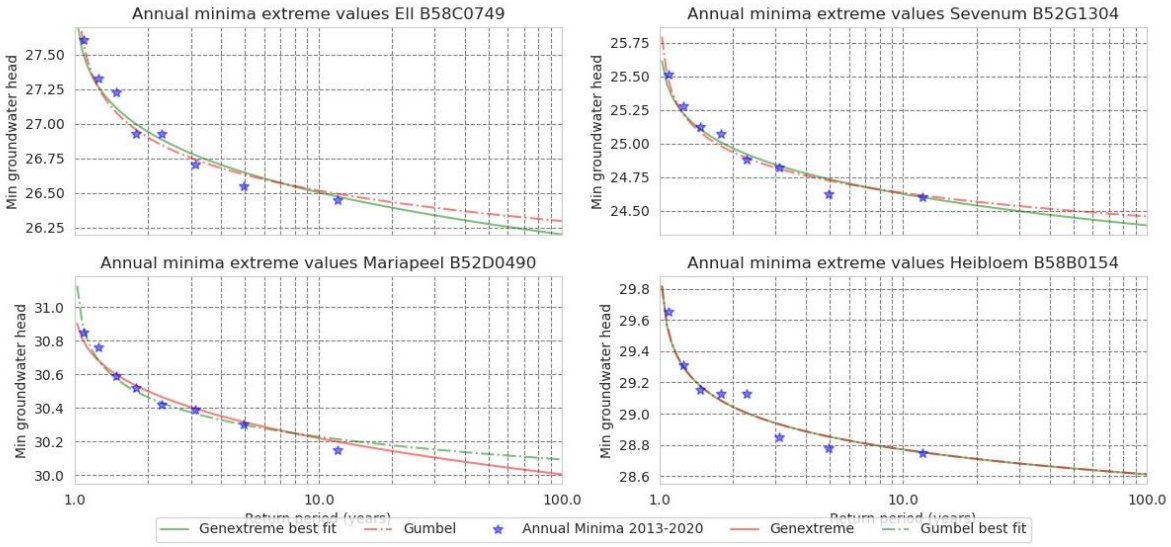


Figure 53: Plot of return periods for annual minima extreme values for both distributions.

Table 17: Estimated return periods fitted extreme value distributions.

T	P	Annual Min Ell [m + NAP]	Annual Min Heibloem [m + NAP]	Annual Min Sevenum [m + NAP]	Annual Min Mariapeel [m + NAP]
2	0.5	26.94	29.04	24.96	30.46
5	0.2	26.65	28.85	24.73	30.30
10	0.1	26.51	28.77	24.62	30.23
25	0.04	26.36	28.69	24.52	30.16
50	0.02	26.28	28.65	24.45	30.12
100	0.01	26.20	28.61	24.39	30.09

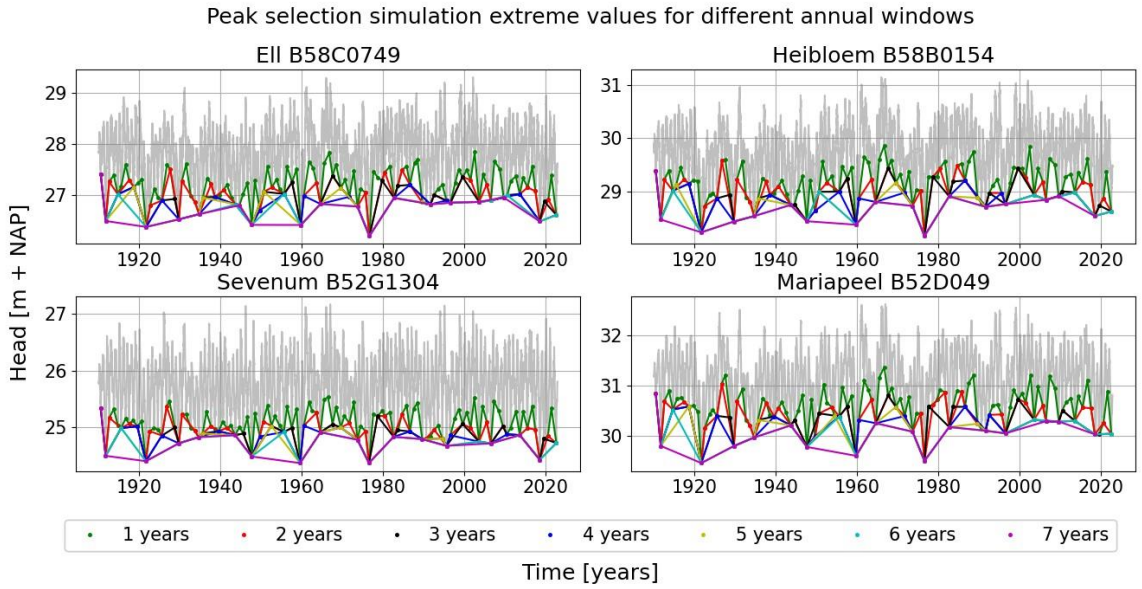


Figure 54: Peak selection extreme value analysis for different annual windows.

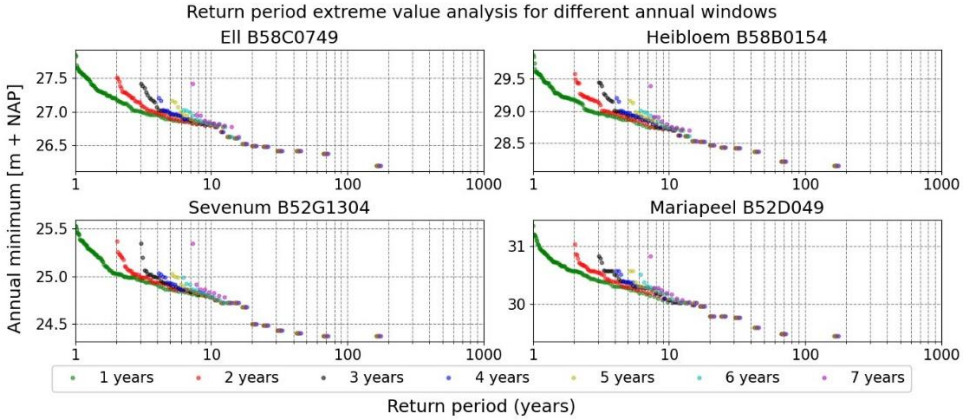


Figure 55: Sensitivity analysis of estimating drought frequency (peak selection in Figure 54)

APPENDIX D: ADDITIONAL RQ2: VALIDATION OF METEOROLOGICAL SERIES

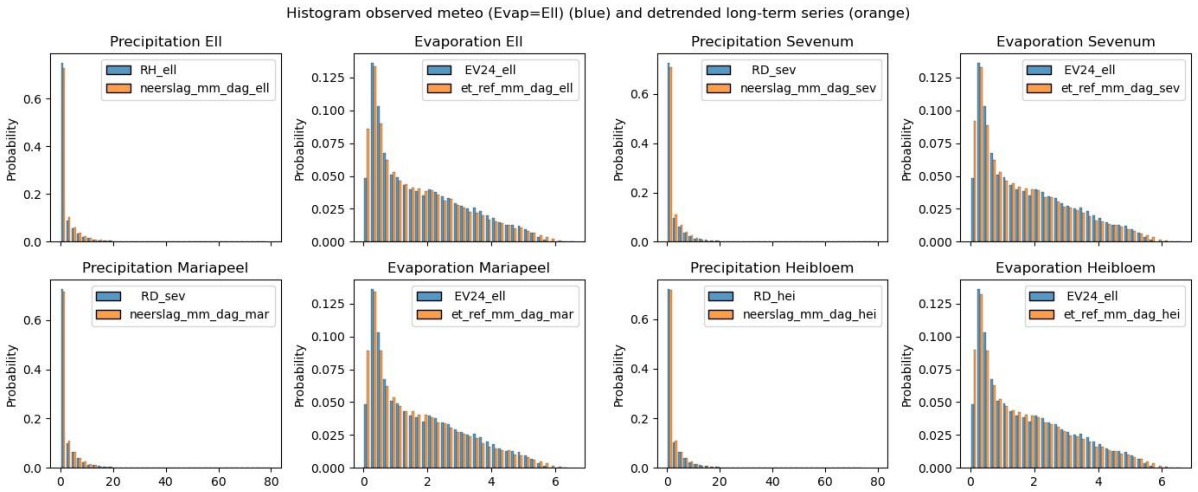


Figure 56: Histogram of observed and historical meteorological data; bins of 2 [mm] for precipitation and 0.2 [mm] for evaporation.

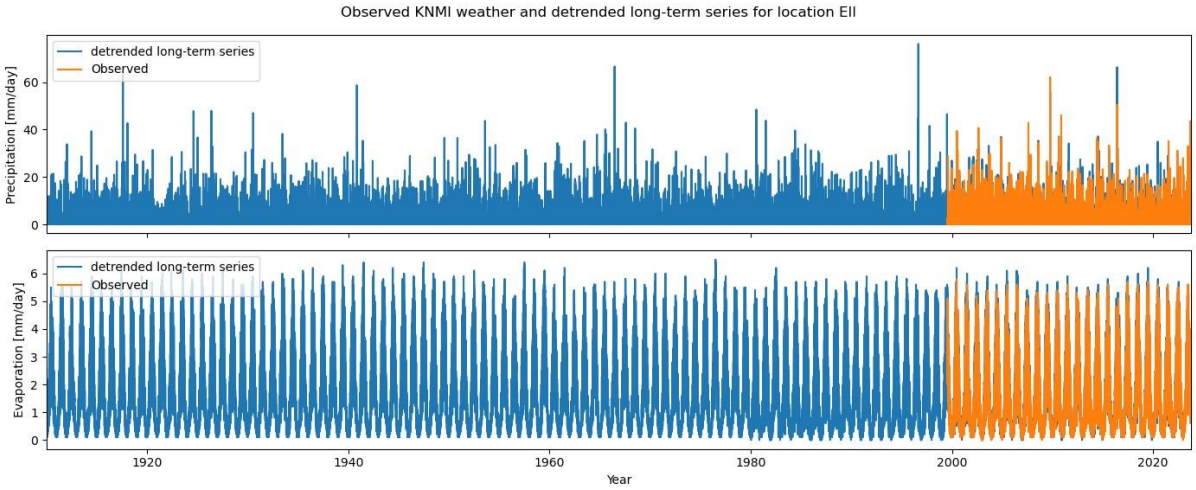


Figure 57: Observed KNMI weather and long-term series for location Ell for the period 1910-2023.

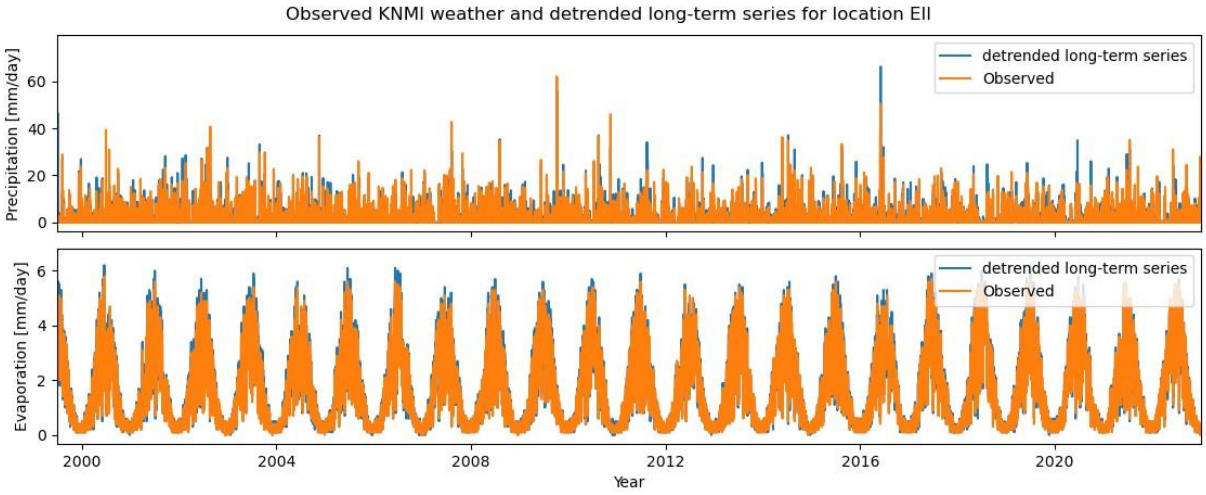


Figure 58: Observed KNMI weather and long-term series for location Ell for the period 1999-2023; long-term series match the observation series.

VALIDATION OF HISTORICAL METEOROLOGICAL DATA BEFORE CORRECTION

The climatological properties are represented in the cumulative density functions. The cumulative density functions (CDF) of the observed and historical climate are shown in Figure 59. The CDF describe the climate for both meteorological datasets (section 3.3), meaning the average and deviations of precipitation and evaporation over a significant period of time. As can be seen in the figures, the CDF's look very similar, indicating that the climate of the historical series represents current climate. Only at the most extreme value of the tail, the historical precipitation and evaporation overestimate the observed precipitation and evaporation. However, since groundwater has a maximum infiltration rate (until the ground is saturated), is it expected that this has minimal impact.

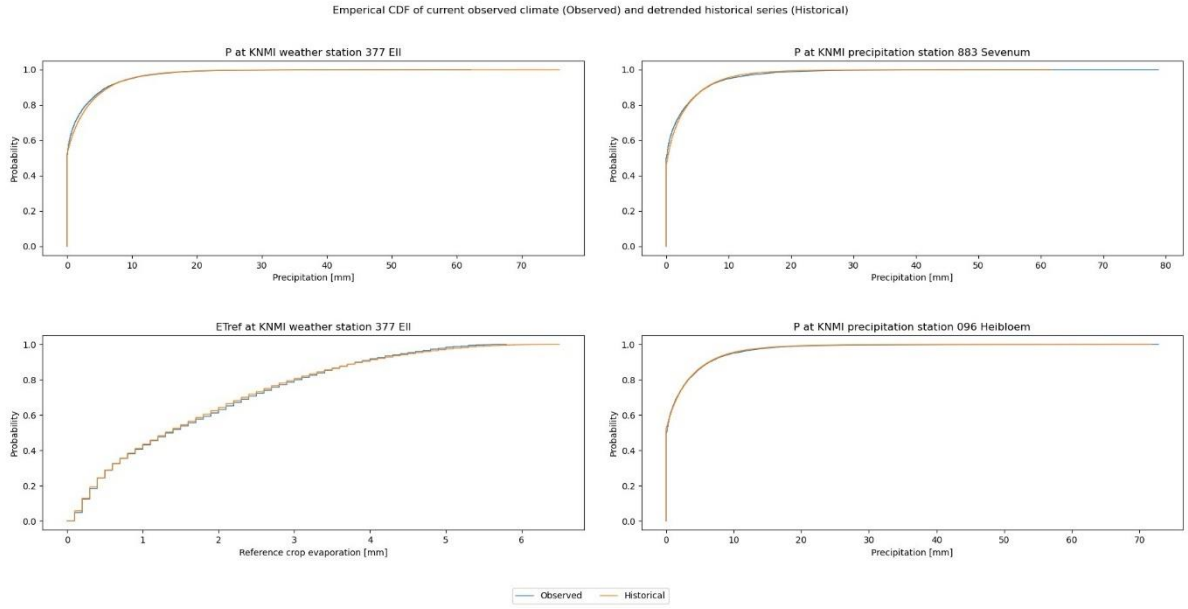


Figure 59: Empirical CDF of historical and observed meteorological data.

The historical precipitation at KNMI locations Heibloem and Sevenum did not have the same distribution as the current precipitation at these locations, according to the KS-test and visual inspection (Table 18, Figure 60). High precipitation events are underestimated in the historical precipitation compared to observed precipitation. Therefore, the bias correction method of quantile mapping is performed to corrected for deviation. The historical precipitation and reference crop evaporation for KNMI station Ell was correct in the historical meteorological data series. This is as predicted, since that KNMI station was used to compile the dataset, while the locations of Sevenum and Heibloem did not. Therefore, the precipitation at these locations were spatially interpolated which meant that the precipitation turns out lower than current climate.

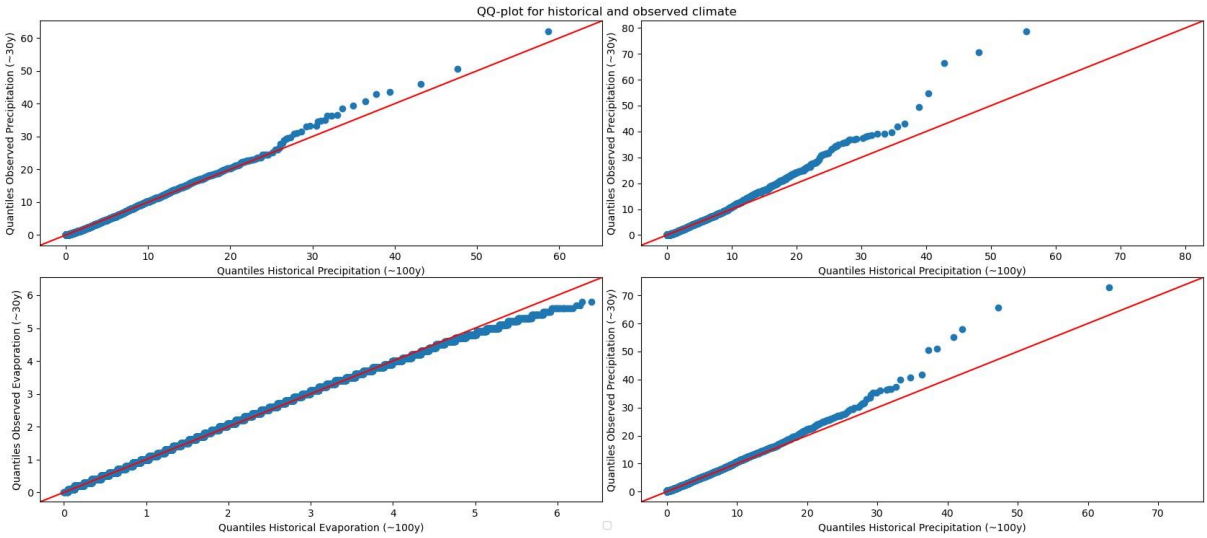
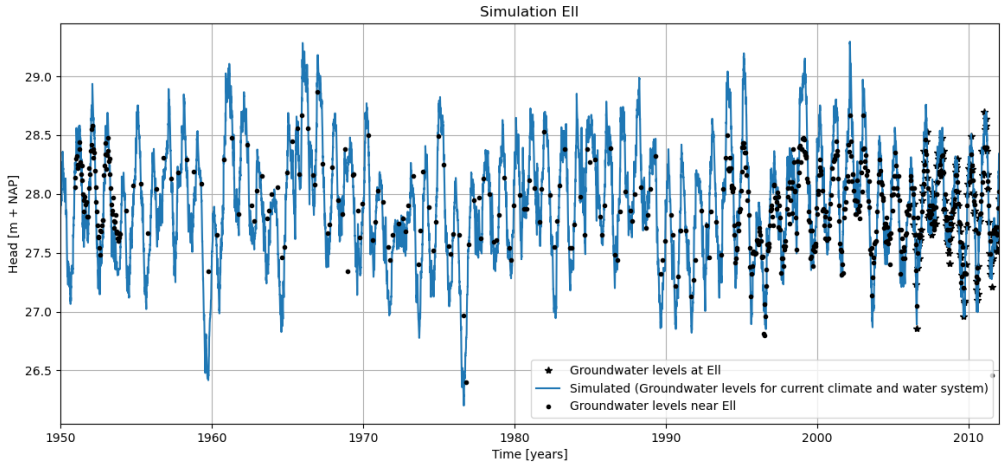


Figure 60: QQ-plot for historical and observed climate at the same location before correction.

Table 18: Kolmogorov-Smirnov test for historical meteorological data before correction.

Test	Ell		Heibloem	Sevenum	**Criteria**
	P	ET	P	P	
K-S	0.052	0.04	0.18	0.261	
P-value	0.13	0.99	1.44e-14	2.43e-30	**> 0.05**

APPENDIX E: ADDITIONAL RQ3: SIMULATION OF GROUNDWATER LEVELS



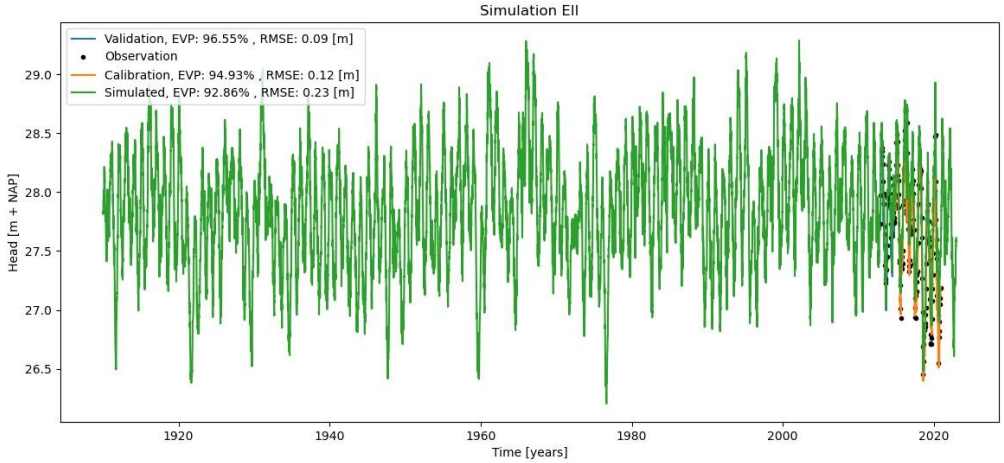


Table 19: Specific for only groundwater levels in the summer.

Model Performance	Ell	Heibloem	Mariapeel	Sevenum	Criteria
Calibration (2014-2020)					
EVP [%]	93.3	90.4	87.7	92	> 70%
RMSE [m]	0.14	0.16	0.19	0.15	<
Validation (2012-2014)					
EVP [%]	97.4	73.1	88.8	94.5	> 70%
RMSE [m]	0.09	0.38	0.13	0.17	<

CALIBRATION IN SUMMER 2018

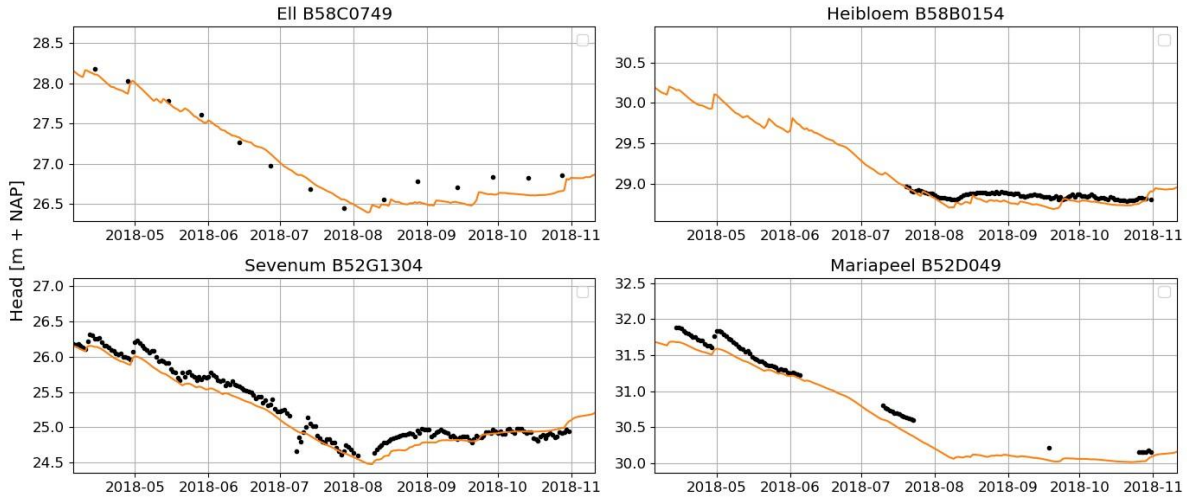


Figure 61: Measurements summer 2018 used in calibration.

VALIDATION OF LONG-TERM GROUNDWATER SIMULATIONS

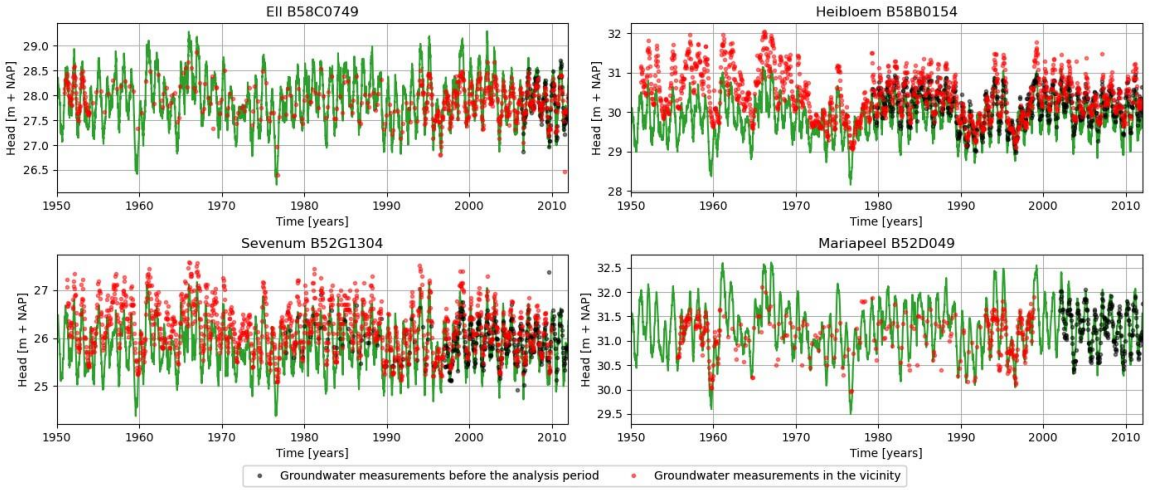


Figure 62: Validation of simulated groundwater levels using groundwater observation outside model calibration period and monitoring stations in the vicinity of the locations.

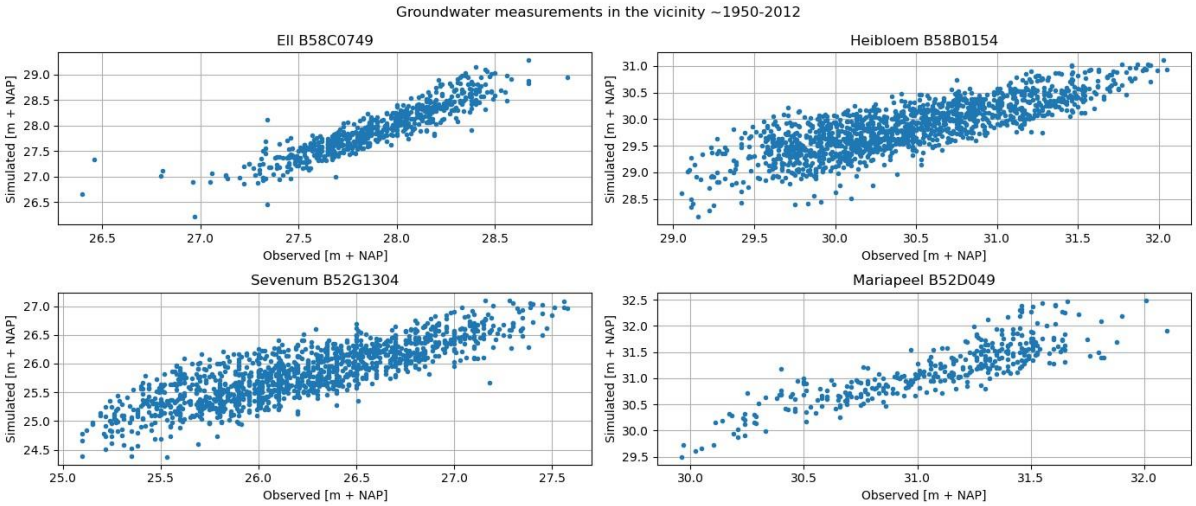


Figure 63: Observed groundwater measurements from monitoring wells in the vicinity alongside groundwater simulations between the period ~1950-2012.

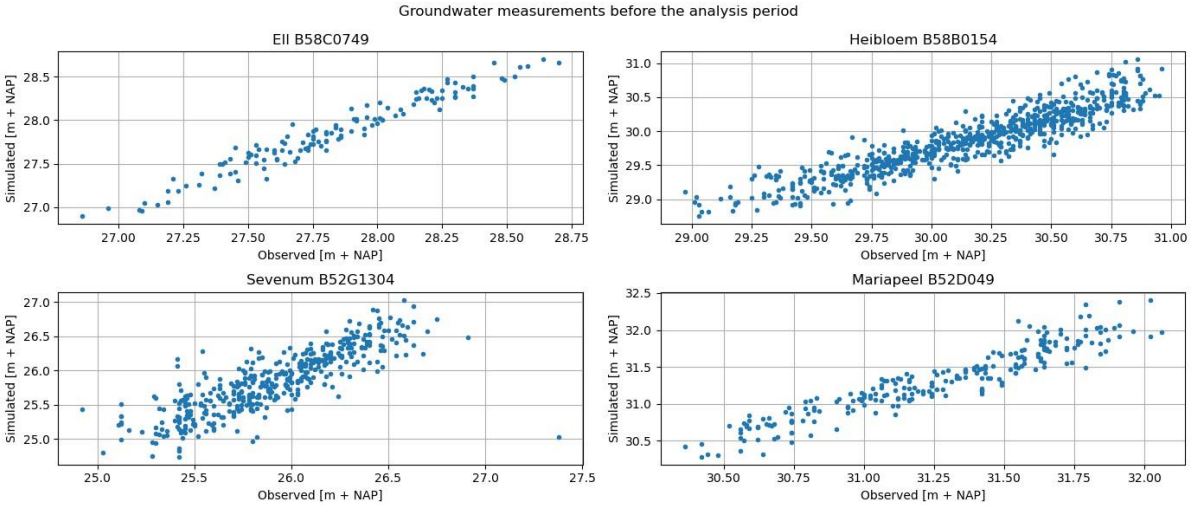


Figure 64: Observed groundwater measurements before the analysis period alongside groundwater simulations between the period 2000-2012.

USE OF DIFFERENT STRESS MODELS

Table 20: Results of model performance using different stress models. Modelling recharge as one input stress of precipitation minus evaporation provides the best results.

Calibration period 2014-2020; Validation period 2012-2014

StressModel:		Goodness of fit:	Ell	Heibloem	Sevenum	Mariapeel
Precipitation and evaporation separate	Calibration	EVP	92.22%	73.61%	83.32%	78.32%
		RMSE	0.14	0.23	0.18	0.2
	Validation	EVP	93.57%	36.08%	61.95%	47.51%
		RMSE	0.08	0.38	0.24	0.24
Recharge (precipitation - evaporation)	Calibration	EVP	94.9%	88.6%	90.3%	92.6%
		RMSE	0.12	0.16	0.16	0.14
	Validation	EVP	96.1%	76.6%	90.3%	89.5%
		RMSE	0.09	0.30	0.16	0.11
FlexModel	Calibration	EVP	92.50%	0.00%	83.10%	86.70%
		RMSE	0.14	0.33	0.18	0.17
	Validation	EVP	93.70%	0.00%	80.47%	89.62%
		RMSE	0.10	0.44	0.17	0.14
Linear	Calibration	EVP	94.61%	86.31%	0.00%	90.82%
		RMSE	0.12	0.18	0.34	0.16
	Validation	EVP	96.22%	80.19%	44.16%	89.89%
		RMSE	0.08	0.32	0.26	0.11
Berendrecht	Calibration	EVP	91.10%	0.00%	0.00%	72.40%
		RMSE	0.15	0.33	0.32	0.2
	Validation	EVP	87.47%	0.00%	0.00%	71.54%
		RMSE	0.16	0.43	0.25	0.17
Peterson	Calibration	EVP	94.53%	83.80%	93.71%	76.65%
		RMSE	0.12	0.19	0.12	0.19
	Validation	EVP	95.12%	56.82%	87.09%	66.43%
		RMSE	0.08	0.26	0.12	0.17

GROUNDWATER MEASUREMENTS

The pre-processed groundwater observations are shown in Figure 65. It is chosen for a period of eight years, since changes in water management can often induce changes in the groundwater regime. Certain climatological events occur consistently across all groundwater observation series, notably the annual groundwater fluctuation pattern. Higher groundwater levels are observed during winter months, due to increased precipitation, while lower levels are observed in summer months. Notably, in the years 2018 and 2019, particularly in Ell and Sevenum, groundwater levels decrease in the summer.

As seen in Figure 65, some measurements expose unrealistic changes to the groundwater head within one observation time step. These measurements are marked as measurements errors since they cause unreliability in further analysis. Therefore, these measurements are removed from the dataset. Whether a measurement is an error is based on visual inspection of the measurement series and comparing outliers with other locations and measured precipitation. The found measurement

errors and a few examples are given in the appendix. The gaps due to the removed errors are not interpolated, since the measurements series already contains data gaps which are too large to interpolate.

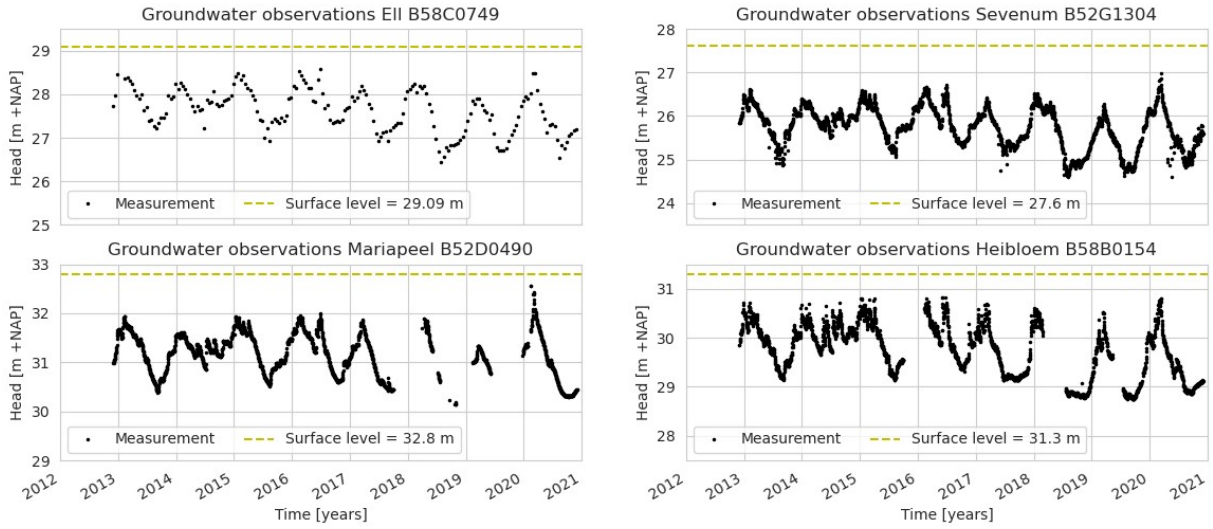


Figure 65: Pre-processed groundwater observations for the monitoring wells.

Distinctive characteristics occur in the annual pattern of the different locations. In Ell and Mariapeel, the peaks display a more rounded profile, whereas in Sevenum and Heibloem, the curves appear more irregular. This deviation is likely influenced by the difference in land uses surrounding the monitoring wells. Ell and Mariapeel are close to natural areas, where precipitation infiltrates quicker into the ground compared to agricultural areas, such as Sevenum and Heibloem (Robinson et al., 2022).

ADDITIONAL HISTORICAL METEOROLOGICAL SERIES

Next to historical meteorological data, there is also synthetic meteorological data. Synthetic meteorological time series refers to artificially generated weather datasets that resemble real-world meteorological data over a specific time period. These synthetic datasets can be created using statistical models or numerical simulations and are designed to capture the statistical properties and dynamics of meteorological variables. While synthetic meteorological datasets can provide valuable insights when real data is unavailable, they might not capture all complexities and nuances of the actual meteorological processes.

Regional Atmospheric Climate Model (RACMO) is a numerical climate model used to simulate the climate and atmospheric processes over a specific area. RACMO is designed to provide detailed information about various climatological variables such as temperature, precipitation and radiation. RACMO is a well-known high resolution climatological model developed by KNMI (Van Meijgaard et al., 2008), based on hydrostatic dynamics originating from a numerical weather forecasting system (HIRLAM) and physics of the European Centre of Medium Range Weather Forecast (ECMWF) (Lenderink et al., 2003).

GRADE is model developed for the KNMI and generates rainfall and discharge extremes for the Rhine and Meuse rivers under current and future climate conditions (Hegnauer et al., 2014). GRADE essentially is a modelling tool that consists of three components: a stochastic weather generator, a rainfall-runoff (hydrological) model and a hydraulic model. The stochastic weather generator is the climate model that could be used to generate synthetic meteorological time series. GRADE can also be combined with RACMO(Van Voorst & Van Den Brink, 2022).

E-OBS comes as an ensemble dataset and is available on a 0.1- and 0.25-degree regular grid for the period of 1950 to 2022 for Europe (Cornes et al., 2018). The ensemble dataset is constructed through a conditional simulation procedure. For each of the members of the ensemble, a spatially correlated random field is produced using a pre-calculated spatial correlation function. The basis consists of 23000 meteorological stations which are gridded using ordinary kriging (Haylock et al., 2008).