# Spatiotemporal development of the 2018-2019 groundwater drought in the Netherlands: a data-based approach

Esther Brakkee[1], Marjolein van Huijgevoort[1], Ruud P. Bartholomeus[1, 2]

[1]KWR Water Research Institute, Nieuwegein, 3433 PE, the Netherlands
5 [2]Wageningen University, Soil Physics and Land Management Group, Wageningen, the Netherlands

*Correspondence to*: Esther Brakkee (esther.brakkee@kwrwater.nl)

**Abstract.** The 2018-2019 drought in northwestern Europe caused severe damage to a wide range of sectors, and has made clear that even in temperate-climate countries adaptations are needed to cope with increasing future drought frequencies. A crucial component of drought management strategies is to monitor the status of groundwater resources. However, providing

10 up-to-date assessments of regional groundwater drought development remains challenging due to the limited quality of available data. This limits many studies to small selections of groundwater monitoring sites, giving an incomplete image of drought dynamics. In this study, a time series modelling-based method for data preparation was developed and applied to map the spatiotemporal development of the 2018-2019 groundwater drought in the southeastern Netherlands, based on a large set of monitoring data. The data preparation method was evaluated for its usefulness and reliability for groundwater drought

15 quantification and prediction. The analysis showed that the 2018-2019 meteorological drought caused extreme groundwater drought throughout the southeastern Netherlands, breaking 30-year records almost everywhere. Drought onset and duration were strongly variable in space, with especially higher elevated areas remaining in severe drought well into 2020. Groundwater drought development appeared to be governed dominantly by the spatial distribution of rainfall and the geological-topographic setting. The time series modelling-based data preparation method was found a useful tool to enable a detailed, consistent record

20 of regional groundwater drought development. Applying a validation step before analysis turned out to be important for good results. The time series simulations were generally found to be reliable; however, the use of time series simulations rather than direct measurement series can bias drought estimations especially at a local scale, and underestimate spatial variability. Finally, time series modelling showed to be a promising tool for regional-scale drought nowcasting and prediction. Further development of time-series based validation and simulation methods, combined with accessible and consistent monitoring

25 data, will be valuable to enable better groundwater drought monitoring in the future.

## 1 Introduction

In the summer of 2018, a severe drought hit large parts of northern and western Europe. Extremely low precipitation coincided with high temperatures, both breaking multiple-decade records in many places (Bakke et al., 2020; Philip et al., 2020; Toreti et al., 2019). Drying soils and declining water reserves caused damage to agricultural production and natural ecosystems,

30 problems with drinking water and energy production, and widespread forest fires, among other impacts (Bakke et al., 2020;

Bastos et al., 2020; Buras et al., 2020; Philip et al., 2020). The kind of 'hot drought' that occurred in 2018 is expected to become more frequent in the future (Philip et al., 2020; Toreti et al., 2019).

The Netherlands was one of the countries most hit by the weather extremes (Bakke et al., 2020). The damage was felt mainly in the southern and eastern parts of the country (Van de Velde et al., 2019; Van den Eertwegh et al., 2019; Witte et al., 2020b).

35    In a country traditionally more focused on getting rid of water surpluses, the drought of 2018-2019 was felt by many water managers as a wake-up call, sparking a widespread search for solutions to prepare water systems for increasingly frequent drought extremes (De Lenne and Worm, 2020; IenW, 2019; Witte et al., 2020a; Van de Velde et al., 2019). In the southeastern Netherlands, as in many other parts of the world, groundwater is a crucial water source. Accordingly, much of the damage in 2018 was directly related to deep drawdowns in groundwater levels (LCW, 2020). Among other effects, the groundwater

40    shortages caused severe damage in peatland and brook ecosystems (Witte et al., 2020b) and concerns over the sustainability of increased irrigation and drinking water abstractions (Van de Velde et al., 2019; Van den Eertwegh et al., 2019). Groundwater is often the most persistent water store in the landscape, reacting latest as a meteorological drought propagates into the hydrological system (Van Loon, 2015). This makes proper management of the groundwater a crucial component of drought management strategies.

45    Previous studies have shown that the response of groundwater to meteorological drought can vary strongly in space. Variations in groundwater response are caused by differences in geology, water management and other catchment characteristics (Bloomfield et al., 2015; Hellwig et al., 2020; Peters et al., 2006; Van Loon and Laaha, 2015). To be able to mitigate and prevent drought damage, it is therefore essential to understand how groundwater drought develops in both time and space. In recent years, water managers in the Netherlands have indeed expressed a need for more up-to-date, locally-specific drought

50    information and predictions to be able to take appropriate measures (IenW, 2019; Pezij et al., 2019; Witte et al., 2020a). Not only the period of meteorological drought itself, but also the recovery of the system after meteorological drought is important to understand.

Multiple recent research efforts have aimed at better understanding the variations in groundwater drought and its impacts at national and European scales (Bakke et al., 2020; Bloomfield et al., 2018; Hellwig et al., 2020; Margariti et al., 2019; Van

55    Loon et al., 2017). The 2018 drought in Europe has so far been mainly studied from a meteorological perspective (Bakke et al., 2020; Philip et al., 2020; Toreti et al., 2019). A large-scale analysis of the hydrological drought has recently been made for Scandinavia (Bakke et al., 2020); while for the Netherlands, some assessments of the drought in the groundwater have been done based on small numbers of measurement sites and physically-based modelling studies (Van den Eertwegh et al., 2019). What is still lacking, is a more detailed image of how the 2018-2019 drought manifested itself in the groundwater and how

60    this varied in space, based on measurement data. This could provide valuable insights into groundwater drought dynamics and mitigation options in the Netherlands and elsewhere.

Groundwater heads are widely monitored in measurement wells. However, analysis of groundwater drought from these data over large areas is often challenged by data quality. Firstly, data usually have to be obtained from multiple organisations and contain errors and other perturbations. Secondly, the length of measured time series is often not sufficient for drought analysis,

65  for which at least 30-year series are recommended (Link et al., 2020). As a result, many data-based groundwater drought
studies have focused on relatively few measurement wells with near-natural, long series (Bakke et al., 2020; Van Loon et al.,
2017); this may give an incomplete image of the true variability in drought dynamics. In addition, available data usually lags
behind the present, hindering the up-to-date drought assessments that water managers need.

To deal with these challenges, several studies have developed methods for automated validation and lengthening of

70  groundwater head time series (Marchant and Bloomfield, 2018; Peterson et al., 2018; von Asmuth et al., 2012). This is usually
done with statistical models, ranging from regression analysis to artificial intelligence methods (Peterson et al., 2018; Van
Loon et al., 2017). One type of statistical modelling that has proven very useful for groundwater data is time series modelling
with impulse-response functions (Bakker and Schaars, 2019; von Asmuth et al., 2002). These models describe groundwater
head variations at a specific location as a function of driving variables, usually weather data, and a fitted impulse-response

75  function. This type of impulse-response time series modelling (TSM) allows accurate simulations to be made without a need
for information on site characteristics. The simulations can be used to identify errors and other atypical behaviour in the data,
and to lengthen and harmonise time series, as shown by e.g. Zaadnoordijk et al. (2019), Bartholomeus et al. (2008) and
Marchant and Bloomfield (2018). Marchant and Bloomfield were the first to develop a full time series model-based method
to study groundwater drought over a large region in the UK. Although TSM-based analyses appear a valuable tool for

80  groundwater drought assessment, their wider applicability for various cases has not yet been well explored. To be able to
widely use TSM data preparation for drought studies, three main questions need to be answered.

Firstly, it is not yet clear what methods are optimal for data validation. Time series modelling can be used to identify time
series influenced by disturbances after analysis (e.g. Marchant & Bloomfield, 2018), but also to remove irregularities from the
data beforehand. In addition, TSM-based data cleaning can be combined with other more basic consistency checks, improving

85  its effectiveness (Peterson et al., 2018). This validation method has not yet been evaluated specifically for the case of
groundwater drought analysis.

Secondly, the reliability of time series simulations for groundwater drought analysis has not been properly tested. To provide
lengthened, gap-filled groundwater head series that are suitable for analysis, researchers have often used TSM simulations
directly as replacement of the data. These simulations often have a very good fit to the data (Bakker and Schaars, 2019;

90  Zaadnoordijk et al., 2019); however, this approach inevitably also strips out part of the external influences that are not explicitly
included in the model drivers (Peterson et al., 2018; Zaadnoordijk et al., 2019). Drought occurrence and development can be
strongly affected by human impacts, as well as by local-scale natural influences such as surface water influence (Margariti et
al., 2019; Van den Eertwegh et al., 2019). As such, excluding such external drivers of groundwater levels may provide an
incomplete image of drought dynamics (Van Loon et al., 2016). It is therefore important to understand how the use of TSM

95  simulations rather than measurement series affects the assessment of drought behaviour.

Finally, the usefulness of TSM for nowcasting and predicting groundwater drought needs to be better understood. Recently,
there has been a surging interest in large-scale drought forecasting, which has shown that physically-based groundwater models
can predict with some skill up to two months ahead (Honingh et al., 2020; Mackay et al., 2015; Sutanto et al., 2020; Wanders

et al., 2019). Time series modelling could provide a very easy alternative to nowcast and predict groundwater drought requiring
100 nothing more than weather data (Bakker and Schaars, 2019; Marchant and Bloomfield, 2018). However, prediction skill may
depend on the location and the delay of the groundwater system (Mackay et al., 2015; Marchant and Bloomfield, 2018). Also,
models may have difficulties to correctly represent groundwater behaviour during abnormal drought conditions (Hellwig et
al., 2020). It is therefore important to better understand how accurate time series models actually are in nowcasting and
forecasting groundwater drought.

105 Given these knowledge gaps, the current study aims to evaluate the usefulness of a time series modelling-based data preparation
method for regional analysis of groundwater drought. The method will be applied to the southeastern Netherlands to
characterise the development in time and space of the 2018-2019 groundwater drought and its recovery. The method will be
evaluated on the usefulness for series validation and simulation, the effect of simulations on drought assessment, and the
potential for drought prediction.

110 The applied method for data validation and drought assessment, as well as the tests used for its evaluation, are set out in section
2 and 3. The performance of the method and the resulting drought analysis are presented in section 4. In section 5, the resulting
insights into the use of time series modelling-based data methods for groundwater drought studies are discussed, followed by
the spatiotemporal drought development of 2018-2019. Final conclusions are given in section 6.

## 2 Study area and data

115 The study area covers six provinces in the southeastern part of the Netherlands (Fig. 1). This area coincides with that part of
the Netherlands that lies above sea level and is dominated by Pleistocene-era deposits. Most of the study area has sandy
sediments at the surface, but part is also covered by river clays and loess deposits (Fig. 1b). Elevation is mostly between 0 and
30 m AMSL (Fig. 1a). Higher elevations occur in the limestone-loess hill landscape of southern Limburg and on glacier-
pushed ridges in Utrecht and Gelderland (areas indicated in Fig. 1a). Land use is dominated by agriculture, while the glacial
120 ridges are covered mainly with forest. The area has a temperate climate with a yearly precipitation surplus. In addition, the
groundwater system is affected by abstractions for drinking water and irrigation, as well as by drainage systems in the lowest-
lying parts of the study area.

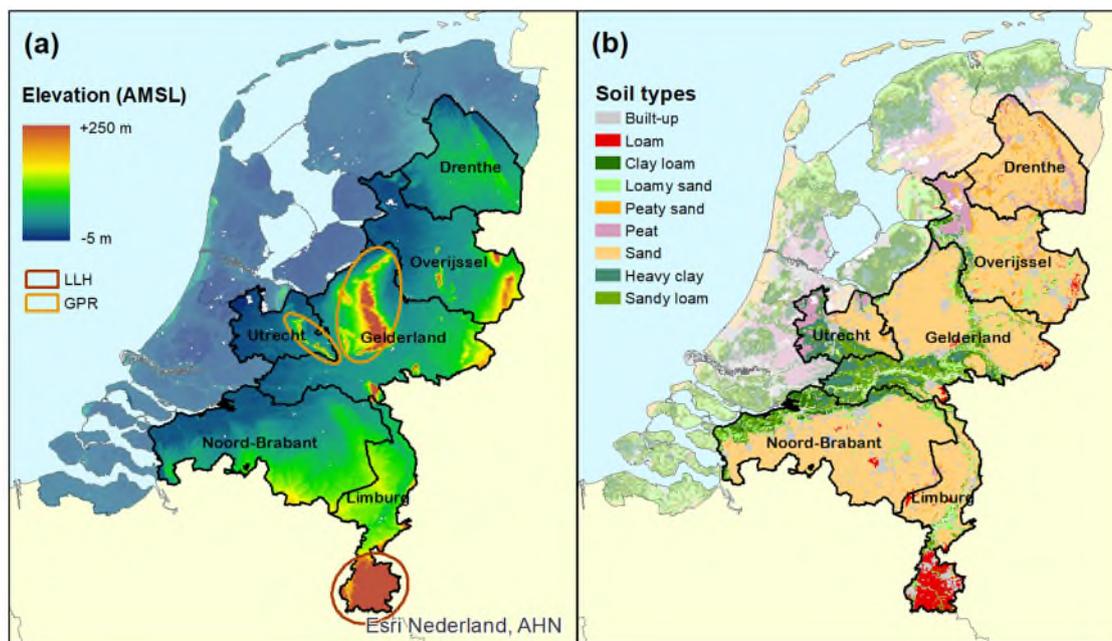



**Figure 1: Study area. a: Elevation (AHN, 2019); b: Soil types (WUR, 2006). The higher elevated limestone-loess hill landscape in**
125 **southern Limburg (LLH) and glacier-pushed sand ridges (GPR) are indicated.**

### 2.2 Input data

Groundwater head data were supplied by several regional water managing bodies. For some areas, additional series were
obtained from the Dutch national groundwater database DINO (TNO). The data consist of groundwater head time series with
a twice-monthly to sub-daily frequency, mostly running until spring 2019. In addition, metadata of the monitoring wells were
130 available, including location, filter depth and surface level. The different data sets were checked for overlapping wells by well
names and coordinates, and overlapping series were combined. Only data from the first filter of boreholes was used, generally
representing the phreatic level. Data were cut off at 1 January 1990 and those series were selected that contained > 10 years of
data and ended after 31 August 2018. This resulted in 2722 series for further analysis.

Daily precipitation and reference evapotranspiration ($ET_{ref}$) were obtained from the Royal Netherlands Meteorological Institute
135 for January 1990 to May 2020 (KNMI, 2020). Data were used from 15 general weather stations (for $ET_{ref}$) and 114 precipitation
stations over the study area. Reference evapotranspiration is determined by KNMI following Makkink (1957). The $ET_{ref}$ series
did not always cover the full period of interest; gaps were filled with the nearest station that did have full data.

### 3 Methods

The method for data preparation and drought analysis consists of three components (Fig. 2): 1) validation of the observed
140 groundwater heads, 2) simulation of groundwater heads, and 3) conversion to a standardised groundwater index (SGI). The

last two steps were also performed for a prediction of the groundwater heads. Finally, each step was evaluated by one or more tests.

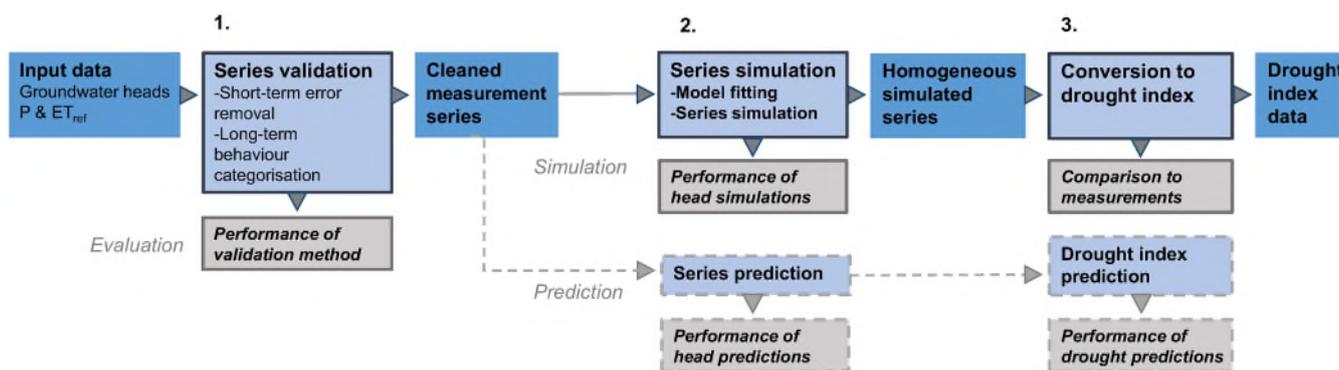



**Figure 2: Method workflow.**

145    **3.1 Series validation**

Data validation in general is a step to check initial data for errors and other deviating behaviour. What validation method is suitable depends on the type of deviations in the data and which of them need to be removed for the intended analysis.

Phreatic groundwater levels typically follow a yearly cycle, overlain by faster fluctuations in response to rainfall and evapotranspiration. An actual series of measured groundwater heads will often show deviations from this expected pattern.

150    Deviations may be of short duration, visible as loose outliers; or they may be structural, shown as level shifts, trends and other abnormal patterns in the data series. Both the short-term and structural deviations can result from errors, such as instrument drift and typing errors. However, deviations can also reflect real groundwater behaviour caused by local natural or human influences, such as rainstorms, flooding or abstraction (Post and von Asmuth, 2013; Zaadnoordijk et al., 2019; Margariti et al., 2019). For one subregion of the study area (Overijssel province, see Fig. 1) log book notes from the data collectors were

155    available; inspection of these showed that the data were influenced by short-term extractions, changes in water management, sensor problems, relocation of wells, well maintenance and other issues. Also Zaadnoordijk et al. (2019) found that probably a majority of series in Dutch groundwater data sets contain such disturbances.

This kind of groundwater head data clearly require a validation step to become suitable for drought analysis. This step needs to remove the errors as much as possible; also real but short-term natural and human disturbances in the groundwater heads

160    are less relevant for a large-scale drought analysis. However, the long-term deviations in the groundwater level, such as caused by long-term abstraction and land use effects, are important for drought studies and should ideally be retained.

Whether atypical behaviour in a series is caused by errors or by real external influences is very difficult to distinguish by automated methods. Our approach is therefore to filter out all short-term deviations and then classify the series according to their long-term behaviour. The short-term deviations will be denoted as outliers: one or a few measurements lying far from the

165    expected level. The use of categories for long-term behaviour allows for retaining some of the potentially influenced series in

the analysis, while acknowledging their lower reliability. The validation itself was done by a combination of time series modelling and simple tests, as explained below.

Time series modelling was done with the PASTAS function package for Python. This package has been developed by Collenteur et al. (2019) to allow impulse-response time series models for groundwater heads to be easily applied and combined into larger workflows. Here the most basic settings of PASTAS were used (see Collenteur et al., 2019). Groundwater heads were modelled as a function of measured precipitation and $ET_{ref}$ series, and a gamma-shaped recharge response function. Recharge itself was modelled as a linear combination of precipitation and $ET_{ref}$. PASTAS also enables nonlinear recharge modelling; we tested the nonlinear model for a subset of series in Brabant, but model fit was improved in only a small minority of series, and so linear recharge modelling was applied for the rest of the study.

Data cleaning was done on the original, raw series in several steps:

1. Removing measurements below the well filter or > 20 cm above the surface level;

2. Removing far outliers. The upper and lower 20 % of the range of a series were identified as potential outliers; if leaving out these measurements reduced the range by more than half, they were classified as outliers and removed.

3. Outlier removal with TSM. A model was fit for each series using precipitation and $ET_{ref}$ from the nearest weather stations and the basic model settings of PASTAS. Fitting was done at a weekly time step to prevent bias towards recent years with more frequent measurements. The simulation interval was calculated as the standard deviation of the residuals, and all measurements outside a range of 4 times this interval around the simulation were removed. This step was repeated a second time to remove smaller outliers.

For the long-term behaviour classification, a new time series model was fitted on the resulting cleaned series. The explained variance percentage (EVP, equal to $r^2 \cdot 100$) and model parameters $A$, $f$ and $d$ were saved. Parameter $A$ represents the long-term response of the groundwater level to a constant recharge input of one unit, in this case 1 mm; $f$ is the influence of $ET_{ref}$ relative to precipitation; and $d$ is the groundwater base level. In addition, a linear trend was fitted through the residual series and the p-value and $r^2$ of the trend were saved. Finally, the series were checked for periods with missing data of > 4 years and, if these were present, only the time period after the last such data gap was used. This check had to be done after data cleaning to ensure all data gaps are recognised. Based on these indicators the series were ordered into four categories:

1. Discarded series: EVP < 60 %, less than 10 years of data, data gaps of more than 4 years, or no data over June-August 2018;

2. Deep-groundwater series: mean water table depth (WTD) > 5 m. These series typically showed a very slow, smoothed behaviour and often poor model fit. They were saved as a separate category to reflect their distinct behaviour and potentially less reliable results;

3. Atypical series: EVP >= 60 %, but containing a trend or atypical parameters. Locations were marked as atypical if they had a trend with p<0.05 and $r^2$ >0.15; an $f$ parameter > -0.05 or < -1.95, very close to the parameter bounds; or an $A$ parameter > 1.5, far from its normal range for the used dataset. These boundaries were chosen by trial and error

testing on a subset of the series, where they were found to properly separate those series with visually somewhat
200     atypical, but not extremely disturbed behaviour.

    4.   Normal series: EVP >= 60 %, no other issues.

The cleaned measurement series of the normal, atypical and deep categories were aggregated to a minimum daily time step
and saved for further use.

The described method for data cleaning and categorising was tested on a representative subset of the data. Around 120 series
205 were randomly taken spread over the different provinces; from these 56 test series were selected: 44 with visual errors or
abnormal behaviour and 12 without. The test series were first visually checked for both short-term outliers and longer-term
abnormal patterns. Then the automated validation results were compared to this reference. The method was scored for its
ability to 1) correctly remove outliers; 2) correctly categorise abnormal longer-term behaviour; and 3) overall, result in a
properly treated (or duly discarded) series. The first two aspects were scored as True or False Positive (deviations correctly
210 recognised) and True or False Negative (absence of deviations correctly recognised). The overall performance was scored as
*Good* if outliers and patterns were properly cleaned and categorised; *Reasonable* if some errors remained or were over-
removed, but without significant influence on the series as a whole; and *Bad* if remaining errors or over-filtering had a
significant influence on the character of the series. Groundwater series with mean groundwater depth > 5 m ('Deep') were
retained as a separate category as the tested validation method was found unsuitable for these series.

215 **3.2 Series simulation**

The cleaned series from the validation step were used to simulate groundwater heads for the drought analysis. It is also possible
to interpolate the original measurement series with TSM (Marchant and Bloomfield, 2018). However, it was chosen to use
simulated series to ensure regular series without sudden level shifts and prevent influence of remaining outliers. First, a
PASTAS model was fit on the cleaned measurement series and the model parameters were saved. Next, for those models with
220 an EVP > 60 %, the fitted model parameters and weather data were taken to simulate daily-step groundwater heads for the
period of interest, in this case 1 January 1990 to 31 May 2020. This setup makes it easy to extend the simulations when new
weather data comes available, without re-doing the model fitting procedure. As most measurement series originally ran until
spring 2019, roughly one year of 'nowcasting' was added to the data.

The performance of the simulations was assessed by the root mean squared error (RMSE) and the mean error (ME) of the
225 simulated groundwater heads over the whole length of the measurement series. The mean error was calculated as
*mean(simulated-observed)* to quantify bias. It is possible that the model performance under dry conditions differs from the
overall performance. Therefore the measurements of each series were subdivided into 'low-head periods' (measured head <
20th percentile), 'medium-head periods' (20th to 80th percentile) and 'high-head periods' (> 80th percentile), and the RMSE
and ME were re-calculated for these periods. Finally, RMSE and ME were determined specifically over July-November 2018,
230 the period in which the meteorological drought was most severe. All performance measures were expressed both as the absolute

value in meters, and as fraction of the mean water table depth (WTD) at the location, as the practical consequences of given error in groundwater head are much larger where groundwater levels are normally shallow.

## 3.3 Drought index

To identify drought periods in time series of hydrological variables, and to compare drought severity between locations,
235 standardised drought indices are used. This concept was first developed for precipitation data with the Standardised Precipitation Index (SPI) (McKee et al., 1993). Later, equivalent indices were developed for other hydrological variables. We quantified the development of meteorological drought over 2018-2020 by the three-month-aggregated Standardised Precipitation Evaporation Index (SPEI) (Vicente-Serrano et al., 2010). The study area was divided into four zones (west, north, mid-east, southeast); the SPEI-3 for each zone was calculated from the average precipitation of all weather stations within the
240 zone and the distance-weighted mean reference evapotranspiration of the three stations closest to the midpoint.

For groundwater heads the Standardised Groundwater Index (SGI) was developed (Bloomfield and Marchant, 2013). To calculate the index values, the measured groundwater heads at a specific location are transformed to a standard normal distribution. This gives a drought index series varying roughly between -3 and 3, indicating conditions from extremely dry to extremely wet compared to the normal situation at the specific location. When analysing multiple locations, a common
245 reference period must be used, recommended to be at least 30 years. Here, the period January 1990 - December 2019 is used throughout as the reference period.

For precipitation, the transformation step is usually done by fitting a gamma distribution function to the data. For groundwater heads, however, distribution shapes vary widely between locations (Bloomfield and Marchant, 2013; Dawley et al., 2019; Loáiciga, 2015). It is possible to fit individual parametric distribution functions to the groundwater series, but the 30-year
250 monthly series used here are probably insufficient to do this reliably (Link et al., 2020), giving a high risk of biased drought index values (Svensson et al., 2017). Groundwater levels were therefore transformed using a normal scores transform (see e.g. Bloomfield and Marchant, 2013). This is a nonparametric transformation method that has the advantage of being simple and transparent and circumvents the risk of bias due to erroneous distribution fits. For each location, the simulated series was first aggregated to a monthly time step. Transformation was then done separately for each calendar month. For each calendar month
255 with $n$ years of data, in this case 30, cumulative probability values are taken, uniformly spaced over the interval $(1/n)$ to $(1 - 1/n)$; the corresponding SGI values are found by applying an inverse cumulative distribution function to these values. The resulting SGI values are assigned to the groundwater head measurements of the given calendar month by their rank from low to high. This method of calculating SGIs results in a limited number of 'discrete' SGI values that have a relatively weak relation with occurrence probability, but correspond directly to the rank of the groundwater level compared to the rest of the
260 reference period. Table 1 gives the SGI values and their corresponding rank and drought severity in this study. Note that the SGI, unlike the SPI, is not aggregated over multiple months; the SGI values given here therefore represent 'SGI-1'.

**Table 1: Used categories for the standardised groundwater index, with the corresponding groundwater level rank in a 30-year record.**

| SGI | Drought category | Rank (dry → wet) |
|---|---|---|
| **> 0** | No drought | >15 *(wettest 15 years)* |
| **0** to **-1** | Mild drought | 6 - 15 |
| **-1.5** to **-1** | Moderate drought | 3 - 5 |
| **-2** to **-1.5** | Severe drought | 2 |
| **< -2** | Extreme drought | 1 *(dryest year)* |

265     The SGI values for the months outside the reference period (January 2020 – May 2020) were estimated by linearly interpolating the groundwater head-SGI relation for the calendar month. If the heads fell outside the range reached in the reference period, they were assigned the most extreme SGI value.

    To test how the use of TSM simulations rather than measurement series affects drought analysis, SGI values were also calculated directly for a selection of locations that had long measurement series. From the cleaned measurement series of the

270     validation step, those were selected that started before 1 January 1993, resulting in 531 series. The SGI values for 2018 were calculated in the same way as for the simulated series, with the SGI of a calendar month calculated only if at least 25 years of data were available. The simulation-based and measurement-based SGI values were compared by regression.

### 3.4 Drought prediction

    Finally an estimate was made of the drought *prediction* performance of the time series models. If weather predictions are

275     available, time series simulations can be easily extended into the future. However, as the predicted period is not included in the calibration, the quality of the predictions is likely to be lower than for the normal 'fully calibrated' simulations. This is especially important for the prediction of extreme drought, which by definition involves system conditions not present in the calibration period. To test the performance of the time series models for predicting extreme drought conditions, models were re-calibrated until the start of the 2018 growing season (1 April 2018); and the groundwater behaviour for the rest of 2018 was

280     'predicted' with the available weather data. As such, an 'upper limit' of the prediction performance is estimated, representing how well the 2018 drought would have been predicted had perfect weather predictions been available. The prediction error in groundwater levels was quantified by the RMSE and mean error ME over July-November 2018.

    Previous studies have found that in slow-reacting groundwater systems and dry situations, initial conditions can control the quality of model predictions many months forward, irrespective of weather forecasts (Honingh et al., 2020; Mackay et al.,

285     2015; Sutanto et al., 2020). The type of time series models used here rely only on transforming of a time series of recharge to deviations in groundwater heads. The predicted heads are therefore insensitive to the initial conditions simulated at the start of the prediction period. For forecasting, it might be advantageous to include the true initial condition into the preditctions. The effect of a simulation residual on subsequent simulated values be estimated as:

$$RE(t) = R_{ini} * e^{\wedge}(-t/\alpha) \,(1)$$

290 Where $RE(t)$ is the remaining effect of the residual at time $t$ after the start of predictions, $R_{ini}$ is the residual at the start of the prediction period, and $\alpha$ is the decay parameter of the noise model (eq. 7 in Collenteur et al., 2019). Adding the series of residual effects to the predicted values thus draws these towards the real values, with the effect decreasing over time.

To estimate how the prediction error in groundwater heads would propagate to the drought quantification, SGI values were calculated both for the predicted series and for the same series with the prediction period replaced by the measurement data.

295 The performance of the SGI predictions was quantified by the mean and maximum absolute error of the prediction- versus the measurement-based SGI values. As such, the difference between the prediction-based and measurement-based SGI values gives an indication of the performance in predicting the occurrence of drought.

## 4 Results

Here, we present the outcomes of the drought analysis and the evaluation of the time series modelling-based data preparation

300 method. We first show the performance of the method for validation and simulation of groundwater head series; then we present the analysis of the 2018-2019 groundwater drought. Finally, the performance of the TSM method for drought quantification and prediction is shown.

### 4.1 Performance of TSM data preparation for groundwater head series

The performance of the applied validation method, based on a combination of basic tests and TSM, was tested on a subset of

305 the groundwater series. The method generally identified longer-term irregularities in behaviour well (Table 2, middle columns). More frequently, it incompletely removed outliers or detected outliers that were visually not clearly erroneous (left columns). Overall, the validation method was successful for a majority of the test series (68 %, right columns). For part of the series (20 %), validation was classified as reasonable, with some errors remaining or slight over-smoothing, but without a substantial effect on the overall series, thus being unlikely to affect drought analysis. For a small fraction of the test series (5 %), validation

310 was unsuccessful and more important irregularities remained. Thus, a small number of potentially erroneous series is likely to remain in the data used for the drought analysis.

**Table 2: Outcomes of the validation method test: performance of the method in cleaning of outliers, long-term patterns, and the series as a whole. Given is the number of test series in each performance category; see 3.1 for a description of the test.**

| Outliers | | Pattern | | Whole series | |
|---|---|---|---|---|---|
| **Performance** | No. series | Performance | No. series | Performance | No. series |
| **TP** | 16 | TP | 20 | Good | 38 |
| **TN** | 12 | TN | 23 | Reasonable | 11 |
| **FP** | 13 | FP | 8 | Bad | 3 |
| **FN** | 15 | FN | 5 | Deep | 4 |

315   When applied to the full study dataset, the validation procedure discarded 31% of the groundwater head measurement series, so that 1869 of the original 2722 series remained for analysis. A poor model fit was the most frequent cause for discarding. 10 % of the series were maintained as atypical series, while another 12 % of locations had a deep groundwater table, being less suitable for the used validation method.

In the simulation step, 1632 locations were modelled with sufficient quality (EVP > 60 %). Overall, these series were simulated

320   with an average error of 14 cm, resulting in a 20 % error in the groundwater table depth (Table 3). The bias is low with -1 mm. Subdivision into dry, normal and wet conditions shows that the errors are larger for more extreme groundwater levels. More precisely, the models tend to underestimate the extremes: there is a positive average bias in periods of low groundwater levels, and a negative bias when groundwater levels are in their high ranges. There was no clear spatial pattern in this bias. Also during the main period of groundwater drought in July-November 2018, the simulation error is above average with 18 cm, but

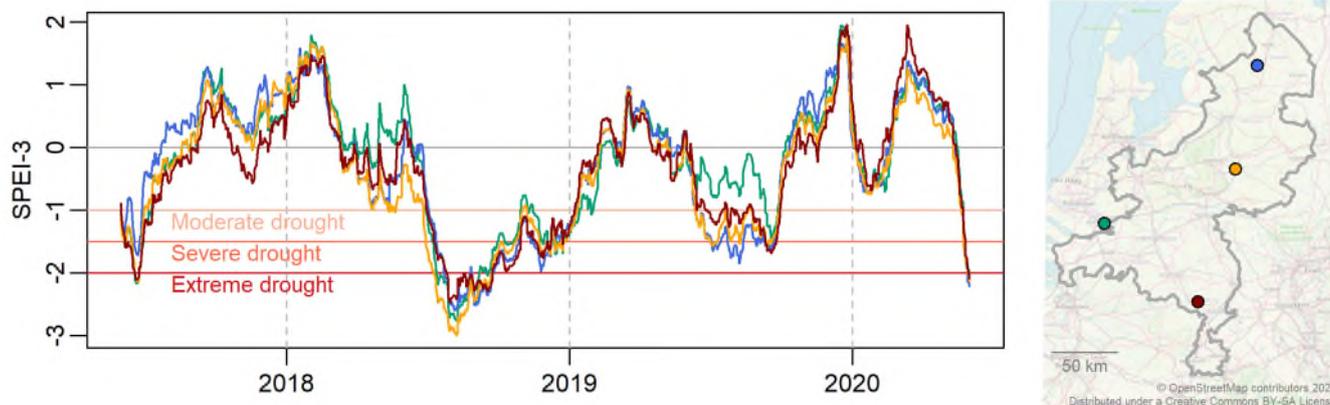
325   there is only a small negative bias of 1 cm.

**Table 3: Performance of the groundwater head simulations for summer 2018. RMSE=Root Mean Squared Error, ME=Mean Error of simulations versus measurements, given in meters and as fraction of the mean groundwater table depth (WTD).**

| | RMSE | | ME (mod-obs) | |
|---|---|---|---|---|
| | Mean value *(range)* [m] | Fraction WTD [-] | Mean value *(range)* [m] | Fraction WTD [-] |
| **Full period** | 0.14 *(0.03...1.7)* | 0.20 | $-1.2 \cdot 10^{-3}$ *(-0.3...0.2)* | $8.1 \cdot 10^{-3}$ |
| **Dry (0 – 20p)** | 0.16 *(0.03...2.0)* | 0.25 | 0.076 *(-0.7...0.6)* | 0.15 |
| **Normal (20 – 80p)** | 0.12 *(0.02...1.6)* | 0.18 | $-3.3 \cdot 10^{-3}$ *(-0.3...0.4)* | 0.039 |
| **Wet (80 – 100p)** | 0.15 *(0.03...1.7)* | 0.20 | -0.070 *(-1.3...0.3)* | 0.089 |
| **Jul-Nov 2018** | 0.18 *($6 \cdot 10^{-3}$...3.2)* | 0.29 | -0.010 *(-2.9...1.5)* | 0.22 |

**4.2 Development and recovery of the 2018-2019 groundwater drought in the Netherlands**

The groundwater drought of 2018-2019 was driven by exceptionally dry weather conditions. The meteorological drought

330   started in spring 2018 and peaked in late summer (Fig. 3). After a relatively normal winter, summer 2019 again showed moderate to severe drought. The winter of 2019-2020 was relatively wet, but exceptionally low rainfall in spring 2020 caused a return to extreme meteorological drought conditions. The meteorological drought was not spatially uniform. Especially in spring 2018 and summer 2019, the western part of the study area experienced less dry conditions than the east (Fig. 3).

**335** **Figure 3: Development of the meteorological drought over 2018-2020 in the west, north, mid-east and southeast sections of the study area given by the 3-month SPEI. The map shows the midpoints of the four sections.**

The development of the groundwater drought in the southeastern Netherlands over 2018 is visualised in Figures 4 and 5. The 2018 growing season started with uniformly normal to high groundwater levels over the study area (Fig. 4). Drought started developing in May and June, with drought onset varying between locations. By July and August, severe to extreme

**340** groundwater drought occurred over most of the area. In September, heavy rain in part of the study area slightly alleviated the drought conditions. However, the drought situation worsened again in autumn, reaching its height in October and November when the simulations show almost uniform extreme drought over the study area. By December, a slow recovery is visible.

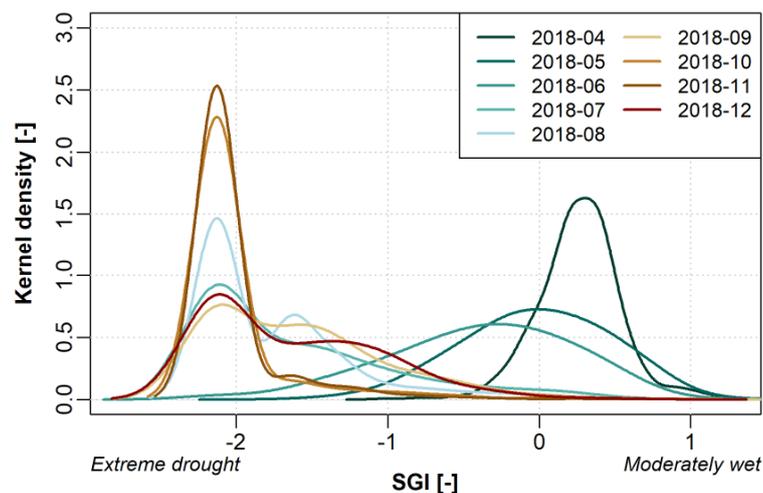
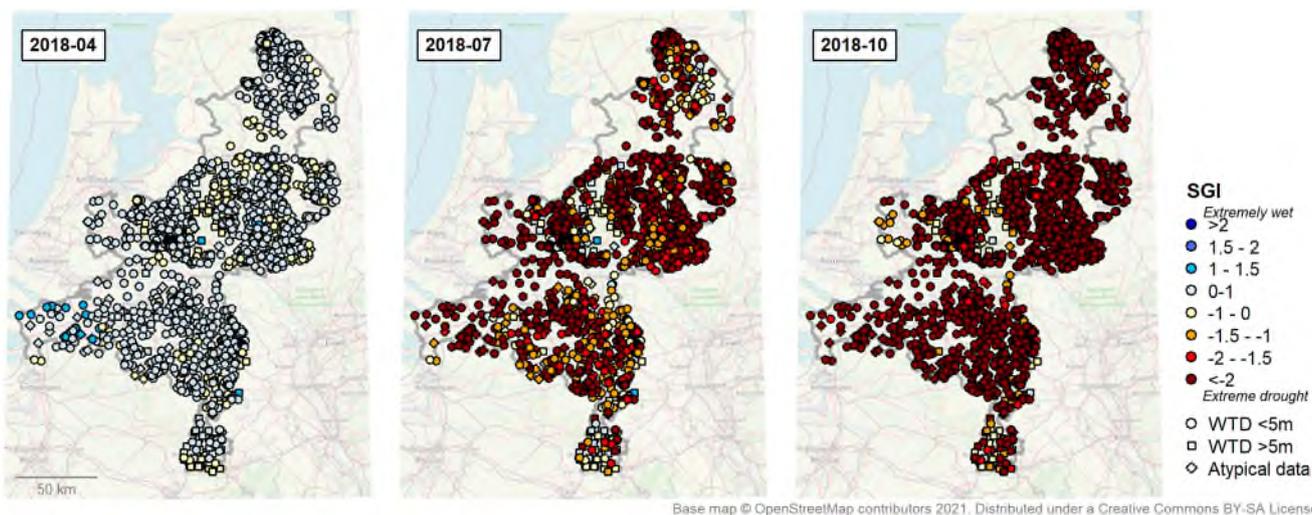



**Figure 4: Monthly distributions of the Standardised Groundwater Index over 2018 in the southeastern Netherlands.**

**345** The simulated data show distinct spatial patterns in the development of the groundwater drought (Fig. 5). Especially southern Limburg and the ridges in Utrecht and Gelderland with their distinct geology and topography (see Fig. 1) reacted more slowly than the rest of the study area and did not experience drought conditions yet in 2018. Also the fast recovery of the low-lying western Utrecht area in autumn stands out.

350 **Figure 5: Groundwater drought development in 2018: SGI of simulated series. WTD: Water table depth.**

The simulations over 2019-2020 show that also drought recovery was strongly variable in space (Fig. 6). In spring 2019, groundwater heads in the west of the study area were again approaching normal levels, but the eastern regions had recovered poorly. By this time extreme groundwater drought had also developed on the high ridges and in southern Limburg. The summer of 2019 again brought severe to extreme groundwater drought, this time clearly concentrated in the east of the study area, 355 corresponding to the differences in meteorological drought. In March 2020, groundwater levels had returned to relatively high levels. However, the exceptionally dry weather in April rapidly resulted in a new severe drought situation by May.

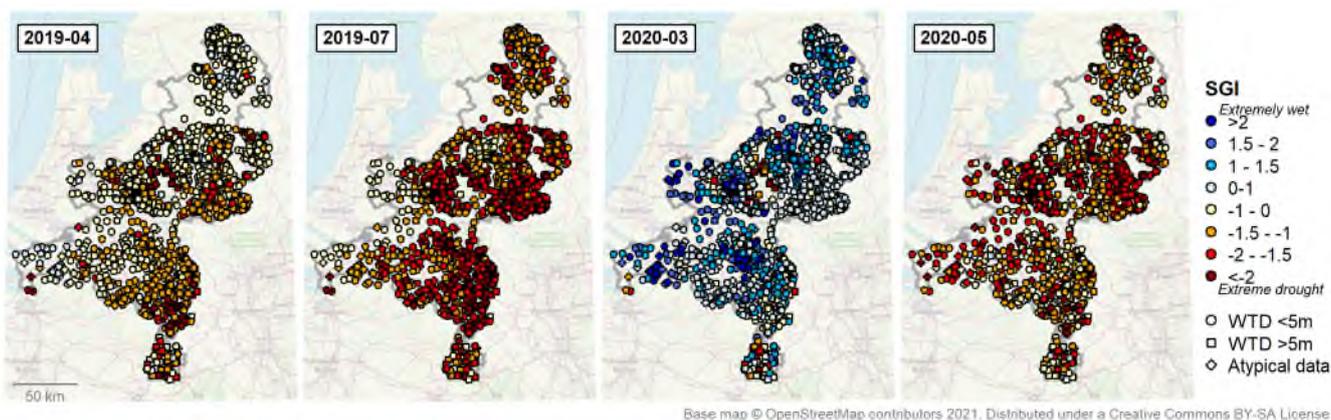



**Figure 6: Groundwater drought over 2019-2020: SGI of extended groundwater series.**

To further explore the spatial variations in drought behaviour, the groundwater response time was mapped for all monitoring 360 locations (Fig. 7). The response time, the time after which 50 % of the groundwater head response to a recharge event has occurred, can be obtained directly from the parameters of the fitted time series model for a specific location. The found response times ranged from a few days to up to two years, but were generally relatively short: 94 % of locations had a response time of less than one year. The spatial distribution of the response time corresponds with the topography and the occurrence of glacial

sand ridges in the landscape (Fig. 1) and with the propagation speed of meteorological drought to groundwater drought in
365  2018-2020 (Fig. 5, 6).

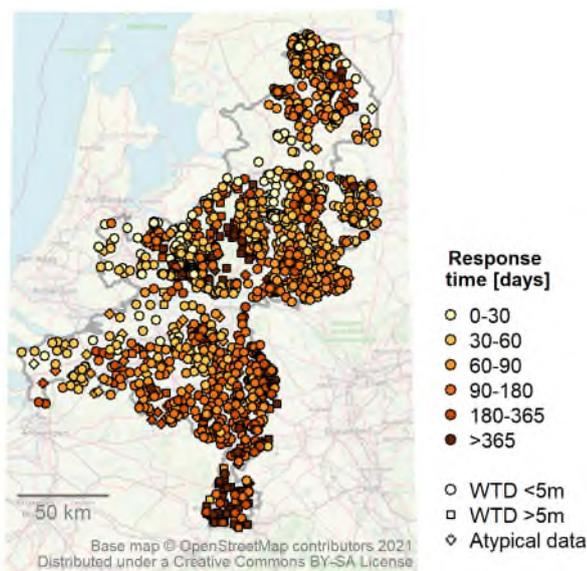



**Figure 7: Response time to recharge as derived from the fitted response functions.**

**4.3 Usefulness of TSM method for drought quantification and prediction**

For a subset of the locations ($n = 531$), a long groundwater measurement series was available and the SGI values resulting
370  from the simulated series could be compared to those obtained directly from the cleaned measurement series (Fig. 8). The
comparison shows that the simulations follow the same general drought behaviour as the measurement series, as the two follow
a 1:1 line (Spearman's $\rho = 0.8$). However, the simulations generally show a smaller spatial variation than the measurements.
This is also visible in measurement-based SGI maps (Fig. 9), which show more scatter and local extremes than the simulation-
based drought maps. In addition, the simulations for 2018 tend to slightly overestimate drought severity in the low ends of the
375  drought (lower left in Fig. 8). This is contrary to the general tendency of the head simulations towards positive bias during
drought periods, shown in Table 3.

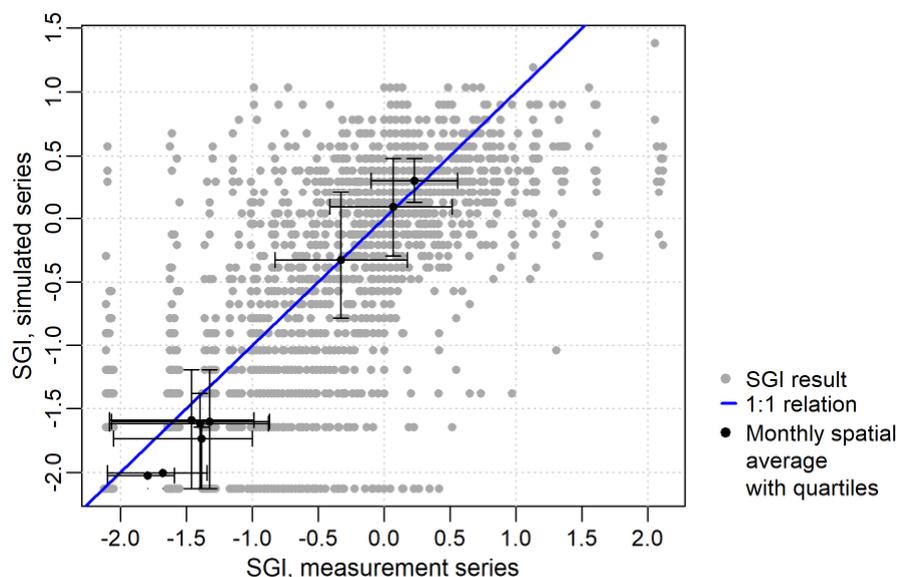



**Figure 8: Comparison of SGI values over April-December 2018 based on cleaned long measurement series (x axis) and based on simulated series (y axis) for all location-month combinations (grey dots). Black dots show the average over all locations for each month, with quartiles (spatial variation).**

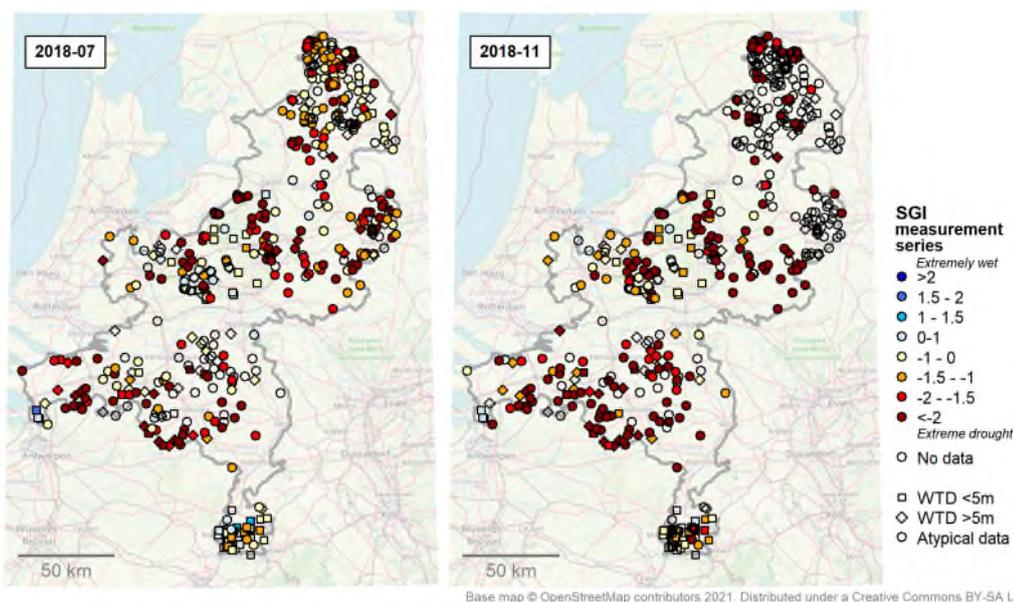



**Figure 9: SGI for two months in 2018 based directly on cleaned long measurement series. 'No Data' symbols indicate insufficient data to enable drought index quantification.**

In addition to the simulations, it was also tested how well the time series models, calibrated until the start of the growing season, would have predicted the 2018 groundwater drought from weather data only (Table 4). The average error in the predicted groundwater heads was 20 cm, giving a relative error in the groundwater depth of 31 %. The (small) negative mean error indicates that the declines in groundwater level over summer, and thus the severity of drought, are slightly overestimated.

This pattern is also visible in the simulations themselves. The prediction performance is poorer than that of the normal simulations with all data included in the calibration (Table 3). This means that, as expected, the prediction of groundwater

390 levels under extreme drought conditions has a relatively high unreliability. However, the difference in average error is only 3 cm, indicating that the effect is relatively small compared to the error already present in the fully calibrated simulations.

The prediction errors in the groundwater head were estimated to result in an average absolute SGI error of 0.34. With the used methods for SGI calculation and n=30 years, an error of 0.48 is required to cause a shift between "extreme drought" and "severe drought", and an error of 0.26 will cause a shift between severe and moderate drought. This means that most locations

395 that were in extreme drought in summer-autumn 2018 were also predicted as such, despite errors in the heads themselves. However, there is a large variation between locations in the size and direction of the drought index errors. Predictions are good for many locations (error ≈ 0), while some are classified at the complete other side of the drought spectrum than the measurements (error of -2.9; see Table 1).

**Table 4: Performance of the drought predictions for summer 2018: means and ranges of performance measures over all model**
400 **locations. RMSE=Root Mean Squared Error, ME= mean error, MAE=mean absolute error.**

| | Prediction of groundwater heads, summer 2018 | | Prediction of SGI, summer 2018 |
|---|---|---|---|
| | **Mean value (range) [m]** | **Mean fraction of WTD [-]** | **Mean value (range) [-]** |
| **RMSE** | 0.20 ($1·10^{-3}$...3.7) | 0.31 | - |
| **ME** | -0.015 (-3.5...1.2) | 0.24 | -0.21 (-2.9...1.0) |
| **Error minimum head** | -0.048 (-4.9...1.5) | 0.34 | - |
| **MAE** | - | | **0.34 (0.0...2.9)** |
| **Maximum AE** | - | | 0.72 (0.0...3.5) |

There are no clear spatial patterns in the RMSE of the predictions, nor in the estimated error in SGI (not shown). The prediction error in the minimum level is also largely randomly distributed, although some larger overestimations occur in part of southern Noord-Brabant, while the minimum levels are underestimated in southern Limburg. The performance of the predictions thus

405 appears independent of specific catchment characteristics.

The initial conditions simulated at the start of the growing season had an average absolute error of 15 cm. Including the true initial groundwater heads on March 31 into the predictions led to small improvements in the prediction on a short term: this reduced the prediction errors after 1 and 7 days on average by 14 cm and 5 cm, respectively. However, after a month or more the effect of the initial residual becomes insignificant, and after 4 months, the inclusion of initial conditions leads to an average

410 increase in error of 0.9 cm. Inclusion of the residual at the start of the growing season therefore did not improve drought predictions in the height of summer.

# 5 Discussion

## 5.1 Usefulness of TSM data preparation for drought analysis

The performed study aimed to evaluate the usefulness of time series modelling-based data processing methods for groundwater
415   drought studies. The application of the TSM-based method to the 2018-2019 drought in the Netherlands provided several new
insights into how TSM methods can be used for data validation; how reliable they are for the quantification of extreme drought
situations; and their usefulness for drought prediction.

### 5.1.1 Validation methods for groundwater data

A validation method was applied that combines basic error tests with time series modelling-based identification of
420   irregularities, while separately treating short-term outliers and long-term atypical behaviour. It was found that the data available
for this study, and this may be true for many cases, contained large numbers of outliers with various sources. Removing these
outliers often led to a good TSM fit that was not possible before this procedure. Therefore, with the kind of data used in this
study, pre-analysis outlier cleaning is likely to improve results compared to performing a more limited validation, or identifying
impacted series only after simulation. The validation method performed reasonably with regard to outlier removal, but not
425   optimally (Table 2). The outlier cleaning may be further improved by using more iterations (e.g. Peterson et al., 2018), although
care must be taken to avoid over-smoothing.

The TSM-based validation also proved suitable for identifying long-term atypical behaviour in the head measurement series.
However, we could not explicitly separate erroneous and real atypical patterns in this study. Many of the long-term non-
weather influences on the groundwater, such as structural abstractions and land use, are likely to be included implicitly in the
430   simulations, as the time series models will model any external influence that correlates with weather or modifies the recharge
response. In addition, the use of a classification for long-term behaviour allowed for retaining moderately 'atypical series'
rather than filtering out all impacted locations. Still, the validation is likely to retain some series with erroneous patterns and
discard some real behaviour.

The separation of errors could be improved by including additional driving factors in the time series modelling, such as surface
435   water fluctuations (Bakker and Schaars, 2019; Van Loon et al., 2016; Zaadnoordijk et al., 2019); however, this requires
substantially more data and modelling effort. Another potential approach is to make use of the spatial coherence of groundwater
behaviour to separate errors. For example, Lehr and Lischeid (2020) and Marchant and Bloomfield (2018) used the spatial
coherence of observed groundwater patterns as an indication of their reliability, and to relate clusters of similar series to some
external (human) impact. It would be valuable to explore such extensions of TSM to improve validation methods for
440   groundwater drought analysis. This will allow for a better understanding of the different natural ánd human drivers of drought
development.

Although the validation-simulation method generally performed well for the given data, it was less suitable for locations with
very deep groundwater tables, dominated by multi-year head fluctuations. Here, model fit was often low, leading to poor outlier

identification and models being discarded for a large proportion of these locations. This issue was also found by Marchant and

445 Bloomfield (2018) and Zaadnoordijk et al. (2019) for locations with thick unsaturated zones. To enable TSM simulation in

such cases, measurement series are needed that are substantially longer than the minimum of 10 years used here.

### 5.1.2 Reliability of TSM groundwater level simulations during extreme drought

It was found that impulse-response function-based time series models on a general level produced reliable groundwater head

simulations. They described most of the groundwater head series very closely, with a low overall bias (Table 3). However, the

450 time series models tended to underestimate extreme conditions somewhat, including dry summer conditions, although the

simulations for summer 2018 interestingly had a small negative bias in the lowest ranges. The overall underestimation of

extreme conditions was also found by Mackay et al. (2015) with a process-based groundwater model. An important

shortcoming of the models is that they assume that the response time to recharge, or the lack of it, is constant for a given

location. In reality, the response may be different during (extreme) drought conditions. This may contribute to less reliable

455 TSM estimations during extreme groundwater drought.

When considering the resulting drought index values, the time series models also appear to give a correct representation of

overall drought dynamics. The comparison of simulation- and measurement-based SGI values (Fig. 8) showed a good

correlation (Spearman's $\rho = 0.8$). The drought estimation is similar or slightly better than found by Van Loon et al. (2017),

who simulated groundwater drought in the eastern Netherlands with a lagged SPEI-SGI correlation (Spearman's $\rho \approx 0.75$).

460 Also, as discussed further below, the general patterns shown by the drought maps match well with the experience of water

managers.

However, the comparison of simulation- and measurement-based SGI values also showed that the simulation-based drought

quantification overestimated drought severity in the lowest ends of the 2018 drought, while also reducing the spatial variability.

In summer and autumn 2018, the models simulated uniform extreme drought over the whole study area, while the

465 measurement-based SGI values showed much more scatter towards less extreme drought (Fig. 8-9). It appears that the weather

conditions were severe enough to trigger a 30-year extreme groundwater drought everywhere, but conditions not accounted

for by the models, such as external water supply, caused the drought to be less extreme locally (Van den Eertwegh et al., 2019).

In addition, the models may have underestimated the severity of earlier droughts; this would inflate the extremeness of the

2018 drought. Similar bias issues can be seen in Hellwig et al. (2020), who analysed groundwater drought in Germany using

470 a physically-based model. Their simulations overestimated drought severity during a drought in 1973, but not in 2003,

confirming that bias can differ between individual drought events. In their case, drought severities and bias therein were

strongly influenced by the parameters related to the response time of the groundwater system. Similar issues may have played

a role for the time series models used here.

Despite these shortcomings, the use of time series model simulations in drought analysis provides large advantages. In this

475 study, it allowed for including a much larger number of series than if only long and disturbance-free series could have been

used. This enabled a detailed image of drought development that would not have been possible by using direct measurement

series only. In addition, the model parameters gave additional insight in groundwater behaviour, for example through the response time (Bakker and Schaars, 2019; Collenteur et al., 2019). Therefore, impulse-response time series simulations are considered a valuable tool for groundwater drought analysis in the Netherlands and other situations, although care must be
480 taken for bias and (local) effects of non-weather influences on the groundwater system.

### 5.1.3 Potential of TSM for groundwater drought prediction

The findings also provide support for the usefulness of time series modelling for groundwater drought prediction. We found that, had perfect weather forecasts been available at the start of the growing season in 2018, the type of time series models used in this study would have reasonably predicted the course of the groundwater drought in summer and autumn, with an
485 error of around 20 cm and low bias (Table 4). The potential for nowcasting is also shown by the lengthened series over 2019-2020, which matched well with general observations provided by other studies and reports (LCW, 2020; Van den Eertwegh et al., 2019). It was expected that the prediction of groundwater heads in periods of extreme drought would perform relatively poorly, as system conditions are being simulated that are not included in the calibration period. However, the errors of our predictions were close to those of the normal simulations that did include 2018 in their calibration (20 vs 18 cm, Table 3-4).
490 The results also suggest that, over the time spans relevant for drought forecasting (months), updating forecasts with measurements may not be of much added value. This supports the idea that time series models, when based on sufficient calibration data, could be a relatively robust means for drought prediction over longer time spans.

Still, the prediction RMSE and mean errors amounted locally to high levels (Table 4). The drought index values in the main drought period of 2018 were predicted with an estimated error of 0.3 units, which can easily shift a location to a different
495 drought category. The magnitude of these errors is similar to that reported by Marchant and Bloomfield (2018), who found that temporal interpolation of head measurement series with TSM led to an uncertainty in SGI values of roughly 0.1 to 0.2 in periods with good data cover, but much larger in longer periods with little data. In this study, we accounted only for the prediction error in drought severity resulting from model inaccuracy in the prediction period. In reality, the prediction uncertainty would obviously be increased by the uncertainty in the long-term weather forecasts (Honingh et al., 2020; Mackay
500 et al., 2015).

Taken together, time series models show potential for groundwater drought nowcasting and forecasting on a general regional scale. However, if water managers are to be provided structurally with up-to-date information on regional groundwater drought, the models have to be calibrated with reliable, recent measurement data. Therefore, improving the accessibility and consistency of recent groundwater data among different organisations and countries remains an important task.

## 5.2 Analysis of the 2018-2019 groundwater drought in the Netherlands

### 5.2.1 Spatiotemporal development of groundwater drought in 2018-2019

The regional-scale analysis of the 2018-2019 groundwater drought in the southeastern Netherlands provided a new, detailed image of the drought development in time and space. The analysis showed that extreme groundwater drought was widespread over the study region in 2018 and 2019, breaking 30-year records almost everywhere (Fig. 4). However, the timing and duration of the drought varied strongly in space. In the western parts of the study area drought was terminated before the end of 2018, while the higher-lying areas reached drought conditions only by 2019 (Fig. 5-6). This image corresponds well with how the drought was generally experienced by water managers. The Dutch Commission on Water allocation (LCW), which regularly surveys the drought status in the Netherlands based on input from water managers, reported widespread exceptional drops in groundwater levels, especially in the higher-lying sandy areas of the country, combined with a relatively fast recovery in the west (LCW, 2020). Also the severe damage observed in groundwater-dependent ecosystems in the southeastern Netherlands (Witte et al., 2020a) and widespread drying of groundwater-fed streams (Van den Eertwegh et al., 2019) match the assessment of an extreme groundwater situation.

The time series models made it possible to extend the available groundwater head measurement series beyond 2018 and obtain an estimate of drought recovery dynamics over 2019-2020. This showed that despite near-normal to high rainfall in the winters of 2018-19 and 2019-20 (Fig. 3) groundwater levels were restored only locally, and severe groundwater drought continued into 2019 or even 2020 over large parts of the study area (Fig. 6). These findings are consistent with general reports of the situation by local water managers (LCW, 2020; Van den Eertwegh et al., 2019). The results confirm that the drought should be viewed as a multi-year event rather than a single-year summer drought. The multi-year character increases the risk of lasting negative effects, especially in natural ecosystems. In addition, it stresses the importance of winter season groundwater management and water retention as determining factors in drought development. Both these issues have already become apparent in the Netherlands after 2018 (De Lenne and Worm, 2020; Witte et al., 2020a; Witte et al., 2020b).

### 5.2.2 Driving factors of spatial drought distribution

The current study did not aim to fully quantify the driving factors of the spatial variations in drought dynamics. However, the results suggest that the variation in drought severity, timing and duration was governed mainly by the spatial distribution in rainfall and the geological-topographic setting. The influence of spatial variations in weather was visible in late 2018 and 2019. In this period, a gradient in meteorological drought towards the east caused groundwater drought to be concentrated in this area. The effect of the geological-topographic setting is visible for the higher elevated parts of the study area, especially the glacial sand ridges and the limestone-loess hills, which clearly had a late, long-lasting drought response compared to lower-lying areas (Fig. 5-6). This pattern is also visible in the groundwater response times (Fig. 7). The slow drought response in the higher elevated areas is likely explained by a thick unsaturated zone and low drainage density, while the fast recovery of the low-lying western parts may have been aided by thinner unsaturated zones and possibly some surface water influence. These

factors have been found by other studies to influence drought behaviour (Bloomfield et al., 2015; Hellwig et al., 2020; Peters et al., 2006; Van Loon and Laaha, 2015). In addition, the topographic-geological setting in the study area correlates with variations in land use, soil type and aquifer characteristics; these factors may have played an additional role.

540 The dominant role of landscape position in shaping drought development calls for locally adapted, but also regionally coordinated mitigation strategies. The response times obtained from the time series analysis form a useful first indicator of the landscape characteristics that control the propagation of meteorological drought to groundwater drought. The response times found here are somewhat shorter than the drought response times reported by Van Loon et al. (2017) for part of the eastern Netherlands. However, they are very similar to those found by Zaadnoordijk et al. (2019), who performed time series analysis

545 on groundwater series from the whole Netherlands. As they used different data and different model quality criteria, the similarity is reassuring and points to the stability of the time series models.

## 6 Conclusions

The performed study aimed to evaluate the usefulness of time series modelling-based data processing methods for regional-scale groundwater drought monitoring and prediction. A TSM-based method was set up for data validation and drought

550 quantification and applied to the regional 2018-2019 groundwater drought in the southeastern Netherlands to test its usefulness and reliability. It was found that the kind of groundwater data often available for drought studies requires a validation-cleaning step before analysis to enable reliable results. The validation method based on a combination of basic tests and time series modelling proved a suitable method for the given situation. Still, improvements in the validation method are desired, especially in the separation of real and erroneous head disturbances. The simulated groundwater head series were generally found to be

555 reliable; however, it was shown that the use of time series simulations may bias drought estimations and underestimate spatial variability, producing large errors at a local scale. Still, the use of time series model simulations in drought analysis provides large advantages, as it enables a spatially detailed record of drought development that may be impossible with direct measurement series only. Time series models were also found to be a promising tool for regional-scale drought nowcasting and prediction, although the same caution for bias and local errors is needed.

560 The drought analysis for the southeastern Netherlands provided a complete, detailed image of the development of the 2018-2019 groundwater drought in time and space. The findings confirm that the meteorological drought in 2018 caused extreme groundwater drought throughout the southeastern Netherlands, starting in late spring and peaking in October-November of that year. The timing of drought onset and the duration of drought varied strongly in space. Drought development appeared to be governed dominantly by the spatial distribution of rainfall and the geological-topographic setting. In much of the area, the

565 drought continued as a multi-year drought into 2019 and 2020, especially in the eastern and higher elevated regions.

Taken together, we conclude that time series modelling forms a useful tool to obtain a quick, detailed and up-to-date image of drought development; however, for a proper understanding of the different driving factors of drought the availability of recent and consistent monitoring data remains crucial.

**Code and data availability**

570    All code for data processing, modelling and visualisation is available on request from the first author. Groundwater data are
freely available from the national groundwater database DINO ([www.dinoloket.nl](www.dinoloket.nl)). The used weather data can be obtained
from the Royal Dutch Meteorological Institute (KNMI) via [http://projects.knmi.nl/klimatologie/daggegevens/selectie.cgi](http://projects.knmi.nl/klimatologie/daggegevens/selectie.cgi) or
through a script via [https://www.knmi.nl/kennis-en-datacentrum/achtergrond/data-ophalen-vanuit-een-script](https://www.knmi.nl/kennis-en-datacentrum/achtergrond/data-ophalen-vanuit-een-script).

**Author contribution**

575    EB, MvH and RB together designed the study. EB performed the study with advice from MvH and RB. The first article draft
was written by EB and reviewed by all three authors.

**Competing interests**

The authors declare that they have no conflict of interest.

**Acknowledgements**

580    The authors thank Sharon Clevers for collecting and pre-processing an important part of the groundwater data.

**References**

AHN:          AHN3          50cm          maaiveld,          Esri          [Data          set],
           [https://www.arcgis.com/home/group.html?id=bd6fbbd182a9465ea269dd27ea985d1a#overview](https://www.arcgis.com/home/group.html?id=bd6fbbd182a9465ea269dd27ea985d1a#overview), 2019
Bakke, S. J., Ionita, M., and Tallaksen, L. M.: The 2018 northern European hydrological drought and its drivers in a historical
585           perspective, Hydrol. Earth Syst. Sci., 24, 5621-5653, 10.5194/hess-24-5621-2020, 2020.
Bakker, M., and Schaars, F.: Solving Groundwater Flow Problems with Time Series Analysis: You May Not Even Need
           Another Model, Groundwater, 57, 826-833, [https://doi.org/10.1111/gwat.12927](https://doi.org/10.1111/gwat.12927), 2019.
Bartholomeus, R. P., Witte, J.-P. M., van Bodegom, P. M., and Aerts, R.: The need of data harmonization to derive robust
           empirical relationships between soil conditions and vegetation, Journal of Vegetation Science, 19, 799-808,
590           [https://doi.org/10.3170/2008-8-18450](https://doi.org/10.3170/2008-8-18450), 2008.
Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., Wigneron, J. P., Weber, U., Reichstein, M., Fu, Z.,
           Anthoni, P., Arneth, A., Haverd, V., Jain, A. K., Joetzjer, E., Knauer, J., Lienert, S., Loughran, T., McGuire, P. C.,
           Tian, H., Viovy, N., and Zaehle, S.: Direct and seasonal legacy effects of the 2018 heat wave and drought on European
           ecosystem productivity, Science Advances, 6, eaba2724, 10.1126/sciadv.aba2724, 2020.
595    Bloomfield, J., Brauns, B., Hannah, D. M., Jackson, C., Marchant, B., and Van Loon, A. F.: The Groundwater Drought
           Initiative (GDI): analysing and understanding groundwater drought across Europe EGU General Assembly, Vienna,
           Austria, 8-13 April 2018, 2018.
Bloomfield, J. P., and Marchant, B. P.: Analysis of groundwater drought building on the standardised precipitation index
           approach, Hydrol. Earth Syst. Sci., 17, 4769-4787, 10.5194/hess-17-4769-2013, 2013.
600    Bloomfield, J. P., Marchant, B. P., Bricker, S. H., and Morgan, R. B.: Regional analysis of groundwater droughts using
           hydrograph classification, Hydrol. Earth Syst. Sci., 19, 4327-4344, 10.5194/hess-19-4327-2015, 2015.

Buras, A., Rammig, A., and Zang, C. S.: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003, Biogeosciences, 17, 1655-1672, 10.5194/bg-17-1655-2020, 2020.

Collenteur, R. A., Bakker, M., Caljé, R., Klop, S. A., and Schaars, F.: Pastas: Open Source Software for the Analysis of Groundwater Time Series, Groundwater, 57, 877-885, https://doi.org/10.1111/gwat.12925, 2019.

Dawley, S., Zhang, Y., Liu, X., Jiang, P., Tick, G. R., Sun, H., Zheng, C., and Chen, L.: Statistical analysis of extreme events in precipitation, stream discharge, and groundwater head fluctuation: Distribution, memory, and correlation, Water, 11, 707, https://doi.org/10.3390/w11040707, 2019.

De Lenne, R., and Worm, B.: Droogte 2018 & 2019: steppeachtige verschijnselen op de 'Hoge Zandgronden'. De gevolgen voor beheer en beleid bij Waterschap Vechtstromen, Stromingen, 25, 1-12, https://www.nhv.nu/wp-content/uploads/2020/06/2020-2_in_press_steppeachtige_verschijnselen_op_de_hoge_zandgronden.pdf, 2020.

Hellwig, J., de Graaf, I. E. M., Weiler, M., and Stahl, K.: Large-Scale Assessment of Delayed Groundwater Responses to Drought, Water Resources Research, 56, e2019WR025441, https://doi.org/10.1029/2019WR025441, 2020.

Honingh, D., Hurkmans, R., and Groot, S.: De toepasbaarheid van ECMWF-seizoensverwachtingen voor het voorspellen van grondwaterstanden, Stromingen, 2, https://www.nhv.nu/stroming/nieuws/de-toepasbaarheid-van-ecmwf-seizoensverwachtingen-voor-het-voorspellen-van-grondwaterstanden/, 2020.

IenW: Nederland beter weerbaar tegen droogte: Eindrapportage Beleidstafel Droogte, Ministerie van Infrastructuur en Waterstaat, https://www.rijksoverheid.nl/documenten/rapporten/2019/12/18/nederland-beter-weerbaar-tegen-droogte, 2019.

KNMI: Klimatologie/daggegevens, [Data set], https://www.knmi.nl/kennis-en-datacentrum/achtergrond/data-ophalen-vanuit-een-script, 2020

LCW: Archief droogtemonitoren 2018-2020, https://waterberichtgeving.rws.nl/LCW/droogtedossier/droogtemonitoren-archief/droogtemonitoren-2018, 2020.

Lehr, C., and Lischeid, G.: Efficient screening of groundwater head monitoring data for anthropogenic effects and measurement errors, Hydrol. Earth Syst. Sci., 24, 501-513, 10.5194/hess-24-501-2020, 2020.

Leunk, I.: Kwaliteitsborging grondwaterstands- en stijghoogtegegevens. Validatiepilot; analyse van bestaande data. , KWR Watercycle Research Institute, Nieuwegein, The Netherlands, 2014.059, https://edepot.wur.nl/317083, 2014.

Link, R., Wild, T. B., Snyder, A. C., Hejazi, M. I., and Vernon, C. R.: 100 years of data is not enough to establish reliable drought thresholds, Journal of Hydrology X, 7, 100052, https://doi.org/10.1016/j.hydroa.2020.100052, 2020.

Loáiciga, H. A.: Probability Distributions in Groundwater Hydrology: Methods and Applications, Journal of Hydrologic Engineering, 20, 04014063, 10.1061/(ASCE)HE.1943-5584.0001061, 2015.

Mackay, J. D., Jackson, C. R., Brookshaw, A., Scaife, A. A., Cook, J., and Ward, R. S.: Seasonal forecasting of groundwater levels in principal aquifers of the United Kingdom, Journal of Hydrology, 530, 815-828, https://doi.org/10.1016/j.jhydrol.2015.10.018, 2015.

Makkink, G.: Testing the Penman formula by means of lysimeters, Journal of the Institution of Water Engineers, 11, 277-288, 1957.

Marchant, B. P., and Bloomfield, J. P.: Spatio-temporal modelling of the status of groundwater droughts, Journal of Hydrology, 564, 397-413, https://doi.org/10.1016/j.jhydrol.2018.07.009, 2018.

Margariti, J., Rangecroft, S., Parry, S., Wendt, D. E., and Van Loon, A. F.: Anthropogenic activities alter drought termination, Elementa: Science of the Anthropocene, 7, 10.1525/elementa.365, 2019.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, Proceedings of the 8th Conference on Applied Climatology, 1993, 179-183,

Peters, E., Bier, G., van Lanen, H. A. J., and Torfs, P. J. J. F.: Propagation and spatial distribution of drought in a groundwater catchment, Journal of Hydrology, 321, 257-275, https://doi.org/10.1016/j.jhydrol.2005.08.004, 2006.

Peterson, T. J., Western, A. W., and Cheng, X.: The good, the bad and the outliers: automated detection of errors and outliers from groundwater hydrographs, Hydrogeology Journal, 26, 371-380, 10.1007/s10040-017-1660-7, 2018.

Pezij, M., Augustijn, D. C. M., Hendriks, D. M. D., and Hulscher, S. J. M. H.: The role of evidence-based information in regional operational water management in the Netherlands, Environmental Science & Policy, 93, 75-82, https://doi.org/10.1016/j.envsci.2018.12.025, 2019.

650    Philip, S. Y., Kew, S. F., van der Wiel, K., Wanders, N., and Jan van Oldenborgh, G.: Regional differentiation in climate change induced drought trends in the Netherlands, Environmental Research Letters, 15, 094081, 10.1088/1748-9326/ab97ca, 2020.

Post, V. E. A., and von Asmuth, J. R.: Review: Hydraulic head measurements—new technologies, classic pitfalls, Hydrogeology Journal, 21, 737-750, 10.1007/s10040-013-0969-0, 2013.

655    Sutanto, S. J., Wetterhall, F., and Van Lanen, H. A. J.: Hydrological drought forecasts outperform meteorological drought forecasts, Environmental Research Letters, 15, 084010, 10.1088/1748-9326/ab8b13, 2020.

Svensson, C., Hannaford, J., and Prosdocimi, I.: Statistical distributions for monthly aggregations of precipitation and streamflow in drought indicator applications, Water Resources Research, 53, 999-1018, https://doi.org/10.1002/2016WR019276, 2017.

660    TNO: DINOloket Ondergrondgegevens, Nederland, T. G. D. [Data set], https://www.dinoloket.nl/ondergrondgegevens,

Toreti, A., Belward, A., Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie, O., Seguini, L., Manfron, G., Lopez-Lozano, R., Baruth, B., van den Berg, M., Dentener, F., Ceglar, A., Chatzopoulos, T., and Zampieri, M.: The Exceptional 2018 European Water Seesaw Calls for Action on Adaptation, Earth's Future, 7, 652-663, https://doi.org/10.1029/2019EF001170, 2019.

665    Van de Velde, I., van der Kooij, S., Van Hussen, K., and Läkamp, R.: Economische schade door droogte in 2018, Ecorys, Rotterdam, the Netherlands, https://www.rijksoverheid.nl/documenten/rapporten/2019/08/31/economische-schade-door-droogte-in-2018, 2019.

Van den Eertwegh, G., Bartholomeus, R., De Louw, P., Witte, F., van Dam, J., Van Deijl, D., Hoefsloot, P., Sharon, C., Van Huijgevoort, M., Hunink, J., Mulder, N., Pouwels, J., and De Wit, J.: Droogte in zandgebieden van Zuid-, Midden-
670    en Oost-Nederland: Rapportage fase 1: ontwikkeling van uniforme werkwijze voor analyse van droogte en tussentijdse bevindingen., KnowH2O, https://edepot.wur.nl/511196, 2019.

Van Loon, A. F.: Hydrological drought explained, WIREs Water, 2, 359-392, https://doi.org/10.1002/wat2.1085, 2015.

Van Loon, A. F., and Laaha, G.: Hydrological drought severity explained by climate and catchment characteristics, Journal of Hydrology, 526, 3-14, https://doi.org/10.1016/j.jhydrol.2014.10.059, 2015.

675    Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A. I. J. M., Tallaksen, L. M., Hannaford, J., Uijlenhoet, R., Teuling, A. J., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., and Van Lanen, H. A. J.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, Hydrol. Earth Syst. Sci., 20, 3631-3650, 10.5194/hess-20-3631-2016, 2016.

Van Loon, A. F., Kumar, R., and Mishra, V.: Testing the use of standardised indices and GRACE satellite data to estimate the
680    European 2015 groundwater drought in near-real time, Hydrol. Earth Syst. Sci., 21, 1947-1971, 10.5194/hess-21-1947-2017, 2017.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, Journal of Climate, 23, 1696-1718, 10.1175/2009JCLI2909.1, 2010.

685    von Asmuth, J. R., Bierkens, M. F. P., and Maas, K.: Transfer function-noise modeling in continuous time using predefined impulse response functions, Water Resources Research, 38, 23-21-23-12, https://doi.org/10.1029/2001WR001136, 2002.

von Asmuth, J. R., Maas, K., Knotters, M., Bierkens, M. F. P., Bakker, M., Olsthoorn, T. N., Cirkel, D. G., Leunk, I., Schaars, F., and von Asmuth, D. C.: Software for hydrogeologic time series analysis, interfacing data with physical insight,
690    Environmental Modelling & Software, 38, 178-190, https://doi.org/10.1016/j.envsoft.2012.06.003, 2012.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., and Wood, E. F.: Development and Evaluation of a Pan-European Multimodel Seasonal Hydrological Forecasting System, Journal of Hydrometeorology, 20, 99-115, 10.1175/JHM-D-18-0040.1, 2019.

Witte, F., Runhaar, J., and van den Eertwegh, G.: Verdroging van de Nederlandse natuur: bijna een halve eeuw goed onderzoek
695    en falende politiek., Stromingen, 26, 65-80, https://www.nhv.nu/stroming/nieuws/verdroging-van-de-nederlandse-natuur-bijna-een-halve-eeuw-goed-onderzoek-en-falende-politiek/, 2020a.

Witte, F., van Deijl, D., and van den Eertwegh, G.: Gevolgen voor de natuur op de hogere zandgronden van de droge jaren 2018 en 2019; Resultaten van een enquête onder deskundigen., FWE, KnowH2O, https://edepot.wur.nl/521008, 2020b.

700    WUR: Grondsoortenkaart 2006, Research, W. U. [Data set], https://www.wur.nl/en/show/Grondsoortenkaart.htm, 2006

Zaadnoordijk, W. J., Bus, S. A. R., Lourens, A., and Berendrecht, W. L.: Automated Time Series Modeling for Piezometers in the National Database of the Netherlands, Groundwater, 57, 834-843, https://doi.org/10.1111/gwat.12819, 2019.