Large-Scale Assessment of Delayed Groundwater Responses to Drought

J. Hellwig[1] (ID), I. E. M. de Graaf[1] (ID), M. Weiler[2] (ID), and K. Stahl[1] (ID)

[1]Environmental Hydrological Systems, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany, [2]Hydrology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

Abstract Groundwater is a vital resource for freshwater supply during extended droughts and also a key storage governing drought propagation through the hydrological cycle. Current drought monitoring lacks large-scale estimates of groundwater droughts, but progress of country-to-global-scale models in the last years suggests that they could now be valuable tools to study and monitor water availability during extended droughts. As a prerequisite the models would need to be able to depict the diverse groundwater response to precipitation well enough to distinguish spatial differences. Here we developed a high-resolution transient groundwater model for Germany and tested its ability for representing the groundwater system dynamics with a focus on droughts. Validation of model results against streamflow-separated baseflow and groundwater head observation confirmed the model's ability to generally represent the groundwater head dynamics over 40 years with lower model performance in mountainous regions where model resolution was too low to capture local valley aquifers. The precipitation accumulation time that has the highest correlation with groundwater anomalies increases with hydraulic conductivity and specific yield from few months in the Central German Uplands to several years in the porous aquifers of northern Germany. Corresponding to these differences, distinct meteorological drought types led to different simulated groundwater reactions across Germany. Given the importance of groundwater as a resource, large-scale groundwater models are important tools for future studies on drought propagation as well as groundwater drought under climate change.

1. Introduction

At the global scale, groundwater is by far the largest freshwater storage of the active hydrological cycle. Accordingly, it serves as an important source for domestic and industrial water supply and agricultural irrigation. Groundwater storage buffers against hydroclimatic variations of water availability, as it retains precipitation and maintains streamflow during periods with low precipitation by groundwater discharge (i.e., "baseflow"). It is therefore the most important water source during extended meteorological drought.

In general, drought events are large-scale phenomena. Precipitation deficits which are commonly the main cause of drought events usually cover hundreds of kilometers. In the theory of drought propagation (e.g., Changnon, 1987; Van Loon, 2015) the meteorological drought signal (i.e., the precipitation deficit) translates into a delayed and attenuated soil moisture deficit (soil moisture drought), discharge deficit (hydrological drought), and decreased groundwater heads (groundwater drought). To compare these different hydrological fluxes and storages and their variability in space it has become common practice to express drought severity as a (standardized) relative deviance instead of an absolute value. Standardized drought indices (e.g., McKee et al., 1993; Palmer, 1965) and percentile thresholds (e.g., Andreadis et al., 2005; Tallaksen et al., 2009) are examples for relative drought metrics that are now frequently used for many variables and applications. While this large-scale comparative approach necessarily neglects small-scale local conditions, the relative metrics allow for a better comparison in space. The natural variability of a hydro(geo)logical variable can be quite large over large distances and hence absolute deficiencies are often not directly comparable. An anomaly-based comparison of drought characteristics for different variables and regions helps to identify large-scale drivers of drought patterns and is a prerequisite for assessments of climate sensitivity, predictions of drought impacts, and proactive drought governance.

Based on large-scale patterns of anomalies, many organizations have developed drought monitoring and early warning maps and systems for different regions worldwide. While almost all of them consider

meteorological drought, groundwater is a rarely considered variable despite its relevance for water supply during drought (Bachmair et al., 2016). For example, the U.S. Drought Monitor (https://droughtmonitor.unl.edu/) and the European Drought Observatory (https://edo.jrc.ec.europa.eu/) rely on data regarding precipitation, streamflow, soil moisture, and vegetation health, but, to date, not groundwater.

Reasons include a lack of readily available data for large-scale comparative drought analysis, uneven spatial distribution of observation networks, that is, clustering in highly productive aquifers, a general diversity of groundwater heads, and a variability strongly affected by human influences. However, for large-scale drought analysis regional estimates of groundwater dynamics are required. As groundwater head dynamics are typically very site specific an upscaling of local head anomaly observations considering different aquifers is often not reliable (Heudorfer et al., 2019; Kumar et al., 2016). An alternative source of information on the groundwater response to drought may be the baseflow in rivers, which can be taken as a proxy for shallow groundwater storage on a catchment scale (Brutsaert, 2008). Whereas groundwater head and catchment-scale baseflow may be affected by very local influences, the satellite-based Gravity Recovery and Climate Experiment (Tapley et al., 2004) is coarse (>200 km) and no data are available before the start of the mission in 2002. Due to these difficulties related to observations, groundwater has been stated to be one of the hydrological research fields where process-based modeling is particularly relevant for scientific progress (Fatichi et al., 2016).

Apart from observational data large-scale drought research often relies on outputs from continental-to-global scale land surface and hydrological models. However, large-scale models struggle to reproduce those surface water dynamics sufficiently that are governed by characteristic groundwater processes such as low flows (Gudmundsson et al., 2012; Stahl et al., 2011, 2012; Van Lanen et al., 2013; Van Loon et al., 2012). Van Loon et al. (2012) concluded that deficiencies in the simulation of storage processes partially prevent the correct representation of drought characteristics in large-scale models. Groundwater storage typically responds smoothed and delayed to precipitation or a lack of it. The magnitude of effects like pooling, attenuation, and lengthening of drought events depends on hydrogeological conditions (Bloomfield et al., 2015) and influence hydrological drought as well (Van Loon & Laaha, 2015). The deficiencies of large-scale hydrological models in the representation of groundwater storage processes prevented the correct simulation of these processes and hampered the investigation of the groundwater's reaction to drought on a large scale as well as large-scale projections of drought under climate change.

To overcome the aforementioned deficiencies explicit groundwater models on a large scale are required. Accordingly, there has been substantial effort to improve continental-to-global-scale groundwater modeling. Fan et al. (2013) was the first to present a global, high-resolution static water table depth map. de Graaf et al. (2015) presented the first global gradient-based groundwater model, including also lateral groundwater flows, which can transport water over large distances and be a relevant part of the water budget of receiving catchments (de Graaf et al., 2015; Krakauer et al., 2014; Schaller & Fan, 2009). Other modeling studies investigated large-scale patterns of water table depth, related to topography (e.g., for the continental United States: Condon & Maxwell, 2015; Condon et al., 2015) or aquifer properties (e.g., de Graaf et al., 2015; Sutanudjaja et al., 2011), and quantified worldwide anthropogenic impacts on groundwater depletion and river discharge (de Graaf et al., 2017; de Graaf et al., 2019). This progress was strongly facilitated by the development of global data sets on hydrogeological properties that become more and more available (e.g., Gleeson et al., 2014; Hartmann & Moosdorf, 2012; Huscroft et al., 2018). Additionally, several groundwater models in the last years have been set up on a national-to-continental scale (e.g., for the continental United States: Maxwell et al., 2015; New Zealand: Westerhoff et al., 2018; Europe: Trichakis et al., 2017). Some national models may run at higher resolutions or use more accurate parameterizations with the aim that they can provide more locally relevant data.

Continental-to-global-scale hydrological models have been useful to study the phenomenon of drought on a large scale. The new hydrogeological models on large scales may extend that capability allowing for studies on the influence of drought on groundwater on this scale. As a prerequisite these models would need to capture the spatially varying groundwater responses to precipitation anomalies (Bloomfield et al., 2015; Kumar et al., 2016; Leelaruban et al., 2017; Van Loon et al., 2017) which are strongly linked to differences in the magnitude of drought propagation features like attenuation and pooling. Hence, the delay of groundwater response to precipitation anomalies is vital for the correct simulation of drought propagation from the
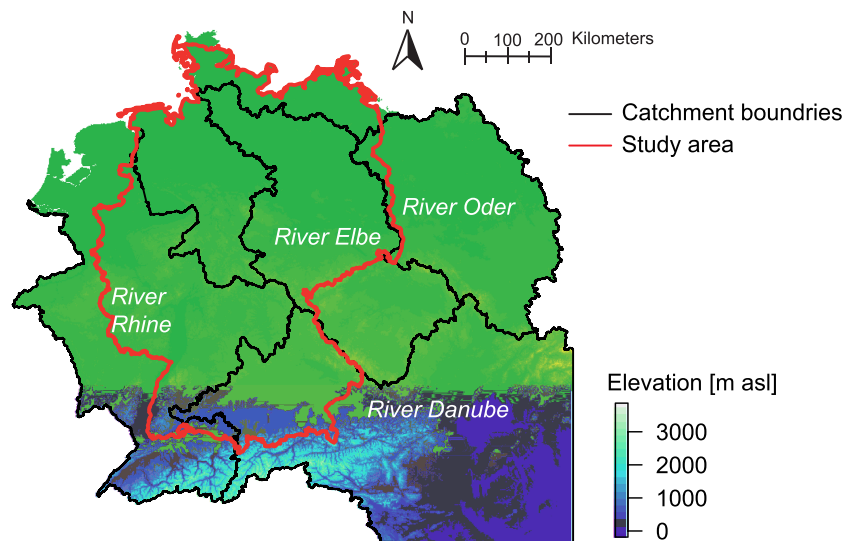
Figure 1. Location and elevation of the study area (Germany) and the model domain. The study area covers only a part of the modeled basins (produced using Copernicus data and information funded by the European Union—EU-DEM layers).

atmosphere through the pedosphere to the hydrosphere and here into surface and groundwater. Hellwig and Stahl (2018) have recently shown that catchment-scale groundwater-baseflow responses to precipitation anomalies are also variable, but highly relevant for the prediction of future drought vulnerability of surface and groundwater resources. The spatial variability in the baseflow data of this study also suggested that groundwater models on large scales are needed to simulate the processes of drought propagation from groundwater drought to hydrological drought.

The aim of this study was therefore to test how propagation and dynamics of groundwater responses to drought at country scale can be represented by a transient groundwater model for Germany. To do this, we first test the general ability of the groundwater model to represent the diversity of water table and river baseflow fluctuations based on different evaluation metrics and a sensitivity analysis. We then specifically investigate the diverse groundwater delay to precipitation anomalies during drought across different hydrogeological settings. Such an assessment may help to assess the potential of preferably simple large-scale groundwater models to be included into drought monitoring and early warning.

2. Methods

2.1. Study Area

For groundwater flow modeling it is important to consider entire aquifers since groundwater heads can be coupled throughout the aquifers. Water tables can also be influenced by neighboring aquifers, for example, by mountain block recharge from bedrock aquifers into porous valley aquifers. While groundwater flows do not necessarily follow exact topographic divides but rather reflect lithological boundaries, for practicality the model was set up for all large-scale river basins intersecting Germany (Figure 1), including the entire basins of River Rhine in the west and River Elbe and River Oder in the east. The southern part of Germany is part of the River Danube basin draining to the Black Sea. The River Danube basin was not completely modeled but cut several hundreds of kilometer downstream of Germany, where it can be assumed to have no more influence on modeled groundwater heads in Germany. This setup, which is larger in scale than the country-scale analysis domain, helps to ensure that all groundwater interactions relevant for groundwater dynamics in Germany can be considered.

2.2. Model Structure and Input Data

The groundwater model used for this study consists of one MODFLOW layer (Harbaugh et al., 2000), simulating groundwater heads in an unconfined aquifer layer, head-dependent groundwater-surface water interactions, and lateral groundwater flows (i.e., flow between the grid cells). This setup is similar to a previously

developed global-scale groundwater model (de Graaf et al., 2015). Grid cells outside the simulated catchments were either classified as no flow (on shore) or as 0 m above sea level constant head (at sea and at the cut of the River Danube basin). The spatial resolution of the model is approximately 1 km (latitudinal: 1/22°, longitudinal: 1/14°) with grid cell area decreasing northward. The simulation for the resulting 900,000 active grid cells was run from 1970 to 2009 (this time span was covered by most of the available evaluation gauges) in weekly time steps.

Model grid cell elevations were derived from a 30-m digital elevation model freely available from the European Environment Agency (2013). Groundwater discharge was routed according to flow directions derived from the minimum values of the digital elevation model per model grid cell ensuring a realistic drainage map for the modeled area.

2.2.1. Aquifer Properties

A main source of uncertainty in groundwater models are aquifer properties (Howard & Griffith, 2009). The GLobal HYdrogeology MaPS (hereafter called GLHYMPS) of Gleeson et al. (2014) as well as the updated version GLHYMPS 2.0 (Huscroft et al., 2018) provide global information on the permeability and porosity and are based on the lithology map of Hartmann and Moosdorf (2012). For Germany, additionally the "Hydrogeologische Übersichtskarte" (Hydrogeological Map of Germany at a resolution of 1:200,000, hereafter called HÜK200; BGR & SGD, 2016) is available, containing information on the potential range of hydraulic conductivity of the subsurface. From the logarithmic mean of these ranges a second country-scale estimate of permeability was derived for Germany. This data set was used to derive hydraulic conductivities for grid cells within Germany. To test the model's sensitivity, comparative model runs using the different estimates of hydraulic conductivity in Germany while keeping all other model parameters fixed were performed (i.e., a "one-at-a-time sensitivity analysis"; Pianosi et al., 2016). The hydraulic conductivity for the grid cells outside Germany as well as the specific yield values for all grid cells in both model versions were derived from permeability and porosity estimates from GLHYMPS. As the porosity is usually slightly higher than specific yield this approximation is rather a small overestimate.

In general, hydraulic conductivity changes with depth. A common approach is to divide the subsurface into several layers/aquifers of a certain thickness and to assign an appropriate uniform conductivity to each layer. To keep the model simple, we use homogeneously only one layer but adjust its conductivity based on the assumption of an exponentially decreasing hydraulic conductivity over depth, following previous studies' approach for a depth distribution (Beven & Kirkby, 1979). Similar to other large-scale groundwater studies the decrease is calculated using the formula of an e-folding depth f (Fan et al., 2007; Miguez-Macho et al., 2008) with the parameters of de Graaf et al. (2015). f is inversely related to terrain slopes z:

$$f = \max\left(\frac{100}{(1 + 150 \ast z)}, 5\right) \tag{1}$$

where the e-folding depth f is set to a minimum of 5 m. Then, the transmissivity T in a grid cell depends on f, the initial values for hydraulic conductivity k_0 (derived from HÜK200 and GLHYMPS), and the water table depth d_{gw}:

$$T = \int_{d_{gw}}^{D} k_0 e^{\frac{-z'}{f}} dz' \tag{2}$$

where D is a maximum aquifer thickness (set here to 100 m) and z' is the depth below surface. Since k strongly declines in steep terrain areas, in these regions the effective aquifer thickness that contributes to groundwater head dynamics is smaller than the maximum aquifer thickness. As the conductivities derived from GLHYMPS are already averages for the aquifer's upper 100 m, k_0 was rescaled to ensure a correct average conductivity in every cell despite the decline of conductivity with depth assumed here (see supporting information). Transmissivity values in the model were updated every modeling time step according to the current water table depth.

2.2.2. Groundwater Interaction With Surface Water

Interactions between surface water and groundwater were implemented using the RIV-package. Each cell contains one of two potential river types, corresponding to large rivers (width >10 m) with two-way groundwater-surface water interactions and small streams with groundwater drainage only. Based on

average streamflow discharge in a cell ($\overline{Q}$ [L^3/T]), for every grid cell a channel width, channel depth, average river head, and river bed conductance was calculated (de Graaf et al., 2015; Sutanudjaja et al., 2011). $\overline{Q}$ was calculated for every grid cell from routed average baseflow Q_{bf} of a previous run using dynamic initialization. To transform average simulated baseflow into average discharge, the ratio $\overline{Q}/\overline{Q_{bf}}$ was calculated based on 338 evaluation gauges. This led to the empirical estimate:

$$\overline{Q} = 1.8 * \overline{Q_{bf}} \tag{3}$$

where the empirical factor 1.8 corresponds to an average baseflow index of 0.56.

Channel width W_{chn} [L] was obtained utilizing Lacey's empirical formula:

$$W_{chn} = 4.8 * Q_{bkfl}^{0.5} \tag{4}$$

where Q_{bkfl} is the bankfull discharge, occurring on average every 1.5 years and 4.8 [T$^{0.5}$/L$^{0.5}$] is an empirical factor. Corresponding to the procedure to estimate $\overline{Q}$ (equation (3)) an empirical factor of 11 was calculated to estimate Q_{bkfl} based on simulated $\overline{Q_{bf}}$.

For large rivers (width >10 m) channel depth D_{chn} [L] was calculated using Manning roughness coefficient n [T/L$^{1/3}$] (set to 0.045 s m$^{-1/3}$) and channel longitudinal slope S [−]:

$$D_{chn} = \left(\frac{n * Q_{bkfl}^{0.5}}{4.8 * S^{0.5}} \right)^{\frac{3}{5}} \tag{5}$$

The average river head above river bed (H_{riv} [L]) was subsequently calculated according to

$$H_{riv} = \left(\frac{n * \overline{Q}^{0.5}}{W_{chn} * S^{0.5}} \right)^{\frac{3}{5}} \tag{6}$$

River bed conductivities c [L^2/T] both for large and small rivers were calculated as proportional to W_{chn}:

$$c = \frac{1}{B_{res}} * W_{chn} * L_{chn} \tag{7}$$

where the channel length L_{chn} [L] was approximated by the diagonal cell length (following Sutanudjaja et al., 2011; de Graaf et al., 2015). Lacking any regional estimates, river bed resistances B_{res} [T] were uniformly assumed to be one day, neglecting a potentially large variability of this parameter. To prevent unrealistic small conductivities for very small streams, a minimum of 10 m was taken for W_{chn} in this formula.

In our model baseflow occurs when groundwater heads H_{gw} [L] in a grid cell are higher than the average river head H_{riv} (large rivers), respectively, surface elevation E [L] (small streams). Rates of surface water groundwater interaction are proportional to streambed conductivity c and the difference of H_{gw} and H_{riv} (respectively, E):

$$Q_{bf} = \begin{cases} c * (H_{gw} - H_{riv}) & \text{if} \quad W_{chn} > 10 \\ c * (H_{gw} - E) & \text{if} \quad W_{chn} \leq 10 \end{cases} \tag{8}$$

For the large rivers groundwater infiltration from the river is possible if H_{gw} drops below H_{riv}. Groundwater infiltration is also proportional to c and the difference of H_{gw} and H_{riv} but does not further increase if H_{gw} falls below river bed elevation.

2.2.3. Groundwater Recharge

The meteorological forcing for the model was derived from the European Climate Assessment & Dataset (Haylock et al., 2008, data set hereafter called E-OBS). Hellwig et al. (2018) found that this data set is highly correlated to a high-resolution precipitation data set that covers Germany, that is, the part of the modeled area which is further investigated in the following section. Even though there might be some small-scale

variability outside of Germany not captured by E-OBS (for example in the Alps), the meteorological forcing (combined with an elevation correction; see next paragraph) is an appropriate estimate for our purpose for the entire model domain. Based on maximum and minimum temperatures (T_H and T_L) potential evaporation E_{pot} was estimated following Hargreaves and Samani (1985):

$$E_{pot} = 0.00023*R*(T_H - T_L)^{0.5}*\left(\frac{(T_H + T_L)}{2} + 17.8\right)$$ (9)

where R denotes the extraterrestrial radiation, derived from latitude and month of the year.

As grid cells of E-OBS did not match model grid cells, the meteorological forcing (precipitation P, average temperature $\overline{T}$, E_{pot}) was re-gridded, that is, interpolated to the model grid cells. The resolution of E-OBS is low compared to the model. Therefore, a bias correction was performed using high-resolution gridded averages of the meteorological variables for Germany. There are several appropriate data sets available for Germany; we selected two of them for their high resolutions comparable to our model. For P and E_{pot}, maps of the "Hydrologischer Atlas Deutschlands" (Hydrological Atlas of Germany with maps at a resolution of 1:2,000,000, hereafter called HAD; BMU, 2003) were used; long-term averages of $\overline{T}$ were derived from yearly $\overline{T}$ maps at a resolution of 1 km^2 of the German Meteorological Service (DWD) for the period 1970–2009. Biases of P and E_{pot} of German grid cells were removed by multiplying the records with the ratio between the long-term E-OBS average and the HAD estimate, for $\overline{T}$ the difference between the long-term E-OBS average, and the DWD estimate was subtracted from the record. For grid cells outside Germany, where no high-resolution estimates are available from HAD and DWD, we used the linear relationships between biases and elevation within Germany to estimate and remove the biases.

Weekly groundwater recharge was modeled for each grid cell including snow and soil storages. The plant available water, which for Germany is taken from the HAD, was used as an estimate for the size of the soil storage. For grid cells outside Germany it was set to the average value (i.e., 123.25 mm).

In the recharge model, P fills either the snow or the soil storage, depending on $\overline{T}$ and the threshold temperature T_{thresh} (set to 0 °C). If $\overline{T}$ is above T_{thresh} and water is currently stored as snow, the soil storage is additionally filled by snow melt M:

$$M = \left(\overline{T} - T_{thresh}\right)*f_{deg}*t$$ (10)

where f_{deg} is the degree-day factor for snow melt (set to 1.8 mm d^{-1} K^{-1}) and t is the time step of the model (7 days). Actual evapotranspiration is limited both by E_{pot} and water availability in the storages. If the soil storage is below net field capacity only the ratio between current storage and net field capacity multiplied with E_{pot} and a soil factor (set to 0.8) can evaporate. Thus, with a lower fill of the soil storage the ratio between actual and potential evaporation decreases proportionally. Only if the soil storage is filled above field capacity, the storage excess drains as groundwater recharge. The application of a soil factor increases the number of recharge events and thus helps to account for additional recharge like it can occur through macropore recharge. As the model ignores other potential flow paths like interflow or surface runoff, the groundwater recharge is overestimated in most areas. This bias was corrected according to the procedure for the meteorological input data, using the high-resolution long-term average groundwater recharge map for Germany from the HAD.

The recharge model was coupled offline to MODFLOW, using time series of modeled recharge as MODFLOW's upper boundary condition. Details of the coupling which depend on our specific modeling environment (PCRaster Modflow which uses MODFLOW 2000) are described in the supporting information.

2.3. Evaluation

Most studies on continental-to-global-scale groundwater models use borehole observations for evaluation (e.g., de Graaf et al., 2017; Fan et al., 2013; Trichakis et al., 2017). However, groundwater dynamics from borehole observations in the upper (shallow) aquifer can be very site-specific and vary over short distances (e.g., Heudorfer et al., 2019; Kumar et al., 2016). On the contrary, modeled groundwater depths are grid cell
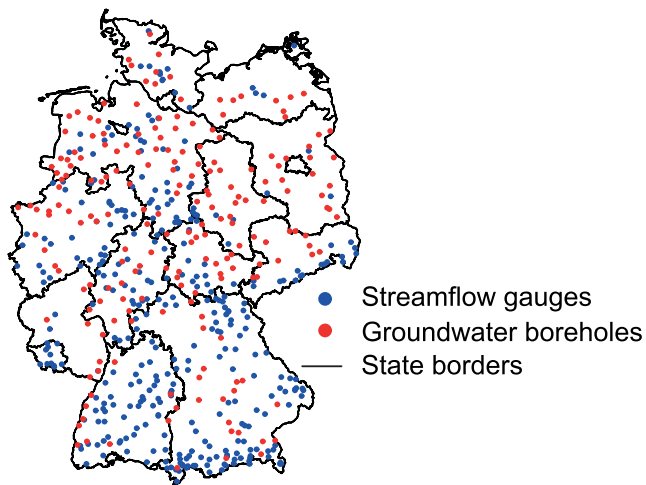
Figure 2. Locations of streamflow gauges and groundwater boreholes used for the evaluation.

averages ignoring subgrid variability, thus requiring regionally representative groundwater observations which are rarely available. Moreover, groundwater boreholes are most often located in larger porous aquifers impeding the evaluation of groundwater models in areas with less productive aquifers.

Another approach to evaluate a transient groundwater model is to focus on groundwater outflow, that is, baseflow. Unfortunately, baseflow is not directly observable on the scale of model grid cells. As an estimate of baseflow at the catchment scale, baseflow can be separated from streamflow observations as the delayed flow component. However, there are also other potential sources than groundwater outflow for delayed flow like snow melt, lake outflow, or anthropogenic sources of discharge. To reduce uncertainty of baseflow separation, streamflow observations of near-natural catchments are preferable, leading in Germany to a bias toward headwater catchments. On the other hand, baseflow provides an integrated measure of groundwater head fluctuations on a larger scale than borehole point data and might therefore be more comparable to modeled grid cell averages.

To balance the (dis-)advantages of different evaluation approaches, in this study a two-step evaluation based on groundwater borehole observations and separated baseflow observations was applied. Both groundwater boreholes and streamflow gauges used for evaluation were located across the study area (Figure 2).

2.3.1. Evaluation Using Groundwater Boreholes

For the evaluation with borehole observations 202 time series from all over Germany were used (Figure 2). Only records with data for more than 90% of the modeling period (1970–2009) were considered. To ensure near-natural behavior of the groundwater heads as well as a good coverage of the study area with a similar density of observation points, stations were selected from a data set which was specifically created for this purpose and has already been used for another study on drought by Kohn et al. (2014). The borehole time series differ in temporal resolution and were all averaged to monthly resolution.

Both modeled and observed groundwater time series were standardized allowing for comparisons between grid cells. The procedure of the Standardized Groundwater Index (SGI; Bloomfield & Marchant, 2013) was used for all standardizations in this study. This nonparametric transformation ensures standardized normal distributed values for all 12 months within the record.

As this study focuses on the transient simulation of groundwater drought periods, three corresponding evaluation metrics were chosen:

1. The Pearson correlation coefficient between simulated and observed standardized time series.
2. A temporal agreement index (TAI) measuring the agreement of two binary time series I_{obs} and I_{sim} (Stahl et al., 2011; Tallaksen & Stahl, 2014). These indicator time series were derived from observed and simulated times series by using a constant threshold denoting drought or normal/wet conditions. The TAI is calculated based on the observed agreement of the binary time series d (i.e., the number of observations with both time series indicating drought):

$$TAI = \frac{\left(\frac{d}{m}\right) - p^2}{p - p^2} \tag{11}$$

where m is the total number of observations and p is the proportion of observations denoted as drought (i.e., the threshold level). As a threshold we used 20-percentile here, that is, the value exceeded for 80% of the time.

3. The difference of simulated and observed precipitation accumulation times that have the maximum correlation with groundwater anomalies (T_{max}). To calculate T_{max} (Figure 3) by grid cell, precipitation and groundwater time series were standardized according to the procedure of the SGI. Referring to the original parametric standardization of precipitation time series (Standardized Precipitation Index (SPI); McKee et al., 1993) the precipitation time series were standardized for different accumulation periods
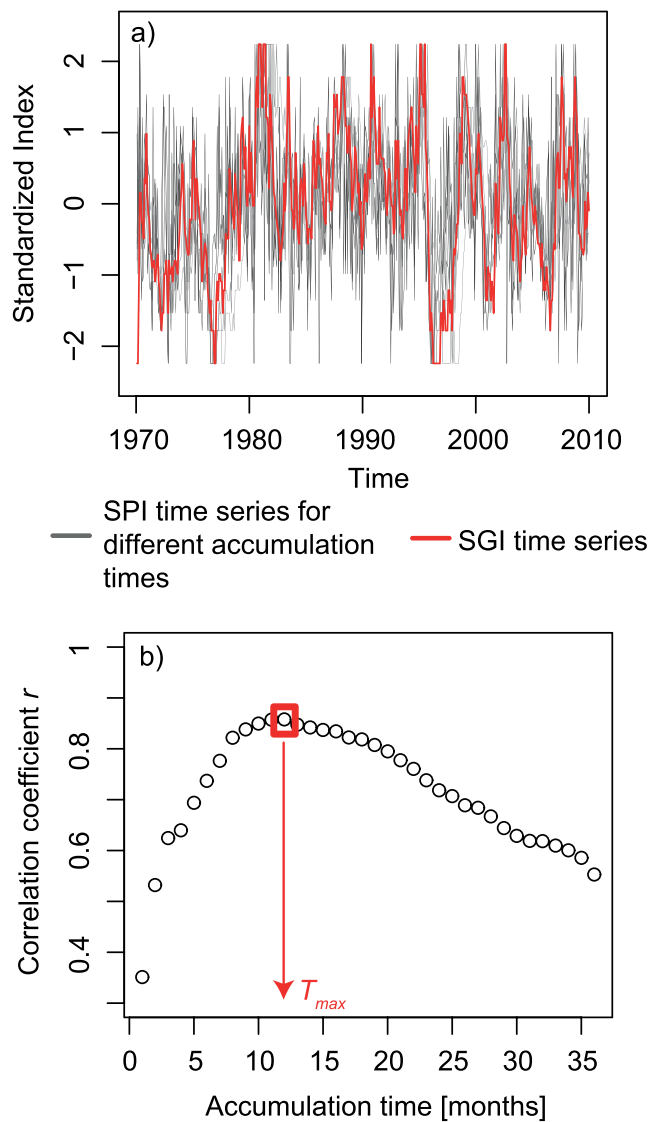
Figure 3. Calculation of the precipitation accumulation time that has the maximum correlation with groundwater anomalies (T_{max}). (a) Standardization of groundwater time series (SGI) and accumulated precipitation times series for different accumulation times (SPI-n). (b) Correlation of all candidate SPI-n with SGI and selection of the accumulation time with maximum correlation as T_{max} (in the example $T_{max} = 12$ months). This selection procedure was performed for every model grid cell.

called SPI-n here, where n stands for the accumulation period in months (ranging from 1 to 36). Then, SGI time series were correlated with all SPI-n time series and the n with the highest correlation was selected as T_{max} (Figure 3; see also Barker et al., 2016; Bloomfield & Marchant, 2013).

The performance of the model was further explored by relating these three evaluation metrics to the four model input parameters hydraulic conductivity, elevation, slope, and specific yield in terms of scatterplots and Pearson correlation coefficients.

2.3.2. Evaluation Using Baseflow

A total of 338 streamflow observation gauges within the study area of Germany was used for baseflow separation (Figure 2). The data set, which was described by Hellwig and Stahl (2018), comprises daily streamflow data from near-natural headwaters smaller than 200 km^2. Baseflow separation from streamflow always contains uncertainties and there are lots of separation techniques available. Here we use a common separation scheme based on a joining of local minima of the hydrograph (Gustard & Demuth, 2008); however, other approaches might be equally valid. Observations of baseflow dynamics were compared to simulated baseflow which was routed according to surface flow directions (see section 2.2). Routing was implemented without time lags as catchments are small and thus travel times can be assumed to be smaller than the temporal resolution of seven days.

For every gauge the model grid cell representing the same stream was determined. Due to model discretization in some cases the stream was not modeled in the grid cell of the gauge's location, but in a neighboring cell. So the original grid cell plus the eight neighboring cells were designated as candidate cells for the corresponding model output. To match the right candidate cell to observations, the cell which had the most similar catchment area based on the baseflow routing compared to the area documented for the gauge was selected as the corresponding one. Like the evaluation using groundwater boreholes, corresponding observed and modeled baseflow time series were standardized and subsequently analyzed using the three evaluation metrics.

2.4. Drought Analysis

The model was used to contrast the effects of two drought events which differ regarding the time scale of the anomaly. A drought event is defined here as a period of negative anomalies in a time series of SPI for a certain accumulation period. Due to differences in the accumulation time of maximum correlation between precipitation and baseflow anomalies across the study area (Hellwig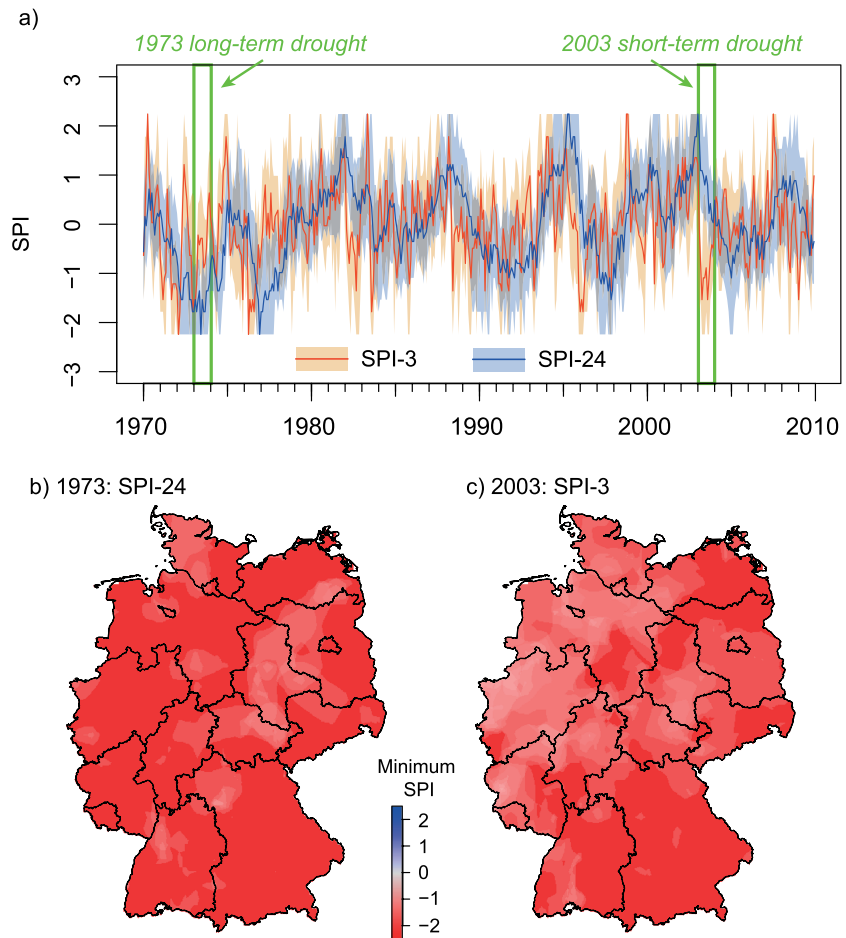 & Stahl, 2018), groundwater resources were expected to show different vulnerabilities to short- and long-term drought events. To compare the different drought time scales, two benchmark drought events were selected from standardized precipitation records of all model grid cells within Germany. Based on the SPI-3 the drought of 2003 was selected as the short-term benchmark drought that influenced all of Germany and was not superimposed on a long-term drought (Figure 4a). The summer drought of 2003 followed a very wet winter; thus, based on SPI-24 the year of 2003 was in fact rather wet. For a long-term drought we selected the event of 1973 based on the SPI-24. Also, SPI-3 was moderately low in 1973 with precipitation continuously below the long-term average. Despite some regional differences in severity all of Germany was affected by both selected drought events (Figures 4b and 4c).

To assess the impact of the meteorological drought periods on water tables, standardized water table depths of every grid cell were averaged for the drought period as an indicator of drought severity. The relationship

Figure 4. Meteorological drought events in Germany. (a) The 90% range (shaded area) and median (bold line) of SPI-3 and SPI-24 for all grid cells covering Germany. (b and c) Spatial distribution of minimum SPI for selected drought years.

between water table depth during drought and $T_{\max}$ as well as potential driving factors was analyzed with the Pearson correlation coefficient.

3. Results

3.1. Model Performance

Moderate correlation coefficients indicate that the groundwater model was able to reproduce the general dynamics of standardized groundwater head and baseflow observations in many locations (Figure 5). More detailed analyses revealed higher correlation coefficients for regions with higher hydraulic conductivity and lower slopes (Figure S1 in the supporting information). For some of the baseflow observations no modeled baseflow could be matched (e.g., due to karst springs), but the median correlation of all records was slightly higher for the evaluation using baseflow (correlation coefficient $r = 0.40$ compared to $r = 0.39$ for the evaluation using groundwater borehole observations). Particularly for baseflow, correlations strongly decrease with elevation (Figure S1). In higher elevations snow becomes a more important potential source for delayed flow and hence baseflow separation is more uncertain in these regions. For the groundwater head evaluation, the spread in correlations is much larger than for the baseflow evaluation with some observations even negatively correlated to simulations. A potential reason is the subgrid variability that might be relevant for a part of the cells. Consistent to that, the model had less power to simulate the anomalies in mountainous regions, where aquifers are mostly fractured rocks and the model cannot capture the small but important valley aquifers which are often smaller than model resolution.
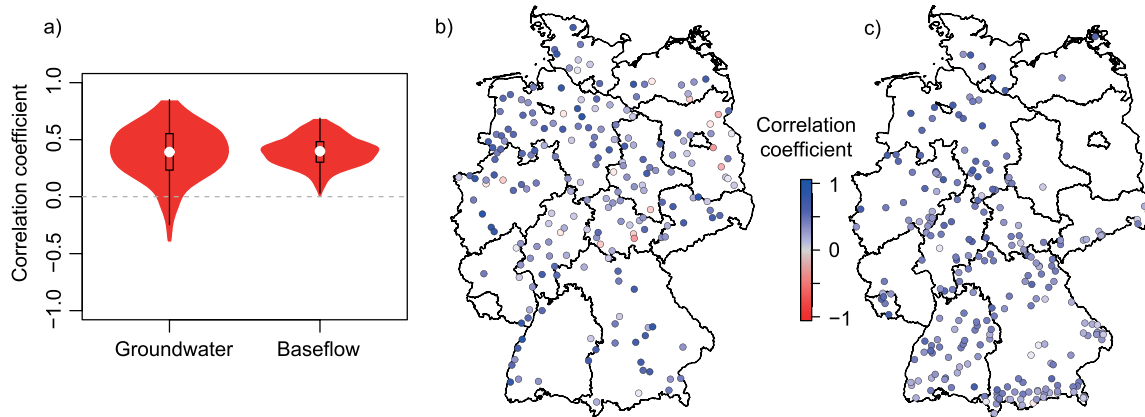
Figure 5. Pearson correlation coefficients between modeled and observed groundwater heads and baseflow. (a) Violin plots illustrating the density distribution of correlations. (b) Spatial distribution for groundwater heads. (c) Spatial distribution for baseflow. Boxes represent the quartiles 0.25 and 0.75 and whiskers last to the most extreme value within a maximum distance of a 1.5-fold interquartile range.

The TAI showed similar results for the evaluations with groundwater borehole observations (median TAI = 0.25) and baseflow observations (median TAI = 0.28). Again, there was considerable spread with some gauges having even negative TAI, that is, no temporal agreement (Figure 6). TAIs were negatively correlated with both elevation (for baseflow) and slope (Figure S2), indicating again a poorer model performance in the mountainous regions of Germany.

Both simulations and observations indicated a much higher diversity of $T_{\max}$ across Germany for groundwater than for baseflow. As $T_{\max}$ for baseflow were only few months in almost the entire study area (Figure 7), the mean absolute error of estimated $T_{\max}$ was low (median = 4.68 months) for baseflow compared to groundwater (median = 10.29 months). None of the potential explanatory factors was clearly related to the difference of simulated and observed $T_{\max}$ for baseflow, but the map indicated rather an overestimation by the model in northern Germany and an underestimation in the southern part (Figure 7f). On the contrary, there was no such spatial pattern for the difference in observed and simulated $T_{\max}$ for groundwater (Figure 7c). However, these differences for groundwater were positively correlated to specific yield and negatively correlated to elevation (Figure S3), so $T_{\max}$ was overestimated by the model in the porous aquifers in the lowlands and underestimated in higher elevations.

All results shown so far are based on the model with the conductivity parameterization from the HÜK200 data. The one-at-time sensitivity analysis using hydraulic conductivities solely derived from GLHYMPS
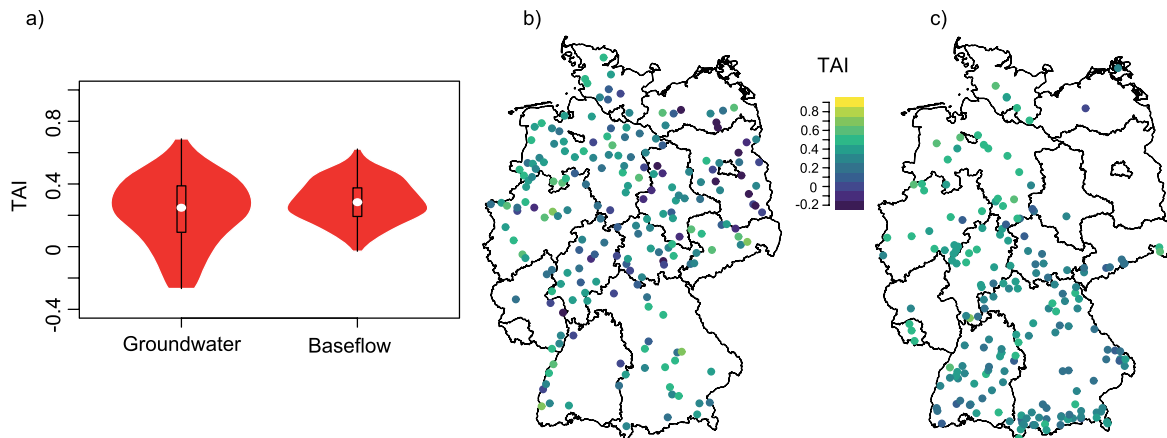

Figure 6. TAI between modeled and observed groundwater heads and baseflow. (a) Violin plots illustrating the density distribution of TAI. (b) Spatial distribution for groundwater heads. (c) Spatial distribution for baseflow. Boxes represent the quartiles 0.25 and 0.75 and whiskers last to the most extreme value within a maximum distance of a 1.5-fold interquartile range.
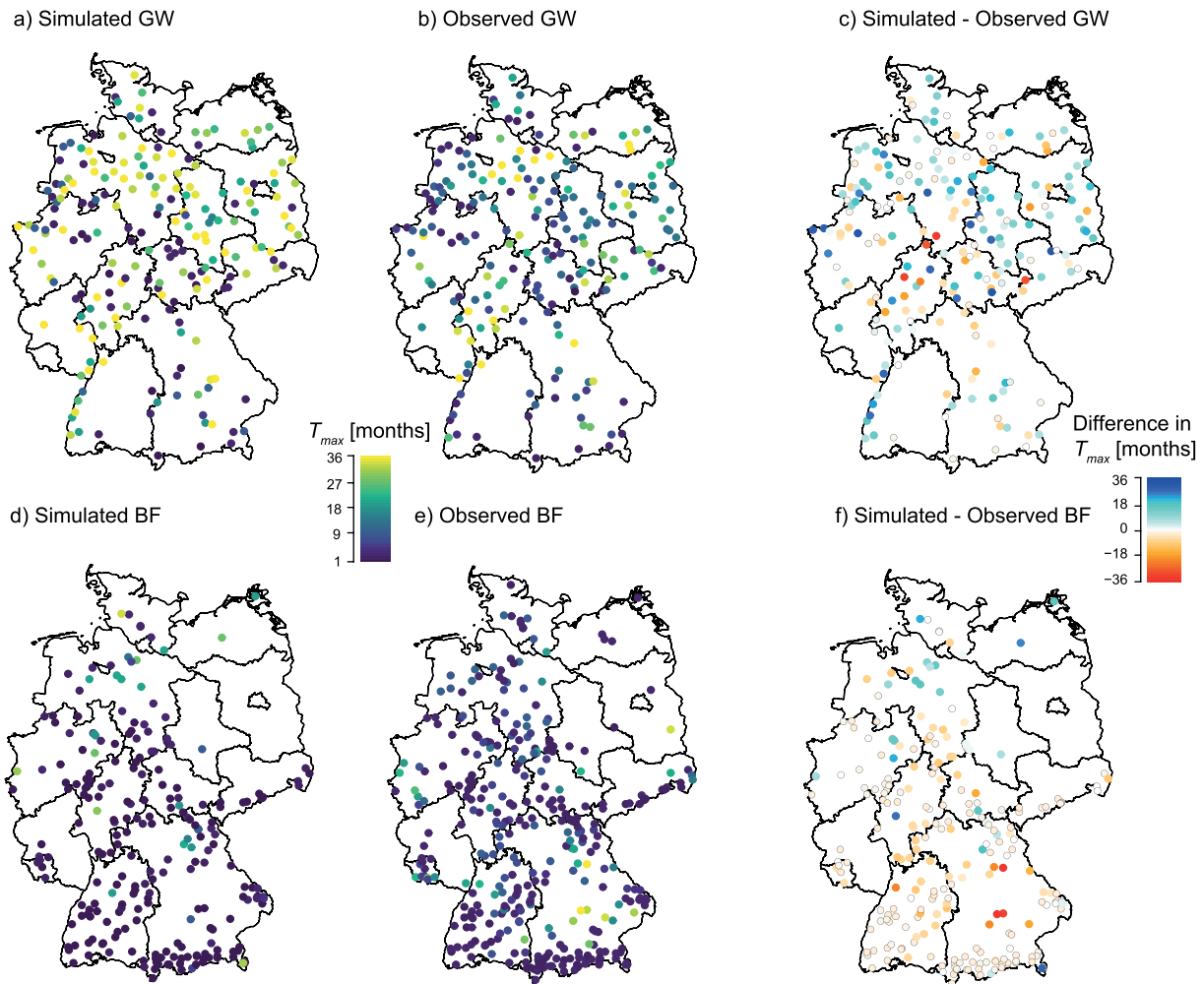
Figure 7. Spatial distribution of T_{max} for the evaluation based on (a–c) groundwater and (d–f) baseflow. T_{max} in (a) and (d) are based on the model and in (b) and (e) on observations. (c) and (f) show the difference of simulated minus observed T_{max}.

revealed an important influence of this parameter. For all evaluation metrics (apart from TAI for baseflow) model performance of the GLHYMPS based model was lower compared to the model based on HÜK200 (Table S1). Hence, using a national conductivity data set instead of a global one improved the results of the groundwater model. Nevertheless, the general relationships between model performance and different factors remained similar for both parametrizations of the hydraulic conductivities (Figures S1–S3).

3.2. Drought Analysis

The drought analysis was performed with the model based on HÜK200. According to the model, the two benchmark drought events 1973 and 2003 led to very different patterns of standardized groundwater heads (Figure 8). During the long-term 1973 drought, groundwater heads were below average across most of the study area (Figure 8, left column). Most pronounced groundwater head declines due to the drought occurred in the largest porous aquifers that are in northern Germany and the river valleys of large streams such as River Rhine and the Danube, whereas the Central German Uplands faced only a moderate drought. On the contrary, from a groundwater perspective the short-term 2003 drought was not that extreme due to the wet winter 2002/2003 (Figure 8, right column). While the groundwater heads of the large alluvial aquifers were above normal in most regions, the short term-drought was more severe for the Central German Uplands. In these regions, within few months the groundwater reacted with strong anomalies whereas in the alluvial aquifers groundwater anomalies reach a minimum not until the year 2004.
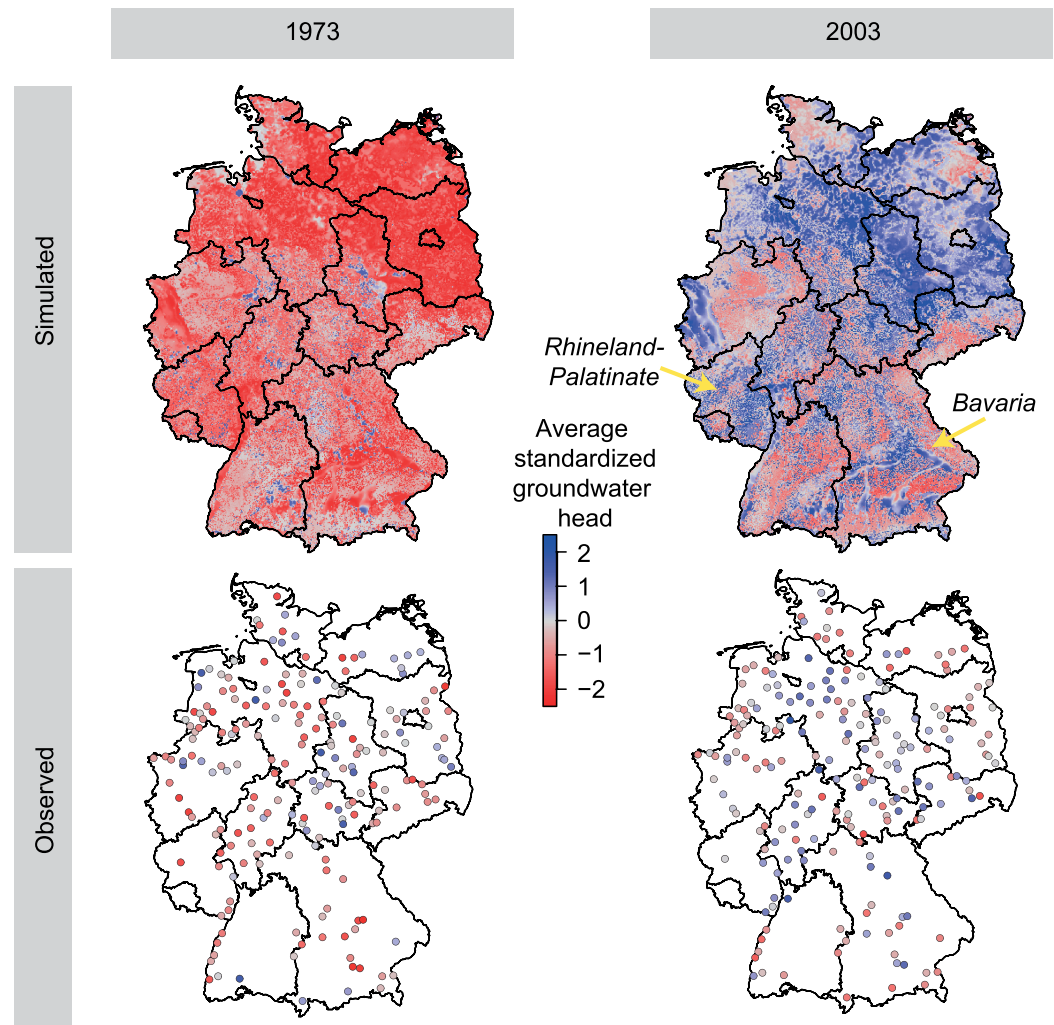
Figure 8. Modeled and observed average standardized groundwater heads across the study area in 1973 and 2003.

The diverse groundwater responses to meteorological droughts on different time scales coincided with patterns of T_{max} across the study area. In general, the Central German Uplands with small aquifers have relatively short T_{max} of few months, whereas the larger alluvial aquifers in the lowlands have much longer T_{max} of up to several years (Figure 9). Correlations between SPI-n and SGI is high for most regions with an average of 0.71 (Figure 9b). There is no relationship ($r = 0.02$) between T_{max} and the correlation between SPI-n and SGI (Figure 8b).

T_{max} was related to the different parameters used in the model. The most important predictor for T_{max} is hydraulic conductivity, with higher T_{max} for higher hydraulic conductivities (Figure 10). The high influence of this parameter on T_{max} demonstrates its importance for the transient modeling of groundwater dynamics. Additionally, T_{max} is higher for grid cells in lower elevations and with higher specific yield (Figure 10).

The different reactions to short- and long-term droughts across Germany were strongly related to modeled T_{max}. Groundwater heads in regions with short T_{max} were highly fluctuant over time with multiple but short-term drought events (Figure 11). Thus, short-term drought events were able to cause severe groundwater droughts as well. On the contrary, groundwater heads in regions with long T_{max} were much less fluctuant, showing only few but longer-lasting droughts. The two regions depicted in Figure 11 differ in many regards including aquifer type, elevation, slope, hydraulic conductivity, and specific yield. All these differences are reflected in the very different T_{max} (29.1 ± 9.8 months for the northern region, 7.6 ± 12.5 months for the southwestern region) governing head fluctuations and the reaction to meteorological drought.
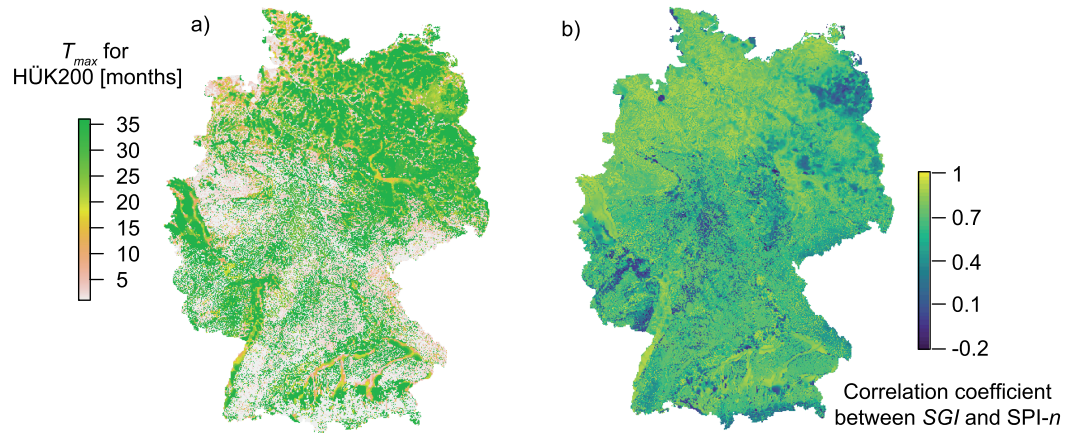
Figure 9. T_{max} across the study area. (a) T_{max}. (b) Maximum correlation between modeled SGI and SPI-n.

The differences in drought response due to T_{max} were statistically significant correlated for both 1973 and 2003. T_{max} explained the observed spread in average SGI (Figure 12) with decreasing SGI for higher T_{max} in 1973 ($r = -0.52$) and increasing SGI for higher T_{max} in 2003 ($r = 0.81$).

4. Discussion

The groundwater model simulated diverse water table fluctuations across Germany, which were related to aquifer characteristics such as hydraulic conductivity and storage properties. Fluctuations in the large alluvial aquifers in the lowlands were smoother compared to the small aquifers in the Uplands leading to modeled water table time series of different characteristics across the study area. This coincides with a large diversity of groundwater dynamics reported for borehole observations in different regions (e.g., Bloomfield et al., 2015; Haas & Birk, 2017; Heudorfer et al., 2019). Accordingly, T_{max} were found to vary from few months to several years. These results match expectations and coincide with previous findings
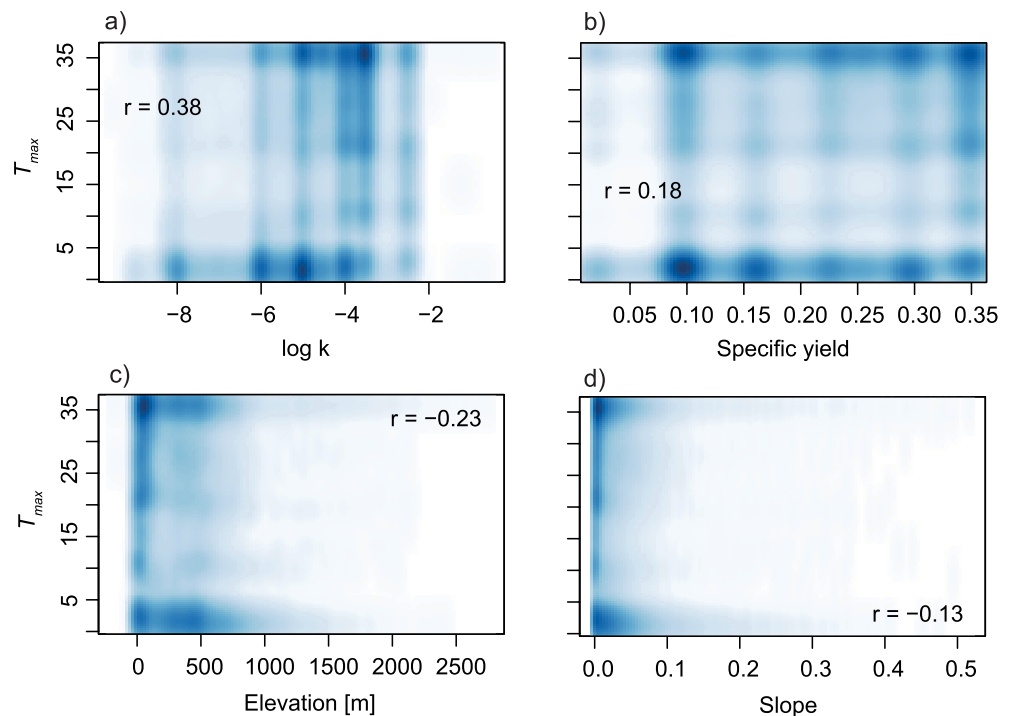

Figure 10. Relationship between modeled T_{max} and model parameters for all grid cells in Germany.
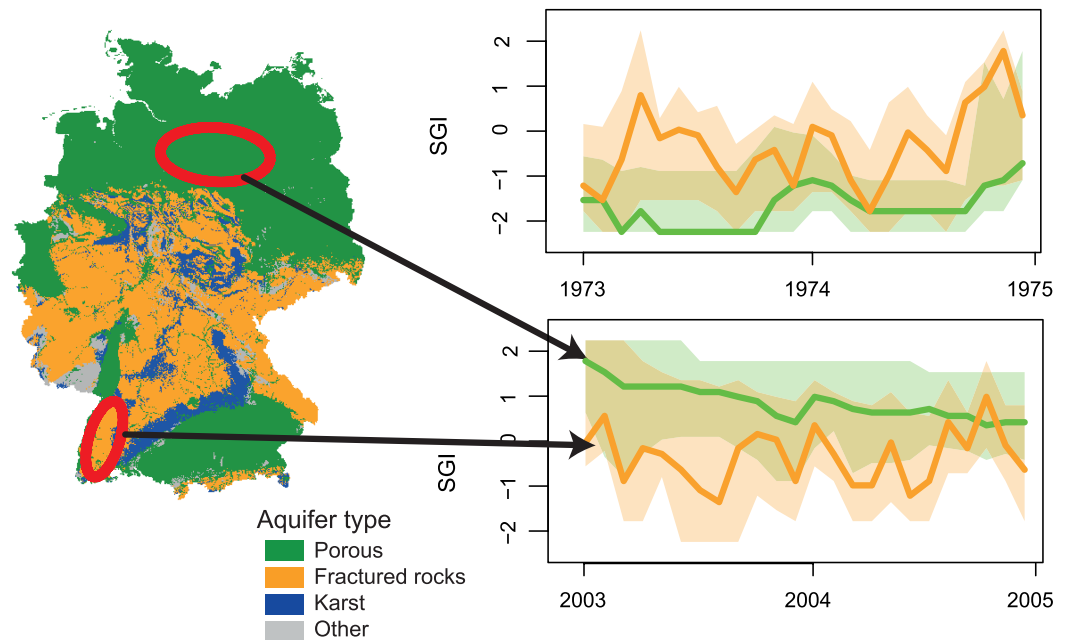
Figure 11. Modeled standardized groundwater heads for 2003 to 2005 and 1973 to 1975 in two regions with distinct groundwater characteristics and T_{max}. The median per region is drawn as a line, and the faint area indicates the 90% range. Map based on BGR and SGD (2016).

based on statistical analysis (Hellwig & Stahl, 2018; Kumar et al., 2016). Heudorfer and Stahl (2017) showed considerable variation in the drought propagation into groundwater head observations of nearby wells within the same basin related to very localized controls that hence do not always follow the theoretical propagation logic that (simplified) models provide. The opposite patterns simulated for the benchmark droughts demonstrate the model's ability to capture distinct characteristics of drought propagation due to hydrogeological differences. The model results are further supported by reports on the 2003 drought event. In Bavaria differences in the groundwater response for different hydrogeological units were observed (Bayrisches Landesamt für Wasserwirtschaft, 2005), coinciding with the differences simulated for this state (Figure 8): there were severe drought conditions simulated in the Alps in the south, the crystalline Bavarian Forest in the east, and the slightly permeable Alpine Foreland in the southern half, whereas the highly permeable quaternary sediments along the rivers and between the Alps and the Alpine Foreland did not show extremely low anomalies. In Rhineland-Palatinate there was no severe groundwater drought reported for 2003 (LfW, 2004) corresponding to our model results as well (Figure 8). Thus, a relatively simple model setup (one groundwater layer, static surface water heads) can simulate realistic patterns of water table fluctuations
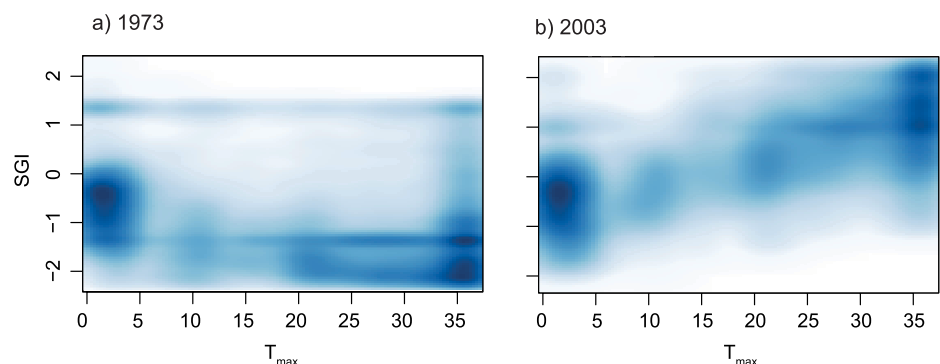

Figure 12. Smoothed density plot of average SGI against T_{max} for (a) 1973 and (b) 2003.

and may therefore improve the range of variability of climate responses of surface and groundwater simulated in large-scale hydrological models currently lacking an adequate representation of the groundwater system (e.g., Gudmundsson et al., 2012; Van Lanen et al., 2013).

The model evaluation revealed some regions, which had a poor model performance. Especially in mountainous regions with local valley aquifers and an important topographic subgrid variability modeling results should be taken with caution. In these regions grid cells are often dominated by fractured bedrock aquifers so that the important porous valley aquifers are not sufficiently depicted. With a more flexible model structure, like a variable grid cell geometry which is made possible in the latest MODFLOW version, and/or improved regional parametrization model performance could probably be increased in these areas. The evaluation with both streamflow and groundwater observations allowed a detailed assessment of the model. Particularly in mountainous regions groundwater head data are usually rarely available, so without an evaluation using baseflow a groundwater model might look better than it really is.

All large-scale groundwater models entail many uncertainties related to model structure, parameters, and boundary conditions. There are many strong simplifications in the model used in this study (e.g., constant maximum aquifer thickness, uniform slope depending depth distribution of k, uniform river bed resistance). Also, our recharge model was highly simplified; however, despite some arbitrary choices for our purposes these simplifications were appropriate. As we focus our analyses on standardized groundwater head anomalies which occur only in the uppermost part of the aquifer, the influence of simplifications such as a constant maximum aquifer thickness is limited. Anyway, the more detailed simulation of the upper boundary or even the online coupling of a groundwater model to a complete hydrological model (i.e., feedback between groundwater heads and other parts of the hydrological cycle are explicitly considered) as well as are more detailed characterization of the aquifers (confined versus unconfined; aquifer thickness; more than one layer) are likely to further improve the model. However, as the most relevant regional differences were already depicted and a more complex model structure largely increases both efforts for model design and computational costs, the specific needs should be carefully considered before model setup.

The results indicate that the hydraulic conductivity is a particularly important parameter for modeled T_{max}. Aquifers with a high hydraulic conductivity usually have smaller gradients of the water table and hence have a larger dynamic storage. A larger dynamic storage allows for more attenuation of recharge anomalies which explains the longer T_{max}. Previous studies also found a high importance of k for simulated water table heads (de Graaf et al., 2015; Foster & Maxwell, 2019; Keune et al., 2016; Sutanudjaja et al., 2011; Westerhoff et al., 2018). However, uncertainties related to this parameter are large, as we found relevant differences between the two data sets used in this study for large parts of the study area. We demonstrated that a national available data set on this parameter can outperform a larger scale one. As the parametrization might be for some regions more relevant for model performance than a more complex model structure, the choice of an appropriate data set should receive special attention during the setup of transient groundwater models.

Particularly for transient groundwater models, storage properties strongly influence model results. Similar to hydraulic conductivity there are no reliable estimates for storage parametrization available at a large scale and hence uncertainties are large. Additionally, existing data sets like those we used usually have a relatively low resolution containing only a few different values. This obscures correlations and causes a banding in scatterplots (Figure 10). More reliable data sets with higher resolution most likely would further emphasize the key role of hydrogeological parameters for the groundwater response to drought. Elevation and slope were further parameters found to be important for T_{max}. The negative correlations indicate shorter T_{max} in mountainous regions, which matches the intuitive expectation of more quick flow paths on the hillslopes. However, Cochand et al. (2019) recently found that groundwater can also buffer in alpine regions effectively against seasonal variations of water availability. They related this unexpected result to the complex geology of the subsurface catchment which is for groundwater more important than the surface topography. However, complex geological features are not represented in our model and hence such behavior cannot be depicted. This might be an important reason for the weaker model performance found for the mountainous regions. Additional reasons for the differences in observed and modeled T_{max} include potential unknown human influences on observations as well as nonrepresentativeness of a local borehole observation for a 1-km^2 grid cell average. Hence, the local interpretability of the results is limited and observed borehole anomalies can differ considerably from modeled ones.

Many past studies on groundwater drought used groundwater borehole observations for statistical analyses (e.g., Bloomfield et al., 2015; Leelaruban et al., 2017; Van Loon et al., 2017). However, the problem of a limited number of appropriate observations and/or the high diversity of groundwater responses impeded a scaling to gain extensive insights into groundwater drought on catchment or even larger scales (Kumar et al., 2016). Other studies more interested in the groundwaters' role in drought propagation through the hydrological cycle focused on groundwater baseflow (e.g., Carlier et al., 2019; Hellwig & Stahl, 2018). However, Käser and Hunkeler (2016) found that groundwater can contribute substantially to catchment outflow during low flow, so the investigation of baseflow records might capture only a part of the catchments' groundwater storage during drought. Additionally, in this study we have shown that baseflow $T_{\max}$ are typically much shorter than those of groundwater. The groundwater model developed for this study can correctly capture these characteristic differences of groundwater and baseflow. Similarly, Huang et al. (2019) showed that the use of a groundwater layer in a hydrological model can improve simulated surface discharge, particularly during low flows when baseflow contribution is high. A large-scale model including lateral flows offers the opportunity to disentangle both the diversity of groundwater responses and the differences of groundwater head and baseflow fluctuations, that is, the difference of groundwater drought and the groundwater's role in drought propagation.

The results indicate that the groundwater reaction to drought is strongly related to the groundwater's $T_{\max}$. For groundwater drought monitoring it is therefore crucial to gain observations that reflect the aquifer's characteristic dynamics. Particularly in densely populated regions, where groundwater is an important source for water demand, groundwater flow paths and fluctuations can differ a lot from near-natural ones (Price, 2011). The model used in this study focuses on near-natural conditions but, for example, for water management anthropogenic changes of the natural system as well as feedback between drought and groundwater pumping (Famiglietti, 2014) also need to be considered.

Climate change is expected to influence all parts of the hydrological cycle, including recharge and groundwater drought (e.g., Chen et al., 2018; Cuthbert et al., 2019; Eckhardt & Ulbrich, 2003; Hunkeler et al., 2014; Taylor et al., 2012). The large diversity of groundwater responses across Germany indicates regionally different vulnerabilities to changes in the precipitation amount, intensity, or seasonality as well as the drought frequency of different drought types. As the future meteorological conditions might go beyond observed past behavior, the complexity of a process-based model can become necessary to predict nonlinear dynamics (Fatichi et al., 2016) like changed $T_{\max}$ resulting from shifts in the recharge period. The groundwater model used in this work is relatively simple—it could be easily set up for other regions too—but still process-based; therefore, it can help to enhance water resource management under climate change. Also, the dynamics of groundwater as an important water resource are important for drought governance and should be included in drought monitoring.

5. Conclusions

In this study, a country-scale groundwater model was tested for its potential to simulate locally different responses of groundwater to meteorological drought across Germany. The model is based on MODFLOW and simulated groundwater heads and head dynamics in an upper unconfined aquifer under natural conditions. The precipitation accumulation time that has a maximum correlation with groundwater anomalies was used as a metric to distinguish where groundwater heads markedly decline either for short- or long-term meteorological droughts. $T_{\max}$ increases with hydraulic conductivity and specific yield, emphasizing that the aquifer parameterization is crucial for simulating the timing and dynamics of groundwater depletion during droughts. We demonstrated that model performance improved using a more accurate national data set on hydraulic conductivity instead of a global data set. However, as accurate national-to-global-scale subsurface data sets are rare and contain large uncertainties, aquifer parametrization remains challenging in general and still limits the local interpretability of large-scale models.

The model evaluation targeting the representation of anomalies in drought situations revealed that the relatively simple groundwater model can reproduce the general regional differences in the groundwater system. Even though model performance is lower in mountainous regions, we consider the model a useful first step toward the representation of the range of drought responses in Germany We conclude that large-scale groundwater models may have potential to expand common anomaly-based drought monitoring by an

important component given they can be validated and/or scaled to the region or aquifer of interest. This addition may enable proactive drought governance as well as early warning. Modeling studies, which assess the large-scale impact of climate change on groundwater, may also benefit from such models. We expect that predictions at large scale may be further improved when models of gradient-based lateral groundwater flow are coupled to more detailed hydrological models and to higher-resolution parameterization in a next step. Locally, more accurate information on groundwater droughts will then be valuable to assess groundwater vulnerability and management in the context of a changing exposure to droughts.

References

Andreadis, K., Clark, E., Wood, A., Hamlet, A., & Lettenmaier, D. (2005). Twentieth-century drought in the conterminous United States. *Journal of Hydrometeorology*, *6*(6), 985–1001. https://doi.org/10.1175/JHM450.1

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., & Svoboda, M. (2016). Drought indicators revisited: The need for a wider consideration of environment and society. *Wiley Interdisciplinary Reviews Water*, *3*(4), 516–536. https://doi.org/10.1002/wat2.1154

Barker, L. J., Hannaford, J., Chiverton, A., & Svensson, C. (2016). From meteorological to hydrological drought using standardised indicators. *Hydrology and Earth System Sciences*, *20*(6), 2483–2505. https://doi.org/10.5194/hess-20-2483-2016

Bayerisches Landesamt für Wasserwirtschaft (2005). *Wasserwirtschaftlicher Bericht – Niedrigwasserperiode 2003*. Bayerisches Landesamt für Wasserwirtschaft: München.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, *24*(1), 43–69.

BGR & SGD (2016). Bundesanstalt für Geowissenschaften und Rohstoffe and Staatliche Geologische Dienste: Hydrogeologische Übersichtskarte von Deutschland 1:200.000, Oberer Grundwasserleiter (HÜK200 OGWL). Digitaler Datenbestand, version 3.0. – Hannover.

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

Bloomfield, J. P., Marchant, B. P., Bricker, S. H., & Morgan, R. B. (2015). Regional analysis of groundwater droughts using hydrograph classification. *Hydrology and Earth System Sciences*, *19*(10), 4327–4344. https://doi.org/10.5194/hess-19-4327-2015

BMU (2003). Federal Ministry for the Environment, Nature Conservation and Nuclear Safety: Hydrologischer Atlas von Deutschland. Lieferung 1-3 mit 51 Kartentafeln. Bonn/Berlin.

Brutsaert, W. (2008). Long-term groundwater storage trends estimated from streamflow records: Climatic perspective. *Water Resources Research*, *44*(2), W02409. https://doi.org/10.1029/2007WR006518

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., & Brunner, P. (2019). Exploring geological and topographical controls on low flows with hydrogeological models. *Groundwater*, *57*(1), 15. https://doi.org/10.1111/gwat.12845

Changnon, S. (1987). Detecting drought conditions in Illinois. ISWS/CIR-169/87: Illinois State Water Survey.

Chen, Z., Hartmann, A., Wagener, T., & Goldscheider, N. (2018). Dynamics of water fluxes and storages in an Alpine karst catchment under current and potential future climate conditions. *Hydrology and Earth System Sciences*, *22*(7), 3807–3823. https://doi.org/10.5194/hess-22-3807-2018

Cochand, M., Christe, P., Ornstein, P., & Hunkeler, D. (2019). Groundwater storage in high alpine catchments and its contribution to streamflow. *Water Resources Research*, *55*(4), 2613–2630. https://doi.org/10.1029/2018WR022989

Condon, L. E., Hering, A. S., & Maxwell, R. M. (2015). Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm. [Article]. *Advances in Water Resources*, *82*, 106–123. https://doi.org/10.1016/j.advwatres.2015.04.008

Condon, L. E., & Maxwell, R. M. (2015). Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model. *Water Resources Research*, *51*(8), 6602–6621. https://doi.org/10.1002/2014WR016774

Cuthbert, M. O., Gleeson, T., Moosdorf, N., Befus, K. M., Schneider, A., Hartmann, J., & Lehner, B. (2019). Global patterns and dynamics of climate-groundwater interactions. *Nature Climate Change*, *9*(2), 137–141. https://doi.org/10.1038/s41558-018-0386-4

de Graaf, I. E. M., Gleeson, T., van Beek, L. P. H., Sutanudjaja, E. H., & Bierkens, M. F. P. (2019). Environmental flow limits to global groundwater pumping. *Nature*, *574*(7776), 90–94. https://doi.org/10.1038/s41586-019-1594-4

de Graaf, I. E. M., Sutanudjaja, E. H., van Beek, L. P. H., & Bierkens, M. F. P. (2015). A high-resolution global-scale groundwater model. *Hydrology and Earth System Sciences*, *19*(2), 823–837. https://doi.org/10.5194/hess-19-823-2015

de Graaf, I. E. M., van Beek, R., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., & Bierkens, M. F. P. (2017). A global-scale two-layer transient groundwater model: Development and application to groundwater depletion. *Advances in Water Resources*, *102*, 53–67. https://doi.org/10.1016/j.advwatres.2017.01.011

Eckhardt, K., & Ulbrich, U. (2003). Potential impacts of climate change on groundwater recharge and streamflow in a central European low mountain range. *Journal of Hydrology*, *284*(1-4), 244–252. https://doi.org/10.1016/j.jhydrol.2003.08.005

European Environment Agency (2013). Digital elevation model over Europe (EU-DEM). Retrieved from http://www.eea.europa.eu/data-and-maps/data/eu-dem

Famiglietti, J. S. (2014). The global groundwater crisis. *Nature Climate Change*, *4*(11), 945. https://doi.org/10.1038/nclimate2425

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global patterns of groundwater table depth. *Science*, *339*(6122), 940–943. https://doi.org/10.1126/science.1229881

Fan, Y., Miguez-Macho, G., Weaver, C. P., Walko, R., & Robock, A. (2007). Incorporating water table dynamics in climate modeling: 1. Water table observations and equilibrium water table simulations. *Journal of Geophysical Research*, *112*, D10125. https://doi.org/10.1029/2006JD008111

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Foster, L. M., & Maxwell, R. M. (2019). Sensitivity analysis of hydraulic conductivity and Manning's n parameters lead to new method to scale effective hydraulic conductivity across model resolutions. *Hydrological Processes*, *33*(3), 332–349. https://doi.org/10.1002/hyp.13327

Gleeson, T., Moosdorf, N., Hartmann, J., & Beek, L. P. H. (2014). A glimpse beneath Earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, *41*(11), 3891–3898. https://doi.org/10.1002/2014GL059856

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., et al. (2012). Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. *Journal of Hydrometeorology*, *13*(2), 604–620. https://doi.org/10.1175/jhm-d-11-083.1

Gustard, A., & Demuth, S. (Eds.) (2008). World Meteorological Organization: Manual on low-flow estimation and prediction. Operational Hydrology Report No. 50, WMO-NO. 1029.

Haas, J. C., & Birk, S. (2017). Characterizing the spatiotemporal variability of groundwater levels of alluvial aquifers in different settings using drought indices. *Hydrology and Earth System Sciences*, *21*(5), 2421–2448. https://doi.org/10.5194/hess-21-2421-2017

Harbaugh, A. W., Banta, E. R., Hill, M. C., & McDonald, M. G. (2000). MODFLOW-2000, The U.S. Geological Survey modular ground-water model-user guide to modularization concepts and the ground-water flow process. Open-file Report. U. S. Geological Survey(92), 134.

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, *1*(2), 96–99.

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*, 37. https://doi.org/10.1029/2012gc004370

Haylock, M. R., Hofstra, N., Tank, A., 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*, *113*, D20119. https://doi.org/10.1029/2008JD010201

Hellwig, J., & Stahl, K. (2018). An assessment of trends and potential future changes in groundwater-baseflow drought based on catchment response times. *Hydrology and Earth System Sciences*, *22*(12), 6209–6224. https://doi.org/10.5194/hess-22-6209-2018

Hellwig, J., Stahl, K., Ziese, M., & Becker, A. (2018). The impact of the resolution of meteorological data sets on catchment-scale precipitation and drought studies. *International Journal of Climatology*, *38*(7), 3069–3081. https://doi.org/10.1002/joc.5483

Heudorfer, B., Haaf, E., Stahl, K., & Barthel, R. (2019). Index-based characterization and quantification of groundwater dynamics. *Water Resources Research*, *55*, 5575–5592. https://doi.org/10.1029/2018WR024418

Heudorfer, B., & Stahl, K. (2017). Comparison of different threshold level methods for drought propagation analysis in Germany. *Hydrology Research*, *48*(5), 1311–1326. https://doi.org/10.2166/nh.2016.258

Howard, K., & Griffith, A. (2009). Can the impacts of climate change on groundwater resources be studied without the use of transient models? *Hydrological Sciences Journal*, *54*(4), 754–764. https://doi.org/10.1623/hysj.54.4.754

Huang, Z., Tang, Q., Lo, M.-H., Liu, X., Lu, H., Zhang, X., & Leng, G. (2019). The influence of groundwater representation on hydrological simulation and its assessment using satellite-based water storage variation. *Hydrological Processes*, *33*(8), 1218–1230. https://doi.org/10.1002/hyp.13393

Hunkeler, D., Möck, C., Käser, D., & Brunner, P. (2014). Klimaeinflüsse auf Grundwassermengen. *Aqua & Gas*, *11*, 43–49.

Huscroft, J., Gleeson, T., Hartmann, J., & Borker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, *45*(4), 1897–1904. https://doi.org/10.1002/2017GL075860

Käser, D., & Hunkeler, D. (2016). Contribution of alluvial groundwater to the outflow of mountainous catchments. *Water Resources Research*, *52*(2), 680–697. https://doi.org/10.1002/2014WR016730

Keune, J., Gasper, F., Goergen, K., Hense, A., Shrestha, P., Sulis, M., & Kollet, S. (2016). Studying the influence of groundwater representations on land surface-atmosphere feedbacks during the European heat wave in 2003. *Journal of Geophysical Research: Atmospheres*, *121*, 13,301–13,325. https://doi.org/10.1002/2016JD025426

Kohn, I., Rosin, K., Freudiger, D., Belz, J. U., Stahl, K., & Weiler, M. (2014). Niedrigwasser in Deutschland 2011. *Hydrologie und Wasserbewirtschaftung*, *58*(1), 4–17. https://doi.org/10.5675/HyWa_2014,1_1

Krakauer, N. Y., Li, H., & Fan, Y. (2014). Groundwater flow across spatial scales: Importance for climate modeling. *Environmental Research Letters*, *9*(3), 034003. https://doi.org/10.1088/1748-9326/9/3/034003

Kumar, R., Musuuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., et al. (2016). Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator. *Hydrology and Earth System Sciences*, *20*(3), 1117–1131. https://doi.org/10.5194/hess-20-1117-2016

Leelaruban, N., Padmanabhan, G., & Oduor, P. (2017). Examining the relationship between drought indices and groundwater levels. *Water*, *9*(2), 16. https://doi.org/10.3390/w9020082

LfW (2004). Sommer 2003. Ein wasserwirtschaftlicher Bericht zur Hitzeperiode im Jahre 2003. Mainz: Landesamt für Wasserwirtschaft Rheinland-Pfalz.

Maxwell, R. M., Condon, L. E., & Kollet, S. J. (2015). A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3. *Geoscientific Model Development*, *8*(3), 923–937. https://doi.org/10.5194/gmd-8-923-2015

McKee, T. B., Doesken, N., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. Eighth Conference on Applied Climatology, Anaheim CA, 6.

Miguez-Macho, G., Li, H., & Fan, Y. (2008). Simulated water table and soil moisture climatology over North America. *Bulletin of the American Meteorological Society*, *89*(5), 663–672.

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30). Weather Bureau Washington, DC: US Department of Commerce.

Pianosi, F., Beven, K., Freer, J., Hall, J., Rougier, J., Stephenson, D., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography*, *35*(4), 465–492. https://doi.org/10.1177/0309133311402714

Schaller, M., & Fan, Y. (2009). River basins as groundwater exporters and importers: Implications for water cycle and climate modeling. *Journal of Geophysical Research*, *114*, D04103. https://doi.org/10.1029/2008JD010636

Stahl, K., Tallaksen, L. M., Gudmundsson, L., & Christensen, J. H. (2011). Streamflow data from small basins: A challenging test to high-resolution regional climate modeling. *Journal of Hydrometeorology*, *12*(5), 900–912. https://doi.org/10.1175/2011JHM1356.1

Stahl, K., Tallaksen, L. M., Hannaford, J., & van Lanen, H. A. J. (2012). Filling the white space on maps of European runoff trends: Estimates from a multi-model ensemble. *Hydrology and Earth System Sciences*, *16*(7), 2035–2047. https://doi.org/10.5194/hess-16-2035-2012

Sutanudjaja, E. H., van Beek, L. P. H., de Jong, S. M., van Geer, F. C., & Bierkens, M. F. P. (2011). Large-scale groundwater modeling using global datasets: A test case for the Rhine-Meuse basin. *Hydrology and Earth System Sciences*, *15*(9), 2913–2935. https://doi.org/10.5194/hess-15-2913-2011

Tallaksen, L. M., Hisdal, H., & Van Lanen, H. A. J. (2009). Space-time modelling of catchment scale drought characteristics. *Journal of Hydrology*, *375*(3-4), 363–372. https://doi.org/10.1016/j.jhydrol.2009.06.032

Tallaksen, L. M., & Stahl, K. (2014). Spatial and temporal patterns of large-scale droughts in Europe: Model dispersion and performance. *Geophysical Research Letters*, *41*(2), 429–434. https://doi.org/10.1002/2013GL058573

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The Gravity Recovery and Climate Experiment: Mission overview and early results. *Geophysical Research Letters*, *31*, L09607. https://doi.org/10.1029/2004gl019920

Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., et al. (2012). Ground water and climate change. *Nature Climate Change*, *3*(4), 322–329. https://doi.org/10.1038/nclimate1744

Trichakis, I., Burek, P., de Roo, A., & Pistocchi, A. (2017). Towards a Pan-European integrated groundwater and surface water model: Development and applications. *Environmental Processes*, *4*(1), 81–93. https://doi.org/10.1007/s40710-017-0216-0

Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., & Van Loon, A. F. (2013). Hydrological drought across the world: impact of climate and physical catchment structure. *Hydrology and Earth System Sciences*, *17*(5), 1715–1732. https://doi.org/10.5194/hess-17-1715-2013

Van Loon, A. F. (2015). Hydrological drought explained. *Wiley Interdisciplinary Reviews Water*, *2*(4), 359–392. https://doi.org/10.1002/wat2.1085

Van Loon, A. F., Kumar, R., & Mishra, V. (2017). Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time. *Hydrology and Earth System Sciences*, *21*(4), 1947–1971. https://doi.org/10.5194/hess-21-1947-2017

Van Loon, A. F., & Laaha, G. (2015). 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

Van Loon, A. F., Van Huijgevoort, M. H. J., & Van Lanen, H. A. J. (2012). Evaluation of drought propagation in an ensemble mean of large-scale hydrological models. *Hydrology and Earth System Sciences*, *16*, 4057–4078.

Westerhoff, R., White, P., & Miguez-Macho, G. (2018). Application of an improved global-scale groundwater model for water table estimation across New Zealand. *Hydrology and Earth System Sciences*, *22*(12), 6449–6472. https://doi.org/10.5194/hess-22-6449-2018