



Brenner, S., Coxon, G., Howden, N. J. K., Freer, J., & Hartmann, A. (2018). Process-based modelling to evaluate simulated groundwater levels and frequencies in a Chalk catchment in south-western England. *Natural Hazards and Earth System Sciences*, 18(2), 445-461. <https://doi.org/10.5194/nhess-18-445-2018>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.5194/nhess-18-445-2018](https://doi.org/10.5194/nhess-18-445-2018)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the final published version of the article (version of record). It first appeared online via The European Geosciences Union at <https://doi.org/10.5194/nhess-18-445-2018> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>



# Process-based modelling to evaluate simulated groundwater levels and frequencies in a Chalk catchment in south-western England

Simon Brenner<sup>1</sup>, Gemma Coxon<sup>2,4</sup>, Nicholas J. K. Howden<sup>3,4</sup>, Jim Freer<sup>2,4</sup>, and Andreas Hartmann<sup>1,3</sup>

<sup>1</sup>Institute of Earth and Environmental Sciences, Freiburg University, Freiburg, Germany

<sup>2</sup>School of Geographical Sciences, University of Bristol, Bristol, UK

<sup>3</sup>Department of Civil Engineering, University of Bristol, Bristol, UK

<sup>4</sup>Cabot Institute, University of Bristol, Bristol, UK

**Correspondence:** Simon Brenner ([simon.brenner@hydrology.uni-freiburg.de](mailto:simon.brenner@hydrology.uni-freiburg.de))

Received: 2 December 2016 – Discussion started: 22 December 2016

Revised: 30 November 2017 – Accepted: 15 December 2017 – Published: 8 February 2018

**Abstract.** Chalk aquifers are an important source of drinking water in the UK. Due to their properties, they are particularly vulnerable to groundwater-related hazards like floods and droughts. Understanding and predicting groundwater levels is therefore important for effective and safe water management. Chalk is known for its high porosity and, due to its dissolvability, exposed to karstification and strong subsurface heterogeneity. To cope with the karstic heterogeneity and limited data availability, specialised modelling approaches are required that balance model complexity and data availability. In this study, we present a novel approach to evaluate simulated groundwater level frequencies derived from a semi-distributed karst model that represents subsurface heterogeneity by distribution functions. Simulated groundwater storages are transferred into groundwater levels using evidence from different observations wells. Using a percentile approach we can assess the number of days exceeding or falling below selected groundwater level percentiles. Firstly, we evaluate the performance of the model when simulating groundwater level time series using a split sample test and parameter identifiability analysis. Secondly, we apply a split sample test to the simulated groundwater level percentiles to explore the performance in predicting groundwater level exceedances. We show that the model provides robust simulations of discharge and groundwater levels at three observation wells at a test site in a chalk-dominated catchment in south-western England. The second split sample test also indicates that the percentile approach is able to reliably predict groundwater level exceedances across all considered timescales up to their 75th percentile. However, when look-

ing at the 90th percentile, it only provides acceptable predictions for long time periods and it fails when the 95th percentile of groundwater exceedance levels is considered. By modifying the historic forcings of our model according to expected future climate changes, we create simple climate scenarios and we show that the projected climate changes may lead to generally lower groundwater levels and a reduction of exceedances of high groundwater level percentiles.

## 1 Introduction

The English Chalk aquifer region extends over large parts of south-western England and is an important water resource aquifer, providing about 55 % of all groundwater-abstracted drinking water in the UK (Lloyd, 1993). As a carbonate rock the English Chalk is exposed to karstification, i.e. chemical weathering (Ford and Williams, 2007), resulting in particular surface and subsurface features such as dolines, river sinks, caves and conduits (Goldscheider and Drew, 2007; Gutiérrez et al., 2014; Stevanovic, 2015). Consequently, karstification also produces strong hydrological subsurface heterogeneity (Bakalowicz, 2005). The interplay between diffuse and concentrated infiltration and recharge processes, as well as fast flow through karstic conduits and diffuse matrix flow, results in complex flow and storage dynamics (Hartmann et al., 2014a). Even though Chalk has a tendency for less intense karstification, for instance compared to limestone, its karstic behaviour has increasingly been recognised (Maurice et al., 2006, 2012; Fitzpatrick, 2011).

Apart from the high water quality, favourable infiltration and storage dynamics which make chalk aquifers a preferred source of drinking water in the UK, their karstic behaviour also increases the risk of fast drainage of their storages by karstic conduit flow during dry years. This also increases the risk of groundwater flooding as a result of fast responses of groundwater levels to intense rainfalls due to fast infiltration and groundwater recharge processes. Groundwater flooding, i.e. when groundwater levels emerge at the ground surface due to intense rainfall (Macdonald et al., 2008), tend to be more severe in areas of permeable outcrop like the English Chalk (Macdonald et al., 2012) as also experienced repeatedly in other karst areas in Europe (Parise, 2003, 2010; Bonacci et al., 2006; Jourde et al., 2007; Gutiérrez, 2010; Naughton et al., 2012; Parise et al., 2015). Groundwater drought indices tend to be more related to recharge conditions in Cretaceous Chalk aquifers than in granular aquifers (Bloomfield and Marchant, 2013). Due to the fast transfer of water from the soil surface to the main groundwater system, chalk aquifers tend to be more sensitive to external changes, as for instance shown by Jackson et al. (2015), who found significant groundwater level declines in 4 out of 7 chalk boreholes in a UK-wide study using historic groundwater level observations.

Climate projections suggest that the UK will experience increasing temperatures, with less rainfall during the summer but warmer and wetter winters (Jenkins et al., 2008). This may stress these groundwater resources and increase the risk of groundwater droughts and potentially winter groundwater flooding. For those reasons, assessment of potential future changes in groundwater dynamics, concerning groundwater droughts, median groundwater levels as well as groundwater flooding is broadly recommended and is the subject of current research around the world (Naughton et al., 2012, 2015; Jackson et al., 2015; von Freyberg et al., 2015; Jimenez-Martinez et al., 2016; Moutahir et al., 2017; Perrone and Jasechko, 2017).

However, present approaches mostly rely on statistical distribution functions to express groundwater dynamics and groundwater level exceedance probabilities (e.g. Bloomfield et al., 2015; Kumar et al., 2016) and it is questionable whether the shapes of these distribution functions remain the same when climate or land use change. Physics-based hydrological simulation models that incorporate hydrological processes in a relatively high detail can be considered to potentially provide the most reliable predictions, especially under a changing environment. However, there are considerable limitations in obtaining the necessary information to estimate the structure and the model parameters, especially for subsurface processes, and this inevitably increases modelling uncertainties (Perrin et al., 2003; Beven, 2006).

The definition of appropriate model structures and parameters from limited information becomes problematic when modelling karst aquifers. In order to achieve acceptable simulation performance they have to include representations of

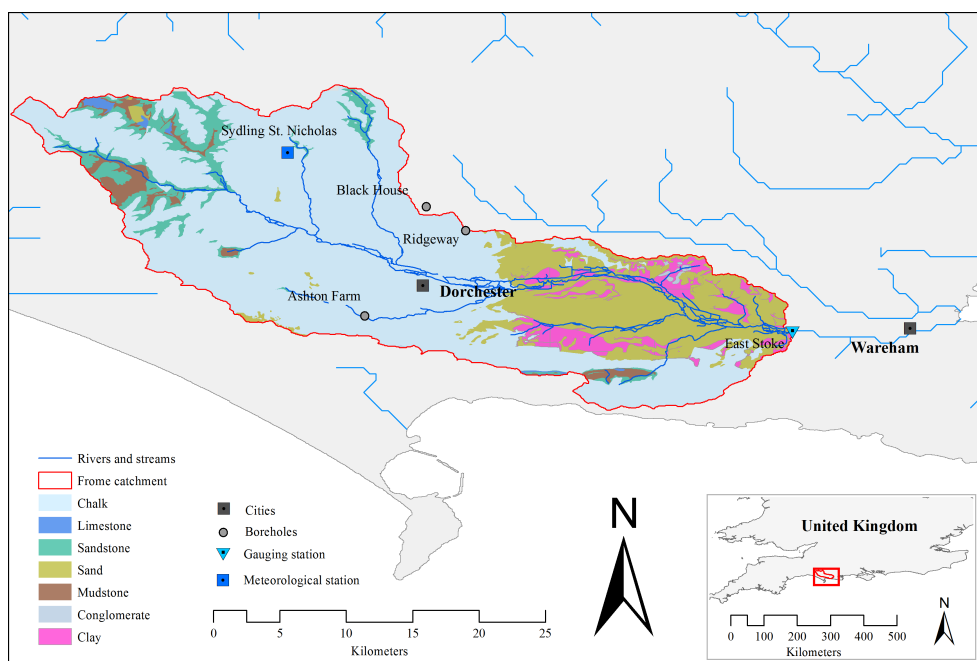
karstic heterogeneity in their structures. Distributed karst modelling approaches are able to simulate groundwater levels on a spatial grid but their data requirements mostly limit them to theoretical studies (e.g. Birk et al., 2006; Reimann et al., 2011) or well explored study sites (e.g. Hill et al., 2010; Jackson et al., 2011; Oehlmann et al., 2015). Lumped karst modelling approaches consider physical processes on the scale of the entire karst system. Although they are strongly simplified, they can include karst peculiarities such as different conduit and matrix systems (Maloszewski et al., 2002; Geyer et al., 2008; Fleury et al., 2009). Since they are easy to implement and do not require spatial information, they are widely used in karst modelling (Jukić and Denić-Jukić, 2009). Simple rainfall–run-off models with more than 5–6 parameters are often assumed to end up in equifinality (Wheater et al., 1986; Jakeman and Hornberger, 1993; Ye et al., 1997); i.e. their parameters lose their identifiability (Wagener et al., 2002; Beven, 2006). For that reason, recent research took advantage of auxiliary data, such as water quality data or tracer experiments (Hartmann et al., 2013b; Oehlmann et al., 2015). These studies showed that adding such information allows the necessary model parameters to be identified, therefore enabling the model to reflect the relevant processes.

Up to now, most lumped karst models have been applied for rainfall–run-off simulations. Groundwater levels were simulated in quite a few studies (Adams et al., 2010; Ladouche et al., 2014; Jimenez-Martinez et al., 2016) but mostly relied on very simple representations of karst hydrological processes and disregarding the scale discrepancy between borehole (point scale) and modelling domain (catchment scale) at which they were applied.

In this study, we present a novel approach to predict and evaluate groundwater level frequencies in chalk-dominated catchments. It uses a previously developed semi-distributed process-based model (VarKarst, Hartmann et al., 2013b) that we further developed to simulate groundwater levels. To assess groundwater level frequencies we formulated a percentile of a groundwater-based approach that quantifies the probability of exceeding or falling below selected groundwater levels. We exemplify and evaluate our new approach on a Chalk catchment in south-western England that had to cope with several flooding events in the past. Finally we apply the approach to simple climate scenarios that we create by modifying our historic model forcings to show how changes in evapotranspiration and precipitation can affect groundwater level frequencies.

## 2 Study site and data availability

Located in West Dorset in the south-west of England the river Frome drains a rural catchment with an area approximately 414 km<sup>2</sup> (Fig. 1). The catchment elevation varies from over 200 m a.s.l. (a.s.l.) in the north-west to sea level in



**Figure 1.** Location map with an overview of the Frome catchment.

the south-east. The topography is very flat with a mean slope of 3.9 % and a mean height of approximately 111 m a.s.l. The climate can be defined as oceanic with mild winters and warm summers (Dorset County Council, 2009). Howden (2006) characterised the Frome as highly groundwater dominated. During the summer months, discharge of the Frome is typically very low, hardly reaching  $5 \text{ m}^3 \text{ s}^{-1}$  (Brunner et al., 2010). The geology is predominated by the Cretaceous Chalk outcrop which underlays around 65 % of the catchment. The headwaters of the Frome include outcrops of the Upper Greensand, often overlain by the rather impermeable Zig Zag Chalk (Howden, 2006). The middle reaches of the Frome traverse the Cretaceous Chalk outcrop followed by Palaeogene strata in the lower reaches, eventually draining into Poole Harbour. The major aquifer Chalk appears mainly unconfined. However, in the lower reaches it is overlain by Palaeogene strata, resulting in confined aquifer conditions. The region around the Frome catchment is known for the highest density of solution features in the UK (Edmonds, 1983) which can be mainly observed in the interfluvium between the Frome and Piddle (Adams et al., 2003). Loams over chalk, shallow silts, deep loamy, sandy and shallow clays constitute the primary types of soils occurring in the study area (Brunner et al., 2010). The soils of the upper parts of the catchment are mainly shallow and well drained (NRA, 1995). In the middle and lower reaches the soils are becoming more sandy and acidic due to waterlogged conditions caused by either groundwater or winter flooding (NRA, 1995; Brunner et al., 2010). Due to its geological setting, the area is prone to groundwater flooding. It has occurred several

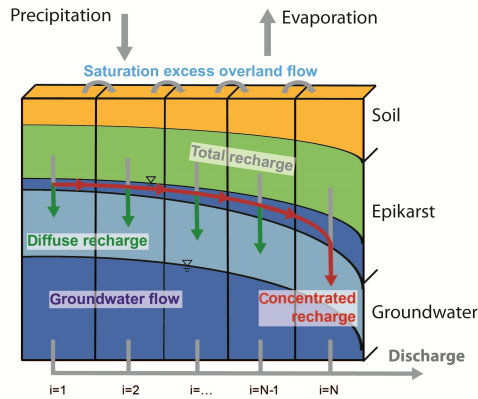
times at different locations, for example in Maiden Newton during winter 2000/2001 (Environment Agency, 2012) and in Winterbourne Abbas during summer 2012 (Bennett, 2013).

### 3 Methodology

In order to consider karstic process behaviour in our simulations we use the process-based karst model VarKarst introduced by Hartmann et al. (2013b). VarKarst includes the karstic heterogeneity and the complex behaviour of karst processes using distribution functions that represent the variability of soil, epikarst and groundwater and was applied successfully at different karst regions over Europe (Hartmann et al., 2013a, 2014b, 2016). We use a simple linear relationship that takes into account effective porosities and base level of the groundwater wells (see Eq. 1) enabling the model to simulate groundwater levels based on the groundwater storage in VarKarst. Finally, a newly developed evaluation approach is used by transferring simulated groundwater level time series into groundwater level frequency distributions and comparing them to observed behaviour at a number of monitored wells.

#### 3.1 The model

The VarKarst model operates on a daily time step. Similarly to other karst models, it distinguishes between three subroutines representing the soil system, the epikarst system and the groundwater system but it also includes their spatial variability, which is expressed by distribution functions that are



**Figure 2.** The VarKarst model structure.

applied to a set of  $N = 15$  model compartments (Fig. 2). Pareto functions as distribution functions have shown to perform best in previous work (Hartmann et al., 2013a, b), as well as the number of 15 model compartments (Hartmann et al., 2012). Including the spatial variability of subsurface properties in this manner, the VarKarst model can be seen as a hybrid or semi-distributed model. All relevant model parameters are provided in Table 1. For a detailed description of VarKarst see the appendix or Hartmann et al. (2013b).

The model was driven by two input time series (precipitation and potential evapotranspiration, PET), and the 13 variable model parameters (see Table 2) were calibrated and evaluated by four observed time series (discharge and the three boreholes,; see Sect. 3.3). Similarly to Kuczera and Mroczkowski (1998) we use a simple linear homogeneous relationship which translates the groundwater storage (mm) into a groundwater level (m a.s.l.):

$$h_{\text{GW}}(t) = \frac{V_{\text{GW},i}(t)}{1000 \times p_{\text{GW}}} + \Delta h. \quad (1)$$

The related parameters are  $h_{\text{GW}}$  (m) and  $p_{\text{GW}}$  (-).  $h_{\text{GW}}$  is the difference between the base of the contributing groundwater storage (that is simulated by the model) and the base of the well that is used for calibration and evaluation.  $p_{\text{GW}}$  represents the average porosity of the rock that is intersected by the well.

### 3.2 Data availability

The daily discharge data for gauge East Stoke were obtained from the Centre for Ecology & Hydrology (CEH, <http://nrfa.ceh.ac.uk/>) and dates back to the 1960s. The borehole data were provided by the Environment Agency (EA) and obtained via the University of Bristol. The total data used for modelling in this study can be seen in Table 1. The three boreholes (Ashton Farm, Ridgeway and Black House) comprised high-resolution raw data which had been collected at 15 min intervals. For further analysis, the data were aggregated to daily time averages. The potential evapotran-

spiration has a strong annual cycle. Since most recent data from 2009 to 2012 was missing, representative PET years were calculated on the basis of the last 50 years. Climate projections were obtained from the UK Climate Projections User Interface (UKCP09 UI, <http://ukclimateprojections-ui.metoffice.gov.uk/>). For more information about the UKCP see Murphy et al. (2009).

### 3.3 Model calibration and evaluation

We use the shuffled complex evolution method (SCEM) for our calibration, which is based on the Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) and the shuffled complex evolution algorithm (SCE, Duan et al., 1992). The Metropolis–Hastings algorithm uses a formal likelihood measure and calculates the ratio of the posterior probability densities of a “candidate” parameter set that is drawn from a proposal distribution and a given parameter set. If this ratio is larger than or equal to a number randomly drawn from a uniform distribution between 0 and 1, the candidate parameter set is accepted. This procedure is repeated for a large number of iterations. If the proposal distribution is properly chosen, the Markov chain will rapidly explore the parameter space and it will converge to the target distribution of interest (Vrugt et al., 2003). In the SCEM algorithm, candidate parameter sets are drawn from a self-adapting proposal distribution for each of a predefined number of clusters. Again a random number [0, 1] is used to accept or discard candidate parameter sets. The SCEM algorithm was applied in default mode using the Gelman–Rubin convergence criteria (Vrugt et al., 2003). In our study, we use the Kling–Gupta efficiency (KGE; Gupta et al., 2009) as the objective function, which can be regarded as an informal likelihood measure or more generally as a monotonically increasing performance metric of model skill (Smith et al., 2008). It was chosen by trial and error, comparing the simulation performances during calibration and validation obtained with different objective functions (RMSE and others). We found that we obtain the most robust results with the KGE. To decide whether to accept or discard a parameter set, we compare the KGEs of the candidate and the given parameter sets. This procedure was already applied in various studies (Engeland et al., 2005; Blasone et al., 2008; McMillan and Clark, 2009) and is possible if the error functions monotonically increase with improved performance. We achieved this in the SCEM algorithm by defining KGE as

$$\text{KGE} = -\sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (2)$$

$$\alpha = \frac{\sigma_s}{\sigma_o}; \quad \beta = \frac{\mu_s}{\mu_o},$$

with  $r$  as the linear correlation coefficient between simulations and observations, and  $\sigma_s \sigma_o$  and  $\mu_s \mu_o$  as the means and SDs of simulations and observations respectively.

The posterior parameter distributions derived from SCEM provide information about the identifiability of the param-

**Table 1.** All available data used in the study.

| Parameter                    | Station                            | Source | Period of time        | Resolution | Unit                           |
|------------------------------|------------------------------------|--------|-----------------------|------------|--------------------------------|
| Precipitation                | Sydling St. Nicolas (44006)        | CEH    | 01.01.2000–31.12.2012 | daily      | mm d <sup>-1</sup>             |
| Discharge                    | East Stoke (44001)                 | CEH    | 01.01.2000–31.12.2012 | daily      | m <sup>3</sup> s <sup>-1</sup> |
| Potential evapotranspiration | Catchment Cut East Stoke           | CEH    | 01.01.2000–31.12.2008 | daily      | mm d <sup>-1</sup>             |
| Groundwater levels           | Ashton Farm, Ridgeway, Black House | EA     | 01.01.2003–31.12.2012 | daily      | m a.s.l.                       |
| Climate delta values         | Grid Box Nr. 1698 (25 × 25 km)     | UKCP   | 2070–2099             | annual     | °C, %                          |

**Table 2.** Model parameters, descriptions, ranges and optimised values.

| Parameter                 | Description  | Unit | Ranges |       | Weighting | Optimised values |
|---------------------------|--|------|--------|-------|-----------|------------------|
|                           |  |      | Lower  | Upper |           |                  |
| $V_{\text{mean, S}}$      | Mean soil storage capacity                             | mm   | 1000   | 2500  |           | 2015.6           |
| $V_{\text{mean, E}}$      | Mean epikarst storage capacity                         | mm   | 1000   | 2500  |           | 1011.7           |
| $K_{\text{mean, E}}$      | Epikarst mean storage coefficient                      | d    | 0.1    | 2.5   |           | 0.7246           |
| $K_C$                     | Conduit storage coefficient                            | d    | 1      | 100   |           | 38.722           |
| $a_{f\text{sep}}$         | Recharge separation variability constant               | –    | 0.1    | 5     |           | 1.1864           |
| $a_{\text{GW}}$           | Groundwater variability constant                       | –    | 1      | 10    |           | 5.9966           |
| $a_{\text{SE}}$           | Soil/epikarst depth variability constant               | –    | 0.1    | 6     |           | 1.8928           |
| $p_{\text{GW, A}}$        | Ashton Farm groundwater level porosity parameter       | –    | 0.001  | 0.5   |           | 0.0069           |
| $\Delta h_{\text{GW, A}}$ | Ashton Farm groundwater level offset parameter         | m    | 0      | 150   |           | 64.167           |
| $p_{\text{GW, R}}$        | Ridgeway groundwater level porosity parameter          | –    | 0.001  | 0.5   |           | 0.0016           |
| $\Delta h_{\text{GW, R}}$ | Ridgeway groundwater level offset parameter            | m    | 0      | 150   |           | 48.718           |
| $p_{\text{GW, B}}$        | Black House groundwater level porosity parameter       | –    | 0.001  | 0.5   |           | 0.0032           |
| $\Delta h_{\text{GW, B}}$ | Black House groundwater level offset parameter         | m    | 0      | 150   |           | 78.448           |
| $KGE_Q$                   | Model performance for discharge                        | –    | 0      | 1     | 0.2       | 0.73/0.58*       |
| $KGE_{\text{GW, A}}$      | Model performance for groundwater level at Ashton Farm | –    | 0      | 1     | 0.4       | 0.94/0.80*       |
| $KGE_{\text{GW, R}}$      | Model performance for groundwater level at Ridgeway    | –    | 0      | 1     | 0.2       | 0.86/–*          |
| $KGE_{\text{GW, B}}$      | Model performance for groundwater level at Black House | –    | 0      | 1     | 0.2       | 0.83/0.74*       |

\* Calibration/validation.

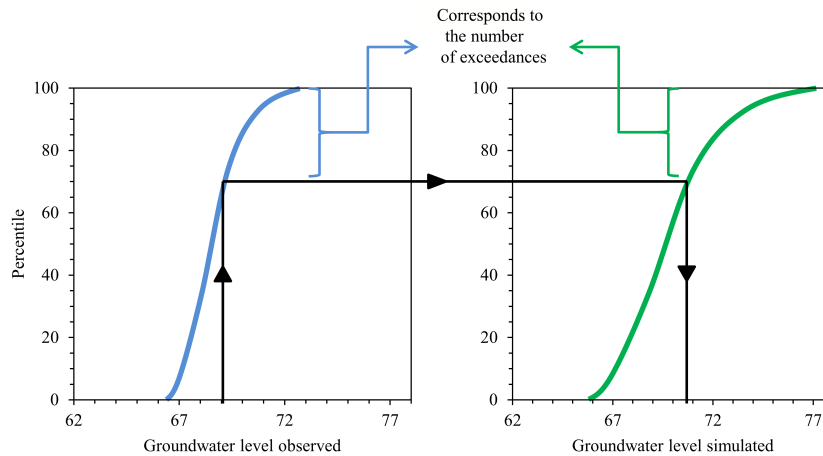
ters. The more they differ from a uniform posterior distribution the higher the identifiability of a model parameter. We present different calibration distributions to show the use of auxiliary data for parameter identifiability.

Parameter ranges were chosen following previous experience with the VarKarst model (Hartmann et al., 2013a, b, 2014b, 2016). Besides the quantitative measure of efficiency, a split sample test (Klemeš, 1986) was carried out. Our data covered precipitation, evapotranspiration, discharge and groundwater levels from 2000 to the end of 2012. We calibrated the model for the period 2008–2012 and used the period 2003–2007 for validation. We chose this reversed order to be able to include the information on three boreholes that was only available for 2008–2012. Three years were used as warm-up for each of calibration and validation. During calibration, the most appropriate of the  $N = 15$  groundwater compartments that represent each groundwater well was found by choosing the compartment with the best correlation to the groundwater dynamics of the well.

This procedure was repeated for each well and each Monte Carlo run and finally provided the three model compartment

numbers that produce the best simulations of groundwater levels at the three operation wells and the best catchment discharge according to our selected weighting scheme. During calibration, we used a weighting scheme which was found by trial and error, as we stepwise added borehole data to our discharge observations. Discharge and the borehole at Ashton Farm were both weighted as one-third as Ashton farm is located in the lower parts within the catchment, while the other two boreholes were located at higher elevation at the catchment's edge and weighted as one-sixth each. In order to explore the contribution of the different observed discharge and groundwater time series during the calibration, we use SCEN to derive the posterior parameter distributions using (1) the final weighting scheme, (2) only discharge, (3) only Ashton farm, and (4) only the other two boreholes (equally weighted). Posterior parameter distributions are plotted as cumulative distributions. The more parameters that show sensitivity, the more information is contained in the selected calibration scheme.





**Figure 3.** Schematic description of the percentile approach.

### 3.4 The percentile approach

Even though the VarKarst model includes spatial variability of system properties by its distribution functions, its semi-distributed structure does not allow for an explicit consideration of the locations of groundwater wells. Its model structure allowed for an acceptable and stable simulation of groundwater level time series of the three wells (see Sect. 4.1), but for groundwater management, frequency distributions of groundwater levels calculated over the timescale of interest are commonly preferred. For that reason we introduced a groundwater level percentile-based approach. Other than Westerberg et al. (2016), who transferred discharge time series into signatures derived from flow duration curves, we calibrate directly with the discharge and groundwater time series in order to evaluate the performance of our approach for selected time periods (see evaluation below). Similarly to the calculation of standardised precipitation or groundwater indices (e.g. Lloyd-Hughes and Saunders, 2002; Bloomfield and Marchant, 2013), we create cumulative frequency distributions of observed groundwater levels and the simulated groundwater levels from the previously evaluated model. Now, the exceedance probability or percentile for a selected observed groundwater level (for instance, the groundwater level above which groundwater flooding can be expected) can be used to define the corresponding simulated groundwater level, and the number of days exceeding or falling below the chosen groundwater level can directly be extracted from the frequency distributions (Fig. 3). Note that this procedure is performed after the model is calibrated and validated with KGE as described in the previous subsection. We avoided a calibration directly to the flow percentiles, as temporal information would have been removed, which would have resulted in a lower prediction performance of the model.

As the approach is meant to be applied in combination with climate change scenarios, we perform an evaluation on multiple timescales and flow percentiles. We assess the 5th,

10th, 25th, 50th, 75th, 90th and 95th percentiles on temporal resolutions of years, seasons, months, weeks and days. The deviation between modelled and observed number of exceedance days of these different percentiles is quantified by the mean absolute deviation (MAD) between simulated exceedances (SE) and observed exceedances (OE):

$$\text{MAD}_p = \text{mean} \left( \text{abs} \left( \sum \text{SE}_{i,x} - \sum \text{OE}_{i,x} \right) \right) \quad [\text{d}], \quad (3)$$

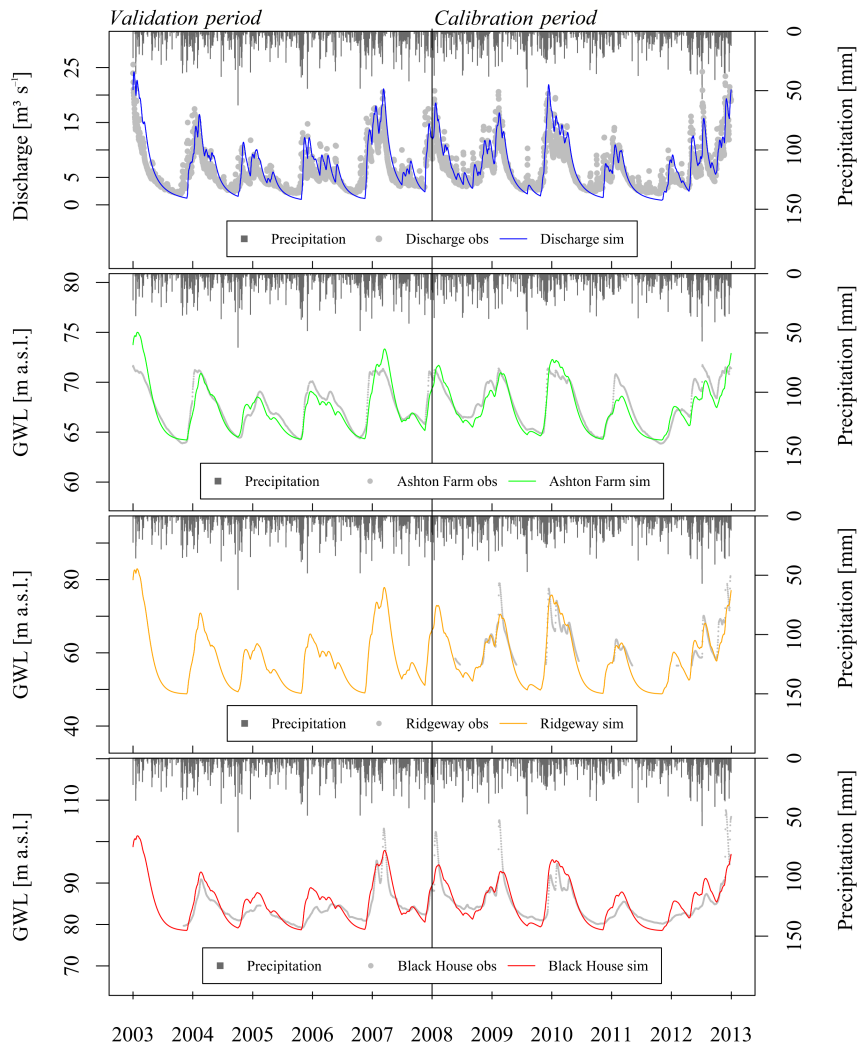
where  $x$  stands for the timescale (years, months, weeks, days) and  $p$  is the respective percentile. To better compare the deviations for different percentiles we normalise the MAD to a percentage of mean absolute deviation (PAD) with the total number of days of the chosen timescale:

$$\text{PAD}_p = \frac{\text{MAD}_p}{dp_x} \times 100 \quad [\%], \quad (4)$$

where  $dp_x$  is a normalising constant standing for the total number of days of the respective timescale and percentile. For example, if we take the timescale of months and the 75th percentile of exceedances we get a  $dp_x$  of  $(100 - 75) \% \times (365.25/12)$  days. To evaluate the prediction performance of the approach, percentiles are derived from the daily data of the calibration period and then applied to the validation period, similarly to the split sample test in Sect. 3.3. In this way we are able to evaluate our model over different thresholds and in terms of temporal resolution.

### 3.5 Establishment of simple climate scenarios and assessment of groundwater level frequency distributions

Given the model performance assessment above, we use our approach to assess future changes of groundwater level frequencies at our study site. We derive projections of future precipitation and potential evapotranspiration by manipulating our observed “baseline” climate data. We extract distributional samples of percentage changes of precipitation and



**Figure 4.** Modelled discharge ( $\text{m}^3 \text{s}^{-1}$ ) and groundwater levels (m a.s.l.) at the boreholes Ashton Farm, Ridgeway and Black House.

evaporation from the UK probabilistic projections of climate change over land (UKCP09) for (1) a low-emission scenario and (2) a high-emission scenario for the time period of 2070–2099. This enables us to capture, in a pragmatic and computationally efficient approach, for the two emission scenarios the general range of changes for the most pertinent variables that we think will most impact changes to monthly seasonal GW responses. We focus on projected median delta values for change in mean temperature ( $^{\circ}\text{C}$ ) and precipitation (%), as well as the respective 25th and 75th percentile from the probabilistic projections, and apply them to our input data. For our model input we transfer projected temperatures into evapotranspiration via the Thornthwaite equation (Thornthwaite, 1948). In this way, we obtain  $3 \times 3$  projections ( $3 \times$  precipitation and  $3 \times$  evapotranspiration) for each of the emission scenarios that also address the uncertainty associated with the projections. The resulting simulations will pro-

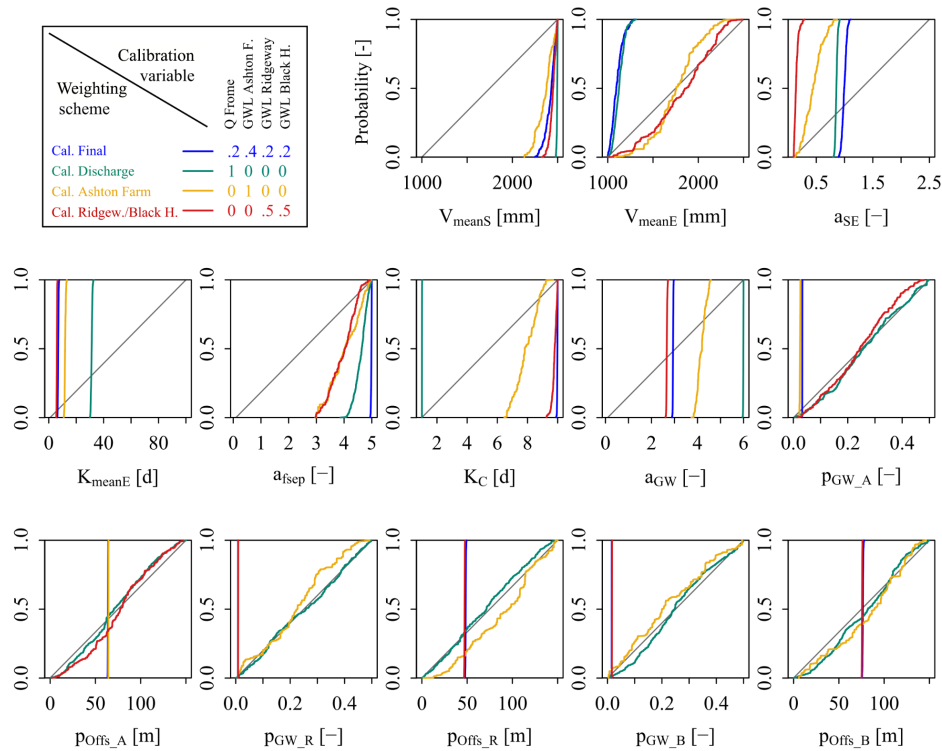
vide an estimate of possible future changes of groundwater level frequencies for the two emission scenarios including an assessment of their uncertainty.

## 4 Results

### 4.1 Model calibration and evaluation

Table 2 shows the optimised parameter values as well as the model performance. The simulation of the discharge shows KGE values of 0.73 and 0.58 in the calibration and validation periods respectively. The borehole simulations show high KGE values and only slight deteriorations in the validation period. The parameters are located well within their predefined ranges. Mean soil storage  $V_{\text{mean}, \text{S}}$  and mean epikarst storage  $V_{\text{mean}, \text{E}}$  are 2015.6 and 1011.7 mm respectively. The porosity parameter at Ashton Farm is the highest, followed





**Figure 5.** Cumulative parameter distributions (blue) of all model parameters; strong deviation from the 1 : 1 (dark grey) indicates good identifiability.

by the borehole at Black House. Ridgeway shows the smallest porosity value. For Ashton Farm and Blackhouse the calibration chose the groundwater storage compartment 7, while for Ridgeway it chose the compartment number 8.

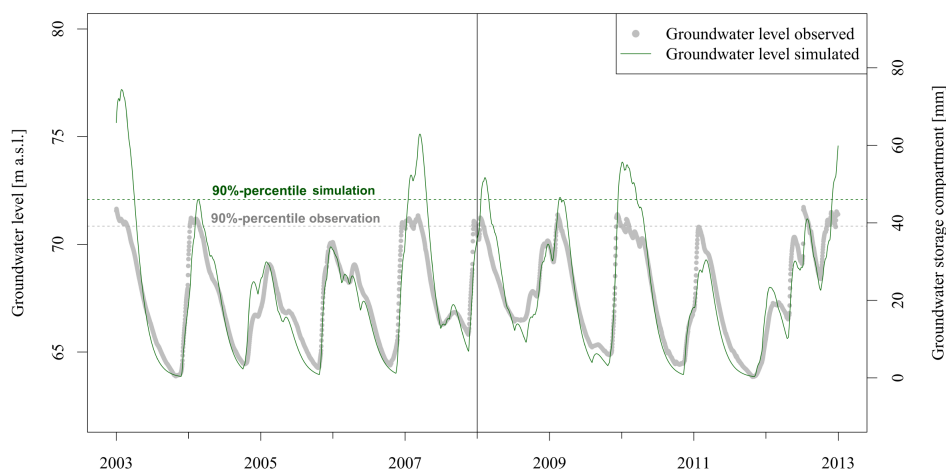
Figure 4 plots the observations against simulations for the calibration and validation periods. Modelled discharge generally matches the seasonal behaviour of the observations. However, some low-flow peaks are not depicted well in the simulation. When looking at the groundwater levels, the simulation of Ashton Farm appears to be most adequate. However, there are considerable periods when differences from the observations can be found for all wells. Simulations at Ridgeway and Black House show moderate performance in capturing peak groundwater levels. Notably, the simulation at Black House is slightly better in the validation period. The cumulative parameter distributions derived by SCEM indicate that the model parameters were easily identifiable when we use all available data (Fig. 5), while some parameters remain difficult to identify when only parts of the available data were used for calibration. Here identifiability of parameters is simply the extent to which the cumulative parameter distributions span the sampled parameter limits, where highly constrained parameter distributions are classed as identifiable. For instance, when only discharge was used for calibration (green lines), the parameters related to groundwater (porosity  $p_{GW}$  and groundwater level offset  $\Delta h$ ) happen

to be unidentifiable. In addition, the groundwater parameters are only identifiable when their respective time series is considered (i.e. the yellow and blue lines at  $p_{GW,A}$  and  $\Delta h_{GW,A}$ ). In turn, the epikarst storage  $V_{mean,E}$  is not identifiable when only the groundwater well data are used (yellow and red lines). We also note, as we would expect, that the final cumulative parameter distributions occur in different parts of the parameter space depending on the combination of performance metrics from different observations.

## 4.2 The percentile approach

When simulated peak values of groundwater levels are compared to the observations, we find a rather moderate agreement. Using the percentile approach we find different thresholds that exceed our selected groundwater level percentiles. This is elaborated for the 90th percentile of simulated and observed groundwater levels of Ashton farm (Fig. 6).

Table 3 shows the mean observed and modelled exceedances of all selected thresholds (the 5th, 10th, 25th, 50th, 75th, 90th and 95th percentiles) at all temporal resolutions in the validation period. By comparing matches in the number days of exceedance we evaluate our model with different percentiles and timescales. The left value is the mean absolute deviation (MAD) and the right value is the percentage of absolute deviation (PAD). We can see that the higher the percentile, the larger the deviation between observed and mod-



**Figure 6.** Illustration of the percentile approach. Time series of the observed (grey dots) and modelled (green line) groundwater levels at Ashton Farm. The dotted lines represent the respective 90th percentile.

**Table 3.** Deviations of simulated to observed exceedances of different percentiles in the validation period (borehole: Ashton Farm). The left value is the mean absolute deviation MAD (d), the right value is the deviation percentage PAD (%).

| Time period   | Percentiles |            |            |             |            |             |              |
|---------------|-------------|------------|------------|-------------|------------|-------------|--------------|
|               | 5           | 10         | 25         | 50          | 75         | 90          | 95           |
| 5 years       | 5.00/0.29   | 30.00/1.83 | 38.00/2.77 | 16.00/1.75  | 1.26/5.04  | 19.00/10.40 | 90.00/98.56  |
| Years         | 2.60/0.75   | 13.60/4.14 | 14.40/5.26 | 21.20/11.61 | 4.33/17.30 | 19.80/54.21 | 26.00/142.37 |
| Years–seasons | 0.65/0.75   | 4.10/4.99  | 3.60/5.26  | 6.90/15.11  | 6.74/26.94 | 6.45/70.64  | 6.50/142.37  |
| Months        | 0.22/0.75   | 1.37/4.99  | 1.20/5.26  | 2.73/17.96  | 7.94/31.76 | 2.58/84.87  | 2.23/146.75  |
| Weeks         | 0.05/0.74   | 0.33/5.27  | 0.27/5.18  | 0.61/17.36  | 7.82/31.27 | 0.58/83.56  | 0.54/153.10  |
| Days          | 0.01/0.75   | 0.05/5.35  | 0.04/5.26  | 0.09/17.96  | 7.94/31.76 | 0.08/84.88  | 0.08/159.91  |

elled exceedances. The same is true for the PAD when moving from lower to higher temporal resolutions. The MAD decreases with higher temporal resolution.

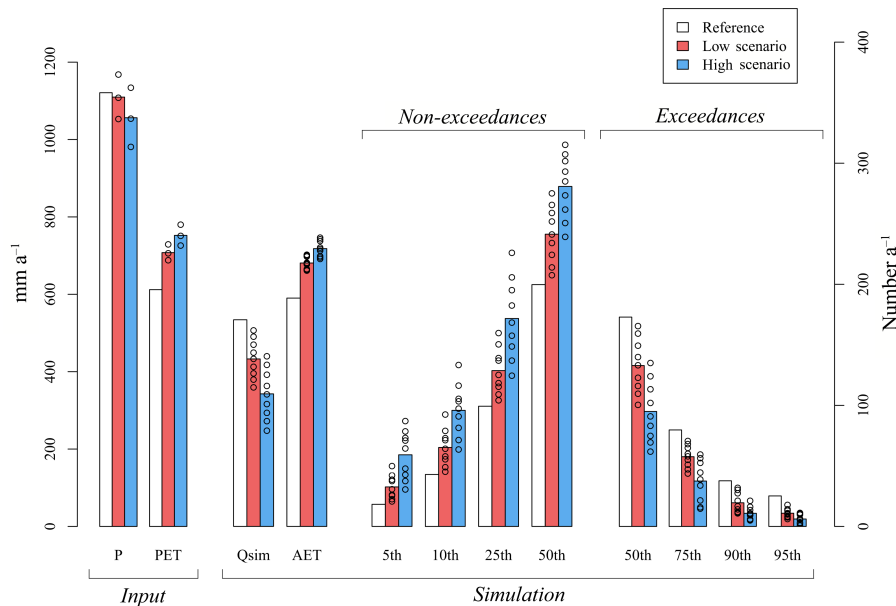
#### 4.3 Impact of simulated climate changes on groundwater level distributions

The results of applying the two climate projections to the model can be found at Table 4 and in Fig. 7. They display the mean model outputs ( $Q_{\text{sim}}$ , AET) and mean exceedances per year, calculated on the basis of our modelled time series. Both emission scenarios (low and high) lead to increased modelled actual evapotranspiration and to decreased discharge simulations. In addition, both emission scenarios show a substantial reduction in exceedances of high percentiles. We also find that the standard error of the exceedances and non-exceedances of high-emission scenario tends to be higher than the standard error of the low-emission scenario.

## 5 Discussion

### 5.1 Reliability of the simulations

A decrease in the simulation performance in the validation period is normally to be expected because there is always a tendency to compensate for structural limitations and observational uncertainties during the calibration. The low decrease in model performance from 11 % (groundwater prediction at Black House,  $KGE_{\text{GW,B}}$ ) to 21 % (discharge prediction,  $KGE_{\text{Q}}$ ) during the validation period indicates a certain robustness of the calibrated parameters and is comparable to split sample tests in other studies (e.g. Parajka et al., 2007; Perrin et al., 2001), although we have to acknowledge that for other applications a higher degree of robustness may be required. In addition, it is corroborated by their generally mainly high identifiability derived by SCEM for the final calibration scheme that used all four available observed discharge and groundwater level time series. By applying the shuffled complex evolution Metropolis algorithm and stepwise increasing the calibration data (only discharge, only groundwater, all together), we show that dis-



**Figure 7.** Mean model input ( $\text{mm a}^{-1}$ ), mean modelled output ( $\text{mm a}^{-1}$ ) and mean (non-)exceeded percentiles (number  $\text{a}^{-1}$ ) in the reference period and both scenarios (borehole: Ashton Farm; future period: 2070–2099). The circles indicate the spread among the nine realisations for each of the two scenarios.

**Table 4.** Model output and (non-)exceedances of percentiles in the reference period and the two scenarios (borehole: Ashton Farm, time period 2070–2099).

| Scenario  | Qsim<br>$\text{mm a}^{-1}$ | AET<br>$\text{mm a}^{-1}$ | 5th<br>non exc $\text{a}^{-1}$ | 10th<br>non exc $\text{a}^{-1}$ | 25th<br>non exc $\text{a}^{-1}$ | 50th<br>exc $\text{a}^{-1}$ | 75th<br>exc $\text{a}^{-1}$ | 90th<br>exc $\text{a}^{-1}$ | 95th<br>exc $\text{a}^{-1}$ |
|-----------|----------------------------|---------------------------|--------------------------------|---------------------------------|---------------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| Reference | 534                        | 590                       | 17.6                           | 41.3                            | 95.6                            | 172.9                       | 79.7                        | 37.7                        | 25.2                        |
| Low       | 433                        | 681                       | 31.4                           | 62.8                            | 123.9                           | 132.9                       | 57.6                        | 19.5                        | 10.9                        |
| High      | 343                        | 718                       | 57.0                           | 92.3                            | 165.3                           | 94.9                        | 37.5                        | 10.9                        | 6.1                         |

charge data alone, as well as groundwater data alone, do not provide enough information to identify all of our model parameters, as the posteriors of some of the model parameters remain close to a uniform distribution. This is similar to the work of Schoups and Vrugt (2010), who found unidentifiable parameter values with their models calibrating only against discharge. The different calibration schemes visualised in the cumulative parameter distributions show that initially unidentifiable parameters become identifiable when the related time series is considered. Using all information, all model parameters are identifiable, which is in accordance with preceding research that showed the usefulness of multi-objective approaches. For instance, Kuczera and Mroczkowski (1998) demonstrated that a combination of groundwater and discharge observations can reduce parameter uncertainty. As we were mostly focussing on the difference among the calibration steps with increasing data, the use of KGE as an informal likelihood measure seems justifiable.

A look at the parameter values reveals an adequate reflection of reality. However,  $V_{\text{mean}, S}$  and  $V_{\text{mean}, E}$  are quite high considering that initial ranges for these parameters were 0–250/0–500 mm (Hartmann et al., 2013a, c). As previous studies took place in fairly dry catchments, the ranges were extended substantially to deal with the wetter climate in southern England. A high  $a_{SE}$  indicates high variability in soil and epikarst thicknesses, favouring lateral karstic flow concentration (Ford and Williams, 2007). Butler et al. (2012) notes that the unsaturated zone of the Chalk is highly variable, ranging from almost zero near the rivers to over 100 m in interflues.

Additionally, the mean epikarst storage coefficient  $K_{\text{mean}, E}$  is quite low, indicating fast water transport from the epikarst to the groundwater storage which is in accordance to other studies (e.g. Aquilina et al., 2006). The value of parameter  $a_{f\text{sep}}$  indicates that a significant part of the recharge is diffuse. A moderately high conduit storage coefficient  $K_C$  and a high  $a_{GW}$  indicate that there is a significant contribution of slow pathways by the matrix system. A rather low value that is sensitive to  $K_C$  was found when calibrating only

by discharge operations, indicating some interactions of  $K_C$  with other model parameters (Saltelli et al., 2008). This is in accordance with the findings by Jones and Cooper (1998) as well as by Reeves (1979), who reported 30 and 10–20 % of the recharge occurring through (macro-) fissures in Chalk catchments. Although groundwater flow in the chalk is dominated by the matrix, given antecedent wet conditions, fracture flow can increase significantly (Lee et al., 2006; Ireson and Butler, 2011; Butler et al., 2012). Overall, a split-sample test, parameter identifiability analysis, realistic values of parameters and plausible simulation results provide a strong indication of reliable model functioning.

## 5.2 Performance of the percentile approach

Based on the idea of the standardised precipitation or groundwater indices (Lloyd-Hughes and Saunders, 2002; Bloomfield and Marchant, 2013) our percentile approach permits the performance of the model to be improved to reflect observed groundwater level exceedances. It yields acceptable performance for years to days up to the 90th percentile. A reduction of precision with the timescale is obvious but in an acceptable order of magnitude when the validation period is considered. Although deviations are considerable both in the calibration and validation periods, they are stable demonstrating certain robustness but also the limitations of our approach. Although the variable model structure of the VarKarst model was shown to provide more realistic results than commonly used lumped models (Hartmann et al., 2013b) it still simplifies a karst system's natural complexity. This can be seen in the simulated time series at Ashton Farm and Black House, which indicate an overestimation of high levels and an underestimation of low levels. The reason for this behaviour might be due to the modelling assumption of a constant vertical porosity, despite the knowledge that there can be a strongly non-linear relation between chalk transmissivity and depth. Several studies acknowledge that hydraulic conductivity in the Chalk follows a non-linear decreasing trend with depth (Allen et al., 1997; Wheeler et al., 2007; Butler et al., 2009). This is mainly attributed to the decrease in fractures because of the increasing overburden and absence of water level fluctuations (Williams et al., 2006; Butler et al., 2012). Hydraulic conductivities in the Chalk can span several orders of magnitude (Butler et al., 2009) and are particularly enhanced at the zone of water table fluctuations (Williams et al., 2006). In addition, cross-flows occurring in the aquifer can lead to complicated system responses in the Chalk (Butler et al., 2009). For the sake of a parsimonious model structure, these characteristics were omitted in this study but their future consideration could help to improve the simulations if information about the depth profile of permeability is available. A decrease in performance was also found for standardised indices that use probability distributions instead of a simulation model (Vicente-Serrano et al., 2012; Núñez et al., 2014; Van Lanen et al., 2016). To im-

prove the approach's reliability for higher groundwater level percentiles, a model calibration that is more focussed on the high groundwater level percentiles may be a promising direction. A consideration of the time spans above the 90th percentile will allow for a better simulation quality. This could be further evaluated by using different percentile weighting schemes and stepwise increasing the weight on the target percentile.

## 5.3 Applicability and transferability of our approach

We prepared two scenarios by manipulating our input data using probabilistic projections of annual changes in precipitation and potential evaporation at 2070–2099 for a low- and a high-emission scenario. This may neglect some of the changes on climate patterns predicted by climate projections but it is based on local and real meteorological values of the reference period, therefore avoiding problems that arise when historic and climate projection data show pronounced mismatches during their overlapping periods. Our results revealed that both scenarios lead to fewer exceedances over higher percentiles and more non-exceedances of lower percentiles, indicating a higher risk of groundwater drought at our study site. However, one problem that arises from our approach is that we do not consider changes in the seasonal patterns of our input variable, for example the increase in winter precipitation. If this increase was considered, the results would probably yield more exceedances of higher percentiles, as for instance found by Jimenez-Martinez et al. (2016). The purpose of the simple climate scenarios was to provide an application example of the new methodology, which is rather hypothetical considering the large uncertainties of current climate projections. We believe that our nine realisations are sufficient to show that different possible future changes have a non-linear impact on groundwater level frequencies. Although quite simplistic, our results are qualitatively in accordance with previous studies indicating increased occurrence of droughts in the UK (Burke et al., 2010; Prudhomme et al., 2014). The risk of drought occurrences might increase depending on the magnitude of change in evapotranspiration. However, more research and the application of more elaborated scenarios are necessary to completely understand the consequences of the change in groundwater frequency patterns in the UK chalk regions.

As the VarKarst model is a process-based model that includes the relevant characteristics of karst systems over range of climatic settings (Hartmann et al., 2013b), our approach can to some extent be used to assess future changes of groundwater level distributions and also be applied in other regions. This may bring some advantage concerning approaches that used transfer functions (Jimenez-Martinez et al., 2016) or regression models (Adams et al., 2010) for estimating groundwater levels if enough data for model calibration and evaluation are available.

As has been noted by Cobby et al. (2009), the likelihood and depth of groundwater inundations is one of the major challenges for future research of groundwater flooding. Since it is a lumped approach it may provide, after Butler et al. (2012), “a good indication of the likelihood of groundwater flooding, but do[es] not indicate where the flooding will take place”. A spatial determination of the groundwater table such as that in Upton and Jackson (2011) would be possible but only in catchments where the borehole network is extensive. Thereby, the possibility to model several boreholes with one single calibration, due to compartment structure in VarKarst, might be also an advantage. Butler et al. (2012) noted that the parameterisation of the unsaturated zone is a major difficulty in the Chalk region. Since this study also struggles with the porosity, future work should take a closer look at this subject.

## 6 Conclusions

We used an existing process-based lumped karst model to simulate groundwater levels in a chalk catchment in south-western England. Groundwater levels were simulated by translating the modelled groundwater storage into groundwater levels with a simple linear relationship. To evaluate our approach we analysed the agreement of observed and simulated groundwater level exceedances for different percentiles. Finally, a simple scenario analysis was undertaken to investigate the potential future changes of groundwater level frequencies that affect the risk of groundwater flooding as well as the risk of groundwater droughts. The model performance for discharge and the groundwater levels was satisfying and

showed the general adequacy of the model for simulating groundwater levels in the chalk. It also revealed shortcomings concerning higher groundwater levels. This was corroborated by the percentile approach that showed a robust performance up to the 90th percentile. A scenario analysis using UKCP projections on expected regional climate changes showed that expected changes may lead to an increased occurrence of low groundwater levels due to increasing actual evaporation. Overall, our study shows that semi-distributed process-based modelling can be a valuable tool for simulating and predicting groundwater frequencies in Chalk regions where information is too limited for the application of distributed models. Here, a thorough model evaluation is essential for obtaining reliable and consistent results. In order to obtain more reliable results we recommend collecting more data about the hydrogeological properties of our study site to improve the structure of our model regarding the porosity and the unsaturated zone. In addition, longer time series and an adapted calibration approach which, in particular, emphasises the > 90th percentiles of groundwater levels could significantly improve our simulations. In addition we propose applying the method to other catchments to test the transferability of our approach and to quantify the variability of climate change impacts over a wide range of Chalk catchments across the UK.

*Data availability.* The sources of all underlying data are described in Table 1. All precipitation, discharge and potential evaporation data are freely available from the CEH website. Groundwater level observations and climate delta values can be accessed via request at EA and UKCP, respectively.

Appendix A

Within the VarKarst model, the parameter  $V_{\text{mean}, S}$  (mm) and the distribution coefficient  $a_{SE}$  (–) define the variation of soil storage capacities across the  $N$  model compartments. They are used to calculate the soil storage capacity  $V_{S,i}$  (mm) for every compartment  $i$  by Eqs. (3) and (4) in Table A1. We apply the same distribution coefficient  $a_{SE}$  when we derive the epikarst storage distribution by the mean epikarst depth  $V_{\text{mean}, E}$  (mm) (Eqs. 6 and 7 in Table A1). We determine actual evapotranspiration from each soil compartment.  $E_{\text{act},i}$  is calculated by reducing potential evapotranspiration using the Thornthwaite equation (Thornthwaite, 1948) and the soil saturation deficit (Eq. 1 in Table A1). Surface run-off is found by the excess of soil and epikarst storage of the previous model compartment (Eq. 2 in Table A1). With surface run-off and actual evapotranspiration, the stored water volume at each soil compartment  $V_{\text{Soil},i}$  (mm) can be calculated by simply applying water balance.

The recharge from the soil to the epikarst  $R_{\text{Epi},i}$  (mm) is calculated by the excess of the soil storage (Eq. 5 in Table A1), while the epikarst outflow follows a linear storage assumption (Eqs. 8 and 9 in Table A1). Again, water balance allows the stored water  $V_{\text{Epi},i}$  (mm) to be determined at each

time step  $t$  and each epikarst compartment  $i$ . The downward flux from the epikarst considers diffuse ( $R_{\text{diff},i}$ , mm) and concentrated groundwater recharge ( $R_{\text{conc},i}$ , mm) component that are found by a variable separation factor  $f_{C,i}$  (–) and a distribution coefficient  $a_f$  (–) (Eqs. 10, 11 and 12 in Table A1). The diffuse component recharges the groundwater compartments beneath the respective epikarst layers ( $i = 1 \dots N - 1$ ). The concentrated component flows laterally to compartment  $i = N$  and therefore recharges the conduit system.

Similarly to the epikarst compartment, variable groundwater storage coefficients  $K_{\text{GW},i}$  (d) are calculated (Eq. 15 in Table A1) and applied to calculate the discharges of the matrix system (Eq. 13 in Table A1) and the conduit system (Eq. 14 in Table A1), which together sum up to the entire discharge of the system (Eq. 15 in Table A1). Knowing groundwater recharge and groundwater discharge for each model compartment  $i$  again allows the stored volume of water to be determined within the groundwater compartment  $V_{\text{GW},i}$  at time step  $t$ , which is used to simulate the groundwater levels (Eq. 1 in Sect. 3.1).

Table A1. Parameters, descriptions and equations solved in the VarKarst model.

| Model routine | Variable                 | Description                                    | Equation  | Unit               | Equation |
|---------------|--------------------------|--|---|--------------------|----------|
| Soil          | $E_{\text{act},i}(t)$    | Actual evapotranspiration                      | $= E_{\text{pot}}(t) \frac{\min[V_{\text{Soil},i}(t)+P(t)+Q_{\text{Surface},i}(t), V_{S,i}]}{V_{S,i}}$          | $\text{mm d}^{-1}$ | (1)      |
|               | $Q_{\text{Surf},i+1}(t)$ | Surface flow to the next model compartment     | $= \max[V_{\text{Epi},i}(t) + R_{\text{Epi},i}(t) - V_{S,i}, 0]$  | $\text{mm d}^{-1}$ | (2)      |
|               | $V_{\text{max}, S}$      | Maximum soil storage capacity                  | $= V_{\text{mean}, S} 2^{\left(\frac{a_{SE}}{a_{SE}+1}\right)}$   | mm                 | (3)      |
|               | $V_{S,i}$                | Soil storage distribution                      | $= V_{\text{max}, S} \left(\frac{i}{N}\right)^{a_{SE}}$   | mm                 | (4)      |
|               | $R_{\text{Epi},i}(t)$    | Recharge to the epikarst                       | $= \max[V_{\text{Soil},i}(t) + P(t) + Q_{\text{Surface},i}(t) - E_{\text{act},i}(t) - V_{S,i}, 0]$              | $\text{mm d}^{-1}$ | (5)      |
| Epikarst      | $V_{\text{max}, E}$      | Maximum epikarst storage capacity              | $= V_{\text{mean}, E} 2^{\left(\frac{a_{SE}}{a_{SE}+1}\right)}$   | mm                 | (6)      |
|               | $V_{E,i}$                | Epikarst storage distribution                  | $= V_{\text{max}, E} \left(\frac{i}{N}\right)^{a_{SE}}$   | mm                 | (7)      |
|               | $Q_{\text{Epi},i}(t)$    | Outflow of the epikarst                        | $= \frac{\min[V_{\text{Epi},i}(t)+R_{\text{Epi},i}(t)+Q_{\text{Surface},i}(t), V_{E,i}]}{K_{E,i}} \Delta t$     | $\text{mm d}^{-1}$ | (8)      |
|               | $K_{E,i}$                | Epikarst storage coefficient                   | $= K_{\text{max}, E} \left(\frac{N-i+1}{N}\right)^{a_{SE}}$   | d                  | (9)      |
|               | $R_{\text{diff},i}(t)$   | Diffuse recharge                               | $= f_{C,i} Q_{\text{Epi},i}(t)$   | $\text{mm d}^{-1}$ | (10)     |
|               | $R_{\text{conc},i}(t)$   | Concentrated recharge                          | $= (1 - f_{C,i}) Q_{\text{Epi},i}(t)$   | $\text{mm d}^{-1}$ | (11)     |
|               | $f_{C,i}$                | Recharge separation factor                     | $= \left(\frac{i}{N}\right)^{a_{fsep}}$   | –                  | (12)     |
| Groundwater   | $Q_{\text{GW},i}(t)$     | Groundwater contributions of the matrix        | $= \frac{V_{\text{GW},i}(t)+R_{\text{diff},i}(t)}{K_{\text{GW},i}}$   | $\text{mm d}^{-1}$ | (13)     |
|               | $Q_{\text{GW},N}(t)$     | Groundwater contribution of the conduit system | $= \frac{\min[V_{\text{GW},N}(t)+\sum_{i=1}^N R_{\text{conc},i}(t), V_{\text{crit}, \text{OF}}]}{K_C} \Delta t$ | $\text{mm d}^{-1}$ | (14)     |
|               | $K_{\text{GW},i}$        | Variable groundwater storage coefficient       | $= K_C \left(\frac{N-i+1}{N}\right)^{-a_{\text{GW}}}$   | d                  | (15)     |
|               | $Q_{\text{main}}(t)$     | Discharge                                      | $= \frac{A_{\text{max}}}{N} \sum_{i=1}^N Q_{\text{GW},i}(t)$  | $\text{ls}^{-1}$   | (16)     |



*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* This publication contains Environment Agency information © Environment Agency and database rights. Thanks to Jens Lange and Sophie Bachmair, University of Freiburg, for their valuable advice. Support for Gemma Coxon, Jim Freer and Nicholas J. K. Howden was provided by NERC MaRIUS: Managing the Risks, Impacts and Uncertainties of droughts and water Scarcity, grant number NE/L010399/1. The article processing charge was funded by the German Research Foundation (DFG) and the University of Freiburg in the funding programme Open Access Publishing.

Edited by: Mario Parise

Reviewed by: Andrew Long and two anonymous referees

## References

- Adams, B., Peach, D. W. D., and Bloomfield, J. P. J.: The LOCAR hydrogeological infrastructure for the Frome/Piddle catchment, British Geological Survey Internal Report, IR/03/179, 1–3, Keyworth, Nottingham, 2003.
- Adams, B., Bloomfield, J. P., Gallagher, A. J., Jackson, C. R., Rutter, H. K., and Williams, A. T.: An early warning system for groundwater flooding in the Chalk, *Q. J. Eng. Geol. Hydrogeol.*, 43, 185–193, <https://doi.org/10.1144/1470-9236/09-026>, 2010.
- Allen, D. J., Brewerton, L. J., Coleby, L. M., Gibbs, B. R., Lewis, M. A., MacDonald, A. M., Wagstaff, S. J., and Williams, A. T.: The physical properties of major aquifers in England and Wales, British Geological Survey Technical Report WD/97/34, 312 pp., Environment Agency R&D Publication 8, 1997.
- Aquilina, L., Ladouche, B., and Dörfliger, N.: Water storage and transfer in the epikarst of karstic systems during high flow periods, *J. Hydrol.*, 327, 472–485, 2006.
- Bakalowicz, M.: Karst groundwater: a challenge for new resources, *Hydrogeol. J.*, 13, 148–160, 2005.
- Bennett, C.: South Winterbourne Flood Investigation, Prepared for Dorset County Council, Parsons Brinckerhoff, Bristol, 2013.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36, 2006.
- Birk, S., Liedl, R., and Sauter, M.: Karst spring responses examined by process-based modeling, *Groundwater*, 44, 832–836, 2006.
- Blasone, R.-S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A., and Zyvoloski, G. A.: Generalized Likelihood Uncertainty Estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, *Adv. Water Resour.*, 31, 630–648, 2008.
- 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, <https://doi.org/10.5194/hess-17-4769-2013>, 2013.
- 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, <https://doi.org/10.5194/hess-19-4327-2015>, 2015.
- Bonacci, O., Ljubenkov, I., and Roje-Bonacci, T.: Karst flash floods: an example from the Dinaric karst (Croatia), *Nat. Hazards Earth Syst. Sci.*, 6, 195–203, <https://doi.org/10.5194/nhess-6-195-2006>, 2006.
- Brunner, P., Dennis, I., and Girvan, J.: River from geomorphological assessment and rehabilitation plan, Tech. Rep., (9V7431/R//Exet), 2010.
- Burke, E. J., Perry, R. H. J., and Brown, S. J.: An extreme value analysis of UK drought and projections of change in the future, *J. Hydrol.*, 388, 131–143, <https://doi.org/10.1016/j.jhydrol.2010.04.035>, 2010.
- Butler, A. P., Mathias, S. A., and Gallagher, A. J.: Analysis of flow processes in fractured chalk under pumped and ambient conditions (UK), *Hydrogeol. J.*, 17, 1849–1858, <https://doi.org/10.1007/s10040-009-0477-4>, 2009.
- Butler, A. P., Hughes, A. G., Jackson, C. R., Ireson, A. M., Parker, S. J., Wheatler, H. S., and Peach, D. W.: Advances in modelling groundwater behaviour in Chalk catchments, *Geol. Soc. London Spec. Publ.*, 364, 113–127, 2012.
- Cobby, D., Morris, S., Parkes, A., and Robinson, V.: Groundwater flood risk management: advances towards meeting the requirements of the EU floods directive, *J. Flood Risk Manag.*, 2, 111–119, <https://doi.org/10.1111/j.1753-318X.2009.01025.x>, 2009.
- Dorset County Council: A Local Climate Impacts Profile for Dorset, Dorset County Council, Dorchester, 2009.
- Duan, Q., Sorooshian, S., Gupta, H. V., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031, <https://doi.org/10.1029/91WR02985>, 1992.
- Edmonds, C. N.: Towards the prediction of subsidence risk upon the Chalk outcrop, *Q. J. Eng. Geol. Hydrogeol.*, 16, 261–266, <https://doi.org/10.1144/GSL.QJEG.1983.016.04.03>, 1983.
- Engeland, K., Xu, C.-Y., and Gottschalk, L.: Assessing uncertainties in a conceptual water balance model using Bayesian methodology/Estimation bayésienne des incertitudes au sein d'une modélisation conceptuelle de bilan hydrologique, *Hydrolog. Sci. J.*, 50, 45–63, <https://doi.org/10.1623/hysj.50.1.45.56334>, 2005.
- Environment Agency: Frome and Piddle Catchment Flood Management Plan, Environment Agency Manley House Kestrel Way Exeter, 2012.
- Fitzpatrick, C. M.: The Hydrogeology of Bromate Contamination in the Hertfordshire Chalk: Double-Porosity Effects on Catchment-Scale Evolution, University College London, London, 2011.
- Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., and Dörfliger, N.: Modelling the hydrologic functions of a karst aquifer under active water management – the Lez spring, *J. Hydrol.*, 365, 235–243, 2009.
- Ford, D. and Williams, P. D.: Karst hydrogeology and geomorphology, John Wiley & Sons, West Sussex, UK, 578 pp., 2007.
- Geyer, T., Birk, S., Liedl, R., and Sauter, M.: Quantification of temporal distribution of recharge in karst systems from spring hydrographs, *J. Hydrol.*, 348, 452–463, 2008.
- Goldscheider, N. and Drew, D.: Methods in Karst Hydrogeology, I. A. of Hydrogeologists, Taylor & Francis Group, Leiden, the Netherlands, 2007.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, 2009.
- Gutiérrez, F.: Hazards associated with karst, in: Geomorphological Hazards and Disaster Prevention, edited

- by: Alcántara-Ayala, I. and Goudie, A., 161–175, <https://doi.org/10.1017/CBO9780511807527.013>, 2010.
- Gutiérrez, F., Parise, M., De Waele, J., and Jourde, H.: Earth-science reviews a review on natural and human-induced geohazards and impacts in karst, *Earth-Sci. Rev.*, 138, 61–88, <https://doi.org/10.1016/j.earscirev.2014.08.002>, 2014.
- Hartmann, A., Lange, J., Weiler, M., Arbel, Y., and Greenbaum, N.: A new approach to model the spatial and temporal variability of recharge to karst aquifers, *Hydrol. Earth Syst. Sci.*, 16, 2219–2231, <https://doi.org/10.5194/hess-16-2219-2012>, 2012.
- Hartmann, A., Weiler, M., Wagener, T., Lange, J., Kralik, M., Humer, F., Mizyed, N., Rimmer, A., Barberá, J. A., Andreo, B., Butscher, C., and Huggenberger, P.: Process-based karst modelling to relate hydrodynamic and hydrochemical characteristics to system properties, *Hydrol. Earth Syst. Sci.*, 17, 3305–3321, <https://doi.org/10.5194/hess-17-3305-2013>, 2013a.
- Hartmann, A., Barberá, J. A., Lange, J., Andreo, B., and Weiler, M.: Progress in the hydrologic simulation of time variant recharge areas of karst systems – exemplified at a karst spring in Southern Spain, *Adv. Water Resour.*, 54, 149–160, <https://doi.org/10.1016/j.advwatres.2013.01.010>, 2013b.
- Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H., and Weiler, M.: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures, *Water Resour. Res.*, 49, 3345–3358, 2013c.
- Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M.: Karst water resources in a changing world: review of hydrological modeling approaches, *Rev. Geophys.*, 52, 218–242, <https://doi.org/10.1002/2013rg000443>, 2014a.
- Hartmann, A., Mudarra, M., Andreo, B., Marín, A., Wagener, T., and Lange, J.: Modeling spatiotemporal impacts of hydroclimatic extremes on groundwater recharge at a Mediterranean karst aquifer, *Water Resour. Res.*, 50, 6507–6521, 2014b.
- Hartmann, A., Kobler, J., Kralik, M., Dirnböck, T., Humer, F., and Weiler, M.: Model-aided quantification of dissolved carbon and nitrogen release after windthrow disturbance in an Austrian karst system, *Biogeosciences*, 13, 159–174, <https://doi.org/10.5194/bg-13-159-2016>, 2016.
- Hastings, W. K.: Monte Carlo sampling methods using markov chains and their applications, *Biometrika*, 57, 97–109, 1970.
- Hill, M. E., Stewart, M. T., and Martin, A.: Evaluation of the MODFLOW-2005 conduit flow process, *Ground Water*, 48, 549–559, <https://doi.org/10.1111/j.1745-6584.2009.00673.x>, 2010.
- Howden, N. J. K.: Hydrogeological controls on surface/groundwater interactions in a lowland permeable chalk catchment: implications for water quality and numerical modelling, Imperial College London (University of London), 2006.
- Ireson, A. M. and Butler, A. P.: Controls on preferential recharge to Chalk aquifers, *J. Hydrol.*, 398, 109–123, 2011.
- Jackson, C. R., Meister, R., and Prudhomme, C.: Modelling the effects of climate change and its uncertainty on UK Chalk groundwater resources from an ensemble of global climate model projections, *J. Hydrol.*, 399, 12–28, <https://doi.org/10.1016/j.jhydrol.2010.12.028>, 2011.
- Jackson, C. R., Bloomfield, J. P., and Mackay, J. D.: Evidence for changes in historic and future groundwater levels in the UK, *Prog. Phys. Geog.*, 39, 49–67, 2015.
- Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall–runoff model?, *Water Resour. Res.*, 29, 2637–2649, 1993.
- Jenkins, G. J., Perry, M. C., and Prior, M. J.: The climate of the United Kingdom and recent trends, Met Office Hadley Centre, Exeter, UK, available at: [http://www.ukcip.org.uk/wp-content/PDFs/UKCP09\\_Trends.pdf](http://www.ukcip.org.uk/wp-content/PDFs/UKCP09_Trends.pdf) (last access: December 2017), 2008.
- Jimenez-Martinez, J., Smith, M., and Pope, D.: Prediction of groundwater induced flooding in a chalk aquifer for future climate change scenarios, *Hydrol. Process.*, 30, 573–587, <https://doi.org/10.1002/hyp.10619>, 2016.
- Jones, H. K. and Cooper, J. D.: Water transport through the unsaturated zone of the Middle Chalk: a case study from Fleam Dyke lysimeter, *Geol. Soc. London, Spec. Publ.*, 130, 117–128, <https://doi.org/10.1144/GSL.SP.1998.130.01.11>, 1998.
- Jourde, H., Roesch, A., Guinot, V., and Bailly-Comte, V.: Dynamics and contribution of karst groundwater to surface flow during Mediterranean flood, *Environ. Geol.*, 51, 725–730, <https://doi.org/10.1007/s00254-006-0386-y>, 2007.
- Jukić, D. and Denić-Jukić, V.: Groundwater balance estimation in karst by using a conceptual rainfall–runoff model, *J. Hydrol.*, 373, 302–315, 2009.
- Klemeš, V.: Operational testing of hydrological simulation models, *Hydrolog. Sci. J.*, 31, 13–24, <https://doi.org/10.1080/02626668609491024>, 1986.
- Kuczera, G. and Mroczkowski, M.: Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, 34, 1481–1489, 1998.
- Kumar, R., Musuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., and Attinger, S.: Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator, *Hydrol. Earth Syst. Sci.*, 20, 1117–1131, <https://doi.org/10.5194/hess-20-1117-2016>, 2016.
- Ladouche, B., Marechal, J. C., and Dorfliger, N.: Semi-distributed lumped model of a karst system under active management, *J. Hydrol.*, 509, 215–230, <https://doi.org/10.1016/j.jhydrol.2013.11.017>, 2014.
- Lee, L. J. E., Lawrence, D. S. L., and Price, M.: Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England, *J. Hydrol.*, 330, 604–620, 2006.
- Lloyd, J. W.: The Hydrogeology of the Chalk in North-West Europe, edited by: Downing, R. A., Price, M., and Jones, G. P., Clarendon Press, Oxford, 220–249, 1993.
- Lloyd-Hughes, B. and Saunders, M. A.: A drought climatology for Europe, *Int. J. Climatol.*, 22, 1571–1592, <https://doi.org/10.1002/joc.846>, 2002.
- Macdonald, D., Dixon, A., Newell, A., and Hallaways, A.: Groundwater flooding within an urbanised flood plain, *J. Flood Risk Manag.*, 5, 68–80, 2012.
- Macdonald, D. M. J., Bloomfield, J. P., Hughes, A. G., MacDonal, A. M., Adams, B., and McKenzie, A. A.: Improving the understanding of the risk from groundwater flooding in the UK, FLOODrisk 2008, Eur. Conf. Flood Risk Manag. Oxford, UK, 30 September–2 October 2008, the Netherlands, available at: <http://nora.nerc.ac.uk/7760/> (last access: December 2016), 2008.
- Maloszewski, P., Stichler, W., Zuber, A., and Rank, D.: Identifying the flow systems in a karstic-fissured-porous aquifer, the Schneealpe, Austria, by modelling of environmental  $^{18}\text{O}$  and  $^3\text{H}$

- isotopes, *J. Hydrol.*, 256, 48–59, [https://doi.org/10.1016/S0022-1694\(01\)00526-1](https://doi.org/10.1016/S0022-1694(01)00526-1), 2002.
- Maurice, L. D., Atkinson, T. C., Barker, J. A., Bloomfield, J. P., Farrant, A. R., and Williams, A. T.: Karstic behaviour of groundwater in the English Chalk, *J. Hydrol.*, 330, 63–70, 2006.
- Maurice, L. D., Atkinson, T. C., Barker, J. A., Williams, A. T., and Gallagher, A. J.: The nature and distribution of flowing features in a weakly karstified porous limestone aquifer, *J. Hydrol.*, 438–439, 3–15, <https://doi.org/10.1016/j.jhydrol.2011.11.050>, 2012.
- McMillan, H. and Clark, M.: Rainfall–runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme, *Water Resour. Res.*, 45, W04418, <https://doi.org/10.1029/2008WR007288>, 2009.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092, 1953.
- Moutahir, H., Bellot, P., Monjo, R., Bellot, J., Garcia, M., and Touhami, I.: Likely effects of climate change on groundwater availability in a Mediterranean region of Southeastern Spain, 176, 161–176, <https://doi.org/10.1002/hyp.10988>, 2017.
- Murphy, J. M., Sexton, D. M. H., Jenkins, G. J., Boorman, P. M., Booth, B. B. B., Brown, C. C., Clark, R. T., Collins, M., Harris, G. R., Kendon, E. J., Betts, R. A., Brown, S. J., Howard, T. P., Humphrey, K. A., McCarthy, M. P., McDonald, R. E., Stephens, A., Wallace, C., Warren, R., Wilby, R., and Wood, R. A.: UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter, 2009.
- Naughton, O., Johnston, P. M., and Gill, L. W.: Groundwater flooding in Irish karst: the hydrological characterisation of ephemeral lakes (turloughs), *J. Hydrol.*, 470–471, 82–97, <https://doi.org/10.1016/j.jhydrol.2012.08.012>, 2012.
- Naughton, O., Johnston, P. M., McCormack, T., and Gill, L. W.: Groundwater flood risk mapping and management: examples from a lowland karst catchment in Ireland, *J. Flood Risk Manag.*, 10, 53–64, <https://doi.org/10.1111/jfr3.12145>, 2015.
- NRA: The Frome & Piddle Management Plan, Consultation Report, South Wessex Area of the National Rivers Authority (NRA), Dorset, 1995.
- Núñez, J., Rivera, D., Oyarzu'n, R., and Arumi', J. L.: On the use of standardized drought indices under decadal climate variability: critical assessment and drought policy implications, *J. Hydrol.*, 517, 458–470, <https://doi.org/10.1016/j.jhydrol.2014.05.038>, 2014.
- Oehlmann, S., Geyer, T., Licha, T., and Sauter, M.: Reducing the ambiguity of karst aquifer models by pattern matching of flow and transport on catchment scale, *Hydrol. Earth Syst. Sci.*, 19, 893–912, <https://doi.org/10.5194/hess-19-893-2015>, 2015.
- Parajka, J., Merz, R., and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling: case study in 320 Austrian catchments, *Hydrol. Process.*, 21, 435–446, <https://doi.org/10.1002/hyp.6253>, 2007.
- Parise, M.: Flood history in the karst environment of Castellana-Grotte (Apulia, southern Italy), *Nat. Hazards Earth Syst. Sci.*, 3, 593–604, <https://doi.org/10.5194/nhess-3-593-2003>, 2003.
- Parise, M.: Hazards in karst, in: Proceedings International Interdisciplinary Scientific Conference “Sustainability of the karst environment. Dinaric karst and other karst regions.”, Plitvice Lakes, Croatia, 2010.
- Parise, M., Ravbar, N., Živanović, V., Mikszewski, A., Kresic, N., Mádl-Szőnyi, J., J., and Kukurić, N.: Hazards in karst and managing water resources quality, in: Karst Aquifers – Characterization and Engineering, edited by: Stevanović, Z., Springer International Publishing, Cham, 601–687, 2015.
- Perrin, C., Michel, C., and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, 242, 275–301, 2001.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, 279, 275–289, 2003.
- Perrone, D. and Jasechko, S.: Dry groundwater wells in the western United States, *Environ. Res. Lett.*, 12, 104002, <https://doi.org/10.1088/1748-9326/aa8ac0>, 2017.
- Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment., *P. Natl. Acad. Sci. USA*, 111, 3262–3267, <https://doi.org/10.1073/pnas.1222473110>, 2014.
- Reeves, M. J.: Recharge and pollution of the English Chalk: some possible mechanisms, *Eng. Geol.*, 14, 231–240, 1979.
- Reimann, T., Geyer, T., Shoemaker, W. B., Liedl, R., and Sauter, M.: Effects of dynamically variable saturation and matrix-conduit coupling of flow in karst aquifers, *Water Resour. Res.*, 47, 1–19, <https://doi.org/10.1029/2011WR010446>, 2011.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, UK, 2008.
- Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, <https://doi.org/10.1029/2009WR008933>, 2010.
- Smith, P., Beven, K. J., and Tawn, J. A.: Informal likelihood measures in model assessment: theoretic development and investigation, *Adv. Water Resour.*, 31, 1087–1100, 2008.
- Stevanovic, Z.: Karst Aquifers-Characterization and Engineering, Springer, Berlin, Germany, 2015.
- Thornthwaite, C. W.: An approach toward a rational classification of climate, *Geogr. Rev.*, 38, 55–94, <https://doi.org/10.2307/210739>, 1948.
- Upton, K. A. and Jackson, C. R.: Simulation of the spatio-temporal extent of groundwater flooding using statistical methods of hydrograph classification and lumped parameter models, *Hydrol. Process.*, 25, 1949–1963, 2011.
- Van Lanen, H., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stage, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, *Hydrol.*

- Process., 30, 3097–3104, <https://doi.org/10.1002/hyp.10838>, 2016.
- Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C., and Morán-Tejeda, E.: Accurate computation of a streamflow drought index, *J. Hydrol. Eng.*, 17, 318–332, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000433](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000433), 2012.
- von Freyberg, J., Moeck, C., and Schirmer, M.: Estimation of groundwater recharge and drought severity with varying model complexity, *J. Hydrol.*, 527, 844–857, <https://doi.org/10.1016/j.jhydrol.2015.05.025>, 2015.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, <https://doi.org/10.1029/2002WR001642>, 2003.
- Wagener, T., Lees, M. J., and Wheater, H. S.: A toolkit for the development and application of parsimonious hydrological models, *Math. Model. Large Watershed Hydrol.*, 1, 87–136, 2002.
- Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, *Water Resour. Res.*, 52, 1847–1865, <https://doi.org/10.1002/2015WR017635>, 2016.
- Wheater, H. S., Bishop, K. H., and Beck, M. B.: The identification of conceptual hydrological models for surface water acidification, *Hydrol. Process.*, 1, 89–109, 1986.
- Wheater, H. S., Peach, D., and Binley, A.: Characterising groundwater-dominated lowland catchments: the UK Lowland Catchment Research Programme (LOCAR), *Hydrol. Earth Syst. Sci.*, 11, 108–124, <https://doi.org/10.5194/hess-11-108-2007>, 2007.
- Williams, A., Bloomfield, J., Griffiths, K., and Butler, A.: Characterising the vertical variations in hydraulic conductivity within the Chalk aquifer, *J. Hydrol.*, 330, 53–62, 2006.
- Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M., and Jakeman, A. J.: Performance of conceptual rainfall–runoff models in low-yielding ephemeral catchments, *Water Resour. Res.*, 33, 153–166, 1997.