

# Response Letter

Dear editor,

Dear reviewers,

5

Thank you very much for your valuable comments and suggestions. In the re-revised manuscript, we incorporated all the comments of referee #3. We respectfully disagree with the negative evaluation of referee #2, which mostly repeated the criticism from her/his last review. Nevertheless, some clarification was added to the manuscript following her/his criticism. Please see our point-by-point response below for more details.

10

We hope that the revised manuscript will be satisfactory to be published in Natural Hazards and Earth System Sciences.

Sincerely,

Simon Brenner on behalf of the co-authors.

15

## Associate Editor

Minor revisions are required to your article. Please, when preparing the revised edition, take into the due account all comments and suggestions from the reviewers. In case you disagree with some of them, please indicate the reasons why. I am  
5 looking forward to receiving your revised paper.

We thank the associate editor for his positive evaluation.

## Referee #3

10

*The manuscript from Brenner et al. is dealing with modeling of fractured and karstified aquifers in England. I found the topic well developed and of interest for an international audience. The contribution is well organized and contains interesting data and discussion. Some minor comments are detailed below:*

We thank the referee #3 for his positive comment on our manuscript.

15

*- Introduction: the literature review on this topic is wider than described in the text. About potential future changes in groundwater dynamics (lines 14-15 p2), several additional examples can be found in recent literature; please add references using recent papers*

20 We agree and added the following papers dealing with the investigation of groundwater dynamics (P2L13-16 in the revised manuscript):

Perrone, D. and Jasechko, S.: Dry groundwater wells in the western United States, Environ. Res. Lett., 12(10), 104002 [online] Available from: <http://stacks.iop.org/1748-9326/12/i=10/a=104002>, 2017.

25 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, , doi:10.1111/jfr3.12145, 2015.

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(October 2016), 161–176, doi:10.1002/hyp.10988, 2017.

30 von Freyberg, J., Moeck, C. and Schirmer, M.: Estimation of groundwater recharge and drought severity with varying model complexity, J. Hydrol., 527(Supplement C), 844–857, doi:<https://doi.org/10.1016/j.jhydrol.2015.05.025>, 2015.

- Results: what are the "hardly identifiable parameters"? See line 22 p7. Please list the parameters which are not well identifiable and discuss this limit in the discussion section

- Discussion: same topic, line 38 p8: "all model parameters are identifiable": It does not seem from the figure

- 5 The parameters are unidentifiable when we use only discharge or the groundwater time series. When using all information, the cumulative parameter distributions show identifiability throughout all parameters (blue lines). We added some clarification in both sections (P7L27-34 in the revised manuscript).

- 10 - line 26-27 p9: the sentence starting with "This is obvious" is not clear to me, please revise and explain better.

We apologize for the confusing sentence. We improved it.

- Conclusions: this chapter needs revision; at the moment it appears more like an abstract than a conclusion. Please identify the main findings and the main limits of your study, and list if possible as "take home messages", clearly and concisely.

- 15 We agree and added some sentences about highlighting the key aspects and limits of our findings (P11L13-16 in the revised manuscript).

## Referee #2

20

*This is a revised version of a manuscript detailing the simulation of groundwater levels in a Karst environment in the Southwest of England. I described in my review of the original manuscript how this topic is highly relevant to the journal and timely given the increasing need to simulate and forecast groundwater levels from limited datasets. The manuscript is suitably concise and the description is clear. I fully agree with the authors' statement in their Abstract that "specialised modelling approaches are required that balance model complexity and data availability". The authors assess whether they have achieved this balance by both exploring the identifiability of the parameters within their model (using a Shuffled Complex Evolution Method; SCEM) and by comparing model performance metrics for calibration and validation datasets (i.e. a split-sample test). They conclude that their modelling exercise had been a success because their analyses suggest that all of the parameters are identifiable and the differences between the calibration and validation metrics take values which they consider to be small.*

30

*Whilst I fully endorse the authors' general approach to assessing the performance of their model I have severe concerns about the exact way in which it has been implemented. I do not believe that the posterior distributions of the parameters yielded from the SCEM accurately reflect the uncertainty of these parameters. Furthermore, I do not believe that the comparison between*

performance metrics upon calibration and validation are particularly meaningful. For these reasons I do not recommend that the manuscript is accepted for publication.

We respectfully disagree with the opinion of referee #2. We already provided a thorough discussion and reply on her/his criticism concerning the reliability of the parameter posteriors, the choice of objective function and the model validation in the first round of review trying to incorporate the constructive elements of the review. For that reason, we will keep our response to this second review a bit shorter.

I first detail my concerns about the analyses of parameter identifiability. Looking at Figure 5, it is apparent that according to the SCEM that when the model is calibrated using only discharge data that the  $K_c$  parameter (for example) is almost perfectly identifiable. This posterior distribution indicates that this parameter definitely has a value less than 1. However, when all of the calibration data is used the parameter definitely has a value greater than 9. This is a clear contradiction and at least one of these two posterior distributions must be incorrect. Similar contradictions are evident for all of the other parameters except for those related to the groundwater level in a specific borehole.

We disagree. As we only consider marginal distributions, parameter interactions (see standard literature, e.g. Saltelli et al., 2008), which may cause an apparent sensitivity of a model parameter, are not visible. Such behaviour is a common feature of hydrological models and should also be familiar to referee #2. We added some elaboration to the discussion of the re-revised manuscript to make sure no other reader gets the wrong impression of a “clear contradiction” (P9L23-25 in the revised manuscript). Our own research, as well as many other colleagues, has shown how models calibrated or sampled using Monte Carlo simulations using different objectives or criteria ‘converge’ to different parts of the parameter space, including scenarios of non-overlapping posterior distributions (i.e. Freer et al., 1996; Freer et al., 2003; Freer et al., 2004; Seibert and McDonnell, 2002, etc.). This is a long standing discussion in hydrology that we have directly participated in a number of times. It remains an important topic ‘in general’ for environmental model evaluation and calibration, but that is not the focus of this paper to re-air all of that discussion again and we have ongoing tailored research that is addressing those issues. We feel we have done enough to discuss this one feature of our results in the context of this paper.

Furthermore, the theoretical justification for the authors’ choice to use the Kling-Gupta efficiency (KGE) as the objective function within the SCEM is rather weak. The formal theory of Markov Chain Monte Carlo methods such as the SCEM require that the objective function is a likelihood (i.e. the probability that the data is realised from the proposed model). The authors indicate that the KGE can be treated as an ‘informal’ likelihood function and refer to a paper by Smith, Bevan and Tawn. This paper does discuss informal likelihood functions and describes sufficient conditions for informal likelihood functions to satisfy the most fundamental axioms of a probability. As far as I can see, the paper does not explicitly mention the KGE. The starting point for satisfying the axioms of probability is that the informal likelihood function can be written as an  $L_p$ -norm. It is not immediately clear to me that the KGE can be written as an  $L_p$ -norm. Therefore, I am unclear of the relevance of the Smith et

*al. paper to the authors' study and I am not convinced that the KGE satisfies the fundamental axioms of a probability. I would have thought these axioms were a necessary requirement for a function to be treated as a likelihood.*

We disagree, we should note the paper title for Smith et al. (2008) is called '*Informal likelihood measures in model assessment: Theoretic development and investigation*'. It fundamentally discusses the issues of applying such measures for environmental problems that are not well constrained by idealised experimental designs that are better approximated by more classical statistical theory and so reason why such informal approaches have value for these circumstances. What the author is arguing is that they philosophically disagree with an entire branch of accepted literature on these issues in hydrology, some of which have shown the danger of making strong formal assumptions in the likelihood measure when the error residuals are misspecified due to epistemic uncertainties (such as Beven et al., 2008). We required from our analysis to allow identifying the information brought by the different types of observation data. As long as our performance metric is fit for purpose, in that it is a useful and valuable monotonically increasing measure of model performance that does not overly weight the very best simulation above all others due to our lack of evidence to fully quantify all the epistemic and aleatory uncertainties within our experimental design (i.e. the over constraining problem). The posteriors clearly indicate that when observations from the different wells are added to the discharge observations during the SCEM analysis, the well parameters (p\_Offs and P\_GW) switch from insensitive (close to a uniformly distributed posterior) to a sensitive (strongly differing from a uniform distribution) mode. This provides indication that our requirements are fulfilled, even though with some uncertainty in determining the shape of the distributions due to the choice of an informal likelihood measure, which we already acknowledged in subsection 5.1 as a result of the first review of referee #2. As noted previously we have, and do, work directly on these problems of criteria and objective choice and the impact on final model outputs.

20

*The authors do not provide any calibration diagnostics which might indicate that the SCEM has converged to a stable posterior distribution.*

The algorithm was applied in default mode as described in Vrugt et al., (2003), which also explains convergence is necessary. A short clarification was added to the methods section (P4, L38-39).

25

*The authors do not conduct any validation tests which might indicate that the posterior distributions reflect the uncertainty of the model parameters.*

We disagree. In fact, adding iteratively data to the SCEM analysis can be seen as a test of the reliability of the distributions: If the groundwater well parameters (p\_Offs and P\_GW) would show sensitivity when only using discharge observations in the SCEM analysis, the validation would be negative. But as they remain insensitive until groundwater observations are added to the analysis, we regard the validation to be positive. Please see our comment above, our response to the previous review of referee #2, and the elaborations in the discussion of the manuscript (subsection 5.1.).

30

I similarly have a number of concerns about the authors split sample tests. First, the authors conclude that decreases in model performance upon validation of 11 and 21% are sufficiently small to indicate robust model performance. They refer to other studies where a similar decrease in performance was observed. However, the decision that 21% is 'sufficiently small' is completely subjective. The expected decrease will be a complex function of the number of observations and the degree of seasonality, variability and autocorrelation realised by the data. Therefore, the comparison with other studies is irrelevant. This being said, if I were to compare these values with the results of modelling exercises I have previously undertaken then I would consider 21% to be a relatively large decrease.

As correctly pointed out by the referee, the acceptance or rejection of a parameters set by prediction is subject to the decision of the individual modeller. In our previous response we provided some studies with comparable changes of model performance in the validation period. By providing the % decrease of model performance we leave the decision of having confidence or doubt about the predictions to the reader. In order to account for this, we replaced the term "acceptable robustness of calibrated parameters" with "certain robustness of calibrated parameter" (P8, L39).

In their 'Responses to comments' document the authors describe how the KGE objective function "was chosen by trial and error comparing the simulation performances during calibration and validation obtained different objective functions (RMSE and other)". This 'trial and error' approach very much concerns me. The validation data should not be involved with the model calibration in any way – this includes the decisions about how the model is calibrated and what objective function is used. In my opinion, the use of the validation data in this manner invalidates the authors' split-sample tests. If the authors have infinite patience it is almost inevitable that they will eventually find a model calibration set up which yields results which they find pleasing. However, this setup is likely to be particular honed to the particular characteristics of the data they have used and is likely to perform less well as other data become available.

We disagree. Testing, which performance measure is most adequate to the modelling purpose, is a standard procedure of model development (Freer et al., 1996; Freer et al., 2003; Beven, 2003; Wagener, 2004).

25

## References

- Beven, K.J., (2003). Rainfall-Runoff Modelling: The Primer. John Wiley & Sons, Chichester.
- Beven, K.J., Smith, P.J. and Freer, J.E. (2008) So just why would a modeller choose to be incoherent? Journal of Hydrology 354(1-4), 15-32.
- 30 Freer, J., Beven, K. and Ambroise, B. (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. Water Resources Research 32(7), 2161-2173.
- Freer, J., Beven, K.J. and Peters, N.E. (2003) Calibration of watershed models. Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A.N. and Turcotte, R. (eds), pp. 69-88, AGU, Water Science and Application Series, Washington.

- Freer, J., McMillan, H., McDonnell, J.J. and Beven, K.J. (2004) Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures. *Journal of Hydrology* 291(3-4), 254-277.
- Seibert, J. and McDonnell, J.J. (2002) On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research* 38(11), art. no.-1241.
- 5 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.
- Vrugt, J.A., Gupta, H. V, Bouten, W., Sorooshian, S., (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* 39, 1201. doi:10.1029/2002WR001642
- 10 Wagener, T., Wheater, H., Gupta, H.V., (2004). *Rainfall–runoff modelling in gauged and ungauged catchments*, Book. doi:10.1142/p335

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

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

<sup>1</sup>Institute of Earth and Environmental Sciences, Freiburg University, 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 to: S. Brenner ([simon.brenner@hydrology.uni-freiburg.de](mailto:simon.brenner@hydrology.uni-freiburg.de))

## 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 to simulate groundwater level time series by a split sample test and parameter identifiability analysis. Secondly, we apply a split sample test on 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 chalk dominated catchment in Southwest England. The second split sample test also indicates that percentile approach is able to reliably predict groundwater level exceedances across all considered time scales up to their 75th percentile. However, when looking at the 90th percentile, it only provides acceptable predictions for the long time periods and it fails when the 95th percentile of groundwater exceedance levels is considered. 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-east 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. the chemical weathering (Ford and Williams, 2013), resulting in particular surface and subsurface features such as dolries, river sinks, caves and conduits (Goldscheider and Drew, 2007). 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, result in complex flow and storage dynamics (Hartmann et al., 2014a). Even though Chalk tends to less intense karstification, for instance compared to limestone, its karstic behaviour has increasingly been recognised (Fitzpatrick, 2011; Maurice et al., 2006, 2012).



1 Apart from the good water quality, favourable infiltration and storage dynamics which make chalk aquifers a preferred  
2 source of drinking water in the UK, their karstic behaviour also increases the risk of fast drainage of their storages by karstic  
3 conduit flow during dry years. This also increases the risk of groundwater flooding as a result of fast responses of  
4 groundwater levels to intense rainfalls due to fast infiltration and groundwater recharge processes. Groundwater flooding, i.e.  
5 when groundwater levels emerge at the ground surface due to intense rainfall (Macdonald et al., 2008), tend to be more  
6 severe in areas of permeable outcrop like the English Chalk (Macdonald et al., 2012). Groundwater drought indices tend to  
7 be more related to recharge conditions in Cretaceous Chalk aquifers than in granular aquifers (Bloomfield and Marchant,  
8 2013). Due to the fast transfer of water from the soil surface to the main groundwater system, chalk aquifers tend to be more  
9 sensitive to external changes, for instance shown by Jackson et al. (2015) who found significant groundwater level declines  
10 in 4 out of 7 chalk boreholes in a UK-wide study using historic groundwater level observations.

11 Climate projections suggest that the UK will experience increasing temperatures, with less rainfall during the summer but  
12 warmer and wetter winters (Jenkins et al., 2008). This may stress these groundwater resources, and increase the risk of  
13 groundwater droughts and potentially winter groundwater flooding. For those reasons, assessment of potential future changes  
14 in groundwater dynamics, concerning groundwater droughts, median groundwater levels as well as groundwater flooding is  
15 broadly recommended ~~(Jackson et al., 2015; Jimenez-Martinez et al., 2016).~~ and is subject of current research around the  
16 world ~~(von Freyberg et al., 2015; Jackson et al., 2015; Jimenez-Martinez et al., 2016; Moutahir et al., 2017; Naughton et al.,~~  
17 ~~2015; Perrone and Jasechko, 2017).~~

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

25 The definition of appropriate model structures and parameters from limited information becomes problematic when  
26 modelling karst aquifers. In order to achieve acceptable simulation performance they have to include representations of  
27 karstic heterogeneity in their structures. Distributed karst modelling approaches are able to simulate groundwater levels on a  
28 spatial grid but their data requirements mostly limit them to theoretical studies (e.g., Birk et al., 2006; Reimann et al., 2011)  
29 or well explored study sites (e.g., Hill et al., 2010; Jackson et al., 2011; Oehlmann et al., 2014). Lumped karst modelling  
30 approaches consider physical processes at the scale of the entire karst system. Although they are strongly simplified, they  
31 can include karst peculiarities such as different conduit and matrix systems (Fleury et al., 2009; Geyer et al., 2008;  
32 Maloszewski et al., 2002). Since they are easy to implement and do not require spatial information, they are widely used in  
33 karst modelling (Jukić and Denić-Jukić, 2009). Simple rainfall-runoff models with more than 5-6 parameters are often  
34 regarded to end up in equifinality (Jakeman and Hornberger, 1993; Wheater et al., 1986; Ye et al., 1997), i.e. their  
35 parameters lose their identifiability (Beven, 2006; Wagener et al., 2002). For that reason, recent research took advantage of  
36 auxiliary data, such as water quality data or tracer experiments (Hartmann et al., 2013b; Oehlmann et al., 2015a). These  
37 studies showed that adding such information allows identifying the necessary model parameters, therefore enabling the  
38 model to reflect the relevant processes.

39 Up to now, most lumped karst models have been applied for rainfall-runoff simulations. Groundwater levels were simulated  
40 in quite a few studies (Adams et al., 2010; Jimenez-Martinez et al., 2016; Ladouche et al., 2014), however mostly relying on  
41 very simple representation of karst hydrological processes and disregarding the scale discrepancy between borehole (point  
42 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. This 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 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 Southwest England that had to cope with several flooding events in the past. Finally we apply the approach on 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> (Figure 1). The catchment elevation varies from over 200 m above sea level (a.s.l.) in the north-west to sea level in 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 typically is very low, hardly reaching 5 m<sup>3</sup>/s (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 (Brunner et al., 2010; NRA, 1995). 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).

### Figure 1: Overview on the Frome catchment

## 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. Similar 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 applied to a set of  $N=15$  model compartments (Figure 2). Pareto functions as distribution functions have shown to perform best in previous work (Hartmann et al., 2013a, 2013b), 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).

**Figure 2: The VarKarst model structure**

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 subsection 3.3). Similar 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.]:

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

The related parameters are  $h_{gw}$  [m] and  $p_{gw}$  [-].  $h_{gw}$  is the difference of 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.  $p_{gw}$  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 was obtained from the Centre for Ecology & Hydrology (CEH, <http://nrfa.ceh.ac.uk/>) and dates back to the 1960s. The borehole data was 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 a 15-minute interval. For further analysis, the data was aggregated to daily time averages. The potential evapotranspiration has a strong annual cycle. Since most recent data from years 2009-2012 was missing, representative PET-years were calculated on the basis of the last fifty 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. (2010).

### 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 (Hastings, 1970; Metropolis et al., 1953) 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 or equal than 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 objective function, which can be regarded as an informal likelihood measure, [or more generally 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 other). We found that we obtain the most robust results with the KGE. To decide whether to accept or discard a parameters set, we compare the KGEs of the “candidate” and the given parameter sets. Such procedure was already applied in various studies (Blasone et al., 2008; Engeland et al., 2005; McMillan and Clark, 2009) and is possible if the error functions are monotonically increasing with improved performance. We achieved this in the SCEM algorithm by defining KGE as:

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

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

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 standard deviations of simulations and observations, respectively.

The posterior parameter distributions derived from SCEM provide information about the identifiability of the parameters. 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, 2013b, 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 on the period 2008-2012 and used the period 2003-2007 for validation. We chose this reversed order to be able including the information of 3 boreholes that was only available for 2008-2012. Three years were used as warm-up for calibration and validation, respectively. During calibration, the most appropriate of the  $N=15$  groundwater compartments to 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 provides 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 one sixth each. In order to explore to contribution of the different observed discharge and groundwater time series during the calibration, we use SCEM 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.

### 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 ground water wells. Its model structure allowed for an acceptable and stable simulation of groundwater level time series of the three wells (see subsection 4.1) but for groundwater management, frequency distributions of groundwater levels, calculated over the time scale of interest, are commonly preferred. For that reason we introduced a groundwater level percentile based approach. Other than Westerberg et al. (2016) that 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). Similar to the calculation of standardised precipitation or groundwater indices (e.g., Bloomfield and Marchant, 2013; Lloyd-Hughes and Saunders, 2002), 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 (Figure 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.

**Figure 3: schematic description of the percentile approach**

As the approach is meant to be applied in combination with climate change scenarios, we perform an evaluation on multiple time scales and flow percentiles. We assess the 5<sup>th</sup>, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup> and 95<sup>th</sup> 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):

$$MAD_p = \text{mean} \left( \text{abs}(\sum SE_{i,x} - \sum OE_{i,x}) \right) \quad [d] \quad (2)$$

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

$$PAD_p = \frac{MAD_p}{dp_x} \cdot 100 \quad [\%] \quad (3)$$

where  $dp_x$  is a normalizing constant standing for total the number of days of the respective time scale and percentile. For example, if we take the time scale *months* and the 75<sup>th</sup> percentile of exceedances we got 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 on the validation period similar to the split sample test in subsection 3.3. That 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 then 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 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 (°C) and precipitation (%) as well as the respective 25<sup>th</sup> and 75<sup>th</sup> percentile from the probabilistic projections and apply them on 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 x 3 projections (3x precipitation and 3x evapotranspiration) for each of the emission scenarios that also address the uncertainty associated with the projections. The resulting simulations will provide 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 period, respectively. The borehole simulations show high KGE values and only slight deteriorations in the validation period. The parameters are located well within their pre-defined ranges. Mean soil storage  $V_{\text{mean},S}$  and mean epikarst storage  $V_{\text{mean},E}$  are 2015.6 mm and 1011.7 mm, respectively. The porosity parameter at Ashton Farm is the highest, followed 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, for Ridgeway it chose the compartment number 8.

**Figure 4**Figure-4 plots the observations against simulations for the calibration and validation period. 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 well identifiable when we use all available data (**Figure 5**~~Figure-5~~), while some parameters remain hardly identifiable when only parts of the available data were used for calibration. Here identifiability of parameters is simply the extent that the cumulative parameter distributions span the sampled parameter limits, where highly constrained or near optimal is classed as identifiable. For instance, when only discharge was used for calibration (green lines), the parameters related to groundwater (porosity  $p_{GW}$  and ~~base-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  $V_{\text{mean},E}$  is not identifiable when only the groundwater well data is 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.



**Figure 4: Modelled discharge [m<sup>3</sup>/s] of the Frome at East Stoke and groundwater levels [m a.s.l.] at the boreholes Ashton Farm, Ridgeway and Black House**

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

## 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 to exceed our selected groundwater level percentiles. This is elaborated for 90<sup>th</sup> percentile of simulated and observed groundwater levels of Ashton farm ([Figure 6](#)).

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

[Table 3](#) shows the mean observed and modelled exceedances of all selected thresholds (the 5<sup>th</sup>, 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup> percentiles) at all temporal resolutions in the validation period. By comparing matches in the number days of exceedance we evaluate our model at different percentiles and time scales. 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 is the deviation between observed and modelled exceedances. The same is true for the PAD when moving from lower to higher temporal resolutions. The MAD gets lower with higher temporal resolution.

**Table 4: 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 [%]**

## 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 [Figure 7](#). They display the mean model outputs (Q<sub>sim</sub>, AET) and mean exceedances per year, calculated on the basis of our modelled time series. Both emission scenarios (low & high) lead to an 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.

**Figure 7: Mean model input (mm/a), mean modelled output (mm/a) and mean (non-)exceeded percentiles (number/a) in the reference period and both scenarios (borehole: Ashton Farm; future period: 2070-2099). The circles indicate the spread among the 9 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)**

## 5 Discussion

### 5.1 Reliability of the simulations

A decrease of [simulation](#) performance in the validation period ~~has-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\_{GW,B}\$](#) ) to 21% (discharge prediction,  [\$KGE\_Q\$](#) ) during the validation period indicates ~~acceptable-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 4 available observed discharge and ground water level time series. Applying the Shuffled Complex Evolution Metropolis algorithm and step wise increasing the calibration data (only discharge, only groundwater, all together), we show that discharge 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 ~~(Kuczera and Mroczkowski, 1998)~~. 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 the reality. However,  $V_{mean,S}$  and  $V_{mean,E}$  are quite high considering that initial ranges for these parameters were 0-250/0-500 mm (Hartmann et al., 2013a, 2013c). 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 a high variability of 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 interfluvies.

Additionally, the mean epikarst storage coefficient  $K_{mean,E}$  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  $a_{fsep}$  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 but sensitive of  \$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 of Jones and Cooper (1998) as well as Reeves (1979) who reported 30 % and 10-20 % of the recharge occurring through (macro-) fissures in Chalk catchments, respectively. Although groundwater flow in the chalk is dominated by the matrix, given antecedent wet conditions, fracture flow can increase significantly (Butler et al., 2012b; Ireson and Butler, 2011; Lee et al., 2006). Overall, split-sample test, parameter identifiability analysis, realistic values of parameters and plausible simulation results provide strong indication for a reliable model functioning.

### 5.2 Performance of the percentile approach

Based on the idea of the standardised precipitation or groundwater indices (Bloomfield and Marchant, 2013; Lloyd-Hughes and Saunders, 2002) our percentile approach permits to improve the performance of the model to reflect observed



groundwater level exceedances. It yields acceptable performance for years to days up to the 90<sup>th</sup> percentile. A reduction of precision with the time scale 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 period, 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 ~~is obvious~~can be seen in the simulated time series at Ashton Farm and Black House ~~indicate~~, which ~~also indicate~~ an over-estimation of high levels and under-estimation 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; Butler et al., 2009; Wheatley et al., 2007). This is mainly attributed to the decrease of fractures because of the increasing overburden and absence of water level fluctuations (Butler et al., 2012a; Williams et al., 2006). 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. Such decrease of performance was also found for standardised indices that use probability distributions instead of a simulation model (Van Lanen et al., 2016; Núñez et al., 2014; Vicente-Serrano et al., 2012). To improve 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 90<sup>th</sup> percentile will allow for a better simulation quality. This could be further evaluated by using different percentile weighting schemes, 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 of 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 less 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 of 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. (2015). 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 9 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 is 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 is 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 as 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 parameterization of the unsaturated zone is a major difficulty in the Chalk. Since this study struggles also 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 Southwest 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 showing the general adequacy of the model to simulate 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 90<sup>th</sup> 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 to simulate and predict groundwater frequencies in Chalk regions where information is too limited for the application of distributed models. Here, a thorough model evaluation is essential to obtain 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, emphasizes on the >90<sup>th</sup> percentiles of groundwater levels could significantly improve our simulations. In addition we propose to apply the method on 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.

## Acknowledgements

This publication contains Environment Agency information © Environment Agency and database right. Thanks to Dr Jens Lange and Dr Sophie Bachmair, University of Freiburg, for their valuable advice. Support for GC, JF and NH 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.

## 7 Appendix

Within the VarKarst model, the parameter  $V_{mean,S}$  [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. (3,4) in [Table 5Table-5](#). We apply the same distribution coefficient  $a_{SE}$  when we derive the epikarst storage distribution by the mean epikarst depth  $V_{mean,E}$  [mm] (Eqs. (6,7) in [Table 5Table-5](#)). We determine actual evapotranspiration from each soil compartment  $E_{act,i}$  is calculated by reducing potential evapotranspiration, which is found by the Thornthwaite equation (Thornthwaite, 1948), by the soil saturation deficit (Eq. (1) in [Table 5Table-5](#)). Surface runoff is found by the excess of soil and epikarst storage of the previous model compartment (Eq. (2) in [Table 5Table-5](#)). With surface runoff and actual evapotranspiration know, the stored water volume at each soil compartment  $V_{Soil,i}$  [mm] can be calculated by simply applying water balance.

The recharge from the soil to the epikarst  $R_{Epi,i}$  [mm] is calculated by the excess of the soil storage (Eq. (5) in [Table 5Table-5](#)), while the epikarst outflow follows a linear storage assumption (Eqs. (8,9) in [Table 5Table-5](#)). Again, water balance allows determining the stored water  $V_{Epi,i}$  [mm] at each time step  $t$  and each epikarst compartment  $i$ . The downward flux from the epikarst considers a diffuse ( $R_{diff,i}$  [mm]) and concentrated groundwater recharge ( $R_{conc,i}$  [mm]) component that are found by a variable separation factor  $f_{C,i}$  [-] and a distribution coefficient  $a_f$  [-] (Eqs. (10,11,12) in [Table 5Table-5](#)). 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.

Similar to the epikarst compartment, variable groundwater storage coefficients  $K_{GW,i}$  [d] are calculated (Eq. (15) in [Table 5Table-5](#)) and applied to calculate the discharges of the matrix system (Eq. (13) in [Table 5Table-5](#)) and the conduit system (Eq. (14) in [Table 5Table-5](#)), which together sum up to the entire discharge of the system (Eq. (15) in [Table 5Table-5](#)). Knowing groundwater recharge and groundwater discharge for each model compartment  $i$  again allows determining the stored volume of water within the groundwater compartment  $V_{GW,i}$  at time step  $t$ , which is used to simulate the groundwater levels (Eq. (1) in subsection 3.1).

**Table 5: Model routines, variables and equations solved in the VarKarst model**

## 9 References

- Adams, B., Peach, D. W. D. and Bloomfield, J. P. J.: The LOCAR hydrogeological infrastructure for the Frome/Piddle catchment, *Br. Geol. Surv.*, 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(2), 185–193, doi: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, edited by D. J. Allen, J. P. Bloomfield, and V. K. Robinson, 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(3), 472–485, 2006.
- Bakalowicz, M.: Karst groundwater: a challenge for new resources, *Hydrogeol. J.*, 13, 148–160, 2005.
- Bennett, C.: South Winterbourne Flood Investigation, , (July), 2013.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrol.*, 320(1–2), 18–36, 2006.
- Birk, S., Liedl, R. and Sauter, M.: Karst Spring Responses Examined by Process-Based Modeling, *Groundwater*, 44(6), 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(4), 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(12), 4769–4787, doi: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(10), 4327–4344, doi:10.5194/hess-19-4327-2015, 2015.
- Brunner, P., Dennis, I. and Girvan, J.: River Frome Geomorphological Assessment and Rehabilitation Plan, , (October), 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(1–2), 131–143, doi: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 ), , 1849–1858, doi:10.1007/s10040-009-0477-4, 2009.
- Butler, A. P., Hughes, A. G., Jackson, C. R., Ireson, A. M., Parker, S. J., Wheater, H. S. and Peach, D. W.: Advances in modelling groundwater behaviour in Chalk catchments, *Geol. Soc. London, Spec. Publ.*, 364(1), 113–127, 2012a.
- Butler, a. P., Hughes, a. G., Jackson, C. R., Ireson, a. M., Parker, S. J., Wheater, H. S. and Peach, D. W.: Advances in modelling groundwater behaviour in Chalk catchments, *Geol. Soc. London, Spec. Publ.*, 364(1), 113–127, doi:10.1144/SP364.9, 2012b.
- 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(2), 111–119, doi:10.1111/j.1753-318X.2009.01025.x, 2009.
- Dorset County Council: A Local Climate Impacts Profile for Dorset, 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(4), 1015–1031, doi:10.1029/91WR02985, 1992.
- Edmonds, C. N.: Towards the prediction of subsidence risk upon the Chalk outcrop, *Q. J. Eng. Geol. Hydrogeol.*, 16(4), 261–266, doi: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

1 methodology/Estimation bayésienne des incertitudes au sein d'une modélisation conceptuelle de bilan hydrologique, Hydrol.  
2 Sci. J., 50(1), 2005.

3 Environment Agency: Frome and Piddle Catchment Flood Management Plan, 2012.

4 Fitzpatrick, C. M.: The hydrogeology of bromate contamination in the Hertfordshire Chalk: double-porosity effects on  
5 catchment-scale evolution, University College London., 2011.

6 Fleury, P., Ladouche, B., Conroux, Y., Jourde, H. and Dörfliger, N.: Modelling the hydrologic functions of a karst aquifer  
7 under active water management--the Lez spring, J. Hydrol., 365(3), 235–243, 2009.

8 Ford, D. and Williams, P. D.: Karst hydrogeology and geomorphology, John Wiley & Sons. 578 pages., 2007.

9 Ford, D. C. and Williams, P. W.: Karst Hydrogeology and Geomorphology, John Wiley & Sons., 2013.

10 von Freyberg, J., Moeck, C. and Schirmer, M.: Estimation of groundwater recharge and drought severity with varying model  
11 complexity, J. Hydrol., 527(Supplement C), 844–857, doi:https://doi.org/10.1016/j.jhydrol.2015.05.025, 2015.

12 Geyer, T., Birk, S., Liedl, R. and Sauter, M.: Quantification of temporal distribution of recharge in karst systems from  
13 spring hydrographs, J. Hydrol., 348, 452–463, 2008.

14 Goldscheider, N. and Drew, D.: Methods in Karst Hydrogeology, edited by I. A. of Hydrogeologists, Taylor & Francis  
15 Group, Leiden, NL., 2007.

16 Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance  
17 criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1), 80–91, 2009.

18 Hartmann, A., Lange, J., Weiler, M., Arbel, Y. and Greenbaum, N.: A new approach to model the spatial and temporal  
19 variability of recharge to karst aquifers, Hydrol. Earth Syst. Sci., 16(7), 2219–2231, 2012.

20 Hartmann, A., Weiler, M., Wagener, T., Lange, J., Kralik, M., Humer, F., Mizyed, N., Rimmer, A., Barberá, J. A., Andreo,  
21 B., Butscher, C. and Huggenberger, P.: Process-based karst modelling to relate hydrodynamic and hydrochemical  
22 characteristics to system properties, Hydrol. Earth Syst. Sci., 17(8), 3305–3321, doi:10.5194/hess-17-3305-2013, 2013a.

23 Hartmann, A., Barberá, J. A., Lange, J., Andreo, B. and Weiler, M.: Progress in the hydrologic simulation of time variant  
24 recharge areas of karst systems – Exemplified at a karst spring in Southern Spain, Adv. Water Resour., 54, 149–160,  
25 doi:10.1016/j.advwatres.2013.01.010, 2013b.

26 Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H. and Weiler, M.: Testing the realism of model structures to  
27 identify karst system processes using water quality and quantity signatures, Water Resour. Res., 49(6), 3345–3358, 2013c.

28 Hartmann, A., Goldscheider, N., Wagener, T., Lange, J. and Weiler, M.: Karst water resources in a changing world: Review  
29 of hydrological modeling approaches, Rev. Geophys., 52(3), 218–242, doi:10.1002/2013rg000443, 2014a.

30 Hartmann, A., Mudarra, M., Andreo, B., Marín, A., Wagener, T. and Lange, J.: Modeling spatiotemporal impacts of  
31 hydroclimatic extremes on groundwater recharge at a Mediterranean karst aquifer, Water Resour. Res., 50(8), 6507–6521,  
32 2014b.

33 Hartmann, A., Kobler, J., Kralik, M., Dirnböck, T., Humer, F. and Weiler, M.: Model-aided quantification of dissolved  
34 carbon and nitrogen release after windthrow disturbance in an Austrian karst system, Biogeosciences, 13(1), 159–174,  
35 doi:10.5194/bg-13-159-2016, 2016.

36 Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57(1), 97–109,  
37 1970.

38 Hill, M. E., Stewart, M. T. and Martin, A.: Evaluation of the MODFLOW-2005 conduit flow process, Ground Water, 48(4),  
39 549–559, doi:10.1111/j.1745-6584.2009.00673.x, 2010.

40 Howden, N. J. K.: Hydrogeological controls on surface/groundwater interactions in a lowland permeable chalk catchment:  
41 implications for water quality and numerical modelling, Imperial College London (University of London)., 2006.

42 Ireson, A. M. and Butler, A. P.: Controls on preferential recharge to Chalk aquifers, J. Hydrol., 398(1), 109–123, 2011.

1 Jackson, C. R., Meister, R. and Prudhomme, C.: Modelling the effects of climate change and its uncertainty on UK Chalk  
2 groundwater resources from an ensemble of global climate model projections, *J. Hydrol.*, 399(1–2), 12–28,  
3 doi:10.1016/j.jhydrol.2010.12.028, 2011.

4 Jackson, C. R., Bloomfield, J. P. and Mackay, J. D.: Evidence for changes in historic and future groundwater levels in the  
5 UK, *Prog. Phys. Geogr.*, 39(1), 49–67, 2015.

6 Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*,  
7 29(8), 2637–2649, 1993.

8 Jenkins, G. J., Perry, M. C. and Prior, M. J.: The climate of the United Kingdom and recent trends, Met Office Hadley  
9 Centre, Exeter, UK. [online] Available from: [http://www.ukcip.org.uk/wp-content/PDFs/UKCP09\\_Trends.pdf](http://www.ukcip.org.uk/wp-content/PDFs/UKCP09_Trends.pdf), 2008.

10 Jimenez-Martinez, J., Smith, M. and Pope, D.: Prediction of groundwater induced flooding in a chalk aquifer for future  
11 climate change scenarios, *Hydrol. Process.*, 30, 573–587, doi:10.1002/hyp.10619, 2016.

12 Jones, H. K. and Cooper, J. D.: Water transport through the unsaturated zone of the Middle Chalk: a case study from Fleam  
13 Dyke lysimeter, *Geol. Soc. London, Spec. Publ.*, 130(1), 117–128, doi:10.1144/GSL.SP.1998.130.01.11, 1998.

14 Jukić, D. and Denić-Jukić, V.: Groundwater balance estimation in karst by using a conceptual rainfall–runoff model, *J.*  
15 *Hydrol.*, 373(3), 302–315, 2009.

16 Klemeš, V.: Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, 31(1), 13–24,  
17 doi:10.1080/02626668609491024, 1986.

18 Kuczera, G. and Mroczkowski, M.: Assessment of hydrologic parameter uncertainty and the worth of multiresponse data,  
19 *Water Resour. Res.*, 34(6), 1481–1489, 1998.

20 Kumar, R., Musuza, J. L., Loon, A. F. Van, Teuling, A. J., Barthel, R., Broek, J. Ten, Mai, J., Samaniego, L. and Attinger,  
21 S.: Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator, , 1117–1131,  
22 doi:10.5194/hess-20-1117-2016, 2016.

23 Ladouche, B., Marechal, J. C. and Dorfliger, N.: Semi-distributed lumped model of a karst system under active management,  
24 *J. Hydrol.*, 509, 215–230, doi:10.1016/j.jhydrol.2013.11.017, 2014.

25 Van Lanen, H., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K.,  
26 Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W.  
27 K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva,  
28 M., Osuch, M., Sauquet, E., Stagge, J. H. and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European  
29 case, *Hydrol. Process.*, n/a-n/a, doi:10.1002/hyp.10838, 2016.

30 Lee, L. J. E., Lawrence, D. S. L. and Price, M.: Analysis of water-level response to rainfall and implications for recharge  
31 pathways in the Chalk aquifer, SE England, *J. Hydrol.*, 330(3), 604–620, 2006.

32 Lloyd-Hughes, B. and Saunders, M. A.: A drought climatology for Europe, *Int. J. Climatol.*, 22(13), 1571–1592,  
33 doi:10.1002/joc.846, 2002.

34 Lloyd, J. W.: The Hydrogeology of the Chalk in North-West Europe, edited by R. A. Downing, M. Price, and G. P. Jones,  
35 pp. 220–249, Clarendon Press, Oxford., 1993.

36 Macdonald, D., Dixon, A., Newell, A. and Hallaways, A.: Groundwater flooding within an urbanised flood plain, *J. Flood*  
37 *Risk Manag.*, 5(1), 68–80, 2012.

38 Macdonald, D. D. M. J., Bloomfield, J. P. P., Hughes, a. G. G., MacDonald, a. M. M., Adams, B. and McKenzie, a. a.:  
39 Improving the understanding of the risk from groundwater flooding in the UK, FLOODrisk 2008, Eur. Conf. Flood Risk  
40 Manag. Oxford, UK, 30 Sept - 2 Oct 2008. Netherlands, ~(1), 10 [online] Available from: <http://nora.nerc.ac.uk/7760/>, 2008.

41 Maloszewski, P., Stichler, W., Zuber, A. and Rank, D.: Identifying the flow systems in a karstic-fissured-porous aquifer, the  
42 Schneealpe, Austria, by modelling of environmental 18 O and 3 H isotopes, *J. Hydrol.*, 256(1), 48–59, doi:10.1016/S0022-

1 1694(01)00526-1, 2002.

2 Maurice, L. D., Atkinson, T. C., Barker, J. A., Bloomfield, J. P., Farrant, A. R. and Williams, A. T.: Karstic behaviour of  
3 groundwater in the English Chalk, *J. Hydrol.*, 330(1), 63–70, 2006.

4 Maurice, L. D., Atkinson, T. C., Barker, J. A., Williams, A. T. and Gallagher, A. J.: The nature and distribution of flowing  
5 features in a weakly karstified porous limestone aquifer, *J. Hydrol.*, 438–439, 3–15, doi:10.1016/j.jhydrol.2011.11.050,  
6 2012.

7 McMillan, H. and Clark, M.: Rainfall-runoff model calibration using informal likelihood measures within a Markov chain  
8 Monte Carlo sampling scheme, *Water Resour. Res.*, 45(4), 2009.

9 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of state calculations by fast  
10 computing machines, *J. Chem. Phys.*, 21(6), 1087–1092, 1953.

11 Moutahir, H., Bellot, P., Monjo, R., Bellot, J., Garcia, M. and Touhami, I.: Likely effects of climate change on groundwater  
12 availability in a Mediterranean region of Southeastern Spain, , 176(October 2016), 161–176, doi:10.1002/hyp.10988, 2017.

13 Murphy, J., Sexton, D., Jenkins, G., Boorman, P., Booth, B., Brown, K., Clark, R., Collins, M., Harris, G., Kendon, L.,  
14 Office, M., Centre, H., Betts, A. R., Brown, S., Hinton, T., Howard, T., McDonald, R., McCarthy, M., Stephens, A.,  
15 Atmospheric, B., Centre, D., Wallace, C., Centre, N. O., Warren, R., Anglia, E. and Wilby, R.: UK Climate Projections  
16 science report : Climate change projections, , (December), 2010.

17 Naughton, O., Johnston, P. M., McCormack, T. and Gill, L. W.: Groundwater flood risk mapping and management :  
18 examples from a lowland karst catchment in Ireland, , doi:10.1111/jfr3.12145, 2015.

19 NRA: The Frome & Piddle Management Plan - Consultation Report, 1995.

20 Núñez, J., Rivera, D., Oyarzún, R. and Arumi, J. L.: On the use of Standardized Drought Indices under decadal climate  
21 variability: Critical assessment and drought policy implications, *J. Hydrol.*, 517, 458–470,  
22 doi:10.1016/j.jhydrol.2014.05.038, 2014.

23 Oehlmann, S., Geyer, T., Licha, T. and Sauter, M.: Reducing the ambiguity of karst aquifer models by pattern matching of  
24 flow and transport on catchment scale, *Hydrol. Earth Syst. Sci.*, 19(2), 893–912, doi:10.5194/hess-19-893-2015, 2015a.

25 Oehlmann, S., Geyer, T., Licha, T. and Sauter, M.: Reduction of the ambiguity of karst aquifer modeling through pattern  
26 matching of groundwater flow and transport, *Hydrol. Earth Syst. Sci.*, 16, 11593, doi:10.5194/hess-19-893-2015, 2015b.

27 Parajka, J., Merz, R. and Blöschl, G.: Uncertainty and multiple objective calibration in regional water balance modelling:  
28 case study in 320 Austrian catchments, *Hydrol. Process.*, 21(4), 435–446, doi:10.1002/hyp.6253, 2007.

29 Perrin, C., Michel, C. and Andréassian, V.: Does a large number of parameters enhance model performance? Comparative  
30 assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, 242(3), 275–301, 2001.

31 Perrin, C., Michel, C. and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*,  
32 279, 275–289, 2003.

33 Perrone, D. and Jasechko, S.: Dry groundwater wells in the western United States, *Environ. Res. Lett.*, 12(10), 104002  
34 [online] Available from: <http://stacks.iop.org/1748-9326/12/i=10/a=104002>, 2017.

35 Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten,  
36 D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y. and Wisser, D.:  
37 Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment., *Proc.*  
38 *Natl. Acad. Sci. U. S. A.*, 111(9), 3262–7, doi:10.1073/pnas.1222473110, 2014.

39 Reeves, M. J.: Recharge and pollution of the English Chalk: some possible mechanisms, *Eng. Geol.*, 14(4), 231–240, 1979.

40 Reimann, T., Geyer, T., Shoemaker, W. B., Liedl, R. and Sauter, M.: Effects of dynamically variable saturation and matrix-  
41 conduit coupling of flow in karst aquifers, *Water Resour. Res.*, 47(11), 1–19, doi:10.1029/2011WR010446, 2011.

42 Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S.: Global sensitivity

1 analysis: the primer, John Wiley & Sons., 2008.

2 Schoups, G. and Vrugt, J. A.: A formal likelihood function for parameter and predictive inference of hydrologic models with  
3 correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46(10), 2010.

4 Smith, P., Beven, K. J. and Tawn, J. A.: Informal likelihood measures in model assessment: Theoretic development and  
5 investigation, *Adv. Water Resour.*, 31(8), 1087–1100, 2008.

6 Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, *Geogr. Rev.*, 38(1), 55, doi:10.2307/210739,  
7 1948.

8 Upton, K. A. and Jackson, C. R.: Simulation of the spatio-temporal extent of groundwater flooding using statistical methods  
9 of hydrograph classification and lumped parameter models, *Hydrol. Process.*, 25(12), 1949–1963, 2011.

10 Vicente-Serrano, S. M., López-Moreno, J. I., Beguería, S., Lorenzo-Lacruz, J., Azorin-Molina, C. and Morán-Tejeda, E.:  
11 Accurate Computation of a Streamflow Drought Index, *J. Hydrol. Eng.*, 17(2), 318–332, doi:10.1061/(ASCE)HE.1943-  
12 5584.0000433, 2012.

13 Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for  
14 optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39(8), 18, 2003.

15 Wagener, T., Lees, M. J. and Wheater, H. S.: A toolkit for the development and application of parsimonious hydrological  
16 models, *Math. Model. large watershed Hydrol.*, 1, 87–136, 2002.

17 Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A. and Freer, J.: Uncertainty in  
18 hydrological signatures for gauged and ungauged catchments, *Water Resour. Res.*, 52, 1847–1865,  
19 doi:10.1002/2015WR017635, 2016.

20 Wheater, H. S., Bishop, K. H. and Beck, M. B.: The identification of conceptual hydrological models for surface water  
21 acidification, *Hydrol. Process.*, 1(1), 89–109, 1986.

22 Wheater, H. S., Wheater, H. S., Peach, D. and Binley, A.: Characterising groundwater-dominated lowland catchments : the  
23 UK Lowland Catchment Research Programme ( LOCAR ) Characterising groundwater-dominated lowland catchments : the  
24 UK Lowland Catchment Research Programme ( LOCAR ), , 11(1), 108–124, 2007.

25 Williams, A., Bloomfield, J., Griffiths, K. and Butler, A.: Characterising the vertical variations in hydraulic conductivity  
26 within the Chalk aquifer, *J. Hydrol.*, 330(1), 53–62, 2006.

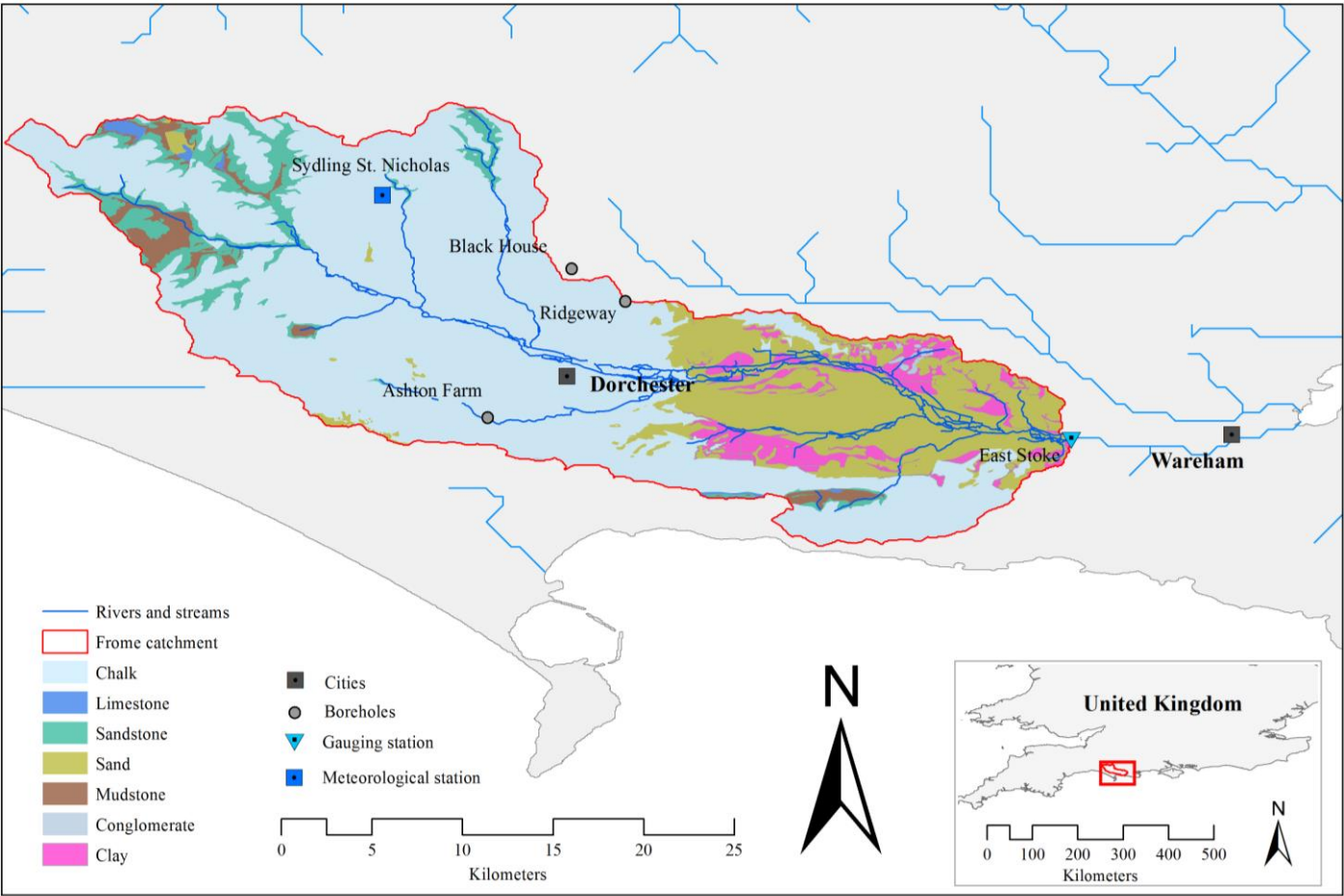
27 Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M. and Jakeman, A. J.: Performance of conceptual rainfall-runoff models in  
28 low-yielding ephemeral catchments, *Water Resour. Res.*, 33(1), 153–166, 1997.

29

30



2

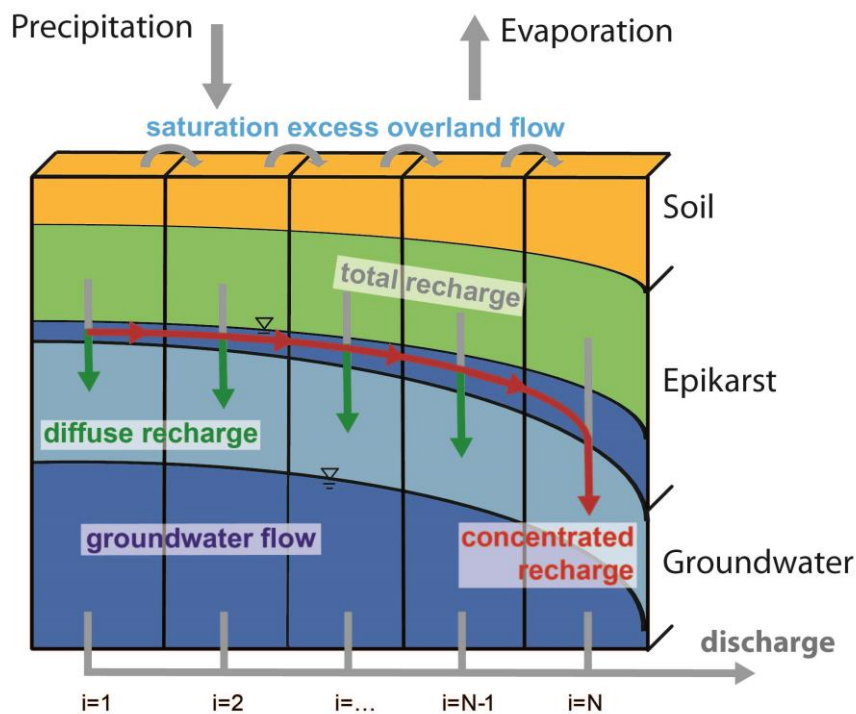


3

4 **Figure 1: Overview on the Frome catchment**

5

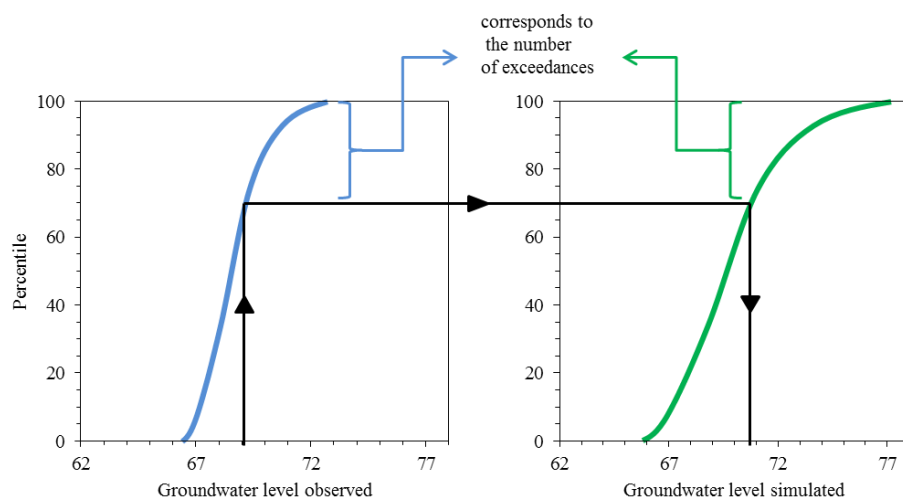
1



2

3 **Figure 2: The VarKarst model structure**

4

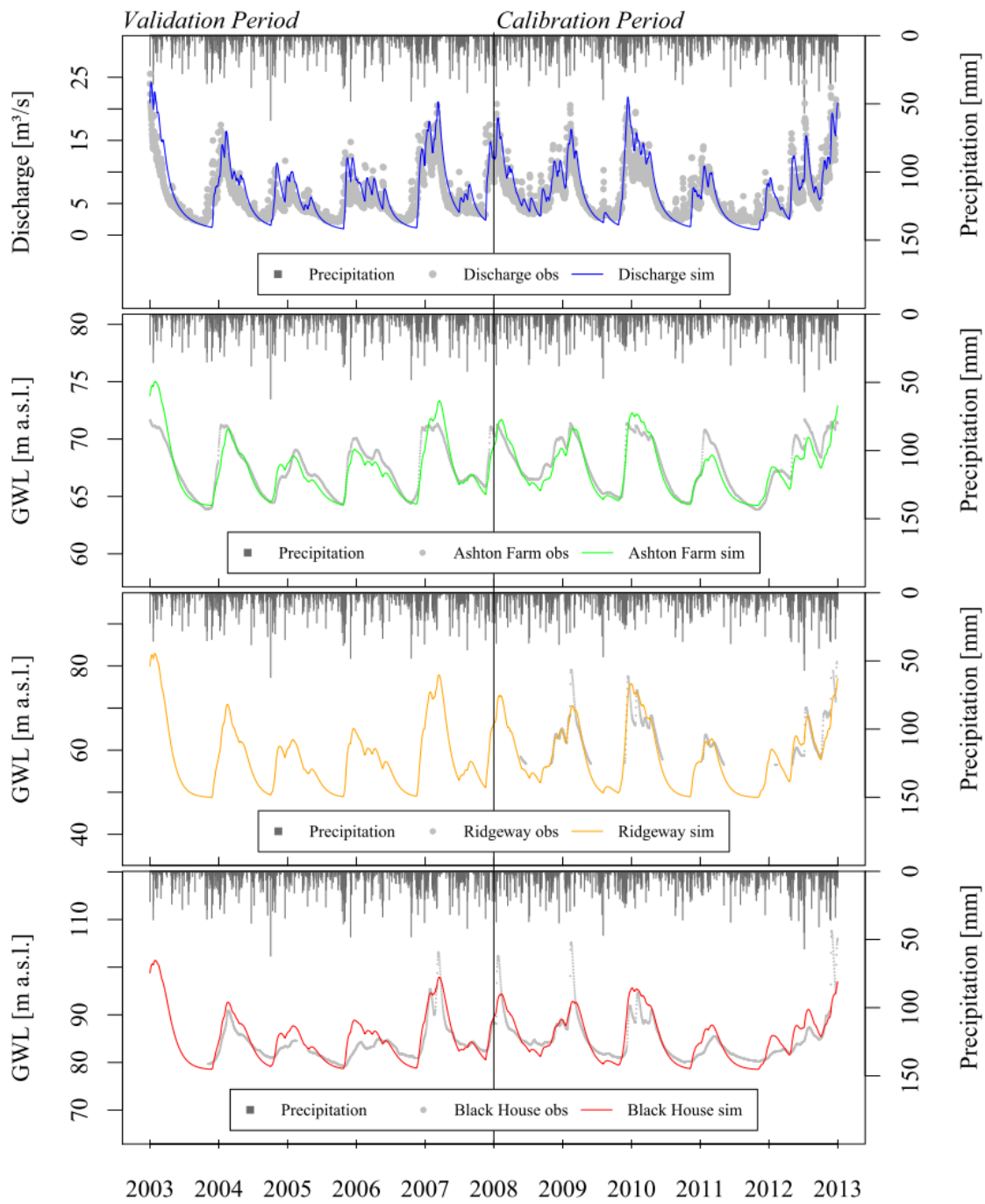


5

6 **Figure 3: Schematic description of the percentile approach**

7

1

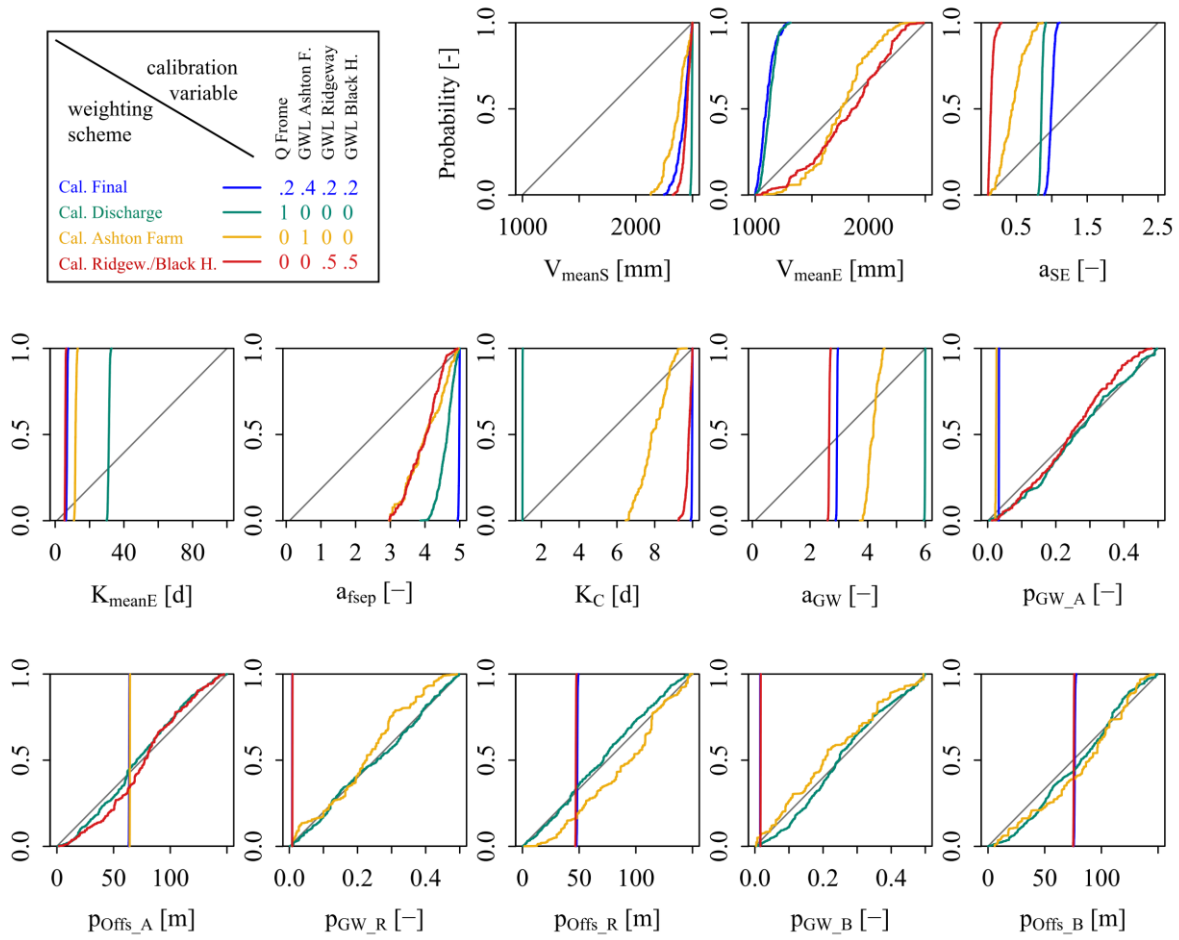


2

3 **Figure 4: Modelled discharge [ $\text{m}^3/\text{s}$ ], and groundwater levels [m a.s.l.] at the boreholes Ashton Farm, Ridgeway and Black House**

4

1



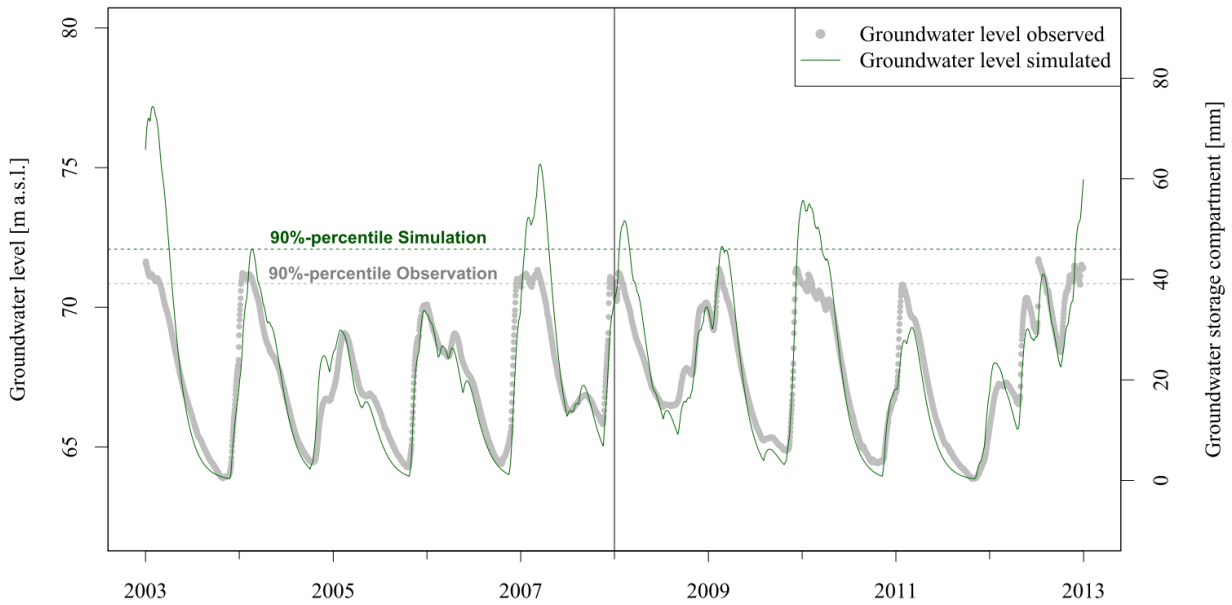
2

3

4

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

5



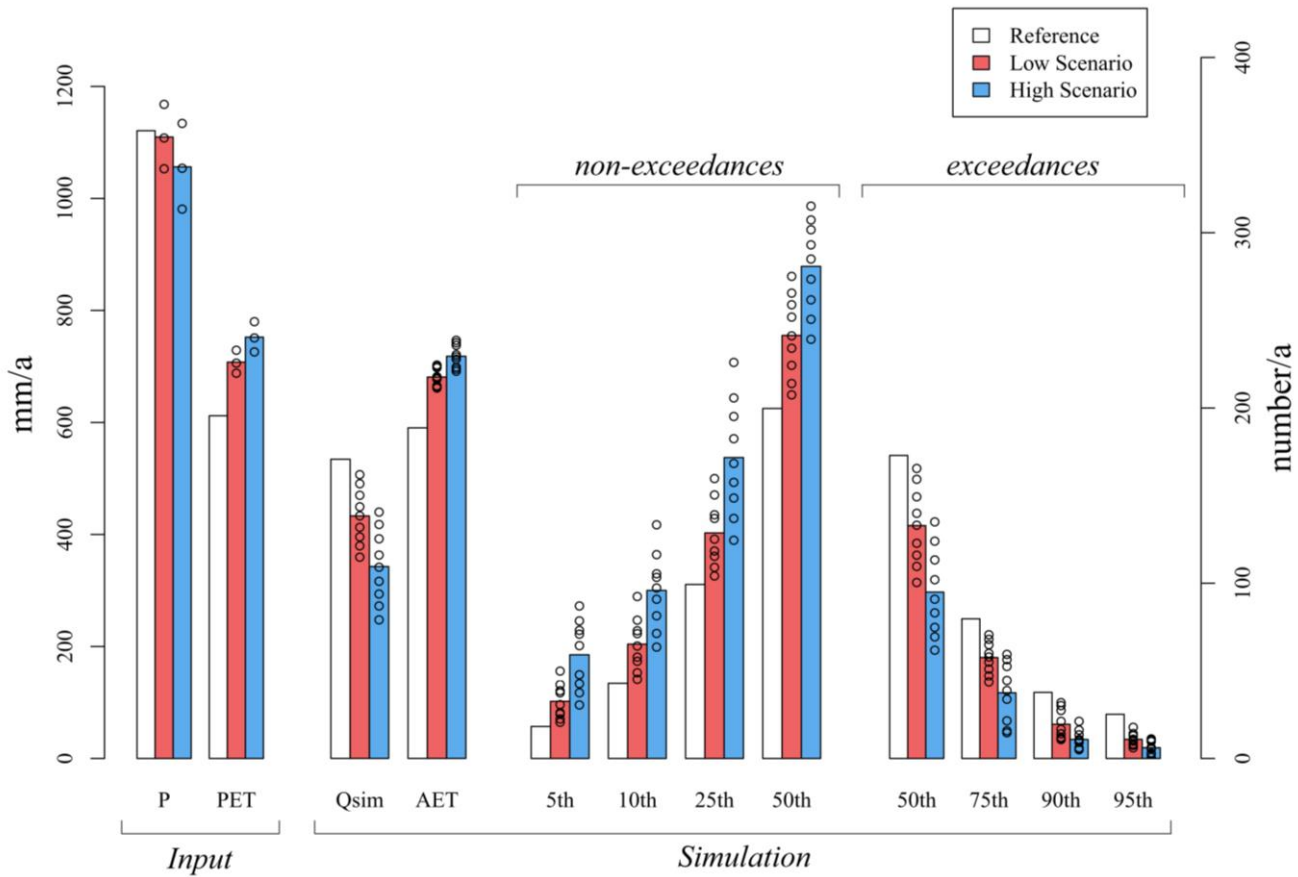
6

7

8

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

9



2

3 **Figure 7: Mean model input (mm/a), mean modelled output (mm/a) and mean (non-)exceeded percentiles (number/a) in the**  
4 **reference period and both scenarios (borehole: Ashton Farm; future period: 2070-2099). The circles indicate the spread among the**  
5 **9 realisations for each of the two scenarios**

6

7

## 11 Tables

2

3 **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>
Pot. 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, %

1

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

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

\*Calibration/validation.

3

**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
year-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

5

**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	AET	5th	10th	25th	50th	75th	90th	95th
	mm/a	mm/a	non exc/a	non exc/a	non exc/a	exc/a	exc/a	exc/a	exc/a
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



**Table 5: Parameters, descriptions and equations solved in the VarKarst model**

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