Response Letter

Dear editor,
Dear reviewers,
Many thanks for your valuable comments and suggestions. In the revised manuscript, we incorporated the suggested changes according to our response in the open discussion. Please find below a point by point response on all your comments followed by the revised manuscript with tracked changes.

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.

Associate Editor

Based upon the reviews, the article needs major revision. Authors are kindly invited to follow the indications by the reviewers when preparing the revised manuscript, or, in case they disagree with their comments, to explain in detail the reasons why.

We thank the associate editor and the two reviewers for their detailed evaluation.

Referee Andrew Long

This manuscript describes the application of an existing hydrologic model for karst aquifers. The approach of evaluating the model calibration in terms of hydrologic exceedance rates appears to be a new and useful approach. Exceedance rates of projected hydrologic simulations are evaluated for human safety or the needs of species (e.g., https://pubs.er.usgs.gov/publication/sir20145089). Therefore, if a model is to be used for this purpose, it makes good sense to evaluate the model directly on the basis of exceedance frequencies.

We thank Dr Andy Long for his positive evaluation and his valuable and helpful comments.

Main comments:

1. *The abstract explains that the approach to simulate groundwater level frequency is novel. I would say that this is not the novel part, because the time-series records were simulated and simply converted into frequency distributions, which is a common way to summarize hydrologic time-series records. However, the novel part is that the model calibration is evaluated on the basis of frequency distributions, which I have not seen before, and I suggest presenting it that way. The tile is more accurate: "A percentile approach to evaluate simulated groundwater levels and frequencies. . ."*

We changed the title to "Process-based modelling to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England" and improved the abstract accordingly.

2. *Section 5.2 discusses the possibility of focusing the calibration on high percentiles. I don't totally agree that longer time-series records would be needed to do this, and this section could benefit from further discussion of this idea. For example, an approach could be developed to evaluate the usefulness and data adequacy of such an endeavor. You could vary the weights within the observed time-series record for individual observations at different exceedances to tailor the calibration to a target percentile. It would be possible to calibrate to a different weighting scheme for each percentile. Further, an uncertainty analysis could be applied on each separate calibration run, and quantifying the presumed decrease in uncertainty as the percentile increases could be useful. Then, when you make predictions for different percentiles, you can also report the differences in uncertainty. This idea also applies to section 5.3, which discusses the prediction of increased drought.*

We added a statement at the end of Section 5.2:

"This could be further evaluated by using different percentile weighting schemes, stepwise increasing the weight on the target percentile."

3. *The Introduction discusses risks to events such as groundwater flooding and drought. I suggest adding a short statement to this effect in the Abstract to emphasize the need for this study in terms of natural hazards.*

We added the sentence: "Due to their properties, they are particularly vulnerable to groundwater related hazards like floods and droughts."

Other comments:

1. *p. 2, lines 24-26: indicates that karst groundwater levels were simulated by lumped models only in a few instances, but see also Long and Derickson (1999), Long and Mahler (2013), and Pinault et al. (2001).*

We added Long and Mahler (2013) and (Derickson and Long (1999) to the literature comparison. We could not find any simulation of groundwater levels in the work of Pinault et al. (2001).

2. *p. 3, lines 22-23: describes a new approach to show groundwater levels as frequency distributions. Showing hydrologic time-series data as frequency distributions is a common method. Please explain how this is new, or describe it differently.*
3. *p. 3, line 35: "PET" should be defined.*

We agree with both comments and improved the text accordingly.

4. p. 4, line 34: discusses a "weighting scheme." I think the calibration weights are applied to observations, but that should be explained here for clarity.

We added "(…), as we stepwise added borehole data to our discharge observations." and further added more detailed information about the model calibration in section 3.3 emphasizing the importance of auxiliary data.

5. figure 7: what is the meaning of "manipulated" in the caption?

This refers to the manipulation of our observed "baseline" data, see chapter 3.5. We clarified the caption of figure 7.

6. table 5: I think these result apply to a particular model time step (e.g., daily), but I'm not sure. Please clarify.

These are the mean model outputs (Qsim, AET) and exceedances per year in the simulation period 2070-2099 which was calculated on the basis of our model which runs at a daily time step. We added:

" They display the mean model outputs (Qsim, AET) and mean exceedances per year, calculated on the basis of our modelled time series. "

in section 4.3

References

Derickson, R.., Long, a. ., 1999. Linear systems analysis in a karst aquifer. J. Hydrol. 219, 206–217. doi:10.1016/S0022-1694(99)00058-X

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

Long, A.J., Mahler, B.J., 2013. Prediction, time variance, and classification of hydraulic response to recharge in two karst aquifers. Hydrol. Earth Syst. Sci. 17, 281–294. doi:10.5194/hess-17-281-2013

Pinault, J.-L., Pauwels, H., Cann, C., 2001. Inverse modeling of the hydrological and the hydrochemical behavior of hydrosystems: Application to nitrate transport and denitrification. Water Resour. Res. 37, 2179–2190.

Anonymous Referee

A percentile approach to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England. The authors use the VarKarst model to predict the variation of discharge and groundwater levels in a catchment in England. The topic is relevant to the journal and the work is timely given a growing interest in the forecasting and characterisation of floods and droughts. It would be very valuable to have a discharge/groundwater level model that gives reliable predictions even when the calibration datasets are small. The paper is suitably concise and the description is generally clear. However, I have a number of serious concerns about the focus of the manuscript and the calculations within it. I am unable to recommend the manuscript for publication unless these concerns are addressed.

We thank the referee for her /his evaluation and detailed comments. We appreciate the concerns about our claims about the novelty of the approach. We hope that we can clarify all of the referees concerns in the following response.

1. *In the title and introduction, the authors promote their 'percentile approach' to assessing the performance of the models as the main novelty in the manuscript. I am afraid that I am not persuaded that the percentile approach is novel enough to merit publication in itself. The approach is a comparison between the realised percentiles of the observed and modelled discharge/groundwater levels. It appears to be exactly equivalent to the standard statistical procedure of comparing the distributions of two variables in terms of their realized quantiles. This is a very well used approach, as evidenced by the Wikipedia page describing the QQ plots that result: https://en.wikipedia.org/wiki/Q%E2%80%93Q_plot*

The referee is right. The novelty of our research is the application of a process-based model instead of a statistical distribution function. We admittedly created a wrong perception by the choice of our title. The revised manuscript will be titled with
"Process-based modelling to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England "
In addition, we provided more reference to the work of others that applied quantile-quantile approaches in groundwater frequency analysis in the introduction.

2. *Furthermore, I am not convinced that the percentiles used by the authors are a good indicator of the performance of a discharge/groundwater level model. The authors are only confirming that the complete set of modelled values are similar to the complete set of observed values. They are not confirming that the groundwater levels are predicted at the correct time. In terms of the authors' percentile criterion, there would be no penalty for a model that predicts a flood at the time of a drought but compensates by predicting a drought at the time of a flood. For these reasons, I believe that a substantial change of theme of the manuscript is required.*

We believe this is a misunderstanding: In this study, we used the percentile approach only for our evaluation. The calibration and evaluation of our model was carried out with continuous flow and water level observations. The error function KGE that we used for comparing model simulations and observations explicitly evaluates the correctness of timing by using the linear correlation coefficient *r* as one of its three components (see also our response to general comment 5).
In the new version of the manuscript we clarified the elaboration of the model calibration and evaluation (see Section 3.3) to avoid further misunderstanding.

3. *The theme that most interests me in the manuscript is the quest to "balance model complexity and data availability" referred to in the Abstract. If the authors could demonstrate that they have achieved this for their study area then they would have a very valuable paper. However, I believe that much more evidence of this is required. The authors calibrate the 13 parameters of the VarKarst model using data from three boreholes and one timeseries of discharge data. In any such modelling exercise I am concerned whether the parameters maintain their physical meaning and whether the internal processes in the model (e.g. the soil and epikarst modules) are reflecting reality. It is entirely possible that the model is acting as a 'black box' where the large number of parameters are giving it the flexibility to reproduce almost any relationship between the input and output*

data with which it is presented. If this were the case, it is unlikely that the model would perform well if the characteristics of the input data were to change (e.g. under climate change).

We understand the concern of the referee. Indeed, models with more than 5-6 parameters are often regarded to end up in equifinality (Jakeman and Hornberger, 1993; Wheater et al., 1986; Ye et al., 1997), i.e. their parameters lose their identifiability (Beven, 2006; Wagener et al., 2002). In such case, the model can be regarded as a “black box” with rather limited prediction skills as correctly stated by the referee.

In order to reflect the complexity of karst hydrology, 5-6 parameters are often not enough to include all relevant processes in a simulation model. For that reason, recent research took advantage of auxiliary data, such as water quality data or tracer experiments (Hartmann et al., 2013; Oehlmann et al., 2015). These studies could show that such information allowed for identifying the necessary model parameters therefore enabling the model to reflect the relevant processes.

In this study, we followed this idea and used a combination of groundwater level observations at three locations and discharge observations to obtain enough information to estimate our model parameters. Applying the Shuffled Complex Evolution Metropolis algorithm (also see our response to general comment 5 below) and step-wise increasing the calibration data (only discharge, only groundwater, all together), we show that discharge alone, as well as groundwater alone, do not provide enough information to identify all of our model parameters (Fig 5 in the manuscript) as the posteriors of some of the model parameters remain close to a uniform distribution.

Using all information, observed discharge and observations of three groundwater levels, all model parameters are identifiable, i.e. their posteriors strongly differ from a uniform distribution (blue lines in Fig 5), which is in accordance with preceding research that showed that a combination of groundwater and discharge observations can parameter uncertainty (Kuczera and Mroczkowski, 1998). Furthermore, the split-sample test indicates a stable performance of groundwater simulations (Table 3, also see our response to general comment 6). We therefore believe that there is enough indication that the model reproduces the system behaviour satisfactorily and that it can be used for prediction.

We added these clarifications to the methodology section (3.3) as well as in the discussion (5.1)..

4. *One piece of evidence of the model reflecting reality rather than acting as a black box would be clearly identifiable parameter values. The authors are therefore quite correct to explore the identifabilty of the parameters using the MCMC approach. Their results (Figure 5) indicate that for their final calibration that the parameters are almost perfectly identifiable. Given the short duration, high seasonality and marked temporal correlation amongst the input data I find this surprising. Indeed when (Schoups and Vrugt, 2010) calibrated their similarly complex river models using an MCMC approach many of the parameter values could not be identified. This makes me question the authors’ implementation of the MCMC approach.*

Thanks for this critical comment. We thoroughly studied the work of Schoups and Vrugt (2010) in relation to our results. Using a hydrological model with seven parameters combined with an error model with 4-5 parameters their calibration problem is indeed similar to the one we present in our study. But there is one important difference: they only use discharge observations for model calibration. As found by many preceding studies (see also our response to general comment 2) simulation models with more than 5-6 parameters typically result in increased parameter uncertainty, which Schoups and Vrugt (2010) also found in their study. Using only discharge information, our study would have resulted in similar problems (green lines in Fig 5). However, the combined use of discharge observations and the observations of three groundwater wells resulted in increased parameter identifiability, as we could also show in Fig 5 (blue lines). Therefore, our results do not contradict Schoups and Vrugt (2010) but they rather show that there are ways to reduce parameter uncertainty by auxiliary data (see also our response to general comment 2).

In the revised manuscript, we put more emphasis on the description of these multiple data sets for parameters estimation (Section 3.3) and refer to the comparison with Schoups and Vrugt (2010) in the discussion (Section 5.1).

5. *Within a MCMC algorithm, a huge number of different sets of parameter values are compared. Those sets that are consistent with the observed data are included in the Markov chain whereas other parameter sets are discarded. These comparisons are normally made by calculating the likelihood function for the different parameter sets (e.g. Schoups & Vrugt, 2010). It is possible to use the calculated likelihoods or probabilities to determine which parameters sets are good enough to be included in the Markov chain. Thus, the inclusion or exclusion of a parameter set is decided by an objective criterion that is consistent with statistical theory.*

   It appears that the authors have compared different parameter sets in terms of their KGE score. This concerns me because it is not clear to me how to decide what magnitude of difference between KGE scores signifies that one set of parameters is not good enough to be included. A threshold on the KGE scores could be set arbitrarily but then the realised distributions of the parameters become meaningless. The apparent identifiability of the parameters could be changed by a simple and arbitrary tweak of this threshold.

   Therefore, the authors must give more detail about the comparison function they included in the MCMC algorithm and demonstrate how it leads to objective estimates of the posterior distributions of the parameters.

The Shuffled Complex Evolution Metropolis algorithm (SCEM, Vrugt et al., 2003) that we used in our study is based on the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) and the Shuffled Complex Evolution algorithm (Duan et al., 1992). The Metropolis-Hastings algorithm uses a formal likelihood measure, i.e. an objective criterion that is consistent with statistical theory, 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. 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 (Smith et al., 2008). To decide whether to accept or discard a parameter 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 monotonically increasing with improved performance. We achieved this in the SCEM algorithm by defining $\mathrm{KGE}_{\mathrm{SCEM}}$ as

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

$$\alpha = \frac{\sigma_S}{\sigma_O}; \beta = \frac{\mu_S}{\mu_O}$$

With r as the linear correlation coefficient between simulations and observations, and σ_S, σ_O and μ_S, μ_O as the means and standard deviations of simulations and observations, respectively.

As stated correctly by the referee, the shape of the posteriors is dependent on the error function and using another likelihood measure, formal or informal, may have resulted in different shapes of the posteriors. However, applying SCEM with KGE in our stepwise procedure we are mostly interested in the relative differences of the posteriors and we can clearly see how some of posteriors translate from a uniform distribution to a well-defined peak when more information is added (see also our response to general comment 4). These results combined with the acceptable multi-objective performance of the model during calibration and validation (see also our response to general comment 6), and the realistic parameters that we finally found (see discussion of parameter values in subsection 5.3) makes us confident that the model reproduces the relevant features of our studied system.

We added these elaborations in the methods section and discuss the consequences of using an informal likelihood measure in the revised discussion.

6. *I'd also like clarification about how the authors decided that their validation results were sufficiently good to conclude that "the model provides robust simulations of discharge and groundwater levels". The authors state that the difference between the calibration and validation KGE scores are small. For each data source, the validation results are worse than the calibration results. Might this indicate that the model is too complex? How big a difference between validation and calibration results would have been required for the authors to conclude that the model had been ineffective?*

Split-sample tests are a common and necessary tool to evaluate the prediction performance of a simulation model (Klemeš, 1986). If the model is compared to a validation period, i.e. a time series of observations that was not used for parameters estimation, a decrease of performance has to be expected because there is always a tendency to compensate for model structural limitations and observational uncertainties during the calibration. If a model contains too many degrees of freedom (model parameters), there is a risk that calibration may overcome all these limitations and uncertainties although the model is a poor choice for the studied system. A split sample-test would indicate such failure by a strong decrease of performance during the validation period.

As correctly questioned by the referee the threshold, from which a decrease of performance is not acceptable anymore, is subject to the individual case of application and the opinion of the modeller. In our case we obtained a decrease of performance from -11% (groundwater prediction) to -21% (discharge prediction). Such ranges are comparable with split-sample tests found by other studies (-4% to -14% by Parajka et al., 2007; -5% to -24% by Perrin et al., 2001). The lower decrease in performance that we found for the simulation of groundwater levels also indicates more stable prediction performance for the groundwater simulations that we later use for our example application with the simplified climate scenarios.

We added these aspects to the discussion (Section 5.1).

7. *There is a great deal of seasonality in the groundwater levels. Can we be sure that the model is going beyond these seasonal trends? Could a simple annual periodic function have given similarly good results and better managed the trade-off between model complexity and data availability?*

Yes, a simple annual periodic function may be able to reproduce observed variability of groundwater levels to some degree. However, such function would not be more than a black box model and it would not be straight forward using it to assess the impact of climatic changes on groundwater levels. The structure of the VarKarst model takes into account the particulates of karst hydrology (see also our response to specific comment 1). We believe that our analysis and evaluation provides some indication that it is also able to reflect the observed processes at our Chalk study site, therefore making it a useful tool to explore the impact of climate changes on groundwater level dynamics.

We added:

" 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."

to the introduction section.

Specific comments:

1. *The introduction provides a clear description of the hydrogeological system with the appropriate level of description and ample references for anyone who wants to delve further (the same can also be said of section 2). More detail could be provided in the paragraph which describes the importance of the work in this study. I appreciate that the authors have made the Methodology section concise by referring to previous papers. However, I think they could give a clearer overview of the VarKarst model whilst leaving the details to the other papers. What do the 15 model compartments correspond to? Are they situated along some sort of gradient in the catchment? If so, is it possible to use knowledge of the hydrogeological system to*

determine the compartment in which each borehole is situated? What do they mean when they say that the spatial variability of the soil, epikarst and groundwater systems are expressed as a Pareto function? - What characteristics of these systems are the authors referring to? - Are these characteristics sampled from a Pareto distribution or do they decay according to a Pareto function?

We thank the referee for these helpful suggestions. We added statements to the abstract and the introduction to highlight the scope and importance of this work (see also general comment 7 above as well as the comment 3 of the other referee).
In addition, we added a more detailed description of the VarKarst model and the meaning of its individual components (see Appendix).

2. *Equation (1). Ensure that all symbols in all equations are defined. Use a multiplication sign rather than '*'.*

We improved our manuscript and eliminated the stylistic flaws.

3. *Section 3.3 – Give more detail about the implementation of the MCMC algorithm to address my concerns above. In particular, explicitly state the function used to decide whether a parameter set is accepted or rejected and explain how these lead to objective and representative samples of the posterior distributions.*

Please see our response to general comment 4.

4. *Section 3.4 The authors state that their percentile approach was motivated by standardised groundwater and precipitation indices. Seasonality is often removed from standardised indices. Did the authors consider removing seasonality from their simulations before assessing them?*

During the period of model development and calibration, we considered calibrating the flow percentiles directly, i.e. removing seasonality. However, removing the temporal information from the time series would have reduced the information content of the data and would have resulted in increased parameter uncertainty (see our response to general comment 2 and 3) with a lower prediction performance of the model.
We added this information at the end of the section 3.4 .

5. *Equation (2) Write words such as mean in standard font rather than italics. State which variable you are summing over. The authors calculated the 5th percentile at a yearly time scale using 10 years of data. Does this mean they attempted to determine the 5th percentile from only 10 observations of yearly data?*

We improved the elaborations on Eq. 2.
The percentiles were derived from the daily data of the calibration period (2008-2012). We then compared the average sum of days exceeding the respective percentile in the respective time scale. We added a sentence to the end of section 3.4 to clarify the origin of the percentiles.

6. *Section 3.5: I am not sure that using nine climate scenarios is sufficient to assess the uncertainty in the effect of climate change on groundwater levels.*

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.
We added this elaboration to section 5.3

7. *Section 4.2. The poor performance of the model when groundwater levels are large could be because the authors are using an objective function that is suited to Normally distributed variables but the distribution of groundwater levels are skewed. Have the authors tried an objective function that is more suited to skewed data?*

The objective function (KGE) is applied to simulated time series of groundwater levels. It was chosen by trial and error comparing the simulation performances during calibration and validation obtained different objective functions (RMSE and other). We found that we obtain the most robust results with the KGE.

We mentioned this trial and error procedure in the section 3.3.

References

Beven, K. J.: A manifesto for the equifinality thesis, J. Hydrol., 320(1–2), 18–36, 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, doi:10.1016/j.advwatres.2007.12.003, 2008.

Duan, Q. Y., Sorooshian, S. and Gupta, H. V: Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, Water Resour. Res., 28(4), 1015–1031, 1992.

Engeland, K., Xu, C.-Y. and Gottschalk, L.: Assessing uncertainties in a conceptual water balance model using Bayesian methodology / Estimation bayÃ©sienne des incertitudes au sein dâ€TMune modÃ©lisation conceptuelle de bilan hydrologique, Hydrol. Sci. J., 50(1), 45–63 [online] Available from: http://www.informaworld.com/10.1623/hysj.50.1.45.56334, 2005.

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(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

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, doi:10.1016/j.advwatres.2013.01.010, 2013.

Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57(1), 97–109 [online] Available from: http://www.jstor.org/stable/2334940, 1970.

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

Klemeš, V.: Dilettantism in Hydrology: Transition or Destiny, Water Resour. Res., 22(9), 177S–188S, 1986.

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

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(4), 1–12, doi: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(6), 1087, doi:10.1063/1.1699114, 1953.

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(2), 893–912, doi: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(4), 435–446, doi:10.1002/hyp.6253, 2007.

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., 241, 275–301, 2001.

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(10), 1–17, doi: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(8), 1087–1100, doi:10.1016/j.advwatres.2008.04.012, 2008.

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(8), 18, 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.

Wheater, H. S., Bishop, K. H. and Beck, M. B.: The identification of conceptual hydrological models for surface water acidification, Hydrol. Process., 1(1), 89–109, doi:10.1002/hyp.3360010109, 1986.

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(1), 153–166, doi:10.1029/96wr02840, 1997.

Process-based modelling to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England ~~A percentile approach to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England~~

Simon Brenner[1], Gemma Coxon[2,4], Nicholas J. K. Howden[3,4], J. Freer[2,4] and Andreas Hartmann[1,3]

[1]Institute of Earth and Environmental Sciences, Freiburg University, Germany
[2]School of Geographical Sciences, University of Bristol, Bristol, UK
[3]Department of Civil Engineering, University of Bristol, Bristol, UK
[4]Cabot Institute, University of Bristol, Bristol, UK.

Correspondence to: S. Brenner (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 ~~of this resource~~. 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 frequenc~~y~~ies~~ distributions with~~ 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 ~~newly developed~~ percentile approach we can ~~simulate~~ assess the number of days exceeding or falling below selected groundwater level percentiles. Firstly, we evaluate the performance of the model to simulate ~~three~~ groundwater level time series by a spilt 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 ~~3~~ 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~~est available~~ time ~~scale~~ 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.

Kommentar [SB1]: AL 3

Kommentar [SB2]: AL 1

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 dollies, 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).

Apart from the good 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). 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, 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 (Jackson et al., 2015; Jimenez-Martinez et al., 2016). 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 (Beven, 2006; Perrin et al., 2003).

Kommentar [AH3]: ANON 7

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 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., 2014). Lumped karst modelling approaches consider physical processes at the scale of the entire karst system. Although they are strongly simplified, they can include karst peculiarities such as different conduit and matrix systems (Fleury et al., 2009; Geyer et al., 2008; Maloszewski et al., 2002). Since they are easy to implement and ~~don't~~ do not require spatial information, they are widely used in karst modelling (Jukić and Denić-Jukić, 2009). Simple rainfall-runoff models with more than 5-6 parameters are often regarded to end up in equifinality (Jakeman and Hornberger, 1993; Wheater et al., 1986; Ye et al., 1997), i.e. their parameters lose their identifiability (Beven, 2006; Wagener et al., 2002). For that reason, recent research took advantage of auxiliary data, such as water quality data or tracer experiments (Hartmann et al., 2013a; Oehlmann et al., 2015). These studies showed that adding such information allows identifying the necessary model parameters, therefore enabling the model to reflect the relevant processes.

Kommentar [SB4]: ANON 3

Up to now, most lumped karst models have been applied for rainfall-runoff simulations. Groundwater levels were ~~only~~ simulated in quite a few~~some~~ studies (Adams et al., 2010; Jimenez-Martinez et al., 2016; Ladouche et al., 2014), however mostly relying on very simple representation 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 ~~simulate and~~ 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 km² (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 m³/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 interfluve 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., 2013c, 2014b, 2016). We use a simple linear relationship that takes into account effective porosities and base level of the groundwater wells (see Eq. 1) ~~to~~ enabling~~e~~ the model to simulate groundwater levels based on the groundwater storage in VarKarst. Finally, a newly developed ~~percentile~~ evaluation approach is used ~~to~~ by transferring simulated groundwater level time series into groundwater level frequency distributions ~~to~~ and ~~compare~~ 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, 2013c), 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 ~~equations and~~ model parameters are provided in Table 2~~, respectively~~. For a detailed description of VarKarst see the appendix or Hartmann et al. (2013b).

Kommentar [AH5]: ANON sc1

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.]:

~~$h_{GW}(t)=\frac{V_{GW,i}(t)}{1000*p_{GW}}+\Delta h$~~

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

Formatiert: Tiefergestellt durch 9 Pt.

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.

~~**Table 2: Model routines, variables and equations solved in the VarKarst model**~~

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

~~The Kling-Gupta Efficiency (KGE) is used as a performance measure to calibrate against the discharge and the three boreholes. The KGE is a result of a decomposition of the NSE (and MSE), emphasizing the importance of the different components of the~~

~~criterion (Gupta et al., 2009).~~ We use the Shuffled Complex Evolution Method (SCEM) for our calibration,~~. This method explores the parameter space using a Monte Carlo Markov Chain and searches for posterior distributions of the model parameters (Vrugt et al., 2003), including the regions with optimum performance.~~ 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. 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 (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:

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

$$\alpha = \frac{\sigma_s}{\sigma_0}\ ;\ \beta = \frac{\mu_s}{\mu_0}$$

With r as the linear correlation coefficient between simulations and observations, and σ_s σ_o and μ_s μ_o as the means and standard deviations of simulations and observations, respectively.

~~In addition, t~~The posterior parameter distributions derived from SCEM provide information about the identifiability of the parameters. The more they differ from a uniform ~~final~~ 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, 2013c, 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).

Kommentar [SB6]: ANON sc7

Kommentar [SB7]: ANON 5

Kommentar [SB8]: ANON 3+4

Kommentar [SB9]: AL oc4

Posterior parameter distributions are plotted as cumulative distributions. ~~Deviations of the posterior distribution (diagonal) indicate a sensitive parameter.~~ 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 a performance indicator~~ as described in the previous subsection. We avoided a calibration directly ~~During the period of model development, we considered calibrating directly~~ to the flow percentiles, as ~~seasonality~~temporal information would have ~~is often~~been removed ~~from standardised indices. However, removing the temporal information from the time series~~, which would ~~have reduced the information content of the data and would~~ have resulted in a lower prediction performance of the model.

Kommentar [SB10]: ANON sc4

Kommentar [AH11]: ANON 2

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 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 **m**ean **a**bsolute **d**eviation (MAD) between simulated exceedances (SE) and observed exceedances (OE):

~~$MAD_p = mean\left(abs\left(\sum obs_{i,x} - \sum sim_{i,x}\right)\right)_p$ [d]~~

$$\mathrm{MAD_p} = \mathrm{mean}\ \left(\mathrm{abs}\left(\sum \mathrm{SE}_{\mathrm{i,x}} - \sum \mathrm{OE}_{\mathrm{i,x}}\right)\right) \qquad [\mathrm{d}] \qquad (2)$$

Kommentar [SB12]: ANON sc5

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 **p**ercentage of mean **a**bsolute **d**eviation (PAD) with the total number of days of the chosen time scale:

~~$$PAD_p = \frac{MAD_p}{dp_x} * 100 \qquad [\%]$$~~

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

Formatiert: Tiefergestellt durch 9 Pt.

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 *75th percentile* of exceedances we got a dp_x of (100-75) % x (365.25 / 12) days. To evaluate the prediction performance of the approach, percentiles are ~~calculated based~~ derived from the daily data of ~~on~~ 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.

Kommentar [SB13]: ANON sc5

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 25th and 75th 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 $V_{mean,S}$ and mean epikarst storage $V_{mean,E}$ 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 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), while some parameters remain hardly identifiable when only parts of the available data were used for calibration. For instance, ~~non identifiable groundwater porosity and base level parameters if only discharge was used for calibration.~~when only discharge was used for calibration (green lines), the parameters related to groundwater (porosity and base level) happen to be unidentifiable.

Kommentar [SB14]: ANON 3

Figure 4: Modelled discharge [m³/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 90th 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 ~~Table 4~~ 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 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~~the~~ higher ~~the~~ temporal resolution ~~is~~.

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 ~~Table 5~~ and in Figure 7. They display the mean model outputs (Qsim, 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.

Kommentar [SB15]: AL oc6

Figure 7: Mean ~~(manipulated)~~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 performance in the validation period has 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) to 21% (discharge prediction) during the validation period ~~for the discharge and groundwater time series~~ indicates acceptable robustness of the calibrated parameters and is comparable to split sample tests in other studies ~~(~~(Parajka et al., 2007; Perrin et al., 2001). In addition, it is~~, which is~~ corroborated by their generally mainly high identifiability derived by SCEM for the final calibration scheme that~~e~~ used all 4 available observed discharge and ground water level time series. ~~Using the different weighting schemes we also see that only the combined calibration with all 4 time series allowed for identifying all model parameters, while using the discharge or the groundwater observations alone would have produced posterior distributions that indicate low sensitivity of some of the model parameters.~~ 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. Using all information, all model parameters are identifiable, which is in accordance with preceding research that showed 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.

Kommentar [SB16]: ANON 6

Kommentar [SB17]: ANON 4

Kommentar [SB19]: ANON 3

Kommentar [AH18]: ANON 5

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., 2013b, 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 interfluves.

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. 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., 2012; 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 ~~new~~ 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 90th 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., 2013a) it still simplifies a karst system's natural complexity. This is obvious in the simulated time series at Ashton Farm and Black House indicate, which also 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; Wheater 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., 2012; 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 90th percentile will allow for a better simulation quality. ~~However, longer time series than available for this study would be needed for a proper evaluation of this idea.~~ This could be further evaluated by using different percentile weighting schemes, stepwise increasing the weight on the target percentile.

Kommentar [SB20]: AL 2

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 ~~might~~ 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.

Kommentar [SB21]: ANON sc6

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., 2013a) 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 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. 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 >90th 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 publications 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 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 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 5). Surface runoff is found by the excess of soil and epikarst storage of the previous model compartment (Eq. (2) in Table 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 5), while the epikarst outflow follows a linear storage assumption (Eqs. (8,9) in Table 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 5). The diffuse component recharges the groundwater compartments beneath the respective epikarst layers ($i = 1 \ldots N\text{-}1$). 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 5) and applied to calculate the discharges of the matrix system (Eq. (13) in Table 5) and the conduit system (Eq. (14) in Table 5), which together sum up to the entire discharge of the system (Eq. (15) in Table 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

Kommentar [AH22]: ANON sc 1

~~8~~9 References

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.

Adams, B., Peach, D. W. D. and Bloomfield, J. P. J.: The LOCAR hydrogeological infrastructure for the Frome/Piddle catchment, Br. Geol. Surv., 2003.

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, doi:10.1016/j.advwatres.2007.12.003, 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 fl ow 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, doi:10.1144/SP364.9, 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(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. 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 methodology/Estimation bayésienne des incertitudes au sein d’une modélisation conceptuelle de bilan hydrologique, Hydrol. Sci. J., 50(1), 2005.

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

Fitzpatrick, C. M.: The hydrogeology of bromate contamination in the Hertfordshire Chalk: double-porosity effects on catchment-scale evolution, University College 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(3), 235–243, 2009.

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

Ford, D. and Williams, P. D.: Karst hydrogeology and geomorphology, John Wiley & Sons. 578 pages., 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, edited by I. A. of Hydrogeologists, Taylor & Francis Group, Leiden, NL., 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(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

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, doi:10.1016/j.advwatres.2013.01.010, 2013a.

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(3), 218–242, doi:10.1002/2013rg000443, 2014a.

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(1), 159–174, doi:10.5194/bg-13-159-2016, 2016.

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(7), 2219–2231, doi:10.5194/hess-16-2219-2012, 2012.

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(8), 6507–6521, doi:10.1002/2014WR015685, 2014b.

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, doi:10.1002/wrcr.20229, 2013b.

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(8), 3305–3321, doi:10.5194/hess-17-3305-2013, 2013c.

Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57(1), 97–109 [online] Available from: http://www.jstor.org/stable/2334940, 1970.

Hill, M. E., Stewart, M. T. and Martin, A.: Evaluation of the MODFLOW-2005 Conduit Flow Process, Ground Water, 48(4), 549–559, doi: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(1–2), 109–123 [online] Available from: http://www.sciencedirect.com/science/article/B6V6C-51SFK71-3/2/8628ab511bef937dce89e9580f1aeb06, 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. Geogr., 39(1), 49–67, 2015.

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(1–2), 12–28, doi:10.1016/j.jhydrol.2010.12.028, 2011.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29(8), 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., 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, doi: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(1), 117–128, doi:10.1144/GSL.SP.1998.130.01.11, 1998.

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

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

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

Kumar, R., Musuuza, 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(3), 1117–1131, doi: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, doi:10.1016/j.jhydrol.2013.11.017, 2014.

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., Stagge, J. H. and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, Hydrol. Process., n/a-n/a, doi:10.1002/hyp.10838, 2016.

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(3), 604–620, 2006.

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

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

Macdonald, D. D. M. J., Bloomfield, J. P. P., Hughes, a. G. G., MacDonald, a. M. 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 Sept - 2 Oct 2008. Netherlands, ~(1), 10, 2008.

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

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 O and 3 H isotopes, J. Hydrol., 256(1), 48–59, doi: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(1), 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, doi: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(4), 1–12, doi: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(6), 1087, doi:10.1063/1.1699114, 1953.

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

NRA: The Frome & Piddle Management Plan - Consultation Report, 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, doi:10.1016/j.jhydrol.2014.05.038, 2014.

Oehlmann, S., Geyer, T., Licha, T. and Sauter, M.: Reduction of the ambiguity of karst aquifer modeling through pattern matching of groundwater flow and transport, Hydrol. Earth Syst. Sci., 16, 11593, doi:10.5194/hess-19-893-2015, 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(2), 893–912, doi: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(4), 435–446, doi:10.1002/hyp.6253, 2007.

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., 241, 275–301, 2001.

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

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., Proc. Natl. Acad. Sci. U. S. A., 111(9), 3262–7, doi:10.1073/pnas.1222473110, 2014.

Reeves, M. J.: Recharge and pollution of the English Chalk: some possible mechanisms, Eng. Geol., 14(4), 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(11), doi:10.1029/2011wr010446, 2011.

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(10), 1–17, doi: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(8), 1087–1100, doi:10.1016/j.advwatres.2008.04.012, 2008.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geogr. Rev., 38(1), 55–94, doi: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(12), 1949–1963, 2011.

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(2), 318–332, doi:10.1061/(ASCE)HE.1943-5584.0000433, 2012.

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(8), 18, 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, doi: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(1), 89–109, doi:10.1002/hyp.3360010109, 1986.

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

Williams, A., Bloomfield, J., Griffiths, K. and Butler, A.: Characterising the vertical variations in hydraulic conductivity within the Chalk aquifer, J. Hydrol., 330(1), 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(1), 153–166, 1997.

~~9~~10 Figures

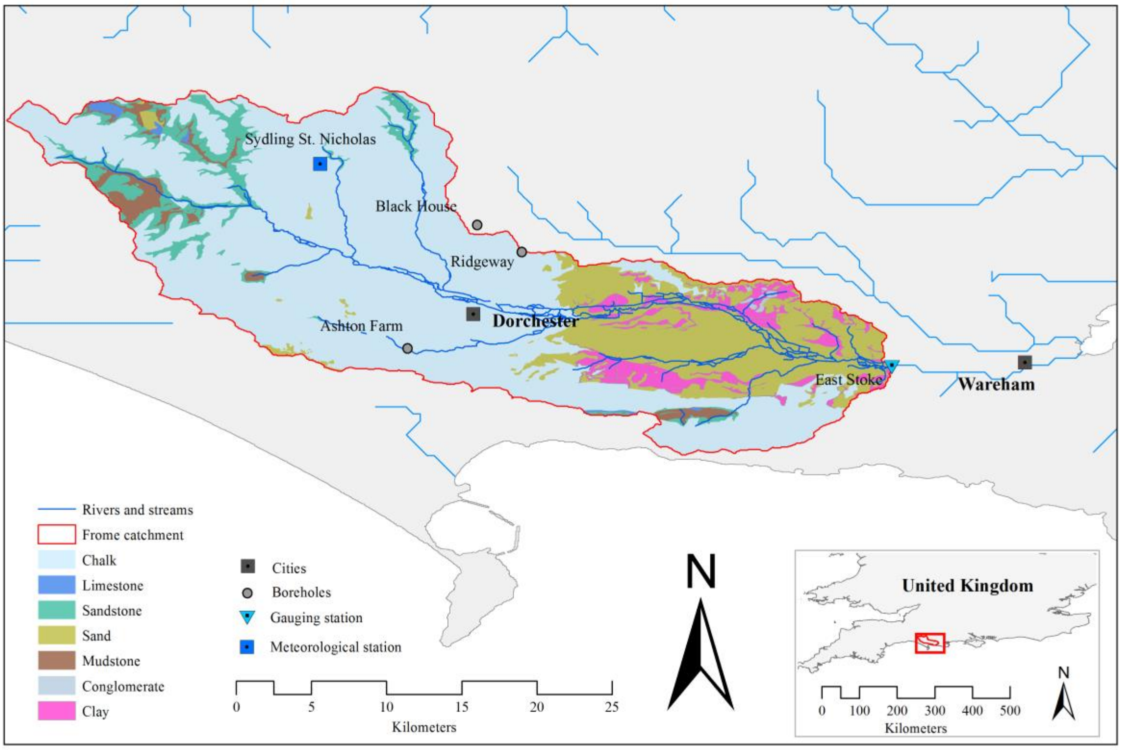

Figure 1: Overview on the Frome catchment

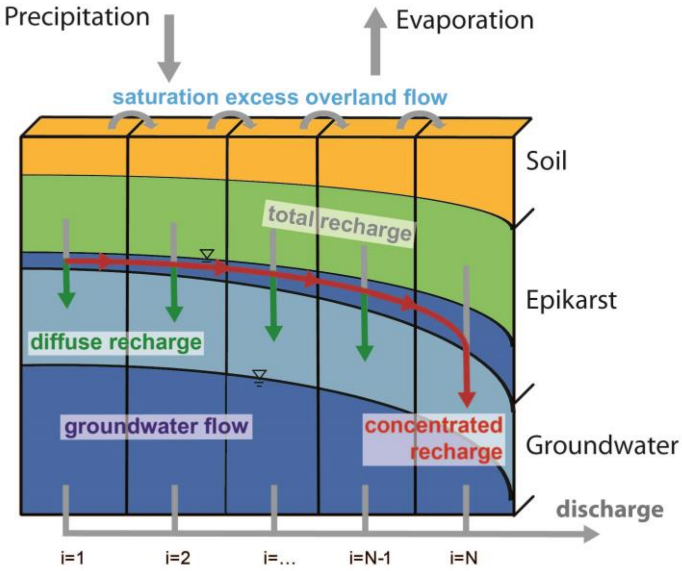

Figure 2: The VarKarst model structure

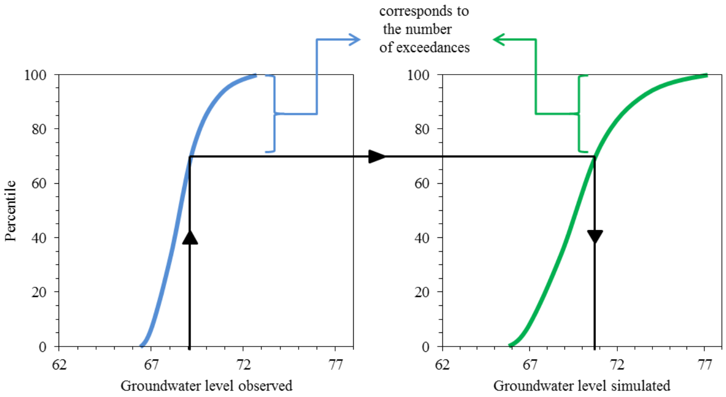

Figure 3: Schematic description of the percentile approach

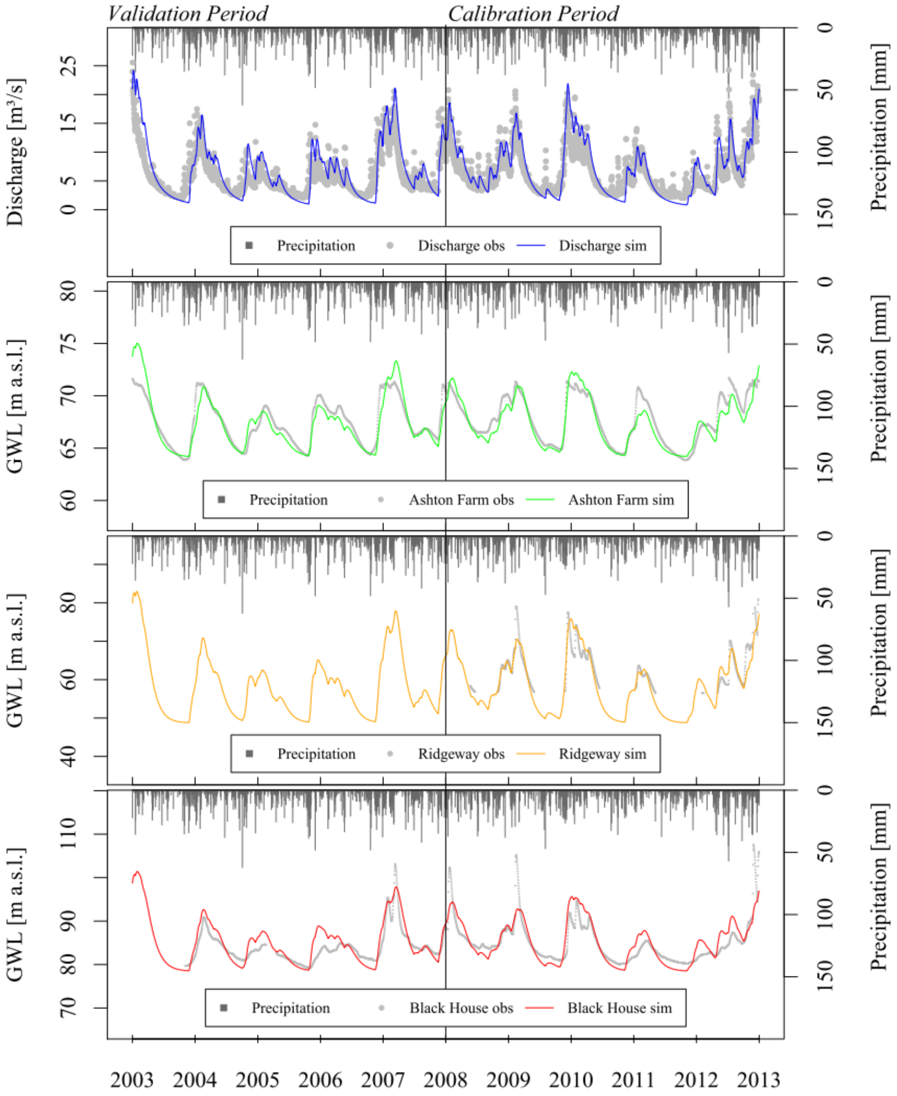

Figure 4: Modelled discharge [m³/s], and groundwater levels [m a.s.l.] at the boreholes Ashton Farm, Ridgeway and Black House

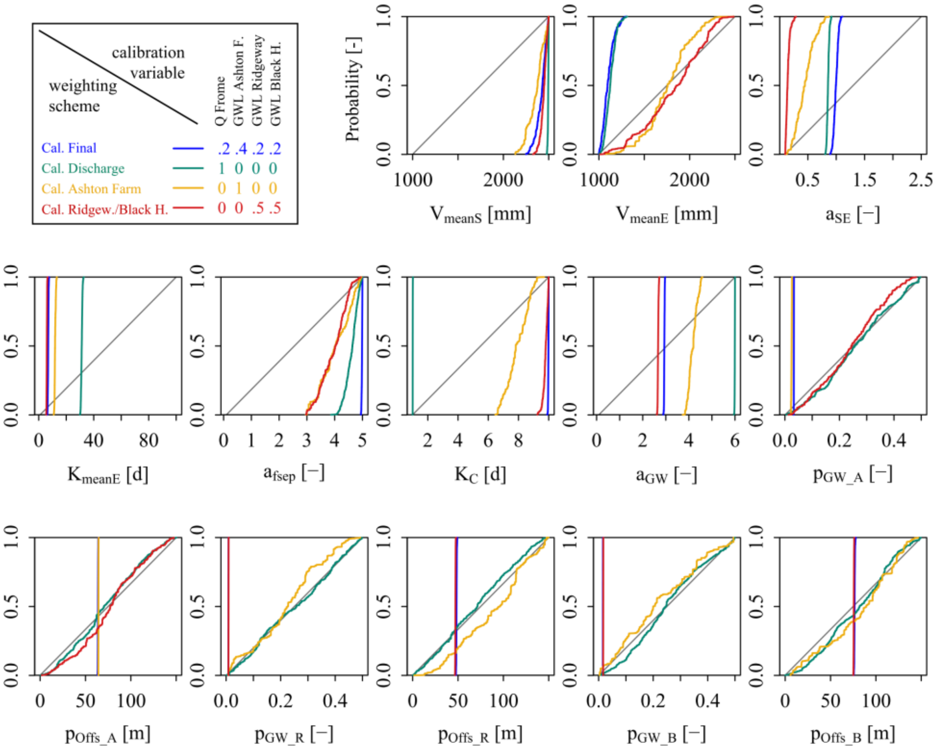

Figure 5: Cumulative parameter distributions (blue) of all model parameters; strong deviation from the 1:1 (dark grey) indicate good identifiability

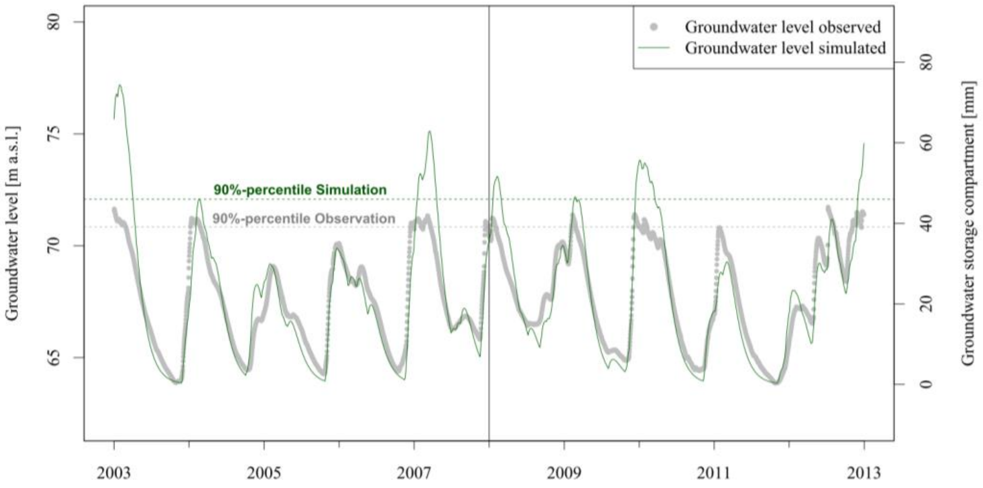

Figure 6: Illustration of the percentile approach. Time series of the oberved (grey dots) and modelled (green line) groundwater level at Ashton Farm. The dotted lines represent the respective 90th percentile

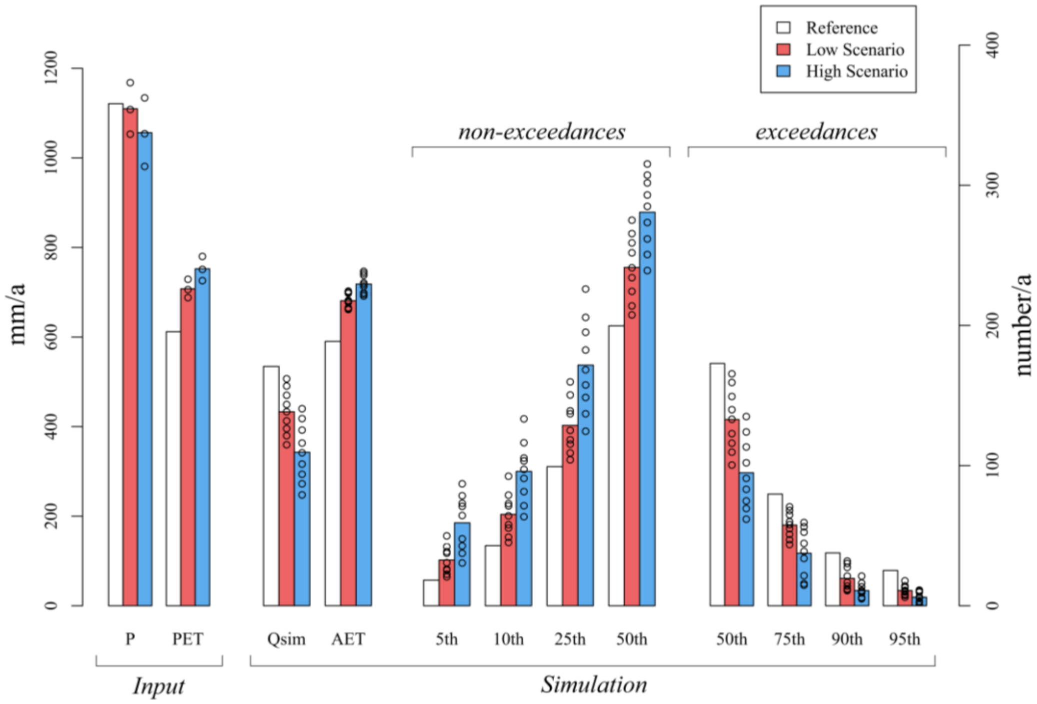

Figure 7: Mean ~~(manipulated)~~ model input (mm/a), mean modelled output (mm/a) and mean (non-)exceeded percentiles (number/a) in the reference period and both scenarios (borehole: Ashton Farm; future period: 2070-2099). The circles indicate the spread among the 9 realisations for each of the two scenarios

Kommentar [SB23]: AL oc5

~~10~~11 Tables

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^{-1}
Discharge	East Stoke (44001)	CEH	01.01.2000-31.12.2012	daily	m^3s^{-1}
Pot. Evapotranspiration	Catchment Cut East Stoke	CEH	01.01.2000-31.12.2008	daily	mm d^{-1}
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: Parameters, descriptions and equations solved in the VarKarst model~~

~~Model routine~~	~~Variable~~	~~Description~~	~~Equation~~	~~Unit~~
Soil	$E_{act,i}(t)$	~~Actual evapotranspiration~~	$= E_{pot}(\mathrm{t})\frac{\min[\mathrm{V_{Soil,i}(t) + P(t) + Q_{Surface,i}(t),\ V_{S,i}}]}{\mathrm{V_{S,i}}}$	~~mm d^{-1}~~
	$Q_{Surf,i+1}(t)$	~~Surface flow to the next model compartment~~	$= max[\mathrm{V_{Epi,i}(t) + R_{Epi,i}(t) - V_{S,i},\ 0}]$	~~mm d^{-1}~~
	$V_{max,S}$	~~Maximum soil storage capacity~~	$= V_{mean,S}2^{\left(\frac{a_{SE}}{a_{SE}+1}\right)}$	~~mm~~
	$V_{S,i}$	~~Soil storage distribution~~	$= V_{max,S}\left(\frac{i}{N}\right)^{a_{SE}}$	~~mm~~
	$R_{Epi,i}(t)$	~~Recharge to the epikarst~~	$= max[\mathrm{V_{Soil,i}(t) + P(t) + Q_{Surface,i}(t)\ - E_{act,i}(t) - V_{S,i},}]$	~~mm d^{-1}~~
Epikarst	$V_{max,E}$	~~Maximum epikarst storage capacity~~	$= V_{mean,E}2^{\left(\frac{a_{SE}}{a_{SE}+1}\right)}$	~~mm~~
	$V_{E,i}$	~~Epikarst storage distribution~~	$= V_{max,E}\left(\frac{i}{N}\right)^{a_{SE}}$	~~mm~~
	$Q_{Epi,i}(t)$	~~Outflow of the epikarst~~	$= \frac{\min[\mathrm{V_{Epi,i}(t) + R_{Epi,i}(t) + Q_{Surface,i}(t),\ V_{E,i}}]}{\mathrm{K_{E,i}}}\Delta\mathrm{t}$	~~mm d^{-1}~~
	$K_{E,i}$	~~Epikarst storage coefficient~~	$= K_{max,E}\left(\frac{N-i+1}{N}\right)^{a_{SE}}$	~~d~~
	$R_{diff,i}(t)$	~~Diffuse recharge~~	$= f_{C,i}Q_{Epi,i}(t)$	~~mm d^{-1}~~
	$R_{conc,i}(t)$	~~Concentrated recharge~~	$= (1 - f_{C,i})\ Q_{Epi,i}(t)$	~~mm d^{-1}~~
	$f_{C,i}$	~~Recharge separation factor~~	$= \left(\frac{i}{N}\right)^{a_{fsep}}$	~~-~~
Groundwater	$Q_{GW,i}\ (t)$	~~Groundwater contributions of the matrix~~	$= \frac{V_{GW,i}(t) + R_{diff,i}(t)}{K_{GW,i}}$	~~mm d^{-1}~~
	$Q_{GW,N}(t)$	~~Groundwater contribution of the conduit system~~	$= \frac{\min[\mathrm{V}_{\mathrm{GW},N}(\mathrm{t}) +\ \sum_{i=1}^{N}\mathrm{R_{conc,i}(t)}\ ,\mathrm{V}_{\mathrm{crit},OF}]}{\mathrm{K_C}}\Delta\mathrm{t}$	~~mm d^{-1}~~
	$K_{GW,i}$	~~Variable groundwater storage coefficient~~	$= K_C\left(\frac{N-i+1}{N}\right)^{-a_{GW}}$	~~d~~
	$Q_{main}(t)$	~~Discharge~~	$= \frac{\mathrm{A_{max}}}{\mathrm{N}}\sum_{i=1}^{N}Q_{GW,i}(t)$	~~l s^{-1}~~

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman), Fett

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert: Schriftart: (Standard) +Textkörper (Times New Roman)

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

Formatiert ...

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.

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 [%]

	Percentiles						
Time period	**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

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^{-1}	(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^{-1}	(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	$= max[V_{Soil,i}(t) + P(t) + Q_{Surface,i}(t)\ - E_{act,i}(t) - V_{S,i},]$	mm d^{-1}	(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^{-1}	(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^{-1}	(10)
	$R_{conc,i}(t)$	Concentrated recharge	$= (1 - f_{C,i})\ Q_{Epi,i}(t)$	mm d^{-1}	(11)
	$f_{C,i}$	Recharge separation factor	$= \left(\frac{i}{N}\right)^{a_{fsep}}$	-	(12)
Groundwater	$Q_{GW,i}\ (t)$	Groundwater contributions of the matrix	$= \frac{V_{GW,i}(t) + R_{diff,i}(t)}{K_{GW,i}}$	mm d^{-1}	(13)
	$Q_{GW,N}(t)$	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^{-1}	(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^{-1}	(16)