Responses to Editor's and Reviewers' comments

Editor

Dear Authors,

Your submission was evaluated by three reviewers who are well-focused on the topic but provided slightly different rates and comments mainly due to their different backgrounds. Overall, the scientific quality and significance of your study were evaluated quite well, whereas the presentation of the investigation requires some improvement. I support the general assessment of these reviewers. However, as also raised by one reviewer, I do suggest the uncertainty issue should be discussed more in-depth as it is an interesting question of your study that definitely can stimulate the interest of a wider readership toward your paper.

Your replies during the discussion phase showed that your group is acquainted with this question, but I think that treating epistemic uncertainty through an assessment of parameter sensitivity might resolve only an aspect of the problem. Maybe I miss something, but the question of parameter correlations and how it reflects on the propagation of uncertainty seems a bit overlooked. One question is to assess parameter uncertainty, whereas another question is to assess how epistemic uncertainty is reflected in the output of a model (a hydrologic one in this case). It is not always the case that uncertainty in the inputs is then lowered in the model outputs (as it can happen when comparing infiltration versus drainage processes). You may find it useful to discuss this latter question in the paper. Due attention should also be devoted to the comments from Reviewer #3.

I recommend reconsideration following major revisions, and invite you to submit your revised manuscript together with detailed point-by-point responses to all comments received thus far. Should you disagree with a comment or suggestion, please explain why clearly.

Response: We thank the Editor and all the Reviewers for their comments and suggestions, which have improved our manuscript. All comments have been thoroughly considered in our revisions. In particular, we included more details on the Bayesian uncertainty analysis by adding

- An improved description of the method
- An additional figure to discuss marginal posterior distributions and address the effect of parameter interdependencies/ correlations between parameter values in posterior samples on propagated predictive uncertainty
- Additional figures and discussion of propagated parameter uncertainties in cumulative and peak recharge estimation

As suggested by the editor, we address the topic of epistemic uncertainty in Lines 339-347 of the revised and marked manuscript:

"Uncertainties in water fluxes were treated as aleatory, derived from stationary statistical characteristics. In addition, the epistemic uncertainty associated with the lack of knowledge about the correct representation of system dynamics (conceptual uncertainty) and forcing data may affect the overall predictive uncertainty and reduce the effective information content of observations (Beven, 2016). Due to their complex and often dynamic nature, epistemic uncertainties pose important conceptual and numerical challenges. For instance,

model conceptual uncertainty can be assessed by comparing different model structures using specific statistical metrics (e.g., marginal likelihood). This was, however, beyond the scope of the present study, which focuses on the inverse estimation of soil water fluxes at multiple monitoring stations to discuss implications on the water balance. In this framework, an appraisal of the model structural adequacy through posterior predictive checks appears sufficient."

In the following, we describe our revisions note by note in response to the Reviewers' comments, indicating the line numbers of the <u>marked manuscript</u>.

We welcome the suggestion of the Editor to reconsider our thoroughly revised manuscript for publication.

Reviewer #1

I enjoyed this paper. The authors made use of an extensive set of water content profile time series in two differing climates within Austria. They used an accepted Bayesian inference algorithm to fit HYDRUS1D to the observations. They found that recharge is highly correlated with precipitation in the shorter term and aridity in the longer term. These results are well aligned with expectations from previous publications. My only complaint is that the method is based on a Bayesian approach, but there is almost no discussion of the resulting uncertainty of the inferences, how these depend on site conditions, how these might affect larger interpretations, and how the data did or did not constrain these uncertainties. The results and conclusions are primarily based on deterministic findings that seem to rely on the maximum likelihood parameter values. (Although this isn't discussed in detail.) I am actually fine with that - as stated above, I think that this makes a useful contribution. But, in the end, I was left wondering why the Bayesian approach was used rather than another parameter estimation tool. I would like to see the authors include some discussion of the special insights that resulted from the Bayesian analysis.

Best

Ty Ferre

Response: We thank Dr. Ferré very much for his appreciation of our work! We agree that a more detailed presentation and discussion of the results of the Bayesian analysis is helpful.

In our revised manuscript, we added a figure to show marginal posterior distributions of SHPs for one exemplary site. In this context we further discussed how parameter uncertainty was constrained by the observations in Lines 249-256:

"A typical example for marginal posterior distributions resulting from SHP estimation on the basis of volumetric soil water content data in this study is shown in Figure 3 for the upper soil layer of the mountain location Zettersfeld. Limits of the plot axes are given by the prior bounds. This representation shows how well the calibration data constrained uncertainties of each parameter: the posterior range of θ_r is only slightly reduced compared to the prior range, indicating that θ_r was least sensitive for simulating soil water content and poorly informed by observations. The parameter K_s has a wide posterior range (although clearly reduced compared to the prior), showing a logarithmic distribution and a clearly defined mode. On the other hand, the parameters α, and especially n and θ_s, show narrow posterior distributions which appear leptokurtic, indicating a higher sensitivity for the soil water content simulations and a high information gain from the calibration data."

We further included two figures to visualize the results of the uncertainty propagation for (1) cumulative recharge and (2) magnitudes of recharge peaks at all sites in the revised manuscript, and discussed them in Lines 318-335:

"Uncertainties in the prediction of long-term and cumulative recharge rates (Figure 5) were generally small in relation to the high sums estimated for mountain and western sites; however, in relation to low absolute values at dry eastern sites Kalsdorf, Lobau and Frauenkirchen, posterior uncertainties played a more important role. The relative uncertainty (IQR/median) in long-term recharge estimates ranged between 1%

(Gschlössboden, Lauterach) and 39% (Lobau). The prediction of peaks in recharge was generally affected by higher uncertainties (Figure 6), especially at western mountainous sites with high maximum rates (Lauterach, Leutasch, Gschlössboden, Sillianberger Alm). In a previous study with similar models and hydrological conditions, we found n to be the most sensitive parameter for cumulative recharge prediction, and K_s to be most sensitive for peak prediction, especially under wet climatic conditions (Schübl et al., 2022). Small uncertainties in the prediction of long-term recharge sums here were related to the generally small uncertainties in the VGM shape parameter n, whereas higher uncertainties in the hydraulic conductivity parameter K_s (sometimes in interaction with uncertainties in α) can be considered the main reason for the greater uncertainty in the peak prediction.

It has often been found that *in-situ* field measurements of soil water content are not sufficient for the accurate and precise estimation of SHPs (Scharnagl et al., 2011; Ritter et al., 2003). Here, we found that while SHPs were partially affected by considerable uncertainties, the precision was still in an acceptable range for groundwater recharge estimation. At dry locations with small absolute recharge rates, model uncertainties could be further reduced e.g., by including additional observations in the calibration. Especially the combination with soil matric potential measurements has been shown to be highly informative for SHP estimation and to considerably reduce uncertainties in recharge estimation (Schübl et al., 2022; Schelle et al., 2012).".

To give better insights on how site conditions and measurements have affected model parameter uncertainties, we have expanded the discussion in Lines 271-290:

"An important factor for parameter uncertainty was soil texture: uncertainties in parameters K_s and n, in terms of the 95% interquantile range (IQR) in posteriors, were significantly positively correlated with the percentage in sand (r=0.43 and r=0.42, respectively). Uncertainty ranges in K_s, α, and n increased significantly with the value of median estimates (Fig. 4). Higher values of these parameters signify a lower water retention capacity of the soil. According to results from Schübl et al. (2022) and Gao et al. (2019), parameter uncertainty from calibration with daily soil water content measurements can be expected to be higher in coarse-textured soils (with a higher soil hydraulic conductivity and lower soil water retention capacity), than in fine-textured soils which was the case in this study. We suppose that the more rapid water flow processes are less efficiently captured in daily soil water content measurements, which are consequently less efficient in constraining uncertainties in SHPs.

We expectedly found high parameter uncertainties for sites where the estimated errors were high ($\sigma > 0.04$ cm^3 cm^{-3} at Karlsdorf and Lobau) or where the error was high in comparison to the temporal variation (more than 90% of the standard deviation in the observations at Zettersfeld and Sillianberger Alm). Xie et al. (2018) observed how the relation between size of the estimated measurement error and temporal variation of the measured variable influences the data's ability to constrain model uncertainties. Brunetti et al. (2019) observed in the estimation of SHPs with remote sensing soil moisture data, that uncertainty in θ_r estimation was low whereas θ_s was highly uncertain. This was related to soil water content values being low in their study and mainly representative for unsaturated conditions. In this study, at Lauterach and Elsbethen, very wet climatic conditions and measurements mainly

in the wet range resulted in the highest uncertainties in the estimation of θ_r. At Kalsdorf, in contrast, soil moisture dynamics were hardly at saturation and resulted in the highest uncertainty in θ_s estimation. At the majority of the Austrian locations, soil water content measurements were more often near saturation and less in the dry range (as for example in Fig. 2(a) at Gumpenstein). The θ_s parameter was therefore mostly better informed by the measurements than θ_r.”

The second part of the paper was aimed at characterizing the Austrian sites based on measured and modeled hydrological variables and site properties. In the revised manuscript, we clarified our approach in Lines 376-379:

“In the following section we characterize the 14 monitoring according to hydrologically relevant properties including model estimations from the previous section. Since uncertainty in long-term actual evapotranspiration and recharge rates were generally low, and to enable the analysis with common statistical tools, we will proceed from here on using the median values without consideration of uncertainty ranges.”

Reviewer #2

General remarks:

The manuscript is very well structured, very well written and represents an interesting scientific contribution to the determination of groundwater recharge rates. Changes to the manuscript are not necessary.

Specific comments:

The sites in Austria used for the study were described comprehensively, as were the statistical methods used. Why Bayes' theorem was used in the statistical analysis was sufficiently explained. In order to be able to better evaluate the results obtained, the methodological limitations were explained in detail.

Technical corrections: no corrections necessary

Response: We are very happy about the positive and motivating feedback on our work from Reviewer#2! We thank him/her for reviewing our manuscript.

Reviewer #3

Dear Editor, dear authors,

Please find below my review of the paper entitled "From soil water monitoring data to vadose zone water fluxes: a comprehensive example of reverse hydrology" by Marleen Schübl, Giuseppe Brunetti, Gabriele Fuchs, and Christine Stumpp.

This article investigates the use of the Bayesian approach to invert water content profiles to derive the soil hydraulic parameters, including their statistical distribution and related statistical parameters. Based on this information, the authors compute the water cycle over large periods and quantify the groundwater recharge, its uncertainty, and its temporal variability at 14 sites in Austria. The authors conclude that there is a West-East gradient with more continuous groundwater recharge at mountainous sites with wetter climates versus seasonal lower groundwater recharge in the Eastern plain.

The article is well-organized, well-illustrated, and well-written. The scientific question is properly stated, the methodology to answer conclusions is clear and straightforward, and the conclusions are quite obvious. The paper addresses an important topic and deserves to be published in the HESS journal. However, I have several concerns that should be addressed prior to publication.

Response: We thank Reviewer#3 for his/her appreciation of our paper and the helpful feedback.

I am not very familiar with the Bayesian approach, and thus hope that my comments do not reveal my lack of expertise in this subject. However, I consider that any paper should be standalone and present clear facts understandable by any scientific reader. Several points deserve to be clarified:

- For the Bayesian approach, the choice of distributions needs to be clarified. If the errors between the modeled and observed data are expected to obey the normal law, no details are given about the laws of the soil hydraulic parameters. I expect most parameters to follow normal laws and hydraulic conductivity to follow a log-normal law. If I understand well, the Bayesian approach allows us to characterize the SHP laws. Then, why not show them in the Result section and state on the alignment of normal laws? Why not state on the multimodality features of the SHPs? In addition, do the SHPS distributions have any consequences on the Bayesian approach and the Monte Carlo method? Is the normality of errors between experimental and modeled data compatible with any statistical law for SHPs?

  Response: The Reviewer raises an important point of discussion to better clarify our approach.

  The Bayesian inference can be applied directly to obtain Soil Hydraulic Parameters (SHPs), if data include pressure head, water content, and conductivity (e.g., laboratory measurements derived from the simplified evaporative method). Instead, our study uses field scale observations of volumetric water content to inversely estimate the most probable distribution of SHPs that generated observations. Therefore, the assumption of homoscedastic and normal errors (reflected in the likelihood function) refers to TDR measurements, not to SHPs. The posterior distribution of SHPs is inferred by combining two components: 1) the **prior** distribution, which reflects the modeler's believe about Soil Hydraulic Parameters (SHPs) **before** considering measurements (in our case, soil water content), 2) the likelihood, which describes the probability

that a parameter set drawn from the prior has generated the observations. By combining the likelihood and the prior with Nested sampling and HYDRUS (or Markov-chain Monte Carlo), we obtain a **posterior** distribution of the most probable SHP values, which reflects the parameters' uncertainty:

1. Prior: As it is frequently the case in vadose zone hydrology (e.g., Brunetti et al. 2020 https://doi.org/10.1016/j.jhydrol.2020.124681, Wöhling et al. 2015 https://doi.org/10.1002/2014WR016292), we assumed bounded uniform priors to avoid making important assumptions on the shape of the posterior, and let the data tell us what is its shape. However, at the same time, we imposed hard boundaries on the parameters to avoid the possibility to obtain physically unrealistic values.
2. Likelihood: We assumed that sensor errors are normally distributed. This is widely established approach in inverse vadose zone modeling (e.g., Schelle et al.. 2012 https://doi:10.2136/vzj2011.0169, Gao et al. 2019 https://doi.org/10.2136/vzj2019.03.0029).

- The problem of equifinality and non-uniqueness needs to be addressed in the paper. The authors inverted all the SHPs, except the parameter "l" fixed at 0.5. However, we know that many parameters may be interrelated, and some may have a poor impact on water fluxes. In particular, the value of the residual water content has no effect (or very little on the water fluxes), so this parameter may not be reachable when inverting. A similar statement may apply to the saturated water content. What is the strategy of the authors regarding this aspect of non-uniqueness? Why not propose a sensitivity analysis that shows the most influential parameters and select those when inverting water content data while suggesting additional information for the others?

Response: Also in this case, we thank the Reviewer for pointing this aspect, which gives us the opportunity to better clarify the utility of the Bayesian approach to address equifinality, and to discuss the limitations of the sensitivity analysis compared to the Bayesian probabilistic approach.

The Bayesian inference is precisely conceived to have a statistical rigorous appraisal of the "equifinality and non-uniqueness". The Bayesian approach infers a marginal posterior distribution that exposes the parameters' uncertainty, and their interaction (e.g., correlation). If the resulting uncertainty is high (i.e., wide posterior), then data are not informative for that parameter. In this case, the modeler has two choices: 1) ask for other measurements (e.g., disk infiltrometer for Ks) to have more informative priors and run again the Bayesian analysis; 2) honestly communicate what is the parameters' uncertainty with the data available, and more important, how the estimated uncertainty propagates in the model simulations. This is what we precisely did in our study.

The sensitivity analysis is certainly a valuable tool, which we use frequently in our research. However, it will not provide any more meaningful information compared to the Bayesian analysis for this work. A global sensitivity analysis will sample the parameters' space (frequently ineffectively as MCMC techniques are much better in finding high-probability regions), and then decompose the variance to identify influential and uninfluential factors. But this is already better targeted in a Bayesian analysis: influential parameters are those that exhibit leptokurtic

posteriors, while uninfluential factors are those that have similar prior-posterior distributions (in our case flat). The sensitivity analysis might have some utility in high dimensions for numerical sampling reasons, but this is not the case and is beyond the purpose of the present study.

In the introduction of the revised manuscript, we have added to the description of the Bayesian approach in Line 65:

"Posterior parameter distributions also reflect the non-uniqueness and equifinality of parameter values."

We further justified our choice of prior distributions and have added some clarification on the Bayesian method in Lines 185-191:

"Prior knowledge, i.e. information available before looking at measured data, is included in the Bayesian inference via the prior distribution which can be chosen as a uniform density bounded by physical limits (e.g., Brunetti et al., 2020b; Gupta et al., 2022; Wöhling et al., 2015). In this study, uniform prior distributions were assumed for all parameters and sites. Their ranges were established based on texture information, literature review, and preliminary testing to prevent truncating posteriors. Final ranges are given in the Appendix in Table A3. By combining the likelihood and the prior, we obtain a posterior distribution of the most probable SHP values, which reflects the parameters' uncertainty."

In our revised results section, we included an example of the marginal posterior distributions of SHPs and discussed them in Lines 249-263:

"A typical example for marginal posterior distributions resulting from SHP estimation on the basis of volumetric soil water content data in this study is shown in Figure 3 for the upper soil layer of the mountain location Zettersfeld. Limits of the plot axes are given by the prior bounds. This representation shows how well the calibration data constrained uncertainties of each parameter: the posterior range of θ_r is only slightly reduced compared to the prior range, indicating that θ_r was least sensitive for simulating soil water content and poorly informed by observations. The parameter K_s has a wide posterior range (although clearly reduced compared to the prior), showing a logarithmic distribution and a clearly defined mode. On the other hand, the parameters α, and especially n and θ_s, show narrow posterior distributions which appear leptokurtic, indicating a higher sensitivity for the soil water content simulations and a high information gain from the calibration data."

Parameter interdependencies in the inverse estimation are reflected in the shapes of bivariate contour or scatter plots of posteriors (see Figure A2 in the Appendix for a representation of posteriors with closer axes ranges). By random sampling from the posterior, the effect of these correlations is propagated in the uncertainty in the prediction of soil water fluxes. Usually, a negative relation exists between the VGM shape parameters (e.g., Scharnagl et al. 2011; Romano and Santini, 1999; Vrugt et al., 2003). Here, both α and n show narrow posteriors and stray very close to the lower physical bounds (0 and 1, respectively). The correlation of posterior samples for α and K_s can be expected to have some effect on the uncertainty in recharge peak prediction, for which both parameters (but especially K_s under wet conditions) are sensitive (Schübl et al., 2022)."

- I also have some concerns regarding the data inverted to derive the SHPs. In their study, the authors invert only water content profiles. However, if I remember well, they also have water pressure head profiles for some sites. I understand they selected the water content profiles because they had those data at their disposal at all sites. However, for a given site (with the two types of data), they could have compared the results when inverting water content and water pressure head. My feeling is that the authors may not have had the same results. Based on this comparison, they might validate the choice of water content for all sites and strengthen their conclusions. That could be the topic of further research.

  Response: In general, we agree that it is helpful to include pressure head data, as it can help to identify SHPs with even less uncertainty than with soil water content data (Schübl et al., 2022). However, at the sites in this study we had some issues with soil pressure head measurements: (1) they were not available for all sites which would have impaired the comparability of results between locations (2) they were composite from different measurement techniques (tensiometers and gypsum blocks) and included sudden shifts and large gaps in time series. Altogether we found the measurements to not be reliable enough to be used in this study. We agree that results for SHP estimates might change when including the available soil pressure head data and have included this in the discussion (see reply to comments in manuscript). We also agree that this would be a very interesting topic for further studies with improved field measurements of soil pressure heads.

- Lastly, I had some questions and concerns about the ACP proposed at the end of the result section. I was surprised by the plots of "individuals" (i.e., sites) and the "variables" on the same plots (Figure 4). Even after searching on R tutorials and finding these types of plots, I am not convinced that we have the right to do so. For ACP, variables and individuals don't have the same nature and should be plotted on separate plots. I also suggest plotting the correlation circles and commenting only on the vectors (variables) that are well represented on the maps, i.e., which vector is close to the correlation circle.

  Response: We used this PCA Biplot here (Figure 7 of the revised manuscript) to visualize the two clusters of hydrologically similar sites in context with the variables according to which they have been characterized. This kind of visualization, with individuals (sites or samples) and variables in the same plot, has been used in several studies, for example by Luna et al., 2018 https://doi.org/https://doi.org/10.1002/eco.1896 (Figure 7), Rodríguez et al., 2020 https://doi.org/10.1007/s10750-020-04201-5 (Figure 3), Gibson et al., 2019 https://doi.org/https://doi.org/10.1016/j.ejrh.2019.100643 (Figure 7).

  Some R tutorials showing this kind of plot are

    1. https://f0nzie.github.io/machine_learning_compilation/detailed-study-of-principal-component-analysis.html (See 4.20 Biplot)
    2. https://www.datacamp.com/tutorial/pca-analysis-r
    3. https://finnstats.com/index.php/2021/05/07/pca/

  Citing from the first tutorial for the use of Biplots with variables and individuals, the focus is "…on the direction of variables but not on their absolute positions on the plot. Roughly speaking, a biplot can be interpreted as follows: an individual that is on the same side of a given variable has a high value for this variable; an individual that is on the opposite side of a given variable

has a low value for this variable." We wrote the code for our plot in Python using the sklearn module (Pedregosa et al., 2011 https://arxiv.org/abs/1201.0490) which is now cited properly in the revised manuscript.

We agree with the comment in the manuscript that the data is not well represented in the Biplot of Figure 4(b) and it does not add further insight, we therefore deleted it.

The authors will find an in-depth review in the enclosed file, with suggestions, comments, and proposals throughout the manuscript. Again, this paper is valuable and promising, and I have no doubts that it will be published after improvements.

Response: We thank the Reviewer for the positive assessment of our work. We are also grateful for the suggestions, which helped to improve the paper.

Comments in Manuscript:

L75: Why not also consider water pressure head profiles? Inverting other types of measures could change the features of estimates.

Response: In general, we agree that it is helpful to include pressure head data, as it can help to identify SHPs with even less uncertainty than with soil water content data. However, at the sites in this study we had some issues with soil pressure head measurements (see manuscript L105/6 and our reply comment). We agree that results for SHP estimates might change when including the available soil pressure head data, therefore we included this in the discussion in Lines 334-337:

"[…] Especially the combination with soil matric potential measurements has been shown to be highly informative for SHP estimation and to considerably reduce uncertainties in recharge estimation (Schübl et al., 2022, Schelle et al. 2012). Including additional measurements in the analysis, however, might not only lead to different shapes in SHP posteriors, but to altogether different estimates. This issue requires further investigation with available soil water monitoring data."

L103: Any rationale for the choice of the period?

Response: We in included the justification in Lines 106-110:

"The length of calibration periods was chosen to be similar for all sites, long enough to be informative for a range of soil water conditions, and excluding the winter season which would require simulation of snow accumulation and melt processes. This exclusion allowed to reduce computational cost and numerical sensitivity of the simulations which often lead to non-convergence or delayed convergence of the sampling algorithm in the Bayesian analysis (described in Sect. 2.3)"

L105/106 That is a good reason. We could have imagined using both water content and water pressure head profiles at some sites to compare the estimates and their distributions when one or the other signal is used. Which signal is the best? That may be a question for further studies.

Response: We investigated this question in Schübl et al. (2022) and came to the conclusions (based on synthetic data scenarios within the tensiometer range) that SHPs are identifiable and in general associated with less uncertainty when using matric potential as compared to soil water content measurements (see comment above). For the reasons highlighted, we did not use water pressure head data here, however, we agree that this would be a topic for further studies with improved field measurements.

In Line 122 we have corrected the symbol.

In Line 134 we have removed the bracket.

L144/145: You simplified the soil profile. However, weren't there some sites with more than three soil horizons? What could be the consequences of such simplifications on soil hydraulic characterization and water fluxes?

Response: We included in Lines 148-152:

"The vast majority of the soil profiles indicated a distinct topsoil overlying deeper soil layers that had low to mild degrees of inhomogeneity. Dealing with 14 monitoring stations, we uniformly adopted two soil layers with varying thickness across different locations, aiming to reduce the overall computational burden of the Bayesian analysis while maintaining a physically realistic description of the soil domain. Simplifications of the soil profile in the model geometry with a mildly heterogeneous soil will usually lead to an acceptably small loss of accuracy in effective parameters (Schneider et al., 2013)."

L147: Could you cite works that justify that the zero flux plane (ZFP) must be above 1.5m depth in all your sites? It seems reasonable to think that 1.5m is enough to include the ZFP and estimate GW recharge. But it is better to cite at least one source.

Response: In Lines 156-158 we included:

"Similar to our approach, Simunek (2015) and Heppner et al. (2007) simulated groundwater recharge with HYDRUS-1D for grass-covered soils as bottom flux at 100 cm profile depth; Assefa and Woodbury (2013) used different profile depths of up to 150 cm."

L158/159: Unclear. Does that concern only the simulation of the six-month period? If yes, please clarify your choice in terms of simulated chronics here.

Response: We rephrased this in Lines 167-169:

"For inverse parameter estimation during the half-year calibration periods, as well as for the model validation periods, we chose boundary conditions with respect to the conditions at the measurement plots, i.e. seepage face for the lysimeter sites and free drainage for sites with natural field conditions."

L166/167: Do you have any reference to justify such linearity between -2 and +2°C? I am not convinced by such a hypothesis given the complexity of processes (freezing rain, isothermal snowfall, etc.).

Response: We included a justification in Lines 178-180:

"This default snow routine in HYDRUS is based on assumptions by Jarvis (1994) and has been found to be suitable for estimating soil water fluxes in unfrozen soils in several studies (e.g., Assefa and Woodbury, 2013; Zhao et al., 2008)."

L169: I am fine with this part, but we should somehow address the problem of parameter equifinality and non-uniqueness of solutions. May we, by increasing a parameter, compensate for the effect of the other? This point is somehow addressed when investigating the correlation matrix between parameters in the result section but not evoked in the modeling section. I suggest giving more details in this section about the distributions of each hydraulic parameter and the random sampling for the Monte Carlo method.

Response: We thank the Reviewer for his/her comments on this section. In the revised manuscript, we clarified the Bayesian approach, justified our choice of prior distributions and discussed resulting marginal posterior distributions (see reply to main Review comment).

L172: Couldn't we note this as P(D | \Omega, M), given that the model is chosen as the first step, and then the data D and the parameter \Omega are related against each other (for a fixed model)? I would prefer the notation P(D| \Omega, M) and P(M | D, M) all together.

Response: We used the common notation for the purpose of Bayesian parameter estimation with a previously chosen model (see e.g., Skilling, 2006 https://doi.org/10.1214/06-BA127).

L182: Do we speak of the distribution of randomly sampled values for each parameter or the distribution of hydraulic parameters on the field? Should we consider uniform distributions for all hydraulic parameters? Alternatively, shouldn't we consider a log-normal distribution for Ks and normal distributions for the other parameters? Which impacts does the choice of the distribution have on the likelihood function and on the Monte Carlo method?

Response: The prior distribution reflects the modeler's prior belief about the parameter before running the analysis, and should be supported by available information. The normal distribution implies that we have data-supported guesses about the mean and the standard deviation of the parameter. However, we didn't have these supporting information, therefore we assumed flat priors to let data inform us about the parameters' values.

Impact of the prior on the Monte Carlo procedure: if the prior is close to the posterior, the convergence will be faster. Otherwise, samples drawn from the prior will be frequently rejected as they don't belong to the posterior, and the convergence will be delayed.

The choice of the prior has no impact on the likelihood.

L193: "multimodal distributions" related to the SHPs?

Response: It relates to the shape of the posterior, which can have (possibly) multiple modes (i.e., high-likelihood regions). These distributions are problematic to sample but MULTINEST does that fairly good.

L215: Why not 0.05, as usual?

Response: E.g., Moeck et al., 2020 (https://doi.org/10.1016/j.scitotenv.2020.137042) use a 90% confidence level for similar questions (even with a much greater data set).

Table 1 caption: Please remind us why the credible intervals are not symmetrical while parameter distributions may be quasi-gaussian (except for Ks).

Response: We do not expect that posterior probability distributions of SHPs are necessarily quasi-gaussian. Therefore, uncertainty bounds below and above the median are given separately (instead of being summarized as standard deviation). We included an example of marginal parameter posterior distributions to better illustrate this (see above comments).

Figure 3: Why not distinguish the dots for the two layers? Why don't you show the trends for these parameters? It could be interesting to discuss the differences between SHPs.

Response: We distinguished the two layers and included residual and saturated water content parameters in the plot (Figure 4 of the revised manuscript). We added in the discussion in Lines 271-279 (see reply to comment of Reviewer#1)

L291: "agreement" instead of "fit" which is more dedicated to the alignment of a given model to experimental data.

In Line 367, we replaced fit by agreement.

L294-295: We have one order of magnitude of difference. Does that stem from the representativeness of the lysimeter, as said below?

Response: We suggested that potential reasons for the larger variability in the lysimeter measurements are spatial heterogeneities and uncertainty in seepage measurements (as stated below).

Figure 4: For PCA analysis, theoretically, it is not correct to superpose the plots of individuals and the correlation circle. You can do it when individuals and variables are of the same nature, which is the case of MCA (Multiple Correspondance Analysis) but definitely not the case of PCA. Then, I reckon that you explain how you did your graphs, or alternatively, I reckon that you replace your figure 4 with a new plot with separated subplots of the plans (F1, F2), (F3, F4) and the correlation circle (F1, F2), (F3, F4). Ensure to include the unity circle that indicates if the variable is well represented in the plan or poorly represented. For instance, GWR/P, GWR, and P are poorly represented in the plan (F3, F4).

Response: We used this PCA Biplot here to visualize the two clusters of hydrologically similar sites in context with the variables according to which they have been characterized (not for further statistical analysis). This kind of visualization with individuals (sites or samples) and variables in

the same plot has been used in several studies (see discussion comment for references and links to R tutorials). We wrote the code for our plot in Python using the sklearn module which is now cited properly in the Figure description. It was based on these tutorials:

https://ostwalprasad.github.io/machine-learning/PCA-using-python.html

https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/

We agree, however, that the Figure 4(b) for PC3 and PC4 does not represent the data very well and does not give much further insight, we therefore deleted it.

Theoretically, in PCA analysis, all the variables are independent. Then, we avoid designing new variables from already considered variables because it reduces the degree of freedom of the system and defines an ill-posed problem (as for inversion). Please, address this point in the revised version of the manuscript.

Response: To our knowledge, there is no requirement that variables included in a PCA must be independent. We used this representation for displaying the site clusters and to describe their characteristics, not for further statistical analysis. We believe this representation is applicable for this purpose.

L327/28: Caution: you have more mountainous sites in the western part of the studied area. Thus altitude and longitude are correlated, which biases the analysis.

Response: We rephrased this statement in Lines 407-410:

"Precipitation and recharge rates were higher in the West than in the East, following both the longitudinal gradient in altitude and the climatic influence of the wet oceanic climate in the West, with high precipitation and recharge rates even at lower altitudes (Lauterach, Elsbethen), versus the dry continental climate in the East."

L333-335: Awkward. Correlation does not give information about the function that relates the variables. Please, elaborate.

Response: We deleted this half-sentence in Line 414:

"The fraction of potential groundwater recharge to precipitation (GWR/P) was strongly correlated with the amount of precipitation (r = 0.91). Similarly, Barron et al. (2012) found […]"

L363: See below my inputs in the annexes. Some events were missed (even if roughly the modeling did well).

Response: We included the discussion on missed events by the model in Lines 295-297:

"Some events were missed by the model: at Lauterach and Elsbethen, the drying of the lower soil layer in summer was underestimated; at Gschlössboden, the peak in soil water content in the early calibration period was missed for both layers."

L380-382: Please, add, afterward, a few words on perspectives of improvement, like the combination with data at lower depths (GW monitoring) and the consideration of additional physical processes (preferential flows and water repellence, etc.).

Response: We added perspectives of improvement in Lines 461-465:

"The approach could be improved by including information on the deeper vadose zone to obtain more insight on temporal variation and seasonality of actual recharge, and to improve the model structure including lower boundary conditions. Especially at dry locations, using improved and additional measurements (e.g. of soil matric potential) could help reduce uncertainty in cumulative recharge estimation. Additionally, consideration of sites with varying slopes and the inclusion of surface runoff simulations in the analysis might improve representativeness for larger scale."

From the methods and results in this study, we cannot draw specific conclusions on which physical processes must be considered in the modeling, as this would require separate case studies for each individual monitoring site and is beyond the scope of this work.

In the Appendix, we improved the Figure descriptions including the 1:1 lines.