Biogeosciences Discuss., author comment AC3
https://doi.org/10.5194/bg-2021-278-AC3, 2022

Reply on RC3

Aaron Smith et al.

Author comment on "Modelling temporal variability of in situ soil water and vegetation isotopes reveals ecohydrological couplings in a riparian willow plot" by Aaron Smith et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-278-AC3, 2022

The authors thank the reviewer for their constructive comments which will be incorporated into the manuscript during revision. Using the reviewers suggestions, the authors will revise the results section to better describe and present the uncertainty of both simulated and measured datasets. The authors will expand the study site description to better encompass necessary information pertaining to data measurement (and uncertainty). Additional clarification will be added to the description of the EcH2O model parameterisation and calibration. Lastly, the presentation of model results using goodness-of-fit measures from calibration will be added to the results section to help with the justification of the results and discussion sections.

Specific Comments: Reviewer 2

R2C1: l.75: I would leave out importantly. It is important, but doesn't need this explicitly here

Response to R2C1: The authors will remove this during revision.

R2C2: Fig. 1: Figure caption is incomplete, in particular d) what are the blue and red bars? What is the grey box?

Response to R2C2: The authors will add further description to Fig 1d. The colored bars correspond to the colors in Fig 1c (location of Site A, B, and AWS).

R2C3: l.144: Sensors were installed until 1 m soil depth. Is that the maximum rooting depth for both willow trees and grass? This is crucial for root water uptake depth determination

Response to R2C3: The maximum rooting depth of each tree was not directly measured. Further measurements of groundwater (~2.2m) were also taken, but it was determined that vegetation source water was not taken from groundwater (Landgraf et al., 2021) and dominant root uptake depth from >50cm.

R2C4: l. 145-160: Even though I understand the method is described in Landgraf 2021, the information on how isotope standards were prepared and measured would be good here. Also, referencing the borehole method because of the short description herein should be considered.

Response to R2C4: The authors will add further descriptions for measurement preparation, standards and methods to the methods section.

R2C5: l. 178-180: and chapt. 3.2.2: how were these parameters determined/calibrated?

Response to R2C5: The authors will add the parameter ranges of the vegetation parameters to the supplementary material. The calibration was conducted as described in section 3.4.2.

R2C6: L.214: the last part of the sentence is unclear, please rephrase and clarify

Response to R2C6: The authors will revise this sentence.

R2C7: L.216: calibration? How was it calibrated?

Response to R2C7: The calibration procedure is in section 3.4, this section is intended only to provide a background of the model and calculation methods. The authors will clarify this section by removing "calibrated" from the section.

R2C8: L.216-223: this approach is interesting, was this used somewhere before? (citation?). It appears like such an approach would completely neglect preferential flow, am I correct? If yes, this should be stated somewhere ('does not account for pref. flow')

Response to R2C8: The approach was used in Smith et al. (2020), the authors will add this reference to the section. This approach is dependent on the structure of the model applied (the approach is not specific to EcH2O-iso). Since EcH2O-iso does not account for preferential flow, the results in this study will additionally not account for preferential flow. The authors will add a statement that preferential flow is not considered for this study.

R2C9: L.229: assumed root distributions…this is a BIG assumption. How were they assumed?

Response to R2C9: The assumption of root distribution here is that the root distributions follow an exponential distribution, which is consistent with empirical observations when the instrumentation was installed. Parameterisation of the exponential distribution is calibrated. The authors will clarify the assumptions made on the rooting distributions.

R2C10: 3.3.1.: How were the root parameters determined/approximated?

Response to R2C10: As with **Response to R2C9** the root parameters are calibrated using transpiration (and sapflux) and isotopic measurements. Description of the calibration of these parameters will be described in more detail in the model calibration section (3.4.2).

R2C11: 3.3.2.: For someone who does not model every day, the explanation on root length determination should be clearer. Coming from the field side of things, I wonder 'how is maximum rooting depth implemented?'; which measured parameters does one actually need (precipitation and sap flow?). I also wonder, if the general root distribution in the model always has the same shape? This is a large simplification that is definitely not true for any given vegetation species. How does it look like if we have a deep-rooter, for instance?

How was the fact handled that there very likely were willow roots present underneath the grass, affecting soil water contents and hence, the modeling efforts?

Response to R2C11: The authors will clarify the parameterisation of the root length parameters in section 3.3.2. In terms of model set-up (and running), all necessary data (forcing data) are presented in Table 2. In terms of measuring additional variables (or parameters), this is dependent on individual study sites, study objectives, and the sensitivity of the model to the output variable. In terms of rooting depth, the maximum rooting depth is the total soil depth. However, parameterisation of rooting distribution may constrain the roots to be shallower. The rooting distribution always follows the same shape (see Kuppel et al. 2018 for further details) and as the reviewer has suggested is not suitable for all vegetation. The model here was adjusted to allow for rooting to occur from outside of each model cell. In this way, willow roots could access water below the grass. This will be clarified in revision.

R2C12: L.277: this is an interesting point, but it should be noted that there is not only an error in simulating, but also measuring soil water isotopes. I am not saying that it should be, but is there a way to include this in such simulations?

Response to R2C12: While outside of the scope for this manuscript, there are methods to account for measurement uncertainty of both forcing and calibration data within model results. This is generally evaluated externally to the model (e.g. GLUE) and included within the uncertainty bounds.

R2C13: L.288-291: Maximum rooting depth is constrained to 100 cm. This needs to be proven/backed up. Stating another paper under review/discussion (here and in many other instances) is sort of cheating, to me. Root water uptake depths shift over a year and it cannot be assumed for the time of experiment (~3 months) that 100 cm max. rooting depth are a given. Please clarify this; I do believe the authors and a quick search tells me that willow trees are generally shallow rooted. However, another citation would help.

Response to R2C13: We are disappointed with the accusation of "cheating". Simply for issues of manuscript length in this modelling-focused paper we referred to the openly available HESS-D paper by Landgraf et al. (2021) for measurement details. Nevertheless, the authors will add further empirical justification and explanation from the maximum rooting depth used.

R2C14:L.305: please explain thoroughly, why 18O was not used in calibration

Response to R2C14: The authors will elaborate on why 18O was not used in calibration. Initial testing of model results did not reveal notably advantages to utilizing $\delta^{18}O$ within the multicriteria calibration with relative differences of simulated to measured δ2H and simulated to measured δ18O showing very similar responses.

R2C15: L.306: What is meant with 'the values for 18O were not greatly different from 2H'? First off, these values are usually very different. Second, the dual-isotope space provides an excellent way of validating the effect of kinetic fractionation. Third, I feel like a comparison of measured and modelled values in dual-isotope space would greatly benefit the trust in the model, apart from the statistical parameters.

Response to R2C15: The authors were referring to the trends of 18O and 2H showing very similar responses rather than the absolute values (**Response to R2C14**). The dual-isotope space for comparison of measured to simulated isotopic data would potentially only reveal some under-enriched shallow soil water below the Willow (as already shown in Fig. 3) where soil evaporation was limited in the model by water availability. As shallow soil isotopes were only one component of the multicriteria calibration, further plotting of additional isotopic variables would not likely reveal more than information already presented within the manuscript.

R2C16: Table 2: Calibration data: Why is only sap flow of 1 tree used? Likewise, Surface Temp and latent heat only from site B? This seems subjectively chosen and is not explained in the text.

Response to R2C16: The authors will revise the text to better indicate why each data were used. Surface temperature and latent heat were measured directed above the grass site (Site B) with the AWS. Sapflow was an average of the sapflow of both willow sapflow (range of sapflow data will be presented in the results during revision) and as both willows experienced the same conditions calibrating both trees to the same sapflow was not deemed necessary.

R2C17: L.324: …starting from likely, it belongs to discussion

Response to R2C17: The authors will move this to the discussion.

R2C18 Results: the subjective phrase like 'adequate' or 'slightly different' should be backed up by some objective measures in the results section.

Response to R2C18: The authors will revise the results section to quantify the descriptors.

R2C19 L.335-338: Just to clarify: The heating cables were not put inside the soil profile, or were they? I am asking this because we did this mistake once in my group and it turned out the cables heated the surrounding soil, hence, producing a heating of the area around the soil gas probes and tdr probes. As a result, one would calibrate data on a totally non-representative dataset that is highly influenced by the heating cables and not representative for the stand.

Response to R2C20: The heated cables were not installed within the soil profile, but were installed from the installed membranes to the soil surface. In this way, the soil heat profile was not impacted by the heated cables.

R2C20: Figure 3: This looks nice indeed, in particular for Site B! However, I repeat my statement from before that the dual-isotope space allows for a more precise evaluation of model performance and further interpretations such as root water uptake depth or kinetic fractionation. Another thing: There is definitely an uncertainty in the in situ isotope measurements, which is almost never incorporated into modeling. However, modeling always incorporates uncertainty in calibration results. I find this odd and not necessarily correct.

Response to R2C20: The authors thank the reviewer for their positive feedback. The authors hold the opinion that with the number of data points presented and the large overlap, differences, particularly temporal, between the simulated and measured may not be as notable. The authors will add the uncertainty bounds to the isotopic measurements during revision.

R2C21: The complete section 4.1 does not make use of any goodness-of-fit criteria and uses subjective and biased statements throughout. For instance, the calibrated sap flow data is judged as 'adequately captured by the model'. If I look at Fig.4 I (subjectively) see that the dynamics are OK (Site A) while the magnitudes are sometimes. For site B, there are no measured values for sap flow. This is not convincing to me. I strongly recommend adding goodness-of-fit criteria here.

Response to R2C21: As with the suggestion by the reviewer in **R2C18**, the authors will add the goodness-of-fit criteria to the results section to better justify the fit of the model.

R2C22 L.343: 'quite' well…objective measure?

Response to R2C22: The authors will add the goodness-of-fit criteria.

R2C23 L.396: simulated day-to-day variability could not reproduce the measured values

Response to R2C23: The authors will revise this statement in revision.

R2C24: 4.3: I find this section well-written and less subjective/biased. The general dynamics are met, but it needs to be said that an offset of 10 in d2H is already a large deviation (in isotope space). Now is that because of a non-perfect model fit or, and I am sure that it also plays a role, uncertainty in the in situ measurements. I feel like including some statements/metrics in regard to the measurement part of the second paper submitted by the authors could benefit the interpretation here. I find the aspect of the time-steps quite interesting: Why temporal resolution do we actually need? In isotope-space, daily is already a great resolution.

Response to R2C24: The authors agree that an offset of 10 ‰ can be quite large even for δ^2H, which here is due to multiple factors as the reviewer has suggested. There is of course uncertainty in the isotopic measurement. The authors will add some statements regarding the measurement uncertainty in the revision. The discussion of model performance is already in the discussion.

R2C25: L.479/480: 'with only minor under-estimation of the transpiration in the willows toward the end of the growing season'…I do not agree that the deviation is minor (>50%) nor that the fit is great for the rest of the period. The dynamics fit, but the magnitudes often do not. And at site B, no comparison is provided.

Response to R2C25: The authors will clarify this in the revision. Given the uncertainty of the range of sapflow measurements (will be added to Fig 4) the under-estimation is quite minor and falls within the measurement ranges. The use of KGE rather than NSE emphasised dynamics (mean and variability) over the absolute value of individual events. Here, the absolute magnitudes were strongly dependent on the soil moisture conditions below the willows. No comments on the sapflow at Site B can be made because there were no measurements of sapflow in the grass.