Anonymous Referee #1

Referee comment on "Coupled modelling of hydrological processes and grassland production in two contrasting climates" by Nicholas Jarvis et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-316-RC1, 2021

This manuscript describes an interesting modelling study of soil water balance components in lysimeters with grassland production, including the simulation of cutting/grazing, performed in two contrasting climate zones in Germany. Six lysimeters were collected in a wetter region (Rollesbroich), three of them were transported to a drier zone (Selhausen) to allow studying the same soil under different climate conditions. This practice, here called “space-for-time substitution”, could be helpful to mimic future climate conditions.

For the modelling, a Richards equation-based hydrological model is developed, using Van Genuchten-Mualem soil hydraulic parameters. Special attention is given to root growth and distribution, and a process-based function is used to predict root water uptake as a function of depth, and actual transpiration rates. Potential grassland growth (dry matter accumulation) is modelled using the concept of RUE. Actual growth rates are calculated from potential rates by reducing for water and temperature stresses. The effect of water stress on dry matter accumulation is included by assuming a linear proportionality between relative transpiration (predicted by process-based root water uptake modelling) and relative growth. Temperature stress is added by establishing a piecewise linear function with zones delimited by specific threshold temperatures. Simplifications, especially with respect to root growth and distribution, are unavoidable in this kind of modelling approach and are well presented and justified. The manuscript is generally well written. There are some issues to be addressed by the authors, among others referring to the soil hydraulic parameterization and clarity about the lysimeter soil contents. See my specific comments below.

We would like to thank the referee for the many constructive comments on our manuscript. We agree with nearly all of these suggestions (see below) and we are sure that this will significantly improve the paper.

Besides these, in my opinion, *the most important shortcoming (making me suggest a major revision) of the manuscript refers to the calibration and validation procedure*. You used the GLUE method to identify the best parameterizations using 6 years of lysimeter observations. You then discuss the model performance based on the 30 best parameter combinations selected by some criteria out of 2000 original combinations (parameter realizations). Posterior results seem fairly good, in terms of soil water content and ET (Figs 4 and 5), as well as model efficiencies (Table 6). But isn´t that to be expected when selecting the 30 best performing parameter sets? I would challenge you to follow a more rigorous calibration-validation protocol, performing the GLUE method on three or four years of your data, selecting the best parameter combinations, and then testing them on the remaining two or three years. This would reveal an unbiased and much more convincing model performance.

We realize now that our objective with the modelling was not very well explained at the end of the introduction at lines 118-126, as pointed out by the referee in a later comment. We will modify this part of the text to make the objective of the modelling exercise much clearer.

The main aim of the modelling exercise in our paper was actually not really to try to "validate" a model, but rather to use the model to investigate likely or plausible reasons for the differences in hydrological and plant response at the two sites. As the referee suggests, the (mostly) acceptable model efficiencies give us confidence that our model is good enough for this purpose. To answer the question posed by the referee: no, we doubt that we would get good results from taking the best 30 simulations of 2000 parameterizations of any given model. We would not expect good results if the model was inadequate. To give an example: we have also run calibrations for a model which is identical in all respects to the one described in our paper, except that it does not account for compensatory root water uptake (the Feddes root water uptake model is used instead). This model performs badly, with large negative model efficiencies for soil water contents and water balance terms, even for the best 30 simulations. This contrasting outcome emphasizes that model "validation" is really only a worthwhile exercise in a relative sense i.e. comparing the performance of different models.

From a practical point of view, the length of our time series (6 years) is anyway much too short in relation to the variations in weather from year to year to make it worthwhile to carry out a split validation. If we applied GLUE on the first two or three years of data only, we would not be including years with very low precipitation. This would mean that parameters related to drought response of the grassland would not be very well identified, due to their lack of sensitivity. In other words, we would get wider bounds for the acceptable parameters, if we applied the model only to the first 2 or 3 years. Many of these acceptable parameter sets would then be considered unacceptable in the remaining years, since they would fail to predict correct responses in the very dry years that were not represented in the first period. A very small proportion of parameter sets would probably be acceptable for both periods. But to get a sufficient number of these acceptable parameterizations for both periods would require very many more simulations (which would just not be practicable). In summary, we can use the model to learn from the data in the most efficient way by applying the GLUE procedure for the entire 6-year period. In fact, split validation would almost certainly not be a good approach even if our time series had been much longer than 6 years. The inefficient and potentially highly misleading nature of the split-sample validation procedure has been demonstrated, for example, by Arsenault et al. (2018, Journal of Hydrology, 566, 346-362) who used a time series of 16 years and came to the conclusion that split validation is …

… futile at best, detrimental at worst, and deceiving in all cases …

The conclusion section contains some discussion but does not mention the interesting fact that plants were able to mitigate water stress in the drier climate by enhancing below ground (root) growth. If this is confirmed, soil organic matter contents could be expected to rise when climate change triggers more water stress. Also, in agricultural (grain) crops, above ground yields might then be expected to diminish to favour belowground development. This is important in several ways, one of them being the expected yield of agricultural crops.

Yes, we certainly agree that this is an interesting and important result, but we did in fact already mention this in the conclusions (at lines 699-700). Related to the last part of this comment, it is worth mentioning that another study with the SoilCan lysimeters containing arable crops has shown the opposite effect i.e. that in the drier climate, water use efficiency increased, with larger grain yield per unit of water consumed. We suppose that putting more resources into seed production is probably a useful evolutionary adaptation for annual crops, whereas it seems that perennials try to survive droughts by enhancing root growth.

Scientific questions/issues:

l118-126 This final part of the introduction appears to refer to Materials and methods. I suggest finalizing the introduction with a concluding statement (possibly containing the objectives), and start the M&M section with the content of these lines. Related to this, please state more clearly that the six lysimeters contain soil from Rollesbroich, but that three of them were transferred to Selhausen. l162 somehow contains this information, but it should be stated more clearly. The sentence in l179-181 is clear with this respect but should be located at an earlier position in the text.

Yes, we agree, and we will modify the text accordingly

l194-196 This net upward flow would imply a depletion of the groundwater body in the long term? I.e., if plants invest more in the belowground biomass to avoid drought stress, large-scale hydrological cycles would be affected? A comment in this respect seems appropriate here.

Yes, actually, the Selhausen site is located on a low-lying relatively flat flood plain of the river Rhine (lines 130-131) and so it seems probable that this net upward flow is sustained by groundwater flow from the surrounding hills. In other words, the site lies in a discharge area (rather than a recharge area). We agree that we should add a sentence or two to explain this.

l210-213 Here you suggest that yield may have been affected by nutrient deficiencies. The performed simulations assume optimum plant nutrition (l356). Shortly discuss if this would affect the fairness of assessing model performance based on experimental data including nutrient issues.

Yes, it would be more accurate to write that the model does not account for variations in plant nutritional status. We will modify the text at line 356 accordingly.

l284-294 No details are given about how you (numerically) solved the Richards equation (13). Was an implicit algorithm scheme used? How about the discretization? Time and space steps? Averaging K and h? Some more details would be important.

Yes, we agree. We will add these details.

l429 Regarding "the fraction of the total root length that is effective for water uptake" (epsilon), you use a fixed value of 0.05 (after Faria et al., 2010). A high model sensitivity to this parameter is to be expected, and it is also plausible to assume that epsilon will increase in periods of more intense root growth and decrease when the root system is shrinking due to stress. Please discuss shortly the effect of assuming a fixed static value for this parameter.

Yes, our sensitivity analysis showed that the effective root fraction is indeed a moderately sensitive parameter (see table S2). However, for practical reasons, we could not include a large number of parameters in the uncertainty analysis. We selected four important parameters to calibrate and the reasons for this choice are already discussed at lines 518-533. One reason for not selecting the effective root fraction is that it would likely be correlated with other sensitive plant parameters, in particular with $f_{bg\ (opt)}$ and $\psi_{o(crit)}$. This is described at line 533.

l481-495 There are a lot of assumptions, guesses, and uncertainties in the soil hydraulic parameterization described here.
Perhaps, but we believe that we have more data on which to base our parameterization than is usually the case in modelling studies of this kind.

It is especially peculiar that the saturated water content of the surface layer is much higher in the lysimeters at Rollesbroich, as “estimated from the data by eye”. Assuming all lysimeters contain a similar soil monolith (collected at Selhausen), can this be made plausible?
Yes, this is interesting. This could be the result of chance spatial variation. But it is also possible that the physical properties of the uppermost layers of the Rollesbroich soil have changed following the move of the lysimeters to Selhausen. One plausible explanation is that the drier soil conditions at Selhausen have led to increased mineralization rates of soil organic matter, leading to a decline in soil organic matter content and consequently, an increase in bulk density (i.e. a loss of porosity). It is also possible that the drier soil conditions at Selhausen have induced water repellency, which reduces soil wettability.

We will add some text on this in the revised version.

In l468-469, the reader is informed that a common parameterization for the soil hydraulic properties in all six lysimeters will be assumed. But apparently, this does not include ThetaS?
Yes, this is true. It is only common among the three replicates at each site. We will change this statement accordingly.

Please improve the description/justification of the soil hydraulic parameterization.
Yes, we will add some further explanation and modify existing statements appropriately

l582 “No differences between the two sites were found for two of the parameters” – this is an interesting result and merits more attention. I think it is corroborated by common sense. Both parameters (the radiation extinction coefficient and the parameter controlling dry matter allocation and leaf loss as a function of water stress) are expected to be mostly genotypical and not affected by environmental conditions. On the other hand, unstressed stomatal conductance (discussed in l591-602) is expected to be a function of sink (atmosphere) conditions, especially VPD, temperature, radiation; rooting depth will probably be affected by soil moisture and temperature distribution in the soil profile, both a function of weather conditions.
These are very interesting observations. We will add some text along these lines.

l627 Here one of the reasons for the high E values at Rollesbroich is assumed to be “the capillary nature of the soil”. This is a too vague description. Soils are very similar at both locations. I suggest you remove this reason.

Yes, we agree. We will delete this.

l700-705 ("A major … growth") These statements are no conclusions from the presented research, they would fit better in the Discussion than in the Conclusion section.
We don't agree. These are conclusions that we can draw from the results of our study.

Some other minor "technical corrections":

l62 SVAT stands for "Soil-Vegetation-Atmosphere Transfer". Please add "Transfer".'
Yes, OK, we will
l79 Johnson et al. (2008) would fit well in this list of citations (doi:10.1071/EA07133)
Yes, we can cite this paper
l121 Delete "In this study".
Yes, O.K.
l235-240 For clarity, please add the units (dimensions) of the parameters of eqs. 1-7 in these lines.
OK, yes, we will do so
l270 To be more precise, one would need pressures and conductances, both unavailable in bucket models.
Yes, true, we will add this
l288 (here and on other occasions) The unit "day" is officially abbreviated as d. I suggest using d instead of day, days throughout the text.
OK, we will do so
l306 I would prefer to say that tau (just like alpha and n) is a (fitting) shape parameter. The way you wrote it here, it seems tau could be independently measured.
We don't think that what we wrote here implies anything about the way these parameters can be estimated. But we will try to write this in an even more neutral way.
l700 after "below-ground biomass" you might add: "thus mitigating drought stress".
Yes, we can do this
Figs. 1 and 2, X-axis label and Figs. 6, 7, 11, and 12, Y-axis: replace yr-1 by y-1
We would prefer not to (but we will do so if the editor insists)
Fig. 4 Symbols in this figure can hardly be distinguished, I suggest using colours for the three lysimeters. It would also be good to clearly identify both columns of figures by Selhausen and Rollesbroich (they are now identified only by the codes of the respective lysimeters).
Yes, we will add the site names to the columns of figures. But adding colours to the symbols does not help. There is so much data (daily time resolution) and the replicates are so close to one another, that the symbols will be indistinguishable regardless of what we do.
Figs. 4, 5, 8, 9, and 10: X-axis label is unclear. I would prefer "Time from onset of simulation (01-Jan-2013) [d]"
Yes, we agree that it is not 100% clear at present. But rather than change the x-axis label, we prefer to add some clarifying information in the caption i.e. Day 1 = 1st January 2013
Figure 7: Interpretation would be easier if Tp and Ta (and Ep and Ea) would appear in the same figure, side by side per location. In the current version of the figure, it is difficult to detect any difference between Tp and Ta.
Yes, we can do this. But actually, the differences between Ta and Tp are very small, as we discuss in the text in the following paragraph in relation to figure 8.