**General comments:**
The manuscript is actually in my opinion pretty good, it gives a very good number of details and the study and methods are described in depths. Some details are missing but overall, it is extensively described.
Significance is also very good, this is a rather important improvement of a model that is used quite a lot.
I do have some concerns though.

[partially required, could be discussed]
My main concern with the comparison between former Yasso07 version and yours is that yours was calibrated, the others if I understand well no. Ok, you calibrated only the scaling function for xi, but still the previous functions were calibrated on different data and might have hit another optimum on this particular dataset, and like this it becomes difficult to understand if the improvement in fitness is because of the structural improvement or because of the calibration. This might impact your Fig. 6 heavily.
I don't consider this a major flaw of the manuscript, since you are anyway declaring properly your methods and the reader can judge, but I would want to elaborate a bit in the discussion about this possible risk, giving some caution to the reader in interpreting your results.
Your results are reasonable. I don't see a reason why a monotonic moisture function could not be much worse than a non-monotonic (more specifically bitonic, even if does sound a bit cacophonic, I agree) one so I really believe the results, it contains important and much needed improvements for a broadly adopted model. But your comparison is at least quantitatively flawed if you affirm your structure was superior, since you cannot tell apart the effect of the structural change and the one of the calibrations. I advise caution here, your structure is superior also according to me but based mainly on inductive reasoning.

[required]
There is some inconsistency with how you refer to figure, sometimes Figure sometimes Fig. (in the text)

[required]
Density of the data: it is not immediate to understand exactly what the time series considered are, I mean how m any points over time each time series considered has. Are they same density or not (I guess not)? How are they spaced, evenly or not? When were the points collected, at what intervals? It is somehow possible to figure this out, but it is a crucial detail for understanding the calibration (the posterior likelihoods from your two objectives might have very different shapes if you compare a sparse time series with a very dense one, the sparse will have many peaks. IUt might also contribute to explain the discrepancies between the calibrations) and I think it should be a detail that stands out clearly in M&M.

[not required, just a suggestion]
I am honestly surprised to see how few models are utilizing a non-monotonic moisture reduction already, it's a decade-old discussion now and seems quite solved. Can you discuss specifically this topic more explicitly in the intro? Like which are the models which already updated the moisture reduction to non-monotonic? Are there some? A bit of state-of-the-art (like 2-3 lines, not more, classifying which models use monotonic and which bitonic if there are, and if there aren't then you can very rightfully claim a very big leap forward in terms of SOC model applicability).

[partially required, at least articulated answer appreciated]
When it comes to the optimum of your calibrated moisture scaling function, my guess is that it is different from other studies because of depth issues. You do not consider subsoil in your model, so you are working with assuming some mean water content, while water content will vary wildly in the profile. Even assuming the same depth of the water table, an organic soil will likely have a different depth/SWC curve than a mineral one, so it will regress to the mean with a different cumulative function. This issue could be discussed more (or other issues you believe could explain that discrepancy if you see others).
In your previous answer to the referees, you state that most respiration comes from the upper layer… I don't think this is necessarily true in an organic soil. In mineral soils this is true because most SOC is there, but an organic soil has a different SOC distribution, and when the water table gets lower than 30 cm you will have respiration also from those layers. I might of course be wrong but I would want to be shown wrong, in that case. Why do you think an organic soil with a low water table would have most of its respiration on 0-30? And how much is it "most", 60%? 95%?
It would be interesting to see the model residuals of your three calibrations (all of them, not just mean) plotted against SOC content, I personally expect them to follow some kind of pattern. I am not requiring this for the manuscript, but it would be something nice to see.

[not required, just a suggestion]
What is the implication on SOC stocks predictions of your three calibrations?
More extensively: it seems from your calibrations that SOC and CO2 time series clearly contains information about different processes. For example, the CO2 could contain information about hysteretic phenomena which are not all captured in the model, hence the two sources do not reconcile fully, as you already discuss.
In terms of applications, depending on the scope of the modeling I would choose one or the other unless the discrepancies are solved. If my aim is to simulate SOC stocks, I should be better off with the SOC only calibration, while if my aim is to simulate both I would accept a likely reduction in SOC fitness to get a better CO2 representation.
It is possible to operate these choices based on table 2 anyway so this is not a required change, just making you aware I would reason this way if I had to apply the model.
A suggestion: plotting the posterior likelihoods of the two calibration objectives might help you understand more. Are they skewed, for example?

[partially required, advise caution to the readers]
Your conclusions are maybe too aggressive. You find a rather different optimum compared to many other studies, 14% to 27% seems quite low, and those studies were based on many data from lab (and also field I guess). There is something weird there, might be some missing processes (which I guess is missing depth), it would be dangerous to extrapolate before understanding what it is. If for example water table is involved, you risk doing wrong extrapolations when you change hydrology radically. Climate change extrapolations might not work so well if they change the hydrology of the sites (if hydrology was involved in the discrepancies between your results and the literature you cite).
If you want to extrapolate such conclusions, I think you would need to discuss a bit more the discrepancies speculating some mechanisms, to then justify that the extrapolation is possible.
I mean, it is possible that what you affirm is true, but I would use some words of caution too.

**Specific comments:**

Line 40: Unimodal. I am not sure about this definition. It is true that an exponential or linear such as what was in Yasso before is not strictly unimodal, since it is strictly increasing and does not have any

distinct peak, so your definition might work. But this definition can be more relaxed, meaning that a function does not have multiple modes, so also the “old” function could be seen as unimodal. I would have personally referred to this concept as non-monotonic (or even better you can use, I think, the term “bitonic”), as opposed to the former monotonic function.
I mean, if you call your non-monotonic function “unimodal”, how do you call then the function previously used? You propose a unimodal function instead of what? Could you define the two functions in a same phrase, this instead of that?
Just a suggestion.

Line 73: What do you mean with “a functional form reaching saturation”? That is monotonic, as in the opposite of (unimodal && multimodal)? Also your proposed function reaches saturation, it saturates at the optimum.

Line 78: modify “all kind” with something like “various”, “a lot of” or similar, such absolute does not work in a scientific context (I get what you mean, though)

Line 84-86: Both statements are true but seem unrelated. One thing is that even if you calibrated a non-monotonic function like Moyano on mineral soils the same calibration won’t probably work on organic soils, another is the fact that a monotonic function cannot represent the anoxic limitation process. It is hard to read and to get what you mean like this.

Line 95-96: if you describe this, then how was the functions improved in this study? Just scaling, or non-monotonic (with oxygen limitation processes)?

Line 102: … I wouldn’t call Yasso particularly “parsimonious” in its class, compared with Century or RothC I mean, they are quite similar in terms of complexity, no? One could say Q is parsimonious, but Yasso should not have at all less parameters than RothC, right? It is a rather simple model class, though, I agree with that. No need to modify this for me, just be aware of how I read it.

Line 105: With SWC are you talking about the whole profile, total mm/m^2 kind of, or the gravimetric/volumetric water content (like g/g^1 or percent of pore space)? I think you should define SWC here to avoid ambiguity, there are many possible way to express it.

Line 108: What does “global” mean in this context? Meaning that the parameter values are considered constant everywhere? I ask because in some context you might be referring to a parameter space for example, as in “local and global optima”.

Line 132-153: it is hard to understand what is the time resolution of these time series. How often did you measure, for each variable?

Line 153-168: Same. Was this one flux measurement each campaign, or more often?

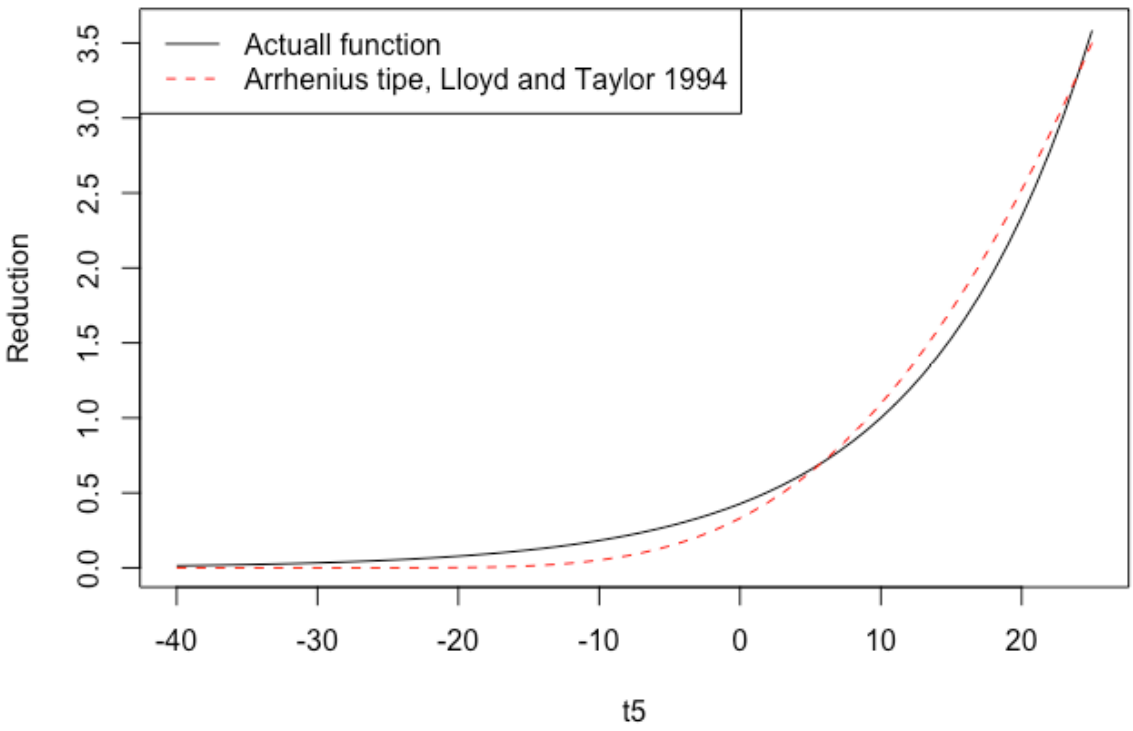
Line197-199: Wait, do you mean that the Yasso07._xi_TW is not calibrated? I see now better what global means here, you mean the optimum of that specific model from previous calibrations on other datasets?

Line 285: “(ter Braak and Vrugt, 2008)” it’s probably a typo

Line 374-375: I would say this phrase is redundant, nowadays Bayesian data assimilation approach has proven useless in countless applications. If you do not judge it redundant, why do you choose these specific studies over many others?

Line 387-390: Do you mean JULES has a constant reduction, not scaling with moisture?!? Just to be sure, I would have assumed these models were already much less rough on moisture reduction.

Line 475-490: I do not see any Arrhenius derived function. In Eq. 3 you have some kind of Q10 function. The Q10 function is quite rough, and it has indeed problems for very low and very high temperatures. See attached plot, where I plot only the lower end. You are using a function that is far from optimal at very low temperature, which in Finnish soils you are going to encounter often, so it's not surprising the model is not very good in those situations. Given the low respiration in those periods though the error should not be a very big issue, but I think you need to update your description if you did not use an Arrhenius or derived functions.




Line 503-504:; what do you mean that SOC stocks had the largest influence on moisture optimum? I guess you mean the opposite. Or do you mean that they were influencing the calibration the most? If so, from where do you derive this extrapolation?

Figure 2: there is probably something wrong in panel a), the smaller boxplots are not lining up. My guess is that you are not taking the right x points when you overlap the second plot (I guess you did this in Base R) as you seem to be doing in panel b)

Figure 2 caption: what kind of mean are you showing for SWC? Annual? Overall?

Figure 5: please describe more precisely the three dimensions you are showing here. What is on the Z axis? Is that the value of the resulting scaling xi_d?

In this case, which is how I interpret the plots, this plot is a bit redundant. It seems to show exactly the same data than Figure 4, since the two modifiers combine linearly, just in a slightly different way.