Comment on gmd-2021-98
Anonymous Referee #2

The manuscript by Shannon de Roos et al. documented an effort on assessing the performance of AquaCrop model simulation at regional scale with benchmark of remotely sensed and in-situ observations of biomass and soil moisture. Generally the current manuscript is lacking scientific interpretation and insights. I have the following concerns for the authors to consider:

First, I suggest the author further consolidate their objectives of their manuscript. It reads like that the authors are going to address the scaling issue from point to global models in the introduction part, but the results only stay at model evaluation at a fixed scale (i.e. 1km). If the objective is scaling, the claim that “the regional AquaCrop model proves to be useful in assessing crop production and soil moisture at various scales and could serve as a bridge between point-based and global models” is not well backed up by the analysis in the manuscript. There is only one scale for model simulation in the current manuscript, i.e. 1km scale. Most importantly, it is unclear how the regional model simulation can serve as a bridge between point-based and global models. The scaling issue from point-based to global models is not touched in this manuscript at all, but deserves a further investigation in the framework the authors developed. For example, when assessing the soil moisture, the authors aggregated the 1km soil moisture simulation to 9km. In other model setups like GGCMI, the model would be running at 9km or even larger scales. How do the performances of different model setups vary and what are the controlling factors for those performance variations? I think those are the key questions to be answered and would also be more interesting for the crop model and land surface modeling communities. The authors can actually test those questions for both biomass and soil moisture simulations with their regional model simulation platform.

If the objective is model evaluation, I suggest the authors rewrite their motivation part in the introduction. Model evaluation with newly remote sensing data is also interesting, especially in the context of further data assimilation (as indicated by the authors in the conclusion part) experiments in which we need to have some information about the model uncertainties.
Second, the authors make many simplifications in their model set up. For example, they set up a generic C3 crop in their simulation. However, this is not well justified. At least, I see a hot spot for corn production in their region. The authors may need to take into consideration of C4 crops or at least quantify the uncertainty of neglecting it (which is not reasonable). They also found that the soil moisture simulation performance is higher at areas with smaller AEI (indicating irrigation area fraction). However, the irrigation is not simulated in their set up. This raise a question: why do we care the performance of an unrealistic model set up? The performance evaluation is only valid when the modelers tried their best to mimic the reality. Otherwise, it is too arbitrary to say anything about the model performance when there is great uncertainty in both model simulations and satellite observations.

Third, it seems that the transpiration simulation in AquaCrop plays a very important role in simulating biomass and soil moisture. Why not do some assessment on transpiration simulation with flux tower and remote sensing ET data?

Forth, the authors directly jumped to conclusion after showing their results. Are there any insights to be discussed from this model evaluation effort? I suggest the authors bring up their most important findings and give more implications about crop model set up and evaluation at regional scale in the discussion part, which is now totally missing. Otherwise, the scientific merit of this manuscript is largely limited.

Other comments:

L80-81: please specify the soil layer depths you used in your regional setup. This is critical information when you want to compare your simulation with satellite-based soil moisture retrievals.

L81-L85: more description about hydrology (runoff, percolation, ...) in the model is required as evaluating soil moisture simulation performance is an important component of this manuscript.

Section 2.2: it would be good to have a flowchart for the regional setup.

L277-L278: how about also aggregating CGLS-SSM to 9km and compare it with model simulations at the same scale with SMAP data? That would be a more fair comparison.