

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2021-98-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-98

Christoph Müller (Referee)

Referee comment on "Performance analysis of regional AquaCrop (v6.1) biomass and surface soil moisture simulations using satellite and in situ observations" by Shannon de Roos et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-98-RC1>, 2021

Dr de Roos and colleagues present a Python wrapper to run the AquaCrop v6.1 model in parallel mode. They present results of a simulation experiment and compare these to reference data from remote sensing and some in-situ measurements.

The paper generally fits the scope of GMD, but I have several concerns that need to be addressed before this manuscript can become a valuable contribution to the scientific literature.

I believe that authors have a right to know their referees' names as a double-open system is superior to double-blind and certainly to half-blind review systems. My name is Christoph Müller of the Potsdam Institute for Climate Impact Research. I hope my comments are helpful to the authors and the journal.

General comments:

Upon request from the executive editors they have made available the source code of the wrapper and the input data, however the source code of the AquaCrop model is not published but only an executable file, not as the source code. This seems to violate the open access policies of GMD.

The manuscript lacks clarity in many cases (see detailed comments below) but also on the objective(s) of the paper. From the source code provided and the setup, it is meant to describe the parallel model framework and to evaluate model performance (not the parallel framework). However, the model performance is evaluated in the manner of individual points (despite that it's a quite large set of points), not in a manner that

addresses the scale and extent, e.g. by addressing the ability to reproduce spatial patterns, which would be a main asset of “regional” model applications compared to a set of field-scale applications.

If the objective is to present the framework that allows for parallel, high-resolution, large-scale applications, the technical skill and spatial properties of the simulation could have stronger emphasis in the evaluation, or how the large-scale setup (e.g. lack of calibration) compares to field-scale setup.

If (one of) the objective(s) is to generally evaluate AquaCrop against novel data (such as the data sets used here), the model description needs to be expanded, the current set of equations does not even address all processes discussed as relevant in the text (see comments below).

I don't understand the claim made that AquaCrop could serve as a bridge between point and global level simulations. It is claimed that AquaCrop was developed for a simplistic representation of crop growth (L46) and performance is good if the model is calibrated for local field conditions (L48). So AquaCrop may actually lack processes that are relevant to capture the heterogeneity of the landscape of environmental and management conditions at larger scales. The calibration of large-scale applications is hampered by lack of data that could serve as calibration targets and is not attempted here.

Even though an eyeball comparison suggests that results hold true, I find the comparison of the AquaCrop performance against the 2 soil wetness datasets a bit biased, as the samples are very different. This could be made more direct if also the statistics would be supplied for the set of pixels that is covered by both reference data sets.

Detailed comments:

L3: good insightS in – or – good understanding of

L19: curious to learn about how that bridge could look like. Many globally applied crop models are, in fact, field-scale models run in a modeling framework to process gridded data

L31: the better GGCMi evaluation reference would be Müller et al. 2017, not 2018. You could add Folberth et al. 2019 (<http://dx.doi.org/10.1371/journal.pone.0221862>)

L34: this “downside” is not only relevant for upscaling field-scale models but holds true for any large-scale crop model application.

L35: do you mean “... and loss of information that is typically available at smaller scales”?

L39: I guess the better reference for the GGCM Project per se is Elliott et al. 2015 (<http://dx.doi.org/10.5194/gmd-8-261-2015>) if you just want to have one reference. If you want to describe the breadth of crop models used in GGCM, you could add Elliott et al. (2015), Müller et al. 2019 (<http://dx.doi.org/10.1038/s41597-019-0023-8>), Franke et al. (2020a) (<http://dx.doi.org/10.5194/gmd-13-2315-2020>) for the different set of models contributing to Phase2 and (Jägermeyr et al. under review) for yet another ensemble of crop models contributing to the current Phase 3 (including AquaCrop).

L40: odd second half of that sentence: response to what? Do you mean “... more insight is needed in the relevancy of different processes represented in crop models applied at different spatial and temporal scales under different management assumptions”?

L52: unclear – what are the difficulties to update to newer AquaCrop versions?

L58: if this is the main objective, what are the side objectives?

L60: I’m surprised by the “generic crop” approach. Crops differ in various aspects and e.g. differences in growing season specifications have been shown to matter quite a lot (Müller et al. 2017, Jägermeyr & Frieler 2018 (<http://dx.doi.org/10.1126/sciadv.aat4517>)). Also above ground biomass differs substantially between crops. I see that this liberates you of having to specify high-resolution inputs that match crop distributions that the satellites see, but some justification of this choice would be needed here.

L79: I don’t understand the discretization of the soil column. It looks like no soil layers are used? But the root zone is divided into compartments? How are these discretized? Is the topsoil/subsoil information of HWSD (described later, L132) assigned to these compartments or are soil properties assumed to be homogeneous?

L101: can you elaborate on what a field and a grid cell really are in AquaCrop. Most gridded approaches (or in fact also field-scale approaches) simulate one single point and assume it is representative for the field/grid cell. Is that similar here or is there any lateral heterogeneity considered within the simulation units (field or grid cell)? If not, this “replacement” merely affects how outputs are interpreted, no?

L116: this section should include the time period that is actually simulated. Or a modeling protocol section should be added that describes the simulation experiment(s) – simulations done in advance for “tuning”, central simulation

L117: unclear why this input data set was selected and no bias correction was performed. See e.g. Ruane et al. 2021 (<http://dx.doi.org/10.1016/j.agrformet.2020.108313>) for the relevancy of input data sets. The AgMERRA data set (based on MERRA and available at 0.25° spatial resolution), would e.g. seem to be more suitable? Or other data sets described there or the ISIMIP3a input data sets that also cover 2012?

L136: but some crops are grown throughout winter in Europe, how are these growing season treated?

L140: is this some form of calibration? Can you better describe how the 30% were approximated? The GAEZ data would also provide some information on soil fertility that could help to abandon a guestimate.

L144: It’s a bit funny to justify the choice of the generic crop with anticipated results. Also, it seems that this is speculation as no counterfactual experiment was conducted. Maybe more specific crop representation would substantially improve model performance? There clearly is still room for improvement in model performance, so we cannot say if this generic crop choice is a good one.

L145: “this file was minimally tuned” this is unclear and needs a clearer description and justification. Keep in mind that the work should be reproducible.

L150: I don’t understand why the canopy development had to be prescribed in calendar days. How is it normally done in AquaCrop? A typical approach to simulate phenological development is based on growing-degree-days (see e.g. Wang et al. 2017 <http://dx.doi.org/10.1038/nplants.2017.102>). Please add a description of how crop phenology is modeled and if/how that relates to transpiration.

L151: the effect of GGD on transpiration is not clear from equation 2. How is this considered? Please elaborate to make sure that the model description is complete and understandable.

L156: what classes? And why 50 as a minimum? Are you saying “if at least half of the pixel is used for rainfed agriculture, it is included in the simulation set”? I wonder if that does not exclude a lot of cropland from the analysis? Is this necessary to have a clear signal from cropland when comparing to satellite data? Please expand.

L187: What is the outcome of that implicit quality screening?

L200: is there a name and reference for the "recommended conservative quality screening"? How important is that for the outcome?

L214: this information should be moved and expanded to the methods section. Modeling domain, selection of grid cells included in simulations, temporal range covered etc. need to be described in the methods part.

L216: all metrics should be described with equations. To me, at least, it is unclear how anomalies were computed or how the bias was removed – it seems there are different options to do so.

L228: multi-year => how many years?

L233: the median of the matching 10 days?

L236: I don't understand the exact processing or the intentions of the correlation analysis of AEI and the satellite products

L237: the metrics section fails to explain how 5cm soil moisture values are made comparable to simulated soil moisture of the rooting zone? Maybe this is not an issue and just a problem of me not understanding how the soil is discretized in AquaCrop (see comment above)

L257: where can we see this? How do you know?

L258: "tons" are abbreviated with small-caps t (also in legend of Fig 1 and titles of Fig 2)

L267: what is "high rate of rainfall"? Do you refer to the precipitation intensity or total amounts? That part of Germany included in your modeling domain and Poland are not exactly areas of high rainfall.

L268: why should the satellite not see water stress effects? Please elaborate

L280: is this a finding by eyeball? Please provide metrics that support your claim (e.g. distribution of errors per texture class or similar).

L285: My understanding of the methods section is that only pixels with at least 50% rainfed agriculture according to the CORINE data set are included in the simulations. How come you now claim that the sandy regions included in the analysis (and thus containing agricultural land) are not suitable for agriculture?

L292: just to avoid ambiguity: these R values refer to AquaCrop vs. in-situ and AquaCrop vs. CGLS-SSM, correct? Not to in-situ vs. CGLS-SSM in the second case?

L294: why 42 pixels? In section 3.4 you describe 32 for SMAP-SSM and 42+3 for CGLS-SSM. The extra 3 HOL points are not available for SMAP-SSM?

L295: is this the mean R of the temporal correlation averaged across the sites or a correlation across time and space?

L306: this seems to miss a reference? Or are these your own findings?

L315: I don't understand this sentence. Irrigation could also dampen the amplitude making the overall weather-driven signal (the only aspect captured by AquaCrop) smaller compared to the noise and thus harder to get any correlation at all. Also for a comparison between the different data sets it seems difficult to compare across different samples? Why not make these claims based on the same spatial mask?

L323: I did not see any analysis separating seasonal, interannual and short-term temporal dynamics here. What results support this claim here?

L330: Is this speculation or can this be shown somewhere? A simple test could be to run the model with uniform soil parameters – the computational costs are claimed to be low?

L338: again here seems to be a reference missing? The comparison of the skill of the AquaCrop model to the 2 different data products should be conducted on the same spatial mask

L341: This point was also made earlier on, but I still don't understand how the explicit focus on "agriculture-only" pixels can include substantial areas with soil parameters unfit for agricultural production?

L342: Speculation or finding? Is that a scientific debate or finding that could be referenced or would e.g. the SMAP-SSM quality flags suggest such relationship?

L345: There are already data on varying crop parameters such as growing seasons and fertilizer available and I guess the abandonment of the idea of a "generic crop" would be a prerequisite to introduce time- and space-variant crop parameters

L348: this is an interesting idea, but can you elaborate on this a bit to make it more tangible?

Fig 2: Panel b shows productivity not production? Small-caps t not T for tones.

Fig 3: What are white areas? In figs 1+2 I was assuming that these are pixels not simulated (could be explained in methods section), but is the SMAP-SSM data set more patchy or are these excluded for some other reason? This should be explained in the caption.

Fig 4: panels b and c don't show information on agreement across the 3 product classes (satellites, AquaCrop, in situ) or if there is always just 2 of the 3 agreeing with each other. That would have implications for the interpretation of results, wouldn't it?