

Geosci. Model Dev. Discuss., author comment AC2
<https://doi.org/10.5194/gmd-2021-98-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Shannon de Roos et al.

Author 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-AC2>, 2021

The authors would like to thank the reviewer, Christoph Müller, for his elaborative comments on our manuscript. We appreciate the taken time to review our work.

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.

Answer: The original source code is exclusively licensed by the Food and Agricultural Organization (FAO). The source code of the executable on our zenodo link is equal to the source code of the AquaCrop windows programme version 6.1 (<http://www.fao.org/aquacrop/software/aquacropstandardwindowsprogramme/en/>), but compiled for a Linux operating system. We would like to emphasize that only the executable is needed to run this spatial version of AquaCrop.

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).

Answer: The model is not only evaluated on individual points, but also regionally, using satellite data. To further investigate the spatial patterns of the simulations, we could look

at time series of (regional) spatial correlation values between regional AquaCrop simulations and satellite retrievals. However, such an analysis is risky by itself because satellite data are not to be taken as reference data in terms of absolute retrieval values, because they depend on local parameter estimates (that might not be anywhere close to the 'truth'). Similarly, the model setup with a generic crop is not meant to correctly estimate the 'absolute' values of biomass. Therefore, the relative time series analysis at all pixels is deemed more important: our crop modelling system is slated by state-of-the-art practices in land surface modelling and data assimilation, where relative variability is much more important (e.g. anomalies) than absolute values.

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.

Answer: We will edit the text as follows:

"The proposed regional AquaCrop system is scalable to any spatial resolution and therefore..."

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.

Answer: All our analyses and performance evaluations are based on a range of objective skill metrics, community-based standards and direct causal/physical relationships – no eyeball comparisons.

We are not entirely sure about the 'bias' in the performance analysis. However, based on suggestions below, we think that (i) there might have been some misunderstanding, which we will correct for in the text, and (ii) that a common spatial mask (crossmasking of datasets) is recommended. Such a crossmasking of datasets would have to be done in space and time, and would reduce our datasets to a very small overlapping sample, because each satellite dataset has very different recommended retrieval quality flags. Crossmasking would thus result in a great loss of information and a consequent bias in our performance analysis (limited to a small subsample). Please see details below.

Detailed comments:

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

Answer: Please see description above.

L31: the better GGCM 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>)

Answer: Thank you for the references.

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

Answer: Agreed, the text will be updated.

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

Answer: Thank you, the text will be updated.

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).

Answer: Thank you for the references, they will be included in the revised manuscript.

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”?

Answer: Thank you for the suggestion, the sentence will be rephrased.

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

Answer: This statement will be removed from the sentence.

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

Answer: This was an unfortunate formulation. The evaluation of high-resolution regional AquaCrop simulations is *the* objective of this paper.

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.

Answer: The text will be updated to make sure that the focus is on the correct relative temporal variability and anomalies.

As briefly mentioned at the end of the manuscript, this model is set up for satellite-based data assimilation at a later stage. Our motivation is that the data-assimilation will correct for the temporal differences in relative biomass production. We wanted to test if the model performs accurately with this generic crop, to see if we can continue with the data-assimilation. Of course, the crop file can easily be replaced by any specific crop for other applications.

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?

Answer: In AquaCrop, output for soil moisture is given for the entire root-zone, but also for the different layers of the root zone, for 10-12 compartments (depending on the root zone depth). The top layer (WC01) was used to evaluate against satellite data, which is more or less equal to the top 5-10cm of the soil. We will clarify this in the text.

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?

Answer: Indeed, each field is considered homogeneous as it would be in the original AquaCrop model. This imperfect mapping adds representativeness error and is mentioned as extra motivation to focus on skill metrics that do not include a bias component (R, anomR, ubRMSD).

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

Answer: A new section will be added to the methods, describing the simulation set-up.

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?

Answer: The text will be updated as follows:

"The MERRA-2 meteorological variables have a 3-hourly temporal resolution, a spatial resolution of 0.5° lat x 0.625° lon, and are readily available at a latency of about a month."

There are many global datasets available for the meteorology. MERRA-2 is a well-established and in depth evaluated long term global re-analysis product, assimilating observations from satellite and gauge stations for precipitation, with a low product latency (Reichle et al. 2017 doi: <https://doi.org/10.1175/JCLI-D-16-0720.1>). AgMERRA is based

on an older version of MERRA, using an outdated modelling system, artificially mapped to 0.25°, and only available until specific years (2010). MERRA-2 has a latency of at most a month, and if we wanted to, we could directly replace MERRA-2 re-analysis with real-time forecasts, which is beneficial for future simulations.

A comparison of MERRA-2 temperature and precipitation and data from field stations (in Austria and Czech Republic), showed satisfactory results. We are looking for reliable datasets at finer resolutions for future applications.

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

Answer: In the methods section we state that this 'generic crop' is suitable for the summer growing season. We therefore exclude analysis over the winter period (Nov-Feb). A different crop file would be needed to simulate the winter crops, with a different starting date of the year (November).

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 guesstimate.

Answer: Yes, this was tuned (manually) by comparing for several locations the biomass production to the CGLS-DMP product. This value was also chosen as in AquaCrop, the recommended soil fertility is in the range of moderate-good instead of perfect (=0% reduction). The reduction of 30% falls within that recommended range.

We will update this in the text as follows:

"..., the value was manually tuned to 30% after initial model evaluation of daily biomass production with the CGLS-DMP (see 3.1) product for several pixels."

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.

Answer: We agree that the use of specific crops would most likely result in better model performances, however this is not within the scope of our research, and even not needed for future data assimilation experiments, if designed to focus on relative temporal variabilities.

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.

Answer: The minimally tuning refers to extending the length of the senescence stage, which was done after the comparison with CGLS-DMP. We will specify this in the updated manuscript. Please note that the final generic crop file used in this research is available in the supplied dataset, and can be used by anyone.

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.

Answer: Indeed, growing degree days are the default option in AquaCrop. However, since we wanted to simulate the biomass production from the 1st to the last day of the year, we had to fix the length of the crop cycle on 365 days to avoid that in a cold year the length would exceed the 365 days of the year (and vice versa in a warm year). However, since in the calculation of crop transpiration growing degree days are considered, an over- or underestimation of the canopy cover will have only a small impact on the simulated crop transpiration (and hence also the effect on the simulated soil water content and biomass production will be limited).

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 in complete and understandable.

Answer: The cold stress coefficient is calculated with growing degree days, which largely eliminates the error of canopy cover in calendar days. This information will be added to the revised manuscript.

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.

Answer: Correct, we will rephrase this.

There would indeed be some crop areas lost. However, since a 1-km pixel should represent an agricultural field, for pixels in which more than 50% is covered by another type of vegetation or not even by vegetation but e.g. an urban class, this would no longer be a realistic representation of an agricultural field.

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

Answer: This sentence will be removed to avoid confusion.

This statement was too trivial: since we are not simulating over areas where satellite soil moisture retrievals do not perform well (e.g. ice, urban, complex terrain), we are also not using any compromised satellite retrievals.

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

Answer: There is; O'Neill, Peggy, et al. "Algorithm Theoretical Basis Document. Level 2 & 3 Soil Moisture (Passive) Data Products." (2018). We will add this reference to the updated manuscript. Microwave satellites are sensitive to various aspects (frozen soils, snow, steep slopes, urban area, water bodies). Quality control is done for both the CGLS-SSM and SMAP-SSM, to minimize the product bias.

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.

Answer: This will be updated accordingly in the revised manuscript.

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.

Answer: The equations described in the attached file will be added to the revised manuscript.

To calculate anomalies, please see Gruber et al., (2020) (<https://doi.org/10.1016/j.rse.2020.111806>) or also <https://github.com/alexgruber/myprojects/blob/master/timeseries.py>. The reference will be added to the updated manuscript.

The anomalies are calculated by first computing for each pixel the climatologies of both the model simulations and the reference dataset (which were crossmasked in time). Subsequently, the data values are subtracted by the climatology values for the matching dates, resulting in the anomaly for these specific dates. A positive value thus indicates a higher biomass/ SSM value compared to the climatological average for that day, whereas a negative value indicates the opposite. The correlation of the model simulated anomalies vs. the satellite product anomalies then gives an indication of the model performance on the interannual variability.

The manuscript will provide more details about the anomaly calculation.

L228: multi-year => how many years?

Answer: Depending on the length of the dataset: 8 years for biomass & CGLS-DMP; ~4 years CGLS-SSM; ~3.5 years SSM- SMAP. We will clarify this in the revised manuscript.

L233: the median of the matching 10 days?

Answer: Yes, exactly. Will be rephrased in the revised manuscript.

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

Answer: The AquaCrop model is run without any irrigation applications. In areas where irrigation is very common, such as northern Italy, this could cause a mismatch between observations from the satellite and model simulations. Since it is impossible to know the exact dates and amounts of irrigation at this scale for each location, the Area Equipped for Irrigation (AEI) dataset was used, to distinguish regions of high irrigation potential. By comparing correlation coefficients to the percentage of AEI, we could say that there is a possible effect of irrigation that influences the correlation between the model and satellite retrieval products. We will repeat in the text that the model is run without irrigation applications.

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)

Answer: Please see answer above, for L79.

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

Answer: This could be visualized with maps presenting water stress and temperature stress over the domain. You would typically find higher water stress in the South and cold temperature stress in the North. We hope that this is self-explaining.

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

Answer: we will make the units consistent.

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.

Answer: This sentence was not well phrased and will be updated in the revised manuscript. In areas where crop growth is expected as there is sufficient water supply by precipitation, this supply will be drained in extremely sandy soils, therefore still create water stress for the plant.

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

Answer: The DMP is not a direct satellite observation, but a product that uses fAPAR from optical satellites and variables from other resources (meteo data from ECMWF) to derive the productivity. It is true that the stress should be observed in fAPAR at some stage (it is a 10-daily product), but the DMP manual still emphasizes that this product contains several limitations, such as omitting water stress and nutrient deficiencies. This will be added in the revised manuscript.

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

Answer: The particular erroneous output stands out in the maps, and corresponds exactly to a certain soil type, with 93% of sand. This is a unique soil class from the HWSO, which is not suitable to be used in AquaCrop. A distribution per texture class would not add much value, and only highlight this one soil class. We will fine tune the language in the manuscript to clarify this.

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?

Answer: This is something we did not expect to see either. However, it is possible that the dominant soil class over an area is not the dominant soil class for the agricultural areas within that pixel. This soil class of 93% sand is simply an outlier in the soil classifications, i.e. in our soil input data set.

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?

Answer: Figure 4.a. shows AquaCrop vs. in situ. 4b AquaCrop vs. CGLS-SSM & AquaCrop vs. SMAP-SSM, 4c. in situ vs. CGLS-SSM & In situ vs. SMAP-SSM. We will change the description to make it clearer.

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?

Answer: Indeed, there is no SMAP-SSM data over the HOAL locations, so we could not make a comparison there. We will check the consistency in the number of data points mentioned (should be 32 for SMAP-SSM and 45 points for CGLS-SSM).

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

Answer: It is a temporal correlation averaged across the sites; we will fine tune the language in the revised manuscript.

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

Answer: A reference will be added (<https://doi.org/10.1109/TGRS.2018.2858004>).

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?

Answer: We will rephrase the sentence as follows:

"...even if the simulations were limited to dominantly rainfed agricultural areas according to the CORINE land use map and therefore did not include irrigation, it is possible that in reality irrigation is applied in the field and seen by the satellite data, resulting in lower correlation metrics."

Limiting the analysis to the same spatial mask would not be beneficial, because active and passive microwave-based soil moisture retrievals each have their own limitations (recommended quality flags) in space and time. Crossmasking both datasets, possibly also with optical satellite-based biomass retrievals, would mean a too large loss of valuable data. We will add a note about this in the manuscript.

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

Answer: With Pearson's R we compared the seasonal dynamics of the model and the product. The anomaly correlations account for interannual temporal dynamics, as the seasonally varying climatology has been subtracted for each year. The time series shown in the manuscript then indicate the short-term and interannual temporal dynamics.

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?

Answer: The TAW map is presented in figure 2.a. as a comparison. We see a clear agreement with the lower anomR biomass map in the North (North Germany/ Poland) and low TAW values and we verified this in the case for extremely sandy soil class, which contained very low TAW values (due to low FC).

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.

Answer: A mask would indeed be better to confirm that statement. However, with the intercomparison of the in situ products, we already make a comparison over the same locations for both SMAP-SSM and CLGS-SSM, showing much better correlations of SMAP-SSM to both in situ observations and AquaCrop simulations. See also above why a full crossmasking of SMAP- and CLGS-SSM would not add value. The same reference of L306 will be added here.

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?

Answer: Please, see comment at L285. In other words: parameter data sets are by design not self-consistent.

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?

Answer: This is a finding (L280).

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.

Answer: Agreed, however due to reasons mentioned in L60 we prefer to stay with the ‘generic crop’.

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

Answer: Of course. We want to apply satellite- based data assimilation to the model, using radar satellite observations (and possibly passive microwave satellite observations). Within our research group at KU Leuven, Sentinel-1 backscatter data has been processed over Europe at 1-km resolution. This dataset is then used to first calibrate a backscatter model, to transform AquaCrop soil moisture and vegetation output into backscatter values. Ensemble realizations will be generated to account for meteorological input uncertainty. Then, we would like to perform the actual data-assimilation, probably using a particle filter approach. This is all work in progress.

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

Answer: Will be changed.

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.

Answer: The white areas are indeed no data areas. We will specify this in the revised text. Due to the strict quality screening for SMAP-SSM, there is less data available.

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?

Answer: With 4.b. and 4.c. we wanted to visualize that SMAP-SSM in most cases correlates better to in situ points and to the AquaCrop measurements than the CGLS-SSM product does.

Please also note the supplement to this comment:

<https://gmd.copernicus.org/preprints/gmd-2021-98/gmd-2021-98-AC2-supplement.pdf>