Response to RC2

We thank the reviewer for the evaluation of our work. All raised points are very helpful and appreciated. Below you will find our collective responses to each part of the reviewer's comment (blue for our reply and purple for the proposed changes to the manuscript).

Summary:

Basically, the authors describe important new model developments for the ACEA model and show novel findings that are of interest for different scientific communities. The results are well described and well-illustrated.

Thanks for the positive response to our work. We are happy to hear that spending a lot of time on those illustrations was worth it.

In my opinion, paper readability should be increased before publication. Therefore, redundancies should be removed from the text. Units are sometimes separated with a blank and sometimes not. Please be consistent throughout the text.

Thanks for pointing this out. We will try to shorten the text and be consistent with units.

Different declarations e.g. about spatial resolution are confusing and the downscaling approach that has been applied is not described in the manuscript.

Since we use different resolutions for modelling (30 x 30 arc minutes) and post-processing (5 x 5 arc minutes), the applied procedures might be confusing for a reader. We briefly explain the reasons and methods used in upscaling of input data and downscaling of results in the manuscript (see Sect. 2.2). For example, lines 210-211: "*The simulation results are downscaled to 5 x 5 arc minutes according to the location of rainfed and irrigated maize production systems in MIRCA2000…*" – meaning that all rainfed/irrigated 5 x 5 arc minute cells within a respective 30 x 30 arc minute cell are assigned the same values. This allows us to have a more precise distribution of maize WFs around the world. We will reformulate the description of the methods and procedures in the revised version to make it clearer.

Important model assumptions and approaches are not described and referred to literature instead. Please summarize the most important model assumptions and approaches, so that this paper stands for its own and the reader doesn't have to read multiple other papers to understand the ACEA model.

Our initial rationale was to exclude detailed explanations of the AquaCrop model, as our work was mostly devoted to the development of a simulation framework and data processing. Instead, we referred to sources that already include the elaborated and well-illustrated explanation of the AquaCrop model. However, in doing so we might have rushed over some important assumptions that the reader would like to see reflected in our manuscript too. We will therefore revise our description of AquaCrop in the Methods and make sure to include an overview of assumptions and approaches.

Generally, I am not sure if one can say that ACEA is a biophysically based mechanistic process model, such as LPJmL or DSSAT. The crop yield calculation is based on AquaCrop-OS, which - to my understanding - is not a very physically based model, since it works with Penman-Monteith and some rough scaling factors.

Indeed, AquaCrop does not have the same level of model complexity in terms of the number and detail of biophysical processes incorporated in the model compared to some other crop models. As AquaCrop's developers aimed to balance simplicity, accuracy, and robustness (Steduto et al., 2009), the model simplifies some processes. Yet AquaCrop does capture the main biophysical processes that are relevant to accurately simulate crop yield response to water (see a short description below). The model is therefore probably best labelled as "*a water-driven process-based crop growth model*" (Vanuytrecht et al., 2014).

As for the scaling factors, they are only used in the post-processing of AquaCrop outputs (see Sect. 2.1.4). We need them to account for technological developments and external disruptions (e.g. political unrest, floods etc) that cannot be modelled.

Short description of AquaCrop

AquaCrop is a "*water-driven process-based crop growth model*" (Vanuytrecht et al., 2014). As words "*water-driven*" indicate, the model works around the assumption that crop development is mainly determined by water availability. The model's core processes are daily soil-water balance and canopy cover (CC) development which is a subject to agronomic management (e.g. irrigation, mulching, plant characteristics) and various abiotic stresses (see the figure below). On a daily basis, CC is used to convert ET_0 (calculated with Penman-Monteith) into crop transpiration. The latter is then converted into crop biomass via the CO2-adjusted water productivity factor (WP). The final yield is calculated by multiplying accumulated biomass with a stress-adjusted harvest index. As AquaCrop aims to balance simplicity, accuracy, and robustness (Steduto et al., 2009), the model considers only water-related biophysical processes. The other processes such as nutrient cycle or carbon dynamics (as in LPJmL and DSSAT) are not considered.

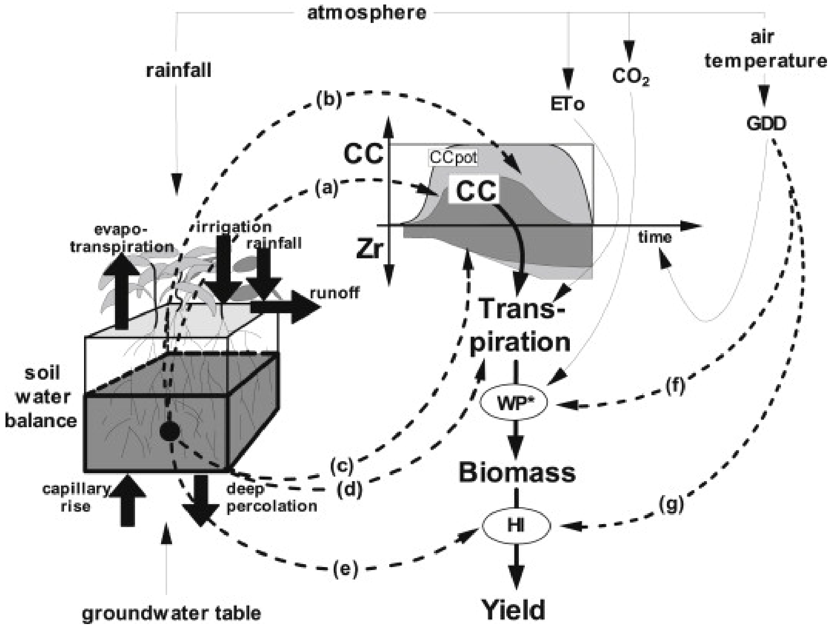


Figure: AquaCrop simulation scheme. Dotted arrows indicate processes affected by water stresses: (a) slows canopy expansion, (b) accelerates canopy senescence, (c) decreases root deepening but only if severe, (d) reduces stomatal opening and transpiration, and (e) affects harvest index; and temperature stresses: cold *temperature reduces biomass productivity (f), hot or cold temperature inhibits pollination and reduces HI (g). Source: Vanuytrecht et al. (2014).*

Literature:

- Vanuytrecht, E., Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., Heng, L.K., Garcia Vila, M., Mejias Moreno, P., 2014. AquaCrop: FAO's crop water productivity and yield response model. Environmental Modelling & Software 62, 351–360. https://doi.org/10.1016/j.envsoft.2014.08.005
- Steduto, P., Hsiao, T.C., Raes, D. and Fereres, E. (2009), AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles. Agron. J., 101: 426-437. https://doi.org/10.2134/agronj2008.0139s

The authors claim several time and in different context that they are the first, which is incorrect every time. I therefore suggest to be a little more modest and careful in the statements, because it could give the impression that you may not be familiar with existing data or literature.
Thanks for pointing this out. We will revise our statements.

Model results are compared against other models, but the study lacks in a model validation (e.g. of ET), as e.g. done in Kimball et al. (2019): https://doi.org/10.1016/j.agrformet.2019.02.037.
We will add the comparison of our ET estimates to literature.

Overall, I would suggest major revisions for the submitted paper.

Specific comments:

- Line 14: The term 'high agricultural development' same as 'low agricultural development' (Line 16) could be misleading, since it is not clear what 'high development' exactly means. I'd suggest using instead 'highly intensive'.
  Thanks for pointing this out. We will revise our statements.

- Line 16: Abbreviation CV (coefficient of variation) could be written out at first appearance to make it easier for the reader.
  Thanks for the suggestion, we will do this.

- Line 17: has reduced by 34.6% until which year? 2016?
  Since it is stated in line 12 that the simulation runs until 2016, we thought it would be clear for the reader that a 34.6 % reduction is by 2016 as well. We will add the year to avoid confusion.

- Line 25: I would be careful with this statement, because the increasing demand is certainly a driver but not the reason for environmental degradation.
  We acknowledge that a "*driver*" does not necessarily mean a "*reason*", and will add more nuance to the statement.

- Line 25: I would keep the term 'planetary boundaries' from the reference instead of using 'environmental limits', since this concept is commonly known under the term 'planetary boundaries'.
  Fair suggestion, we will do this.

- Line 26: There are also large uncertainties and different values exist for the global water consumption of ‘crop production’. It would be interesting for the reader and also nice for the introduction (also with respect to your new approach) to describe the range from different approaches (maybe between 70 and 90%). Another question in this context: Do you mean agriculture or exclusively ‘crop production’ here?

  Thanks for pointing this out. We will consider adding a range instead of one number. In that paragraph, we refer to crop production only.

- Line 51: The coupling of grid cells is only required if it is necessary to consider lateral water flows, what you don’t do?

  We do not model the lateral flows as it is not possible in AquaCrop.

- Line 51: Since there is a lot of new literature available for GGCMs, I’d suggest to additionally cite the following publications here to give a broad overview of existing models and latest approaches:
  - Zabel F, MuÌller C, Elliott J, et al. Large potential for crop production adaptation depends on available future varieties. Glob Change Biol. 2021;00:1–13. https://doi.org/10.1111/gcb.15649
  - Minoli, S., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Zabel, F., Dury, M., Folberth, C., François, L., Hank, T., Jacquemin, I., Liu, W., Olin, S., Pugh, T. A. M. (2019): Global response patterns of major rainfed crops to adaptation by maintaining current growing periods and irrigation. - Earth's Future, 7, 12, 1464-1480. https://doi.org/10.1029/2018EF001130
  - Müller, C., Franke, J., Jägermeyr, J., Ruane, A. C., Elliott, J., Moyer, E., Heinke, J., Falloon, P., Folberth, C., Francois, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Liu, W., Olin, S., Pugh, T., Williams, K. E., Zabel, F. (2021): Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios. - Environmental Research Letters, 16, 3, 034040. https://doi.org/10.1088/1748-9326/abd8fc

  Thanks for providing the references. We will consider adding them to the Introduction.

- Line 52: I’d suggest to mention the Global Gridded Crop Model Initiative (GGCMI) within the Agricultural Model Intercomparison and Improvement Project (AgMIP). In the context of climate impact assessments, it would be great to include Jägermeyr et al. (2021), in which the new CMIP6 scenarios are applied to a large ensemble of global gridded crop models. The publication is currently still under review in Nature Food but could be accepted soon.

  Thanks for this suggestion. We will refer to AgMIP and GGCMI as well as add Jägermeyr et al. (2021) to the references.

- Line 50-54: I disagree that GGCMs have never been used so far to estimate WFs. Maybe that depends on the definition of a GGCM and also of what you mean with WFs. Since a general definition of a GGCM does not exist, this is difficult. In general, a GGCM must not necessarily be physically based. There are a lot of studies that look e.g. at Evapotranspiration (ET) for crop models, e.g. Lui, W. et al. (2016): Global investigation of impacts of PET methods on simulating crop-water relations for maize. Agricultural and Forest Meteorology, 221, 164-175. https://doi.org/10.1016/j.agrformet.2016.02.017.

Thanks for pointing this out. We will revise our statements.

- In GGCMI, models have simulated ET in phase 1, phase 2 and the latest phase 3. There has been approaches, e.g. by Jägermeyr et al. to investigate water flows, irrigation demands, and crop water productivity with crop models. Deryng et al. e.g. investigated crop water productivity in context with increasing CO2 concentrations:
    - Jaegermeyr, J. et al. Reconciling irrigated food production with environmental flows for Sustainable Development Goals implementation. Commun. 8, 15900 doi: 10.1038/ncomms15900 (2017).
    - Jaegermeyr,J. et al. (2015): Water savings potentials of irrigation systems: global simulation of processes and linkages. HESS, 19, 3073–3091, 2015 . doi:10.5194/hess-19-3073-2015
    - Deryng D. et al. (2016): Regional disparities in the beneficial effects of rising CO2 concentrations on crop water productivity. Nature Climate Change. DOI: 10.1038/NCLIMATE2995

  Thanks for providing the references. We will consider adding them to the Introduction.

- Line 78: I am not sure if one can say that ACEA is a process-based model, such as LPJmL or DSSAT. Crop yield calculations are based on AquaCrop-OS, which - to my understanding - is not a biophysical process based mechanistic crop model.

  See our response above. ACEA is a process-based model, but indeed not on the same level of detail as for example LPJmL, which we will clarify in a revised version of the paper.

- Line 78: Redundant. The model abbreviation ACEA has already been introduced, so not necessary to do it again. Same with the abbreviation GGCM. Please only write out the complete name at first use and use the abbreviation in the following (without the abstact, in which abbreviations should generally be avoided).

  Fair suggestion, we apply this throughout the paper.

- Line 79: What means high temporal resolution? In line 92, you say daily. Most GGCMs use daily temporal resolution, but some global gridded crop models run at hourly resolution. Accordingly, what means 'high' in your sentence? I would suggest to delete high and write daily instead. Finally, to reduce redundancy, this should be deleted in line 92.

  We meant daily indeed. We will revise these statements according to your suggestion.

- Line 86 and Figure 1: Since the 'scenarios' only refer to different water supply systems, I'd suggest to call them 'water supply' or 'water supply scenarios' or 'water supply assumptions' instead of 'scenarios'.

Good suggestion, we will consider changing it to water supply scenarios.

- Line 89: Is fertilizer (N,P,K) a possible input for management in ACEA?
  No, as nutrient cycles are not considered in the current version of AquaCrop.

- Line 91: Why are grid cells iterated, when lateral flows are not considered? You could parallelize the grid cells (as you actually say in line 65).
  There are both options available. The first one is to run grid cells consecutively in a *for* loop. The second one is to run the same loop parallelized so the iterations (or tasks) will be distributed among a user-defined number of CPUs. The first one is generally enough for small applications but for large scales, as in this paper, the parallelized option is preferred.

- Line 92: How is crop growth simulated? What approaches are used? I think it is required to describe the main approaches and processes of the model (e.g. how is atmospheric CO2 concentration considered?). The reader has to understand the most important underlying approaches without having to read the other Aquacrop publications!
  As mentioned in response to another comment, we will add the description of the AquaCrop model in the paper.

- Line 94: Redundant, parallelization is already mentioned in line 65.
  Thanks for pointing this out. We will revise this statement.

- Line 107: Since there are many different GDD approaches available, which one has been implemented to the model?
  AquaCrop provides three methods to calculate GDDs, so it is user-defined in ACEA as well. For our simulation, we chose the default AquaCrop method - Method 3 (see the excerpt from the AquaCrop manual below):

3.3.3 Method 3

As in method 2, the comparison to T_{base} and T_{upper} occurs before the calculation of the average temperature. However the check is only on the maximum air temperature. The average temperature is given by:

$$T_{avg} = \frac{\left(T_x^* + T_n\right)}{2} \qquad \text{(Eq. 3.3d)}$$

where T_x^* is the adjusted maximum air temperature and T_n the minimum air temperature. The following rules apply:

- T_x^* is the maximum air temperature ($T_x^* = T_x$)
  If T_x is greater than T_{upper}, then $T_x^* = T_{upper}$,
  If T_x is smaller than T_{base}, then $T_x^* = T_{base}$
- T_n is not adjusted. However if T_n exceeds T_{upper}, T_n will be set equal to T_{upper}.

Once T_{avg} is calculated, it is checked if the average air temperature is above the base temperature. If T_{avg} is less than T_{base}, then T_{avg} is taken as T_{base} (resulting in 0 °C day on that day).

Reference Manual, Chapter 3 – AquaCrop, Version 6.0 – 6.1 May 2018 3-20

Figure: Excerpt from D. Raes, P. Steduto, T. C. Hsiao, and E. Fereres, 'AquaCrop Version 6.0 - 6.1: Reference manual (Annexes)', Rome, 2018.

- Line 150: Is this correction factor, derived by a bias correction of yields, used also to scale evapotranspiration? This is not explained here. But if yes, is a linear relationship between yield and ET realistic?

  First, we would like to point out that the main drivers for the historical increase in maize yields are better crop varieties (improved biotic and abiotic stress resistance), higher plant density, and better field management (fertilizers, chemical control of weeds and insects etc) (Duvick, 2005; Lorenz et al., 2010). In ACEA, we attempt to represent the combined effect of these drivers by scaling the simulated yields to the national statistics from FAOSTAT.

  We only scale the yields but not the seasonal ET of maize (i.e. CWU) due to several reasons:

  1. There is no strong evidence that maize CWU increased along with improvements in **maize varieties**. For example, Nagore et al. (2014) found that two new maize varieties have similar CWU as an old one in Argentina, and Xu et al. (2018) showed only a minor difference in CWU between large and small canopy maize varieties in China.

     One of the main conclusions from these studies is that differences in crop varieties may change the ratio of transpiration (T) to soil evaporation (E) but overall CWU would be only minorly affected as a decrease/increase in T is compensated by an increase/decrease in E. Both E and T consume green and blue water so we do not expect significant changes to overall green and blue CWU either.

2. To address the effect of increased **plant density**, we carried out sensitivity analysis and found that it does not lead to significant changes in CWU. We illustrate this by simulating maize production in one of the cells in northern Italy during 1986-2016 with three plant density values (see table below):

Plant density [plants ha^{-1}]	Average yield [t ha^{-1} y^{-1}]		Average CWU green [mm y^{-1}]		Average CWU blue [mm y^{-1}]	Average WF [m^3 t^{-1} y^{-1}]	
	rainfed	irrigated	rainfed	irrigated	irrigated	rainfed	irrigated
50000	7.2	13.5	343.5	309	187.6	477.1	367.9
75000	7.2	14	341.7	308.3	194.1	474.6	358.9
100000	7.4	14.3	341.4	310.6	195.7	461.4	354.1

As you can see from the table, CWU values barely change with an increase in plant density. Thus, the difference in maize WFs comes from an increase in yields rather than in CWUs. Also, Barbieri et al. (2012) concluded that a change in maize row spacing does not affect CWU. Hence, having the same plant density value for the whole simulation period worldwide seems to be a reasonable assumption (in our paper we assumed 75 000 plants ha^{-1} as it is the most common value in literature).

3. As for the **fertilizer application** rates, Rudnick et al. (2017) showed that nitrogen inputs might increase maize CWU up to 13 % compared to no nitrogen input. Also, the authors demonstrated that CWU does not show the consistent direction of changes with different nitrogen application rates (from 84 to 252 kg ha^{-1}) but they were always in a range of -10 % to +10 %. In our study, we have to assume no nutrient stress (i.e. optimal nutrient supply) as AquaCrop cannot simulate the nutrient cycle. This might lead to the overestimation of CWU in the grid cells with no fertilizer application. However, we expect no significant impact on global maize WFs. First, because the potential overestimation of CWU by even 13 % is still minor compared to the effect of yield scaling. Second, because fertilizer application is a common practice among the big maize producers, and thus we think it is safe to assume that the (production-weighted) global average maize WFs would hardly be affected.

To sum up, the literature seems to indicate that historical changes in crop varieties and field management only minorly affect green and blue maize CWU compared to crop yields. Therefore, we conclude that using the yield scaling factors is sufficient to represent historical dynamics in maize WFs at the global level. We will further substantiate our assumptions in the Methods and Discussion chapters according to the mentioned above reasoning. Also, we will add a comparison of our maize ET estimates to the existing literature which, in case of minor differences, would again indicate that scaling of ET is not needed.

Literature:

- D. N. Duvick, ‘The Contribution of Breeding to Yield Advances in maize (Zea mays L.)’, in Advances in Agronomy, vol. 86, Elsevier, 2005, pp. 83–145. doi: 10.1016/S0065-2113(05)86002-X

- A. J. Lorenz, T. J. Gustafson, J. G. Coors, and N. de Leon, 'Breeding Maize for a Bioeconomy: A Literature Survey Examining Harvest Index and Stover Yield and Their Relationship to Grain Yield', Crop Sci., vol. 50, no. 1, pp. 1–12, Jan. 2010, doi: 10.2135/cropsci2009.02.0086.
- M. L. Nagore, L. Echarte, F. H. Andrade, and A. Della Maggiora, 'Crop evapotranspiration in Argentinean maize hybrids released in different decades', Field Crops Research, vol. 155, pp. 23–29, Jan. 2014, doi: 10.1016/j.fcr.2013.09.026.
- G. Xu et al., 'A lysimeter study for the effects of different canopy sizes on evapotranspiration and crop coefficient of summer maize', Agricultural Water Management, vol. 208, pp. 1–6, Sep. 2018, doi: 10.1016/j.agwat.2018.04.
- P. Barbieri, L. Echarte, A. Della Maggiora, V. O. Sadras, H. Echeverria, and F. H. Andrade, 'Maize Evapotranspiration and Water-Use Efficiency in Response to Row Spacing', Agronomy Journal, vol. 104, no. 4, pp. 939–944, Jul. 2012, doi: 10.2134/agronj2012.0014.
- D. R. Rudnick, S. Irmak, K. Djaman, and V. Sharma, 'Impact of irrigation and nitrogen fertilizer rate on soil water trends and maize evapotranspiration during the vegetative and reproductive periods', Agricultural Water Management, vol. 191, pp. 77–84, Sep. 2017, doi: 10.1016/j.agwat.2017.06.007.

- Line 159: The assumption that maize harvested areas experienced the same dynamics as croplands seems arbitrary to me and must lead to large regional errors. Is there any evidence that maize areas behave similar than total cropland areas? The term 'extrapolation' in this context seems wrong as it seems to be a scaling. To me, the procedure is not yet clear. Does the irrigation fraction in each pixel remain constant in your scaling approach? If not, can you explain where the change in irrigation fraction comes from? If all values are scaled with FAOSTAT in the end, why not directly scaling MIRCA2000 with FAOSTAT trends for maize for each country?

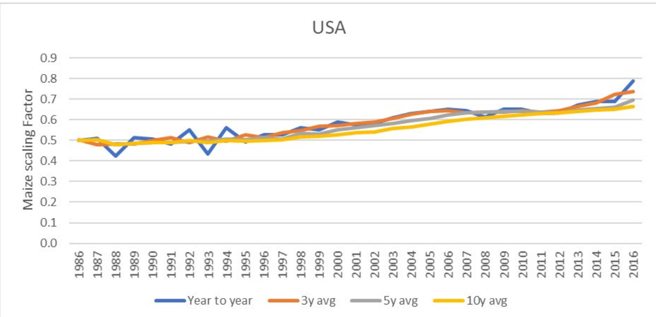
  First, we would like to note that FAOSTAT does not differentiate between rainfed and irrigated production systems. Therefore, if we scale MIRCA2000 directly to national FAOSTAT data, the irrigated areas would have the same % change as rainfed ones in each grid cell within a country. However, the literature suggests that many countries substantially changed the fractions of irrigated areas within their territories since the 1980s (e.g. in India, former USSR countries, China, Brazil) (Ambika et al., 2016; Nagaraj et al., 2021; Siebert et al., 2015).

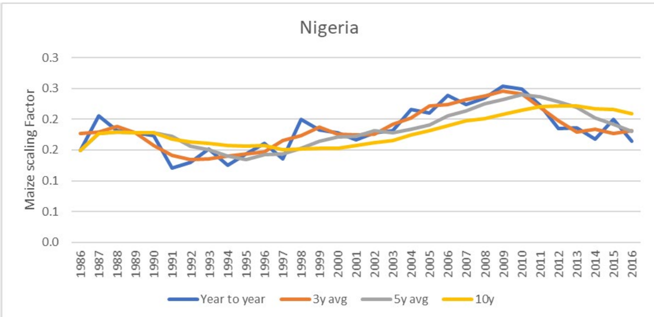
  To address this issue, we add one more step before scaling MIRCA2000 to FAOSTAT: we project MIRCA2000 data – which reports gridded maize irrigated and rainfed harvested area separately – into the past (till 1986) and future (till 2016) by assuming that irrigated and rainfed harvested areas generally followed the same trend as irrigated and rainfed croplands from HYDE did (see lines 158-161). This allows to account for historical change in the fraction of irrigated maize in each grid cell, but of course, it may lead to regional errors as maize is only one of the crops considered in cropland dynamics. However, regional errors would be also present if we just scale MIRCA2000 to FAOSTAT (as explained in the first paragraph), and thus both approaches would lead to inevitable uncertainties. Therefore, in the revised paper, we will try to validate our historical rainfed/irrigated harvested area estimates with external data (such as GAEZ v4 (Fischer et al., 2021) and SPAM2010 (Yu et al., 2020)) and compare them to the alternative scaling approaches of the MIRCA2000 dataset.

Literature:

- Ambika, A. K., Wardlow, B., and Mishra, V.: Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015, Sci Data, 3, 160118, https://doi.org/10.1038/sdata.2016.118, 2016.
- Nagaraj, D., Proust, E., Todeschini, A., Rulli, M. C., and D'Odorico, P.: A new dataset of global irrigation areas from 2001 to 2015, Advances in Water Resources, 152, 103910, https://doi.org/10.1016/j.advwatres.2021.103910, 2021.
- Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, Hydrol. Earth Syst. Sci., 19, 1521–1545, https://doi.org/10.5194/hess-19-1521-2015, 2015.
- Fischer, G., Nachtergaele, F. O., Velthuizen, H. van, Chiozza, F., Franceschini, G., Henry, M., Muchoney, D., and Tramberend, S.: Global Agro-Ecological Zones (GAEZ v4) Model Documentation, FAO & IIASA, 2021.
- Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., Xiong, W., Lu, M., Wu, W., and Yang, P.: A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps, Earth Syst. Sci. Data, 12, 3545–3572, https://doi.org/10.5194/essd-12-3545-2020, 2020.

- Line 165: Why do you use a three year moving average and not 5 or 10-year? Have you made a sensitivity analysis or is there any assumption that gives arguments for taking a 3-year average? Yes, we carried out a sensitivity analysis with 3-, 5-, and 10-year moving averages for four countries that represent a range of annual maize production quantities and environmental conditions: USA, Nigeria, Australia, and Algeria (see the graphs below).





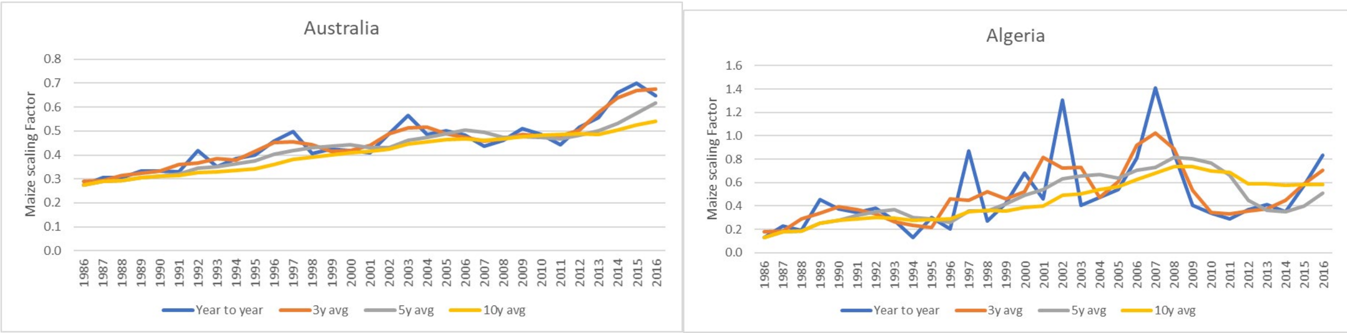


Figure: Sensitivity analysis of moving averages of yield scaling factors for maize in USA, Nigeria, Australia, and Algeria (blue - no moving average, orange – 3-year, grey – 5-year, and yellow – 10-year moving average).

Among those three moving average options, the 3-year one resulted in the most favourable representation of interannual variability and trends, namely: "*This allows to keep the overall trend in historical crop yields and attenuate extreme responses to short-term external developments*" (line 165).

- Line 178: Now I am confused. In line 13 (abstract) and in line 78, you said that you allied the model at 5 x 5 arc minutes. Now you say, you run ACEA at 30x30 arc minutes. What is correct?
  Almost all input data for crop modelling are obtained at 30 x 30 arc minute resolution. Therefore, running ACEA at 5 x 5 arc minutes instead of 30 x 30 arc minutes would make no difference to the final results (except for the cells with shallow groundwater levels, see lines 200-204). However, the time spent for running the simulation would differ significantly. Therefore, we decided to run ACEA at a coarser resolution and then distribute the outputs to respective 5 x 5 arc minute grid cells. The following data-processing (WF calculation, yield scaling etc.) is then performed at 5 x 5 arc minute grid cells as well.

- Line 179: Please be aware that MIRCA2000 harvested areas for the maize class includes maize (corn), maize for forage and silage, and pop corn. How do you deal with the different maize usages that also go along with different plant characteristics and harvest characteristics (e.g. for silage, the complete overground biomass is harvested)?
  Thanks for pointing this out. We considered all maize types (under the maize class in MIRCA2000) as maize (corn) only. The reason for this is a lack of input data to simulate them separately. We will mention this in Methods.

- Line 180: When you consider harvested areas according to MIRCA2000, you implicitly consider multiple growing seasons that are included in the harvested area (if a physical area of 1 ha is harvested twice per year, the harvested area is 2 ha).
  We indeed simulate only the main (one) growing season if the subsequent growing seasons (i.e. sub-crops in MIRCA2000) are minor. In the case of maize, sub-crop 2 area / sub-crop 1 area = 0.53%, so the simulation of the second growing season is not relevant to represent the global trends.

- Line 185: Is the GSWP3-W5E5 data is based on bias-corrected reanalysis data? If yes, that would be important to mention here.

  Thanks for pointing this out. We will add this information.

- Line 190: The same methodology is also used within GGCMI, you could refer to Minoli et al. (2019):
  - Minoli, S., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Zabel, F., Dury, M., Folberth, C., François, L., Hank, T., Jacquemin, I., Liu, W., Olin, S., Pugh, T. A. M. (2019): Global response patterns of major rainfed crops to adaptation by maintaining current growing periods and irrigation. - Earth's Future, 7, 12, 1464-1480. https://doi.org/10.1029/2018EF001130

  Thanks, we will add this reference.

- Line 205: I know studies that assume irrigation to be triggered below 70% of field capacity. There seems to be some a range of values in the literature that could be discussed as another source of uncertainty.

  We also saw a 70 % threshold in some studies, but 50 % is the most common value in literature. This is indeed a source of uncertainty as we mention in line 403.

- Line 207: To be clear: You always assume full irrigation and don't consider e.g. deficit irrigation, right?

  Yes, correct, we only consider full irrigation since it is the most common approach and there is a lack of global data on irrigation strategies.

- Line 210: How is the downscaling applied? If you downscale the results to 5x5 arc minutes, you cannot say that the model is applied at 5x5 arc minutes (see e.g. abstract).

  We run the model at 30 x 30 arc minutes and, based on more spatially detailed information on maize growing areas, we then downscale results to 5 x 5 arc minutes as described in Sect. 2.2. We will choose other wording in the abstract.

- Figure 3: For the right panel of Fig. 3, I would suggest to use an area weighted mean to consider the different maize areas (e.g. the US corn belt should weigh more than small areas), instead of using the median of all data points along the latitude. Is the color bar logarithmically scaled? Please explain in figure caption. Additionally, I think the 10thpercentile is the correct formulation (10% percentile would be doubled).

  Thanks for the suggestion, the colour bar was scaled to optimise the colour distribution (so not logarithmic). We will consider adjusting this and other similar graphs according to your suggestions.

- The Table 2 is good to have and helpful.

  We are glad that you like it.

- Line 293: There seems to be high uncertainties about global maize areas and expansion.

  That is true and we aim to discuss these in more detail as indicated in the responses above.

- Figure 7 and 10: For improving this figure, one could set the dot size relative to the maize area or maize production in the country to visualize the importance of the respective country.
  Thanks for the suggestion. We will consider this, but the main point of those graphs is to show the number of countries around -30 % and + 30% dotted lines, so scaling the dots to production amounts might look too messy.

- Line 343: Please be aware that the applied crop calendar also includes high uncertainties, and regions have been identified that do not well represent local phenological data from observations. A new updated crop calendar is currently being processed in ISIMIP.
  Thanks for pointing this out. We will add this to limitations in the Discussion.

- Line 346: Can you explain why it is less accurate to calculate green and blue CWU in the post-processing? Isn't that a question on how it is implemented?
  As explained by Hoekstra (2019), there are two main approaches to calculate green and blue CWU in post-processing. The first one is to assume that blue CWU is equal to the difference between irrigated and rainfed CWU. This approach is problematic since the rooting depths and soil moisture dynamics are different if irrigation is applied. The second one is to estimate blue CWU based on the relative addition of water to the soil via irrigation and rainfall. This approach is better, but the problem here is the lack of tracing of green and blue water within the soil profile, which leads to the same composition of soil moisture at different soil depths. Both approaches result in a less accurate estimation of green and blue water loss through evaporation and transpiration, which leads to a less accurate estimation of CWU. Moreover, it is not known how much green and blue water is entered or lost the soil profile during the fallow period. Hence, the composition of water in the soil profile at the start of the next growing season is not known either. This may increase the uncertainty in green and blue CWU estimates of each subsequent growing season.

  - Hoekstra, A. Y.: Green-blue water accounting in a soil water balance, Advances in Water Resources, 129, 112–117, https://doi.org/10.1016/j.advwatres.2019.05.012, 2019.

- Line 382: I would be careful with this statement. The fact that you don't know any study that has shown this doesn't mean that no other GGCM can do this, as most existing GGCMs have the ability to simulate that on a daily base.
  Thanks for pointing this out. We will revise this statement.

- Line 394: Again, it is not explained, which downscaling methodology is applied. If this is just interpolated, I wouldn't call it a 'downscaling' approach.
  We already explained this in the previous responses.

- Line 396: Actually all data except the crop calendar are available at 5x5 arc minutes. For climate input, one could use e.g. WFDE5, HWSD soil data is also available at 0.00833° spatial resolution.

We are not aware of the 5 x 5 arc minute version of WFDE5. We kindly ask the reviewer to provide a reference to this dataset as we are only able to find the 30 x 30 arc minute version online. If this dataset is available, we will revise our statement in line 396.

- Line 398: Indeed, this is a strong limitation, since cultivar variations and improvements over time play a big role, especially for maize.
  As we already explained before, the historical changes in maize cultivars mainly lead to higher crop yields while CWU stays approximately the same. We account for the increase in maize yields by applying the scaling factors, so the improvement in maize varieties is indirectly considered. Consequently, this limitation has only a minor impact on global WF trends. However, the impact at regional scales might be more pronounced.

- Line 410: Please delete this statement. An extrapolation of historical maize areas based on FAO trends has been performed e.g. by Iizumi, T., Sakai, T. The global dataset of historical yields for major crops 1981–2016. Sci Data 7, 97 (2020). https://doi.org/10.1038/s41597-020-0433-7.
  Thanks for pointing this out. We will revise this statement and add this reference to the Methods.

- Line 423: What means high spatial and temporal resolution? Please avoid subjective statements such as ‘good’, ‘big’ or ‘high’ in scientific articles.
  Well noted. We will revise this statement.

- Line 423: To me, the conclusion mainly reads like a summary and can be shortened.
  Thanks for this suggestion. We will consider shortening the conclusion.