Dear RC2,

We sincerely appreciate the time and effort you dedicated to reviewing our work. Below, we provide our detailed responses to your concerns:

A. Model validation

"1. It is recommended to use multiple gridded yield datasets to validate global gridded crop models (GGCMs) because grid-level yields can vary significantly between datasets, which is a significant source of uncertainty when assessing GGCM performance (Müller et al., 2017, Lin et al., 2021). Currently, annual gridded yield datasets are available for the globe and major producing countries at a 5-arcmin resolution (Su et al., 2022, Cao et al., 2025). In addition to comparing their model simulation with the Global Dataset of Historical Yields (GDHY), authors are encouraged to account for yield dataset uncertainty by comparing their model simulation with a family of recent gridded yield datasets."

Reply: Thank you for this helpful suggestion. We agree that relying on a single yield dataset can lead to bias evaluation. Hence, we will add the comparison with other datasets as well for figure 8 and add the explanation "Simulated maize yields and three references in Figure 8 of MATCRO-Maize, Global Dataset of Historical Yield (GDHY by Iizumi and Sakai, 2019), GlobalCropYield (Cao et al., 2025), and Spatial Production Allocation Model (SPAM by IFRI, 2019) were harmonized into 0.5° resolution. The value was averaged over 1981–2014 for GDHY, averaged over 1982-2014 for GlobalCropYield, and year 2010 for SPAM.". In addition, we will include spatial distribution maps showing the differences between MATCRO-Maize yields and each reference dataset to make the comparison clearer in the revised manuscript.

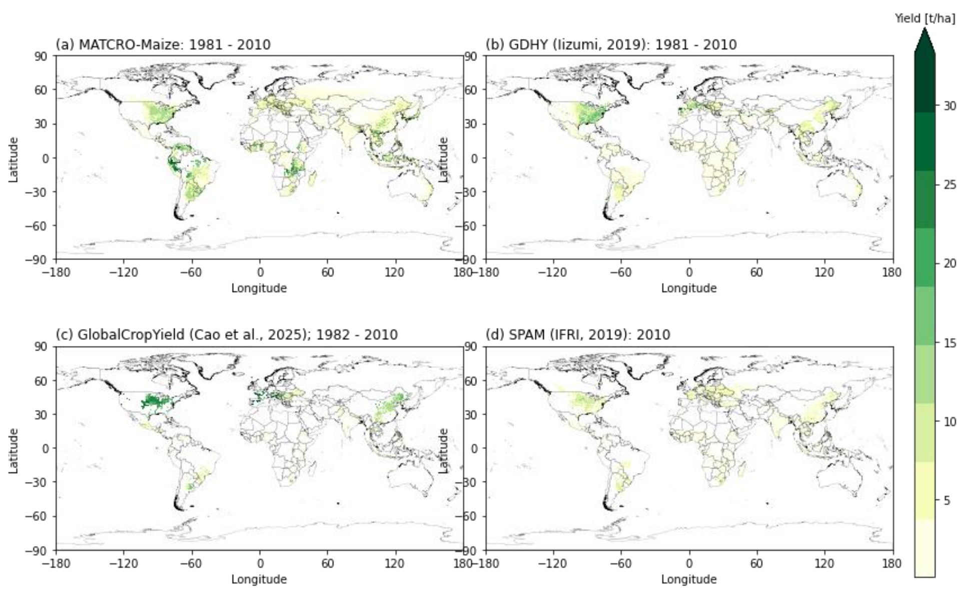

Figure 8. Global distribution of the 30-year average (1981-2010) maize yield by (a) simulations from the MATCRO-Maize and (b) the GDHY dataset. For comparison, yield estimates from shorter periods are also shown from (c) GlobalCropYield for 29-year average (1982-2014) and (d) SPAM2010 for year 2010. The simulated yield is aggregated based on the harvested area from MIRCA2000.

“2. The validation of crop phenology at the global level is currently lacking. I’m happy to see the model validation result at the site level (Fig. 4). However, the data compared are for only four sites and one year, which is inadequate for concluding the model performance. Gridded crop phenology datasets have recently become available for the globe and some major countries (Luo et al., 2020, Yang et al., 2020, Mori et al., 2023). I strongly encourage the authors to compare their simulation with these datasets.”

Reply: Thank you for insightful comments. In this study, we evaluated crop phenology at the point scale using a single growing-time dataset. As data availability is still limited for phenology, we focused on simulating the yield and relied on the gridded crop calendar from Jägermeyr et al. (2021) for planting and harvest dates. In response to your suggestion, we will expand the evaluation by comparing simulated phenology with recently available of global dataset of crop phenological events (GCPE by Mori et al., 2023) and present discrepancies in the figure for simulated harvest time compared with GGCMI and GCPE.

“3. In relation to Comment#2, in the current form, it is unclear how the model parameter values related to crop phenology were determined before the model simulation. The authors state that “We used local daily climate data … and phenological data (planting, flowering, and maturity dates) for model input data at each site. (Line 276-278)”. Did you calibrate the parameter values using the site data and then run the model? If so, this does not constitute model validation because no independent data were used for comparison. I would ask the authors to clarify this point and rerun the model validation if necessary.”

Reply: We are truly sorry for the confusion. You are correct, the current figures reflect the same data with parameterization which is not independent validation. We will use different term from “validation” and replace it with “evaluation”. Figure 4 will be described as an evaluation of model fit for phenology (flowering and maturity dates) with caption: ”Figure 4. Model-fit comparison of the flowering and maturity date simulations (SIM on the y-axis) and observations (OBS on the x-axis).”

B. Modeling

“1. How did you determine Gdd,m (eq.22; the growing degree days at maturity)? Is this a universal value across grid cells worldwide? It is well-documented that Gdd,m varies spatially, with higher values in warmer regions and lower values in cooler regions (Deryng et al., 2011, Mori et al., 2023). I would ask the authors to clarify this.”

Reply: Pardon us for the confusion. We use different value of growing degree days in each grid cell as noted in Deryng et al,. (2011) and Bouman et al., (2001). We will revise the Eq. 22 with adding the subscript of i for each grid cell where i means the grid cell number as stated below:

$$D_{vs,i} = G_{dd,i}/G_{ddm,i}, \tag{22}$$

“2. The leaf area index (LAI) simulated by the MATCRO-Maize model appeared to be lower than the site observations (Fig. 5). It is also noticed that the difference in maximum LAI between the sites is smaller in the simulation than in the observations (Fig. 5). It leads to the thought that the maximum value of the specific leaf nitrogen parameterized with annual nitrogen application rate (*Nfert*) (eq. 29) is rather site-dependent and cannot be applied universally in its current form. This does not mean that publishing this preprint is unjustified. However, readers at least want to know whether underestimation of the seasonal maximum LAI correlates with environmental conditions, such as soil carbon content, soil total nitrogen content, water holding capacity of the soil and so on, in order to seek a possible scaling factor to convert specific leaf nitrogen to LAI. The equation (8) of Hasegawa et al. (2008) for the fraction of canopy cover may help the authors relate specific leaf nitrogen to seasonal maximum LAI (though this equation was developed for rice). If such a correlation analysis provides no insight, then calibrating the scaling factor for each country is another option, as was done by Ai and Hanasaki (2023)."

Reply: Thank you for raising this topic. We agree that the simulated LAI is lower and shows less variation across sites compared to observations. In Figure 5, we applied the same universal parameters (SLW and leaf partitioning) across all sites, as our aim was global-scale application. Under low nitrogen conditions (e.g., Brazil), this can lead to underestimated LAI because the universal parameters do not represent no-fertilizer situations in the site scale simulation, as leaf morphological traits are known to vary with nitrogen availability (Ciampitti et al., 2013a,b; Hokmalipour and Darbandi, 2011). A sensitivity test in MATCRO confirmed that varying SLW strongly affects simulated LAI. The SLN–Vcmax relationship itself is applied globally because site-specific data are not available. Moreover, the soil water balance in MATCRO tends to underestimate water availability in deeper soil layers, which may contribute to yield underestimation under rainfed conditions. However, this could not be confirmed due to the limited availability of observational data. Other factors not considered in the current model framework may also contribute to this bias. We will clarify in the manuscript that the underestimation of LAI is more likely due to using universal morphological parameters at the site scale.

"3. The presentation of the relationship between *Nfert* and yield, as presented in Fig. 12, is a bit misleading and could be improved. As can be seen in Fig. 12 (a), yield increases with an increase in *Nfert*, but then saturates. The yield response to *Nfert*, as derived from FAOSTAT, is consistent with literature which attributes recent maize yield growth to delayed leaf senescence (staygreen), morphological change from horizontal to vertical leaf type and increased drought tolerance, and resulting increase in planting density, rather than an increase in N input (Duvick, 2005). These genetics and management improvements have changed maize yield response to N input (Fig. 3 of DeBruin et al. 2017). Therefore, liner regression is inappropriate to describe the *Nfert*-yield relationship. Consider using a nonlinear regression or locally estimated scatterplot smoothing (LOWESS) instead. More importantly, the presented version of MATCRO-Maize imperfectly represent the *Nfert*-yield relationship (regardless of whether the data for Egypt is included or omitted). Rather than presenting Fig. 13, I would suggest the authors discuss this limitation of the model."

Reply: Thank you very much for this constructive suggestion. We agree that the relationship between Nfert and yield cannot be adequately described by linear regression. We will use nonlinear regression (or LOWESS) in the revised manuscript and remove Fig. 13. We will also add a statement in the limitation section noting that MATCRO-Maize can generally reproduce *yield responses to different N fertilizer levels, but the model tends to overestimate yields under certain conditions (e.g., Egypt), which indicates opportunity for further improvement as follow:*

"The current version of MATCRO-Maize can reproduce yield responses to nitrogen fertilization across a range of fertilizer levels, but it tends to overestimate yields under certain conditions (e.g., Egypt) likely because the model assumes higher nitrogen use efficiency and idealized irrigation conditions where actual yields are constrained by soil quality, management, and local cultivar traits that are not explicitly represented. This suggests that the representation of nitrogen effects in the model remains simplified, and further refinement is needed for region-specific scale simulation."

"4. The simulated aboveground biomass was lower than the site observations (Fig. 7). However, the simulated yields at the country level were substantially overestimated. This discrepancy may be due to inaccurate partitioning to harvested organ or to stress factors reducing yield formation. I do understand that there are many factors not considered in the model, such as biotic stresses (pests and diseases, weeds, etc.), as described in Line 554. Nevertheless, recent crop models that are embedded in Earth system models that operate at a global level are encouraged to incorporate some form of parameterization to handle major drivers of historical yield growth even in a simple way (Lombardozzi et al., 2020). Alternatively, please consider calibrating some of the existing parameters to better reproduce historical yields (Ai and Hanasaki, 2023)."

Reply: Thank you for this thoughtful comment and explanations. We acknowledge the discrepancy between site-level and country-level simulations due to the use of universal parameter in the site-level simulation. While additional calibration could improve agreement with historical yields, our approach emphasizes physiological mechanisms and universal parameters rather than statistical fitting. We will clarify this distinction in the manuscript and note the limitation that stress factors and other drivers of yield formation are not yet explicitly represented as follow:

"A limitation of the current study is the use of universal parameters at the site scale leads to discrepancies between site-level and country-level simulations. It partly arises from applying universal parameters across different environments. Moreover, genetic variation among cultivars is not considered, and key factors of yield formation are not yet explicitly represented (e.g. plant nitrogen balance)."

C. Technical corrections

"1. 'Production' is generally measured in tones and is calculated by multiplying yield (production volume per unit harvested area and cropping season) by area harvested, in the case of single-season maize (see Box 1 of Wei et al., 2023). However, as MATCRO-Maize does not harvest area, the ""yield model" is more appropriate than the "production model".

"2. Line 264. I think the correct citation for the GDHY is "Iizumi and Sakai, 2020" rather than "Iizumi, 2019". Please check what recent literature describes this point (for instance, Data Availability and references of Iizumi et al., 2025)."

"3. Line 281. Do you mean "AgMERRA" (Ruane et al., 2015), a bias-corrected version of the MERRA reanalysis designed for agricultural applications, rather than the original "MERRA"?"

"4. Line 173. In agronomic literature, the flowering of maize is generally referred to as 'silking'. The first time it appears, you should mention this, for example, "flowering (known as silking; *Dvs,flw*)"."

Reply: Thank you for clarification on the technical corrections. We agree with your review in point 1-4 and we will adopt them in the revised manuscript.

References:

- *Bouman, B. A. M., Kropff, M., Tuong, T., Wopereis, M., ten Berge, H., and van Laar, H.: Oryza2000: modeling lowland rice, International Rice Research Institute, and Wageningen: Wageningen University and Research Centre, Philippines and Wageningen, the Netherlands, 235 pp., 2001.*
- *Ciampitti, I. A., Murrell, S. T., Camberato, J. J., Tuinstra, M., Xia, Y., Friedemann, P., and Vyn, T. J.: Physiological dynamics of maize nitrogen uptake and partitioning in response to plant density and N stress factors: I. Vegetative phase, Crop Sci, 53, 2105–2119, https://doi.org/10.2135/cropsci2013.01.0040, 2013a.*
- *Ciampitti, I. A., Murrell, S. T., Camberato, J. J., Tuinstra, M., Xia, Y., Friedemann, P., and Vyn, T. J.: Physiological dynamics of maize nitrogen uptake and partitioning in response to plant density and nitrogen stress factors: II. reproductive phase, Crop Sci, 53, 2588–2602, https://doi.org/10.2135/cropsci2013.01.0041, 2013b.*
- *Hokmalipour, S., Darbandi, M. H. Effects of nitrogen fertilizer on chlorophyll content and other leaf indicate in three cultivars of maize (Zea mays L.). World Applied Sciences Journal, 15(12), 1780-1785, 2011.*
- *International Food Policy Research Institute (IFPRI), Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0, https://doi.org/10.7910/DVN/PRFF8V, Harvard Dataverse, V4, 2019.*

- Ai, Z. and Hanasaki, N.: Simulation of crop yield using the global hydrological model H08 (crp.v1), Geosci. Model Dev., 16, 3275–3290, https://doi.org/10.5194/gmd-16-3275-2023, 2023.
- Cao, J., Zhang, Z., Luo, X. et al. Mapping global yields of four major crops at 5-minute resolution from 1982 to 2015 using multi-source data and machine learning. Sci Data 12, 357 (2025). https://doi.org/10.1038/s41597-025-04650-4
- DeBruin, J.L., Schussler, J.R., Mo, H. and Cooper, M. (2017), Grain Yield and Nitrogen Accumulation in Maize Hybrids Released during 1934 to 2013 in the US Midwest. Crop Science, 57: 1431-1446. https://doi.org/10.2135/cropsci2016.08.0704
- Deryng, D., W. J. Sacks, C. C. Barford, and N. Ramankutty (2011), Simulating the effects of climate and agricultural management practices on global crop yield, Global Biogeochem. Cycles, 25, GB2006, doi:10.1029/2009GB003765.
- Duvick, D. N. The contribution of breeding to yield advances in maize (Zea mays L.). Adv. Agron. 86, 83–145 (2005).
- Hasegawa, T., Sawano, S., Goto, S. et al. A model driven by crop water use and nitrogen supply for simulating changes in the regional yield of rain-fed lowland rice in Northeast Thailand. Paddy Water Environ 6, 73–82 (2008). https://doi.org/10.1007/s10333-007-0099-1
- 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
- Toshichika Iizumi, Toru Sakai, Yoshimitsu Masaki, Kei Oyoshi, Takahiro Takimoto, Hideo Shiogama, Yukiko Imada, David Makowski, Assessing the capacity of agricultural research and development to increase the stability of global crop yields under climate change, PNAS Nexus, Volume 4, Issue 4, April 2025, pgaf099, https://doi.org/10.1093/pnasnexus/pgaf099

- Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., Ciais, P., Deryng, D., Folberth, C., Glotter, M., Hoek, S., Iizumi, T., Izaurralde, R. C., Jones, C., Khabarov, N., Lawrence, P., Liu, W., Olin, S., Pugh, T. A. M., Ray, D. K., Reddy, A., Rosenzweig, C., Ruane, A. C., Sakurai, G., Schmid, E., Skalsky, R., Song, C. X., Wang, X., de Wit, A., and Yang, H.: Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications, Geosci. Model Dev., 10, 1403–1422, https://doi.org/10.5194/gmd-10-1403-2017, 2017.
- Lin, T.-S., Song, Y., Lawrence, P., Kheshgi, H. S., & Jain, A. K. (2021). Worldwide maize and soybean yield response to environmental and management factors over the 20th and 21st centuries. Journal of Geophysical Research: Biogeosciences, 126, e2021JG006304. https://doi.org/10.1029/2021JG006304
- Lombardozzi, D. L., Lu, Y., Lawrence, P. J., Lawrence, D. M., Swenson, S., & Oleson, K. W., et al. (2020). Simulating agriculture in the Community Land Model Version 5. Journal of Geophysical Research: Biogeosciences, 125, e2019JG005529. https://doi.org/10.1029/2019JG005529
- Luo, Y., Zhang, Z., Chen, Y., Li, Z., and Tao, F.: ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products, Earth Syst. Sci. Data, 12, 197–214, https://doi.org/10.5194/essd-12-197-2020, 2020.
- Akira MORI, Yasuhiro DOI, Toshichika IIZUMI, GCPE: The global dataset of crop phenological events for agricultural and earth system modeling, Journal of Agricultural Meteorology, 2023, Volume 79, Issue 3, Pages 120-129, Released on J-STAGE July 10, 2023, Advance online publication May 16, 2023, Online ISSN 1881-0136, Print ISSN 0021-8588, https://doi.org/10.2480/agrmet.D-23-00004,
- Alex C. Ruane, Richard Goldberg, James Chryssanthacopoulos, 2015. Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. Agricultural and Forest Meteorology, 200, 233-248. https://doi.org/10.1016/j.agrformet.2014.09.016
- Su, H., Willaarts, B., Luna-Gonzalez, D., Krol, M. S., and Hogeboom, R. J.: Gridded 5 arcmin datasets for simultaneously farm-size-specific and crop-specific harvested areas in 56 countries, Earth Syst. Sci. Data, 14, 4397–4418, https://doi.org/10.5194/essd-14-4397-2022, 2022.
- Yang Y, Ren W, Tao B et al., 2020: Characterizing spatiotemporal patterns of crop phenology across North America during 2000-2016 using satellite imagery and agricultural survey data. ISPRS Journal of Photogrammetry and Remote Sensing 170, 156-173. https://doi.org/10.1016/j.isprsjprs.2020.10.005
- Wei, D., Gephart, J.A., Iizumi, T. et al. Key role of planted and harvested area fluctuations in US crop production shocks. Nat Sustain 6, 1177–1185 (2023). https://doi.org/10.1038/s41893-023-01152-2