Response to Reviewers

Reviewer #1

Liu et al. present a well motivated analysis that could provide the climate and crop community with a lightweight tool to apply to a variety of climate impacts studies. Their manuscript is well written and clearly presented, but their cross validation analysis, which provides the basis of the manuscript, is flawed. To their credit, they admit this flaw, but recognizing it is not sufficient.

RE: Thanks for your suggestions and comments. We have revised the text according to your suggestions and respond to your comments point-by-point as follows.

The 10-fold cross validation should be withholding the entire domain for selected years, not just random samples for the exact reason that they state (e.g. spatial autocorrelation). The Köppen–Geiger approach within a 10-fold cross validation could also work (withhold entire Köppen–Geiger class for 10% of years). But the 10-fold cross validation that is presented is not valid and should not be the default metric presented throughout the paper. Provided the strong spatial autocorrelation of climate and yields in this model setup in many locations at the grid-cell level, the 10-fold cross validation cannot be considered an out-of-sample analysis and should not be presented as such. Few of the graphs specify whether the 10-fold cross validation is the basis of the results, but I assume that is the default cross validation chosen to compute the model evaluations based on Figure 2. I encourage the authors to correct their cross-validation and revise their results accordingly.

RE: Thanks for your suggestions. We have now used a new cross-validation strategy and updated the corresponding results.

In the method section, we revised the cross validation strategy as "Considering the spatiotemporal autocorrelation of simulated crop yield given by GGCM, we now used a "held out years and regions" strategy for leave one-year-out cross-validation (Roberts et al., 2017; Sweet et al., 2023). Specifically, the all grid-year samples are split into N folds. N is determined by the number of Köppen–Geiger (KG) classes, which have more than 100 grid cells with harvested areas. If there are too few harvested areas in one KG class, it will not be included in the cross-validation process. For each fold of emulator training and validation, we withhold 10% of years (the last 3 years) and one entire KG class for validation, and the other grid-year samples are used for training the emulator. We think selecting continuous years for validation can avoid temporal autocorrelation. If we randomly select 10% of years, the correlation between adjacent years still exist. Actually, any continuous three years are able to solve this problem, such that we just use the last years according to the choice of (Sweet et al., 2023)."

In results section, the Figure 2-11 are revised by using the new cross-validation strategy. The revised Figure 2 is presented as follows:

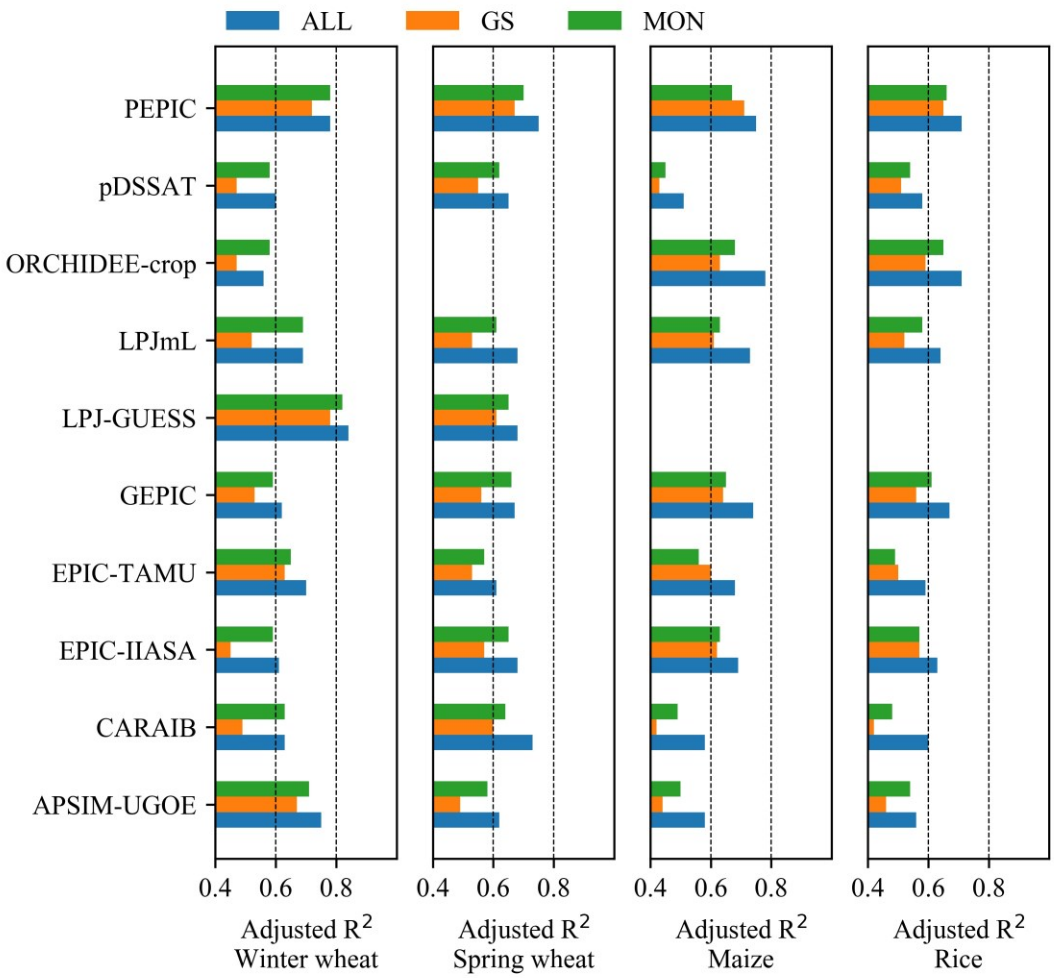

Figure 2 Adjusted R^2 of emulators (10-fold cross validation with randomly selected samples) with different strategery of predictors. All: "Full model", GS: "GS model", MON: "Mon model". Emulators for ORCHIDEE by spring wheat, and LPJ-GUESS by Maize and Rice were not fitted due to the lack of simulation of raw GGCM.

The results have changed with this new cross-validation approach and the presentation of results and the discussion of the results have been adjusted accordingly. Although the new emulators obtain lower accuracy than originally reported, the revised emulators still facilitate lightweight estimates of yields and their year-to-year variability.

L75-76 grammar typos

RE: Revised as suggested. In the text, we revised the phrases as "However, the scenario-based future crop yield projection is not a systematic perturbation of climate factors.".

L169: CDD is consecutive dry days not drought days

RE: Revised as suggested. In the text, we revised the phrases as "maximum consecutive drought dry days (CDD)".

L194-196: I'm not sure what a "spatial difference term" means or what the clarification of a "temporal constant growing season length" means. Is this a location fixed effect applied at the grid-cell level? Could you clearly write out an equation for the covariates going into you XGBoost model? This would help to clarify which variables are location invariant, which are time invariant, and which vary both with location and with time. Table 2 is good, but it doesn't clarify which covariates change in space vs time. Alternatively you could add this information into Table 2.

RE: Thanks for your suggestions. We intended to say "time invariant" variables but used "spatial difference term". And we have used "temporal constant growing season length" to indicate "the length of days from planting date to maturity date given by GGCMI phase2 crop calendar input". Those two phrases have been corrected in the revision. In the text, we revised the phrases as "To reproduce the length of days from planting date to maturity date given by GGCMI phase2 crop calendar input, we added a temporal constant growing season length as a predictor, i.e. temporal constant growing season length.”

Meanwhile, we have added a column to Table 2 to clarify the time/space variant information. A new column “Time” is used to clarify the spatiotemporal dynamics of predictors. The new Table 2 is revised as follows:

Table 1 Predictors of emulation. For rainfed yield emulators, we used all these predictors but for fully-irrigated yield emulators, the precipitation predictors were not included. Full, GS and Mon were three strategies to develop emulators. Full: developing emulators with all the climatic predictors; GS: developing emulators with climatic predictors during growing season scale; Mon: developing emulators with climatic predictors during monthly scale.

Predictor abbreviations	Descriptions	References	Full	GS	Mon	Time
	Temperature related predictors					
$\mathrm{GDD}_{\mathrm{low\text{-}high}}$_GS	Growing degree day during growing season (winter wheat: low=0°C, high=30°C; spring wheat: low=5°C, high=30°C; maize: low=8°C, high=30°C; rice: low=10°C, high=35°C)	(Frieler et al., 2017; Jägermeyr et al., 2020; Lobell et al., 2012)				1
$\mathrm{EDD}_{\mathrm{high+}}$_GS	Extreme degree day during growing (winter and spring wheat, maize: high=30°C; rice: high=35°C	(Lobell et al., 2012)				1
Tmax_GSmean	Average daily maximum temperature during growing season	(Zhu and Troy, 2018)				1
Tmin_GSmean	Average daily minimum temperature during growing season	(Zhu and Troy, 2018)				1
Tmax_GSstd	Standard deviation of daily maximum temperature during growing season	(Zhu and Troy, 2018)				1
Tmin_GSstd	Standard deviation of daily minimum temperature during growing season	(Zhu and Troy, 2018)				1
Tmax_MONmean	Harmonized monthly average daily maximum temperature (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)				1
Tmin_MONmean	Harmonized monthly average daily minimum temperature (MON=1–10 for winter wheat, MON=1–8 for spring	(Folberth et al., 2019) (Jägermeyr et al., 2020)				1

	wheat and maize, MON=1–7 for rice, since planting date)					
Precipitation related predictors						
Pre_GSsum	Total daily precipitation during growing season	(Troy et al., 2015)				1
Pre_GSstd	Standard deviation of daily precipitation during growing season	(Zhu and Troy, 2018)				1
CDD_GS	Consecutive drought day (daily precipitation=0)	(Troy et al., 2015)				1
Pre_MONsum	Harmonized monthly total precipitation (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)				1
Solar radiation related predictors						
SRAD_GSmean	Average daily solar radiation during growing season	(Folberth et al., 2019)				1
SRAD_GSstd	Standard daily solar radiation during growing season	(Folberth et al., 2019)				1
SRAD_MONmean	Harmonized monthly average daily solar radiation (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)				1
Greenhouse gas concentration						
CO_2	CO_2 concentration	(Franke et al., 2020a)				2
Non-climatic predictors						
N	Nitrogen fertilizer application	(Franke et al., 2020a)				2
Soil_type	Soil type	(Blanc, 2017)				3
GSL	Growing season length	(Folberth et al., 2019)				3

*The colored the row denotes the predictors was included in the emulator. The column "Time" is defined to clarify the spatiotemporal dynamics of predictors: "1" represents both time and space variant predictors, "2" represents space invariant predictors, "3" represents time invariant predictors.

Line 270: remind readers what A0 and A1 simulations mean

RE: Revised as suggested. In the text, we revised the phrases as "The A0 denotes no adaptation and A1 denotes adaptation of the growing season to regain the original growing season length under warming scenarios that otherwise lead to accelerated phenology and thus shorter growing seasons."

Table 3: consider adding a key for A0 and A1 as you have for W and Winf

RE: We have added a key for Table3 “The A0 denotes no adaptation and A1 denotes cultivar adaptation to regain original growing season length under warming scenarios.”

Minor point: While the manuscript is generally clear and well written, grammar should be checked throughout the manuscript. There are typos throughout.

RE: Thanks for pointing this out. We have checked the typos throughout the manuscript.

Reviewer #2

The content of this study meets the requirements of this journal, and the paper is highly complete with exquisite images. However, there are certain deficiencies in using process-based models to simulate the impacts of current climate change. Therefore, exploring the use of artificial intelligence algorithms to improve the limitations of the current process-based models is a worthwhile direction, in my opinion, and has the potential for publication.

RE: Thanks for your comments. We have revised the manuscript and responded to your comments point-by-point as follows.

Line 34: To be frank, I didn't know what "per se" meant, so I looked it up and found that it is a commonly used term in written language. However, I am unsure if it is appropriate to use it here. I have been using English for over twenty years, and this is the first time I have come across this usage.

RE: Thanks for pointing this out. We have replaced the phrase “models per se” as “raw models”. In the text, we revised the phrases as “While crop model emulators are believed to be lightweight tools to replaces the raw models”.

Line 59-60: I am somewhat confused about the author's statement here. The author mentions that "the relationship between climate factors and crop yield is constrained by the current climate conditions." What does this mean? Aren't both of these models also constrained by the current climate conditions?

RE: Thanks for pointing it out. We intended to say that the statistical relationship can only reproduce the historical climate-yield relationship, such that using historical climate-yield relationship to project the crop yield in future are not convincing.

We have revised the sentence as “the relationship between climate factors and crop yield is based the historical climate conditions and their effects on crop yields, which can hardly be used for future projection with new, unprecedented climate conditions”.

Line 75: The readers are already aware that the scenario-based future crop yield projection is not a systematic perturbation of climate factors. What other limitations are there in the scenario-based future crop yield projection that should be mentioned here?

RE: Thanks for your suggestions. We have added some limitations of scenario-based future crop yield projection here. In the text, a new sentence has been added "For instance, the scenario-based yield projection can only provide the simulated crop yield driven by simultaneous changes in climate factors. The dependency of temperature and precipitation will be kept in scenarios, such that the impact of temperature and precipitation cannot be clearly separated."

Line 114: I think you can avoid mentioning soybean here. It is confusing why soybean is not included in your subsequent research.

RE: Thanks for your suggestions. We have removed the mentioning of soybean.

Regarding Figure 3 and the following 1:1 graphs, while they are visually appealing, they provide limited information. We can see that the points align along the 1:1 line, but no statistical indicators are presented.

RE: Thanks for your suggestions. We have added correlation coefficient (R) and mean absolute error (MAE) in Figure3 and Figure 6.

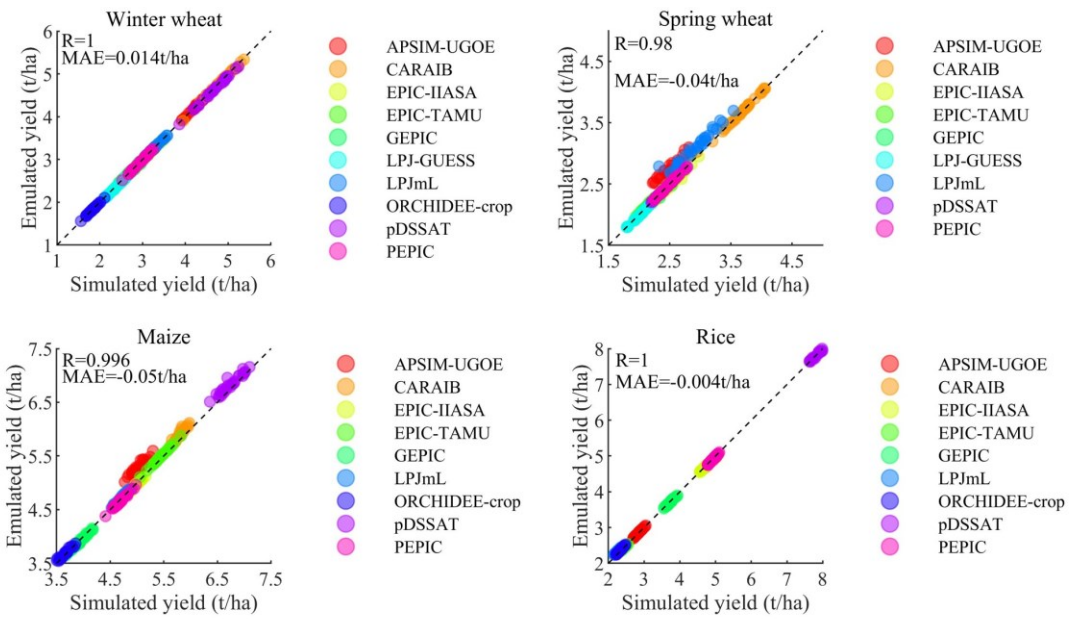

Figure 3 Emulator performance to reproduce the year-to-year variation of global average yield (1981 – 2010) over current cropland. As ORCHIDEE-crop has not simulated yield under C360T0W0N200, we used the C360T0W10N200 as the baseline. Each point with the same color is yield in 30 year. R is correlation coefficient and MAE is mean absolute error.

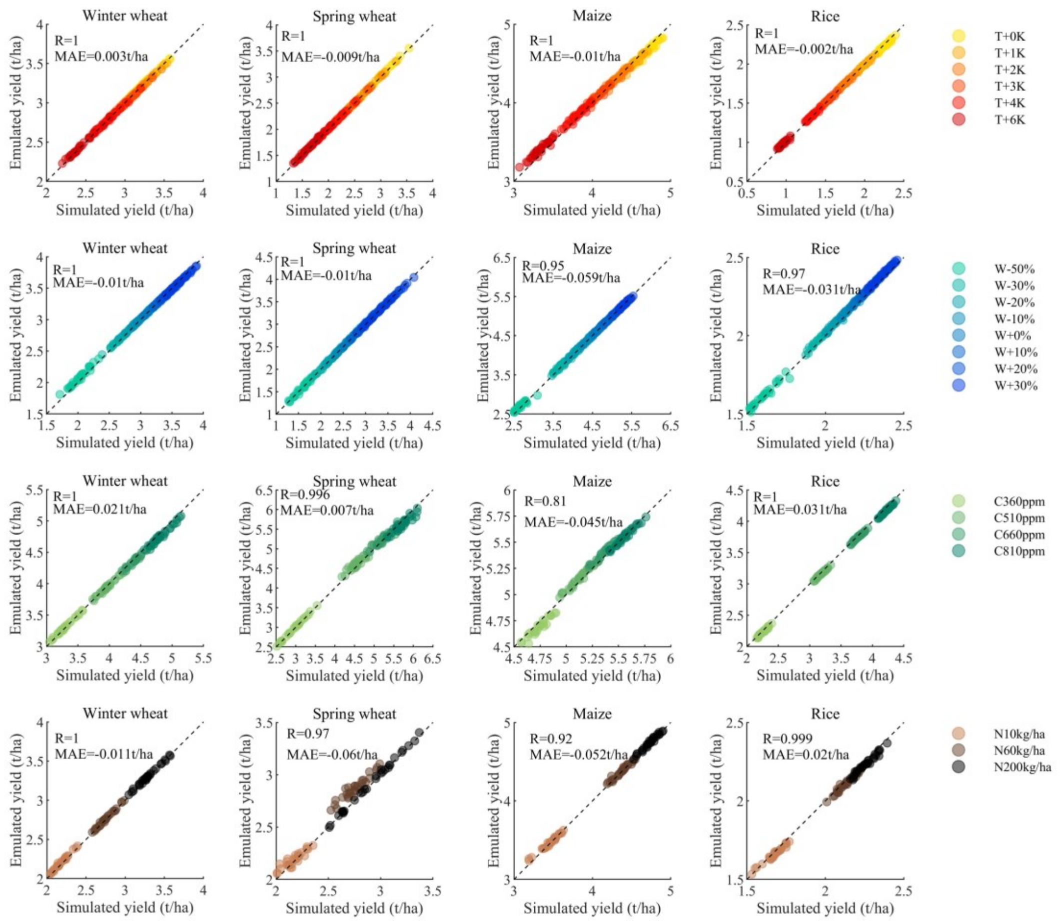

Figure 6 Performance of one exemplary emulator (LPJmL-A0) in reproducing the year to year variation of global mean yield from 1981 to 2010 under varied individual CTWN perturbations. Each point with the same color is yield in one year. The performances of other emulators are similar to LPJmL-A0, which can be referred in the Table S1 and Table S2.

References

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography (Cop.). 40, 913–929. https://doi.org/10.1111/ecog.02881

Sweet, L., Müller, C., Anand, M., Zscheischler, J., 2023. Cross-validation strategy impacts the performance and interpretation of machine learning models. Artif. Intell. Earth Syst. 1–35. https://doi.org/https://doi.org/10.1175/AIES-D-23-0026.1