

General comments

I appreciate the authors' efforts devoted for this revision (and thank you for listing my comments one by one, which are more readable. There was a technical barrier for me to do so). The methodological details for the SPAM2010 are much clearer than the previous manuscript. The authors' responses regarding the validation of SPAM2010 are sufficiently strong to justify the publication of the paper, although I comment on the way of presentation in this review report. Nevertheless, some methodological details remain unclear and their clarifications are still needed. I should mention all my questions at the first round of review (sorry for this), but some descriptions added in the revision require further clarifications. I pointed out several concerns below, which are relatively major but not substantively major.

Relatively major concerns

1. L67-71. This is misleading and needs edit carefully. I understood that SPAM2010 product offers global crop production maps for 2010, which has the latest base year compared to other products, such as M3 and MIRCA for 2000. However, as the authors discuss in L811-816, it is inappropriate to compare SPAM2010 with its predecessors (SPAM2000 and SPAM2005) to explore historical change in area, yield and production. I admire the authors' honesty to show this result and agree with this statement. Therefore, it is misleading to emphasize the importance of area change over time because SPAM products cannot be used for this purpose.
2. L161. It is still unclear which production and area you intend to refer when you state "share". I found some explanations for this distinction in Supplementary Information (SI) Section 4, but a clearer distinction between production share and area share throughout main text is needed. This comment is applied throughout the manuscript, for instance, L334, L347, Table 2 (e.g., Percentjlk) and Table S5 (the title).
3. L323-324. Although FAO country statistics are the most reliable source of data for global agriculture, these include many data values from unofficial sources or estimation. This is especially true for developing countries. This arise a question whether adjusting national and subnational statistics against FAO data is always valid. I don't request the authors recalculating SPAM2010 because this type of uncertainty is widely observed not only FAO data versus subnational statistics but also different global yield maps (Anderson et al. 2015 (this is already cited); Fig.9 of Müller et al. 2017; Schauburger et al. 2017, Iizumi et al. 2018). However, the authors are encouraged to explain a bit more (in addition to L323-324) why FAO data are used as the baseline in the adjustment of country statistics despite the possible uncertainties in FAO data.
4. For Figs. 6, 8, 9,10, 11, 12, 14 and 15 presenting grid-cell level comparisons between SPAM2010 and other products, I would strongly suggest the authors using the density scatter plot instead of the ordinary scatter plot to improve the visibility of the agreement between two data sources. For instance, see Scatter plots with rectangular bins at <http://www.sthda.com/english/wiki/ggplot2-scatter-plots-quick-start-guide-r-software-and-data-visualization> for R script. And some of these figures show RMSE (e.g., Figs. 8, 9 and 10) but others don't. For consistency and informativeness, I ask the authors adding RMSE values for all of these scatter plots.

Technical corrections

5. L48. “geo-political boundaries”. I think, “administrative units” would be more suitable to be consistent with the remaining portion of the manuscript.
6. L58. “the harvested area and yield” should read “the potential harvested area and yield” as GAEZ estimates potential levels but not actual ones.
7. L119. It would be nice if a brief introduction of SPAM2005 (e.g., “SPAM2005 expands its coverage to xx crops with the updated base year of 2005” or similar) could be added here to give readers a short history on the improvements to SPAM products.
8. L131-132. This is a bit hard to understand when I read for the first time. Crop statistics are disaggregated to what? Perhaps, you want to say here that national-level statistics are disaggregated into sub-national level; statistics for crop aggregates are divided into individual crop types; and area, production and yield statistics are separated for each of rainfed and irrigated conditions (rainfed conditions are further disaggregated into input levels – high, low and subsistence).
9. L179. “the statistical demand of a crop”. I don’t understand this term. Please consider rephrasing. I suspect that the authors intend to indicate a measure of the completeness in disaggregation in terms of area extent or production quantity. The same comment is applied to SI Section 4.
10. L352. I realized that there are two types of yield conversion factor from Response to Reviewer Comments, (i) rainfed yield to irrigated yield and (ii) rainfed low-input yield to rainfed high-input yield. This information is important but currently lacking in main text. Please consider adding it to main text around here.
11. L386. “optimal potential” can simply read “potential”.
12. L395-401. Why did you use GMIA to derive information on irrigation equipped area? MIRCA can provide crop-specific irrigated area, whereas GMIA cannot (but MIRCA is for 2000). Indeed, MIRCA also used for some countries to collect irrigation information, as described in Table S4. And HID (Siebert et al. 2015) also provide global irrigation area for 2005 and that for 2010 is likely estimated using historical trends. I would appreciate if the authors’ underlying thought on the use of GMIA could be explained briefly here.
13. L480-490. Here the procedure is hard to follow. To me, it seems that yield and production at different scales are adjusted several times (maybe my understanding is incorrect). Can you edit a bit more to increase readability? More importantly, it seems that yield at statistical reporting unit is used as the first guess of grid-cell yields. However, yields at finer scales (e.g., farm field level) could largely differ from subnational statistical yield and there is a space for better modeling (Gerlt et al. 2014, Porth et al. 2017). Therefore, I’d like to ask the authors to add a brief justification and possible limitations of the current method here (or elsewhere in Discussion).
14. L511. “high in than region”. Probably, this is a typo.
15. L537. “I\$”. Do you mean international dollar? If yes, please explicitly state so. What is the base year?
16. L577. “predicting crop areas”. In general, the term “prediction” is used to derive a value of a variable of interest in the time t using inputs at the time t-1. Here, inputs at the time t are used to derive value of crop area at the time t. In this case, “estimation” is an appropriate term.

17. L667. “231 and 307”. What is the unit? ha? The same comment is applied to RMSE value in L672.
18. L668. Why is the agreement for wheat in the US between SPAM2010 and CDL worse than that for maize and soybean? This is interesting because the result suggests that a key factor is likely lacking in wheat modeling.
19. L749. “the correlation”. This should read “the coefficient of determination”.
20. L813. “Izumi et al. (2020)”. Do you mean “Iizumi and Sakai (2020)”?
21. Table S2. What is the difference between zero (e.g., ADM2 harvested area for Afghanistan) and – (e.g., ADM2 harvested area for Albania)? Can you add a brief explanation in footnote of the table? And the difference between #N/A (Liechtenstein) and – is unclear as well.
22. The third paragraph, SI Section 4. Which “harvested area shares for different production systems” and “production quantity shares sourced from different production systems” does “production system shares” indicate?
23. Table S5. Please be specific whether the values indicate either production share or area share.
24. Table S7. I feel difficulty in interpreting the cropping intensities. If I take double rice cropping in monsoonal Asia (India, Indonesia, Philippines etc.) as the example, it operates rainfed condition in wet season and irrigated condition in dry season (e.g., Koide et al. (2013) for Philippines). Therefore, in my interpretation, the sum of cropping intensities over rainfed and irrigated condition can exceed one. However, the cropping intensity values exceed one for both rainfed and irrigated conditions (for instance, cereals in Indonesia). Can you explain a bit more what is the definition and how you calculate these values?
25. SI Section 16. In the figures in this section, what is the area in gray for SPAM harvested area maps?
26. Thailand, SI Section 16. “F on 12” should read “F on 13”, isn’t it?

References

- Gerlt, S. et al. 2014. Exploiting the relationship between farm-level yields and county-level yields for applied analysis. *J Agric Resour Econ.* 39, 253–270.
- Iizumi, T. et al., 2018. Uncertainties of potentials and recent changes in global yields of major crops resulting from census- and satellite-based yield datasets at multiple resolutions. *PLoS ONE* 13(9): e0203809. <https://doi.org/10.1371/journal.pone.0203809>
- Koide, N. et al., 2013. Prediction of Rice Production in the Philippines Using Seasonal Climate Forecasts. *J. Appl. Meteor. Climatol.*, 52, 552–569, <https://doi.org/10.1175/JAMC-D-11-0254.1>
- Müller, C. et al., 2017. 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>,
- Porth, L. et al., 2017. Farm-level crop yield forecasting in the absence of farm-level data. <https://www.soa.org/research-reports/2016/2016-farm-level-forecasting>
- Schauburger, B. et al. 2017. Global evaluation of a semiempirical model for yield anomalies and application to within - season yield forecasting. *Glob Change Biol.* 23, 4750–4764. <https://doi.org/10.1111/gcb.13738>

- Siebert, S. et al., 2015. 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>