Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-11-RC1, 2020


Open Access Earth System Science Data Discussions



# *Interactive comment on* "A cultivated planet in 2010: 2. the global gridded agricultural production maps" *by* Qiangyi Yu et al.

**Anonymous Referee #1**

Received and published: 5 April 2020

General comments SPAM products are one of well-known spatially-explicit global agricultural production datasets. An update of SPAM products can be potentially a great contribution to scientific communities (Earth system modeling and global food security monitoring in particular). However, I think, the current form of the Discussion paper is not sufficiently persuasive for some aspects. An evaluation of the validity of the spatial disaggregation method is lacking. Particularly, although the method estimates harvested area and yield for each of the four farming systems (irrigated, rainfed high input, rainfed low input and subsistence) and this is the most unique characteristics of SPAM products, no evaluation is presented in this Discussion paper (because SPAM products are model estimates, earlier papers (You et al. 2006, 2014) cannot justify skipping evaluation in the paper). A comparison between the latest SPAM product and other



Printer-friendly version

Discussion paper

independent datasets is partly presented, but there is a space for improvements. For these reasons, I would suggest major revision. My comments are elaborated below.

Specific comments 1. An evaluation of the spatial disaggregation model is required. The most prominent uniqueness of SPAM products, including the latest one (i.e., SPAM2010), is a distinction in harvested area and yield across the farming systems. Currently, global datasets other than SPAM products provide no information on area and yield specific to farming system. However, area and yield for each farming system in SPAM products are "estimates" derived using a spatial disaggregation model optimized using the entropy method. Although the authors may claim that this is a data-fusion approach but not a model prediction approach, a model evaluation against the validation subset (that is independent of the training subset) is essential even for a data-fusion approach. This is a common practice across studies using models even in global crop yield dataset compilation (Iizumi et al. 2014; grid-cell yield estimates derived using national yield statistics as the model input are compared with reported subnational yield statistics which are not used as the model input). Note that M3 and MIRCA2000 use a simple allocation rule rather than modeling; and GAEZ is a model output but for "potential" geographic distribution of crop suitable area. However, the purpose of this Discussion paper is to present "actual" distributions of area and yield for specific farming systems. Therefore, an evaluation of the model used is a mandate. Probably, for some crop-region combinations, the authors have farming-system-specific area and yield statistics at subnational levels. I strongly encourage the authors testing and reporting the performance of their model in disaggregating national agricultural statistics into subnational ones when national statistics are used as the model inputs. 2. A comparison of SPAM products and other independent datasets has a space for further improvements. The key shortfalls in the current Discussion paper are: (1) although CDL2010 for the United States and NLCD2010 for China are compared with SPAM2010, these are for harvested area and no comparison is presented for area and yield for the specific farming systems; and (2) although the relative changes in area between 2005 and 2010 are presented in the paper (Fig. 8),

these need be compared with other independent datasets (for instance, CDL2005 and CDL2010 for the United States). The updated M3 dataset which offers the average harvested area and yield for three time points, 1995 (1993–1997), 2000 (1998–2002) and 2005 (2003–2007) is a candidate for the independent dataset and is available online at: http://www.earthstat.org/ (see the dataset labelled “Harvested Area and Yield for 4 Crops (1995-2005)”). For a consistent comparison, if possible, I would encourage the authors updating the earlier SPAM2000 and SPAM2005 products by utilizing the model used for SPAM2010. Such updating is a common practice in global agricultural dataset compilation and important to ensure the continuity of data in products (Iizumi and Sakai 2020, Sloat et al. 2020). 3. Related to the comment #2, Zhang et al. (2017) provides annual paddy area time series from 2000 to 2010 based on satellite remote sensing for China and India. Because recent satellite-based paddy area estimates are quite accurate, this dataset can be a useful source of information to evaluate the relative changes of paddy area in SPAM products. 4. Related to the comment #2, a distinction between average irrigated and rainfed yields for the 1998–2002 period at the global scale is made in Siebert and Doll (2010). These estimates are also used in recent study (Sloat et al. 2020). I think, these estimates can be a useful source of information when evaluating the reliability of farming-system-specific estimates in the SPAM products once updating of SPAM2000 and SPAM 2005 using the latest model is done. 5. A more in-depth discussion on advantages, disadvantages and limitations of the spatial disaggregation model is required. Although the authors hypothetically assume that the use of economic factors, including prices and access to markets, in the disaggregation model is superior to other methods, such as the proportional allocation. However, this working hypothesis has never been tested (at least, I could not find any result neither in this Discussion paper nor in earlier work (You et al. 2009, 2014)). “garbage in garbage out” is a well-known behavior of models. In general, price statistics are less reliable than other variables (e.g., production). I have the same concerns for the quality of data on production share by farming system and the indicator of market access. If some of model inputs are not reliable, model outputs are expected to be unreliable, depending





Printer-friendly version

Discussion paper

on the sensitivity of model output to specific inputs. I like the idea that economic factors are considered in disaggregation, but the idea does not automatically guarantee that model outputs (disaggregated area and yield by farming system) is correct. I think, the advantages of the model relative to simpler methods are stated too bold throughout the Discussion paper. The authors' claims might be true, but need be tested in a standard way of model evaluation (e.g., by using the cross-validation technique).

Technical corrections 6. L71-73. I strongly suggest removing this description. Researchers would use the latest version once global agricultural dataset is updated, but no such update is available to date. This is the reason why the studies cited here use an earlier version. The authors' criticism made here is inappropriate. 7. L107. The current text is a bit misleading. This text should read "M3 has no distinction across farming systems . . ." or similar. 8. L156. Country crop-specific production costs for a specific year (e.g., 2011) are available via GTAP9 database (Aguiar et al. 2016). Just for your information. 9. L158. GAEZ only provides "potential" crop suitability area. Please consider keeping precise terminology in the Discussion paper. 10. Eq. 7. What is "CE"? The abbreviation suddenly appears without definition. And I would appreciate it if the authors could provide a brief explanation what is the difference between {s ln s} versus {s ln $\pi$}. 11. Eq. 16. AdjCropY suddenly appears in main text although it is explained in Supplement. A brief explanation need be added in main text for readability. 12. L304-305. Are the yield conversion factors in the text same with those shown in Table S6? Table S6 shows only for irrigated versus rainfed. Where is rainfed high input versus rainfed low input? 13. 372-373. This assumption is too crude. Dong et al. (2017) presents a nice global dataset in specifying urban areas. It can be useful to distinguish rural and urban areas more accurately. 14. L532. I do not understand "methodological-cum-data". Please consider rephrasing. 15. L626-627. This is true but has not been demonstrated yet. I would suggest removing this statement unless a comparison in area and yield for each farming system against subnational statistics is presented. 16. L636. Zhang et al. (2017) reports the northward shift of paddy area in China and the westward shift of paddy area in India for the 2000-2010 period. These

tendencies seem be inconsistent with the upper panel of Fig. 8. 17. L679-680. Global roads and railways database used in Koks, E.E. et al. (2019) is maybe of your interest to more accurately define accessibility to markets. Just for your information.

References ïĆă Aguiar, A. et al. (2016) An Overview of the GTAP 9 Data Base. Journal of Global Economic Analysis, 1,181–208, http://dx.doi.org/10.21642/JGEA.010103AF ïĆă Dong, Y. et al. (2017) Global anthropogenic heat flux database with high spatial resolution. Atmospheric Environment, 150, 276-294, https://doi.org/10.1016/j.atmosenv.2016.11.040. ïĆă Koks, E.E. et al. (2019) A global multi-hazard risk analysis of road and railway infrastructure assets. Nature Communications, 10, 2677, https://doi.org/10.1038/s41467-019-10442-3 ïĆă Iizumi, T. et al. (2014), Historical changes in global yields. Global Ecology and Biogeography, 23, 346-357, doi:10.1111/geb.12120 ïĆă Iizumi, T., Sakai, T. (2020) The global dataset of historical yields for major crops 1981–2016. Sci Data 7, 97, https://doi.org/10.1038/s41597-020-0433-7 ïĆă Siebert, S. & Doll, P. (2010) Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. Journal of Hydrology, 384, 198–217, https://doi.org/10.1016/j.jhydrol.2009.07.031 ïĆă Sloat, L. L., et al. (2020) Climate adaptation by crop migration. Nature Communications, 11, 1243, https://doi.org/10.1038/s41467-020-15076-4 ïĆă Zhang, G, et al. (2017) Spatiotemporal patterns of paddy rice croplands in China and India from 2000 to 2015, Science of The Total Environment, 579, 82-92, https://doi.org/10.1016/j.scitotenv.2016.10.223.

---

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-11, 2020.



Printer-friendly version

Discussion paper