ESSDD

Interactive
comment

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

Patrick Meyfroidt (Referee)

patrick.meyfroidt@uclouvain.be

Received and published: 10 April 2020

************** General comments: **************

This paper presents the latest update of the SPAM global gridded crop maps for 2010. Overall, this is a very valuable effort. Yet, in its current form, I have several general remarks:

1/ The method is insufficiently explained and unclear in some places. A series of expert judgments are used along the way, and although this is acknowledged in the description of the methods, this seems insufficiently acknowledged in the Abstract and Introduction. Overall, this raises concerns about the transparency and reproducibility of the work, but also makes it very unclear what is the same and what is changed

compared to previous versions, thereby justifying a new paper.

2/ Validation: This is a model (mixing reproducible rules and expert judgments), and as such, one would expect more rigorous and transparent validation efforts. Here it appears very thin and, in the words of the Authors themselves, the uncertainty assessment "is not a scientific, rigorous" one.

3/ Beyond operational uses for agencies focusing on crop production, the paper does not discuss how can these efforts serve more broadly scientific agendas regarding an improved understanding of the role of land management in global environmental change, earth system dynamics and other global sustainability issues (e.g., see Erb et al. 2016 in GCB for a discussion)? This would be useful to make the paper more valuable in itself beyond "just" presenting the dataset (no offence here, this is of course a great achievement!).

I return to these main comments below.

************** Methods: **************

* If this is an update with just purely the same methodology, it should not be an extra scientific paper. If there are substantial changes (improvements) in the methodology, then previous validations should not be taken for granted. Here, it is not totally clear what is new and should be validated, versus what is standard.

* Overall, the explanation of the method is unclear in many places. Being familiar with many of these gridded products, but not very much with the previous versions of SPAM in particular, I really have a hard time understanding the approach here. I am dubious that a reader that has not read the previous methods papers can understand what the Authors have done here.

* The method is insufficiently explained: 3.1: The 4 farming systems are explained, but not how the disaggregation between these 4 is done. The answer seems to be actually in Section 4.1.3, but here the answer is essentially "we do it, based on multiple

information and stuff, trust us". Figure 2 is supposed to present an "illustration" ("We present an illustration for obtaining the farming system shares by crop j and administrative unit k (Percent jlk) in Figure 2"), but Figure 2 doesn't give any information on how this disaggregation is done.

* Then Section 3.2 explains the optimization but honestly, I understand the equations but it doesn't allow me to understand the process itself.

* Section 3 does not clarify explicitly what methodological aspects are the same as in previous versions, versus those that have been modified or are new.

* 4.1.3 Crop statistics disaggregated by farming systems: » This seems to be a mix of various approaches. Can you at least clarify the share of cropland disaggregation achieved based on statistics versus some expert knowledge or assumptions?

* p.5: "The rainfed subsistence farming system (S), which is also low input as well, and is introduced to account for situations where cropland and suitable areas do not exist, but farmland is still present in some way." » This is very unclear.

* Accessibility: This comes in Eq. 1 in Section 3.2, and then is detailed in 4.2.5, but I don't understand what is the rationale for creating / using an "accessibility" to market dataset based only on rural population? Is there an assumption that urban populations are fed from anywhere on the planet through global supply chains without this creating any particular incentive for farmers in surroundings (so that only rural population create a revenue incentive as per Section 4.2.6)? This would be quite a strong assumption. What is the rationale behind?

* Overall, there is a lot of expert judgment and wiggling with the data (see S4, S13-15 etc) (e.g., Section 4.3.1 "Under these circumstances, we set several options to "force" a solution, including adjusting the entropy conditions, and adjusting the data harmonization rules. We elaborate on the details for adjusting areas (Section S13), entropy conditions (Section S14) and harmonization rules (Section S15) respectively in

the SI."). So this is far from resulting from a clean and reproducible algorithm based on simple economic rules. I don't want to distrust the work done by the Authors, but given this large amount of expert-driven decisions, this should be very clearly stated in the abstract and main results / Conclusion, so that the reader understands clearly that this is largely an expert-driven process, with multiple human decisions and assumptions, more than a simply reproducible algorithmic work that produces a transparent output.

************** Validation: **************

* Same as above and general comment: First, this is a model; and thus it should be validated properly as far as possible. I understand of course that by the nature of the work done, there is no simple, global, adequate validation data ready to be used. But still, (i) there are ways to do more & better, and (ii) the current efforts are reported in an unclear manner.

* If, as you explain, you run most countries with data at ADM0 level, but you do have incomplete data at finer administrative levels, then you can at least validate against these incomplete subnational data. This is explored in Figure 5 but given the breadth of the map, just one example is not sufficient.

* Partial validation could also be achieved through a sampling of points, with visual interpretation of high-resolution imagery to at least identify irrigated systems versus non-irrigated intermediate categories versus the subsistence category. Even some specific crops could be assessed, at least some perennial crops like oil palm, banana, or others.

* You can't just say (l.539): "As the coverage, quality and spatial precision of data input are much better for SPAM2010 than for its predecessors (see Section 4), the reliability of the data product is believed to improve as well."

* l.548: "Firstly, we evaluate the results by sending the crop maps to collaborators and users alike for comments or assessment. For example, the CGIAR..." » I don't

understand how this is an "example". Either you did it and you report the results, or you explicitly state that this is something that you have not done but could do.

* "We took advantage of their vast network of field offices and local expertise to help us to validate the SPAM results. Many researchers from these institutes have been involved in the production of SPAM2010, which increases the reliability of the results." » If this has been done, then you should report in more details the outcome of this process, the validation data collected...

* "The validating information could either be collected by" » "Could be", or it has been done? If the former, then it's not useful. If the latter, then provide the results.

* "We take these feedbacks and re-run SPAM model and release updated versions of SPAM. The complete validation process could take a great deal of effort and time, but these users' feedbacks are quite important and valuable." » Same, not clear, is this something you plan to do, or something you have done and can provide data about? The use of present tense makes it confusing.

* "The current product, i.e., SPAM2010v1.1, is also expected to have major updates" » Then is it the right time to release it? Wouldn't it be better to have this round of validation - improvement first?

* "Secondly we do a regional validation in case that the third-party independent crop maps are available," » Same, present time: Does that mean you have done it? Or does that mean this is an aspirational goal that at some point you hope you can do it? Here, as you provide the comparison with US data in Figure 5, it appears that this is something that you have actually done. But (i) we have to guess it, and (ii) it's not clear for all the above.

* Figure 8: Differences are huge. I understand that this mixes real changes on the ground and changes in the methods. But over - nominally - 5 years, this appears to be predominantly dues to changes in the methods. Please elaborate further (note, this is

C5

in relation to the above point on Methods, as it is not fully clear what is stable and what has changed in the Methods).

* l.604: "In addition, we collect feedback and comments from users, local experts and collaborators as discussed above. They are sporadic but very useful. We combine all the information together to give a subjective rating on how confidence we, SPAM team, think of our final crop maps (both area and yield). This is the uncertainty rating we provided here. It is not a scientific, rigorous rating and so we put it only into 1 to 5 categories (1 represents the lowest uncertainty, 5 the highest)." » If this is not a "scientific" rating does it belong to a "scientific" paper?

************** Minor comments: **************

* Abstract: I don't understand this sentence: "but also dedicates as platform providing archived global agricultural production maps for better targeting the Sustainable Development Goals by making proper agricultural and rural development policies and investments"

* Overall the writing is good, but there's a series of weird words, typos and stuff like l. 363: "protected areas. but if the" or l.371 "rural pulation density" (just to give examples, there's plenty of these). Please triple-check through.

****

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