

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2021-125-RC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-125**

Anonymous Referee #1

Referee comment on "Harmonized in situ datasets for agricultural land use mapping and monitoring in tropical countries" by Audrey Jolivot et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-125-RC1>, 2021

The paper introduces a dataset of 27000 polygons in 7 countries (9 study areas) with information on land cover, crop type and cropping systems. This information can then be used to validate land use maps or train models for land use classification. It results from a large-scale international initiative supported by GEO and gathers scientists with a long experience in crop mapping with remote sensing data. It is also noteworthy that the 9 study areas represent very diverse agricultural areas.

The paper is clearly written and I advise for publication. My only minor revisions concern the following points :

- In the introduction : similar efforts in northern countries should be mentioned
- In the methods : I think a little bit more details on the « logistics » issues would be interesting (how many people ? how many days of field campaigns ? etc). It is important that the reader understands how difficult it is to do field campaigns in southern countries.
- In the discussion/conclusion : add additional comments on future perspectives :
 - will the dataset will be updated regularly ?
 - how to add reference samples from non-JECAM colleagues ?
 - how to improve the data collection in the future (UAV data, crowdsourcing)?

Other comments are:

In the abstract,

« Altogether, the datasets completed 27 074 polygons (20 257 crop and 6 817 non-crop) documented by detailed keywords. »

Depending on the authorized length of the abstract, it may be good to complete with additional information : how much maximum polygons/year and/or polygons/class and/or polygons/study area?

L61 to 70

The list of reference datasets is interesting. I think it lacks some information on each of them (mainly, I am not sure they all discriminate various cropping systems or just croplands from other LC classes). In addition, I think it would be great to have this information in a table but that would imply to move it out from the introduction section to another section.

In addition, it would be interesting to get more info on such initiatives in northern countries (in Europe and in the US mainly)

Figure 1. It looks like there is no croplands in Europe, US and Australia. Maybe you could add a word on that in the figure caption, to explain it only focuses on developing/emerging countries

L103 . « 60 x 60 km² area ». I am not English but I would say « 60 x 60 km area » (thus removing the ²) : 60 x 60 km = 3600 km². Please have a check.

L103. « commune ». sounds very French. Maybe put in italics ?

L154 : March instead of Mars

L.184. « Field surveys were conducted yearly ». Yearly is not very well chosen since you have only one year of data for a few sites.

L197 « filling » instead of « filing »

Section 2.2. I would be interested in reading more information on the number of colleagues who participated to the data collection and if it was necessary to train local colleagues to collect the data. (This may also appear later, around L 250). The capacity building part is important to ensure future update of the database by local partners.

In Table 2, a fourth column with an example for a given polygon of the database would be welcome

In figure 3, I guess the pasture class mentioned in the description of some study areas are included in the grassland class. Yet I wonder if there is a discrimination between natural grasslands and managed pastures. can you please clarify that point?

L254 . « photographs » instead of « photos »

L255 « Finally, the fact that the same person performed the whole acquisition and processing chain - from waypoint collection to polygon labelling - minimizes errors and contributes to the overall quality of the datasets. »

This point is questionabale. If the operator is not « good », he may repeat the same error N times. More generally, operators working with photointerpetation usually work with a cross-checking protocol to minimize errors. But I think that in your case it is a bit different since the class labelling is done by at least two people (L249) while this step mentioned in that sentence (L255) only regards the geoprocessing part (polygon delineation).

I think you should rephrase slightly to clarify this point.

L270 « In Table 3, are given the type and extent of the zones where are located our JECAM study sites, for both maps.»

I would rephrase as follos : « For both maps, the type and extent of the zones corresponding to our JECAM study sites are given in Table 3 »

Table 3. There is a double parenthesis in line 1.

L289 First, « , » is missing

L303 « valorized » rephrase « used » ?

L303 Tocantins . Do not separate the S.

L314. I guess the citation should be Jolivot et al. (2021)

L321 JECAM (not JEAM)