

This study has presented the ground data collected from nine selected sites in tropical countries within JECAM initiative. The area of site ranges from 250 to 11700 according to the location. This data is collected for either one or two years except Burkina Faso (7 years), Madagascar (4 years) and Brazil (3 years). As it is ground data collection manually – there is not accuracy assessment except set up of standardized procedures to collect data.

Overall, this work has potential to add value in the problem area of availability of ground data for agricultural monitoring and can be considered for publication with major edits. I have major and minor comments as follows:

Major comments

1. Why is the dataset called harmonized although it only depicts to be ground data?
2. What is the clear definition of classes used to label? Author may add these definitions for more clarity and avoid confusion? For example, what is mean by croplands in this study? (do croplands include agroforestry, rangelands and horticultural crops too)
3. For every site and region – summer crop and winter crop seasons are separate – if author is using this term – I would recommend to use it carefully as the seasons although name is same but months are different location-wise. In short , provide definition of winter crop and summer crop – and in the label – add unique name if possible
4. Figure 1 is totally misleading. The selected site cannot possibly represent the entire tropical farming systems. I would suggest to remove it as it is not adding any value. Main important is to highlight the sites and author can do that by showing site zoomed regions rather than misleading with unnecessary presentation.
5. Study site – the explanation of study sites in section 2.1 is repetitive of what's there is in table 1. I would recommend rewriting to avoid repetition.
6. Table 1 name is synthesis of database – which is not quite correct – I would recommend just naming it as “study area description” or related.
7. Table 1 can have additional columns such as season, temperature, major crops, average precipitation etc.
8. Data collection protocol need to provide with more details as it is important step in this data. I would recommend to explain it with sample examples of data points and showing the standardized format along with flowchart if possible.

9. Post processing of data may add many additional errors to the raw data point collection with the provided steps by author. I would suggest to provide more details and explanation about how the manual error were avoided? In short, provide the framework in methodological format. (although – the step is performed by same personnel who did survey – it is not valid explanation of expertise or scientific explanation)
10. In section data records – one of the column is data source – and there are three data sources – As it is important information – I would recommend author to provide a number on how many samples are “0” , “1” and “2” as data is mainly labelled as in situ
11. Detailed explanation on data source or data collection is needed regarding minimum size unit (MMU), labeling strategies, mix land use class and other details related to land use.
12. Total crop samples are ~20,257 and non-crop are related very low – how did you decide this number? Need further explanation on classes, their sampling size, sampling and labeling strategy.
13. What is the overarching goal and novelty in this dataset? I understand it is very important dataset – but author need to add its novelty and goal of research in introduction for more clarity.

Minor comments:

14. Line 40 and 41 – can be split to two sentences to avoid complexity
15. Validation through study cases – is confusing section – is it application of dataset or validation?
16. Overall, writing needs to improve for spelling and grammar – I would recommend professional English proof-reading – I had real difficult time in reading this paper.
17. Title of article contains “JECAM” – which needs to be expanded?
18. Abstract is misleading in many aspects such as – data time from 2013-2020 (which is not true as most of the sites has data from 2 years only). I would recommend author to be careful and precise facts in the abstract for more clarity and description of work.
19. For small field sizes – what was the strategy to collect data – how would it be homogeneous to the data collection strategy/
20. Overall, I would recommend author to improve readability of the article.