Comments from Referee 1:

Thank you for your positive comments and interesting suggestions. This document's intent is to provide point-by-point answers to your remarks, directly proposing, where possible, modifications to the original paper that will be integrated in the revised version. We have worked in particular on:

- mentioning similar efforts in northern countries with examples in Europe and USA
- describing in further details other available LULC reference data products
- providing more information about logistic details for field missions (field team, duration, partner training,…)
- putting forward future perspectives for the proposed database (BD update, non-JECAM data providers, …).

In remainder of the document, lines in **bold** echo your comments for ease of reading, lines in red provide direct answers to your comments, followed in case by proposed modifications to our paper (with new elements in green).

We sincerely hope that these corrections will match your expectations.

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 introduction, we have added several paragraphs which quotes some similar efforts made in northern countries, along with an example on the way these data sets can be used for LC mapping (a reference has also been added).

Land use and land cover (LULC), and their changes, are key information to study and monitor carbon and water cycles, threats to biodiversity, but also to set up land use planning and public policies. In particular, accurate mapping of cropland and associated cropping practices is of primary importance for food security, agricultural and environmental monitoring as well as land management. However, cropland and crop type mapping using Earth observation data is still challenging as it requires large sets of training and validation data, and as the land use (field limits and content) generally changes annually, even seasonally. Large data sets on cropping practices are available in the Global North, mainly thanks to agricultural policies that support annual census and provide tools for the digitization at field level using Very High Resolution remote sensing imagery (e.g. the Land Parcel Identification System designed to implement common agricultural policy in the European Union, or the Cropland

Data Layer of the National Agricultural Statistic Services of the United States Department of Agriculture). Such data sets provide a very large number of annotated surface samples reporting yearly crop types, which can often easily be integrated in reference data sets for land cover mapping systems at the cost of a relatively simple "cleansing and harmonization" procedure (Inglada et al., 2017). Despite the fact that the declarative nature of such annotations makes them error-prone, such "noise" is typically compensated by the large number of available crop type samples. As arguable, no such large scale data base currently exists in most of the developing and emerging countries. Matter of facts, in these countries cropland and crop types can be particularly difficult to map (Waldner et al., 2015) because the fields are often small to medium size (Fritz et al., 2015), the crops are easily confused with natural vegetation and fallows, and cropping systems are typically highly variable in time and space. Each farming system has its own specificities in terms of crop type and composition, field size, cropping calendar, irrigated/rainfed mode and other practices (Bégué et al., 2018). It is thus necessary to adapt the classification approaches (satellite data and algorithms as well as training and validation in situ data) to the large variability of the farming systems in the world (Dixon et al., 2001), and thus to have access to appropriate training data.

- **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 Data collection section, we have added a paragraph with more detailed logistic information about the field team, duration, and partner training.

During a field mission, the team is composed by an agronomist with geoprocessing skills, accompanied by a national researcher or technician with expertise in the local farming systems and a local driver. In some countries (Burkina Faso, Senegal, Madagascar, Kenya), local partners were trained to collect data. The training sessions were carried out directly in situ to be the closer as possible of the reality. The data acquisition duration depends on the visited area: in Brazil (large fields and good road infrastructures), 300 plots can be visited in one day while for other sites (small to very small fields), it is possible to collect between 50 and 150 plots per day (depending on the roads state and fields accessibility). Usually, the mission for a smallholder's site of ~3600km² is one week with around 700 plots visited.

- **In the discussion/conclusion: add additional comments on future perspectives:**

➔ These are indeed very important remarks concerning the potential growth of the database in the future. We will provide a brief discussion on these points in the revised version.

  - **will the dataset will be updated regularly?**

First, recall that data collections have been performed in the framework of different projects and funded initiatives, which constitutes a significant part of our mission as well as of other research and development institutions working on tropical agriculture. In this sense, this paper's aim is also to provide evidence that such independent field efforts can be mutualized to durably contribute to the extension of the database. To date, several field campaigns are already planned on some of the presented sites, and projects are being built which will lead to the inclusion of several new ones.

- **how to add reference samples from non-JECAM colleagues?**

➔ The paper also proposes a set of technical guidelines for potential non-Jecam contributors, which might in turn lead to a significant extension of the geographic extent of the database and hence of its representativity with respect to the diversity of tropical agrosystems. Although no technical solution is proposed yet to facilitate external contributions, which will be pointed out as future work, we are definitely open to such contributions and willing to take care of the technical issues.

- **how to improve the data collection in the future (UAV data, crowdsourcing)?**

➔ As we will mention in the revised introduction (see answer about details on existing land cover reference databases), we consider crowdsourcing a valuable approach for providing very large data sets, at the cost of a larger "noise" due to the loosely controlled quality of the contributions. Our field strategy is more meant to preserve quality at maximum, and provide a less extensive, yet reliable set of field observations. So we consider our approach as being somehow complementary with respect to crowdsourcing initiatives, and no such strategy is planned in the future.

UAVs (drones) can be, and already are in few of our sites, used to validate acquisitions and extend / accelerate the surveys over areas which are less accessible (far from roads, across fields / flooded areas, etc.). However, both the availability of UAVs as well as of the required competences to drive them is generally costly, so that, to date, it is very unlikely for such acquisition means to be systematically included in future campaigns. In some sites, the strong resemblance of some crop types when observed from and airborne sensors also prevents the use of such solutions.

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?

➔ We have added some details about the minimum and the maximum plots visited in a year.

These quality-controlled datasets are distinguished by in situ data collected at field scale by local experts, with precise geographic coordinates, and following a common protocol. Altogether, the datasets completed 27 074 polygons (20 257 crop and 6 817 non-crop, ranging from 748 plots in 2013 (one site visited) to 5515 in 2015 (six sites visited) documented by detailed keywords. These datasets can be used to produce and validate agricultural land use maps in the tropics, but also, to assess the performances and the robustness of classification methods of cropland and crop types/practices in a large range of tropical farming systems. The dataset is available at https://doi.org/10.18167/DVN1/P7OLAP.

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.

➔ Indeed, more information about each product is interesting to understand the lack of agronomic data. We have finally chosen to provide further details on these products in the text (Introduction), since the different nature, approaches and objectives of the different data sets makes it difficult to make a rigorous summary through the use of a table.

At a global and continental scale, initiatives that freely distribute land cover reference datasets exist (see review by Tsendbazar et al. (2015)). The GOFC-GOLD (Global Observation for Forest and Land Cover Dynamics; see http://www.gofcgold.wur.nl/sites/gofcgold_refdataportal.php for further details and access to data) regroups and consolidates existing reference datasets used for the validation of legacy global land cover products (prior to 2015) at moderate spatial resolution (300m-1km) such as GLC 2000 and GlobCover 2005. All referenced databases are provided at global scale, ranging from few hundreds to around 2,000 samples each. Except for GlobCover 2005, which contains a "rainfed cropland" class, other referred LC nomenclatures only contain a single cropland class, sometimes referred to as "cultivated".

Other data collection experiences reached a sensibly higher number of samples through the use of crowdsourcing campaigns, a notable example being the LULC reference dataset presented in Fritz et al., 2017, and its companion work from Laso Bayas et al., 2017b: thanks to the Geo-Wiki tool providing an easy-to-use interface for the photo-interpretation of very high spatial resolution satellite images, it was possible to collect up to 150,000 samples of different LULC classes. This includes over 36,000 cropland locations, distributed over contrasted areas in terms of cropland density. As in the previous case, a single cropland class is referenced in the nomenclature, alone or mixed with natural vegetation ("mosaic" class). Although crowdsourcing confirms as a valuable strategy to collect reference cropland data at larger scales, it still remains unsuited when precise information has to be collected, both spatially (resolution, plot boundaries, etc.) and in terms of crop type nomenclatures. Matter of facts, most of the crowdsourcing initiatives are based on visual image interpretation, which prevents the precise localization and identification of cropping practices. Shifting to a crowdsourced field strategy will not be suitable as well, both because of the very specific agronomic and GIS competences needed and the limited accessibility to cultivated areas in tropical countries.

More lately, the LandCoverNet dataset has been released for the African continent (Alemohammad et al., 2020), with the specific aim to foster the use of recent machine and deep learning approaches for automatic land cover classification. Here, samples are provided in the form of densely annotated image chips (256x256 pixels at 20m resolution) accompanied by the corresponding Sentinel-2 observations over the reference year (2018). A total number of 1,980 fully annotated chips, accounting for more than 30 million of labelled pixels, are provided, spanning 66 tiles of Sentinel-2 over the entire African continent. Although such dataset could allow a finer spatial validation of LULC products at high resolution, it still provides a single “cultivated land” class, making it unsuitable for the assessment of LULC products specifically conceived for the monitoring of agricultural systems.

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

➔ See answer above

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

➔ The caption has been completed as suggested

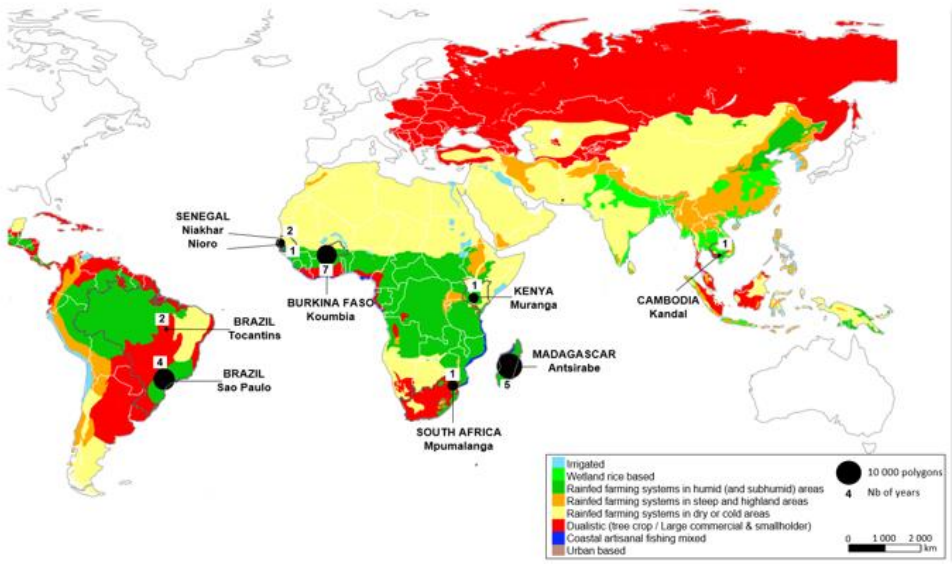

Figure 1. Location map of the study sites, and the associated number of collection years and sampled plots (symbolized by the size of the red circles), displayed on the FAO (broad) farming system map focused on developing / emerging countries (Dixon et al., 2001)

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.

➔ Indeed, it was a shortcut and we have corrected the sentence.

The acquisition protocol is based on the JECAM guidelines (Defourny et al., 2014) with adaptations to consider some characteristics of tropical agriculture (mainly small field size and accessibility). Field surveys were conducted at least once in each study zone, with several sites revisited over multiple consecutive years (up to 7 for the Burkina Faso site). Campaigns took place either around the growing peak of the cropping season, for the sites with a main growing season linked to the rainy season such as Burkina Faso, or seasonally, for the sites with multiple cropping (e.g. São Paulo site). Except for Senegal where a stratified sampling plan for field surveys was used (Ndao et al., 2021), the GPS waypoints were gathered following an opportunistic sampling approach (called the "windshield survey") along the roads or tracks according to their accessibility (that can be difficult during the rainy season, leading to less surveys in secondary roads or tracks in some study areas), while ensuring the best representativity of the existing cropping systems in place (Defourny et al., 2014; Waldner et al., 2019).

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.

➔ See answer above

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

➔ We have added a 4th column with an example

Attribute Name	Data Type	Description / available arguments	Example
Id	Numeric	Unique ID	*26413*
Country	Text	Country name	*Burkina Faso*
SiteName	Text	Site name (generally related to the biggest city around or to the region name)	*Koumbia*
DataSource	Numeric	Discrimination between land uses acquired from *in situ* surveys or satellite image CAPI (computer assisted photointerpretation) 0: Land use from *in situ* survey 1: Land use from satellite image interpretation 2: Land use from aircraft observation	*0*
AcquiDate*	Date	*In situ* survey acquisition date or satellite image acquisition date (when the land use is photointerpreted, see "DataSource" attribute) – Format: yyyy-mm-dd	*2020-10-21*
LandCover	Text	Land cover of the polygon. If value is "Cropland", see CropType 1, 2 and 3 attributes for more information	*Cropland*
CropType1	Text	Main crop type of the polygon	*Cotton*

CropType2	Text	Secondary crop type of the polygon (in case of intercropping)	*Maize*
CropType3	Text	Tertiary crop type of the polygon (in case of intercropping)	*NULL*
SOS*	Date	Start of season date in the site (if empty, this means that no specific season exists in the study area) – Format: yyyy-mm-dd	*2020-05-01*
EOS*	Date	End of season date in the site (if empty, this means that no specific season exists in the study area) – Format: yyyy-mm-dd	*2020-11-30*
Irrigated	Numeric	Presence/absence of an irrigation system 0: No information available 1: Rainfed 2: Irrigated Empty: For polygons other than cropland	*1*
Intercrop	Numeric	Presence/absence of intercropping 0: Single crop 1: Mixed crop or row inter-crop 2: Agroforestry Empty: For polygons other than cropland	*1*
Weeding	Numeric	Presence/absence of weeds 0: No information available 1: Presence of weeds Empty: For polygons other than cropland	*0*
Area_ha	Numeric	Polygon area in hectares	*0.446*
KeyWords	Text	Set of terms associated to the land use of the polygon (separated by semicolons ";")	*Agricultural land ; Cropland ; Arable land ; Temporary crop ; Cash crop ; Fiber crop*

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?

➔ Indeed, Pasture should be overally considered as a particular type of Grassland cover whose vegetation dynamic may be affected by the regular passage of livestock, and should not be confused with cultivated pastures for which the database eventually reports the specific crop type. In order to preserve at maximum the information about the possible use of grassland for animal feeding, our global strategy across the different sites is to note grassland areas as pastures whenever such use can be assumed with certainty, either from prior knowledge on the visited areas or by the direct observation of the presence of livestock.

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.

➔ Again, this phrase resumes too briefly the concept of per-site field supervision, which is actually set up in a more structured way in order to ensure a reliable quality checking. In order to clarify this point, we have added a short paragraph in the Quality checking section, to explain that each site has a referee person who supervises the whole field collection chain, whose different steps, from the acquisition to the integration in the database, are generally performed by a dedicated specialist.

4.1 Quality Checking

Finally, each site has a referee person who knows very well the area. He supervises all the chain from the data collection to the database integration. Following this approach, each step is generally conducted by one or more “discipline” specialists (agronomy, GIS, database) whose work is coordinated by the referee in order to minimize errors and contribute to the overall quality of the data sets.

For the other comments:
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
L197 « filling » instead of « filing »
L254 . « photographs » instead of « photos »
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)

➔ All these remarks will be corrected in the revised text. Thanks very much again for your valuable implication and comments.