We have answered point-by-point to all the comments raised by the referee in the attached PDF. The answers are provided below each comment in blue font. We would like to thank the reviewer for the quality and detailed reviewing provided. We would also like to notify the referee that this manuscript was selected as an Executable Research Compendia (ERC) pilot (see here https://o2r.info/pilots/) and the code to process the data and the table and figures will be available and reproducible after reviewing.

Referee Comment 1

General comments This paper presents a unique dataset of major interest for the scientific community of EO in the EU. The protocol is rigourous which make me trust the quality of the dataset. I made however a few comments that could be worth to consider for the future collection. Concerning the paper itself, I am not a native speaker so I cannot judge the language, but the complexity of the embedded datasets made it sometimes difficult to follow and some small changes in the structure would, in my opinion, make it more readable. For the sake of completeness, I suggest to add the full description (including threshold that seprate "mixed" classes) of the land cover classes in the annex of the paper.

Specific comments

The title only mentions the land cover component, then the paper describes both land cover and land use attribute. The paper could focus on Land Cover only because it is already quite complex. Otherwise more details about Land use is necessary (e.g. what if the land use extent is not compatible with the land cover extent ?)

This comment is valid and the title was modified to include “land use” especially as the land use extent is compatible with the land cover.

Line 49 (and after): it is not clear to me what is referred to by "EO limitation". From this paragraph, I was expecting limitation of the LUCAS dataset in order to be used by in EO workflow, but the three limitation are presented as typical shortcomings in "operationnal EO projects". I recommand to first focus on the reasons why in situ data in necessary for EO, then talk about the inherent shortcomings of a dataset designed to collect statistics (the latter is not explained IMHO) and that need to be addressed in the Copernicus module. Further details are given in section 3, but I think that this information is relevant in the introduction.

We thank the reviewer for this remark and have rephrased this section of the manuscript as the motivation of having a Copernicus module. The inherent limitation of the classical LUCAS for EO are described only in section 3.

Line 85: so ~1/3rd of the points of the grid have been surveyed in 2018 ? And they are selected according to a land cover based stratification ? Please clarify how you end up with 337854 points.

Indeed, the points were selected based on a land cover stratification. The stratification methodology is described in detail in “Redesign sample for Land Use/Cover Area frame

Survey (LUCAS) 2018” (https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-TC-18-006). This reference was missing and thus was added in the manuscript. We have also rephrased this section to increase readability.

About the resulting number of points (337854), it is also described in the document. To summarize, a sample can be defined as optimal both in terms of its costs (i.e. the number of units to be interviewed) and its accuracy (related to the sampling variance of target estimates) (Ballin, M. and Barcaroli, G., 2013). The approach followed in the optimisation process of LUCAS sampling design is based on the joint determination of the optimal stratification of a sampling frame, together with the optimal sample size determination and allocation. This approach is the most general one, as it can operate in the full multivariate case and its implementation is based on the use of the genetic algorithm.

Ballin, M. and Barcaroli, G., 2013. Joint determination of optimal stratification and sample allocation using genetic algorithm. Survey Methodology, 39(2), pp.369-393.

Line 95: How were the 90620 points selected ? random or stratified sampling ?

The sample of the Copernicus module was a third-phase sampling nested in the two-phase LUCAS core sampling scheme. This was clarified in the manuscript.

Line 120: what is meant by "exact location" of the observation ? If not defined by the 2km*2km grid, how is the point identified on the ground ? Is there a mark ? What is the precision of the geolocation (centimetric ? decimetric ?). What is the precision of the distance measurement from the point.

The terminology “exact location” is not accurate. By this, we mean, the gps-measured location. The surveyor is aiming to reach the grid point but there is no mark on the ground. To locate it, the surveyor is using both a GPS and a ground document with the points displayed with an orthophoto in the background. The precision of the GPS varies according to devices, area and time but is recorded and provided along with the survey data. To avoid confusion, we have replaced the term “exact location” by “measured location”.

How are the cardinal directions determined ? is it the geographic North, cartographic North or the magnetic North? For future work, I would suggest UTM north with a 45◦ shift (NE, NW, SE, SW) to be as close as possible to the standard Sentinel-2 grid (I see on line 156 that LAEA is used to build the polygons, but for me the polygons should be created in UTM then projected in LAEA).

A traditional magnetic compass has to be used to indicate a particular direction. North is indicated on both the topographic maps and the orthophotos. With the help of the compass, the surveyor can correctly orient the map and the orthophoto before examining them. Moreover, the compass is helpful to correctly take the photos in the cardinal directions (N, E, S, W).

We thank the reviewer for the suggestion. Indeed LAEA was used here and the authors will suggest to use UTM for the next survey.

What is the minimum mapping unit of the distance estimate ? On figure 1 the polygon is obviously crop, but there are areas of bare soil in this crop field. My example is trivial, but what if there is a small shrub in a grassland or when is a gap in a forest considered as "not a tree". This needs to be specified in order to be used appropriately with respect to the spatial resolution of the EO data. The issue of heterogeneity at different spatial resolutions should be discussed at the end of the paper because it could have an impact on some studies.

For the Copernicus module, one of the conditions to do the Copernicus survey is that " it is necessary that the extent of the Copernicus LC is at least 5m in any possible direction.". Therefore, the MMU of copernicus is 78.53 m² (i.e. a circle of 5 m radius).

For the core points, the LUCAS minimum mapping unit is about 7 m² (a circle of 1.5 m radius) for points falling in a homogeneous area. When the land cover is not homogeneous, for example when it is composed of trees or shrubs interspersed with grass, the scale of observation has to be changed to classify it. In these cases a systematic observation of the "environment" in the vicinity of the point, which in LUCAS is called the extended window of observation, has to be adopted. The extended window of observation expands to a radius of 20 meters of distance (or 40 m diameter) from the point (representing an area of 0.13 ha).The window of observation has to be extended whenever the land cover at the point is heterogeneous. This occurs regularly in areas such as:
- permanent crops (B7X, B8X, except nurseries B83): parcels of permanent crops where the trees or other plants alternate with bare soils and/or grassland or another crop
- woodland (CXX)
- shrub land (DXX): where a mix of e.g. shrubs and trees might occur
- grassland (EXX), where land features may alternate (e.g. grassland with trees)
- bare land (FXX)
- wetland (HXX)

Line 131: what if there are more than one land covers (depending on the direction)

See the next question for an answer..

Line 159: the homogeneity of the quadrilaterals is not guaranteed by the protocol. In the (unlikely) event that an item from another land cover is located inside the equilateral, is there a protocol to reduce the radial distance in one or two directions in order to exclude it ?

Specifically for the Copernicus module, the protocol reduces the radial distance is case of another land cover (note that Linear features with width < 3m are not considered a land cover change). The detailed field collection protocol of the Copernicus module including specific cases is available pp.57-67 of the C1 documents (https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C1-Instructions.pdf)

Line 164: do you mean that the point was unreachable ? With a 2 by 2 km grid, 11 km is really far away.

No, the aim of this filtering is to prevent potential encoding problems such as decimal or sign problems and thus remove macro error coming from encoding.

technical correction

Lines 1 and 20: Please replace "regular" by "evenly spaced" (or any more precise word), because regular could also be related to the temporal dimension.

Done.

Line 22: please clarify if there are 1351 293 unique geographic points point or if that number is the number of records (up to 5 records at each location)

The text was clarified by adding “at 651780 unique locations” indicating that the 1351 293 observations could be done at the same point for different years.

Line 83: what is the purpose of the stratification if the sample is systematic ?

The

Line 96: 69.02 should be 69.92%

Modified.

Line 97: I think that planned is better that programmed

Modified.

Line 111 : double "with"

Corrected.

Line 177 : because of the orientation of the quadrilateral, I think that less than 50 pixels are fully included.

The text was clarified “corresponding thus to almost 50 10-m pixels depending on the orientation”.

Line 178 : fo "subdecametric" sensors, the MMU becomes important.

As explained in the previous MMU comment, they are 3 different MMU according to which data are selected:

- Copernicus module: 78.5 m² (a circle of 5 m radius)
- LUCAS points homogenous area :7 m² (a circle of 1.5 m radius)
- LUCAS points homogenous area :1256.6 m² (a circle of 20 m radius)

This was detailed in the manuscript.

Figure 4: what is the coordinate system of the map (Equirectangular ?) ? Please note that the orientation of the polygons is quite different from the cartographic North (LAEA effect ?)

The Fig is in EPSG:3857. Indeed, the polygon was created in LAEA and is then reprojected.

Referee Comment 2

In addition to my previous comments, I suggest to add the sampling probability of each point in an additionnal attribute field. This is indeed necessary for an appropriate use of the point in a statistically rigourous accuracy assessment.

This is a good suggestion and will be taken up for the LUCAS 2022 Copernicus campaign. The LUCAS Copernicus sampling was a third phase in the two phase LUCAS core sampling scheme. In addition, there was an additional requirement that the points sampled for the LUCAS grassland module (3500), and the LUCAS soil module (xxxxx), were also covered by the LUCAS Copernicus module. The stratification methodology is described in detail in “Redesign sample for Land Use/Cover Area frame Survey (LUCAS) 2018” (https://ec.europa.eu/eurostat/web/products-statistical-working-papers/-/KS-TC-18-006). The authors will pursue this in a forthcoming update.