Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-178-RC1, 2020

Earth System Science Data Discussions

Interactive comment on "LUCAS Copernicus 2018: Earth Observation relevant in-situ data on land cover throughout the European Union" *by* Raphaël d'Andrimont et al.

Radoux Julien (Referee)

julien.radoux@uclouvain.be

Received and published: 14 September 2020

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 ?)

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.

Line 85: so $\sim$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.

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

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.

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).

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.

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

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 ?

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

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.

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)

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

Line 96: 69.02 should be 69.92%

Line 97: I think that planned is better that programmed

Line 111 : double "with"

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

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

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 ?)

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-178, 2020.