

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2022-300-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-300**

Anonymous Referee #1

Referee comment on "The first map of crop sequence types in Europe over 2012–2018" by Rémy Ballot et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-300-RC1>, 2022

Major comments

The manuscript proposes an interesting approach to map the crop rotation at EU-28 level by reconciling LUCAS surveys happening every three years and annual dataset from farmer's declaration. While the approach is promising and the authors provided interesting statements along with open source code, major improvements and clarifications are needed before being considered for publication.

First of all, there is an inherent flaw in determining crop rotation from a 3 year (LUCAS) observation set and then comparing it to a subset (France) of annual rotations, and extrapolate that comparison to the EU. For instance, a common 3 year rotation like cereals - sugar beet - potatoes (e.g. in NW Europe) will only appear as a sequence of a single crop class in LUCAS, while being explicit in the annual France data set. This can only be resolved (for France) by comparing 3-year cycles starting each year (e.g. 2012 - 2015 - 2018, 2013 - 2016 - 2019 and 2014 - 2017 - 2020) and then aggregate co-located sequences into a LUCAS-like indicator. There is no discussion on this anywhere in the paper. This should be deeply clarified and discussed.

The groups of crop sequences (crop sequence types) identified by the PCA should be described in detail (which crops, how many of each of them) in the method part. This comes only in the result part (Table 2) and thus makes the understanding of the follow up of the methods very difficult as this result of the clustering is not described properly before.

Another issue is the decision of the authors to include permanent grassland (E20) in the temporary grassland class. Especially as in the results they show that this is the predominant class (Table 2) and that "LUCAS dataset overestimate the proportion of the "grasslands" crop sequence". This should be clarified in the manuscript.

RPG is a so-called block system (a particular implementation of a LPIS for use in the CAP). A block can enclose several (1 or more) agricultural parcels with different crop types. Thus, an RPG parcel does not necessarily link to a single crop type. In earlier RPG use (e.g. 2012) the agricultural parcel was only indicated as an area estimate without specific geolocalization within the RPG parcel. Also, RPG has undergone significant change in this period, replacing the use of outdated cadastral parcels with a block system. This progress is regionally specific in France. This all makes spatially and temporally consistent point comparison difficult. It is relatively easy to highlight such issues in a graphical analysis. It is essential to include this analysis in the paper and discuss possible impact.

Crop attributes in RPG are not the same as (aggregated) crop class in the LUCAS nomenclature. The paper states that more than 300 crop types are included in the most recent RPG. There was likely some translation step to map 300 crop types into the N LUCAS groups. This should be explained. The problem with harmonizing at EU level is not so much related to spatial or temporal coherence (all national sets are large scale ortho-photo based, full territory and annual) but with consolidating the parcel attributes (reference to TUM effort).

Farmer declared crop attributes are not 100% correct (otherwise, there would be no need for controls). A key factor is the quality of the LPIS and the registration process. While that quality has improved significantly in the period 2012 - 2018, it is still prudent to expect a 2-4% "material error" in the data even in 2018. Again, it would be important to understand how errors propagate into the rotation pattern results.

In section 3.2, relative importance of crops are compared between LUCAS and LPIS. LUCAS is designed to be used as a regressor estimator of area at NUTS2 level. Therefore such comparison of looking at the distribution of the occurrence of LUCAS points VS LPIS has a lot of caveats. It would be better to compare the area from the LPIS with the area estimates from LUCAS point.

We encourage the authors to make the improvements proposed to improve their interesting manuscript and we are providing further minor comments below.

Minor comments

Language use is somewhat peculiar. Native English review would be beneficial.

Montenegro is not an EU Member State. Remove all references to it, including the discussion on the low LUCAS point density.

Abstract and line 21: use of "essential linchpin" is odd. Probably "key element" is meant. It is "assumed to be a key element", there is not so much evidence that it actually is.

L 10: "temporally-incomplete" maybe mention here the LUCAS data years (2012,2015 and 2018) that were used in the study?

L19: The Zenodo link (<https://doi.org/10.5281/zenodo.7016986>) provides only a png low quality map of the points and a CSV table. I would have expected to have a georeferenced dataset.

L 21: Crop *diversification* is the process (action) that leads to crop *diversity* (status). Check the review paper by Hufnagel et al., 2020. They note that *diversification* is interpreted and defined differently in the scientific literature. Hufnagel et al define: "Crop diversification can be considered as an attempt to increase the diversity of crops through, e.g. crop rotation, multiple cropping or intercropping compared to specialized farming with the aim to improve the productivity, stability and delivery of ecosystem services". In the scientific literature, the most studied aspect of diversification is crop rotation/intercropping.

L31-34: Maybe add also this reference from Bohan et al. "Designing farmer-acceptable rotations that assure ecosystem service provision in the face of climate change"

<https://www.sciencedirect.com/science/article/abs/pii/S0065250421000027>

L 43: harmonization of nomenclature is another issue

L 48: "European" should be changed to "EU"

L 51-52: Suggestion to rephrase and to add sentence: "The purpose of this work is to map current dominant crop sequences from the European Land Use Cover Area frame statistical Survey (LUCAS). For this study, the multi-year harmonised data by d'Andrimont et al. (2020) were used."

It is important to clarify that LUCAS is carried out by EUROSTAT and what you have used is the multi year harmonised dataset.

L 54: "preprocessed" should be changed by "filtered"

L 66: "from" should be "since"

L 66: EU-28, Europe, European Union. Different wording in the manuscript. I would suggest using EU-28 everywhere in the manuscript to avoid confusion.

L 74 : "over the period 2012-2018" The RPG Explorer Crop provides 2007-2019 according to <https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.15454/XH84QB>

L 77 : This section should be more detailed and should provide reference to the Table 1

L 83: six-year (not seven). Also, I would not say "the three most recent campaigns" as there was a campaign this year in 2022.

L 85 : The following statement “and thus we do not consider older campaigns (i.e. 2006 and 2009) which may be outdated to represent current crop sequences” is not exactly the reason you explain afterwards at the end of the paragraph, I would rephrase for the sake of consistency.

L 101: “the temporal frequencies over three years” is non-sense. The LUCAS time span is 6 years, with only 3 observations. What you really do is tabulate for each point the crop (group) occurrence sequence and then map the spatial pattern for the 8 most important ones. The key weakness in this approach is that different sequences may actually relate to the same 3 year rotation (e.g. as in the example above: if one point shows wheat - wheat - wheat and the next closest point to it: potato - potato - potato, it may be exactly the same rotation, but just shifted one year). This is why the comparison to annual rotation in France is impossible to extrapolate to LUCAS-based rotation in the EU.

L 105-18: presenting this as a table would ease the reading.

L 108: “(viii) temporary grassland (LC B53, B55, and E20)” E20 is NOT a temporary grasslands. Check here p.58 <https://ec.europa.eu/eurostat/documents/205002/8072634/LUCAS2018-C3-Classification.pdf> . “This class excludes Temporary grassland and fodder crops (B5X)”. This has to be clarified

L 113-116: The groups of crop sequences (crop sequence types) identified by the PCA should be described in more detail (which crops, how many of each of them)... This comes only in the result part (Table 2) and thus makes the understanding of the follow up of the methods very difficult.

L 120: the use of an RF model is poorly understood. The LUCAS sequences are directly

given by the observation data, there is no need for any model to predict those.

L 123-224: applying the RF model trained on 3 year LUCAS observations to annual sequences is even less comprehensible, as the input to the RF inference is not of the same nature as the training. This is a very weak part of the paper, which would normally lead to rejection, as all the subsequent "patterns" are derived from it. As suggested above, the spatial pattern comparison between LUCAS vs France data should be derived from France data that is also spaced in 3 year intervals. This would also help to resolve how dominant regular rotations (e.g. 2, 3, 4 or 5 year rotation) are. Longer rotations are potentially of more interest in a biodiversity context (which is the "linchpin" etc.).

In light of the weakness of the approach, the results are likely less meaningful than proclaimed in the manuscript. They are more likely derived directly from the 3 year LUCAS (grouped) observation set, without the need for the RF model and "validation" with an annual data set. For sure, there is currently little added value in the extrapolated results from the France set. This simplification could probably be addressed in a major revision of the paper.

L165: statement "LUCAS dataset overestimate the proportion of the "grasslands" crop sequence." should be reconsidered when taking into account temporary/permanent grassland properly in LUCAS.

L 380: Figure 1. Typo. "Average" and not "Avergae".