Review of: “GMIE-100: a global maximum irrigation extent and irrigation type dataset derived through irrigation performance during drought stress and machine learning method.”

essd-2023-536

General comments

This article describes a dataset purporting to describe maximum irrigation extent and irrigation type with global scope at a 100-metre resolution. This dataset would have broad applicability for agricultural, economic and other analyses at global and more localised levels.

The authors make an attempt at providing this dataset at such a refined resolution, however there are some fundamental issues that need to be addressed before it could actually deliver what the authors promise in the article. I believe currently the authors give a flawed sense of accuracy in their estimates of irrigated and non-irrigated land. In its current form I do not recommend this manuscript/dataset be accepted for publication in ESSD.

Major comments

1. **Areas and cropland definition:** This dataset/manuscript needs better clarification of what areas of irrigated and non-irrigated land are included. For instance the title suggests the dataset is global, implying all irrigated and non-irrigated land are included. In the abstract they state ‘In our study, we present a robust methodology that leverages irrigation performance during drought stress as an indicator of crop productivity and water consumption to identify global irrigated *cropland.*’ The latter implies it includes only cropland. Cropland has different definitions to different authors (see Tubiello et al 2023: https://www.nature.com/articles/s43016-022-00667-9) and can be very tricky to differentiate properly. In section 2.3 the authors state they use the JECAM definition of cropland which includes land used for seasonal crops (sowed/planted and harvested at least once within the 12 months) such as cereals, root and tuber crops, oil crops as well as economically significant crops like sugar, vegetables, and cotton. Additionally land occupied by greenhouses was considered as cropland. Greenhouses in cropland is a strange inclusion and needs explaining. The authors then go on to say they used “The cropland mask at 30-meter resolution could be obtained from International Research Center of Big Data for Sustainable Development goals via https://data.casearth.cn/thematic/cbas_2022/158”. They state the overall accuracy of this dataset is 89.4%, but when I look at maps from these data it appears as though they include a lot of non-cropland area esp. pasture and meadow land (see Fig 1 below). I therefore do not have confidence that this dataset is suitable for supporting the authors assertion that their dataset has 100 metre resolution. Furthermore, the title of this manuscript implies this dataset is for ‘maximum irrigation extent’ i.e. all irrigation. They assess centre pivot irrigation, but it is not clear if the authors include lateral irrigators which is much the same technology as centre pivot, only it could be harder to distinguish lateral irrigation due to the patterns of NDVI (see figure 12). Finally, as per section 2.1 the research relied on evapotranspiration data at a 500 m resolution. Shouldn’t the authors state that the resolution of their irrigation dataset is equivalent to the lowest

resolution of their input data? Otherwise you are giving a false sense of accuracy.

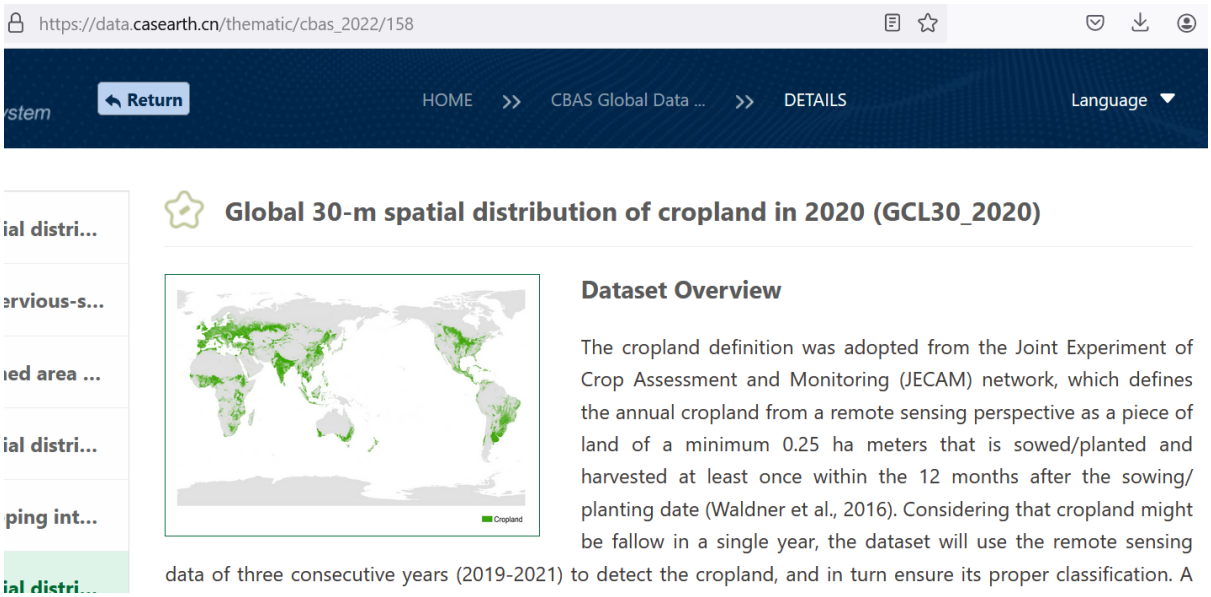

Figure 1: Screenshot of casearth.cn datset.

2. Given the above uncertainties in cropland categorisation I suggest the authors use a definition of cropland that aligns to something like that used by the FAO. This will improve the applicability of the dataset.
3. This manuscript needs to be edited heavily before it is resubmitted. I made a note of some of these edits in minor comments in the first few pages. Note, the list I provide is not exhaustive as there were many other changes to make.
4. Lines 134-140. A better plain language description of how irrigated and non-irrigated land was categorised is needed.
5. Section 3.4. The uncertainty in estimates of cropland used in the authors models needs to be better explained. Differences in classification of 'cropland' for instance can contribute to variation in estimates in irrigated cropland mentioned in section 3.1.

Minor comments

1. Abstract 1st line 11. "primary sector of human water…"; Use other word than sector such as form.
2. Line 26: What is the DL method? Define when you first use an abbreviation.
3. Line 27: What is Pivot-Net?
4. Line 29: "The GMIE-100 dataset containing both or irrigated extent…". What does the both relate to?
5. Line 40 use reference to back up claim that highest resolution maps are 500m to 10km.
6. Line 60 use space between croplands and (Thenkabail et al 2009)
7. Line 106. Use reference to back up claim of 80% efficiency.
8. Throughout references and tables, make sure abbreviations are defined esp in title of Figure 1 and 2.
9. Line 175. What is GVG?
10. Line 251. Spelling mistakes in Nirrgated and Nnon-irrgated.
11. Line 268. Spelling mistake exemple.
12. Line 377: belt_Mexican coastal plain. Error.

13. Line 469-471: How does looking at if an area of land has been cultivated during the driest month over a span of three year help determine if it is irrigated land? What if the cultivation occurs in one of the regular wet seasons of the year but irrigated is still needed thereafter?