

Comment on essd-2022-336

Andy Nelson (Referee)

Referee comment on "Estimating local agricultural gross domestic product (AgGDP) across the world" by Yating Ru et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-336-RC3>, 2022

1 The manuscript:

1.1 Are the data and methods presented new?

The AgGDP dataset is unique. The methods used to generate the dataset are not new per se, but it is the combination of diverse input data and the use of multiple methods to generate AgGDP that is new. I commend the authors for taking on the challenge of producing this important dataset

1.2 Is there any potential of the data being useful in the future?

Yes I am sure the layer will be used in many future global modelling exercises that require detailed agricultural economic activity/output information. See comments in the data quality section that would aid future users in their understanding and usage of the dataset.

1.3 Are methods and materials described in sufficient detail?

The methods and data section is comprehensive.

I recommend some discussion on the impact of choices/assumptions such as those made on line 144. This is just one example, other assumptions should also be addressed in the

discussion.

1.4 Are any references/citations to other data sets or articles missing or inappropriate?

I did not miss anything.

1.5 Is the article itself appropriate to support the publication of a data set?

Yes with modification following the recommendations below.

2 The data quality:

2.1 Is the data set accessible via the given identifier?

The data is accessible at the following location
<https://datacatalog.worldbank.org/search/dataset/0061507>

2.2 Is the data set complete?

The data and metadata are incomplete

- The metadata is quite sparse (perhaps limited by the WB Data Catalog format)
- The units are not mentioned in the metadata.
- Why is the dataset floating point? This level of precision seems unjustified – is it float because integer formats cannot deal with the large range of values? Even so, reporting GDP in USD to decimal places seems unjustified.
- The metadata does not link back to the preprint.
- The *first published date* is after the last updated date – please check and correct.
- Sea areas where no AgGDP data is possible (because marine-based AgGDP is allocated to land) and territories where there is no data available (due to a lack of data for now) are treated the same – this is not very elegant. Consider using different pixels values to distinguish these two "no data" types.

2.3 Are error estimates and sources of errors given (and discussed in the article)?

No. There are no error estimates or validation data in the dataset, though they are discussed in the article. See below for comments on validation and sensitivity analysis.

2.4 Are the accuracy, calibration, processing, etc. state of the art?

The pre-processing steps to generate the AgGDP dataset are appropriate. The section on uncertainties in each input layer and how those uncertainties may be compounded when they are combined is rather brief.

Some quantification via a sensitivity analysis would be a welcome addition to the paper. The lack of a quantitative uncertainty assessment is a weakness of the paper and the dataset in the absence of a robust validation.

2.5 Are common standards used for comparison?

The AgGDP dataset is correlated with night time lights (section 2.5). This somewhat contradicts the introduction that states that night timelights are not always a good indicator of agricultural economic activity. The choice is not well justified. I understand that the AgGDP dataset is unique and depends on many input datasets on production value (limiting the availability of possible datasets against which to validate), but I would like to see a much better choice of validation/comparison with a strong justification too.

The statements on lines 263 onwards are not sufficient indicators of quality. This section can be substantiated by reference to national studies that have also spatially decomposed AgGDP or similar measures of agricultural economic output.

The correlation table (Table 2) is a very high-level aggregation, which does not reflect the highly spatially disaggregated AgGDP and Night Lights data. It is not very convincing or useful.

The validation section seems to compare this cross-entropy model against spatial allocation model based on rural population. The description of the rural population based comparison dataset is not sufficient for a reader to fully understand what it is, how it was made and thus what is being compared against what. I assume that national and subnational Ag GDP is disaggregated based on head count giving every rural person an

equal share of the AgGDP? But I am guessing.

Either way, the validation is a case of comparing one model against another with the argument that the assumptions in one model are more valid than those in another. This is not a very satisfactory validation and I am not sure what message is intended by showing the two have different degrees of correlation in different parts of the world.

Why exclude areas from the analysis with values that are less than 200,000 (USD?) ?. No summary or regional statistics on the correlation are provided. This correlation map and the rural per capita GDP should be provided as spatial datasets with the AgGDP dataset.

2.6 Is the data set significant – unique, useful, and complete?

From my perspective the AgGDP dataset is unique.

I have given recommendations above to make it both useful and complete. In addition to that, I recommend that a table of the production values (priors) per country and the collated GDP data would be very valuable additions to the dataset. Where appropriate these layers should also be provided in spatial data formats. This would help users understand the spatial patterns and artifacts in the AgGDP (alloc.tif) dataset and help ensure appropriate use

3 Article and data set:

3.1 Are there any inconsistencies within these, implausible assertions or data, or noticeable problems which would suggest the data are erroneous (or worse). If possible, apply tests (e.g. statistics). Unusual formats or other circumstances which impede such tests in your discipline may raise suspicion.

I note the edge effects above which could give potential users pause for thought before using this data

The authors could subnational representations of the data, both spatial and tabular which would make it easier to assess whether there are any noticeable problems due to modelling or assumptions.

3.2 Is the data set itself of high quality?

The effort is impressive; detailed estimates of Ag GDP are valuable.

The datasets value is detracted from by the lack of validation data, spatial artifacts and the presentation of the data (level of precision is not justified, file name choice, sparse metadata, lack of tabular summaries).

4 Presentation quality:

4.1 Is the data set usable in its current format and size?

The format is suitable for use in both open and proprietary GIS software and can be easily read in open source software such as R or Python for statistical analysis

There are edge effects in northern latitudes and on some country and subnational boundaries that are rather inelegant – can these be dealt with better?

Recommend to change the file name from alloc.tif to something more meaningful

Recommend the dataset is converted to integer not float – both to reduce size and to be more realistic about the precision of the GDP estimates. Estimates could even be rounded up to the nearest 1000 USD.

Recommend that the different types of no data are treated differently

4.2 Are the formal metadata appropriate?

See previous comments on metadata completeness – this needs to be addressed.

5 The publication:

5.1 is the length of the article appropriate?

Length is fine, but more space can be given to (i) a more robust validation or (ii) a sensitivity analysis to understand the impact of choices in methodology and/or the contribution of the uncertainties in the input layers.

5.2 Is the overall structure of the article well structured and clear?

Structure is fine though the natural hazards component seems like an add-on that does not add much value to the paper and dataset, which is really about AgGDP. The hazards part is one of many possible applications. Is it essential to the paper to focus on one use case like this? If a use case is a requirement of the journal then fair enough.

Starting the conclusions section with a paragraph on hazards is a curious choice given that this is not the core purpose of the paper. Again if this is a requirement of the journal then fair enough.

5.3 Is the language consistent and precise?

The language would benefit from professional English editing. The text is largely understandable but many lines in the text jar due to non-standard English. This reduces the readability. I had to pause and re-read some lines several times, e.g., lines 17 and 18. The dataset and documentation is an extremely valuable resource and I commend the author's effort for developing it; please bring the text up to the same level of value as the data.

Line 8

The paper estimates the exposure of areas with at least one extreme drought during 2000 to 2009 to agricultural GDP is an estimated US\$432 billion of agricultural GDP circa 2010, where nearly 1.2 billion people live.

Alternative We estimate that US\$432 billion of agricultural GDP (circa 2010) was exposed to at least one extreme drought during 2000-9.

If hazard exposure is important, consider adding it to the title.

Line 2 of the abstract is hard to parse

Line 6 – consistency needed – either small “a” on agricultural GDP throughout the paper or capital A.

Line 12 remove “the” , same in line 15, same in line 92 and many other instances of non-standard use of the definite article Check and correct throughout the text.

Line 15 – location variation in what?

Line 15 – the possible implications of the mismatch are not clear

Line 72 –AgGDP not agricultural GDP – check paper that this abbreviation is used henceforth.

Line 79 – what does efforts varied mean?

Line 93/94 is the repetition necessary? Aim to be concise.

Line 118 – clarify the pixel areas. Is this land area or simply the total area of each 5 min pixel? Depends on how the densities were computed, but this is not clear from the paper.

Line 125 – first sentence seems superfluous. Also was the start of civilisation really the first use of forest resources? What about hunter/gatherer societies?

Some statements are superfluous and can be removed. For example

Line 270 The correlation of AgGDP with night light varies across world regions as it

requires areas to emit light (Table 2).

Line 279 The exposure to drought is not uniform across the world.

Lines 303-304 are more or less repeated in lines 317-318.

5.4 Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

Yes

5.5 Are figures and tables correct and of high quality?

Maps are clear.

Figure captions do not need to start with "This map..." Just state what the figure shows.

Recommend an Equal Area projection and remove the E and N coordinates and graticules—they do not add useful information.

Check capital letter usage in map legend title; production instead of Production

Table 1 is shown before it is referenced in the text. Check capital letter usage in column headings in tables. Table 1 caption is not self explanatory – conversion factor should be explained.

Rating

On a scale of 1 (excellent) to 4 (poor) I would give the datasets and paper a 2.5 at the moment with the potential to be closer to 1 than 2. The dataset is **unique, potentially significant** and will be **widely used**. To reach it's full potential users need to fully grasp how it was made, what the inputs were, where the major uncertainties are and thus how

to properly use the dataset in further research. These are areas for improvement (**completeness and data quality**) in the manuscript and associated datasets that could be included with the AgGDP layer. The **presentation quality** of the manuscript can be improved - see comments above to do justice to the impressive work conducted so far to produce this unique global spatial dataset.