Earth Syst. Sci. Data Discuss., referee comment RC3
https://doi.org/10.5194/essd-2022-297-RC3, 2022

Comment on essd-2022-297

Anonymous Referee #3

Referee comment on "GlobalWheatYield4km: a global wheat yield dataset at 4-km resolution during 1982–2020 based on deep learning approaches" by Yuchuan Luo et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-297-RC3, 2022

The manuscript of Luo et al. describes a new database providing gridded wheat yield at the global scale for the 1982-2020 time period. The authors used agricultural census data and advanced machine learning combined with remote sensing information and other ancillary data for the construction of GlobalWheatYield4km.

Disseminating a gridded wheat yield database is promising and the usefulness of such dataset is undoubted. However, I have serious concerns with the applied methodology. Given the fact that the presented method uses census based yield data that is disaggregated by remote sensing NDVI signal, the overall robustness of the product is almost guaranteed. In other words, the dataset shows good performance against census based yield data (Fig. 4) since census data is used to train the machine learning model. In regions where yield is high the model will provide similar high yield, and in regions where it is low it will definitely provide low overall yield, so the explained variance will be high. If the underlying remote sensing information is completely noisy (which is possible; see below) the performance of the model will be still good. And this is the case when the results are right for wrong reasons.

The methodology, as it is presented in the manuscript, is very brief and not reproducible. In order to get more information I read the cited Luo et al. (2022) paper (L108) published in International Journal of Applied Earth Observations and Geoinformation (https://doi.org/10.1016/j.jag.2022.102823). Surprisingly, the Luo et al. (2022) paper is very similar to the present study that in fact can be considered as the extension of the previous work to 54 countries (in the original paper 8 countries were included). I do not see any other added value here. The machine learning model is the same, the methodology is the same, the climate data is the same, the remote sensing information is the same. For this reason the novelty of the presented manuscript is minor.

I do not question the validity of the LSTM model. I just state that the method uses problematic input data and the results (including the correlation between the yield and the climate variables that is presented in the supplement of the Luo et al. (2022) paper) are

questionable.

Overall, also because of other issues detailed below, the manuscript is not qualified for publication in ESSD. As such, I would suggest a rejection.

Major issues

The temporal coverage of GLASS LAI is not clear from the manuscript. In the Luo et al. (2022) paper it is stated that it covers 2006-2012. Due to this short temporal coverage it seems that most of the time the AVHRR-based NDVI is used which is in fact not suitable for crop type identification due to known issues with geolocation and accuracy of the several AVHRR sensors onboard the NOAA satellites. Some studies explicitly mention issues with phenology detection based on AVHRR (see e.g. Atzberger et al. 2014, https://doi.org/10.3390/rs6010257). The authors state (L85-87): "In addition, the 8 d composite Global Land Surface Satellite (GLASS) Leaf Area Index (LAI) at 1-km spatial resolution and Global Food Security-support Analysis Data (GFSAD) 1 km Crop Mask product (GFSAD1KCM) were used to map spatial distributions of wheat." For me it means that actual (annually changing) wheat area was not estimated from the AVHRR data but rather it was handled as static; but in reality it changes year by year due to crop rotation in many places worldwide. I found this approach unacceptable. There is a major global effort to map crop types using Sentinel imagery supplemented with SAR data. It uses very high resolution and multiple data streams that are needed for accurate crop type mapping. Although I appreciate the effort of the authors, I do not see any justification that they really detect wheat cropping area and wheat phenology accurately.

In Luo et al. (2022) the authors themselves state: "Thus our wheat maps have not really characterized the dynamic variability in wheat-planting areas over time partly from avoiding the uncertainties of remote sensing data." In my understanding this exactly means that AVHRR-based crop type identification is happening that is not justified because of the above reasons.

I have problems with small (subpixel) parcel size that is typical in many regions worldwide. Did you study parcel size distribution? Using the 0.05 x 0.05 degree resolution AVHRR data the majority of the signal will be mixed by other crops/grasses/trees/shrubs etc. I do not think that it is possible to extract usable phenology profiles from that signal in a region that is characterized by small parcels. As AVHRR NDVI is noisy and problematic even for homogeneous areas, this is just an additional source of error.

I do not see the validation of the crop type detection using ground truth. Misclassification of winter wheat is an issue even using Sentinel data. For example, barley and what have

very similar phenological patterns (see e.g. Harfenmeister et al., 2021, https://doi.org/10.3390/rs13245036). So even if we assume that NDVI3g is applicable to phenology detection (which is questionable; see above), the accurate identification of the crop type is not guaranteed.

Some minor comments

L12: information on crop yield will definitely not ensure food security

L20: at this point it is unclear what is SPAM

L21: will play -> might play

L26: we are not sure what is going to happen by 2050. do not use 'will' here

L41: I do not agree that statistical models are less dependent on calibration data

L54: check English

L63: check English

L83: what dataset is it? NDVI3g? please specify

L98: explain the terminology. how is it related with the NUTS regions used within the European Union?

L103: PET is not actual evapotranspiration but potential evapotranspiration

L106: organic carbon content

L112: this is a key point. this statement is not true. there is very large similarity between the phenology of winter wheat and some other winter crops. see general comments

L116: I would mention harvest explicitly

L122: what statistics? I guess you mean census data. see also line 124 and other occurrences in the manuscript. check carefully and be more specific

L239 and section 2.3.2: provide loss-curve for the deep learning model. providing explicit information on the network used would have added value.

L133: provide reference to the Python package

L165: see the above comment on the terminology (statistical data). see also L166

L173: what is Kha? is it kilohectares? use standard abbreviation

L175: check English

L196: see the above comment (statistics)

L201: I guess quantifying bias would be more useful

L204: are you sure that this is the reason for the issue?

L216 and onwards: does SPAM provide annual data or is it aggregated yield for multiple years?

L254-255: I do not understand the sentence about SPAM

L234-235: check English

L251: see above; you mean agricultural census data