

Comment on **essd-2022-273**

Anonymous Referee #2

Referee comment on "AsiaRiceYield4km: seasonal rice yield in Asia from 1995 to 2015" by Huaqing Wu et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-273-RC2>, 2022

High-spatial and high-temporal resolution rice yield datasets are lack especially over large regions. The manuscript employed machine learning algorithms to generate long-term high-resolution rice yield over the South Asia, Southeast Asia, and East Asia. Undergoing a study at continental scales like this is a huge project. The 5km rice yield map over the major rice producing countries in Asia from 1995 to 2015 fills the data gap for assessing the impacts of climate change and the sustainable development. However, I have a few major concerns to be addressed so that the manuscript could be more solid.

(1) The rice cultivated area is the fundamental information for rice yield estimation. The manuscript used rice map for each year from 2000 to 2020 while the yield model was developed and used to estimate spatial distribution of rice yield during 1995 to 2015. Since most input dataset used for rice yield model in the study are available for the year 2000 to 2020, why not generating rice yield for 2000 to 2020 so that the map and the rice yield coincided with each other for the same year?

(2) Another concern is the way of predictor selection. The authors selected the predictors based on the correlation analysis between indicators and the yield at each administrative unit. While this is in general logic, it might be a problem when great differences existed in cropping patterns and the rice management in an administrative unit. The correlations may fail to achieve a significant level when an improper unit was targeted. This needs more clarification. Please also specify the administrative unit. Is it national level or sub-national level administrative units?

(3) The authors only used one vegetation indicator LAI as the inputs. It is assessed by several research that LAI products are of high uncertainty even for the improved GLASS

LAI products. The product still has some abnormal values and unrealistic seasonality especially in winter. From my understanding, using LAI products might introduce high uncertainty in yield model which is unable to be solved.

(4) According to the importance of the indicators, static indicators (Year, Lat, Long, Ele) are much higher than other indicators. For some countries, the proportioned importance of CEC+TI indicators could be higher than 90%. And for the whole study area, the CEC+TI are the most important indicators. How to explain this? Does this mean there are no need to add other indicators for yield mapping?

(5) When the model is applied for yield estimation during different growing season, does the pixel level cropping intensity map used or it is mainly based on the majority of rice cropping patterns in each administrative unit? The uncertainty of season rice yield might exceeded the uncertainty of the model due to the biased seasonal rice map.

(6) Any possibility to use some in-situ collected actual yield data to validate the yield map?

Specific comments:

(1) Page 4 Line 106, what do you mean by 27 seasons?

(2) The authors collected many rice yield data from different sources. Please add more detailed information of the yield data including the spatial units, temporal extent, etc.

(3) Page 10, Line 229 – 234, the dataset was first divided into two parts according to the administrative units. 80% of the administrative units were randomly selected as training and validation among which 70% of samples were used for training and 30% were used as validation sets. In this case, the training samples were not 56% of the whole dataset. Same for validation and testing. Please make it more clear for readers.

(4) Add more testing results for other years. The authors estimated rice yield for Asia for 1995-2015 but was insufficiently validated and tested for different years. Also, the temporal changes of rice yield should be added to result and discussion sections.