Earth Syst. Sci. Data Discuss., author comment AC1
https://doi.org/10.5194/essd-2022-16-AC1, 2022

Reply on RC1

Xunhe Zhang et al.

Author comment on "Mapping photovoltaic power plants in China using Landsat, random forest, and Google Earth Engine" by Xunhe Zhang et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-16-AC1, 2022

In this manuscript, authors developed a workflow combining machine learning and visual interpretation methods to map the Photovoltaic power plants in China. This topic is very important to assess the environmental and social impacts of these estabilished photovoltaic power plants. In fact, there are a number of papers on remote sensing target information extraction, the paper is not particularly novel. And, there are some problems in this paper:

Response: We are grateful for referee #1's recognition of this study's importance. Although there are a number of previous studies for mapping land properties, PV power plant mapping has not been widely conducted, and there still lack the open dataset for PV power plant in China. Our dataset offers the latest public dataset for the spatial extent of PV power plants in China. In this study, we integrate the advantage of cloud computing, machine learning, visual interpretation and freely available remote sensing imagery to map the PV power plants in China.

1, The introduction is inconformity with the objectives of study. For example, the one and two paragraphs are talking about the Photovoltaic power plants, machine learning, which can be wrote in a more refined way and introduce the main topic quickly. The references in the Introduction section are too limited, the authors should refer more works of relation analysis on deep learning methods, PV power plants and Remote sensing images, about the mechanism of deep learning and remote sensing image extraction of PV power plants, this paper hasn't given more details description.

Response to comment 1: Following referee #1's suggestion, we have rewritten our introduction part to fit the objectives of this study. We have streamlined the introduce part. We added sentences to describe the mechanism of machine learning and deep learning. We also added more references about learning. We further explained the advantages of our methods.

2, Figure 2 (page 7): the specific meanings of 1,2,3,4,5, and 6 in this figure should be explained.

Response to comment 2: We are sorry for the unclear description of Figure 2. We have refined the figure 2 with accurate description. The specific meanings of 1, 2, 3, 4, 5, and 6 are the 6 example sites to show the true-color images from Landsat-8, Sentinel-2, and Google Earth for visual interpretation.

3, There are too many texts in the discussion section and need to be further streamlined.

Response to comment 3: We have streamlined and shortened the discussion in the revised manuscript.

4,In addition, did the authors consider how to validate the results? how can we believe the results? e.g. what validation, more specific about ground truthing etc. Without this information, I can not trust the results of this paper.

Response to comment 4: In this study, there were two stages for mapping the PV power plants in this study. In the first stage, we used a pixel-based random forest model with selected features to map the PV power plants in China. We further validated the model and the performance of the random forest model using kappa coefficient, overall accuracy, producer's accuracy and user's accuracy (Table 3). In the second stage, we used visual interpretation method to filter the misclassified PV power plant due to commission errors in machine learning process. We carefully inspected each polygon with the latest Landsat-8, Sentinel-2, and Google Earth true color images by visual interpretation. The entire visual interpretation step took us about 2 weeks. While visual interpretation is time consuming, it generally offer validation with high accuracy.

In the revised manuscript, we added extra validation by comparing our dataset with the Dunnett's dataset and Kruitwagen's dataset in China. Dunnett et al. (2020) provided a harmonized solar plants dataset obtained from OpenStreetMap, an open-access map. The PV power plants in the open-access map were annotated by volunteers. The total area of PV power plants in China from Dunnett's dataset is 897.4 km^2, of which 842 km^2 have spatially intersected with our dataset. The no intersected solar panels area is 55.4 km^2. Some of them are too small for our method to recognize.

Kruitwagen's dataset (Kruitwagen et al., 2021) was classified by deep learning methods. The total area of PV power plants in China from Kruitwagen's dataset is 2169.8 km^2 by 2018, of which 1873.5 km^2 have spatially intersected with our dataset. The PV power plants in Kruitwagen's dataset that do not intersect with our dataset are 296.3 km^2, some of which are too small to be identified by our method and some of which are misidentified in Kruitwagen's dataset.

We found our methods could fail sometimes to recognize these PV power plants situated in mountainous areas that typically have unique installation spacing and installation angles for their solar panels. Small size PV power plants (less than 0.04 km^2) was potentially another reason for mis-classification in this study.