

Author Response to the Reviewer Comments to the manuscript "Mapping photovoltaic power plants in China using Landsat, Random Forest, and Google Earth Engine" [essd-2022-16] submitted to Earth System and Science Data.

We thank the referee #3 for his/her constructive comments, which are listed below (shown in blue) along with our response (shown in black). Accordingly, we have incorporated all the requested changes in the revised manuscript. We have also added the data DOI (<https://doi.org/10.5281/zenodo.6849477>) to the end of Abstract.

### **RC3: 'Comment on essd-2022-16', Anonymous Referee #3 and our response**

This a well conducted study, and the comments of R1 and R2 have improved the paper.

Response: We very much appreciate Reviewer#3's positive comments on our study.

I think one clarification is necessary before publication. R2 highlights the bias of relying on Dunnett's data, to which the authors respond: "So we further manually selected and edited the extent of different PV power plants that were not annotated in Dunnett's dataset" - does this mean there are some PV power plants in the training data that are not from this dataset? If so, it is not clear at the moment where these extra data come from. Are they manually searched for? Are they from some other national dataset?

Response: We are sorry for the unclear description of the training set. In addition to the Dunnett's dataset, we also manually collected data points for the training set. To do this, we firstly searched for the approximate distribution area of PV power plants through the web pages and then manually labelled the PV power plants on Google Earth high resolution images in 2017. These extra labelled PV power plants (approx. xx) are mainly located in eastern China, where PV power plants are rarely labelled in Dunnett's dataset.

We improved the sentence in Section 2.1.3 "*With high resolution Google Earth images of 2017, we further enriched the training dataset by manually selecting and labelling xx PV power plants over regions of eastern China, where PV power plants are rarely labelled in Dunnett's dataset.*

One of the objectives to the paper is to provide as replicable a process as possible for those without the computational means etc. so I feel like a full account of the training data is necessary.

Response: Thank you for the advice! We have uploaded our training points with training features for Landsat-8 imagery in this study as well as the labelled PV power plants to a data sharing website. It is publicly accessible at <https://doi.org/10.5281/zenodo.6849477>.

As an aside, I notice the RF model does poorly identifying "mountainous" PV sites, and I note that the authors purposefully excluded pixels with a >30deg slope. Could these two concepts be linked?

Response: Thanks for the comment. These two concepts are not really connected. The RF model didn't perform well for identifying "mountainous" PV sites mainly due to the unique installation spacing and installation angles of the solar panels in such area. Additionally, the mountainous terrain also affects the reflectance of the PV power plants, causing a distinct reflectance of PV power plants

in mountain areas from that in other flat areas. Therefore, the accuracy of our model's identification of PV power plants in mountainous areas is generally lower.

The reason we excluded pixels with the slope being over 30° for the analysis of this study was because the PV power plants are generally not built in locations with large slopes due to construction difficulty and soil erosion protection. As a result, we excluded very slopy areas that wouldn't have PV power plants to reduce the workload and computing costs of this study.