Open Access Earth System Science Data Discussions

Interactive comment on "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery" *by* Xiao Zhang et al.

Anonymous Referee #2

While I appreciate the authors' tremendous efforts in this global-scale mapping project, I have several major concerns. From the remote sensing perspective, the novelty of this project is low. Almost all the methods have been developed and used somewhere in the previous land-cover mapping projects. The classification system proposed in the study looks relatively simple. The study is not targeting the issue - "a fine land-cover system is still lacking" - as described at the end of the Introduction. However, the construction of the training database is a great effort that should be given more emphasis in the description of methods (e.g., adding a flowchart) and in the discussion (e.g., effects of sample outliers on mapping accuracies across land cover classes). See details below. I also feel there is a lack of in-depth discussion. For a large-scale project, data

uncertainties, model calibration, and land-cover heterogeneity could have a significant effect on mapping accuracy. But the current form of discussion is superficial and needs to add a comprehensive evaluation of the developed database.

Detailed comments: Line 10: add 'a' before 'lack'. L15: Include full names with the acronyms when they are first introduced. L99: The "lack of global satellite data coverage" is no longer a challenge for MODIS and Landsat that have been free of charge for over a decade. In fact, we are now in a data-rich era, which is why supercomputing and effective data mining are critical. L145-146: Why not directly using the ASTER GDEM product? The most recent version 3 of GDEM has better accuracy than SRTM. Section 2.3: What are your criteria for deriving how many points for each land cover class? L170: Where did you get the high-resolution imagery? How many points did you check? Following what criteria? L177: great -> big. Section: 3.1: There are multiple steps. I suggest a flowchart to describe your process. Also, how many samples did you collect for the study and for each class? What were your criteria? L214: land-covers -> land cover. L303-304: I do not agree that "classification accuracy was insensitive to these parameters". Please see a review of RF in RS classification by Belgiu and Dragut (2016). L320-322: It is vague how you balanced performance, efficiency, and sample volumes. What criteria did you use? Section 5.1: "huge training samples". Exactly how many samples were used? It is vague to use "exceeded 20 million points". Since building the training sample database is the most important contribution of the project, it is critical and would certainly benefit the users through discussing how the number of the training samples and how sample balance (across classes) have affected the results. The authors lightly touched on the outlier effect, but there is a lack of in-depth analysis and discussion using the data from the present project.