

Response to comments

Title: A 30-meter terrace mapping in China using Landsat 8 imagery and digital elevation model based on the Google Earth Engine

MS No.: essd-2020-157

5

Referee #1

General comments

Comment 1:

10 Terrace extraction is one of the important applications in remote sensing imagery processing in recent decades. Very few research focuses on terrace extraction in large scale areas. The paper suggested a new strategy that collected national, regional and local training samples that could be applied to large-scale areas. However, the effectiveness of the strategy is needed to clarify (Please refer Specific comments, No. 4).

Response 1:

Thank you very much for the comments and suggestions. Please see the detailed point-by-point responses below.

15

Specific comments

Comment 1:

Since the title of the paper is to classify terrace, the accuracy of terrace should be mentioned in the abstract section.

Response 1:

20 Thank you for your advice. We added the accuracy of terrace class in **abstract**: “For terrace class, the producer’s accuracy (PA) was 79.945% and the user’s accuracy (UA) was 71.149%.”

Comment 2:

25 Line 147-148: How did you use the national samples “for training general classification rules and identifying terraces with typical features” in the paper?

Response 2:

The random forest classifier can learn the terrace/non-terrace characteristics and summarize the classification rules based on the features of samples. So, the characteristics and rules RF learned were similar to those of the samples. Because the national terrace samples were collected from the terraces with typical features, the trained classifier can identify the typical terraces.

30 The specific principle of training and classification using random forest classier was supplemented in **Section 2.4**: “It is consisted of multiple decision trees, all of which perform classification separately and vote for the final results. During the training process of decision tree, each tree node is split based on the most contributing feature among the randomly selected input features of the training sample subset. After training, each decision tree judges the pixel class according to the established tree rules (Breiman, 2001; Gislason et al., 2006).”

35

Comment 3:

Line 168-169: What is “the feature number”? Is it the number of training samples?

Response 3:

In machine learning, feature is the selected property or characteristic of samples. The feature number is 39 in this study.

40 Detailed information about features used in the study can be found in Section 2.2.

To make it clearer, we added the specific number of features: “the number of decision trees and the number of variables per split, were set to 200 and the rounded square root of the feature number (39)” (**Section 2.4**).

Comment 4:

45 Line 171-172: Is it necessary to collect national and regional training samples instead of only local training samples since you merged the three sample sets and applied each random forest model for each province individually? Please try to apply your model using only the local training samples and compare the results.

Response 4:

Thank you for the suggestion. In the study, we adopted the strategy of collecting national, regional and local samples in order to improve the sampling efficiency, i.e., reduce the workload of sampling through reusing the national and regional samples. As described in Section 2.3, we first collected the national and regional samples (801 samples and 54 samples respectively), and then supplemented some local samples (3989 samples) to improve the mapping accuracy of all provinces except Macao, where there is no cropland. However, the local sample size was very small ($N < 10$ for either terrace or non-terrace) in some provinces (Gansu, Guizhou, Heilongjiang, Hongkong, Jilin, Macau, Ningxia, Shandong, Shanghai, Tibet, Xinjiang, Yunnan).

55 So, only using the local samples collected in the study is insufficient for training the classifier and comparing the results.

We supplemented some sentences in **Section 2.3** to clarify the purposes of collecting national, regional and local samples: “Through reusing the national and regional samples, smaller total sample size was required and the workload of sampling was minimized effectively.”

60 **Comment 5:**

Table 2, 3; line 214-216: Please clarify your point by describing Statistically significant (P value) of the results of using the two test sample types.

Response 5:

65 We calculated the P-value as the reviewer suggested to prove the accuracies of terrace class evaluated by the two test sample sets are similar.

We added the related texts to describe statistically significant in **Section 3.2**: “The accuracy evaluation result using the 301 test samples of known terraces (Table 4) was numerically similar to the above result using the 10875 random test samples (Table 3). The Chi-square tests (Mantel, 1963) were carried out for the two PAs and UAs of terrace class respectively to further prove the similarity quantitatively. The P-values of both tests were greater than 0.05, indicating there was no statistically significant difference between the terrace accuracies using the two test sample sets.”.

Reference:

Mantel, N.: Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure, J. Am. Stat. Assoc., 58, 690-700, <https://doi.org/10.1080/01621459.1963.10500879>, 1963.