Comments on "Mapping the Vegetation of Lake Tana Basin in Ethiopia Based on Google Earth Images"

This MS introduces a vegetation dataset of Lake Tana Basin for potential uses in vegetation management and conservation. Vegetation types were delineated using the method of human visual interpretation. The accuracy of the dataset was validated by ground survey. The authors have put great labor efforts to delineate such a detailed vegetation map. It is a valuable dataset to document vegetation in this specific area. It may be too much to ask the authors to improve the quality of the dataset at this stage, but more details on methodology and discussions on accuracy of the dataset should be provided for future uses.

Manuscript

1. This is a new dataset that has potential uses in landscape management and land process modeling in the basin of Lake Tana, Ethiopia.
2. The method of producing the dataset is not new, but valid.
3. The mothed is not described in sufficient detail.
   3.1. Even though the paper cites Shimelis et al. (2008) for the vegetation classification system. I suggest describe this important information for the convenient of data use.
   3.2. L132: Number of the plots should be clearly provided, even though you have shown in Figure 1 in the MS.
   3.3. L144: Validation accuracy matrix should be provided (accuracy of each land use/cover class), not just an overall accuracy (90.6%). The data quality has not been fully assessed regarding different vegetation classes.
4. Potential error sources should be evaluated and discussed, such as errors in visual interpretation: How many people? Who were they? Were they from local areas? How did they work (time, place, machines, etc.)? How were they trained to do the work? How did they consult to the geobotanist in Ethiopia (L143)?
5. The manuscript should focus on methodology and data accuracy. The Results, Discussions, and Conclusions belong to interpretation of data, which is outside the scope of ESSD articles (https://www.earth-system-science-data.net/about/aims_and_scope.html). However, the possible uses of the dataset should be discussed.

Data quality

1. The dataset is accessible.
2. Readme file (meta data)
   2.1. "In the attributes table of this °∞VegCover°± dataset…": unexpected text codes (it could be my computer problem).
   2.2. "Because, except the cover type 1-9, lake tana basin was mostly under cultivation": grammar: "except for", "types", "Lake Tana Basin". Please check the rest of the file.
   2.3. Suitable software for visualization of the dataset should be pointed out as required by the journal.
3. Cultivated lands were not identified in the dataset. That is not consistent with the manuscript (Figs. 2 & 3 in the MS). Without identification of cultivated lands, any statement

in Section “4.7 Cultivated Land” is not accurate. This also limited the use of dataset in land use/cover change: human disturbance indicated by cultivated lands.

4. I understand that it is labor-intensive for visual interpretation. Most boundary delineation is OK. But some polygon boundaries do not have the accuracy that is claimed in L139: polygons were added in the scale of 1:5000 (Fig 1 here). It should be revealed and discussed.
5. There are big or small gaps between polygons. They can be cultivated lands that were not identified, or gaps that should be merged to neighbor polygons that have the same attributes (Fig 2 here). It is not usual leaving gaps in maps of land use/cover.
6. It is hard for me to visually identify 2-plantation forest from 3-natural forest (Fig 2 here). Figure 2 in the MS is helpful, but more detailed description about the difference between the classes and identification criteria is recommended for the sake of accuracy assessment and future uses. The gaps are not necessarily cultivated lands, implied by the “readme” file.
7. There might be a mismatch between the time of Google map images (undocumented in the MS, meta data show 2016-2017) and validation time (i.e., 2015 and 2016, L144). My random samples indicate some misclassifications of 2-plantation forests, which look like cultivated lands (Fig 4 here). The misclassifications may be due to land use changes and the time difference of the basemap I used here and the basemap in Google Map that was used for visual interpretation for this dataset. Please document the years of images that were used for this dataset. Please discuss the validation issue related to the mismatch of time. Even though it is costly practical using Google Images, it is not the best practice using unknown (time) images to for interpretation of land use/cover.

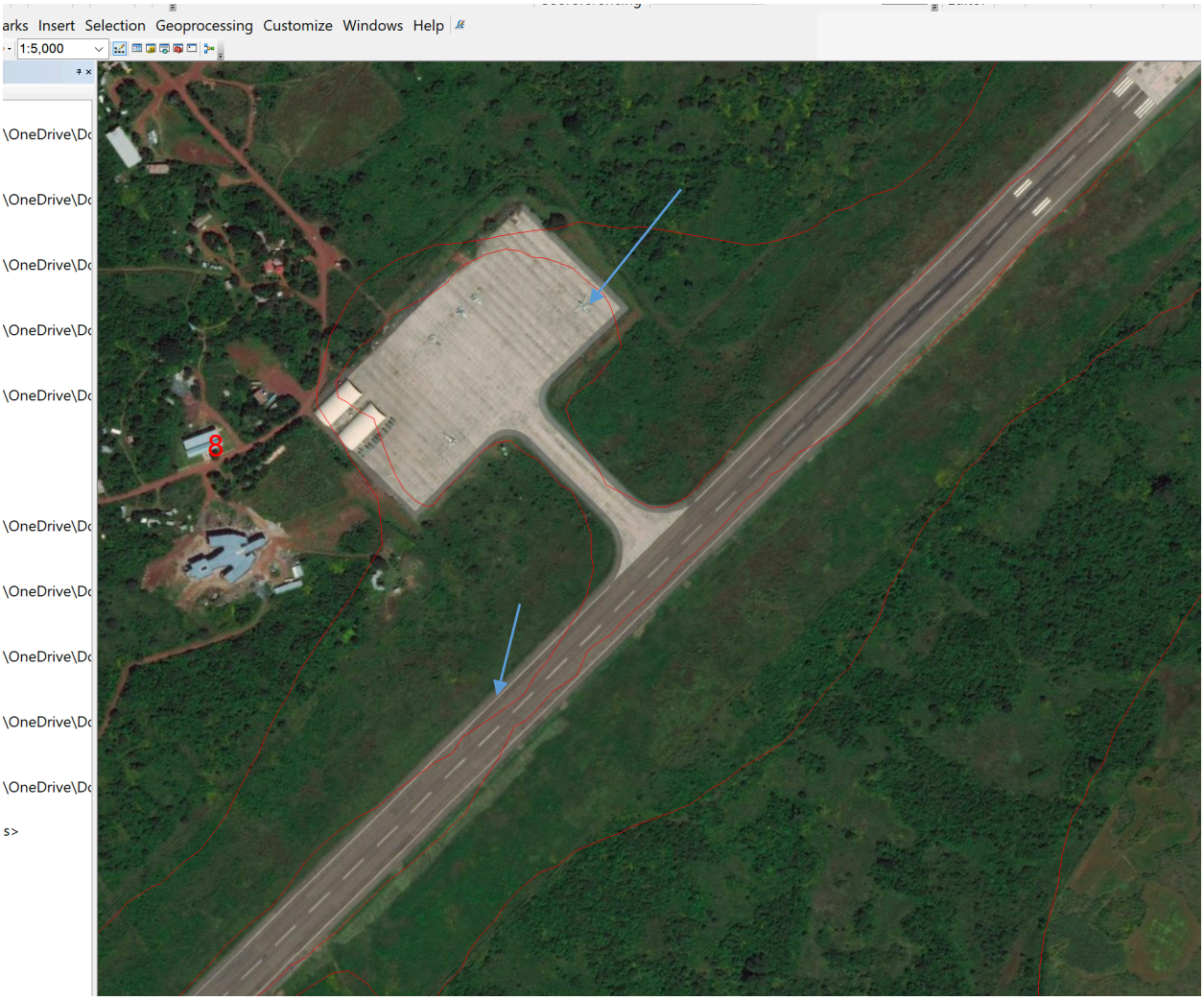

Fig 1. Screenshot of polygon boundary on ArcMap 1:5000 scale. Basemap is ArcGIS online base images (imported on Jun 1, 2018, same for Fig 2, 3, and 4 here).

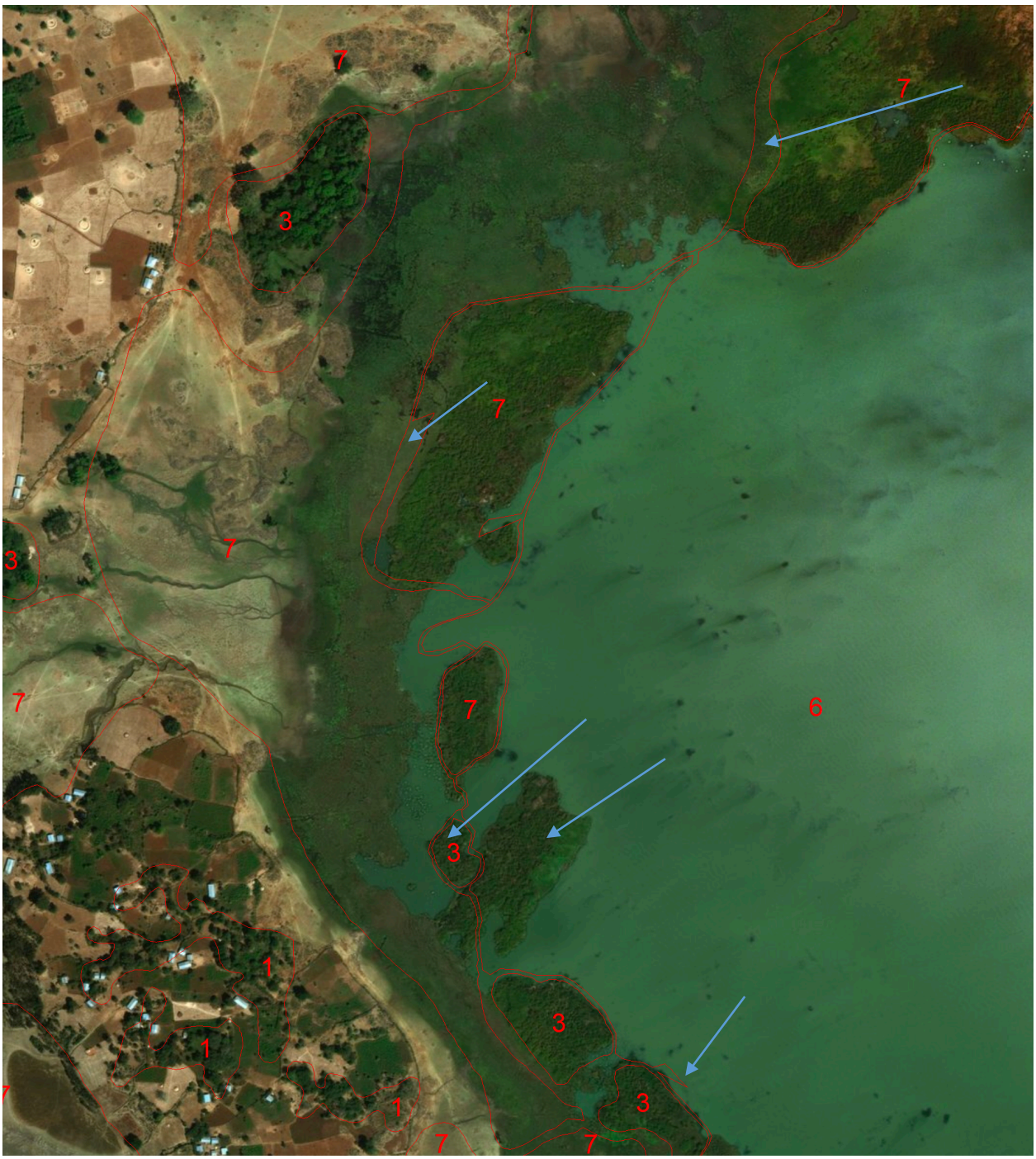

Fig 2. Gaps and classification accuracy. Map is on ArcMap 1:5000 scale. Arrows from the top to bottom indicate: the 1st and 2nd: unnecessary gaps between class 7 (wetlands); the 3rd, miss classification between 3 (natural forest) and 7; the 4th, unidentified island; the 5th, boundary error.

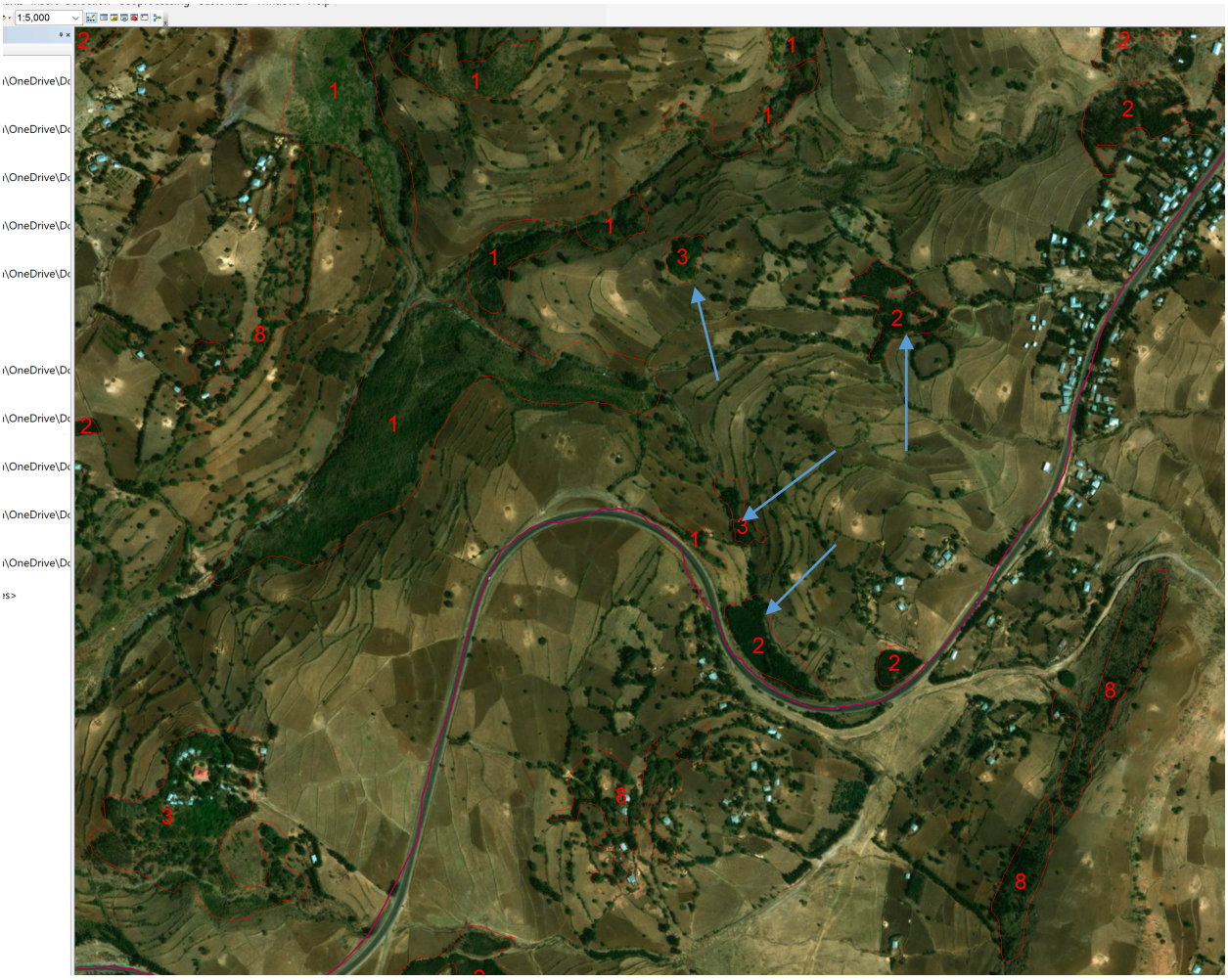

Fig 3. The visual difference between 2-plantation forest and 3-natural forest is very small. Map is on ArcMap 1:5000 scale.

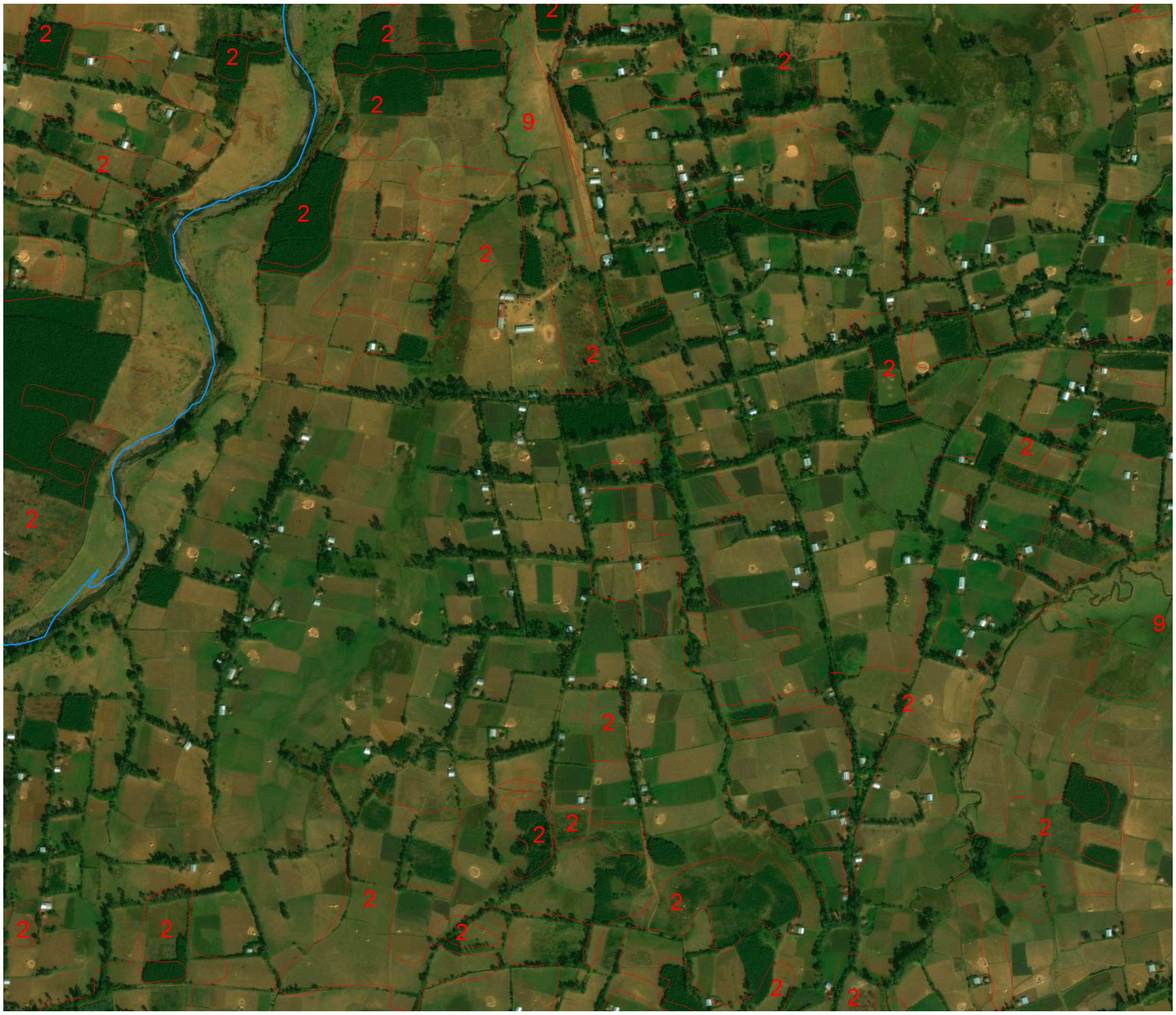

Fig 4. Some cultivated lands were classified as 2-plantation forests. ArcMap 1:5000 scale.