Comment:

In this study, the authors developed a lake level dataset with dense samples for large lakes in 2000–2017 in the Tibetan Plateau (TP). The lake level product is validated by in situ water level measurements for Yamzhog Yumco. The water volume changes of 52 lakes with lake level were also estimated. This dataset is very valuable for studies of lake variations and their response to climate change in the TP and lake water balance. I recommend this manuscript to publish in ESSD, but some improvements based on comments below are necessary.

Response:

We really appreciate these overall comments and recommendation by this reviewer. Our point-by-point responses to the reviewer's comments are given as follows.

General comments:

1) The uncertainties for lake volume changes and other number should be added through the manuscript.

Response:

Thanks for this comment. They will be added into the revised manuscript.

2) What is optical water level? It is estimated by the correlation between lake area and level, and then to reconstruct the corresponding lake level using known lake area?

Response:

Thanks for this comment. As illustrated in Page 10, line 20, the generation of optical water levels is similar to the description in this comment but is based on changes in lake shoreline observed in a smaller ROI (region of interest) rather than the whole lake area (e.g., the yellow square shown in Page 12, Figure 3 (b)). The reason for this is due to the increasing computational cost and probability of cloud/gap contamination with increasing areas of ROI. On the other hand, it is pointed out in section 4.2 that if the ROI had a larger width (here 'width' is in the direction parallel to the shoreline), the uncertainty of optical water level would decrease. Therefore, the choice of the ROI is a trade-off between the accuracy and data availability or computational cost.

3) How all the lake level datasets are converted to same geoid?

Response:

Thanks for this comment. For altimetry water levels, the initial reference ellipsoid and geoid are different for different satellite missions/products. Information on the reference ellipsoid and geoid is listed in Supplement Table 1:

Supplement Table 1

Altimetry mission	Reference Ellipsoid	Geoid
Envisat	WGS84	EGM2008
ICESat	T/P	EGM96
CryoSat-2	WGS84	EGM96
Jason-1/2/3	T/P	EGM96

As mentioned in the manuscript, different altimetry water levels were merged by comparing the overlap period (more details are available in the response to short comment 1). Systematical biases caused by the geoid and reference ellipsoid were removed during this process.

For optical water levels, they were generated using linear regression with a certain source of altimetry water level data so they have the same reference ellipsoid and geoid with the respective altimetry data used in the regression. And they can be merged with other altimetry data by comparing the overlap period as well.

However, there is a correction that must be made to the manuscript in Page 8, Line 18. "…all water levels were with respect to EGM96…" was incorrect. For the 12 lakes with Jason data, all kinds of water levels were converted into T/P, EGM96, because the Jason-1/2/3 data were used as the baseline (i.e., the longest records will be used as the baseline). For the rest of the lakes we mainly used Envisat data as the baseline to merge all the water levels. Therefore, for lakes without Jason data but having Envisat data, the water levels were converted into WGS84, EGM2008 (see Supplementary Table 1 above). For lakes without either Jason or Envisat data, Cryosat-2 data were used as the baseline, so water levels for these lakes were converted to WGS84, EGM96. We will provide a supplement document to mark out the Reference Ellipsoid and Geoid for each lake.

4) In this study, lake boundaries were extracted using GEE. The visual checking and manual editing of delineated lake boundaries with original Landsat images are very necessary. How this was done at GEE platform?

Response:

Thanks for this comment. Lake areas or lake boundaries were used in two situations in this study: the first is during the process of generating hypsometric curves and the second is during the selection of altimeter footprints. We used lake areas derived from GEE in the first situation but used an existing dataset based on manual delineation produced by Wan et al. 2016 in the second situation. Therefore, problems raised in this comment may only exist in the first situation.

Visual checking can be hard to perform on GEE due to a large number of images we used, but we did visually check and preclude some of the images that resulted in outliers in the extracted lake surface areas (e.g., the entire ROI was covered by snow resulting in the failure of the Otsu method). It is true that manual editing of lake boundaries is important if we only use a small number of images (e.g., less than 10 images) to derive hypsometric curves (which is common in similar studies, e.g., most

hypsometric curves provided by Hydroweb used less than 10 data pairs).

Nevertheless, for most lakes (42 out of 52) in our study, we used more than 20 data pairs (i.e., lake water area and corresponding lake water level) to fit hypsometric curves. More data pairs we used make the hypsometric curves more robust, even though there may be some misclassification in a single image. This is evidenced by the fact that most R^2 values for the hypsometric curves are higher than 0.9. In addition, all the images used in the regression analysis met the criterion of cloud contamination less than 5%, which has largely reduced uncertainty in the extracted lake water area.

5) How the in situ water level for Yamzhog Yumco is converted to consistent reference ellipsoid with satellite altimetry data? For validation of lake water classification with UAV, how about classification accuracy?

Response:

Thanks for this comment. As illustrated in our response to short comment 2, the in situ water levels of Yamzhog Yumco were made comparable with the satellite altimetry/optical water levels by calculating the anomalies of each water level time series.

Lake water classification with UAV images was performed by manually identifying the lake shoreline using ArcGIS. Therefore, the uncertainty in the UAV derived lake shoreline is considerably small, because the spatial resolution of the UAV image is ~5 cm.

Specific Comments:

1) Page1: "There are more than 1,200 alpine lakes larger than 1 km^2 (Zhang et al., 2017a)" This result should come from Zhang, G. et al., 2014. Lakes' state and abundance across the Tibetan Plateau, Chinese Science Bulletin, 59(24):3010−3021. Please correct this cite here.

   Page 2: ETM should be ETM+, not a superscript symbol of +, others are similar.

Response:

Thanks for this comment. They will be modified in the revised manuscript.

2) Page 4: "examine long-term", 2000–2017 is not long-term.

Response:

Thanks for this comment. "Long-term" will be changed into "multiyear".

3) "The TP can be generally divided into 12 major basins... ". Two suggested reference here:

   Wan, W. et al., 2016. A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014, Scientific Data, 3:160039.

Zhang, G. et al., 2013. Increased mass over the Tibetan Plateau: From lakes or glaciers?, Geophysical Research Letters, 40(10):2125−2130.

Response:

Thanks for this comment. They will be added into the manuscript.

4) Lake Selin Co-> Selin Co, others are similar

Response:

Thanks for this comment. They will be modified in the revised paper.

5) Page 5: "a lake shape data set generated by Wan et al. (2016) was used". This lake shape data was derived from GF data. How about the shift of lake outline? Did you check it with original Landsat images or Google Earth?

Response:

Thanks for this comment. Yes, we did notice that lake outlines derived from GF data that were only used for generating the 2014 subset of the data set by Wan et al. (2016)) have a shift relative to those derived from Landsat ETM+ and CBERS-1 that were used for the 2005 subset of the data set). Therefore, we only used the 2005 subset of the data set to select altimetry footprints in our study.

6) "We managed to make use of some images with gaps in generating lake shore changes." How to understand it?

Response:

Thanks for this comment. As shown in Figure 3 (b), the ROI we used to derive lake shoreline changes is small enough to be fitted into an ETM+ image strip (valid pixels) between two gaps (no-value pixels). In this way, we can make use of some images with gaps. However, the gaps are shifting so we set a criterion (no-value pixels in the ROI should be less than 2%) to remove those images for which the ROI is contaminated by shifting gaps.

7) "A half of them were excluded from the final results due to cloud contamination or gaps." How this is determined? Some lakes are missed? How to make sure a high-quality output of lake boundary, especially lake with little ice or turbid water?

Response:

Thanks for this comment. As illustrated in the manuscript, we have excluded images which have more than 5% cloud pixels or 2% no-value pixels in the selected ROI. "A half" is an approximation for the portion of effective images when generating optical water levels. To further show this, we randomly chose five lakes to present the portion of effective images:

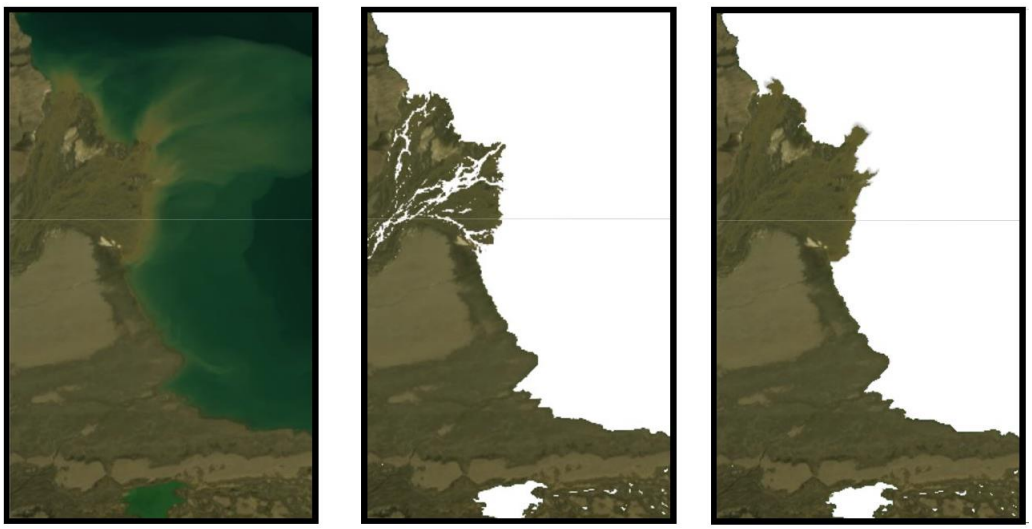
Supplement Table 2

Lake name	TM	ETM+	OLI	Total
Jingyu	90/287	225/360	58/128	373/775
Zhari Namco	127/308	47/229	178/371	352/908
Mapam Yumco	50/57	216/274	85/121	351/452
Lumajiangdong Co	74/183	78/561	68/250	220/994
Aqqikkol Lake	169/387	108/557	109/246	386/1190
Total	510/1222	674/1981	498/1116	1682/4319

Effective ETM+ images have a lower portion due to gaps. But for Landsat TM and OLI images, the portion is appropriately 1/2. However, the portion of effective images that can be used to derive lake water areas, instead of optical water levels mentioned earlier, is much smaller, sometimes less than 10%, which is due to the increasing probability of cloud/gap contamination with increasing areas of ROI.

We did consider the impact of lake ice or snow on the accuracy of lake area/lake shoreline extraction. Both MNDWI and NDWI were not able to well discriminate lake ice and water as what we had expected (i.e., if lake ice was eliminated, the extracted lake area would be smaller than its real size). We noticed that the MNDWI cannot completely discriminate snow and water either, resulting in artificial increases in lake areas in winter when the lake bank was covered by snow. Therefore, the NDWI was used to better discriminate snow from water/floating lake ice in winter. However, if there was snow cover on the lake ice, the NDWI could also produce artifacts in the derived lake area and we had to remove these outliers manually as we mentioned earlier.

As for the turbid water problem which mainly occurred in summer, we examined the study lakes and found that both MNDWI and NDWI can precisely locate the lake boundary, even though the near-shore water color was affected by turbid inflow as shown in Supplement Figure 1:



Landsat OLI 2017/10/07 MNDWI water classification NDWI water classification

Supplement Figure 1: Water classification results at the estuary of Lake Kusai based on the MNDWI and NDWI during the flood season.

The difference between the MNDWI and NDWI is that the MNDWI is able to detect shallower turbid water than the NDWI (e.g., shallow rivers can be detected by MNDWI in Supplement Figure 1), which is important in determining the accurate position of lake shorelines. If the NDWI was used in summer, less information on changes in lake shoreline (i.e., optical water level) would be detected. On the other hand, rivers and other small water bodies near the lake can lead to noise to the extracted lake area due to the sensitivity of the MNDWI. Therefore, we carefully chose the ROI to avoid rivers or small water bodies. A comparison between the MNDWI and NDWI was performed by (Huang et al., 2018) based on UAV images, which also shows that the MNDWI has better performance than the NDWI under the condition without snow cover.

8) Table 2: "d, m, km" can be put first row of table, then others below can be removed.

Response:

Thanks for this comment. Yes, it will be modified in the revised manuscript.

9) "Either comparing the mean water level of the overlap period or comparing the two water level time series with changes in lake shoreline" How about the uncertainty and it is reasonable?

Response:

Thanks for this comment. The uncertainty of this method is important, because errors induced by this data merging method will evolve into the merged water levels and become remaining systematical biases. Such biases will cause artificial rises or falls of the merged lake water levels, jeopardizing the consistency of the merged lake water levels between different time periods and sensors. Therefore, the consistency of the merged water levels can reflect the remaining systematical biases and the uncertainty of the data merging method that caused these biases.

However, it is a dilemma that evaluating the consistency of the merged water levels is difficult to perform without continuous in situ observations over multiple years. As far as we know, there are few continuous measurements of lake water levels in the Tibetan Plateau due to the equipment failure in the frozen season (e.g., caused by fierce winds, waves, and freezing process). For instance, several water level sensors have been set up in Nam Co since 2005 (Song et al., 2015), but the in situ water level measurements of Nam Co presented in the literature were discontinuous in the frozen season.

Therefore, the best available reference we used to assess the consistency of the multiple altimetry water levels <u>when there is no overlap period</u> is the optical water level. Optical water levels are generally continuous in our study period and could

even be more reliable than intermittent ground observations. Given the fact that continuous ground observations do not exist, are not accurate enough, or are not accessible if any, the altimetry data merging method proposed in this study is a reasonable and effective way to generate longer and denser time series on lake water levels.

10) As the differences of extracted lake outlines, it is better to use a unique NDWI or MNDWI in classification of water and other land-cover in the study period? In addition, the differences from NDWI or MNDWI are not apparent?

Response:

Thanks for this comment. Based on response to Comment 7, it is clear that either NDWI or MNDWI has pros and cons and may perform quite differently. Therefore, a combination of the two water indices is a reasonable solution and has been used in this study.

11) "We selected images with less than 5% cloud cover". Some images with free-cloud coverage on lake shorelines are still useful?

Response:

Thanks for this comment. Yes, they are. Moreover, the cloud mask algorithm imbedded in the Landsat QA band is quite sensitive. Sometimes an image with light cloud in ROI slightly higher than 5% is still useful, because water indices are not largely affected. A 20% threshold was used by (Huang et al., 2018), which also produced satisfactory results.

12) Figure 11: background of this figure is not clear?

Response:

Thanks for this comment. Yes, the background has been changed into green now.

13) Figure 12: What is a high peak in Figure 12 in about 2010?

Response:

Thanks for this comment. It may have been caused by an outlier that was not removed prior to uploading the generated data set. In the uploaded data set, such a peak does not exist. It will be corrected in the revised paper.

14) Figure 13: The trend of lake storage change is more robust than the result from Yao et al (2018) from Yao, F. et al., 2018. Lake storage variation on the endorheic Tibetan Plateau and its attribution to climate change since the new millennium, Environmental Research Letters:1-16. What is the cause for this difference?

Response:

Thanks for this comment. As illustrated in Line 5–7 in the manuscript Page 23, our data (a combination of the merged optical water levels and altimetry water levels)

have higher sampling frequency than (Yao et al., 2018), resulting in a more robust estimation of the trend in the lake water levels. As shown in Figure 12, there are several abrupt changes with magnitudes up to ~3 km^3 in lake storage observed by (Yao et al., 2018), which is not likely to happen, given that there is no report on basin flood/upstream lake overflow. This could be due to the uncertainty in the lake area they derived and applied to estimate changes in lake storage. We also noticed that lake areas derived from Landsat archives could be much noisier than lake shoreline changes, due to cloud contamination/temporary small water bodies within the ROI. Therefore, we have calculated changes in lake storage with water levels and hypsometric curves, instead of directly using water levels and lake areas to reduce the uncertainty in derived lake areas.

15) Figure 15: How to understand the difference of lake level between these different datasets, especially polylines for optical water level?

Response:

Thanks for this comment.

First, the difference between altimetry water levels in our data set and the Hydroweb data set mainly comes from following processes: (1) different reference ellipsoids and geoid models, (2) different retracking methods, and (3) different schemes of removing systematical bias. The last process is the most significant difference that could make our data set more consistent compared with the optical water levels as we explained in response to specific comment 9.

Second, the difference between the optical water levels and altimetry water levels mainly comes from different mechanisms of observations. Altimetry water levels are based on the time delay between the generated and received signal measured by altimeters. Each cycle corresponds to one water level value averaged from several footprints across the lake. Therefore, the number of footprints in a cycle is crucial to the accuracy of altimetry water levels. Footprints falling on a study lake are determined by the orbit of the satellite altimetry and the size of the lake, both of which are fixed.

On the other hand, optical water levels are derived from optical images, which could be affected by cloud cover. Therefore, there is not a fixed temporal resolution for optical water levels. As illustrated in section 4, optical water levels are mainly affected by the slope of lake shore, the width of ROI, the spatial resolution of the optical image, and the accuracy of the water classification method. Some of these factors, such as the width of ROI, spatial resolution, and slope can be well handled. Therefore, optical water levels are less noisy than altimetry water levels.

16) "5.3 Lake overflow flood monitoring". Many similar Chinese papers have been published. It is not need to include in the Title of this manuscript and put some in discussion is enough? In addition, some sentences such as equation can be moved into Method section?

Response:

We have revised the title of this manuscript according to this comment. Content associated with lake overflow flood monitoring is no longer reflected in the title, but has been put in the method and discussion sections. These modifications will be shown in the revised paper after considering all reviewers' suggestions.

17) Xiaojun et al., 2012 -> Yao et al., 2012

Response:

Thanks for this comment. It will be modified in the revised manuscript.

18) "Water loss was more likely to be found among the southern TP lakes. In the Selin Co basin, a more complicated spatial pattern of lake storage changes was detected, as small lakes were slowly losing water whereas the large lake was gaining water, which we speculated to be caused by lake-river interactions that need further investigation." These conclusions have found in previous studies. The summary here should more focus on the lake level data developed in this study.

Response:

Thanks for this comment. It will be modified in the revised manuscript.

19) Section 4 is too long? It can be shortened?

Response:

Thanks for this comment, we plan to move part of the content in section 4 into the supporting information.

Reference:

Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A., and Hong, Y.: Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River, Remote Sensing of Environment, 219, 115-134, 2018.

Song, C., Ye, Q., and Cheng, X.: Shifts in water-level variation of Namco in the central Tibetan Plateau from ICESat and CryoSat-2 altimetry and station observations, Science bulletin, 60, 1287-1297, 2015.

Wan, W., Long, D., Hong, Y., Ma, Y., Yuan, Y., Xiao, P., Duan, H., Han, Z., and Gu, X.: A lake data set for the Tibetan Plateau from the 1960s, 2005, and 2014, Scientific data, 3, 160039, 2016.

Yao, F., Wang, J., Yang, K., Wang, C., Walter, B. A., and Crétaux, J.-F.: High resolution data set of annual lake areas and water storage across the Inner Tibet, 2002-2015. In: Supplement to: Yao, F et al. (2018): Lake storage variation on the

endorheic Tibetan Plateau and its attribution to climate change since the new millennium. Environmental Research Letters, 13(6), 064011, https://doi.org/10.1088/1748-9326/aab5d3, PANGAEA, 2018.