Anonymous Referee #1

This dataset of Tibetan Plateau lake-catchment characteristics fills a data gap for conducting a variety of potential scientific studies on the lakes in the region, which itself is of great importance to understanding the earth systems. Development of the dataset is well justified and well designed. The dataset is comprehensive and contains many aspects of information on a large number of lakes over a long period of time (including time series data). Besides, potential users of the dataset would appreciate the distinction between full-lake-catchment and inter-lake-catchment statistics.

The only suggestion from the reviewer is to add a section to briefly discuss or highlight the uncertainties of such a dataset. For example, given that most of the environmental statistics were obtained from existing datasets (DEM, Land cover, climatic data, etc.), highlighting sources/magnititude of uncertainties of such datasets, and discussing how the uncertainties propagate to the developed TP lake-catchment dataset would be beneficial to future users of this dataset.

Overall, the reviewer thinks the dataset is scientifically sound (adding discussion on uncertainty is a plus) and presented in a very clear manner. This dataset could be of great interest to the earth scientfic reserach communities.

Reply:

We appreciate the reviewers' insightful and helpful comments on our manuscript. We have revised the manuscript according to the reviewer's suggestion. A new section about the uncertainties of the dataset has been added in the revised manuscript (L330-366) to facilitate the usage of this dataset.

"5 Uncertainties of the dataset

Since the catchment-scale attributes in this dataset were mostly derived from existing datasets by calculating zonal statistics (such as sums, means, and medians), uncertainties of source datasets were propagated to the results and determined the uncertainties of this dataset. We have done our best to collect the most reliable datasets to date and will regularly update the related datasets in the future to ensure their timeliness. Still, users of this dataset need to be aware of the uncertainties of the main source datasets which are listed as follows.

1) Lake water level and volume. The RMSE (root mean square error) of the Landsat-derived water levels from Li et al. (2019) was 0.11 m. The water level data from Xu et al. (2022) had $R^2 > 0.80$ and RMSE < 0.12 m in Qinghai Lake. The uncertainties for each value in the time series of Li et al. (2019), Zhang et al. (2021), and Xu et al. (2022) can be found in corresponding uncertainty files (Table S2).
2) Topographic data. Most topographic attributes in this dataset were derived from MERIT DEM and MERIT Hydro (flow direction map) datasets. MERIT DEM was produced by eliminating main error components (e.g. absolute bias, stripe noise, speckle noise, and tree height bias) from existing DEMs (SRTM3 DEM, AW3D DEM, and VFP-DEM). It has a resolution of 3″ (~90 m at the equator) and the land areas mapped with ±2 m or better vertical accuracy were 58% (Yamazaki et al., 2017). MERIT Hydro was derived from MERIT DEM and water body

datasets (G1WBM, Global Surface Water Occurrence, and OpenStreetMap). The relative error of MERIT Hydro in drainage area delineation was less than 0.05 for 90% of Global Runoff Data Center (GRDC) gauges.

3) Climatic data. The CMFD meteorological dataset used in this study was produced through fusion of remote sensing products, reanalysis datasets, and in-situ observations from a larger number of stations. Its accuracy in western China was validated based on independent observations, and the results showed that CMFD had closer-to-zero MBE (mean bias error), lower RMSE, and higher R^2 than the Global Land Data Assimilation System (GLDAS) for almost all meteorological variables (He et al., 2020).
4) Land cover/use data. The land cover/use data used in this study came from the fusion of six popular land use products, with an accuracy of 88.71% (Xu, 2019). The GPP and NPP data came from the MODIS products (MOD17A2H.006 and MOD17A3HGF.006). The R^2 between monthly MODIS GPP and eddy covariance measurements was reported to be 0.64 on average, and the RMSE was 2.55 g C m^{-2} day^{-1} in alpine grassland, which is the most widely-distributed biome on the TP (Zhu et al., 2018), and the R^2 between MODIS NPP and in-situ observations in 23 stations across China was reported to be 0.81, and the RMSE was 73.44 g C m^{-2} (Sun et al., 2021). The RMSE of fractional snow cover data from Jiang et al. (2022) was 0.14 taking the results from high-resolution Landsat images as reference. The R^2 between snow depth data from Che et al. (2021) and in-situ observations was 0.81, and the RMSE and MAE were 7.7 cm and 2.7 cm.
5) Soil data. The SoilGrids 2.0 dataset used in this study was generated by machine learning methods, using approximately 240 000 soil observations worldwide and over 400 environmental variables as inputs. It provides a spatial distribution map of data uncertainty generated by the quantile regression forest prediction model, which is the ratio of the interquartile range (i.e. the difference between 0.95 quantile and 0.05 quantile) over the median (Poggio et al., 2021). The catchment-level average uncertainty for each soil variable was calculated and included in this dataset. For the maximum freezing depth of seasonal frozen-soil, the R^2 in the four periods of 1980s, 1990s, 2000s and 2010s were 0.77, 0.83, 0.73 and 0.71, respectively (Wang and Ran, 2021)."