Authors reply to the comments by Jan Beutel of the manuscript essd-2020-106

"An integrated observation dataset of the hydrological-thermal deformation in permafrost slopes and engineering infrastructure in the Qinghai-Tibet Engineering Corridor"

by Lihui Luo et al.

We thank Dr. Jan Beutel for valuable feedback, which helped us improve the manuscript. Please find below the Referee comments in black, Author responses in green, and Changes to the manuscript in blue.

Response to referee comment 1:

Dear authors,

as stated in the previous review report your submission is lacking w.r.t. the dataset preparation. I acknowledge your efforts done for this the revision, especially the preparation of an extensive README document as guidance, a working R codebase and the inclusion of the InSAR data, however data files mentioned are still missing and/or the data described cannot be found. Also, some of the files I downloaded from Zenodo are labeled as corrupt.

Response: Thank you for the insightful comments. We endeavored to improve the manuscript and datasets by your comments. To address these concerns, we have made the following modifications: (1) We have merged multiple meteorological or ground data files into one meteorological or ground data file, and it was updated to 2020 because a weather station contains multiple meteorological or ground data in different

years; (2) We have updated the README file to ensure that it is consistent with the naming in the data repository; and (3) we have added frozen soil classification and borehole figures, and added vector data for frozen soil classification in the Qinghai-Tibet Plateau. Because the Zenodo repository does not support directories, we have compressed some datasets. We guess that the data corruption may be caused by the decompression software of different operating systems. We have updated and checked every dataset in the data repository. We have no intention of concealing any marked data files. We are not only data producers, but also data users. Ensuring the integrity of data so that it can be used has always been our goal. If there is any data missing or corrupted, please feel free to contact us. For more details, please see our replies below.

In your response you add two figures R2/R3 as well as argumentation about certain subjects. If these figures/argumentation is important for the understanding of the manuscript they should be included there, and not only in the response of the authors. IMHO the figures R2 and R3 would be nice to have in the full paper.

Response: We agree. We have redrawn these two figures and put them in the appendix of the manuscript.
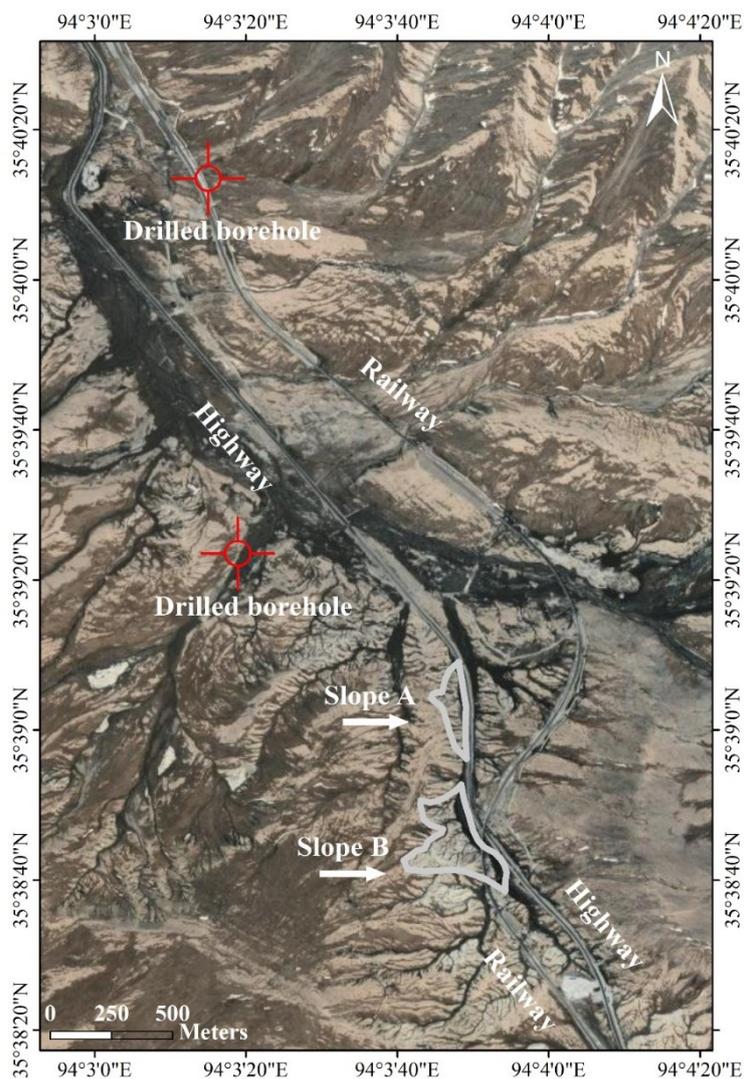
Appendix B: Drilling data source

Figure B1: The location of two drilled boreholes.
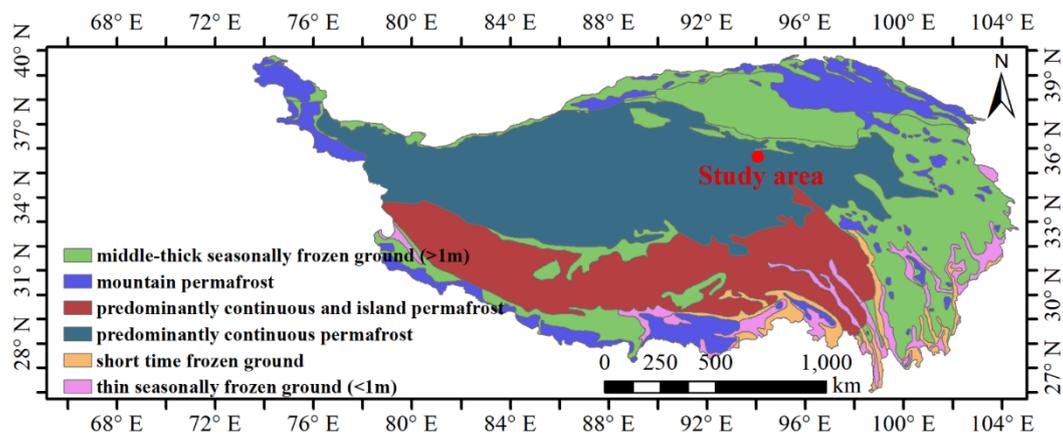
Appendix C: Classification of frozen soil

Figure C1: The frozen soil distribution in the study area.

Concerning data cleaning/validation you mention a column "corrected_p" and give a table 3 on correction values. however i cannot find these in any data file provided.

Response: Sorry, due to our mistake, the prepared data has not been uploaded to the data repository correctly. In fact, the "corrected_p" column is included in the data file that we uploaded before starting with "TRPFUPS" (DOI: 10.5281/zenodo.4588099). In order to solve the problem of multiple meteorological data files for a weather station, we merged multiple weather files from a weather station into one file.

You mention a multitude of data corrections as well as gap filling. The details remain unclear and especially w.r.t. figure 2 the utility is not given. you simply interpolate over large gaps of ~a year of data. That is not correct here. remove the green daily means at least for the large gaps.

Response: Gap fill is better for applications that lack a few data in the middle and have data before and after. It is unreliable for long-term lack of data. The quality control code 8 for meteorological data is used to fill in missing data. The missing data in Figure 2 mainly occurred from July 7, 2017 to October 3, 2017, which was caused by instrument failure. Filling in the data for this period is a difficult task. The drawing program in Figure 2 is in the 40-69 lines of the Meteorological.R code in "R code for permafrost indices and visualization". We use ggplot2 library to draw, and we need to convert multiple columns of data into single column data. The converted data has also been placed in the Data directory. The visualization of precipitation is a bar graph, and the other graphs are line graphs, so the precipitation data is visualized separately and then spliced with the previous graph. The files are XDTMS2014-2018_GT.csv and XDTMS2014-2018_PREC.csv. We have given instructions in the README file.

The README files you have created greatly simplify the data access. However the contain incorrect/incomplete information w.r.t. the data files provided. A number of the files listed in table 2 cannot be found in the repository or they are named slightly different. Furthermore the data variables listed in table 3 cannot be found. As such the

data description paper as well as the dataset are not in agreement and should be rejected.

Response: Sorry because of our incorrect naming and uploading. The data in Table 2 is divided into different directories according to different dates. In fact, the two data files (DOI: 10.5281/zenodo.4588099) Golmud_52818_2010_2018.xlsx and Wudaoliang_52908_2010_2018.xlsx contain part of the QC quality control in table 3. We merge multiple data files of each weather station into a ground and a weather csv or xlsx format data, and we checked and added all missing data files.

As a last comment i don't think it is necessary to make a copy of the readme file in each directory as well as in the supplement and it's subdirectories. rather than having many (copied) README file, concentrate on generating one that contains full and correct information.

Response: We kept a README file with complete information and updated this README file.

(In the README to your data)

Data Sources and Terms of Use

The use of data is conditional on citing the original data sources. Full details on how to cite the data are given at the bottom of each page. For research projects, if the data are essential to the work, or if an important result or conclusion depends on the data, co-authorship may need to be considered. Permafrost engineering and slope monitoring facilitate the acquisition of data to encourage its use and promote understanding of the potential impact of freeze-thaw cycles on Permafrost engineering. Respecting original data sources is key to help secure the support of data providers to enhance, maintain and update valuable data.

Response: All data in this manuscript follow the Apache License 2.0 (thanks to the editor for suggesting that we choose this license in the early stage of the manuscript), so we deleted this term in the README.