Responses to Reviewer #1

Text in red are the reviewer's comments; **those in black** are the authors' replies and explanations to the reviewer's comments; and those in blue are the revised texts appeared in the revised manuscript.

Dear authors,

this paper develops a refined model for permafrost distribution on the Qinghau-Tibet Plateau (QTP) based on a prior publication in Permafrost Periglac. (Hu et al. 2020). The extensions of the model contain the introduction of further metrics (F) aiming at guaranteeing boundary consistency and ensemble simulation. The paper is very extensive w.r.t. the model description (7 pages) and discussion of the results (15+ pages). The output data, i.e. a permafrost map of QTP as well as thawing/freezing indices and soil clusters in the form of 1km grid raster files are provided as data artefacts. As such this paper is very much a methods paper and not a typical data description paper targeted for publication in ESSD. It is commendable however that this paper makes use of recently published consolidated ground surface data from QTP (Zhao 2021) for model validation.

Response:

Many thanks for your comments on our work. This paper aims to provide a QTP permafrost map as a potential benchmark for modeling results, with elaborate methodological descriptions and relevant discussion. A benchmark map requires very high accuracy ensured by robust, reproducible methodology and a solid data base. During the writing process, we also learned from similar papers published in ESSD (e.g. Friedl et al., 2021; Chen et al., 2021; and Ran et al., 2022).

In this revision, as requested by the reviewers, we have shortened the paper by moving descriptions of some sub-procedures that are not core to the methodology to the supplementary materials in the hope of focusing more on the map and its specific methodology.

The moved contents include 1) the introduction to the solar-cloud-satellite geometry (SCSG) based interpolation approach (lines 226-239 in the original manuscript), 2) two approaches for estimating the annual thawing index (lines 252-277 in the original manuscript) and the corresponding results (lines 386-393 in the original manuscript), and 3) the instance of 'boundary cell' to help understand our proposed concept of boundary consistency (lines 305-345 in the original manuscript).

Refs:

Chen, Y., Liang, S., Ma, H., Li, B., He, T. and Wang, Q.: An all-sky 1km daily land surface air temperature product over mainland China for 2003-2019 from MODIS and ancillary data, Earth Syst. Sci. Data, 13, 4241-4261, 10.5194/essd-13-4241-2021, 2021.

Friedl, P., Seehaus, T. and Braun, M.: Global time series and temporal mosaics of glacier surface velocities derived from Sentinel-1 data, Earth Syst. Sci. Data, 13, 4653-4675, 10.5194/essd-13-4653-2021, 2021.

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J., Hori, M., Yu, Q. and Chang, X.: New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere, Earth Syst. Sci. Data, 14, 865-884, 10.5194/essd-14-865-2022, 2022.

Concerning the refined model I have the following comments:

You extend your the model by ensemble simulations. Yet the chosen parameter of 1000 runs seems arbitrarily picked. It is unclear how the number of runs affects your solution (accuracy/performance etc). In order to judge the utility of your proposed method this should be investigated.

Response:

As values of parameter E were estimated by the particle swarm optimization approach and one single time of optimization could suffer from inappropriate initial values and local optima, we chose to run the parameter optimization many times and let the result be determined by the majority voting of these estimates.

However, if the number of runs is not big enough, it is still possible that most estimates of these runs are inappropriate, compromising the efficiency of majority voting. Therefore, the number of runs should be large enough. We have investigated how the mean values and standard deviations of E estimates change with the number of runs (Fig. R1). The mean values of E become stable rapidly when the number of runs is over 100. While the standard deviations of eight soil clusters all become stable when the number of runs is over 700. It indicates that 700 runs could be enough to produce a suitable set of estimates for majority voting. Therefore, we chose 1000 times runs in this study.

We then investigated how the number of runs affects the final map. We produced permafrost maps voted by one to ten hundred runs, and find there are only very minor differences (<0.005%) between them. Though 100 seems enough, we still choose a larger number of runs (e.g. 1000) in our study to ensure stable standard deviations of E estimates (Figure R1).

The above discussion has mentioned in the manuscript briefly.

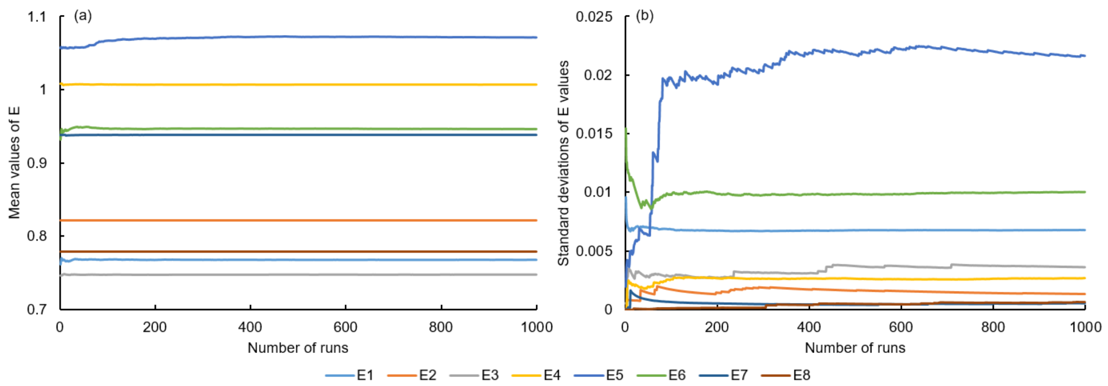

Figure R1. Changes in mean values and standard deviations of soil parameter E along with number of parameter optimization runs (not shown in the revised manuscript)

The distribution of ground control data used (weather stations/GST) in QTP is very inhomogeneous with many stations in the East and few in the west. This is nicely illustrated in (Zhao 2021). However your model approahc seems to not take this into special consideration. How do you deal with very sparse ground control data?

Response:

As a fact, the observation sites on the QTP are indeed unevenly distributed, and this situation is quite inevitable due to the harsh environment, which prevents the observation of more ground control data, especially in the west QTP regions. In our estimation of ground surface temperature (GST) from land surface temperature (LST), the established method, however, can effectively handle the effects of spatially inhomogeneous distribution of observation sites on the QTP.

In our approach, 0cm ground temperature (GST) measurements from weather stations were used to correct for the thermal offset between LST and GST. Based on our and others' (Wang et al., 2011) investigation, the GST-LST thermal offset is only significant on the eastern QTP during the growing season because of the vegetation cover. In contrast, in the western QTP where vegetation cover is rather low (NDVI < 0.1) (Fig. R2) or in seasons other than the growing season, the thermal offset is almost negligible. This is why we didn't carry out a correction for the freezing index. Our multilinear regression model established to correct for the thermal offset is a function of NDVI and latitude, predicting a small thermal offset when NDVI is small. Therefore, our model can effectively work for both eastern and western QTP even though the relationship is built from sites mostly on eastern QTP.

What's more, though thermal offsets between GST and LST cannot be fully eliminated through the regression model, the effects of residual offsets can be further reduced in the optimization phase of our approach, as the effects of GST-LST thermal offsets can be compensated by adjusting values of soil parameter E to achieve a best possible agreement with the survey-based subregion permafrost maps. Two survey-based subregion permafrost maps (West Kunlun and Gaize) can represent the environmental conditions of permafrost in the western QTP.

We improved relative descriptions on how this method works also for western QTP where few weather sites are available.

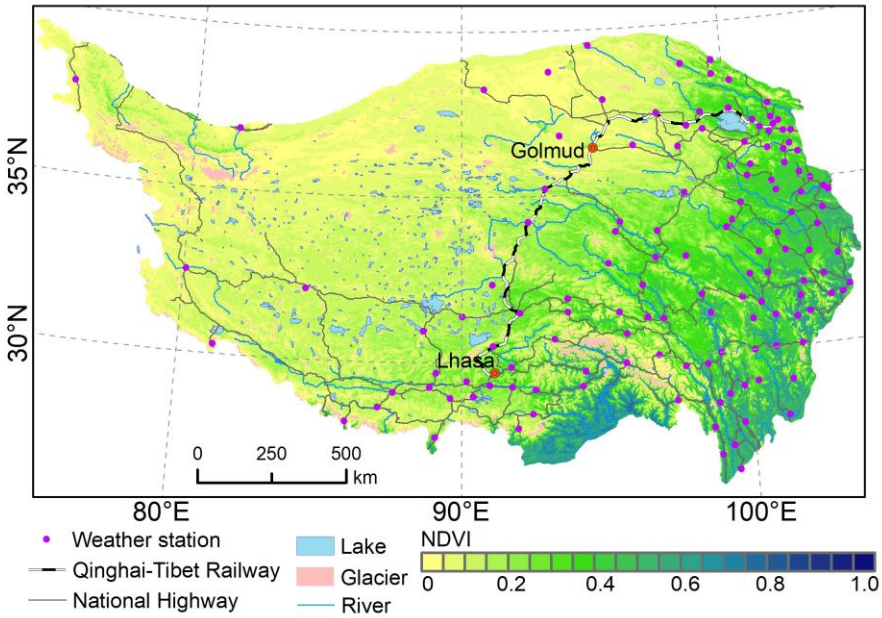

Figure R2. Map showing the normalized difference vegetation index (NDVI) distribution on the Qinghai-Tibet Plateau, the locations of meteorological stations. (not shown in the revised manuscript)

Ref:

Wang, Z., Nan, Z., and Zhao, L.: The applicability of MODIS land surface temperature products to simulating the permafrost distribution over the Tibetan Plateau. Journal of Glaciology and Geocryology, 33(1), 132-143 ,2011. (in Chinese)

There exist a number of other models to estimate permafrost (extent) and thermal regime covering e.g. the whole northern hemisphere and available at 1km grid cell size: Youhua Ran, Xin Li, Guodong Cheng, Jingxin Che, Juha Aalto, Olli Karjalainen, Jan Hjort, Miska Luoto, Huijun Jin, Jaroslav Obu, Masahiro Hori, Qihao Yu, and Xiaoli Chang: New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere. https://doi.org/10.5194/essd-14-865-2022. How does your approach and results derived compare to these "global" models? Especially since when considering the whole northern hemisphere much more ground data exists?

Response:

Our study aims to provide a historical reference for permafrost simulation studies. With this a goal, we believe that our map and the map of Ran et al. (2022) (hereinafter, Ran map) are not entirely comparable for the following reasons.

First of all, to provide a historical reference of 2010, the forcing data in our study are all from 2005 to 2010, but 99% of measurements used to produce Ran map were made during 2000-2016 (Ran et al., 2022). It means that Ran map may represent an averaged permafrost status of a longer period of time, which is not intended to serve as a historical reference. In addition, Ran map is estimated by the ensemble mean of 1000 runs of four machine learning methods trained by in-situ MAGT and ALT, which is methodologically similar to the Wang map (Wang et al., 2019) (We compared ours with Wang map in our study). Such mapping methods can be satisfactory in circum-Arctic areas (e.g. Siberian and north Canada), because in these areas, evenly-distributed in-situ observations are available and land surface is more homogeneous than Plateaus. But on the QTP, machine learning based approaches may not be that effective due to inadequate ground data and insufficient consideration of local factors.

We have downloaded the dataset of Ran et al. (2022) and compared the permafrost extent (defined as permafrost probability > 0 according to fig.7 in Ran et al. (2022)) with our map and other existing QTP permafrost maps cited in our study (Fig. R3, Table R1, Table R2). Compared with survey-based permafrost maps and borehole observations, Ran map has similar a performance to Wang map which is also based on machine learning methods. Both of them overestimated permafrost extent on the QTP.

Considering there are much more ground data exists in the whole northern hemisphere (especially in circum-Arctic areas), we indeed plan to extend our approach to mapping the permafrost distribution in the whole northern hemisphere.

We have no plan to include Ran map in our manuscript as a comparison reference for our map, as ours and Ran map were created for different purposes.

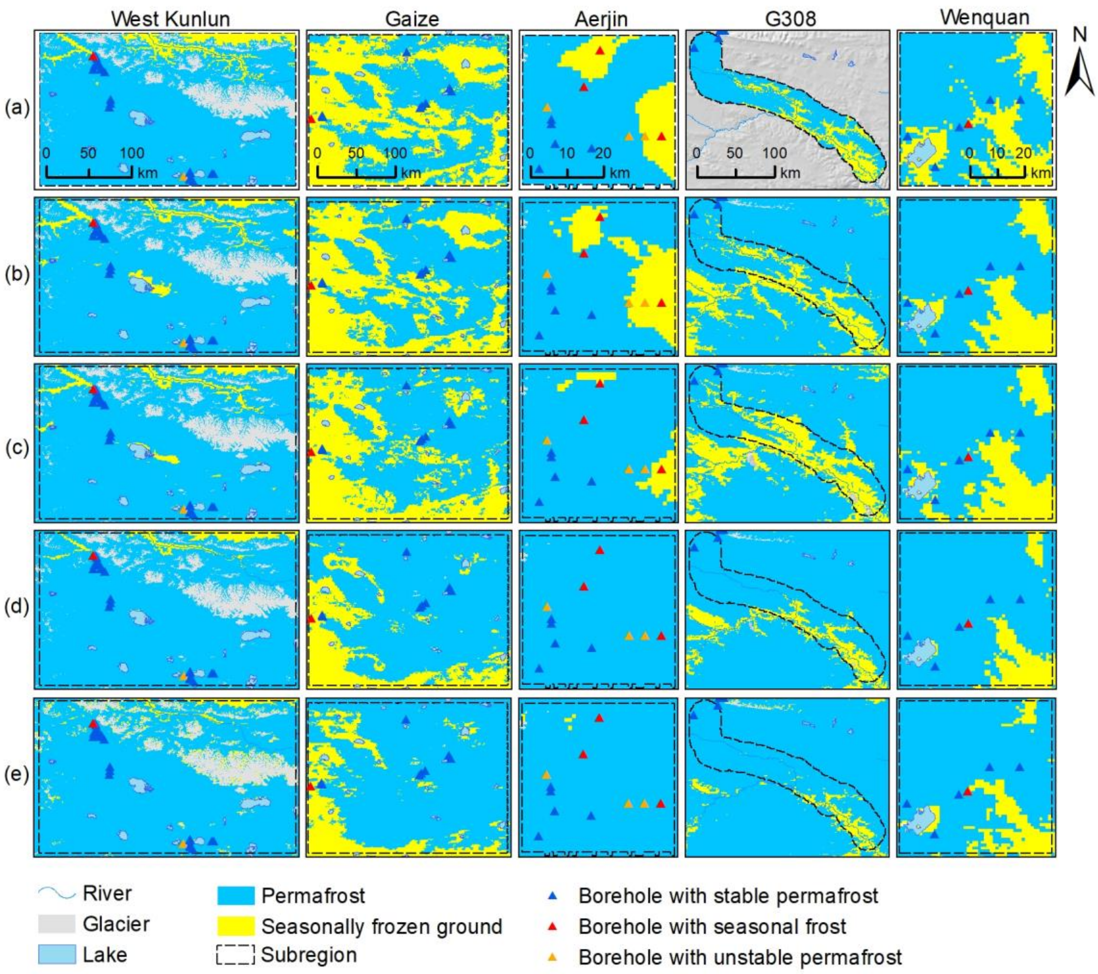

Figure R3. Spatial distributions of frozen ground in subregions from (a) survey-based maps, (b) our map, (c) the map from Zou et al. (2017), (d) the map from Wang et al. (2019), and (e) the map from Ran et al. (2022). Triangle symbols mark the locations of boreholes drilled around 2010 where the type of frozen ground was identified.

Table R1. Kappa values measured between the evaluated maps (our map, Zou map, Wang map, and Ran map) and survey-based permafrost distribution maps in the subregions.

	West Kunlun	Gaize	Aerjin	G308	Wenquan	All subregions
Our map	0.62	0.71	0.71	0.68	0.70	0.74
Zou map	0.63	0.48	0.38	0.46	0.65	0.55
Wang map	0.63	0.38	0.00	0.68	0.46	0.50
Ran map	0.36	0.32	0.01	0.49	0.49	0.39

Table R2. Measures of confusion matrices describing the performance of the evaluated maps (our map, Zou map, Wang map, and Ran map) at the borehole locations. To fit the binary classification, permafrost is regarded as positive and seasonally frozen ground negative. n=72.

	Our map	Zou map	Wang map	Ran map
True positives (rate)	55 (90.2%)	53 (86.9%)	61 (100.0%)	61 (100%)
False positives (rate)	5 (45.5%)	6 (54.5%)	10 (90.9%)	8 (72.7%)
True negatives (rate)	6 (54.5%)	5 (45.5%)	1 (9.1%)	3 (27.3%)
False negatives (rate)	6 (9.8%)	8 (13.1%)	0 (0.0%)	0 (0%)
Accuracy	84.7%	80.6%	86.1%	84.7%
Cohen's Kappa	0.43	0.30	0.14	0.39

Refs:

Ran, Y., Li, X., Cheng, G., Che, J., Aalto, J., Karjalainen, O., Hjort, J., Luoto, M., Jin, H., Obu, J., Hori, M., Yu, Q. and Chang, X.: New high-resolution estimates of the permafrost thermal state and hydrothermal conditions over the Northern Hemisphere, Earth Syst. Sci. Data, 14, 865-884, 10.5194/essd-14-865-2022, 2022.

Wang, T., Yang, D., Fang, B., Yang, W., Qin, Y. and Wang, Y.: Data-driven mapping of the spatial distribution and potential changes of frozen ground over the Tibetan Plateau, Sci. Total Environ., 649, 515-525, https://doi.org/10.1016/j.scitotenv.2018.08.369, 2019.

Zou, D., Zhao, L., Sheng, Y., Chen, J., Hu, G., Wu, T., Wu, J., Xie, C., Wu, X. and Pang, Q.: A new map of permafrost distribution on the Tibetan Plateau, The Cryosphere, 11, 2527-2542, https://doi.org/10.5194/tc-11-

2527-2017, 2017.

You specifically mention that your dataset/model is representative for the year 2010 and tries to not incorporate all historic data but only such observations that are relevant to the specific mapping year. yet in multiple places you incorporate data or discuss processes over longer periods of time, e.g. 2005-2010. How did you arrive at this again arbitrarily picked boundary of 2005? and why not incorporate data beyond 2010?

Response:

In our approach, the optimization targets subregions maps based on field works conducted in 2009-2010. Freeze/thaw indices over 2005-2010 are used as forcing data to drive the model. There are several reasons why we derived freeze/thaw indices based on a short period before the target year (2010) instead of the single target year (2010).

First, permafrost is defined as the frozen ground for at least two consecutive years and buried underground, permafrost situation in a specific year can be affected by climate conditions of several years before but cannot be affected by years after, thus we didn't incorporate data beyond 2010.

Second, a previous study on the mapping approach (Hu et al., 2020) show that forcing data from five to six years of are suitable for permafrost mapping.

Third, we actually used an average climatic condition over 2005-2010, to avoid the possible influence of abnormal single-year meteorological conditions.

Last, many automatic weather stations which provide GST measurements were put into operations since 2005, so we chose a period 2005-2010.

We provided justifications in the revised manuscript regarding the selection of 2005-2010 for deriving freeze/thaw indices as forcing to our mapping approach.

Ref:

Hu, J., Zhao, S., Nan, Z., Wu, X., Sun, X. and Cheng, G.: An effective approach for mapping permafrost in a large area using subregion maps and satellite data, Permafrost Periglac., 31, 548-560, https://doi.org/10.1002/ppp.2068, 2020.

Concerning the dataset I have the following comments

You present a number of comparisons of your resulting map with other data (Zhou map). In order to continuously be able to improve this modeling work (and reuse your data) it would be helpful if the comparison data was made available as well.

Response:

Currently, our open dataset coming with this paper does not include these comparison data in order to avoid possible copyright infringement. To facilitate access to these data, we created a special section in the supplementary materials to provide information on where these comparison data can be downloaded, as well as links to our open dataset.

It would be very helpful for the community if your code would be made available with this model study.

Response:

Yes, codes will be made publicly available on github (https://github.com/nanzt), as well as on our research group website at https://permalab.science/publications/codes, where codes of a modified Noah LSM model associated with a previous paper (Earths Futures) were already freely available.

Similarliy it would be very helpful if you could list the exact sources of all input data used in one place, similar to your data products.

Response:

Thanks for the suggestion. We created a special section in supplementary materials to summarize accessible sources of all input data, comparison data, as well as resulting datasets and codes of mapping.

I have downloaded your dataset and was able to load this into QGIS.

Response:

Thank you for your attention to our work and dataset.