Response to Referee 3

The authors would like to thank the reviewer for the constructive feedback, and the thorough assessment of the manuscript. Below, we are providing a point-to-point response to each comment: Reviewer comments are given in black, and our responses are given in blue. Additionally, we have included details of how we address these changes in the revised submission.

General Comments

This study conducted extensive ground surface temperature measurements in the Headwater Area of the Yellow River on the Qinghai-Tibet Plateau, providing abundant and valuable data for permafrost research in the QTP region. Based on the acquired data, the authors also conducted a detailed analysis and provided readers with insights into the possible applications of the current data in soil freeze/thaw research. The paper is well-organized, and the writing is clear and easily readable. However, there are still some issues that should be addressed before final publication.

Thank you for your kind summary and appreciation!

Specific comments:

1. In page 5, line 121, “local-scale sites are established in a flat peat plateau”. In page 6, line 137, “some sites are covered by coarse gravel”. Peat soils or gravel soils have distinct properties compared to fine mineral soils, and QTP is generally characterized by widespread gravel soil and generally low soil organic content (SOC). Therefore, more information should be included. For example, are there any measurements of topsoil organic content? Is

the SOC in topsoil related to the intra-plot differences at sites all covered

by same vegetation? Does the site covered with coarse gravel have any

influence on the analysis results?

We do not have measurements of topsoil organic content but we have grain-size analysis and measurements on the water content for 11 sites situated in alpine meadow and bare ground. Unfortunately, we do not have soil samples from both plots of a site and only from one plot per site. Thus, we cannot consider the soil texture in the fine-scale analysis of the intra-plot differences regarding the ground surface temperature (GST). However, this suggestion is an interesting point that we will consider in the next field campaign, and will collect additional soil samples from key sites and plots to perform more analysis on soil characteristics.

Regarding the last question, we performed in the past grain-size analysis on the available soil samples and added more information in the manuscript. The following changes have been made:

L168-169: "From 11 plots, soil samples were collected (Table 1) for grain size and water content analysis. Samples were weighed before and after being dried at 105 °C for 16 hours to determine the water content. The coarse texture represented by gravel (> 2 mm) was quantified by sieving, while the fine texture (< 2 mm) representing sand, silt, and clay was measured with a Malvern Mastersizer-2000 laser diffraction particle size analyzer." has been added.

L304-311: "To better understand the variability of MAGST under the influence of landcover, even under the same landcover type, the MAGST was compared with the soil texture and soil water content (Fig. 10, Table 3). All the samples from the local and landscape scales collected from bare ground and with a fine texture of above 75% (Fig. 10a) revealed a low MAGST ranging between –2.2 and –1.2 °C (Fig. 10b). The samples from the lower part of the elevational transect (up to 4432 m) revealed a MAGST between 1.3 and 1.7 °C, regardless of landcover type (meadow or bare ground) or texture. Three of them had a fine texture between 70 and 98%, except plot B3A with the lowest fine texture of 59%. Only plot B8A from bare ground with a fine texture of 73% had the lowest MAGST of 0.2 °C among the elevational sites sampled for soil texture. However, the elevation of plot B8A is still relatively low, only 4473 m a. s. l." has been added.

L314-317: "Figure 10. Textural classification of the soil samples from selected plots (a) and their corresponding mean annual ground surface temperature (MAGST) for the period August 2019 to July 2020 in south-central HAYR (b). Notes: Coarse soil texture is represented by gravel while fine soil texture by sand, silt, and clay, with the threshold between them set at 2 mm in diameter." has been added.

L320-321: "Table 3. Grain size distribution and water content of the soil samples from selected plots." has been added.

2. The authors mentioned that this dataset can be useful as inputs or validations for permafrost and SFG models. Given the high spatial resolution of GST monitoring, providing information about soil texture or soil type at each site would be beneficial for model simulations and further analysis.

Indeed, the soil type and texture are very useful for a better parameterization of physical models. Unfortunately, only from 11 plots were carried out analysis on soil texture as we detailed in the response to the previous comment. Results of this grain-size analysis were added in Table 3 and Figure 10. Because the soil analysis does not cover all the sites we emphasized only the utility of the GST dataset for validating other models or remote sensing products. Furthermore, the GST dataset can be also used to improve the upper boundary conditions in simulations of soil temperature or permafrost distribution by using physical models.

3. The elevation cross-section is located on the northern side of the Bayan Har Mountains. Is there any information available regarding the slope and aspect of these locations? Does it have any impact on the results?

The information regarding the slope and aspect of the monitoring sites are detailed in Table 1 from Şerban et al., 2023. In the respective work, statistical tests (Pearson correlation, linear models, and analysis of variance – ANOVA) were performed to identify the environmental controls on GST variability in this area. The results showed that slope and aspect did not have any statistically significant influence on the GST variability and only the landcover and elevation. In fact, in this region the topography is relatively smooth. In this data paper, we avoided repeating the same information and analysis and we focused more on the intra-plot variability of GST and added only a reference.

The following changes have been made:

L323-325: "The intra-site MAGST variability has been mainly controlled by elevation and landcover types (Şerban et al., 2023), similar to observation from the Swiss Alps (Gubler et al., 2011)." has been replaced with "The intra-site MAGST variability has been mainly controlled by elevation and landcover types (as is shown in Figs. 4 and 8 of Şerban et al., 2023), similar to observation from the Swiss Alps (Gubler et al., 2011). Slope angles and aspects do not play a relevant role because the monitoring plots are located mostly in flat areas (Şerban et al., 2023)."

4. One of the multiscale settings is the "fine scale," ranging from 2 to 16 meters. The authors stated that the fine-scale measurements were set for backup reasons and to identify the variations in GST. What were the criteria for setting two plots at each site? This scale is hardly matching the modeling or remote sensing applications. What are the potential applications of observations at the fine scale?

Indeed, in line 140 we said “This was done due to backup reasons and to identify the variations in GST at a fine scale.” The main reason was to identify the variability of GST across various short distances ranging from 2 to 16 m under similar topographical conditions and differences only in terms of the landcover type. For several sites, even the landcover type was similar and only the intra-plot distance was different. Complementary, to the fine scale comparisons of the GST evolution in this data paper, in Șerban et al., 2023 were emphasized the intra-plot differences in MAGST according to the intra-plot distance. The differences in MAGST were higher when the intra-plot distance was above 8 m. However, there was no statistical significance probably due to the low number of samples for the statistical test. However, the intra-plot difference in MAGST was more clear when the plots from bare ground were compared to vegetated ones.

The second reason for two plots in each site was for backup reasons because as can be seen in Table 1 that in several plots the sensors failed to acquire a complete timeseries of data. Therefore, the data available in the other plot was used for the intra-sites comparison, and therefore in all sites, there was at least one plot available with a complete timeseries of GST. An exception was only in site B18 where the data is not complete in both plots. Details on the missing data and incomplete timeseries are in subchapter 3.1 Data quality check (L.203-224).

The scale does not match the spatial resolution of the most common free remote sensing products from satellite images but matches the special resolution of the thermal images from unmanned aerial vehicles (UAVs) and airplane images. This scale of 2 to 16 m is relevant for high-resolution modeling and remote sensing products. For example, the spatial resolution of both optical and thermal bands is between 0.3 and 4 m for many commercial high-resolution sensors, such as WordView, GeoEye, KOMPSAT, Quickbird, Pléiades, SkySat, IKONOS, or GaoFen. The free and common satellite images of Sentinel and Landsat products also have a spatial resolution between 10 – 100 m). High spatial resolution modeling approaches often go to 10 m resolution. The fine scale variability can give an estimation of the internal variability at pixel scales for such applications. However, the spatial resolution of satellite images and computational power for numerical models is in a continuous improvement, and sooner or later more products will reach the resolution of our dataset.

Finally, an important potential application of this fine scale database of GST is for simulations of the energy exchange in this dynamic environment at the land and atmosphere boundary. The 1D simulations of the energy fluxes in the process-based models will benefit from this dataset both for validation and for parameterization. These models could help to better understand this fine scale variability of GST under the influence of landcover.

GST is important not only for characterizing periglacial processes and permafrost and seasonally frozen ground evolution but could be useful for various geosciences and economic applications. For example, soil science still relies on the air temperature as input while GST is more adequate for microbiology studies and ecosystem monitoring. Even for modeling soil temperature and permafrost distribution the air temperature and land surface temperature are often used, while studies have shown the large surface offsets with GST. In precision agriculture, the GST can be used for scheduling irrigations, combating droughts and freezing for a sustainable development and management of resources.

5. The intra-plot differences at most sites are usually larger during the freeze-thaw transition period (Figs. 3-6), but at site B6 and B7, the same pattern is

not observed and the differences are large throughout the entire year (Fig.

5). What are the possible reasons? Is there anything special about these

two sites?

Now that we uniformized the Y scale with the same temperature range, for these two sites we can see that the daily differences are still reaching the maximum values during the freeze-thaw transition period. However, for the other periods of the year, even though the differences are lower they still show a spiky pattern. We believe now that the fine scale variability of the water content in the shallow soil could explain this pattern. This might be more relevant for site B6 placed in swamp meadow where even the surface water was present in a patch pattern. However, the water content can also show high variability in the bare ground sites. For example, the water content was one of the lowest in plot B7B (8%) compared to other plots from the bare ground sites that reached even 44% (Table 3). Therefore, as we stated in the reply to the first comment, additional soil samples should be taken from these key sites and plots that raised more questions. Having more parameters to compare concerning the soil properties, water content, and organic matter content, could better explain the fine-scale differences in GST even under similar landcover types.

6. Table 2. It's not surprising that R or R2 values are close to 1, but the RMSE

or MAE provide more insightful information regarding GST variation at different scales. Additionally, investigating potential relationships between

GST differences and environmental factors like elevation might be helpful.

Including a figure to visualize these relationships could enhance the clarity

of the analysis

Indeed, as also stated by the first two referees, the main aim of this data paper is just to make available new GST data for the scientific community. A detailed analysis of the controlling factors on GST variability was performed in our previous work. The control of elevation on GST spatial variability has been assessed in detail by statistical tests and including a graphical representation of the decrease of GST with elevation (please see Figures 4 and 8 from Şerban et al., 2023). In this data paper, we avoided repeating the same analysis and figures and we focused more on the intra-plot variability of GST. We only briefly mentioned:

L83-84: "The variability of MAGST at other scales and their environmental controls have been assessed in detail by Şerban et al. (2023)."

L323-324: "The intra-site MAGST variability has been mainly controlled by elevation and landcover types (as is shown in Figs. 4 and 8 of Şerban et al., 2023), similar to observation from the Swiss Alps (Gubler et al., 2011)."

Technical corrections:

Figure 1: add the lat/lons infromation, and adding a permafrost map as the

background may be also helpful.

The lat/long coordinates have been added as suggested on the inset map that shows the study area in the south-central Headwater Area of Yellow River (HAYR). On the same map, in the background, it is already added the permafrost distribution after Wang et al., (2005) as described in the caption. The permafrost distribution layer is represented with dots and a brown boundary and described in the legend as "Plateau discontinuous permafrost".

line 119-121: are these data from site CLP-1 or CLP-2?

These data are from borehole CLP-2 because that is the deep borehole of 100 m in depth, while CLP-1 is 20 m in depth. More details are in Luo et al., 2018b.

Line 156: “Photographs were taken at each site and plot”. I would suggest the authors add some photos to better present sites condition.

Photographs illustrating the site and plot conditions have been added as suggested.

The following changes have been made in the manuscript:

L163-166: “Figure 3. Photographs presenting the monitoring plots of GST in different landcover types: alpine steppe and bare ground – B1 (a); earth hummocks in alpine swamp meadow – A8 (b); fine bare ground – A4 (c); coarse bare ground – D1 (d); fine bare ground in the depression of a drained thermokarst pound and in the nearby alpine meadow – A2 (e); alpine meadow – B4 (f).” has been added.

Line178: please briefly describe what AIC is.

L195-196: “The AIC is a statistical test used to assess how well the model fits the data (Akaike, 1974).” has been added.

Line 222: change “both” to “these two”

L234-235: “At both sites, the plots are situated in a …” has been replaced with “At these two sites, the plots are situated in a ...”.

Figure 3-7: I would suggest the authors using same Y scale to better show the differences.

A Y scale ranging from –4 to 4 °C has been used for all the plots in Figures 3-6 (now Figures 4-7), while a Y scale ranging from –15 to 15 °C has been used for all plots in Figure 7 (now Figure 8).

Figure 7d: the color difference between the two lines is minimal, making it difficult to distinguish the line representing "steppe."

The color of the steppe has been changed from yellow to purple to better be distinguished from the orange of the bare ground.

Figure 8: I would suggest sorting the sites in transect by elevation to better

present if there are elevation effects.

The sites in the transect have been sorted by elevation in both Figures 8 and 9 (now Figures 9 and 11).

References

Akaike, H.: A new look at the statistical model identification, IEEE Trans. Automat. Contr., 19, 716–723, https://doi.org/10.1109/TAC.1974.1100705, 1974.

Gubler, S., Fiddes, J., Keller, M., and Gruber, S.: Scale-dependent measurement and analysis of ground surface temperature variability in alpine terrain, Cryosphere, 5, 431–443, https://doi.org/10.5194/tc-5-431-2011, 2011.

Luo, D., Jin, H., Jin, X., He, R., Li, X., Muskett, R. R., Marchenko, S. S., and Romanovsky, V. E.: Elevation-dependent thermal regime and dynamics of frozen ground in the Bayan Har Mountains, northeastern Qinghai-Tibet Plateau, southwest China, Permafr. Periglac. Process., 29, 257–270, https://doi.org/10.1002/ppp.1988, 2018b.

Șerban, R. D., Bertoldi, G., Jin, H., Șerban, M., Luo, D., and Li, X.: Spatial variations in ground surface temperature at various scales on the northeastern Qinghai-Tibet Plateau, China, Catena, 222, 106811, https://doi.org/10.1016/j.catena.2022.106811, 2023.