Response to Referee 1

The authors would like to thank the reviewer for the constructive feedbacks and kind advice, 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 paper describes a ground surface temperature (GST) monitoring network established in a specific region of the Qinghai-Tibet Plateau. Temperature sensors were deployed across areas of varying surface characteristics to monitor changes in GST under different landcover conditions. The collected monitoring data is abundant and of reasonably high quality. The authors have conducted a thorough analysis of the acquired data, providing readers with a more in-depth understanding about the freeze-thaw state during that period. Overall, the English writing in this paper is clear and coherent, and the obtained data can serve as valuable input for modeling or validation of surface processes. However, there are still some issues that the authors should consider. I would be highly appreciated if the authors could address them.

Thank you for your kind summary! We have tried our best to address the raised issues as follows.

Specific comments:

1. What is the difference between the ground surface temperature (GST) mentioned in the paper and the land surface temperature (LST) commonly referred to in the remote sensing field, as well as soil temperature? Additionally, the description "topsoil temperature" in the data website provided by the authors raises questions about the physical meaning of the variables discussed in the paper. It is recommended that the authors either standardize their terminology or provide additional explanations within the text to ensure a clearer representation.

   The definition of the ground surface temperature (GST) is at lines 65-66: "GST is usually measured at approximately 5 cm into the ground but in literature, the GST depth was varying from 2 to 10 cm (Ferreira et al., 2017; Grünberg et al., 2020; Oliva et al., 2017; Onaca et al., 2015)".

   The land surface temperature (LST), measured either by remote sensing sensors or in-situ sensors, is the temperature at the surface of the landcover or on the top of the landcover. Thus, it is directly exposed to solar radiation and it is also known as the "skin temperature". We didn't add explanations of LST in the manuscript because it is not measured and does not represent the subject of this paper.

   We use the term "soil temperature" only when we refer to other studies with the temperature measured at different depths than the depth of GST at 5 cm. For example, at depths of 20 cm (line 306) or 10 cm (line 392).

In the data repository, we used the term "topsoil temperature" only in the title of the data more as a synonym of GST. In the summary of the dataset, we described that we refer to GST, which is measured at a depth of approximately 5 cm.

2. In the Introduction section, the authors mentioned that some scholars have already deployed GST monitoring networks in the northeastern part of the Qinghai-Tibet Plateau (e.g., Luo et al., 2020; Serban et al., 2023). What distinguishes the observational data in this study from those previous efforts? Perhaps the authors placed their monitoring network in mountainous regions? However, it seems that the data analysis by the authors did not include a specific analysis of mountainous characteristics. Despite some sections discussing elevation, the more unique features of mountainous regions such as three-dimensional structure and illumination conditions were not addressed.

   Indeed, Luo et al. (2020) measured the GST but only at a few sites and on a small flat area (3.5 km^2) at Chalaping close to our sites from the local scale. This newly established monitoring network is covering a larger range in terms of elevational range and landcover types. However, not so much in terms of slope and aspect because the monitoring plots are located mostly in flat areas. The study area is on a high plateau with smoothed interfluves and peaks and the illumination conditions do not differ substantially.

   Serban et al. (2023) analyzed the GST from this database but focused more on the intra-site comparison and detecting the environmental controls on GST variability. In that paper are included specific analysis and statistical tests regarding the environmental variables of topography (elevation and slope angle and aspect) and landcover types. The mountainous regions of the study area are described as well. We briefly referred to that in lines 83-84: "The variability of MAGST at other scales and their environmental controls have been assessed in detail by Serban et al. (2023)."

   L306-308: "The intra-site MAGST variability has been mainly controlled by elevation and landcover types (Șerban 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 Fig. 4 of Șerban et al., 2023), similar to observation from the Swiss Alps (Gubler et al., 2011). Slope and aspect do not play a relevant role because the monitoring plots are located mostly in flat areas (Șerban et al., 2023)."

   L.97-98: "The study area is on a high plateau with smoothed interfluves and peaks and the illumination conditions do not differ substantially." has been added.

3. The title of the paper mentions a "multiscale observation network…" but typically, multiscale implies different sensor observation fields (e.g., ground stations, drones, satellites). However, in this study, all sensors used for observations are ground-based and have the same

observation field, with differences only in their placement. Additionally, it cannot be claimed that the sensors observed data at "local scale", "landscape scale", and "regional scale" because the instruments still provide sparse point observations and do not comprehensively cover an area. In summary, I am concerned about the validity and accuracy of the description "multiscale observation" in the paper.

If we refer to different sensor fields then is not a multiscale but we used this term more to have a way to spatially divide our study area. We also were inspired by similar repositories, such as:

https://data.tpdc.ac.cn/en/data/b6269aeb-8b44-4d03-b514-2c804c2cfc26/?q=soil%20temperature

We used the terms local and landscape scales, as well as the transect just to divide our study area based on its size, the differences in the environmental conditions, and the density of our sites. For example, the local scale represents an area of just 2 km^2 with homogeneous topographical conditions over a flat peat plateau with an elevational difference of only 18 m. In this area, GST is measured at 9 sites. Indeed, these measurements are point observations but because of their density and this relatively small area, we considered them representative of that area and the landcover type where they are placed.

To avoid confusion, we removed the term "multiscale". The following changes have been made:

L.1-2: "Multiscale observation network of ground surface temperature under different landcover types on NE Qinghai-Tibet Plateau" has been replaced with "An observational network of ground surface temperature under different landcover types on northeastern Qinghai-Tibet Plateau"

L.459: "The multiscale observational network" has been replaced with "The observational network"

4. The authors mentioned that some sensors were malfunctioning. What is the current status of these sensors? Are they now operational, or are they still not functioning correctly? Is there a possibility of acquiring more comprehensive observational data in the future?

   The sensors that were malfunctioning were replaced with new ones. It is planned to visit again the sites this October to check their status and download the data for the second time after a long-term measurement due to the COVID-19 epidemic.

5. Page 6, line 154. The authors mentioned a data collection interval of 3 hours for ground observations. Does this mean that data is recorded once every 3 hours, or is it recorded

multiple times and then averaged using a specific algorithm? I suggest providing a brief explanation in the paper for clarity.

The data is recorded once every 3 hours, not multiple measurements and averaged. The following changes have been made:

L.162-163: "…at a 3 h interval" has been replaced with "and the temperature was recorded once every three hours."

6. Page 9, line 206. How were the 165 "biased" data points mentioned in the paper determined? Were they identified through manual inspection or using a specific criterion (e.g., three times the standard deviation screening)?

   The biased values were determined through manual inspection, plotting the timeseries, and checking the minima and maxima. These values were easily detected because represent extreme values, such as –41 ºC or 87 ºC. The 165 biased values are described in the following paragraph:

   L.219-223: "From these, the most severe one was found in plot A3A, with a period of 10 days from 1 to 19 September 2019, with 151 measurements blocked at –41 ºC. In addition, there were another 13 erroneous measurements with temperatures of –41, –39.5, and 87 ºC on 23 and 26 February 2023. The sensor from plot B16B had only one wrong measurement of –7.7 ºC on 17 October 2019, while the other temperature readings during that period ranged from 0.1 to 2.6 ºC."

7. Page 14, line 266. Why is it that a 14-meter distance can observe larger GST differences for the same type of landcover type?

   When we said larger GST differences at intra-plot distances of 14 m, we were referring to all sites and these differences mainly occurred when vegetated plots were compared to the bare ground.

   There were only two sites with both plots in alpine swamp meadow with several days of larger intra-plot GST differences. From them, only site C4 had an intra-plot distance of 14 m, while site D2 had an intra-plot distance of 8 m. Timeseries of these sites are represented in Figs. 7 e and 7f. This was also observed for the mean annual ground surface temperature (MAGST) where the intra-plot differences were below 0.5 °C, especially for sites with the same landcover in both plots.

   Even though we observed these differences at 14-m distances, especially for comparing bare ground to vegetated sites, the statistical tests did not show a significant influence of the intra-plot distance (Please see Serban et al., 2023).

These observations were summarized in lines 291-294: "The intra-plot variability of MAGST was mainly below 0.5 °C, especially for sites with the same landcover in both plots (Fig. 8a). Variations between 1 and 2 °C were observed at the sites where the bare ground was compared to vegetated plots (A10, C3, and B11). This variability was observed mainly at all analyzed distances between plots and according to the analysis of variance and linear regression, it was insignificantly influenced by distance (Șerban et al., 2023)."

For the sites with swamp meadow in both plots and with several days/periods of higher differences in GST (e.g., C4) we assumed the cause is the variability in moisture content. Some plots had a higher moisture content and were oversaturated, even with the presence of surface water around them. While in the nearby plots, soil moisture content was lower without the presence of surface water. Like comparing to the drier vegetation plots from the alpine meadow (L.310-311) the evaporative cooling and the variability of the moisture content may cause a higher thermal offset.

More detailed explanations we provided in Șerban et al. (2023), such as:

"The high soil water content from oversaturated swamp meadows assures a high heat capacity and thermal conductivity of the soils than those in the drier meadows…."

"Rich soil moisture contents or presence of surface water body will retard the ground freezing or thawing due to the huge fusion heat of phase change either ice/water (melting), water/vapor (evaporation), or ice/vapor (sublimation). Freeze-up of icy soils in the active layer or in lakes/wetlands will release heat more efficiently in winter. In the meantime, lower thermal conductivity of dry, thawed/unfrozen organic soils and higher thermal conductivity of ice-rich frozen soils result in higher thermal offsets. At the same time, intense evapotranspiration will cool the ground more effectively in summer along water surfaces that may also absorb more heat, but they are not in the same order of magnitude."

"A reduction in soil temperature variations was also observed in the Arctic caused by the higher thermal conductivity of wet soils and the high heat capacity of water (Aalto et al., 2013)."

"…higher moisture boosted evapotranspiration, which in turn lowered GST (Aalto et al., 2013)."

"Detailed *in-situ* observations on snow cover and soil conditions (texture, moisture, and organic content) are needed to better understand the controls of GST in the HAYR. These soil properties are strongly influencing the soil thermal conductivity, heat capacity, and hydraulic conductivity that affects the soil freeze/thaw processes (Jiang et al., 2020) and subsequently the GST variability."

8. Page 17, line 290. Although the authors have provided some explanations regarding the relationship between MAGST and elevation, it might be more intuitive to include a graphical representation of the MAGST and elevation relationship.

   Indeed, the influence of elevation on GST spatial variability has been detailed assessed, including a graphical representation of the decrease of GST with elevation (please see Figure 4 from Șerban et al., 2023). In this data paper, we avoided repeating the same analysis and we focused more on the intra-plot variability of GST, and added only a reference:

   L306-307: "The intra-site MAGST variability has been mainly controlled by elevation and landcover types (as is shown in Fig. 4 of Șerban et al., 2023), similar to observation from the Swiss Alps (Gubler et al., 2011)."

   Please also see the reply to the comment number two.

9. Page 19, lines 341-348. While it is understandable that the authors compare the results of FDD calculations with previous satellite-based calculations, is it meaningful to compare the results with very distant regions like Antarctica or other islands (especially when the timeframes are not consistent)?

   We considered that besides comparing the FDD and TDD to other works on the QTP, it is worth to also compare it to other permafrost environments for a global overview of these indices. QTP is part of the "Third Pole" region, thus a comparison to the Arctic and Antarctica areas is deemed necessary to better fit with the special issue "*Extreme environment datasets for the three poles*" to each the manuscript is submitted.

10. Page 20, lines 374-376. While the authors mention that GST monitoring can provide a better assessment of the presence or absence of permafrost, they also note the high spatial variability of permafrost thaw. In my view, for an accurate determination of permafrost status, even when using GST as an indicator, a highly dense sensor network would be necessary, which does not seem to be currently feasible. Therefore, the authors need to further explain why they chose GST monitoring for assessing permafrost status over other methods such as borehole measurements (considering factors like cost, convenience, data uncertainty, etc.).

    Boreholes are more precise but expensive and invasive. The heavy machine destroys the grasslands and ecosystems, and the drilling requires a lot of water. Drilling also affects the ground at deeper depths, it thaws the permafrost and it requires several years to become stable again and to record concluding measurements. GST monitoring is a non-invasive method, low cost, and with faster results/measurements. The following changes have been made:

    L.396-403: "It is a low-cost and non-invasive method that can cover even the most inaccessible and remote areas in the rough mountainous terrains. In terms of uncertainty, it is similar to geophysical methods, which could be complementary, but way more convenient in terms of logistics. On the other hand, borehole drilling and followed ground temperature measurements are more precise but prohibitively expensive, inaccessible for rough terrains, consume large quantities of water, and are heavily invasive to the local ecosystems (Noetzli et al., 2021). Moreover, permafrost around the borehole is thawed during the drilling process and it requires several months to years to be able to record concluding ground temperatures (Kutasov and Eppelbaum, 2018). Furthermore, deep boreholes can increase the risk of gas escape to the surface with consequences to local populations, ecosystems, and the climate system (Gizatullin et al., 2023; Klotz et al., 2023)." has been added.

11. Page 22, line 425. While the authors mention the potential significance of this dataset for improving modeling methods, the entire paper analyzes the relationship between GST and freeze-thaw without specifying the advantages of higher spatial resolution GST monitoring data for model improvement (compared to using satellite data). Considering that large-scale snow and ice state analysis typically relies on satellite observations, is there a genuine necessity for such dense sensor deployment?

    Yes, a dense sensor network will help to validate at a higher spatial resolution the satellite data products (e.g., land surface temperatures, snow distribution, reanalysis climatic grid datasets) and the models of permafrost spatial distribution.

    These products are still too coarse to reproduce this high spatial variability of the ground temperature as was observed in the monitoring of GST. Moreover, the actual permafrost models rely on these coarse data as inputs and are not able to detect the fine scale patterns of permafrost thawing and thermal status. The increasing availability of high resolution remote-sensing derived products need an increasing variability of accurate datasets for their validation. It is important to highlight that every remote-sensing derived product is the result of a complex modelling chain of the signal, which often requires strong assumptions on the physics of the soil surface that require validation data.

    Moreover, it is important to underline that remote sensing approaches measure LST and not GST. Deriving GST from LST requires a physical or statistical modelling approach (Endrizzi et al., 2014). Particularly for the process-based numerical models, a dense network of GST can be used as input to better parameterize and calibrate the model. These will improve the upper boundary conditions of the ground profile and will help to better represent the fluxes of energy exchange between the dynamic interaction of land and atmosphere. The GST is the key parameter controlling all the bio-physical processes at the land-atmosphere boundary due to its central position in the Earth Critical Zone.

    To better elaborate the necessity of a dense observational network of GST with emphasis on its usefulness for modelling approaches, the following changes have been made:

    L.455-457: "Particularly for the process-based numerical models, a dense network of GST can be used as input to better parameterize and calibrate the model. These will help to better represent the fluxes of energy exchange between the dynamic interaction of land and atmosphere due to the central position of GST in the Earth Critical Zone." has been added.

L.457-458: "Furthermore, is helpful for understanding the effect…" has been replaced with "Furthermore, a dense observational network helps to understand the effect…"

Technical corrections:

1. Page 2, line 46. The term "permafrost areal extents" is also a component of "model accuracies", so there is no need to repeat it.

   L.47-48: "…but with significant differences in permafrost areal extents and in model accuracies…" has been replaced with "…but with significant across-model differences in model accuracies …"

2. In Figure 1, there is an issue with the legend labels. "locale" should be "local". Additionally, please confirm whether "Qingshui'he" should be "Qingshuihe".

   Thank you for notifying that. "locale" has been replaced with "local".

   "Qingshui'he" is the correct term and has also been used more often in previous publications.

3. Page 9, line 194. Are the mentioned four failed sensors included among the previously mentioned 11 sensors, or are they an additional set of four sensors?

   They are included. Thank you for pointing out this unclarity. The following changes have been made:

   L.205-207: "Four sensors had become malfunctioned and have not recorded any measurements, while three sensors stopped recording the measurements after seven months." has been replaced with "Among the 11 malfunctioned sensors, four sensors had become malfunctioned without any recorded measurements, while three sensors stopped recording the measurements seven months after installations. ."

4. Page 10. In the title of Figure 3, there is no need to repeatedly provide the full term of "GST".

   The full term of "GST" has been removed from the title of Figure 3, as well as from Figures 4, 5, 6, and 7.

5. Page 12. I suggest adding a legend to Figure 5.

   The legend has been added to Figure 5. In the legend in Figures 3, 4, and 6, the "swamp meadow / bare ground" has been replaced with "vegetation / bare ground".

6. Page 17, line 318. "… TDD of 320 m and 180 °C day", remove "m".

   L.336: "… TDD of 320 m and 180 °C·day, respectively …" has been replaced with "…TDD of 320 and 180 °C·day, respectively …"

7. Page 18, line 328. When the authors mention "… most of the sites", I suggest giving the exact percent.

   L.345-346: "… while for most of the sites, …" has been replaced with "…while for 88% of the sites, …"

References:

Aalto, J., Le Roux, P. C., and Luoto, M.: Vegetation mediates soil temperature and moisture in arctic-alpine environments, Arctic, Antarct. Alp. Res., 45, 429–439, https://doi.org/10.1657/1938-4246-45.4.429, 2013.

Endrizzi, S., Gruber, S., Dall'Amico, M., Rigon, R.: GEOtop 2.0: Simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. Geosci. Model Dev. 7, 2831–2857. https://doi.org/10.5194/gmd-7-2831-2014, 2014.

Jiang, H., Zheng, G., Yi, Y., Chen, D., Zhang, W., Yang, K., Miller, C.E.: Progress and challenges in studying regional permafrost in the Tibetan Plateau using satellite remote sensing and models. Front. Earth Sci. 8, 1–17. https://doi.org/10.3389/ feart.2020.560403, 2020.

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.