Main changes made to the revised manuscript:

- The abstract was modified to highlight some of the main outcomes of the statistical analysis of the ground temperature dataset
- The entire dataset and repository was restructured and all loggers were renamed following the suggestions of the three reviewers
- Additional information (e.g. linear regression equations) regarding the gap-filling of missing data in the time series are now provided in the information sheet included in the repository
- Additional figures are provided in Appendix B to better illustrate the performance of the gap-filling procedure
- The statistical analysis has been expanded to quantify the number of frost hours per month and to estimate the maximum frost penetration depth (the respective figures are included in Appendix C)
- A statement was included in Section 6 “Data availability” to emphasise that individual access to the meteorological data from the Bale Mountains can be granted on request as we are unfortunately not able to publish the meteorological data along with the ground temperature data

General response

We would like to thank the associate editor for obtaining three valuable reviews and the anonymous referees for their thorough and constructive comments on our manuscript. We are pleased that all three referees are convinced of the usefulness of the data set and generally support the publication of the manuscript if their concerns are properly addressed. On the following pages, we respond to the reviewers’ comments point by point. The reviewers’ comments are highlighted in grey and the responses in white. We hope that our responses qualify us to submit a revised version of the manuscript.

Response to Referee Comment 1 (RC1)

The manuscript of Groos et al. presents a dataset on ground temperature data: At 5 + 14 locations in the Ethiopian highlands, ground temperature was monitored at three depths (5) or near the surface (14 stations) for nearly three years. The manuscript contains a first analysis of the data with focus on the occurrence of frost, annual and seasonal temperature variations.

I am not an expert in the area of ground temperature dynamics. Thus, my comments are mostly restricted to the overall content of the paper, its structure and technical details.

That being said, the authors convincingly lay out the potential use of the data set. Moreover, the logistic and technical challenges of the respective measurements in this remote area underline its uniqueness.

We appreciate that the reviewer acknowledges the usefulness and uniqueness of the presented ground temperature dataset from the Ethiopian Highlands.

However, with effectively only three station providing true temperature profiles (albeit 3 depths only), and timespan of just 3 years and many data gaps (although reasonably-well filled), the data set is not excessively rich. Therefore, I strongly suggest to include the concomitant measurements at the meteorological stations into the dataset. The paper often relates to them (e.g. ll), and even includes their analysis (p. 12)., as they greatly enhance its value or seems even mandatory (as stated in the Discussion) for its analysis. I was even somewhat surprised not to find them. As for meteo-data of secondary provider that cannot be included, at least the respective reference would of great help to the potential user.

We would like to stress that 5 sites (and not only 3 sites as stated by the reviewer) provide “true” temperature profiles, although some longer gaps in the time series had indeed to be filled. In comparison with datasets from the lowlands or other continents, the presented data set may “not [be] excessively rich”, but in view of the logistical and technical challenges related to measurements in remote and high-mountain environments, we think the presented ground temperature dataset (comprising multi-annual time series from different depths and elevations between 3500 and 4400 m) from this data-scarce region is comprehensive enough to be of interest for the wider research community. Apart from Kilimanjaro, a similar dataset does not exist for any other high mountain on the entire continent.

We agree that the meteo data from the automatic weather stations are of benefit for the interpretation and analysis of the ground temperature data, but we cannot publish the meteo data along for multiple reasons: 1. The data from the weather stations are not automatically transferred from the study site to the database of the research unit and some of the data still need to be saved and postprocessed. 2. The meteo data available so far are currently analysed by PhD students within the research unit. Once the meteo data are processed and analysed (we don't know when this will be exactly the case), they will be made publicly available as well. However, before the final publication of the meteo dataset, personal access could be granted on request to researchers who are interested in the data. We added a respective statement regarding the data availability in the text (see Section 6 Data availability, last paragraph).
The core data of the publication are available from a Zenodo repository. However, they seem incomplete in some aspects, while redundant in others. I recommend some restructuring: - Naming the folders "raw data" (i.e. log files) and "processed data" would seem more intuitive to me. - Some files (e.g. Hourly_Ground_Temperatures_Corrected.csv) have a deviating formatting of the date column. Highly impractical for automatic use / scripts - Information_Sheet_Data_Correction.ods: For each logger, add a column "flag" with indicator(combinations) for "interpolated to full hour", "missing/filtered", "interpolated/gap filled based on logger nn" This potentially makes the other files (except, perhaps, Information_Sheet_Data_Interpolation) obsolete. - The duplication of *.txt and *.ods seems unnecessary. I suggest using txt for the actual data, ods for the meta-data - Please add a GIS file (or at least table with the station coordinates), the information of Table 1 plus further site attributes (vegetation, etc.) - Please add some reference to the ESSD-manuscript (e.g. with an overall readme.txt and/or in the Zenodo desciption field)
Thanks for the helpful advice. We have made the following changes according to your suggestions to restructure and modify the repository: - we renamed the mentioned folders to "raw_data" and "processed_data" - we removed all redundant or unnecessary files (e.g. " Hourly_Ground_Temperatures_Corrected.csv" and "Information_Sheet_Data_Correction.ods") - we combined all post-processed data into one file (i.e. table) and added a "flag" column to provide additional information on the treatment of each hourly measurement of each time series (see comment further below) - we included a readme-file - we included a GeoPackage (GIS file) with the requested attributes - we added a reference (DOI) to the ESSD-manuscript
The overview data analysis in the paper provide an adequate first insight into the data. As this is not

the core of a data publication, I recommend not extending them. Please consider shortening by excluding the analysis of the effect of slope and orientation. However, I made some suggestions how to make improve some points, if these are considered essential. Some conclusions drawn should probably be formulated more carefully, as they do not seem to be fully backed up by the analysis.
We decided to keep the analysis of the effect of slope and orientation in the discussion as this section is important to understand the spatial variations of ground temperature in the Bale Mountains. Moreover, this section is relatively short anyway. We considered the other recommendations to rephrase some of the conclusions more carefully.
The formal quality of the paper is high, with logical structure, adequate style and (mostly) helpful figures. As a technical suggestion I strongly urge to decrease the file size of the resulting PDF - 20 MB for a text document is inacceptable. I assume this comes from photos with excessively high resolution.
Thanks for the hint. We noticed the large file size too late. Figures and photos have been compressed (where possible bitmaps were converted to vector graphics) in the revised version to reduce the file size of the PDF to a few MB.
The dataset and paper merits publication, especially when supplemented with the meteo-data. However, I recommend moderate revision before. Further details in the annotated PDF.
We outlined before why we are not unable to publish the meteo-data along. Please find below our responses to the annotations in the PDF.
Page 1, Line 6: Why "climate" in this context? The current dataset concerns soil temperatures. Would "temperature observations" be more appropriate? Does "on the continent" mean "in Africa"? If so, I suggest to state this explicitly.
We replaced “climate observations” by “temperature observations” and “on the continent” by “on the African continent”
Page 1, Line 7: Please add monitored depths for temperature.
Monitored depths were added.
Figure 1: The upward-pointed triangle is commonly used to demark peaks. Please consider another symbol for the weather stations
Symbol for weather stations was changed (from triangle to square)
Page 4, Line 6: sorry to be nitpicking: Would't "cold phase" be more appropriate than "cooling", as those formations testify the former, not (necessarily) the latter?
Yes, “cold phase” would be more appropriate. Revised accordingly.
Page 5, Line 16: What is "pan" in this context?
“logger-pan-to-USB cable” was replaced by “logger-to-USB adapter»
Page 5, Line 30: Please add some kind of overview or table to illustrate the selection and choice of locations. Could also be added e.g. as some colour code to the labels of Fig. 3
Sorry, but we do not understand this comment as an overview of the distribution of the GT and TM data loggers is already provided in Fig. 1 and Table 1.
Page 5, Line 32: specify the temperature range where this RMSE was obtained

The comparative measurement was performed at 12~°C and at 4~°C in a fridge over seven hours with logger GT00 as reference. The temperature range for the accuracy of the TM data loggers is -10 °C to 65 °C as stated in Section 3.1 “Data loggers”. We added this information here as well.
Figure 2: Panels a) and b) do not seem to convey any necessary information apart from the general impression of the landscape. Consider removing or reducing to one. c) and d) seem redundant, too. It would be more informative how the readily installed site looked like, i.e. what (if any) disturbance was necessary to access the loggers for downloading the data. I suggest removing e) in favor for an enlarged version of f), showing only the logger and the measuring tape.
Panel (a) shows the characteristic relief and vegetation of the Sanetti Plateau and panel (b) shows the northern slope of Mount Wasama with its sparse vegetation cover. We think that such overview photos are essential for potential end users to get an impression of the afro-alpine environment in which the ground temperature loggers were installed. However, we restructured panels (c-f) following your suggestions. Panels (c-e) show a sequence of the logger installation process. Panel (f) was enlarged.
Figure 3: Please indicate which loggers continue measuring, e.g. by arrows at the end of the bars. In line graphs, try including vector graphics in favor of bitmaps.
We added arrows to indicate which loggers continue measuring. The bitmaps (Figures 3-5) have been replaced by vector graphics. Figures 6-7 were not replaced by vector graphics, but the resolution was reduced as well to reduce the size of the PDF.
Caption Table 1: I suggest to write "29" instead (as with all number above 12).
Revised accordingly.
Page 9, Line 10: I cannot see this period of interpolation in figure 3. I would expect the bar of GT07 having a grey part at the start. Please check.
We reserved the grey colour for “real” data gaps in the time series. The first part of the time series of GT05t (former GT07) was indeed adjusted to account for the wrong installation depth. Since this adjustment only slightly affected the temperature amplitude and not the long-term ground temperature variations, we do not think it’s necessary to emphasize this moderate modification in Fig. 3.
Page 9, Line 17: I suggest not to use the term "interpolation" here, as this is different from actual interpolation of "adjacent" values. Please consider using the introduced term "gap filling" consistently.
We replaced “interpolation” by “gap filling” throughout the manuscript
Page 9, Line 34: meteorolgoical => meteorological
Corrected
Page 10, Line 2: I don't understand the tenses used in this sentence. The data had not been available so far, but are available now, I expect. Please clarify.
We rephrased this sentence for clarification. The meteorological data from the Bale Mountains are shared within the research unit we are part of for internal usage, but they have yet neither been published nor made publicly available as some PhD students are still working on the data: “The meteorological data are stored in an on-demand processing database system - currently with

restricted access to the members of the joint Ethio-European Research Unit 2358 "The Mountain Exile Hypothesis" (Wöllauer et al., 2020). The data will be made publicly available at some point in the future (see Section 6)”
Page 10, Line 26: I suggest to prepare a more informative plot on the performance of the *gap filling procedure*: installation depth vs. R2 (or RMSE). Colour the resulting dots by the length of the time series reconstructed this way.
We included two additional plots in the appendix (B) of the revised manuscript to better illustrate the performance of the gap-filling procedure. One shows scatter plots of measured vs. modelled ground temperatures for each gap-filled time series (see Fig. B1) and the other shows R^2 and RMSE vs. distance between the predictor logger and gap-filled logger (see Fig. B2).
Page 10, Line 29: I suggest to locate the cross-comparison exercise to the beginning of this section, since consistency of TMs and GTs is a prerequisite for all subsequent steps. I expect this was performed on non-gapfilled data? Please confirm.
We shifted the concerned paragraph to the beginning of this section. Yes, the cross-comparison was performed on non-gapfilled data. We added this information here.
Page 10, Line 31: Please report RMSE and length of time series used in this comparison.
The RMSE was 1.7 °C and the time series spanned four month. Both information are now included.
Page 11, Line 4: For all depths?
This statement is based on measurements from a depth of 2 cm. We added this note.
Page 11, Line 7: please specify, which stations are these (Logger IDs)
We specified the stations (GT03, GT04, GT05) in the revised version.
Page 11, Line 10: specify the logger-IDs
We specified the logger location (GT02) in the revised version.
Page 11, Line 22: I agree that data analysis is not the focus of this paper. Still, providing a plot of T_ground vs. T_air and colouring the dots by season could help to illustrate the described phenomenon. In the current way, this all very plausible, but cannot really be seen by just looking at the data.
As ground temperatures are in principle controlled by meteorological parameters as well as ground properties (e.g. porosity) and conditions (e.g. moisture), a bivariate statistic (e.g. T_ground vs. T_air) does not provide much insight in the processes governing the ground temperature variations. We think a multivariate statistical analysis would be interesting/necessary, but that’s beyond the scope of this manuscript.
Page 11, Line 26: Clarify at which depths.
2 cm. Clarified in the revised version.
Page 11, Line 28: Replace "superficial" with "near-surface"
Replaced.
Page 11, Line 33: Would it be possible to give an estimate on the maximum frost penetration depth based on the gradient between 2 and 10 cm? If so, I suggest to state this value here.
We used the two complete time series at 2 and 10 cm at location GT05t (southern Sanetti Plateau)

to estimate the maximum frost penetration depth assuming a linear ground temperature gradient (see new Fig. C2). Based on this simple analysis, we can conclude that frost usually does not penetrate deeper than 5 cm into the ground. The mean frost penetration depth is about 3 cm.
Page 11, Line 35: Please use protected whitespace between number and unit to prevent unfortunate linebreaks.
Thanks, we tried to use protected whitespaces, but we forgot it here. It's now corrected.
Page 12, Line 3: tend => tends
Corrected.
Page 12, Line 4: replace "constant" with "(near) zero"
Done.
Page 12, Line 6: I suggest to treat the altitude gradients om meant T first, as they concern the primary data. After this, move to profile gradients and seasonal varaibility.
We restructured the paragraph accordingly.
Page 12, Line 12-14: hard to understand, consider rephrasing.
Rephrased
Page 12, Line 20: replace "long-term" with "annual". I cannot make any inference on diurnal amplitude from the given plots. If you want to make this point, consider plotting the (smoothed) daily temperature range of the stations over time.
We replaced "long-term" with "annual" and added a plot (Fig. 7d) to illustrate differences in the diurnal amplitude.
Page 12, Line 30: Please state the intended update interval of the repository.
We note in the revised version that we intend to update the repository on an annual basis.
Page 13, Line 5: Please clarify if you mean "denser compared to solitary measurements" (which I suppose) or "denser than the network presented here".
We clarified in the revised version that we mean a "network consisting of numerous loggers"
Page 13, Line 9: replace "final" with "post-processed
Modified accordingly.
Page 13, Line 12-13: "sporadic permafrost" seems to be a contradiction in terms. Rephrase.
"Sporadic permafrost" is an established term in the scientific community for landscapes where permafrost represents 0/10 to 50 % of the total area. So it has a spatial and not a temporal meaning. https://ipa.arcticportal.org/publications/occasional-publications/what-is-permafrost
Page 13, Line 14: replace "order" with "range"
Modified accordingly.
Page 13, Line 20: Neither *surface* measurements nor an analyses of diurnal temperature amplitude have been shown.
"at the ground surface" was replaced by "ground temperature amplitude near the surface" We added two exemplary plots of the diurnal temperature amplitude in the revised manuscript and appendix (Fig. 7c and Fig. B3).

Page 13, Line 24-27: I don't see how your analyses "reveals" this: While this assumption is very plausible, without measurements of net radiation and soil moisture this remains a hypothesis. Please correct.
We weakened our statement and replaced “reveals” by “suggests” in the revised version.
Page 13, Line 27-28: not shown
We added a respective figure (Fig. 7c). See previous comment before.
Page 13, Line 31-33: References to respective studies or examples would strengthen this claim.
References to respective studies were added
Page 13, Line 35-36: I am very skeptical how even a very dense network of temperature stations alone should be able to provide high resolution maps of precipitation and (air?) humidity. Please weaken this statement.
This was a misunderstanding. We didn’t intend to claim that maps of precipitation or air humidity can be inferred from temperature data (alone). We just wanted to provide a selection of potential methods that can be used to create maps from numerous point data. We modified the sentence as follows: “Geo-statistical and machine learning techniques have been applied in other studies to create high-resolution maps of temperature, precipitation, and humidity for Mount Kilimanjaro on the basis of meteorological data from different sites”
Page 14, Line 3-5: As these will be highly site-specific, I find this claim quite optimistic. I'd rather envision regionalizing soil temperatures (or profiles) with the help of remote sensing. Please comment.
We did not claim that the statistical relationship between air and ground temperature at one site (e.g. the Bale Mountains) is transferable to other sites (in the Ethiopian Highlands). We just wanted to outline the following: “For mountain regions where both air and ground temperature are measured simultaneously, their statistical relationship can principally be used to generate air temperature maps from remotely-sensed land surface temperatures”
Page 14, Line 8: remove quotes.
Quotes were removed.
Page 14, Line 29: this seems to be "gap filled" rather than "interpolated"? File is possibly obsolete (see extra remarks)
“interpolated” was replaced by “gap-filled” (throughout the manuscript) We merged the files “Hourly_Ground_Temperatures_Corrected" and “Information_Sheet_Data_Correction” to one file (“Hourly_Ground_Temperatures”) using a flag to provide additional information regarding each measurement (i.e. if it was interpolated to the full hour, gap-filled, corrected, removed, etc.). See Section 6 “Data availability)
Page 15, Line 2-3: "...with the data download date encoded in the filename". Please specify delimiter character.
We added the suggested information.
replace "superficial" with "near surface"
The sentence was rephrased (see comment below).
Having measured at a few dozen sites, I find it a bit strong to deny the existence of permafrost

"anywhere" in the EH. Please consider weakening your statement or strengthening this claim by better presenting that you monitored all/the most frost-prone sites existing.
We weakened the statement as follows: “Moreover, the data confirm the frequent occurrence of nocturnal ground frost in the afro-alpine belt to a depth of about 5 cm (Fig. C2). However, the mean annual ground temperatures of more than 7 °C on the highest peaks suggest that permafrost in the southern Ethiopian Highlands is either absent or restricted to isolated patches.”
strong claim. The performance of regional climate models will hardly be evaluated on ground temperature data.
The sentenced was rephrased as follows: “The dataset may serve a wide range of scientific applications, ranging from the validation of remote sensing products and the modelling of spatial ground temperature variations to the investigation of certain natural processes such as the formation of periglacial landforms or the geographic distribution of species that live underground such as the giant root rat.”
Figure 5e: Replace line plot by bar plot, as daily prcp is not a continuous variable
The line plot was replaced by a bar plot.
Figure 6, headline first column: replace "annual" with "whole year"
“annual” was replaced by “full year”
Figure 6, Caption, Line 2: add "dry" and "rainy", respectively.
“dry” and “rainy” was added in the caption
Figure 7, Caption: replace "temperature variations" with "smoothed temperature". Add dashed line to legend.
The caption was modified accordingly and the meaning of the colour-coded dashed lines was included in the figure caption as well.

General response

We would like to thank the associate editor for obtaining three valuable reviews and the anonymous referees for their thorough and constructive comments on our manuscript. We are pleased that all three referees are convinced of the usefulness of the data set and generally support the publication of the manuscript if their concerns are properly addressed. On the following pages, we respond to the reviewers’ comments point by point. The reviewers’ comments are highlighted in grey and the responses in white. We hope that the responses qualify us to submit a revised version of the manuscript.

Response to Referee 2 (RC2)

This paper outlines a unique dataset on ground temperatures recorded over a 3 year period on the high plateau of the Bale Mountains. The area is underrepresented in terms of past climate data records and the data is of clear use in understanding the current environment of tropical high mountains. It is also of interest as a dataset collected in remote circumstances and I think therefore that this paper and the data are of value and should be published. There are some concerns however. The data is somewhat messy, contains a lot of gaps, and the time period is not extensively long. The value would be significantly enhanced therefore if it was to be combined with attendant meteorological measurements (particularly air temperature) which have apparently also been measured in many of the same locations at similar hourly resolution.
We agree with all reviewers that the structure of the repository itself as well as the numbering of the locations and loggers could be improved (see comments below and response to Referee 1). Due the logistical and technical challenges related to the operation of a long-term (ground) temperature monitoring in remote high-mountain environments, data gaps in the time series can hardly be avoided. We therefore tried to fill the gaps in a reasonable way to provide a consistent and complete three-years dataset. Compared to ground temperature datasets from the lowlands or other continents, our time series may not be extensively long, but to our knowledge the presented dataset is the most comprehensive one from any African mountain above 3,500 m. Moreover, the timeseries will be continued over the next years at five sites (i.e. 15 loggers). We agree that the meteo data from the automatic weather stations are of benefit for the interpretation and analysis of the ground temperature data, but we are currently not able to publish them along as the dataset is still processed and analysed by other members of the research unit (see comment to Referee 1).
Since this is a data paper the communication of the dataset and its organisation are important. I am a little confused by the numbering of the locations/loggers and the rather unsystematic organisation of this aspect. I know several loggers were stolen and there are no observations as a result. I would recommend ignoring these and numbering the sensors with data starting with TM1 and GT1, and somehow creating a more logical order (maybe high elevation to low elevation sites in order?). Loggers that were lost don’t really need numbers in the final dataset. Figure 3 can then be ordered in the correct order (starting with TM1 and GT1)... You could even number the loggers at the same location but different depths GT1h (2 cm), GT1m (10 cm) and GT1l (50 cm) - i.e, high, mid, low

(or something similar, a, b and c) to make it more obvious which are in groups of three. A more intelligent numbering system would make the dataset easier to navigate and make it appear less a collection of disparate sub-experiments.

The loggers were originally numbered in the order they were installed. Since we did not know in advance which locations would be suitable and could be revisited on a regular basis, the numbering seems unsystematic. We did not want to change the numbering later on as this could lead to confusion in the long term. However, we have been convinced by the reviewers that a systematic numbering would help the potential end users. We therefore decided to renumber the loggers on a geographical basis – more or less from northwest to southeast. Furthermore, we now use one number per site and differentiate between different depths using the letters t(op), m(iddle), and b(ottom) as suggested. All logger acronyms were revised throughout the manuscript.

Figure 1 needs to show all the weather stations (two are off the bottom edge of the map) and I think some contours would help show the hypsometry more clearly – the current shading is somewhat confusing and concentrates on landforms rather than elevation bands... perhaps 500 m contour interval?

As we are not able to publish the meteo data along, there is no need to show the other two weather stations in the map. We reduced the transparency (i.e. shading) and added 500 m contour lines to show the hypsometry more clearly.

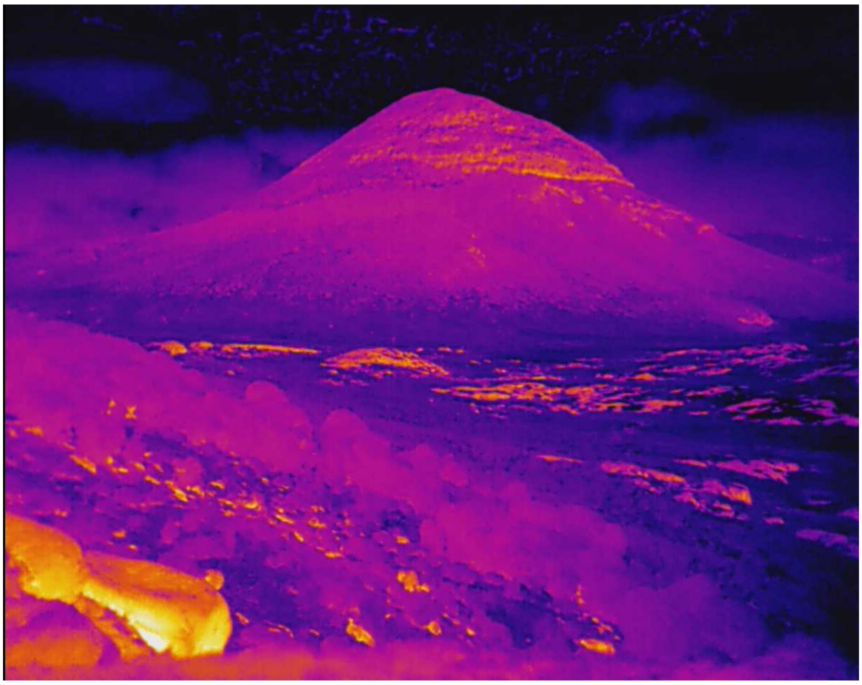
There are two sets of calibration data between the high and low cost loggers in the lab and in the field. Again it would be best if this was clearly accessible and perhaps made available in a sub-directory, since this is important information. Having said this I am somewhat concerned that the field calibration is of limited use, since one logger was installed slightly lower than the other (page 10: line 32). This seems like an error. The details of the lab calibration in the text are vague (page 5, line 31). It says “several hours” and does not say under what controlled temperatures for example. If there is too much information for the main text put it in a text file with the calibration data.

We added the following information regarding the comparative experiment: *“Due to the much lower accuracy of the TM data loggers compared to the GT data loggers, we performed a comparative measurement at 12~°C and at 4~°C in a fridge in the lab over six hours (see Fig. A1) with logger GT00 as reference (the logger was stolen in the field before the first readout and therefore does not appear in Fig. 3 and Table 1)”*
The six-hour comparative measurement did not show any offset between the low-cost TM data loggers and the reference GT data logger that was greater than the accuracy of ±0.5 °C stated by the manufacturer. We included a figure of the comparative measurement in the supplements (see Fig. A1). We did not integrate the raw-data from the comparative measurement in the repository as we think they do not provide any added value (on top of the figure).
We did not conduct any calibration in the field. We just compared the time series of logger TM06t (former GT08) and GT03t (former GT07), which were installed at the same location and almost the same depths (difference of max 1-2 cm) to detect larger deviations. However, as outlined in the manuscript, the comparison shows that both loggers measured ground temperature consistently and do not show any larger deviation.

I think the data itself might be clearer in 3 directories, a) raw data, b) corrected data and c) complete time series with gaps filled. The current structure is a little confusing. Put the readme files

concerning each stage in the relevant directory.
We split the repository in 2 directories: “raw_data” and “processed data”. Following the idea of reviewer 1, we combined the two datasets “corrected” and “complete” in one table and used a numeric flag to provide additional information regarding each measurement (i.e. if it was interpolated to the full hour, gap-filled, corrected, removed, etc.). Readme files were also included.
I also have some comments about the analyses and findings. Much of the data analysis concerns comparing ground temperatures with equivalent meteorological variables measured at Tuluka (3848 m). Since this AWS is not adjacent to most of the sensors a more logical choice would be Tullu Dimtu (4377 m) which apparently also has an AWS. Can you explain why this station is not chosen?. If it has similar data I would suggest using this site. Having said this, Figures 4 and 5 are very useful. Vertical lines separating each season (NDJF, MAM, JASO) would help the reader see the seasonal changes each year much more clearly. I am also wondering whether June is a short (but cloudy) break in the two wet seasons since it does appear that there is a short dry period around this time (any comments?). Maybe this is a fourth season?
That is true. Tullu Dimtu would be the logical choice as it is located in proximity of the ground temperature loggers, but unfortunately the meteorological time series of Tullu Dimtu contains longer data gaps. We therefore decided to use data from the Tuluka station, which has an almost-complete record. Vertical lines were added in Fig. 4 and 5 to highlight the different seasons. Yes, there is usually a little less precipitation during the “transition” of the consecutive rainy seasons, but it is not a typical dry season in the sense of the absence of convective clouds and rain.
The analysis of slope aspect is good (I would keep it) but it is not just the timing of the peak soil temperature that is changed due to aspect (page 12, line 14). The peak is much subdued on the north facing slope because the sky is cloudy during June when the sun is at its most northern point in the sky. The cloudless period coincides with when the sun is near its most southern trajectory. Thus, this explains the much higher readings recorded on the south face.
Thanks for the additional note. We included these information in the revised manuscript.
The lapse rate relationships in Figure 6 appear to be skewed by an outlier which is much warmer than expected given the elevation (I guess it is TM12 since its elevation is just below 3800 m), particularly in NDJF (when it is often sunny). I suspect therefore that this site is south facing (or has a distinct microclimate) and I would drop it from the lapse rate calculation.
Yes, it is logger TM02t (former TM12) as it is located near to a basalt cliff and probably heats stronger during the dry season (because of higher solar radiative input) than other sites at similar elevation. We removed this logger from the lapse rate calculation.
Some specific comments: Page 2, line 1. It is not always true that high mountains in the tropics receive more precipitation than adjacent lowlands (see Kilimanjaro for example) and precipitation often decreases on the highest summits. This statement is a bit misleading.
Yes, this statement is not necessarily valid for all regions in the tropics, so we have removed the half sentence.
Page 2: line 12: elevation-dependent warming (and places elsewhere)

thanks, corrected throughout
Page 2: line 16: define longer records (I think the original context was >20 years)
yes, we included “>20 years”
Page 5: line 31: TM04 was used as calibration but then no longer used in the field. Any reason? This whole section is a bit vague on detail.
Thanks for the hint. We forgot to remark that (former) GT04 was one of the loggers that were stolen in the field before the first readout. We clarified this in manuscript.
Table 1: It strikes me that the elevation range of 3493 m to 4377 m (marketed in the abstract) is rather optimistic since the lowest station was only recorded for a year, and without this station the range is only ~600 m.
Yes, the lowest logger recorded only for a limited period, but we could expand the time series at this site using our gap-filling approach to generate a complete three-years dataset that covers an elevation range from 3493 to 4377 m. Instead, we could state in the title only the elevation range of the GT data loggers (3877-4377 m) that continue measuring, but in this case we would ignore all the other data presented in the manuscript. Thus, we keep the title as it is.
Page 9: line 14 ff: It would be good to have the regression equations listed somewhere in the metadata files, rather than just r2 values and RMSE. This enables someone else to replicate your work.
We added a column with the regression equation for each gap-filled logger in the “Information_Sheet_Data_Gap-Filling.ods” file (see Section 6 “Data availability”). Moreover, we included additional plots in Appendix B to provide more information on the performance of the gap-filling procedure.
Page 10: lines 23-28. This paragraph seems out of place... it is about reliability of method and should come after everything else about the dataloggers – or in the method section.
We shifted the two relevant sentences to the methods section.
Page 11, line 6: where does the 2°C figure come from (data source?)
The data come from the AWS on top of Tullu Dimtu. The information was added.
Page 11, line 23: Also the solar angle is lowest in Dec/Jan – with a maximum elevation of only around 60° at the December solstice – yet it is overhead in Apr/Aug.
That’s all true, but we do not understand what the comment refers to.
Page 11, line 27: This is so much higher than the sites shown in Figure 4 (which are also at 2 cm) and must be a result of specific soil properties or the datalogger becoming exposed to radiation at the surface? Can you comment?
Fig. 4 shows daily mean ground temperatures and in the text we describe the hourly ground temperature maxima. This explains the large difference.
Page 11, line 30: cold air ponding is an interesting hypothesis but do you have any evidence? i.e. from air temperatures?
Unfortunately, no weather station is operated in the Wasama Valley or at a similar location. However, we have a thermal-infrared time-lapse video from one night that documents/illustrates the process quite nicely (the thermal-infrared photo on the next page serves as an example; dark blue to

black colours indicate low temperatures whereas purple to orange colours indicate higher temperatures). We will try to make the video available along with the publication of the manuscript.  Thermal infrared photo of the Wasama Valley and Mount Wasama in the background from 23 January 2020 at about 5:30 AM.
It strikes me that an analysis of frost incidence at each site would be really interesting. Perhaps a histogram showing the number of hours below freezing and its seasonal distribution at each site would be a useful graph. This is especially important, given the context given for the research which is about permafrost and peri-glacial landforms.
Thanks for the advise, a barplot showing the number of hours below freezing was indeed missing. We included such a plot exemplarily for logger GT05t in the Appendix (Fig. C1). We chose this site as ground temperature was measured here continuously over the three-years periods. It is evident that frost near the surface occurs predominantly during the dry season "Bega" from November to February.
Page 12, line 14: the sun is not in its zenith in Jan/Feb.... it is overhead in the southern hemisphere. It may have a high local angle of incidence on the south facing slope, but that is not the same thing.
Thanks, corrected.
Page 12, line 28: the five sites which are being continued need identification in Figure 1, Table 3.
As suggested by reviewer 1, we added arrows in Fig. 3 to highlight those sites where the measurements are being continued. The five sites are already marked in Fig. 1 as measurements at all GT-sites (white circles) will be continued. We added a note in the caption.
Page 13, line 22: the mean annual air temp is how much lower than soil temp?.... some figures would be useful here
We specified the difference in the manuscript: the mean annual difference between air and ground temperature is 5.6 °C at Tullu Dimtu.
Page 13, line 28: not just the timing, but the amplitude of the seasonal cycle

Added.
Page 14, line 14: elevation-dependent warming (as earlier)
Thanks, corrected throughout the manuscript.
The conclusions are a bit similar to the abstract. I note that ground frost is again mentioned here as a major finding, yet there is no analysis of this aspect (frost frequency).
Plots showing frost frequency and frost penetration depth have now been included exemplarily for the site GT05 in the Appendix C.
I hope that these suggestions will improve the organisation and communication of the findings.
The suggestions definitely helped to improve the manuscript. Thanks!

General response

We would like to thank the associate editor for obtaining three valuable reviews and the anonymous referees for their thorough and constructive comments on our manuscript. We are pleased that all three referees are convinced of the usefulness of the data set and generally support the publication of the manuscript if their concerns are properly addressed. On the following pages, we respond to the reviewers’ comments point by point. The reviewers’ comments are highlighted in grey and the responses in white. We hope that the responses qualify us to submit a revised version of the manuscript.

Response to Referee 3 (RC3)

The manuscript “ A multiannual ground temperature dataset covering sixteen high elevation sites (3493–4377 m a.s.l.) in the Bale Mountains, Ethiopia” by Groos et al presents a dataset from 29 ground temperature loggers that were installed at 16 sites and operate since 2017. The data was measured from 2 to 50cm depth, using single-channel miniloggers, UTL-3 Scientific Dataloggers providing high resolution and accuracy temperatures, and tempmate.®-B2 providing low accuracy temperarures. The measurements were obtained at hourly intervals. The data that is made available in the repository Zeondo is relevant for research and applications, mainly due to the scarcity of ground temperature observations from mountains and highlands in Africa. It has the potential to be used for soil climate characterization, geomorphic dynamics analysis, as well as to locally validate climate modeling or thermal remote sensing imagery. However, the dataset shows several relevant gaps that occurred due to logger malfunction. These gaps were filled by simple statistical techniques based on correlation. Both the original and the full infilled series are presented. The series were characterized using statistical measures, including some widely used indexes in environmental studies, but the results are not analysed in depth. The manuscript is, in general, well organized, clear and the language is precise. I must, however, note that English is not my native Language.
We thank the reviewer for his constructive feedback and his contribution to improve the manuscript.
Despite the overall clarity if the manuscript, I think that there are some significant issues that the authors must address to improve the manuscript and the applicability of the dataset before it may be accepted for publication.
We address all concerns point by point below.
The coding system used to identify the data is cumbersome. It is based solely on the logger IDs (given as numbers, which are in some cases, non-consecutive spatially). I strongly suggest using an ID system allowing to identify the location, depth of measure and the type of logger. This would facilitate data analysis.
Thanks for the comment. The loggers were originally numbered in the order they were installed. Since we did not know in advance which locations would be suitable and could be revisited on a regular basis, the numbering seems unsystematic. We did not want to change the numbering later on

as this could lead to confusion in the long term. However, we agree with all reviewers that a systematic numbering would help the potential end users. We therefore decided to renumber the loggers on a geographical basis – more or less from northwest to southeast. Furthermore, we now use one number per site and differentiate between different depths using the letters h(igh), m(iddle), and l(ow) as suggested by reviewer 2. The type of logger could already be identified before by the first two characters: GT = Geotest dataloggers; TM = Tempmate dataloggers.
The characterization of the soil at the different sites where the instruments were installed is highly needed and the manuscript needs this information before its acceptance. This issue is of most importance, both for the analysis of the datasets, but also because data from different sites are used to estimate missing temperature values.
Thanks for the hint. We included some general information on the ground/soil type in Section 2 "Study area": *"Most of the basaltic and trachytic rocks, especially on the plateau, are covered by a fine regolith layer that is only sparsely vegetated (see Fig. 2). In areas where soils have developed, Andosols are the most widespread soil type (Lemma et al., 2019)."* Moreover, we added the respective information in the Section 3.2 "Ground temperature monitoring" and Section 3.3. "Data post-processing".
The procedures used for data-filling must be described in more detail. In particular, all correlations, p-values and also scatterplots of soil temperatures between the sites (reference vs corrected) should be presented. The limitations of the procedures should be also described, especially in what concerns to the distance between sites and the potential impacts of the types of vegetation cover, soils and topography in the data. It seems that some depths were infilled with regression using temperature profiles from other sites. How were soil type and moisture differences considered for the procedure? These are factors that may strongly influence temperature change with depth and dynamics and they are especially relevant in environments with ground frost.
Some of the requested information (e.g. R^2 and RMSE) were already provided in the supplementary file "Information_Sheet_Data_Gap-Filling.ods" in the Zenodo repository (see Section 6 "Data availability"). We expanded this table and included also the distance to the predictor logger as well as the individual linear regression models (i.e. equations). Moreover, we created scatterplots (measured vs. modelled ground temperature) for all time series that were gap-filled (see new Fig. B1). Another figure (R^2 and RMSE vs. distance to predictor logger) was included in the Appendix (see new Fig. B2). It shows that the RMSE tends to slightly increase with increasing distance between the predictor and target logger while R^2 is also very high when a logger from a further location is chosen for the gap-filling procedure. Figure B3 was added to provide examples of gap-filled hourly ground temperature time series at different depths. As we used statistical (simple linear regression models) for the gap-filling procedure, local differences in the ground type and moisture are considered implicitly to a certain degree. However, we didn't measure ground moisture and, thus, are not able to characterise local differences. Nevertheless, the high R^2 of more than 0.8 at most sites (see Fig. B1) suggests that the local differences are of minor relevance (apart from maybe loggers TM07m and TM11t).
I also have some more specific comments: - In the title, the data period (start-end) should be indicated.
We refrain from indicating a static observation period in the title as some measurements start in

2017 and some in 2018. Moreover, the low-cost loggers were collected while the scientific dataloggers continue measuring.
- The abstract should be improved. For instance, in line 9, the authors mention that the series provide new insights, but what they say is too vague. They should concretize some of the results of the temperature analysis that they have done.
We highlighted some of the results in the abstract more specifically.
- In the last paragraph of page 3, It should be mentioned the type of climate according to Köppen classification, and it should be mentioned why the wet season is bimodal. It is confusing to mention that there are two rainy seasons and only one dry season. It could be mentioned that this is a transitional regime and that the two peaks in precipitation follow the movement of the sun and it can also be indicated the importance of the Intertropical Convergence.
We are not convinced that a reference to the climate classification of Köppen-Geiger makes sense here as it does not include an appropriate category for the tropical mountains (in Eastern Africa). However, we revised the section regarding the rainy seasons accordingly.
- In page 14, line 8, it doesn't make sense the mention of the endemic giant root rat. The reference should be contextualized.
We revised the sentence to provide a bit more context: *"In the Bale Mountains, distributed meteorological, ecological, and ground temperature data are of particular interest to better understand the relationship between spatial ground property variations and the scattered distribution of Erica trees across the Sanetti Plateau (Miehe and Miehe, 1994; Lemma et al., 2019; Mekonnen et al., 2019) as well as the scattered occurrence of endemic giant root rats (Tachyoryctes macrocephalus) that spend most of their time in large underground nests in the afro-alpine belt (Vlasatá et al., 2017)."*
- Figure 1 could be improved if the type of font/type/size of letters discriminate the information that they relate to.
We are not quite sure if that's what you meant, but we coloured the labels of the weather stations and data loggers to match the colour of the respective symbols in the map.
- in fig 3, as mentioned before the acronyms of the loggers should be modified in order to simplify the reading.
All logger acronyms are revised according to the reviewers' suggestions (see comment above and responses to reviewer 1 and 2)