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.

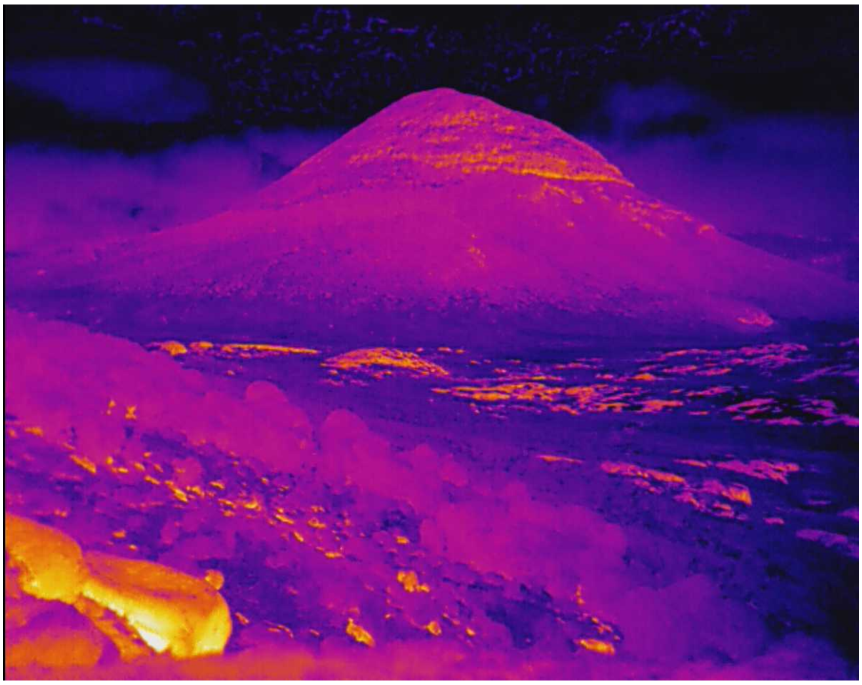
Yes, it is logger TM02t (former TM12) as it is located near to a basalt cliff and probably heats stronger during the dry season (because of higher solar radiative input) than other sites at similar elevation. We removed this logger from the lapse rate calculation.

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!