

Response to the reviewers' comments

Camilo Melo-Aguilar ^{1,2}, J. Fidel González-Rouco ^{1,2}, Elena García-Bustamante ³, Jorge Navarro-Montesinos ³, and Norman Steinert ^{1,2}

¹Universidad Complutense de Madrid, 28040 Madrid, Spain

²Instituto de Geociencias, Consejo Superior de Investigaciones Científicas-Universidad Complutense de Madrid, 28040 Madrid, Spain

³Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), 28040 Madrid, Spain

Correspondence: Camilo Melo-Aguilar (camelo@ucm.es)

1 Anonymous Referee 1

GENERAL COMMENTS:

GC01: *REVIEWER'S COMMENT:*

5 *The manuscript presents a rather detailed investigation of the coupling between air surface temperature and ground temperature in a series of climate simulations over the past millennium with an Earth System model. The analysis is focused on two main aspects: their co-variability and inter-annual time scales and their relative trends over the period of anthropogenic warming. The main conclusions are that in general terms both variables are strongly coupled, but with regional particularities. The main mechanism that modulate the coupling is snow cover. Considering the long-term trends over the past century, varying snow cover due to warming causes a change of the coupling over time that should be taken into account when using ground temperature from boreholes to reconstruct past near surface-air temperature. Deforestation over the past century also affects the coupling between surface temperature and ground temperature. In my view the manuscript is very well written, and I have just a few minor suggestions that the authors may want to consider.*

10

AUTHORS' RESPONSE:

15 The authors welcome the positive perspective of the reviewer on the paper. We are grateful for the reviewer's comments. Please find below our reply to your review in the discussion process of our manuscript. This reply is intended to give an overall view of the analysis we have done of your comments and the way we are going to include them in the revised manuscript. The comprehensive point-to-point response will be done in a separate document, including the original manuscript with track changes so the specific modifications can be found in the text.

20 GC02: *REVIEWER'S COMMENT:*

Sometimes, the text is a bit too wordy, with long paragraphs that pack a lot of information. This is admittedly a matter of style, but I guess that a typical reader would prefer some of the very long paragraphs broken up in smaller pieces, so that, after a first reading, the information can be located more rapidly if needed.

AUTHORS' RESPONSE:

5 A thorough revision of the text is going to be done in order to re-edit and to identify which of the long paragraphs can be broken up in smaller pieces or shortened, and the changes will be implemented in the revised manuscript. We hope that the reviewer finds the changes satisfactory and the quality of the manuscript improved.

GC03: REVIEWER'S COMMENT:

10 *The explanation of the negative correlation between air temperature and ground temperature in Tibet is intriguing, but I am not sure whether it is complete. The authors argue that the intermittent heat flux - modulated by seasonal snow cover - can cause a negative correlation between the time derivative of both temperatures, but does this translate in a negative correlation between the temperature anomalies themselves ? This may be true depending on the spectrum of the temperature variations. If the spectrum is white, i.e. little multidecadal variability, I can see that this explanation can be correct. However, if the temperature variations are mostly decadal, changes in the time derivative of temperature may*
15 *not be enough to change the sign of the temperature anomalies. All in all, I found this mechanism interesting, but I am not completely sure that it is the sole explanation. Unfortunately, I cannot suggest new ideas.*

AUTHORS' RESPONSE:

Our explanation of the negative correlation between air and ground surface temperature is based on annual time series. Indeed, it refers specifically to anti-correlation between the anomalies themselves. Longer air and ground temperature
20 variations, such as decadal or multi-decadal, are also anti-correlated, and intermittent heat flux modulated by seasonal snow cover is the dominant factor. A negative correlation between the time derivative of both temperatures is also present. We think that perhaps a more clear explanation of the time scales used to describe the main mechanisms impacting the SAT-GST coupling may be needed in the text. This will be included in the revised manuscript, trying not to increase the length of the text.

25 ***SPECIFIC COMMENTS:***

Other minor points:

SC01 : 'Improving our knowledge of the LM climate variability' LM has been defined in the abstract, but I think not in the
main text

30 Answer: LM is going to be defined in the main text of the revised manuscript

SC02 : The second, postulate is that variations in SAT propagate downward to the subsurface through conduction (Pollack et al., 1998; Smerdon et al., 2003, Delete comma after postulate

Answer: The comma will be deleted in the revised manuscript

SC03 : ‘In the present work the LM refers to the period from 850 to 2005 CE while the periods from 850 to 1850 and 1851 to 2005 CE refer to pre-industrial and industrial, respectively. 2 m air temperature is used for SAT and the first land model level.’ I my understanding, a sentence in English may not start with a number. Change to ‘two-meter temperature..’

Answer: The text will be changed as suggested in the revised manuscript

SC04 : CESM-LME supports the assumption that SAT is tightly coupled with GST at global scales and above multi-decadal scales. May be change ‘above’ to ‘longer than multidecadal temporal scales’

Answer: This will be changed as suggested in the revised manuscript

10 SC05 : ‘the thermal difference between SAT and GST. In order to explore the influence of such processes on the global SAT-GST relationship this analysis is extended considering winter and summer seasons (DJF and JJA hereafter) independently (Fig. 1c,e).- Specify boreal summer and boreal winter

Answer: This will be changed as suggested in the revised manuscript

15 SC06 : The correlation maps (Fig. 2b) provide a similar pattern with high and significant values in most the globe (>0.8 in regions located below 45° north) and lower correlations over NH mid- and high-latitudes; especially over the east of Siberia. ‘below’ is here ambiguous (or even incorrect): does it mean between 45N and 90S ? or between 45N and the equator. Maybe simply say ‘for northern latitudes smaller than 45 N

Answer: The reviewer is right in pointing out that this may be ambiguous. Actually, "below" here refers to land areas between 45N and 90S. This will be changed in the revised manuscript to "regions located between 45° N and 90° S.

2 Anonymous Referee 2

GENERAL COMMENTS:

5 GC01: REVIEWER'S COMMENT:

This is overall a well written and interesting manuscript. It is a logical next step, given the work that some of these authors have done to test SATGST coupling and the implications for borehole paleoclimatology in LM simulations. It goes into detail to explain the various coupling behavior that is observed and has some important conclusions. I therefore think that it should ultimately be published, but I have significant concerns about some of the analysis as outlined below.

10 *A major revision is therefore necessary before the manuscript should be accepted for publication.*

AUTHORS' RESPONSE:

We appreciate the positive perspective on the manuscript as well as the comments and suggestions. We are considering each comment of the reviewer and we are sure this will help improving the next version of the manuscript.

15 Please find below our reply to your review. This reply is intended to give an overall view of the analysis we have done of your comments and the way we are going to include them in the revised manuscript. The comprehensive point-to-point response will be done in a separate document, including the original manuscript with track changes so the specific modifications can be found in the text.

GC02: REVIEWER'S COMMENT:

20 *The biggest analysis issue that concerns me is the two-phase regression model. The authors do not give a reasonable justification for selecting this regression model and I think it has several significant impacts on their conclusions. In particular, the breakpoints influence their results in ways that can be hard to interpret. One way in which this is manifest is in their physical interpretations of the coupling influences. In Figure b, d, and f we are provided time series and regression results for SAT-GST and snow cover. The author's interpretation is that snow cover has influenced the coupling, but it appears that the breakpoint comes earlier in the SAT-GST time series than it does for snow cover. If we are to take*

25 *the breakpoints as physically meaningful (and I think that should be done very cautiously), why should we conclude that snow cover is driving the decoupling when the major break in its trends comes after the breakpoint in SAT-GST? The answer is that we probably should not take the breakpoints very seriously and it is a shortcoming of the analysis that the regressions are confined to the two-phase model.*

AUTHORS' RESPONSE:

30 We agree that improving the description of the two-phase regression and its use in this paper will be good for the text. Reviewer #3 has also shown some concern about this. We will include a more detailed motivation in Sect 3. We think that a better description of the physical interpretation as well as for the timing of the changes is necessary so it will be included in the revised document.

Figure 1b, d, and f illustrates that globally, snow cover is the main mechanism determining the SAT-GST offset during the LM. However, there is not an actual breakpoint in snow cover since it remains relatively stable during the LM. Thus, the breakpoint in SAT-GST cannot be solely explained by the influence of snow cover. We used the breakpoint in SAT-GST as a reference for depicting the linear regression of snow cover, for the same time periods, in order to show that snow cover does not vary at the time when SAT-GST does. Hence, some other mechanisms should be influencing the long-term variation in the SAT-GST offset. With this in mind, we explored different mechanisms in the subsequent Sections for explaining the breakpoint in SAT-GST

We think that the description of this argument needs to be improved. Therefore, we will revise how it is presented, and the required changes will be included in the revised document.

GC03: *REVIEWER'S COMMENT:*

This is also a problem in the comparison of trends in Figure 9. The upper panels indicate that the break points vary significantly spatially, with some locations yielding breakpoints in the 1000-1200 period and many in the 1700-1800 period. All of these trends of course end at the beginning of the 21st century, meaning that the second-period trends displayed in the middle and bottom panels of Figure 9 are taken over vastly different time periods. It is very difficult to determine how to compare such trends, many of which will be impacted by strong 20th-century changes that are either much stronger because of a later breakpoint or much weaker because of an earlier breakpoint. One might justify this approach if the breakpoints could more reliably be interpreted, but in many cases, indeed as evidenced across the ensemble, there is a lot of variability in the determined timing of the breakpoint (look at Figure 15, for instance, and try to convince yourself that the break points in the upper panel for ALL-F or OZ/AER shouldn't be several centuries in either direction of where the breakpoints were determined to be). I therefore have a very hard time interpreting the magnitude of the trends and what they actually mean. As such, I think the two-phase regression model is both poorly justified and probably a very confusing interpretation of the data.

AUTHORS' RESPONSE:

This is an interesting point raised by the reviewer. There is actually a large spatial variability of the breakpoints depicted in Fig. 9. It is important to bear in mind that only the breakpoints delivering statistical significance indicate an actual change in the long-term SAT-GST relationship. These points are mainly concentrated within the 19th century, with some exceptions that are indicated at earlier times as a response of land use land cover changes occurring before the beginning of the industrial period. This issue is discussed in Sect. 4.1 (page 13, line 24-30 of the submitted manuscript). According to this, the trends that should be considered refer consistently to the same time period.

We think that this argument may be better depicted in Fig. 9. For instance, by considering only the grid-points where significant variations in the long-term SAT-GST relationship are present. In that way, the distribution of those points around the 19th century would be visible and the comparison of such trends could be done directly. We will evaluate the

best way of presenting this information reordering both the related text and figures and the changes will be included in the revised manuscript.

GC04: *REVIEWER'S COMMENT:*

5 *Related to the above, the overall magnitude of the impact is hard to parse in the context of what the authors have done. They look specifically at the 1850-2005 period and note that the overall difference in SAT and GST trends looks to be about 0.05C. First of all, this should be couched in the context of the 20th-century trend estimate from the global borehole reconstruction of ~0.5C/century for the 20th century. If we take the model estimate at face value, the error introduced is therefore less 10%, given the authors' estimate is over a longer period of time (it should be reported in C/century*
10 *for better comparison). But the other analyses complicate this number and I am not sure how to interpret them. The trends estimated in Figure 9 for SAT-GST are +/- 0.1C per DECADE, which is a much larger regional impact than the number determined for the global difference (I understand these regional impacts would be muted in a global average, but the numbers are complicated by my subsequent point). Part of the confusion is because of an apples-to-aardvarks comparison, namely the global trend difference is from 1850-2005, while the trends in Figure 9 are for widely different*
15 *intervals determined by the breakpoint in the regression analysis. In some ways it would be much cleaner if the authors determined trends over set centuries to make sure the comparisons were consistent in time and to make the comparisons to the actual borehole reconstructions simpler. Such an analysis would have its own shortcomings, but as things stand the presentation is ambiguous on the most important point of the paper, namely if and how big a bias might be present in the borehole reconstructions based on the authors' analysis.*

20 *AUTHORS' RESPONSE:*

As discussed in the previous point (GC03), the bulk of the significant variations take place within the 19th century. We will consider changing Fig. 9 to better illustrate this fact. According to this, the trends in Fig. 9 that actually should be considered are consistent with the time interval of the global trend.

The use of a fixed time period, as for instance from 1850 to 2005, to determine the trends as the reviewer suggests is indeed an interesting approach that we also considered. On the other hand, allowing for the trends to optimally show when changes related to each forcing occurs, provides arguably more complete information and of interest for a broader community. We are not aware of any text, and there may be few, having shows the temperature changes in response to each forcing within decadal timescales and in the context of the last millennium. Results also express the spatial dependence of the response without imposing bounds. Nevertheless, we will reconsider both approaches and evaluate
25
30 the best way of presenting the results. We think that by improving this, the text will gain clarity

GC05: *REVIEWER'S COMMENT:*

Another issue related to the interpretation of the authors' results in the context of the borehole reconstructions is the actual spatial sampling of the borehole database and how it may actually be impacted by the biases the authors are reporting. This is particularly the case regarding snow cover and the other cryogenic effects that the authors report for

the high latitudes. While these impacts are reasonably described and there are clear differences between GST and SAT particularly in the high northern latitudes, there simply aren't that many boreholes used for the global reconstructions above about 55-60 degrees N latitude.

5 *While the biases are certainly a concern and relevant to any regional studies of high-latitude boreholes, it is not clear that the regionally very large impacts play much of a role in the existing NH and global borehole reconstructions. The authors could do a lot to quantify the effect by performing an analysis that restricts the assessment to locations where boreholes exist in the reconstruction database. I would advocate for that, but it would also add to what is already a long and detailed paper. It would suffice to simply mention this issue and *perhaps* provide at least some numbers on the*
10 *impact of the hemispheric or global estimates if, for instance, latitudes north of 60 degrees were excluded.*

AUTHORS' RESPONSE:

The reviewer raises a very interesting point that we have been considering. However, the analysis of the impacts that this kind of long-term SAT-GST decoupling processes may have on the actual spatial distribution of the borehole dataset goes beyond the scope of the present work. As the reviewer states, this manuscript is already long and detailed. Additionally,
15 its main purpose is to give an overall view of the different processes that may play a role in impacting the long-term SAT-GST coupling assumption. With this in mind, we are currently working on a paper that aims to explore in detail the impacts on borehole reconstructions in the context of the actual borehole distribution. Thus, the specific numbers will be addressed in subsequent work.

SPECIFIC COMMENTS:

20

The paper is generally well written, but there are language issues scattered throughout. I have noted some of those below, but the paper would be improved by a thorough evaluation of typos and language choices.

SC01 *Pg 3, ln 9:* has been supported by observations...

25 Answer: This will be corrected in the revised manuscript.

SC02 *Pg 4, ln 2:* PPEs experiments is redundant

Answer: We agree with the reviewer. Therefore the redundant expression will be deleted in the revised manuscript.

SC03 *Pg 6, ln 1:* It composes a total...

Answer: The text will be changed as suggested in the revised manuscript

30 SC04 *Pg 6, ln 12:* covariance during...

Answer: This will be corrected in the revised manuscript

SC05 *Pg 6, ln 16*: may help characterize a potential...

Answer: The text will be changed as suggested in the revised manuscript

SC06 *Pg 6, ln 28*: associations

5 Answer: The expression refers to a singular form therefore it will not be changed.

SC07 *Pg 7, ln 4*: I don't think that the OZ/AER (ozone/aerosol) abbreviation has been defined in the text.

Answer: It is defined in Table 2. which is mentioned in Section 2. However, we will check whether it needs to be defined in the main text. The changes will be included in the revised manuscript.

SC08 *Pg 7, ln 5*: allows identification

10 Answer: The text will be changed in the revised manuscript.

SC09 *Pg 7, ln 7*: through industrial times

Answer: This will be corrected in the revised manuscript.

SC10 *Pg 7, ln 3*: I don't think the subscript notation for the ensemble member or the layer depth has been explained anywhere. It should be defined for completeness.

15 Answer: The subscript notation for the ensemble member is explained in Table 3. and it is mentioned in Sect. 2. page 6, line 4 of the submitted document. The subscript notation for the layer depth is explained in Table 1. However, we think that this could be specifically mentioned in the main document, specifically in Section 3. Therefore, a brief explanation will be included in the revised manuscript

SC11 *Pg 7, ln 20*: While the correlations with GST and the lower depths are perhaps useful to explore, it should be explained that because of the phase shift correlations will diminish even if the system is purely conductive. All borehole inversions take into account the phase shift and this kind of time series comparison is a bit misleading. The authors should clearly point this out if they are going to leave this analysis in the manuscript. For what it is worth, some of the length and detail of the manuscript would be improved by removing these specific analyses because I do not think they add much.

25 Answer: Indeed the reviewer is right in pointing out the fact that the phase shift of the downward propagating signal diminishes the correlation between GST and temperature at lower depths even if the system is purely conductive. This may be the general argument by which depth impacts correlations by amplitude damping and phase shift. Actually, we stressed this issue in the first part of Sect. 4, (page 7, line 21) where the discussion about the correlation between SAT and ST at different depths is presented. Nevertheless, Fig. 6 shows that, in spite of the phase shift, correlations within the upper meter of the subsurface are large in general over the areas where the regime is purely conductive. Additionally, 30 Fig. 6 clearly shows that the covariance of the signal weakens and even disappears during the boreal summer over the northernmost part of the globe where the regime is not purely conductive. We consider that this result deserves being shown in the manuscript because it gives some relevant information about the influence of non-conductive processes on

the subsurface heat transfer and was discussed in page 11, line 15-35 of the submitted manuscript. In addition, Fig 6. gives a spatial view of the amplitude attenuation with depth of the temperature signal at annual time scales that may be of interest for some readers. Nevertheless, we agree that some clarification may be needed in the text as the reviewer points out. Therefore, this part will be revised and the modifications will be included in the revised manuscript.

5

SC12 *Pg 7, ln 26*: very cold air temperatures

Answer: The text will be changed in the revised manuscript.

SC13 *Pg 8, ln 30*: consider "In contrast" instead of "On the contrary"

Answer: It will be replaced as suggested in the revised manuscript.

10 SC14 *Pg 9, ln 28*: relationship at relatively

Answer: The text will be corrected in the revised manuscript.

SC15 *Pg 10, ln 3*: during JJA consistent with

Answer: It will be corrected in the revised manuscript.

SC16 *Pg 10, ln 8*: high rates

15 Answer: This will be corrected in the revised manuscript.

SC17 *Pg 10, ln 24*: such partitioning determines

Answer: It will be corrected in the revised manuscript.

SC18 *Pg 10, ln 27*: because the water, relative to the continental air, is warmer in winter.

Answer: The text will be corrected in the revised manuscript.

20 SC19 *Pg 10, ln 33*: The paragraph that ends here is an example of where things get a little too detailed. I think that this paragraph in particular could be scaled down a bit.

Answer: We agree. A thorough revision of the text will be done in order to identify also other paragraphs that can be shortened. The changes will be implemented in the revised manuscript.

SC20 *Pg 11, ln 14*: during these months

25 Answer: Thanks, the text will be corrected in the revised manuscript.

SC21 *Pg 12, lns 13-14*: It is relevant that other GCMs may suggest different levels of impacts, but this seems a little strange in the context of the real world. The real question is how well the GCMs represent the real-world impacts. It would be useful here to make a statement along those lines, in connection to what consistency or disagreements across models might imply for the real world.

Answer: It is indeed true that different GCMs could indicate different level of impacts in the SAT-GST relationship during the LM. This difference would depends on how different land surface models represent the physical processes at the surface as snow cover, and soil moisture among others. For the shorter timescales even internal variability will play a role.

In spite of the possible differences, the overall impact may be expected to be consistent among the available simulations of the LM with different GCMs, since the external forcings considered in the PMIP3/CMIP5 LM simulations are similar (Schmidt et al. , 2011) and should have a similar contribution. Nevertheless, not all model simulations consider exactly the same set of forcings. Some consider land use land cover changes, and aerosols, and other don't (Masson-Delmotte et al. , 2013). This will arguably have some more profound impacts as shown in our manuscript. Even if the role of external forcings should be dominant on a long-term context, Addressing this issues from a multi-model ensemble approach would help understanding the uncertainties associated to all the factors of variations describe above.

This work describes the most relevant physical mechanisms and feedbacks that may affect SAT-GST coupling in different parts of the world, many of which have not yet been measure systematically, and how these processes interact with long-term changes in external forcing factors. For this purpose we have used the largest and only ensemble of model simulations including several all-forcing and single forcing realizations. We would expect that, largest ensembles with a more systematic samplings of forcings and processes will be available in CMIP6/PMIP4 (Eyring et al. , 2017; Jungclaus et al. , 2017). This will allows for exercises exploring better sensitivity to different models.

We will consider a brief discussion on this issue and the required changes will be included in the revised manuscript.

20 SC22 *Pg 12, ln 16:* trends of both temperatures independently

Answer: The text will be corrected in the revised manuscript.

SC23 *Pg 13, ln 1:* Why not report pattern correlations in this discussion? A quantification of the spatial similarity would seem helpful.

Answer: The pattern correlation will be included as a measure of the spatial similarity

25 SC24 *Pg 15, ln 15:* the driving mechanisms

Answer: Thanks, the text will be corrected in the revised manuscript.

SC25 *Pg 15, ln 26:* The authors say nothing about simulated increases in LAI here. Instead of transitions to different plant functional types, it is also possible that increases in CO₂ during the 20th century drive LAI increases in the CESM that in turn impact evapotranspiration. This has been shown for the model projections and also may play a role in the historical portion of the runs that the authors are analyzing. See the following paper for some background: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018GL077051>

Answer: The reviewer is right in pointing out the role of increased LAI on increasing evapotranspiration rates. It is indeed very interesting that increases in LAI, as a response to enriched CO₂ environment, lead to increases in

evapotranspiration, and possibly to long-term variations in the SAT-GST relationship. Actually, the Community land model version 4 (CLM4), the land component included in the CEMS-LME, incorporates a Carbon-Nitrogen module. We will address this issue of the potential effects on LAI and report back to the reviewer in the final response.

5 SC26 *Pg 15, ln 32*: stable state

Answer: The text will be corrected in the revised manuscript.

3 Anonymous Referee 3

GENERAL COMMENTS:

5 GC01: REVIEWER'S COMMENT:

The paper examines the relationship between surface air temperature and ground surface temperature within a set of model simulations. The idea is to use the SAT-GST correlation as a metric to examine – or as a proxy, for the ground surface energy balance. This is done with the CESM-LME model ensemble that includes a set of experiments with all forcing as well as simulations with individual forcing. It is expected that comparison of the resulting simulations would yield information regarding the importance of each of these forcing on the SAT-GST relation over spatial scales and its temporal evolution. The paper briefly mentions that this may be important for borehole climatology based climate reconstructions.

This is a generally well-written papers, but it would benefit from a full language revision.

I have several minor clarifications that I believe need to be implemented for completeness before the paper is published.

15 AUTHORS' RESPONSE:

We appreciate the positive perspective on the manuscript as well as the comments and suggestions. We are considering each comment of the reviewer and we are sure this will help improving the next version of the manuscript.

Please find below our reply to your review. This reply is intended to give an overall view of the analysis we have done of your comments and the way we are going to include them in the revised manuscript. The comprehensive point-to-point response will be done in a separate document, including the original manuscript with track changes so the specific modifications can be found in the text.

GC02: REVIEWER'S COMMENT:

What do "regional" and "local scale" mean in the context of this model resolution?

AUTHORS' RESPONSE:

25 Considering the horizontal resolution of the CESM-LME $\sim 2^\circ$, in this work local scale refers to the scale of resolution a single or a few grid points while for the regional scale the range between 10^4 to 10^7 km² can be considered as defined in (Giorgi et al. , 2001). From this perspective it is really broad and refers in a general sense to sub-continental structures that can be as small as the province of Madrid and as large as the U.S. We will consider the use of this concepts and make sure that it is senseful in the revised text.

30 GC03: REVIEWER'S COMMENT:

How is the SAT defined within the CESM? Is it always the surface air temperature at a height of 2 m above the bare ground? Is it defined differently over a (say) forested area? If so, what could be the potential problems for the SAT-GST correlation?

AUTHORS' RESPONSE:

The variable we used for SAT is the temperature near the surface or 2-m air temperature identified by the model output "TREFHT".

5 TREFHT is calculated by interpolation between the surface temperature and the lowest atmosphere model level temperature. The surface temperature depends on the energy fluxes at the surface. For non-vegetated areas, the surface temperature is obtained from the energy balance at the ground surface. For vegetated surfaces, the energy fluxes are partitioned into vegetation and ground fluxes (Oleson et al. , 2010). Therefore, the surface temperature is the result of the balance between both the vegetation and the ground fluxes.

10 Since in both cases the energy fluxes from the different surface components are considered in the calculations of the surface temperature, and consequently on the 2-m air temperature, there should not be "potential problems" for the SAT-GST correlation.

15 From the surrogate reality of the GCMs simulated air temperature represents the best available alternative for exploring the air and ground temperature relationship. Previous works of this type have also used 2-m air temperature from different GCMs for investigating the covariance structure between the air and ground temperatures (e.g. González-Rouco et al. , 2009; García-García et al. , 2009).

GC04: *REVIEWER'S COMMENT:*

Could the author comment on the role played by evapotranspiration on the SAT-GST metric?

AUTHORS' RESPONSE:

20 Evapotranspiration plays an important role on heat transfer from the surface to the atmosphere. It is involved in the energy balance at the surface. However, evapotranspiration does not necessarily influence the SAT-GST metric by itself. The SAT-GST long-term relationship may be affected if the means by which heat is transferred from the surface to the atmosphere experiences a long-term variations. There are a couple of examples in the document describing how long-term variations in evapotranspiration, as a response to changes in the surface characteristics, lead to long-term changes
25 in the SAT-GST metric. For instance in Fig. 12 of the submitted manuscript it is presented how a significant increase in evapotranspiration impacts the SAT-GST relationship.

On the contrary, if the energy fluxes at the surface are relatively stable in time, the SAT-GST metric would remain also stable. We will consider including comments of this topic in the revised text.

GC05: *REVIEWER'S COMMENT:*

30 *What is the justification for using the two-phase regression model?*

AUTHORS' RESPONSE:

This comment is also raised by anonymous referee #2. A more detailed motivation will be included in the revised manuscript.

GC06: *REVIEWER'S COMMENT:*

5 *Why do authors use the “change points” obtained from the SAT-GST regression for the analogous analysis for snow cover? Shouldn't these snow cover “change points” be determined independently in order to see whether Is a relationship exists?*

AUTHORS' RESPONSE:

10 This is actually a point that also concerns anonymous referee #2. The change points from the SAT-GST regression were used as a reference for snow cover because there is not a significant change in the global snow cover. Thus, the use of the change points from the SAT-GST allows for exposing that besides snow cover there should be other mechanisms influencing the long-term variation in the SAT-GST offset. With this in mind, we explore different mechanisms that may be influencing such long-term variation, as is presented in the subsequent Sections of the manuscript.

As it is a common concern for the reviewers, we think that the description of this argument needs to be improved. Therefore, we will revise the how it is presented. The required changes will be included in the revised manuscript.

GC07: *REVIEWER'S COMMENT:*

15 *DJF and JJA refer to northern Hemisphere seasons, but are used in the global context. There is some discussion later in the text to acknowledge these restrictions, but they should be given up front or simply restrict the analysis to the northern hemisphere; likely there would be no difference in the results.*

AUTHORS' RESPONSE:

20 This will be revised to clarify it from the beginning of the document that DJF and JJA refer explicitly to the northern hemisphere seasons. The modifications will be included in the revised manuscript.

GC08: *REVIEWER'S COMMENT:*

Regarding the discussion about the Tibetan Plateau (Fig 4) , could the authors discuss whether the CESM resolution can account for the topographic variations and whether such high elevation variability could introduce unforeseen effects on SAT-GST?

25 AUTHORS' RESPONSE:

The topographic variations over complex mountainous regions are roughly represented in most state-of-the-art GCMs due to its relative coarse horizontal resolution. In the case of the CESM-LME, the elevation of each grid point is considered as the mean elevation for the whole area of the grid point. Thus, the orographic variations are smoothed out.

30 The limited characterization of the topography in most GCMs, in addition to the sub-grid-scale parameterizations, does not allow for an accurate representation of some hydrological processes over complex terrain (Huning & Margulis , 2018; Rhoades et al. , 2018). This situation may indeed lead to unforeseen effects on the SAT-GST relationship since the simulated snow cover fraction plays an important role on the effect described over the Tibetan Plateau in the CESM-LME.

Consequently, a brief mention on this issue deserves being included in the manuscript, since the specific effect on the SAT-GST relationship over complex mountainous regions could be model-dependent. We will consider this and the required changes will be included.

5 GC09: *REVIEWER'S COMMENT:*

In Figure 9, are all trends for the "post-change" interval? If so, why are the values given in K/decade? Most change points are post 1850; there should be at least a century post change. i.e. could you give the changes in K/century?

AUTHORS' RESPONSE:

The trends will be indicated in K/century in the revised manuscript as suggested.

10GC010: *REVIEWER'S COMMENT:*

I am not convinced that the results from this work necessarily imply that borehole climatology reconstructions would have to be revised to account for SAT-GST uncoupling. The effects are small and given the uncertainties inherent to the estimates of the quasi-steady state of the thermal regime of the ground, these additional errors are likely very small. Perhaps a simple theoretical experiment with +/- the LUC uncertainty imposed on a typical borehole temperature profile and its resulting inversion may illustrate the effect well, if any.

15

AUTHORS' RESPONSE:

This is an interesting point. Actually the effects that the results of this work could have on borehole reconstructions are being explored in a subsequent work. At global scales the effects resulting from this analysis are minor. At regional and local, they may not be so.

References

- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., Taylor, K. E.,: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 5,1937–1958, 10.5194/gmd-9-1937-2016, 2016
- 5
- García-García, A., Cuesta-Valero, F. J., Beltrami, H, Smerdon, J. E., Simulation of air and ground temperatures in PMIP3/CMIP5 last millennium simulations: implications for climate reconstructions from borehole temperature profiles, *Environmental Research Letters*, 11, 4, 044–022, 2016
- Giorgi F., and Coauthors: Regional Climate Information-Evaluation and Projections In J.T. Houghton et al (eds.), *Climate change 2001: The Scientific Basis*, Cambridge University Press, 583-638 , Cambridge, United Kingdom and New York, NY, USA.
- 10
- González-Rouco, J. F., Beltrami, H., Zorita, E., Stevens, B.,: Borehole climatology: a discussion based on contributions from climate modelling, *Clim. Past*, 5, 99-127, 2009
- Huning, Laurie S. and Margulis, Steven A., Investigating the Variability of High-Elevation Seasonal Orographic Snowfall Enhancement and Its Drivers across Sierra Nevada, California,*Journal of Hydrometeorology*,19(1),47-67,2018,<https://doi.org/10.1175/JHM-D-16-0254.1>
- 15
- Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J. F., Goosse, H., Hurtt, G. C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., LeGrande, A. N., Lorenz, S. J., Luterbacher, J., Man, W., Maycock, A. C., Meinshausen, M., Moberg, A., Muscheler, R., Nehrbass-Ahles, C., Otto-Bliesner, B. I., Phipps, S. J., Pongratz, J., Rozanov, E., Schmidt, G. A., Schmidt, H., Schmutz, W., Schurer, A., Shapiro, A. I., Sigl, M., Smerdon, J. E., Solanki, S. K., Timmreck, C., Toohey, M., Usoskin, I. G., Wagner, S., Wu, C.-J., Yeo, K. L., Zanchettin, D., Zhang, Q., Zorita, E.,: The PMIP4 *past1000* simulations, *Geoscientific Model Development*, 10-11, 4005–4033, 10.5194/gmd-10-4005-2017, 2017
- 20
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González-Rouco, J.F., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., Timmermann, A.,:Information from Paleoclimate Archives, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, chapter 5, 383-464, ISBN 978-1-107-66182-0, 10.1017/CBO9781107415324.013, 2013
- 25
- Oleson, K. W., and Coauthors:Technical Description of version 4.0 of the Community Land Model (CLM),2010
- Rhoades, A. M., Ullrich, P. A., Zarzycki, C. M., Johansen, H., Margulis, S. A. Morrison, H., Xu, Z., Collins, W. D.,: Sensitivity of Mountain Hydroclimate Simulations in Variable-Resolution CESM to Microphysics and Horizontal Resolution,*Journal of Advances in Modeling Earth Systems*,10,6,1357-1380,10.1029/2018MS001326,2018
- 30
- Schmidt, G. A., and Coauthors, 2012: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.1), *Geosci. Model Dev.*,5,185-191,2012.
- 440