

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 Cientificas-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)

The authors would like to thank the reviewers for their constructive suggestions and the time they devoted in reading and proof-reading the manuscript. We have tried to integrate all suggestions and think that the manuscript has improved with them. We do appreciate their contribution. The original manuscript with track changes is included so the specific modifications can be found in the text.

5 The next sections contain a detailed point by point response to the reviewers comments.

1 Anonymous Referee 1

GENERAL COMMENTS:

GC01: *REVIEWER'S COMMENT:*

10 *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 bot variables are strongly coupled, but with regional particularities. The main mechanism that modulate the coupling is snow cover. Considering the long-term*
15 *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.*

AUTHORS' RESPONSE:

20 The authors welcome the positive perspective of the reviewer on the paper. We are grateful for the reviewer's comments. Please find below the comprehensive point-to-point response to your review. The original manuscript with track changes is included so the specific modifications can be found in the text.

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.

5 *AUTHORS' RESPONSE:*

After a thorough revision of the text we have identified some paragraphs that have been broken up in smaller pieces or shortened. The main changes affect Sections 4 and 4.1. See pages 7, 8, 9, 11, 14 and 15 of the manuscript with track changes.

We hope that the reviewer finds the changes satisfactory and the quality of the manuscript improved.

10 GC03: *REVIEWER'S COMMENT:*

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 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.

15

AUTHORS' ANSWER:

20

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-term air and ground temperature 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.

25

We have included a brief mention in the text stressing the fact that the anti-correlation refers specifically to the annual temperatures. See page 9, line 30 of the manuscript with track changes. Additionally, we have included in this document a figure depicting the influence of simulated snow cover and sensible heat fluxes on the anti-phase SAT-GST behavior at multi-decadal to centennial timescales (Fig. 1). Note that the negative correlation between SAT and GST is also evident for the 31-yr filter outputs. Additionally, snow cover and sensible heat fluxes also present this anti-correlation at the multi-decadal timescales. Further, the anti-correlation is also notable at centennial timescales and trends at the end of the millennium. Also see response and changes related to reviewer #3 GC08.

30

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.

5 Answer: LM is now defined in the main text. Page 1, line 20 of the manuscript with track changes.

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 has been deleted. Page 2, line 22-23 of the manuscript with track changes.

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 has been changed accordingly. Page 6, line 8 of the manuscript with track changes.

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'

15 Answer: This has been changed as suggested. Page 7, line 15 of the manuscript with track changes.

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 has been changed as suggested. Page 8, line 1-2 of the manuscript with track changes.

20 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 45°N and 90°S ? or between 45°N and the equator. Maybe simply say 'for northern latitudes smaller than 45°N

25 Answer: The text has been changed to "(>0.8 in regions located between 45°N and 90°S). Page 9, line 8 of the manuscript with track changes.

2 Anonymous Referee 2

GENERAL COMMENTS:

GC01: REVIEWER'S COMMENT:

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

10 AUTHORS' RESPONSE:

We appreciate the positive perspective on the manuscript as well as the comments and suggestions. We have considered each comment of the reviewer and we are sure that the new version of the manuscript have been improved. Please find below the comprehensive point-to-point response. The original manuscript with track changes is included so the specific modifications can be found in the text

15 GC02: REVIEWER'S COMMENT:

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 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

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.

Changes in the text have been included to incorporate these improvements affecting the Methods and the Results section (3 and 4 respectively), modifying the test related to Figs. 1 and 9; Fig. 9 has also been changed (see description below and Fig. 2 of this document). See also page 6, line 33 and page 7, line 9-12 of the manuscript with track chances.

Indeed, the use of the two-phase regression has an impact on the text, to our opinion improving the understanding of

the results. The reason is that no condition is imposed on the time interval to evaluate trends and this allows for a better characterization of the SAT-GST responses to external forcings and feedbacks. This has been stated in the Methods (Sect. 3) part and also in the text related to Fig. 9 in Sect 4.1 (see page 14, line 4-10 of the manuscript with track changes). The modified text and Fig. 9 show better now that for many areas, changes in response to external forcing occur before 1850 CE, during the 17th and 18th centuries. We think that allowing the responses of SAT-GST to freely describe contributions in the timing and magnitude of the various single forcing factors to the ALL-F is interesting. We hope this is more clear now in the text.

The reviewer shows understandable concern for the physical interpretation of the timing of changes in snow cover in Fig. 1. We have made changes in the text describing this figure regarding the long-term trends at the end of the LM. We can recall that snow cover does play a role in multi-decadal/centennial in-phase (JJA) and anti-phase (annual and DJF) changes in pre-industrial times (see page 8, line 7-9 and 20-35 of the manuscript with track changes). Figure 1 shows two-phase trend results for SAT-GST and trends for snow cover time-locked to the changes in SAT-GST. The purpose of this is to show the same in-phase/anti-phase consistency described earlier for multi-decadal changes, acting here for long-term changes. We hope the text is more clear now regarding the interpretation of snow cover timing of changes and that these do not occur earlier or after those in SAT-GST. Further, in all examples provided elsewhere in Sect. 4.1 (e.g. Figs. 11, 12, 13, 14) a physical interpretation is provided that supports the timing of two-phase regression changes.

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

We hope the changes in Fig. 9 make the interpretation to this part more clear now. The dates of change are now shown as an example for one member in each ensemble (Fig. 9a) and also exclusively for the significant areas in at least 80% of all members (Fig. 9b). It becomes obvious, as the reviewer indicates, that changes spread over the millennium, but

the important ones only take place in the last centuries, although not confined to post-1850. This is interesting and better highlights the physics of the time response of SAT-GST than if the trends were locked to the post-1850 period. The timing of the regional changes and its amplitude depends on the timing in the forcing changes and the influence of internal variability in each run. The fact that changes are consistent in 80% of the experiment members (Fig. 9b,d of the revised manuscript) indicates results are robust regarding the temporal and spatial variability of the changes described.

GC04: *REVIEWER'S COMMENT:*

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 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 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.

AUTHORS' RESPONSE

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 and is adopted in Fig. 8. On the other hand, allowing for the trends to optimally show when changes related to each forcing occur (Fig. 9), provides arguably more complete information and of interest for a broader community. As mentioned above, it also shows that the dates and magnitude of changes are spatially and temporally consistent; also that changes pre-date 1850. We are not aware of any text, having shown the temperature changes in response to each forcing in the context of the last millennium, describing the spatial dependence of the response without imposing time bounds. Indeed, the text states that implications for borehole reconstructions are marginal at global scales. Some editing has been done in this direction in page 8, line 26-35 of the manuscript with track changes. This is also stated in the Conclusions in page 20 lines 25-35. However, the influences at subcontinental scales could be relevant. A more quantitative evaluation requires a more detailed implementation of borehole inversion methods (González-Rouco et al. , 2006, 2009). This will be done elsewhere.

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.

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 impact of the hemispheric or global estimates if, for instance, latitudes north of 60 degrees were excluded.

AUTHORS' RESPONSE:

We have been considering 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 database. However, this goes beyond the scope of the present work, as indicated in the response to GC04. We are currently working on a paper that explores in detail the impacts of decoupling in the context of the actual borehole distribution may have on borehole reconstructions at different spatial scales. Thus, the specific numbers will be addressed in the subsequent work.

SPECIFIC COMMENTS:

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...

Answer: This has been corrected. Page 3, line 9 of the manuscript with track changes.

SC02 Pg 4, ln 2: PPEs experiments is redundant

Answer: The text has been deleted. Page 4, line 2 of the manuscript with track changes.

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

Answer: The text has been changed as suggested. Page 6, line 1 of the manuscript with track changes.

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

Answer: This has been corrected. Page 6, line 12-13 of the manuscript with track changes.

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

Answer: The text has been changed as suggested. Page 6, line 16 of the manuscript with track changes

SC06 *Pg 6, ln 28*: associations

Answer: The expression refers to a singular form therefore it is not changed in the document

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

Answer: It is defined in Table 2, which is mentioned in Section 2 specifically in page 6, line 3 of the manuscript with track changes.

SC08 *Pg 7, ln 5*: allows identification

Answer: The text has been corrected accordingly. Page 7, line 5 of the manuscript with track changes.

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

Answer: This has been corrected. Page 7, line 8 of the manuscript with track changes.

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 and for the layer depth is explained in Table 3 and Table 1 respectively (Section 2). Additionally, the manuscript with track changes (page 6, line 9-10) includes now further comments on this.

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.

20 Answer: We agree with the reviewer and actually nor the text nor the figures contradict this. We have included a brief note stressing this issue. See page 7, line 23 and 26-27 of the manuscript with track changes. We think that the analysis of the correlation with GST and the lower ST_{L8} deserves being included since it gives some relevant information about the propagation of the temperature signal with depth. On the one hand, the time series allow for illustrating that the CESM-LME represents the characteristics of the conductive regime, as amplitude attenuation and phase shift of the temperature signal with depth. On the other hand, the spatial analysis presented in Fig. 6 permits illustrating the influence of non-conductive processes on the subsurface heat transfer over certain areas and the amplitude attenuation with depth of the temperature signal at annual time scales that may be of interest for some readers.

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

Answer: The text has been changed. Page 7, line 31 of the manuscript with track changes.

- SC13 *Pg 8, ln 30*: consider "In contrast" instead of "On the contrary"
- Answer: This has been replaced. Page 9, line 13 of the manuscript with track changes.
- SC14 *Pg 9, ln 28*: relationship at relatively
- Answer: The text has been changed. Page 10, line 14 of the manuscript with track changes.
- 5 SC15 *Pg 10, ln 3*: during JJA consistent with
- Answer: The text has been changed. Page 10, line 24 of the manuscript with track changes.
- SC16 *Pg 10, ln 8*: high rates
- Answer: This has been corrected. Page 10, line 33 of the manuscript with track changes.
- SC17 *Pg 10, ln 24*: such partitioning determines
- 10 Answer: It has been changed. Page 11, line 11 of the manuscript with track changes.
- SC18 *Pg 10, ln 27*: because the water, relative to the continental air, is warmer in winter.
- Answer: We have included some changes affecting this sentence. Page 11, line 14-15 of the manuscript with track changes.
- 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.
- 15 Answer: We have done a thorough revision of the text we have identified some paragraphs, including this specific one, that have been broken up in smaller pieces or shortened. The main changes affect Sections 4 and 4.1. See pages 7, 8, 9, 11, 14 and 15 of the manuscript with track changes.
- SC20 *Pg 11, ln 14*: during these months
- 20 Answer: The text has been corrected. Page 12, line 2-3 of the manuscript with track changes.
- 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.
- 25 Answer: Regarding how well CGMs represent the real world, this is a broad issue (Flato et al. , 2013). 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 measured 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 sampling of forcings and processes will be available in CMIP6/PMIP4 (Eyring et al. , 2017; Jungclaus et al. ,
- 30

2017). This will allow for exercises exploring better the sensitivity to using different models.

We have included a discussion on this issue in the conclusions part of the revised manuscript. See page 20 line 25-35 of the manuscript with track changes.

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

5 Answer: The text has been corrected. Page 13, line 6 of the manuscript with track changes

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 has been included. See page 13, line 26-27 of the manuscript with track changes.

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

10 Answer: This has been corrected. Page 16, line 18 of the manuscript with track changes.

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:
15 <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018GL077051>

Answer: 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 CESM, incorporates a Carbon-Nitrogen (CN) module (Oleson et al. , 2010). This module permits the allocation of the available carbon to new plant growth. In fact,
20 due to the increase in CO₂ concentration during the industrial period, there is an increase in gross primary productivity in the CESM-LME simulations that consider the GHG forcing. The CN module also calculates the CO₂ atmospheric concentrations originating in land use changes and other sources. However, it is radiatively inactive since the radiative forcing is prescribed (Lehner et al. , 2015).

We have identified that there are indeed some areas where increases in LAI during the industrial period lead to variations
25 in the long-term SAT-GST relationship in the CESM-LME. Nonetheless, the impact of this mechanism is relatively small compared to others described in the manuscript. We have included a brief mention of this issue. See page 18, line 17-19 of the manuscript with track changes.

SC26 *Pg 15, ln 32*: stable state

Answer: This has been corrected. Page 17, line 1 of the manuscript with track changes

3 Anonymous Referee 3

GENERAL COMMENTS:

GC01: REVIEWER'S COMMENT:

5 *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*
10 *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.

AUTHORS' RESPONSE:

15 We appreciate the positive perspective on the manuscript as well as the comments and suggestions. We have considered each comment of the reviewer and we are sure that the new version of the manuscript has been improved.

Please find below the comprehensive point-to-point response. The original manuscript with track changes is included so the specific modifications can be found in the text

GC02: REVIEWER'S COMMENT:

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

AUTHORS' RESPONSE

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 that range can refer in a general sense to sub-continental
25 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.

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?
30

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 may play 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 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.

25 On the contrary, if the energy fluxes at the surface are relatively stable in time, the SAT-GST metric would remain also stable.

GC05: *REVIEWER'S COMMENT:*

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

AUTHORS' RESPONSE

30 This comment is also raised by anonymous referee #2. Please for the answer to this comment refer to GC02 of reviewer#2

GC06: *REVIEWER'S COMMENT:*

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 a relationship exists?

AUTHORS' RESPONSE

Please for the answer to this comment refer to GC02 of reviewer#2

5 GC07: REVIEWER'S COMMENT:

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

10 We have included an explanation indicating that DJF and JJA refer specifically to the boreal winter and summer respectively. Page 8, line 1-2 of the manuscript with track changes.

GC08: REVIEWER'S COMMENT:

15 *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?*

AUTHORS' RESPONSE

Horizontal model resolution does not seem to be an issue since a higher resolution version of the model (CCSM4) produces a similar behavior and other models of similar resolution do not show it (García-García et al., 2009). This comment has been included in page 10, line 8-11

20 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 are indicated in C/century in the Manuscript with track changes.

25GC010: REVIEWER'S COMMENT:

30 *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.*

AUTHORS' RESPONSE

We are currently working on a paper that aims to explore in detail the impacts in the context of the actual borehole distribution may have on borehole reconstructions at different spatial scales. Thus, the specific numbers will be addressed in the subsequent work. See also response to GC04 of reviewer #2.

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 Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S.C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., Rummukainen, M.,: Evaluation of Climate Models, *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, 9, 741-866, ISBN 978-1-107-66182-0, 10.1017/CBO9781107415324.020, www.climatechange2013.org, 2013.
- 10 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.
- 15 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
- González-Rouco, J. F., Beltrami H., Zorita E., von Storch, H.,: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, *Geophys. Res. Lett*, 33, L01703,doi:10.1029/2005GL024693, 2006.
- 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.
- 20 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
- Jungclauss, 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.,
- 25 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 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 *past1000* simulations, *Geoscientific Model Development*, 10-11, 4005–4033, 10.5194/gmd-10-4005-2017, 2017
- 30 Lehner, F., Joos, F., Raible, C.C., Mignot, J., Born, A., Keller, K.M., Stocker, T.F.,: Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, *Earth System Dynamics*, 6, 2, 411–434, 10.5194/esd-6-411-2015, 2015
- 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
- 35 5, 383-464, ISBN 978-1-107-66182-0, 10.1017/CBO9781107415324.013, 2013
- Oleson, K. W., and Coauthors:Technical Description of version 4.0 of the Community Land Model (CLM),2010

- Pongratz, J., Reick, C., Raddatz, T., Claussen, M., A reconstruction of global agricultural areas and land cover for the last millennium, *Global Biogeochemical Cycles*, 22, 3, 1944-9224,10.1029/2007GB003153, 2008.
- , Ramankutty, F., and Foley, J.A.: Estimating historical changes in global land cover: croplands from 1700 to 1992, *Global Biogeochemical Cycles*, 13(4), 997-1027, 1992.
- 5 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
- 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.
- 10 Solow, A.R.: Testing for Climate Change: An Application of the Two-Phase Regression Model, *Journal of Climate and Applied Meteorology*, 26, 10, 1401-1405,10.1175/1520-0450(1987)026<1401:TFCCAA>2.0.CO;2,1987.
- Trenberth, K.E.: Some effects of finite sample size and persistence on meteorological statistics. Part I: autocorrelations, *Mon. Wea. Rev.*, 112, 2359-2368, 1984.

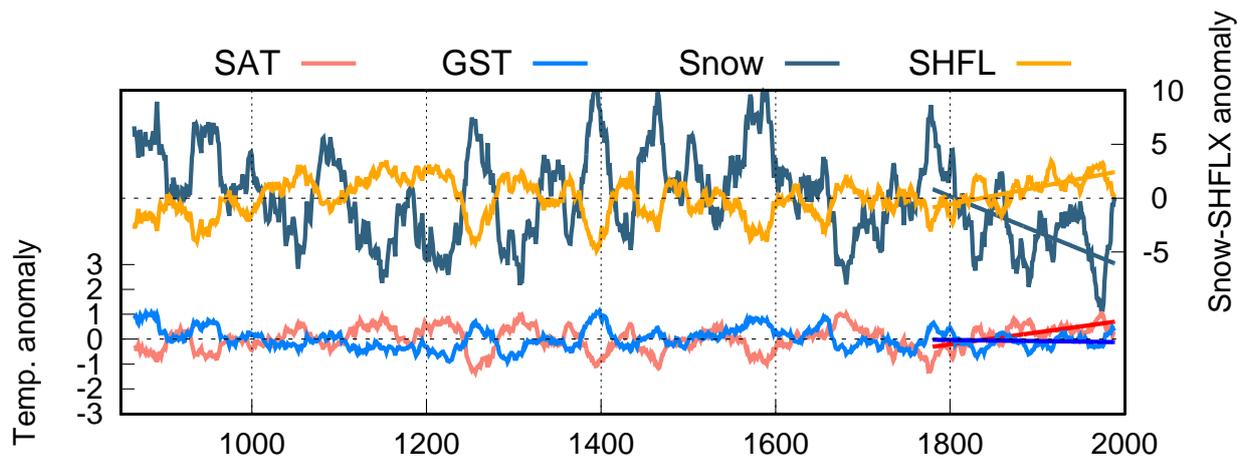


Figure 1. LM evolution of surface air temperature (SAT), ground surface temperature (GST), snow cover fraction (Snow) and sensible heat flux (SHFL) for a grid-point located over the Tibetan Plateau from the ALL-F₂ ensemble. All series are 31-yr moving average filter outputs. The linear trends for the last 200 years of the simulated period is also presented.

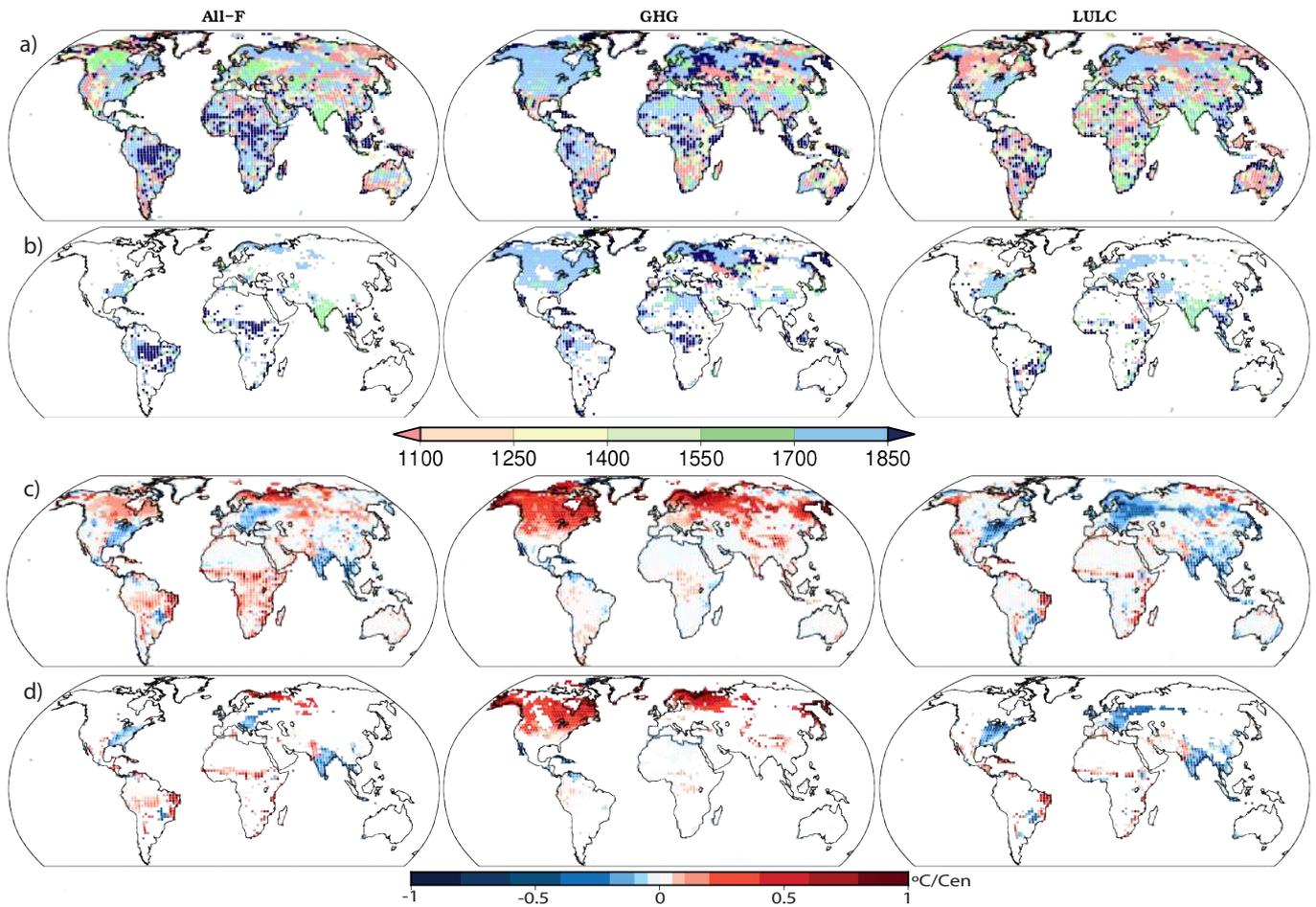


Figure 2. Spatial distribution of the two-phase regression results for SAT minus GST in the ALL-F₂, LULC₁, GHG₁ ensembles as examples: (a) dates (years) of change; (c) trends after the year-of-change. (b,d) are the same as (a,c) but only for areas where at least 80% of the ensemble members show significant changes. Trends before the year-of-change (not shown) are not significant in any of the ensembles. Significance ($p < 0.05$) is obtained based on a F-test (T-test) for year-of-change (trends) following Solow (1987). Significance also accounts for autocorrelation (Trenberth, 1984).

Influence of radiative forcing factors on ground-air temperature coupling during the last millennium: implications for borehole climatology

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)

Abstract. Past climate variations may be known from reconstruction methods that use proxy data as predictors. Among them, borehole reconstructions is a well-established technique to recover the long-term past surface air temperature (SAT) evolution. It is based on the assumption that SAT changes are strongly coupled to ground surface temperature (GST) changes and transferred to the subsurface by thermal conduction. We evaluate the SAT-GST coupling during the last millennium (LM) using simulations from the Community Earth System Model LM Ensemble (CESM-LME). The validity of such premise is explored by analyzing the structure of the SAT-GST covariance during the LM and also by investigating the evolution of the long-term SAT-GST relationship. The multiple and single-forcing simulations in the CESM-LME are used to analyze the SAT-GST relationship within different regions and spatial scales and derive the influence of the different forcing factors on producing feedback mechanisms that alter the energy balance at the surface. The results indicate that SAT-GST coupling is strong at global and above multi-decadal time scales in the CESM-LME, however, a relatively small variation in the long-term SAT-GST relationship is also represented. Although at global scale such variation does not impact significantly the SAT-GST coupling, at local to regional scales this relationship experiences considerable long-term changes mostly after the end of the 19th century. Land use land cover changes are the main driver for decoupling SAT and GST locally and regionally since they modify the land surface properties such as albedo, surface roughness and hydrology, and thus the energy fluxes at the surface. Snow cover feedbacks due to the influence of other external forcing are also important for corrupting the long-term SAT-GST coupling. Our findings suggest that such local and regional SAT-GST decoupling processes may represent a source of bias for SAT reconstructions from borehole measurement since the thermal signature imprinted in the subsurface over the affected regions is not fully representative of the long-term SAT variations.

1 Introduction

Improving our knowledge of the **LM** last millennium (LM) climate variability is key for a better understanding of the mechanisms that determine the Earth System response to natural (solar, volcanic and orbital) and anthropogenic (land use and

atmospheric composition changes) external forcings (Masson-Delmotte et al., 2013). Instrumental records represent the most adequate alternative to study past climate variations. However, they only provide coverage since the mid of the 19th century (e.g. Hansen et al., 2010; Jones et al., 2012). Therefore, to understand the nature of climate variability operating on longer temporal scales, LM reconstructions from a variety of proxy data (e.g. tree-rings, corals, preserved pollen, ice cores, etc.; Jones et al., 2009) and simulations using general climate models (GCMs) are generally employed.

The current available LM temperature reconstructions agree in depicting the general pattern of the temperature evolution going from a relatively warmer period at the beginning of the LM (MCA; Medieval Climate Anomaly) to a colder period from about 1450 to ca. 1850 (LIA; Little Ice Age), interrupted by the industrial warming in the 19th century (Masson-Delmotte et al., 2013). Despite this general agreement, there is still a large range of uncertainty that stems from different sources, including different reconstruction methods, various calibration and verification processes, spatial and temporal coverage, the different proxy locations (land, land and ocean, etc.) and the alternative statistical methods employed (e.g. Jones et al., 2009; Fernández-Donado et al., 2013). For instance, the range of uncertainty during the MCA in reconstructing Northern Hemisphere (NH) temperature is of about 0.6°C. Also, Ammann and Wahl (2007) estimate that the cooling during the Maunder Minimum (1645-1745) relative to present is about 0.7°C while Christiansen and Ljungqvist (2012) report a colder LIA (1.4°C) similarly to findings by Pollack and Smerdon (2004).

Addressing the range of uncertainties in reconstructing past temperature changes is relevant not only for assessing our understanding of past temperature changes and the confidence we have on available estimates (Masson-Delmotte et al., 2013), but also for model-data comparison exercises (e.g. Fernández-Donado et al., 2013) and for constraining the range of estimates of the system response to changes in the external forcing, i.e. the sensitivity of the climate system (Hegerl et al., 2006; Fernández-Donado et al., 2013)

Borehole temperature inversion is a well-established reconstruction technique and leans on two main assumptions: First, surface air temperatures (SAT) are expected to be closely coupled to ground surface temperatures (GST); ~~The second~~; The second postulate is that variations in SAT propagate downward the subsurface through conduction (Pollack et al., 1998; Smerdon et al., 2003, 2004). As a result, a thermal signature of the surface temperature is imprinted in the subsurface. This reconstruction technique is limited to recovering only the low-frequency information (decadal, and longer time-scales) since the soil acts as a low-pass filter and progressively filters out the higher-frequency variations with depth. Oscillations with periods on the order of days penetrate only ~50 cm deep (Smerdon and Stieglitz, 2006), seasonal cycles are solely observed close to the surface (<10 m depth), decadal variations propagate within the upper 50 m and multi-centennial changes within the LM are observed in the upper 500 m below the surface (Mareschal and Beltrami, 1992; Pollack and Huang, 2000).

As every type of reconstruction method, the borehole technique is also subject to diverse sources of uncertainties. One of them is that the assumption of a conductive heat transfer may not always be substantiated, since non-conductive processes such as advection and convection may influence or even dominate the subsurface thermal regime (Kukkonen and Clauser, 1994) in areas with important groundwater flows or geothermal activity. Nevertheless, a number of studies suggest that the impacts of such processes can be reduced with an appropriate treatment of the affected borehole temperature logs (e.g. Kohl, 1998; Bodri and Cermak, 2005; Ferguson et al., 2006)

Additionally, the most important contributions to uncertainty arise from ~~the impacts of~~ changes in surface processes that affect the SAT-GST coupling. Snow cover is especially important since it insulates the ground surface from the cold winter air causing large differences between the soil and air temperatures (e.g. Pollack and Huang, 2000; Stieglitz et al., 2003; Beltrami and Kellman, 2003; Bartlett et al., 2004). The impact of this effect on borehole theory has caused considerable discussion.

5 Mann and Schmidt (2003) argued that borehole-based reconstructions may be substantially biased by this seasonal influence of snow cover. Hence, under changing snow cover conditions GST may be not completely representative of SAT variations. Chapman et al. (2004) argued against based on the assumption that SAT and GST coupling is strong at longer than seasonal timescales, therefore snow biases would influence only the high-frequency oscillations. Additionally, the long-term SAT-GST coupling has been also ~~supported by both observations~~ supported by observations (Smerdon et al., 2003, 2004) and modeling

10 assessments (González-Rouco et al., 2003, 2006, 2009; Bartlett et al., 2005). Besides snow cover, other land surface and soil properties such as soil water content, vegetation and land use land cover (LULC) changes have also the potential of impacting the SAT-GST coupling. For instance, deforestation, afforestation and other land cover changes modify the land-surface properties including albedo, roughness, and evapotranspiration altering the energy transfer between the atmosphere and the ground (Anderson et al., 2011). In a recent study, MacDougall and Beltrami (2017) found that deforestation tends to warm the ground

15 surface mainly by reducing the transport of heat away from the surface. They found that such continuous vegetation changes would drive to long-term surface temperature anomalies and thus, deforestation should be considered as a possible source of bias for temperature reconstructions from subsurface temperatures.

One way of addressing uncertainties of paleoclimate reconstruction methods is using climate model simulations as a surrogate reality in which pseudoproxy records of varying complexity are created and reconstruction methods are replicated in pseudo

20 proxy experiments (PPEs) mimicking real-world cases (Smerdon, 2012). The robustness of using the borehole method for reconstructing past SAT variations has been tested in PPEs (e.g. González-Rouco et al., 2006, 2003, 2009; Beltrami et al., 2006). González-Rouco et al. (2003) investigated the relationship between simulated SAT and GST from interannual to centennial timescales in a forced climate simulation of the LM (1000-1900 common era; CE) with the ECHO-G GCM (Legutke and Voss, 1999). They found that in spite of the seasonal and longer-term variability of snow cover, the coupling was stable

25 for decadal and larger timescales, so that GST should be a good proxy for long-term SAT variations. González-Rouco et al. (2006) extended this analysis by implementing the borehole method in the simplified reality of the same model. They simulated underground temperature perturbation profiles using a heat-conduction forward model driven by simulated GST. Then, they applied an inversion approach to reconstruct ground surface temperature histories from the simulated profiles. Their results supported the overall performance of the borehole methodology.

30 Likewise, García-García et al. (2016) employed a similar approach as in González-Rouco et al. (2006) to retrieve past GST variations from GCMs applying an inversion method as well. They simulated global synthetic temperature profiles using a one-dimensional conductive model driven by simulated GST as an upper boundary condition from the CMIP5/PMIP3 (Coupled Model Intercomparison Project Phase 5 - Paleoclimate Modeling Intercomparison Project Phase 3; Taylor et al., 2012) LM simulations. Unlike the GCMs used in the early works that do not incorporate some external forcings, the PMIP3/CMIP5

35 simulations include a larger representation of LM forcings than pre-PMIP3 experiments (Fernández-Donado et al., 2013; Tay-

lor et al., 2012) including a variety of land-surface model components. Their results reinforced the reliability of recovering past global surface temperature variations from subsurface temperature measurements ~~in PPEs experiments~~ using current state-of-the-art GCMs.

The previous analyses support estimations of global/hemispheric past temperature obtained from borehole temperature inver-
5 sions. First, they support the overall performance of the methodology in retrieving past GST histories from borehole profiles. Second, the use of up-to-date forcing representations in CMIP5 model ensembles also ensures that long-term alterations of sur-
face properties like those induced by LULC changes, the effect of anthropogenic aerosols cooling and the potential long-term
snow cover feedbacks induced by both forcings, do not seem to bias inversion results at global/hemispheric scales. In spite of
these positive results, the analysis of last generation model experiments including the complete set of agreed CMIP5 forcings
10 (Schmidt et al., 2011, 2012) in the so-called all-forcing experiments does not allow for insights into the individual effect of
each of the forcing and the related feedbacks, both at global/hemispherical or at continental/regional scales. This would be
desirable in order to quantify the effect of the individual forcings on SAT-GST coupling and to obtain estimates of its partic-
ular temporal evolution that would allow for disentangling its contribution from that of other forcings. Therefore, the present
work considers all-forcings and single-forcing types of experiments in order to address how state-of-the-art climate models
15 simulate SAT-GST coupling from global to regional scales and to evaluate the potential influence of the external forcings on
the SAT-GST relationship. This will also provide information about where and when a decoupling of SAT-GST may exist,
with implications for borehole inversion practices at different spatiotemporal scales. For this purpose, we use the Community
Earth System Model-Last Millennium Ensemble (CESM-LME; Otto-Bliesner et al., 2016), which to date is the largest existing
ensemble of LM simulations with a single model and has not been used in previous assessments of this kind. The CESM-LME
20 includes all- and single-forcing experiments that consider jointly or separately the transient evolution of solar variability, vol-
canic activity, orbital changes, GHGs, anthropogenic aerosols and LULC (see details in Sect. 2)

Using all-forcing simulations (ALL-F hereafter) allows for evaluating the SAT-GST relationship throughout the LM in a realis-
tic representation of the real world conditions. Additionally, the single-forcing experiments are suitable to identify the specific
role that each forcing might play and its ‘fingerprint’ (Hegerl et al., 2011) on the presence of biases in the SAT-GST coupling.
25 Most external natural (solar and volcanic variability) and anthropogenic (GHGs, LULC and aerosols) forcings have the poten-
tial to indirectly affect SAT-GST thermodynamics through snow cover feedbacks (Hartmann et al., 2013). Moreover, LULC
changes can also have a direct influence on SAT-GST. During the LM the Earth's land cover has been substantially modified by
the replacement of natural ecosystems with agricultural land, especially since the industrial period (Pontgratz et al., 2008; Hurtt
et al., 2009). Such land use changes have the potential of altering the surface energy balance by the modification of energy
30 fluxes and moisture budgets (Betts, 2001; Betts et al., 2007; Brovkin et al., 2006; Anderson et al., 2011; Zhao and Jackson,
2014) driving to direct impacts on the atmosphere and ground-surface temperature relationship (MacDougall and Beltrami,
2017). Furthermore, vegetation changes may additionally lead to indirect non-linear effects that are relevant in the borehole
context. The most important one is the fact that deforestation at high latitudes leads to an increase in snow cover since low veg-
etation accumulates continuous snow cover more readily than forests in early winter favoring its permanence longer in spring
35 (Myhre et al., 2013). The local, regional and large-scale implications of these interactions on long-term climate variability and

specifically on the validity of the borehole method assumptions have not been explored so far.

The first part of the manuscript (Sect. 2) describes the CESM-LME simulations and forcings used in the experiment. Subsequently, Sect. 3 describes the methodological approach employed for the analysis of the coupling between SAT and GST during the LM. Section 4 presents the main results in the analysis of the global covariance **structure** of SAT, GST and soil temperature (ST) at different model depths throughout the LM. This analysis includes the global SAT-GST long-term coupling at annual and seasonal timescales. The latter helps identifying the impacts of seasonality on the SAT-GST coupling. In addition, the spatial distribution of the covariance structure, as well as the SAT-GST offset, are also illustrated. This allows for the detection of possible failures in the air/ground temperature coupling at regional and local scales. The spatial analysis is extended to investigate the heat transfer within the shallow subsurface by comparing GST relative to deep model layers. Section 4.1 specifically addresses the long-term trend of the SAT-GST relationship in order to evaluate whether this association experiences variation with time during the simulated LM. Finally, Sect. 5 provides a discussion about the implications of decoupling processes at different temporal and spatial scales for the borehole temperature reconstruction method.

2 Simulations and forcings

This analysis considers LM simulations produced with version 1.1 of CESM (CESM1; Hurrell et al., 2013). The Community Atmosphere Model version 5 (Neale et al., 2012) is used as the atmospheric component, the Parallel Ocean Program version 2 (Smith et al., 2010) represents the ocean component, including the Los Alamos sea ice model (Hunke et al., 2015). The horizontal resolution of the CESM-LME is $\sim 2^\circ$ over the atmosphere and land and $\sim 1^\circ$ in the ocean and sea ice (Otto-Bliesner et al., 2016).

The land surface component in the CESM1 is the Community Land Model Version 4 (CLM4; Lawrence et al., 2011) which incorporates some improvements relative to the previous version (CLM 3.5; Oleson et al., 2008) in the representation of land surface processes that are important for the energy transfer between the atmosphere and the soil (Oleson et al., 2010). Some of these include a better description of ground evaporation, thermal and hydrologic properties of organic soil, the ground depth, snow albedo, snow cover fraction and burial fraction of vegetation by snow (Lawrence et al., 2011). In addition, The CLM4 has the deepest bottom boundary condition placement (BBCP) among the current land surface models to date, placed at 42.1 m depth, discretized into fifteen layers (see Table 1) and including up to five additional layers in the overlying snowpack. This is relevant for borehole reconstruction applications because previous works have shown that shallow BBCs may not represent reliably the downward propagation of the temperature signal corrupting the amplitude attenuation and the phase shift (Smerdon and Stieglitz, 2006; Nicolsky et al., 2007; Alexeev et al., 2007), particularly for long-term trends as decadal from the comparison of energy storage in model simulations and borehole profiles (e.g. MacDougall et al., 2010). Therefore, this improved and deeper land surface model allows for analyzing the coupling between SAT and GST in a more realistic atmosphere-subsurface heat transfer scheme.

The CESM-LME incorporates ALL-F simulations with natural and anthropogenic forcings that were chosen following those

used in Landrum et al. (2012). ~~It is composed~~ It composes by a total of 30 simulations including a subset of 10 simulations that incorporate all the LM external forcings and smaller subsets of single-forcing simulations considering each forcing individually (see Table 2 for details and references therein). The simulations used in this study and how they trace back to the original experiment names produced by Otto-Bliesner et al. (2016) are shown in Table 3.

5

3 Methods

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~~ Two-meter air temperature is used for SAT and the first land model level (Table 1) represents GST. Additionally, the ST at different depths, denoted by the subscript notation: ST_{L2} , ST_{L3} , ..., ST_{L15} , is used for addressing different aspects of the subsurface heat transport within the CESM-LME. The simulated ST does not include data over the Antarctic region therefore this region is excluded from the analysis.

The relationship between SAT and GST is analyzed from two complementary perspectives. First, the ~~covariance structure during~~ covariance during the LM of the global SAT and GST anomalies relative to 1851-2005 CE is assessed at interannual scale. Likewise, the spatial patterns of the correlation between SAT and GST are also analyzed. This spatial analysis is extended to the differences between SAT and GST LM mean values as this gives an additional measure of the energy exchange across the air-ground interface (Bartlett et al., 2004) that ~~may inform about a potential~~ may help characterizing a potential air-ground temperature decoupling. In addition, the same spatial analysis is done for the differences between GST and ST_{L8} in order to get insights into processes affecting the pure conductive subsurface transport assumption within the shallow subsurface. For specific cases where SAT and GST exhibit some kind of decoupling, a description of the main processes that lead to this decoupling at local and regional scales is presented. For this purpose, the evolution of SAT, GST and ST at different depths is assessed during the last 105 years of the simulation period (1900-2005 CE) for the cases of interest. In some particular examples, additional variables (e.g. latent and sensible heat fluxes or snow cover) are also included for a better description of the physical processes. The purpose of such analyses is providing some examples rather than detailing the minutiae of the processes that lead to different SAT-GST responses at many different locations. For the analysis described above the ALL-F ensemble will be used, thus, providing from the point of view of the forcings considered a more complete and realistic representation of real world conditions than the single forcing runs. This allows for using the surrogate reality of the CEMS-LME model as a test bed for detecting potential sources of deviations in the SAT-GST relationship during the LM.

Second, we focus on the assessment of the long-term trend in the SAT-GST relationship. Non-stationarity in such association at long timescales (multi-decadal to multi-centennial) might imply that GST variations would not be representative of the SAT variations, thus rendering unreliable inferences of past climate change (Bartlett et al., 2004). A two-phase regression model (Solow, 1987) is applied to address this issue, that allows for detecting changes in the long-term trend of the SAT-GST relationship without imposing any temporal a priori condition on their occurrence. The year-of-change is identified, as well as the trend before and after the change. The timing of change and the magnitude of trends are suggestive of long-term changes

in the surface coupling. Initially, this assessment is performed at a global scale by analyzing the LM evolution of global SAT minus GST anomalies. The analysis of the long-term trend is then extended starting from the evaluation of SAT and GST linear trends independently during the industrial period since the largest changes are expected to occur within this period. In this part, besides the ALL-F ensemble, we have also considered the GHG-only, LULC-only and OZ/AER-only ensembles because they

5 have the highest potential of altering the land surface characteristics. This ~~allows for identifying~~ allows identification of the influence of each forcing on the LM SAT-GST coupling, including changes in snow cover, soil moisture, latent and sensible heat variations. The linear trend analysis of SAT and GST gives a general view of their evolution ~~through the industrial times~~ through industrial times and a first glance at the most important areas where long-term SAT-GST decoupling may exist in response to changes in external forcings and related feedbacks. Although these types of processes have taken place mostly during

10 industrial times, GHGs and land use have also experienced notable changes before 1850 CE (Ramankutty and Foley, 1992; Pongratz et al., 2008; Otto-Bliesner et al., 2016). The application of the two-phase regression to the complete LM, permits identifying changes that occur before 1850 CE.

4 Results

15 CESM-LME supports the assumption that SAT is tightly coupled with GST at global scales and ~~above~~ longer than multi-decadal scales. Figure 1a illustrates the stability of the coupling at global scales during the LM in the ALL-F₂ ensemble member as an example. Results are comparable for other ALL-F ensemble members. The LM evolution of global continental SAT, GST, ST_{L8} and ST_{L15} anomalies relative to the 1850-2005 mean are shown. For ST_{L15} non-filtered model output is represented and evidences the low pass filter influence of the heat conduction below the surface, whereas for SAT, GST and

20 ST_{L8} 31-yr running mean low pass filter outputs are shown. Subsurface temperature anomalies closely track SAT anomalies with relative small differences between them, thus indicating that air and soil temperature are coupled above multi-decadal timescales. The correlation coefficients for the 31-yr filtered series in Table 4 indicate high correlation ($p < 0.05$) for the soil layers close to the surface that diminishes slightly with depth, as expected, due to the phase shift of the signal. Table 4 also indicates that the correlation considering the high-frequency variations (yearly; left column) is only high at the levels close to

25 the surface, whereas at the deepest layer the correlation is low since the high-frequency variations are progressively filtered out and phase shifted as depth increases. Hence, the assumption of conductive heat transfer within the subsurface is realistically represented in the CESM-LME simulations.

Despite the strong coupling between air and subsurface temperatures at global scales, the existence of a relatively small offset between SAT and GST that grows backwards in time is evident for the annual (Fig. 1a) averages. This indicates a slight long-

30 term decoupling between SAT and GST. Previous works have argued that the nature of this offset arises from the changes in snow cover and its influence insulating the ground from ~~very low air temperatures~~ very cold air temperatures (e.g. Mann and Schmidt, 2003; Bartlett et al., 2004). ~~González-Rouco et al. (2009) suggested that beside snow cover, soil moisture has also the potential of accounting for 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~~ considering boreal winter and boreal summer seasons (DJF and JJA hereafter) independently (Fig. 1c,e). The offset between SAT and GST observed in the annual plot is apparent only in the DJF season. Indeed, the differences are larger during this season than in the annual data whereas the JJA evolution of SAT and GST anomalies is virtually identical. ~~In addition, the correlation coefficients (Table 4)~~
5 ~~suggest slightly lower correlation for DJF than for JJA, even if SAT and GST are highly correlated and significant for both seasons, especially for the filtered values. The largest offset occurring in DJF reinforces also the important role of snow cover in decoupling air and soil temperatures~~ In addition, even if SAT and GST are highly correlated and significant for both seasons, Table 4 suggest slightly lower correlation for DJF than for JJA. This feature and the largest offset occurring in DJF suggest an important role by snow cover. Figure 1b,d,f further illustrates the strong relation between the SAT-GST offset and snow cover
10 by displaying the 31-yr low pass filter outputs of the LM evolution of global snow cover and SAT-GST differences. Note that on DJF and, due to its influence on annual averages, the decrease (increase) in snow cover leads to a decrease (increase) in the SAT-GST offset. This is due to the insulating effect of snow that keeps GST close to zero while SAT can reach large negative values. Thus, an increase of snow cover leads to larger negative SAT-GST differences. For JJA on the contrary, an in-phase relationship is found at all time scales. Long-term trends change both in snow cover and in SAT-GST after the end of the 19th
15 century. During the boreal summer increases of snow also enhance SAT-GST differences due to the insulation of the ground from the warmer summer SAT, and conversely for snow cover decreases. This effect dominates the global average over that of the JJA austral winter during which SAT-GST and snow cover changes experience an anti-phase relationship as described above. Therefore, the NH influences anti-phase covariability of snow cover and SAT-GST during DJF (detrended correlations, $r=-0.52$; $p<0.05$) and annual ($r=-0.67$; $p<0.05$) and in-phase covariability during JJA ($r=0.62$; $p<0.05$).
20 ~~Dashed lines in Fig. 1b,d,f indicate the long-term trend in both SAT-GST offset and snow cover. Interestingly, the SAT-GST offset remains constant within the pre-industrial period but its long-term trend experienced a change during the industrial period. The behavior is similar for the annual and seasonal cases with some difference in the timing. Such variations in the long-term relationship should be taken into account since they play an important role in decoupling SAT and GST at time scales relevant for the borehole theory. Nonetheless, this relative small change in the long-term trend has limited impact for the~~
25 ~~global SAT-GST coupling (Sect. 4.1) .~~ Beyond the anti- and in-phase covariability at multi-decadal to centennial timescales, changes in the longer-term relationship between SAT and GST can play an important role in decoupling with implications for the borehole theory. At global scales, the long-term offset is relatively small, as shown in Fig. 1, and therefore with very limited implications. Nevertheless, it is interesting to asses the consistency and relative magnitude of changes in snow cover and SAT-GST. Two-phase regression in
30 annual and seasonal SAT-GST (Fig. 1b,d,f) shows consistent date of change during the 18th century. All changes are towards smaller SAT-GST differences and significant for annual and DJF ($p<0.05$). In snow cover trends, using the same dates of change as for SAT-GST, indicate anti-phase (in-phase) relationships for annual and DJF (JJA) as in the multi-decadal timescales described above. However, changes in snow cover trends during the last centuries are small and can hardly be invoked to solely account for the comparatively larger SAT-GST trend changes. This calls for considering other possible mechanisms and a more
35 spatial perspective.

The spatial variability of the relationship between SAT and GST gives further insights about the role of different processes on the SAT-GST coupling. Figure 2 shows the spatial distribution of the differences between mean LM SAT and GST as well as the correlation coefficients for annual, DJF and JJA averages in the ALL-F ensemble. ~~Maps have been obtained from the ALL-F₂ member; dots indicate that 80% of the ensemble members agree in delivering significant differences or correlation for a given grid-point.~~ In the case of the annual temperatures (Fig. 2a), GST is generally warmer than SAT being the differences low in most of the globe (less than 2 °C) except in the NH mid- and high-latitudes areas where the differences are higher (up to 15 °C). 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~~) (>0.8 in regions located between 45° N and 90° S) and lower correlations over NH mid- and high-latitudes; especially over the east of Siberia. A similar behavior is seen in the DJF season although more pronounced. During these months, GST is much warmer than SAT reaching differences up to 30 °C at the northernmost part of North America and Eurasia (Fig. 2c). Differences are smaller at mid- and low-latitudes. The influence of the ocean over coastal areas providing larger SAT relative to GST is noticeable. Likewise, the correlation is lower over the northern snow covered areas (Fig. 2d) while in the rest of the globe it remains high. ~~On the contrary,~~ In contrast, Fig. 2e,f shows that during JJA, when the snow cover is scarce, the SAT-GST coupling is strong globally with temperature differences lower than 2 °C, and high correlation coefficients (>0.9). Consequently, the role of the snow cover in decoupling SAT and GST is highlighted. Positive correlation values are low over borderline areas where snow cover is more variable, thus producing variability in the SAT-GST offset and thereby altering the covariance structure. Close to these areas in central Asia the high negative correlation within the Tibetan Plateau is noteworthy (see discussion below).

Figure 3 illustrates the SAT-GST decoupling due to the snow cover at the local scale for a particular grid-point as an example. The grid-point is located over a region with considerable snow cover during the boreal cold season (northeastern Russia). The DJF and JJA evolutions of SAT, GST, ST at different depths and the snow cover for the last 105 years from the ALL-F₂ simulation are shown. Note that in DJF the snow covers 100% of the grid-cell during almost the whole period. Thus the soil is insulated and the difference between SAT and GST is ~-15 °C on average. The temperature of the deeper layers is presented in order to illustrate the amplitude attenuation and phase shift with depth of the temperature signal. Note that during DJF, GST is only slightly below 0 °C and the agreement of its variations with those of SAT is only noticeable in the largest changes of both while ST_{L6} and deeper STs are above 0 °C. In JJA, SAT and GST are very similar and their low frequency variability propagates to deeper levels, all above 0 °C.

Some aspects of the SAT-GST spatial distribution deserve further attention. One of the most noteworthy is that over the Tibetan Plateau region, temperature differences between SAT and GST are as large as in the mid- and high-NH latitudes for both annual and DJF (Fig. 2a, c). However SAT and GST are negatively correlated for annual and DJF seasonal resolution data (Fig. 2b,d), being the only region of the globe where this occurs. For JJA, the correlation nonetheless is positive and high (Fig. 2e). The nature of this opposite phase arises from discontinuous snow cover over this region and the very low air temperatures during DJF. Usually, the snow cover insulates the soil from the colder air, avoiding the heat exchange with the atmosphere (as shown in Fig. 3). Nevertheless, discontinuous snow cover insulates the soil only partially, leading to this particular SAT-GST interaction. Figure 4 displays this behavior for a grid-point located over the Tibetan Plateau. SAT, GST,

ST at different depths, snow cover and surface sensible heat flux (SHFLX) are shown. In DJF during periods of low snow cover the fraction of surface exposed to the atmosphere allows for exchanging energy from the warmer soil to the colder air. Conversely, when snow cover is high, the large fraction of insulated soil reduces almost completely the heat transfer from the soil to the atmosphere. Therefore, with lower (higher) fractions of snow cover, higher (lower) heat transfer takes place with GST decreasing (increasing) and SAT increasing (decreasing). Indeed, there is a high negative correlation (-0.84 ; $p < 0.05$) between the snow cover fraction and SHFLX which is the main way for energy dissipation within this region since latent heat fluxes (LHFLX) are negligible. Higher/lower albedo due to variations in snow cover fraction also contribute to the negative SAT-GST correlation over this region (not shown). Horizontal model resolution does not seem to be an issue since a higher resolution version of the model (CCSM4; Landrum et al., 2012) produces a similar behavior and other model of similar resolution do not show (García-García et al., 2016). During JJA on the other hand, snow cover is negligible thus SAT-GST coupling is strong. Note that GST and all ST are above zero and GST is higher than SAT as it is warmed by radiative gain and transferring heat to the atmosphere, thus the positive correlation (0.5 ; $p < 0.05$) between SAT-GST and SHFLX.

The spatial SAT-GST differences during JJA (Fig. 2e) depict other relevant aspects that have an influence on the SAT-GST relation-at-relative relationship at relatively short timescales. SAT is generally colder than GST globally. However, for JJA, there are large areas inland, mainly located at the south-east of the United States, some parts of central and eastern Europe and eastern Asia, with warmer SAT relative to GST. Variations in LHFLX from DJF to JJA are driving this effect. Figure 5 shows that the areas where LHFLX increase in JJA relative to DJF are related to the same areas where SAT is higher than GST in JJA (Fig. 2e). Therefore, there is a direct relation between the increase of evapotranspiration in JJA and the ground temperature response at these locations. The time series in Fig. 5 display this behavior for a grid-point located over southeastern China.

During JJA the surplus of energy due to higher solar radiation reaching the surface is mostly dissipated as latent heat leading to a net heat loss at the ground surface. Note the anti-correlation between surface soil moisture and LHFLX (-0.46 ; $p < 0.05$) during JJA as well as the large SHFLX values that take place when soil moisture is lowest and evapotranspiration is limited. Therefore, the high rate of LHFLX contributes to cool the surface and GST tends to be lower relative to SAT. Soil water content also exhibits large changes during JJA in consistence with during JJA consistent with the large evapotranspiration (Fig. 5, top), thus providing the source of moisture that contributes to temperate SAT and cool GST. Large variations in evapotranspiration from DJF to JJA are also present at mid-latitudes of the SH summer continents (America, Africa and Australia), although only a very limited impact on Fig. 2e is perceived, especially over the western coast. Over these regions, high incoming energy impinging the surface during SH summer (not shown) supports high-rate high rates of latent heat fluxes. However, as soil water becomes a limiting factor, more energy is dissipated as sensible heat and the ground surface is warmed. Therefore, GST experiences higher temperature than SAT on average. Figure 5 also shows large evapotranspiration over the tropical rainforest in America and Africa, both in DJF and JJA, that does not translate to positive SAT-GST differences. Over the rainforest, the energy fluxes at the surface do not vary significantly from DJF to JJA since incoming radiation is relatively constant throughout the year, and transferred to evaporation and evapotranspiration within the canopy while soils are well watered by precipitation to support the large amounts of evapotranspiration. This situation leads to a small range of variation both in SAT and GST with very small differences between them and higher GST relative to SAT.

Figure 2a,c,e shows similar higher SAT relative to GST also at some coastal areas both in JJA and in DJF. During DJF the effect is mainly present in the NH mid-latitudes as, for instance, in most of the coasts of Europe, both the east and the west coasts of North America, as well as in Japan and the east coast of China. During JJA, on the other hand, this behavior is seen mostly in the SH mid-latitudes, as in southern South America, South Africa and the south of Australia. ~~Therefore, there is a clear relationship between the winter season for each of the hemispheres and this SAT-GST temperature response in the CESM-LME simulations.~~ Interestingly, higher SAT relative to GST is also evident at some coasts over tropical regions both in JJA and in DJF mainly over southeastern Asia, the south of the Indian subcontinent, the Gulf of Guinea and some areas of South and Central America. In the CESM-LME the atmospheric grid-box of the coastal areas is partitioned into land and ocean fractions. For the areas with sea ice formation, an additional sea ice fraction is considered (Neale et al., 2012). This configuration of the coastal grid-points leads to a partition of the energy fluxes at the surface into those of the land fraction and those of the ocean part. During the cold season, ~~such partition is determining~~ such partition determines the higher SAT warming relative to GST since the relatively low net radiation that impinges the surface at mid- and high-latitudes, limits the ground surface heating as well as the energy fluxes out of the land surface fraction. In contrast, the energy fluxes from the ocean surface to the air above are large, mainly as a result of the temperature difference between the water and the comparatively colder air above. ~~since the water is relatively warmer than the continental air in winter. Strong winds blowing predominantly onshore over most of the coast also contribute with advection of latent heat from the relatively warmer ocean.~~ The dissipation of energy from the ocean fraction to the atmosphere warms the air, so the net effect is the higher SAT relative to GST in the winter season. Over the tropical coasts that exhibit the same behavior, the energy fluxes out of the ocean fraction of each grid-point also contribute to the higher warming of the air relative to the ground surface. Nevertheless, although at these locations high rates of evapotranspiration all year long also play an important role since they generate evaporative cooling of the ground surface as in the example described in Fig. 5.

The different examples used to illustrate the most important processes that may influence the air and soil temperature relationship at short time scales depict also relevant information about the propagation with depth of the annual cycle. For instance, at the grid-point located over the south-east of China in Fig. 5, both in DJF and JJA, the temperature offset between contiguous levels is noticeable with a gradient of about 5 °C in the first meter of the ground and of about 10 °C down to the lowest level. Comparable pictures with some differences in the magnitude of gradients can be seen in the previous figures.

Therefore, it is interesting to understand also the propagation of temperature below the surface. GCMs simulate purely conductive regimes, and the temperature variations that propagate to deeper soil layers are established at or near the ground surface (Smerdon et al., 2003). Thus it is important to assess the propagation of the temperature signal within the shallow subsurface. This issue is addressed by analyzing the relationship between GST and ST_{L8} .

Figure 6 (left) provides a spatial view of the temperature differences between GST and ST_{L8} for annual, DJF and JJA. The correlation is also shown in the right panels. SAT-GST differences, for DJF and JJA show the yearly cycle of temperature with negative (positive) SAT-GST for the NH (SH) in DJF and vice-versa for JJA, illustrating the conductive regime within the shallow subsurface. The annual temperature differences are low and the correlation is high almost globally as it is the balance between the respective patterns in JJA and DJF. However, the northernmost part of the globe exhibits larger temperature

differences (between 4 and 5°C) and lower correlation coefficients. The DJF and JJA patterns show that the annual offset and correlations for this part of the globe are mostly the result of the larger weight of those in JJA given that, **during this months during these months**, the temperature differences for a latitudinal band at ca. 60-70 N° are as large as 15°C and the correlation coefficients are close to 0. García-García et al. (2016) describe a similar behavior in some of the GCMs used in their analysis, 5 detected over areas where frozen ground persists during JJA. Indeed, the nature of the large departure in the temperature response at the shallow subsurface at these locations arises from non-conductive processes related to latent heat release/uptake of freezing and thawing of the water content above 1 m depth that may account for the subsurface heat transfer (Kane. et al., 2001).

To illustrate this mechanism, Fig. 7 shows the temperature evolution of SAT, GST and ST at different depths as well as the 10 soil ice content (SIC) in the upper soil layers for a grid-point located at the north of Canada. Over these areas the SIC in the upper 1 m depth increases (decrease) during the cold (warm) season. During JJA, SAT and GST increase/decrease at the same rate since no ice is present at the ground surface so it is warmed by radiative gain and heat is transferred to the atmosphere. However, for deeper soil layers, the energy available is employed in melting the SIC (note the lower SIC in L6 during JJA relative to DJF) and latent heat is required so these layers do not experience a temperature increase as the shallowest ones. 15 Therefore the temperature at L8 (~1 m depth) is kept below/near 0°C during the warm season due to the zero-curtain effect (Outcalt et al., 1990), while GST is centered around 12°C, leading to differences of ca. 15°C between GST and ST_{L8} and low correlation coefficient (0.28 for this grid-point) during JJA. In turn, during DJF, SAT distributes around -35°C and the frozen ground experiences skin temperatures of about -8°C and of about -4°C at 1 m depth. Consequently, the temperature offset between GST and ST_{L8} is largest in JJA. According to this, there are some non-conductive processes associated with 20 permanent frozen soils in the shallow subsurface that are included in the CLM4 parametrization (Lawrence et al., 2011) and that play an important role on the heat transport. The higher SIC in L8 during JJA relative to DJF (Fig. 7) arises from the fact that freezing of the active layer begins in late autumn and, due to the release of the latent heat of fusion, the freezing front is inhibited (Hinkel et al., 2001) and only reaches L8 in spring when SIC at this layer peaks. Then, as the thawing front penetrates downward in JJA, ice in the shallowest soil is melted, but it does not reach L8 until autumn when SIC at this layer is lowest. 25 Therefore, due to the thawing/freezing processes, seasonal changes at the upper and deeper subsurface levels are phase-shifted.

4.1 SAT-GST long-term changes

The mechanisms that have been described have an impact on the coupling between SAT and GST at short-time scales but they do not affect the long-term SAT-GST association if they are stationary since its influence would be constant at long timescales. 30 However, if such mechanisms experience variations with time, the SAT-GST relationship would also change with time. Thus, the thermal signature imprinted in the subsurface would not be representative of the long-term SAT variations (Bartlett et al., 2005). Figure 1b illustrates the existence of a constant offset between SAT and GST within the preindustrial period that changes during the industrial times indicating a variation in the long-term SAT-GST relationship. This may be relevant in the interpretation of borehole climate reconstructions because it may induce a long-term decoupling between SAT and GST in the

CESM-LME. At global scale, the changes in the long-term SAT-GST offset have an impact of about 0.05°C (Fig. 1a,b) and thus, do not seem to be very relevant at these scales. However, the impact could be larger for other GCMs with higher climate sensitivity or a different representation of surface processes that may contribute to decouple GST from SAT (e.g. snow cover). Likewise, within the CESM-LME simulations, impacts on decoupling may be important at regional or local scales.

5 To examine the spatial distribution of the long-term SAT and GST evolution during industrial times, we evaluate the linear ~~trends of both of them independently~~ trends of both temperatures independently during this period at every land model grid-point. Besides the ALL-F ensemble, we have also considered the anthropogenic single-forcing ensembles (Fig. 8), bearing in mind their potential influence on the processes that modulate the relationship between SAT and GST, as for instance: variations in snow cover, soil moisture and albedo, among others. The results in this section are shown considering information from
10 all members of the ensembles (see details in Table 3). For specific examples, one of the members will be used as indicated accordingly in figure captions.

Figure 8a,b describes a predominant warming for both SAT and GST in the ALL-F ensemble with the largest values distributed over north-west North America, north and central Eurasia, northeast Africa and southern South America. Interestingly, there are also regions showing negative trends like southeastern China, the north of the Black and Caspian sea regions, Pakistan,
15 some relatively small areas of central and south of Africa and Brazil. The warming trend pattern can be explained to a large extent if the 1850-2005 trends are calculated on the basis of the GHG-only ensemble (Fig. 8c,d) which is consistent with the global warming pattern due to the influence of GHGs (Hartmann et al., 2013). Indeed, if only the contribution of GHGs is considered, the warming would be higher and globally distributed. Figure 8 also indicates that the cooling in the ALL-F ensemble is mainly driven by the contribution of the LULC and OZ/AER external forcings. For instance, the cooling trends over
20 the Baltic sea and the north of the Black and Caspian seas that dominate the SAT and GST cooling trends during the industrial period in the ALL-F ensemble are the result of the influence of LULC changes (Fig. 8e,f) with additional contributions of OZ/AER (Fig. 8g,h). In addition, the negative trends of both SAT and GST over some areas of Africa, as well as over the northeast of Brazil, are also detectable in the LULC-only ensemble (Fig. 8e,f). Likewise, the OZ/AER-only ensemble also contributes to the cooling over Brazil, and the strong negative trends observed in south-east China in the ALL-F ensemble are
25 clearly identifiable in this ensemble (Fig. 8g,h).

Although the general pattern of cooling/warming during industrial times is broadly similar for SAT and GST, with a spatial pattern correlation of 0.60, 0.64, 0.38 and 0.53 in the ALL-F, GHG-only, LULC-only and OZ/AER-only ensembles, respectively,

it differs substantially in some regions. Note that there are considerable differences in the amplitude of warming trends in SAT and GST over Fennoscandia as well as at the northernmost part of North America in the ALL-F ensemble (Fig. 8a,b).

30 Similarly, over some areas of central and eastern Europe, SAT and GST industrial trends have different sign. There are also considerable differences in the amplitude of the cooling in SAT and GST over northeastern Brazil. Such dissimilar behaviors of SAT and GST during the industrial period are connected to variations in the energy fluxes at the surface in response to changes in the land surface characteristics due to the influence of the external forcings during this period.

~~For a more detailed analysis of SAT and GST long-term relationship, we applied a two-phase regression model (see Sect. 3) during the complete LM at every land model grid-point to the differences between SAT and GST in the ALL-F ensemble~~

35

as well as in the GHG-only and LULC-only ensembles (Fig. 9). For a more detailed analysis of SAT and GST long-term relationship, a two-phase regression model (Sect. 3) was applied at every land model grid-point to the SAT-GST differences in the ALL-F, GHG-only and LULC-only ensembles (Fig. 9). This allows to analyze long-term changes in the coupling without assuming any *a priori* condition on their time of occurrence and also to separately analyze the contribution of the different forcing factors and the temporal consistency among them. In the case of the OZ/AER-only ensemble, since this set of simulations spans from 1850 to 2005 CE, linear trends are only shown for the industrial period (Fig. 10). Figure 9a,b,c shows the year-of-change. Figure 9d,e,f (9g,h,i) represents the magnitude of the change before (after). Note that except for some areas, most continental regions experience statistically significant dates of change in phase with the start of the industrial period, during the 19th century (darker blue). For all of them, SAT-GST offset trends before the change are negligible or slightly negative in the ALL-F ensemble (Fig. 9d), indicating a stable relationship between SAT and GST globally before industrial times. This characteristic is also evident in the GHG and LULC-only ensembles (Fig. 9e,f). The strong positive or negative trends before the change (Fig. 9d,e,f) correspond to early dates of change around the 10th and 11th centuries (light green colors in the year-of-change maps), as a result of a change in the simulated differences between SAT and GST early in the LM and a stable behavior in GST-SAT differences thereafter. Note that for all of these regions the trend after the change is negligible and none of them is statistically significant (Fig. 9e,f,i). Figure 9a shows the year-of-change for the three ensembles; changes significant for the 80% of the ensemble members are shown exclusively in Fig. 9b. The three ensembles show dates of change that span the whole millennium. However, significant changes occur only during the last centuries. Two-phase regression allows for identifying that times of change in most regions take place prior to 1850, during the 18th and even in the 17th century (e.g. India) in the ALL-F. Trends before the change are not significant in any of the ensemble (not shown). Trends after the change are shown in Fig. 9c; significant areas in at least 80% of the ensemble members are exclusively shown in Fig. 9d. Note that large positive and negative trends in Fig. 9c coincide with the significant dates of change occurring during the last centuries of the LM.

The trends after the change, on the other hand, indicate significant regional variations in the long-term SAT-GST relationship. In general, annual SAT minus GST yields negative values (as shown in Fig. 2a) with the exception of the coastal areas as explained in Sect. 4. Thus, for continental areas ($SAT-GST < 0$), positive (negative) trends indicate that differences tend to get smaller (larger) in absolute values, and conversely for the limited coastal areas where SAT-GST differences are positive. Figure 1b allows for visualization of this behavior. Note the positive trend after the change when the difference between SAT and GST anomalies becomes smaller with time. Regionally, several circumstances account for impacting SAT-GST long-term coupling. On the one hand, decreasing SAT-GST differences over land may emerge from two conditions. Firstly, when there is a higher warming rate of SAT relative to GST as depicted in Fig. 8a,b over the northernmost part of North America, Fennoscandia, northeast Russia and some areas of central Eurasia. Secondly, when there is a cooling of both SAT and GST but the latter decreases at a higher pace as described in Fig. 8a,b for the region of northeastern Brazil and some areas of Africa. These two scenarios are represented in Fig. 9g Fig. 9c for the ALL-F, with positive trends over these regions after the change. On the other hand, the increase in SAT-GST difference arises either from the effect of rising GST in the presence of stable/decreasing SAT or due to the higher warming rate of GST relative to SAT. The former case is appreciated in Fig. 8a,b for the areas of

central and eastern Europe as well as eastern U.S whereas the latter is found over the Indian subcontinent and southeastern Asia. Note that both cases are represented in [Fig. 9g](#) [Fig. 9c \(ALL-F\)](#) with negative trends.

Trends from the GHG-only and the LULC-only ensembles help understanding the relative contributions to the long-term variations seen in the ALL-F simulations. For instance, the GHG-only ensemble ([Fig. 9h](#)) shows similar positive trends as [Fig. 9g](#)

5 [the ALL-F \(Fig. 9c\)](#) over northern North America, Fennoscandia, northeast Russia and central Eurasia, though with a much larger magnitude and geographical extension. Correspondingly, negative trends after the change in the LULC-only ensemble ([Fig. 9i](#)) are comparable to those in [Fig. 9g](#) [the ALL-F](#) for the regions of central and eastern Europe, eastern U.S., the Indian subcontinent and southeastern Asia. Additionally, the positive values over Brazil, as well as over central and southern Africa in [Fig. 9g](#) are also depicted in [Fig. 9i](#). [the ALL-F, are also depicted in the LULC-only. Most of these changes are robust in](#)

10 [80% of the ensemble members \(Fig. 9d\).](#)

Interestingly, the two-phase regression analysis does not expose any variation in SAT-GST long-term relationship over southeastern China, where the linear trends during the industrial period show a relatively strong decrease in both SAT and GST in the ALL-F and the OZ/AER-only ensembles (Fig. 8a,b,g,h). Furthermore, the linear trend of SAT-GST differences during industrial times for the OZ/AER-only simulations (Fig. 10) does not exhibit any SAT-GST decoupling over this region either.

15 This suggests that the dominant effect of OZ/AER forcing on the SAT and GST responses over this region is not affecting their long-term coupling. Nonetheless, Fig. 10 illustrates some interesting aspects of the SAT-GST relationship in the OZ/AER-only ensemble such as the negative contribution to the SAT-GST trends over North America, northern Europe, the Tibetan Plateau and central Asia. Additionally, the positive trends over northern Siberia are also notable, as well as the positive values over some relatively small areas of central and eastern Africa, the coast of Angola and eastern Brazil. Although the bulk of these

20 SAT-GST responses depicted in Fig. 10 does not translate to SAT-GST long-term decoupling in the ALL-F ensemble, they play an important role either in counteracting the influence of other external forcings or contributing to decoupling-related processes over some regions.

The following paragraphs aim at providing an insight into the relative contribution of the individual forcings and the associated physical mechanisms to the variations of the long-term SAT-GST association detected in Figs. 9 and 10.

25 In the cases of the long-term variations due to the LULC influence, changes in vegetation cover alter the radiative fluxes and water cycling at the surface due to the modification of the physical properties such as albedo, roughness and evapotranspiration (Pongratz et al., 2010). Figure 11 gives an example of how long-term changes in the energy fluxes at the surface due to LULC changes do impact the SAT-GST coupling at long timescales. It shows the 31-yr low pass filter outputs of SAT, GST, reflected shortwave radiation (RSW) and SHFLX evolution for a characteristic grid-point over the Great Lakes region (U.S.)

30 where a warming of GST relative to SAT is simulated in CESM-LME during the industrial period. Results are shown for one of the members of the ALL-F, LULC-only, GHG-only and OZ/AER-only ensembles. Around 1800 CE SAT tends to decrease whereas GST tends to increase in both the ALL-F and LULC-only simulations, while GHG-only and OZ/AER-only simulations do not display the same behavior that produces larger differences between SAT and GST represented by negative trends in [Fig. 9d](#). [Fig. 9c](#). At the same time, RSW and SHFLX exhibit large long-term variations in the ALL-F and LULC-only

35 simulations. Therefore, this modification of the long-term SAT-GST relationship is clearly the response to LULC changes.

Such variations in the surface energy fluxes over this region are likely a response of vegetation replacement from forested areas to grassland or croplands. Forested landscapes dissipate SHFLX more efficiently to the atmosphere due to a higher surface roughness than open fields (Jackson et al., 2008). In addition, lower vegetation types have higher reflectivity than forests. All the previous contribute to SAT decreases over these regions, especially in DJF. Furthermore, deforestation at mid- and high- latitudes tends to positively feedback with increases in snow cover (Anderson et al., 2011). This type of changes in LULC contribute to increase albedo, which is reinforced by changes in snow cover at these latitudes. Additionally, higher DJF snow cover tends to increase the insulation of the soil from the cold overlying air. ~~Therefore, the different temperature response of SAT and GST is related to higher albedo that leads to cooling the atmosphere, the reduction of SHFLX from the ground surface to the atmosphere and the additional snow cover positive feedback.~~ The combination of these mechanism leads to this temperature response of SAT and GST. This particular LULC process is important for corrupting the SAT-GST coupling at time scales relevant for the borehole theory (centennial) since the thermal signature recorded in GST during the industrial period would not be representative of the past long-term SAT variations in regions where this effect is dominant.

For the areas of central and eastern Europe, where a GST warming relative to SAT is also observed in Fig. 9c, the mechanisms are similar to those described in Fig. 11, because these areas were also subject to an intense transformation from forested areas to cropland ~~at the~~ prior to the beginning of the industrial period according to the LULC forcings considered in the CESM-LME (Pongratz et al., 2008; Hurtt et al., 2009).

Changes in vegetation cover are also important for long-term SAT and GST temperature differential response over tropical regions although ~~the driver mechanisms~~ the driving mechanisms are different from those at mid- and high-latitudes (Lee et al., 2011), and they deserve being considered. At the northeast of Brazil, both SAT and GST have negative trends in both the ALL-F and the LULC-only ~~simulations~~ ensembles during the industrial period (Fig. 8a,b,e,f). However, the decrease of GST is much larger than the one of SAT as represented in ~~Fig. 9e,i~~ Fig. 9c with positive trends after the change in both ensembles.

Figure 12 shows the temporal evolution of SAT, GST, RSW and LHFLX for a grid-point located at the northeast of Brazil. At the end of the 18th century, GST drops sharply whereas SAT slightly decreases. Likewise, RSW and LHFLX experience significant changes at the same time as a result of the modification of the surface characteristics. Such changes are present solely in the All-F and LULC-only simulations whereas GHG-only and OZ/AER-only ones show no differences in their evolution. The changes observed at this location in the energy fluxes likely correspond to transitions from open lands to a forested area (reforestation or afforestation) leading to lower albedo and higher evapotranspiration rates as is shown in Fig. 12. This situation leads to an apparent long-term cooling of GST relative to SAT at this location. The temperature response over this area is influenced by different mechanisms. First, the conversion from lower- to higher-type vegetation reduces the solar radiation that impinges on the surface and GST decreases by a radiative effect. Second, forested lands usually have lower albedo thus absorb more shortwave radiation (Zhao and Jackson, 2014). This surplus of energy is balanced by the increase in transpiration, and consequently, GST decreases also by a non-radiative process. The latter is especially important in humid climates (von Randow et al., 2004) as the one in this example. According to this, there is a net heat loss at the ground surface with a higher decrease in GST relative to the overlying air. The SAT-GST coupling becomes strong again after some period when the new vegetation

cover reaches a stable-estate stable state by the mid 20th century. For the areas of Africa, where there is also a cooling of GST relative to SAT (Fig. 9g,i); (Fig. 9c in the ALL-F and LULC-only ensembles), the mechanisms are comparable to those described for Fig. 12.

Over some of these tropical regions, there is also a contribution from the OZ/AER forcing to the SAT-GST response. Note that Fig. 10 shows positive trends over Uganda, the coast of Angola and over the east of Brazil, also noticeable in Fig. 9c. At these regions, the incoming solar radiation is reduced due to the effect of aerosols as they are an important element for cloud formation, thus contributing to a higher reflectivity of solar radiation (Tao et al., 2012). The description of the specific processes related to aerosol-cloud interaction goes beyond the scope of this study. Therefore, only the influence on the energy balance at the surface is addressed. As lower shortwave radiation impinges the surface, the energy gain decreases and therefore the ground surface heating is lower. The reduction in the energy gain at the surface is compensated by a lower dissipation via sensible heat whereas the fluxes of latent heat remain relatively constant or even increases in some areas due to a higher moisture as a result of increased precipitation. This mechanism is illustrated in Fig. 13 for a grid-point located at the region of Uganda. Note in the ALL-F simulation, the reduction in incoming shortwave radiation (SSW) after 1900 CE that leads to a decrease in SHFLX. Interestingly, for this location LHFLX experiences an increase at the same time as a result of increased precipitation (not shown) thus providing the source of moisture for evapotranspiration. This situation leads to a higher decrease in GST relative to SAT due to a net loss of energy at the ground surface. Similar SAT and GST responses are observed in the OZ/AER-only and LULC-only simulations, whereas in the GHG-only, SAT and GST increase.

The SAT-GST decoupling processes described above for individual grid-points are also important at larger spatial scales. Figure 14(left) shows an extension of the mechanism depicted in Fig. 12 including a larger area over the northeast of Brazil (between 1°S-11°S and 47°W-35°W). The negative trend since the 18th century is less accentuated for SAT than for GST (-0.14 and -0.53 °C/decade °C/century respectively in the All-F simulation and -0.07 and -0.33 in the LULC-only one) indicating a strong contribution of past LULC changes. The same air-subsurface temperature response occurs in other tropical and subtropical areas such as the east of Africa.

The regional analysis is extended to the area of southeastern China in order to illustrate additional information about the influence of different external forcings on the SAT-GST relationship. As discussed before, the negative trends for both SAT and GST over this region during the industrial period represented in the ALL-F and OZ/AER-only ensembles (Fig. 8a,b,g,h) do not entail a corruption of the SAT-GST long-term coupling. Figure 14(right) allows for an insight into this particular effect and the role of the OZ/AER forcing on the air and soil temperature responses over this region during industrial times. The negative long-term trend within the industrial period is seen only in the ALL-F and the OZ/AER-only simulations. Similarly, in both of them, there is a reduction in the RSW as well as in both the LHFLX and SHFLX. On the other hand, the SAT and GST evolution at industrial times in the GHG-only and LULC-only simulations does not follow the same path depicted in the ALL-F, thus highlighting the dominant influence of OZ/AER forcing. In this case, the variations of the energy fluxes at the surface depend on the reduction of the incoming shortwave radiation as a response of the anthropogenic aerosol-cloud interaction rather than by modifications of the land surface properties. Therefore, the decrease in the energy that impinges the surface is balanced by a decrease of both the sensible and latent heat fluxes. Hence, the air-soil interactions are not significantly

altered and the SAT-GST relationship remains stable.

Although the majority of the important long-term variations in the SAT-GST relationship at regional and local scales observed in the ALL-F ensemble (Fig. 9c) are induced by LULC changes, there are some regions in which the GHG forcing is the main driver for long-term SAT-GST decoupling. For instance, the positive trends after the change over Fennoscandia, northeast

5 Russia and north of North America observed in the ALL-F ensemble (~~Fig-9e~~) can be to a large extent explained by the influence of GHG-only ensemble inasmuch as a broadly similar picture is ~~also~~ portrayed over these regions ~~in Fig-9f~~ in both ensembles (Fig. 9c). In the case of the GHG-only ensemble, the strong warming of SAT relative to GST over these regions is driven by the increasing air temperature during industrial times due to the positive radiative forcing of GHGs in the presence of a considerable long-term reduction in simulated snow cover. González-Rouco et al. (2009) showed that such a

10 scenario would lead to a higher exposure of soil to cold winter air, therefore the soil would partially record colder temperatures, previously prevented by the snow cover insulating effect. In the ALL-F ensemble, this effect is damped for the fact of considering additional forcings that keep the snow cover relatively constant during industrial times. For instance, the contribution of the OZ/AER forcing is particularly important for counteracting the effect of GHGs since it leads to colder climate conditions due to its negative radiative forcing. Note the strong negative trends in Fig. 10 over North America, northern Europe and the

15 Tibetan Plateau that partially balance the effect of the GHGs over these regions. Nonetheless, in the ALL-F there is still an overall SAT warming relative to GST since the relatively stable snow cover is insulating the soil from a warmer SAT reducing the overall offset between them (Bartlett et al., 2005). Additionally, other processes can play some local roles on SAT-GST changes. One example is CO₂ fertilization generating increases in Leaf Area Index (e.g. Amazon rainforest), thus leading to higher evapotranspiration (Mankin et al., 2018) and GST cooling relative to SAT.

20 Figure 15 gives further insights into the interactions between anthropogenic forcings, their influence on global snow cover and consequently on the long-term SAT-GST relationship in the CESM-LME. The LM evolution of global SAT minus GST (top) and the annual global snow cover (bottom) for the ALL-F, LULC-only, GHG-only and OZ/AER-only simulations are shown for one member of each ensemble. Similar results are obtained if other members are selected. On the one hand, when the GHG-only ensemble is considered the global SAT-GST offset experiences a sharp long-term decrease in absolute value at the start of

25 the industrial period as well as a strong long-term reduction in global snow cover. In fact, the correlation between changes in the SAT-GST offset and snow cover in the GHG-only ensemble member is high -0.93 ($p < 0,05$). On the other hand, the overall effect of the LULC-only ensemble is a relatively small increase in the global snow cover mainly due to deforestation at mid and high latitudes as well as due to the negative radiative forcing of LULC as Earth's albedo increases (Myhre et al., 2013), leading to a small increase in the global SAT-GST offset. In the same way, the OZ/AER-only ensemble shows an increase in global

30 snow cover while the SAT-GST offset in industrial times exhibits a relatively slight increase. The interaction between different external forcings in the ALL-F ensemble leads to a relatively stable snow cover during the industrial period since the sharp decrease in snow, induced by the GHG forcing, is partially compensated by the counteracting effect of the LULC and OZ/AER forcings. Additional forcings such as volcanic may also contribute to counteracting GHGs effect at multi-decadal timescales (not shown). Consequently, in the presence of a warmer climate, there is a difference in the warming rate of SAT and GST in

35 industrial times at global scale in the ALL-F ensemble member (0.25 and 0.18 °C/century respectively). This scenario leads to

a net effect of a long-term decrease in the SAT-GST differences starting around 1800 CE as discussed in Fig. 1 and also evident in Fig. 15.

5 Conclusions

5 This work evaluates the stationarity of the coupling between SAT and GST temperatures as simulated by the CESM in an ensemble of experiments spanning the LM. The initial motivation for this work roots on previous literature (González-Rouco et al., 2006, 2003, 2009; García-García et al., 2016) that addresses the realism of the borehole hypothesis for climate reconstruction, namely, that SAT and GST vary synchronously and that reconstructing past GST changes from borehole temperature profiles is a good proxy for past SAT variations. The use of the CESM-LME allows for analyzing the influence of forcing changes on the SAT-GST covariability, both individually and as a group, by considering the different all-forcing and single-forcing ensembles. Additionally, having several experiment ensemble members for each given forcing type allows for disentangling the effects of internal variability from that of the forcing response. Ultimately, the coupling between SAT and GST is assessed at global and also at regional/local scales. This last purpose needs considering different mechanisms that contribute to SAT-GST variability within different climate types. In doing so, a variety of factors and conditions that contribute to the surface energy balance with different mechanisms is provided.

The CESM-LME shows that at global scale the SAT-GST coupling is strong above multi-decadal timescales since GST tracks SAT throughout the LM, as found in previous works. However, in spite of the strong coupling, the CESM-LME also reflects that the SAT-GST relationship has not remained constant along the whole LM at these spatiotemporal scales. Hence, the nature of such variation is evaluated.

20 Globally, snow cover is the most important agent in modulating the connection between SAT and GST. Therefore the variation of the SAT-GST relationship described by the CESM-LME simulations should be in principle driven by variations of global snow cover. Nevertheless, the simulated snow cover remains relatively stable at the time when SAT-GST coupling varies, thus this change cannot be solely explained by the influence of the snow cover. With this in mind, we explored with some detail different processes that may influence the SAT-GST relationship at different spatiotemporal scales. Firstly, we address processes acting at seasonal time scales that were identified from a spatial analysis of the SAT-GST differences and correlations. Secondly, the long-term evolution of the SAT-GST relationship is evaluated in the ALL-F, LULC-only, GHG-only and OZ/AER-only ensembles.

Several processes over different regions either relevant during DJF/JJA play an important role in impacting the SAT and GST coupling. Some examples are: snow cover over mid and high latitudes, discontinuous snow cover over the Tibetan Plateau region and seasonal variations in the energy fluxes at the surface. Although these processes are important for disrupting the SAT-GST relationship at seasonal scales, they have no implications on the long-term coupling if they are stationary. Nonetheless, if they experience variations with time the SAT-GST long-term relation may be impacted.

As discussed in Sect. 4.1 some of the anthropogenic external forcings have the potential to impose long-term variations on

processes that regulate the relationship between SAT and GST. Among them, the LULC changes are the most important ones since they modify the energy fluxes at the ground-air interface, and consequently corrupt the SAT-GST coupling locally and regionally at various time scales. One example is the response to the deforestation processes triggered by the expansion of agriculture mainly during the industrial period at mid and high latitudes, where SAT and GST long-term coupling is impacted
5 due to the variations in the albedo, surface roughness and hydrology. Similar decoupling processes related to LULC changes are found over different regions around the globe, as the one described in Sect. 4.1 over northeastern Brazil, some areas over Africa and over the Indian subcontinent. All of them driven by the long-term modifications of the energy fluxes at the surface either from increased evapotranspiration, reduced energy dissipation via sensible heat and others. Besides these kinds of decoupling processes induced by individual forcings, the interactions of a variety of mechanisms and feedbacks from different
10 external forcings can also exert an influence on the long-term SAT-GST relationship at different spatial scales. For instance, the effect of GHGs leads to a reduction of the snow cover during industrial times that is counterbalanced by the opposite effect of both LULC and OZ/AER forcings. As a consequence, the snow cover remains relatively stable over some regions during the industrial period in the presence of a warmer climate. This situation leads to a difference in the SAT and GST long-term evolution during the industrial period, since the snow cover is insulating the soil from a warmer SAT. This effect is present
15 over the NH high-latitudes of North America, Fennoscandia and northeastern Russia. Indeed, at global scale, the combination of a steady snow cover under warmer climate conditions is the dominant effect for explaining the variations in the long-term SAT-GST relationship.

Our findings indicate that the assumption of a strong relationship between SAT and GST may be impacted from local to regional scales by different mechanisms especially by the influence of LULC changes due to the modification of the energy
20 balance at the surface. Therefore, the interpretation of temperature reconstructions from borehole measurements at this spatial scales must consider LULC changes as a source of possible bias. The effects of additional external forcings may also exert some influence on processes such as variations in the snow cover, hydrology and other land surface properties, that may in turn feedback on the SAT-GST long-term coupling. At global scale, the influence of such local and regional decoupling processes is only ca. 0.05°C, hence the SAT-GST coupling at this spatial scale is supported by the CESM-LME.

25 The analysis using different GCMs may yield different levels of impacts. Overall, some consistency in the impacts among the available simulations of the LM with different GCMs may be expected, since the external forcings considered in the PMIP3/CMIP5 LM simulations are similar (Schmidt et al. , 2011) and should have a similar contribution to the extent that the impacts on the SAT-GST long-term relationship is mainly driven by the external forcings. 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). In addition, different model climate sensitivities and differences in the representation of surface processes (e.g. snow cover or soil moisture) may also contribute to produce different responses. 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 described above. The CMIP6/PMIP4 (Eyring et al. , 2017; Jungclaus et al. , 2017) offers an opportunity for exploring better sensitivity to different models since
30 largest ensembles of the LM with a more systematic sampling of forcings and processes would be expected.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research was supported by an FPI grant: BES-2015-075019, from the the Spanish Ministry of Economy, Industry and Competitiveness. We gratefully acknowledge the IIModels project, CGL2014-59644-R. We also thank Bette L. Otto-Bliesner for some personal communications regarding several aspects of the methodological design of the CESM-LME, [as well as the CESM1\(CAM5\) Last](#)

5 [Millennium Ensemble Community Project and supercomputing resources provided by NSF/CISL/Yellowstone.](#)

References

- Alexeev, V. A., Nicolsky, D. J., Romanovsky, V. E., and Lawrence, D. M.: An evaluation of deep soil configurations in the CLM3 for improved representation of permafrost, *Geophysical Research Letters*, 34, n/a–n/a, <https://doi.org/10.1029/2007GL029536>, <http://dx.doi.org/10.1029/2007GL029536>, 109502, 2007.
- 5 Ammann, C. M. and Wahl, E.: The importance of the geophysical context in statistical evaluations of climate reconstruction procedures, *Clim. Change*, 85, 71–88, doi:10.1007/s10584-007-9276-x, 2007.
- Anderson, R., Canadell, J., Randerson, J., Jackson, R., Hungate, B., Baldocchi, D., Ban-Weiss, G., Bonan, G., Caldeira, K., Cao, L., Diefenbaugh, N., Gurney, K., Kueppers, L., Law, B., Luysaert, S., and O'Halloran, T.: Biophysical considerations in forestry for climate protection, *Frontiers in Ecology and the Environment*, 9, 174–182, <https://doi.org/10.1890/090179>, 2011.
- 10 Bartlett, M. G., Chapman, D. S., and Harris, R. N.: Snow and the ground temperature record of climate change, *J. Geophys. Res.*, 109, F04008, doi:10.1029/2004JF000224, 2004.
- Bartlett, M. G., Chapman, D. S., and Harris, R. N.: Snow effect on North American ground temperatures, 1950-2002, *J. Geophys. Res.*, 110, F03008, doi:10.1029/2005JF000293, 2005.
- Beltrami, H. and Kellman, L.: An examination of short- and long-term air–ground temperature coupling, *Global and Planetary Change*, 38, 291 – 303, [https://doi.org/http://dx.doi.org/10.1016/S0921-8181\(03\)00112-7](https://doi.org/http://dx.doi.org/10.1016/S0921-8181(03)00112-7), <http://www.sciencedirect.com/science/article/pii/S0921818103001127>, 2003.
- 15 Beltrami, H., González-Rouco, J. F., and Stevens, M. B.: Subsurface temperatures during the last millennium: Model and Observation, *Geophys. Res. Lett.*, 33, L09705, doi:10.1029/2006GL026050, 2006.
- Berger, A., Loutre, M.-F., and Tricot, C.: Insolation and Earth's orbital periods, *Journal of Geophysical Research: Atmospheres*, 98, 10341–10362, <https://doi.org/10.1029/93JD00222>, <http://dx.doi.org/10.1029/93JD00222>, 1993.
- 20 Betts, R. A.: Biogeophysical impacts of land use on present-day climate: near-surface temperature change and radiative forcing, *Atmospheric Science Letters*, 2, 39 – 51, <https://doi.org/http://dx.doi.org/10.1006/asle.2001.0023>, <http://www.sciencedirect.com/science/article/pii/S1530261X01900234>, 2001.
- Betts, R. A., Falloon, P. D., Goldewijk, K. K., and Ramankutty, N.: Biogeophysical effects of land use on climate: Model simulations of radiative forcing and large-scale temperature change, *Agricultural and Forest Meteorology*, 142, 216 – 233, <https://doi.org/http://dx.doi.org/10.1016/j.agrformet.2006.08.021>, <http://www.sciencedirect.com/science/article/pii/S0168192306002887>, the Contribution of Agriculture to the State of Climate, 2007.
- 25 Bodri, L. and Cermak, V.: Borehole temperatures, climate change and the pre-observational surface air temperature mean: allowance for hydraulic conditions, *Global and Planetary Change*, 45, 265 – 276, <https://doi.org/http://dx.doi.org/10.1016/j.gloplacha.2004.09.001>, <http://www.sciencedirect.com/science/article/pii/S0921818104001225>, 2005.
- 30 Brovkin, V., Claussen, M., Driesschaert, E., Fichefet, T., Kicklighter, D., Loutre, M. F., Matthews, H. D., Ramankutty, N., Schaeffer, M., and Sokolov, A.: Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity, *Climate Dynamics*, 26, 587–600, <https://doi.org/10.1007/s00382-005-0092-6>, <https://doi.org/10.1007/s00382-005-0092-6>, 2006.
- Chapman, D. S., Bartlett, M. G., and Harris, R. N.: Comment on 'Ground vs. surface air temperature trends: implications for borehole surface temperature reconstructions' by M. E. Mann and G. Schmidt, *Geophys. Res. Lett.*, 31, L07205, doi:10.1029/2003GL019054, 2004.
- 35 Christiansen, B. and Ljungqvist, F. C.: The extra-tropical Northern Hemisphere temperature in the last two millennia: reconstructions of low-frequency variability, *Climate of the Past*, 8, 765–786, doi: 10.5194/cp-8-765-2012, 2012.

- Ferguson, G., Beltrami, H., and Woodbury, A. D.: Perturbation of ground surface temperature reconstructions by groundwater flow?, *Geophys. Res. Lett.*, 33, L13 708, doi:10.1029/2006GL026 634, 2006.
- Fernández-Donado, L., González-Rouco, J. F., Raible, C. C., Ammann, C. M., Barriopedro, D., García-Bustamante, E., Jungclaus, J. H., Lorenz, S. J., Luterbacher, J., Phipps, S. J., Servonnat, J., Swingedouw, D., Tett, S. F. B., Wagner, S., Yiou, P., and Zorita, E.: Large-scale temperature response to external forcing in simulations and reconstructions of the last millennium, *Climate of the Past*, 9, 393–421, <https://doi.org/10.5194/cp-9-393-2013>, <http://www.clim-past.net/9/393/2013/>, 2013.
- Gao, C. C., Robock, A., and Ammann, C.: Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models, *J. Geophys. Res.*, 113, D23 111, doi: 10.1029/2008JD010 239, 2008.
- García-García, A., Cuesta-Valero, F. J., Beltrami, H., and 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, 044 022, <http://stacks.iop.org/1748-9326/11/i=4/a=044022>, 2016.
- González-Rouco, J. F., von Storch, H., and Zorita, E.: Deep soil temperature as proxy for surface air-temperature in a coupled model simulation of the last thousand years, *Geophys. Res. Lett.*, 30, 2116–2119, 2003.
- González-Rouco, J. F., Beltrami, H., Zorita, E., and von Storch, H.: Simulation and inversion of borehole temperature profiles in surrogate climates: Spatial distribution and surface coupling, *Geophys. Res. Lett.*, 33, L01 703,doi:10.1029/2005GL024 693, 2006.
- González-Rouco, J. F., Beltrami, H., Zorita, E., and Stevens, B.: Borehole climatology: a discussion based on contributions from climate modelling, *Clim. Past*, 5, 99–127, 2009.
- Hansen, J., Ruedy, R., Sato, M., and Lo, K.: GLOBAL SURFACE TEMPERATURE CHANGE, *Reviews of Geophysics*, 48, n/a–n/a, <https://doi.org/10.1029/2010RG000345>, http://dx.doi.org/10.1029/2010RG000345_rG4004, 2010.
- Hartmann, D., Klein Tank, A., Rusticucci, M., Alexander, L., Brönnimann, S., Charabi, Y., Dentener, F., Dlugokencky, E., Easterling, D., Kaplan, A., Soden, B., Thorne, P., Wild, M., and Zhai, P.: Observations: Atmosphere and Surface, book section 2, p. 159–254, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.008>, www.climatechange2013.org, 2013.
- Hegerl, G., Zwiers, F., and Tebaldi, C.: Patterns of change: whose fingerprint is seen in global warming?, *Environmental Research Letters*, 6, –, <https://doi.org/10.1088/1748-9326/6/4/044025>, 2011.
- Hegerl, G. C., Crowley, T. J., Hyde, W. T., and Frame, D. J.: Climate sensitivity constrained by temperature reconstructions over the past seven centuries, *Nature*, 440, 1029–1032, 2006.
- Hinkel, K., Paetzold, F., Nelson, F., and Bockheim, J.: Patterns of soil temperature and moisture in the active layer and upper permafrost at Barrow, Alaska: 1993–1999, *Global and Planetary Change*, 29, 293 – 309, [https://doi.org/https://doi.org/10.1016/S0921-8181\(01\)00096-0](https://doi.org/https://doi.org/10.1016/S0921-8181(01)00096-0), <http://www.sciencedirect.com/science/article/pii/S0921818101000960>, inference of Climate Change from Geothermal Data, 2001.
- Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., and Elliott, S.: CICE: the Los Alamos Sea Ice Model Documentation and Software User’s Manual Version 5.1, 2015.
- Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, *Bulletin of the American Meteorological Society*, 94, 1339–1360, <https://doi.org/10.1175/BAMS-D-12-00121.1>, <http://dx.doi.org/10.1175/BAMS-D-12-00121.1>, 2013.

- Hurt, G. C., Chini, L. P., Frohling, S., Betts, R., Feedema, J., Fischer, G., Goldewijk, K. K., Hibbard, K., Janetos, A., Jones, C., Kindermann, G., Kinoshita, T., Riahi, K., and Smith, E. S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D., and Wang, Y.: Harmonization of Global Land Use Scenarios for the Period 1500-2100 for IPCC-AR5, Integrated Land Ecosystem-Atmosphere Processes Study (iLEAPS) Newsletter, 7, 6–8, 2009.
- 5 Jackson, R. B., Randerson, J. T., Canadell, J. G., Anderson, R. G., Avissar, R., Baldocchi, D. D., Bonan, G. B., Caldeira, K., Diffenbaugh, N. S., Field, C. B., Hungate, B. A., Jobbágy, E. G., Kueppers, L. M., Noretto, M. D., and Pataki, D. E.: Protecting climate with forests, *Environmental Research Letters*, 3, 044 006, <http://stacks.iop.org/1748-9326/3/i=4/a=044006>, 2008.
- Jones, P. D., Briffa, K. R., Osborn, T. J., Lough, J. M., Van Ommen, T. D., Vinther, B. M., Luterbacher, J., Wahl, E. R., Zwiers, F. W., Mann, M. E., Schmidt, G. A., Ammann, C. M., Buckley, B. M., Cobb, K. M., Esper, J., Goose, H., Graham, N., Jansen, E., Kiefer, T., Kull, C., Küttel, M., Mosley-Thompson, E., Overpeck, J. T., Riedwyl, N., Schulz, M., Tudhope, A. W., Villalba, R., Wanner, H., Wilff, E., and Xoplake, E.: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects, *Holocene*, 19, 3–49, <https://doi.org/10.1177/0959683608098952>, <https://doi.org/10.1177/0959683608098952>, 2009.
- 10 Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010, *Journal of Geophysical Research: Atmospheres*, 117, n/a–n/a, <https://doi.org/10.1029/2011JD017139>, <http://dx.doi.org/10.1029/2011JD017139>, d05127, 2012.
- 15 Kane, D., Hinkel, K., Goering, D., Hinzman, L., and Outcalt, S.: Non-conductive heat transfer associated with frozen soils, *Global and Planetary Change*, 29, 275–292, 2001.
- Kohl, T.: Palaeoclimatic temperature signals — can they be washed out?, *Tectonophysics*, 291, 225 – 234, [https://doi.org/http://dx.doi.org/10.1016/S0040-1951\(98\)00042-0](https://doi.org/http://dx.doi.org/10.1016/S0040-1951(98)00042-0), <http://www.sciencedirect.com/science/article/pii/S0040195198000420>, heat Flow and the Structure of the Lithosphere - IV, 1998.
- 20 Kukkonen, I. T. and Clauser, C.: Simulation of heat transfer at the Kola deep-hole site- implications for advection, heat refraction and paleoclimatic effects, *Geophys. J. Int.*, 41, 175–191, 1994.
- Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N., and Teng, H.: Last millennium climate and its variability in CCSM4, *J. Climate*, xxx, yyy–yyy, 2012.
- 25 Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, *Journal of Advances in Modeling Earth Systems*, 3, n/a–n/a, <https://doi.org/10.1029/2011MS00045>, <http://dx.doi.org/10.1029/2011MS00045>, m03001, 2011.
- Lee, X., Goulden, M. L., Hollinger, D. Y., Barr, A., Black, T. A., Bohrer, G., Bracho, R., Drake, B., Goldstein, A., Gu, L., Katul, G., Kolb, T., Law, B. E., Margolis, H., Meyers, T., Monson, R., Munger, W., Oren, R., Paw U, K. T., Richardson, A. D., Schmid, H. P., Staebler, R., Wofsy, S., and Zhao, L.: Observed increase in local cooling effect of deforestation at higher latitudes, *Nature*, 479, 384–387, <https://doi.org/10.1038/nature10588>, <http://dx.doi.org/10.1038/nature10588>, 2011.
- 30 Legutke, S. and Voss, R.: The Hamburg Atmosphere-Ocean Coupled Circulation Model ECHO-G, Tech. Rep. 18, DKRZ, Hamburg, 1999.
- MacDougall, A. H. and Beltrami, H.: Impact of deforestation on subsurface temperature profiles: implications for the borehole paleoclimate record, *Environmental Research Letters*, 12, 074 014, <http://stacks.iop.org/1748-9326/12/i=7/a=074014>, 2017.
- MacDougall, A. H., Beltrami, H., González-Rouco, J. F., Stevens, M. B., and Bourlon, E.: Comparison of observed and General Circulation Model derived continental sub-surface heat flux in the Northern Hemisphere, *J. Geophys. Res.*, 115, D12 109, doi:10.1029/2009JD013 170, 2010.

- MacFarling Meure, C., Etheridge, D., Trudinger, C., Steele, P., Langenfelds, R., Van Ommen, T., Smith, A., and Elkins, J.: Law Dome CO₂, CH₄ and N₂O ice core records extended to 2000 years BP, *Geophys. Res. Lett.*, 33, L14 810, 2006.
- Mankin, J. S., Seager, R., Smerdon, J. E., Cook, B. I., Williams, A. P., and Horton, R. M.: Blue Water Trade-Offs With Vegetation in a CO₂-Enriched Climate, *Geophysical Research Letters*, 45, 3115–3125, <https://doi.org/10.1002/2018GL077051>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2018GL077051>, 2018.
- 5
- Mann, M. E. and Schmidt, G. A.: Ground vs. Surface air temperature trends: implications for borehole surface temperature reconstructions, *Geophys. Res. Lett.*, 1, 1–1, 2003.
- Mareschal, J.-C. and Beltrami, H.: Evidence for recent warming from perturbed geothermal gradients: examples from eastern Canada, *Climate Dynamics*, 6, 135–143, <https://doi.org/10.1007/BF00193525>, <http://dx.doi.org/10.1007/BF00193525>, 1992.
- 10
- Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., González Rouco, J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X., and Timmermann, A.: Information from Paleoclimate Archives, book section 5, p. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.013>, www.climatechange2013.org, 2013.
- Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, book section 8, p. 659–740, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, <https://doi.org/10.1017/CBO9781107415324.018>, www.climatechange2013.org, 2013.
- 15
- Neale, R. B., Gettelman, A., Park, S., Conley, A. J., Kinnison, D., Marsh, D., Smith, A. K., Vitt, F., Morrison, H., Cameron-smith, P., Collins, W. D., Iacono, M. J., Easter, R. C., Liu, X., Taylor, M. A., chieh Chen, C., Lauritzen, P. H., Williamson, D. L., Garcia, R., francois Lamarque, J., Mills, M., Tilmes, S., Ghan, S. J., Rasch, P. J., and Meteorology, M.: Description of the NCAR Community Atmosphere Model (CAM 5.0), Tech. Note NCAR/TN-486+STR, Natl. Cent. for Atmos, pp. 2009–038 451, 2012.
- 20
- Nicolsky, D. J., Romanovsky, V. E., Alexeev, V. A., and Lawrence, D. M.: Improved modeling of permafrost dynamics in a GCM land-surface scheme, *Geophys. Res. Lett.*, 34, L8501,doi:10.1029/2007GL029 525, 2007.
- Oleson, K. W., Niu, G.-Y., Yang, Z.-L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A., and Qian, T.: Improvements to the Community Land Model and their impact on the hydrological cycle, *Journal of Geophysical Research: Biogeosciences*, 113, n/a–n/a, <https://doi.org/10.1029/2007JG000563>, <http://dx.doi.org/10.1029/2007JG000563>, g01021, 2008.
- 25
- Oleson, K. W., Lawrence, D. M., B, G., Flanner, M. G., Kluzek, E., J, P., Levis, S., Swenson, S. C., Thornton, E., Feddema, J., Heald, C. L., francois Lamarque, J., yue Niu, G., Qian, T., Running, S., Sakaguchi, K., Yang, L., Zeng, X., Zeng, X., and Decker, M.: Technical Description of version 4.0 of the Community Land Model (CLM), 2010.
- 30
- Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G.: Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model, *Bulletin of the American Meteorological Society*, 97, 735–754, <https://doi.org/10.1175/BAMS-D-14-00233.1>, <http://dx.doi.org/10.1175/BAMS-D-14-00233.1>, 2016.
- Outcalt, S. I., Nelson, F. E., and Hinkel, K. M.: The zero-curtain effect: Heat and mass transfer across an isothermal region in freezing soil, *Water Resources Research*, 26, 1509–1516, <https://doi.org/10.1029/WR026i007p01509>, <http://dx.doi.org/10.1029/WR026i007p01509>, 1990.
- 35
- Pollack, H. N. and Huang, S.: Climate reconstruction from subsurface temperatures, *Annu. Rev. Earth. Planet. Sci.*, 28, 339–365, 2000.
- Pollack, H. N. and Smerdon, J. E.: Borehole climate reconstructions: spatial structure and hemispheric averages, *J. Geophys. Res.*, 109, D11 106, doi: 10.1029/2003JD004 163, 2004.

- Pollack, H. N., Huang, S., and Shen, P. Y.: Climate change record in subsurface temperatures: a global perspective, *Science*, 282, 279–281, 1998.
- Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, *Global Biogeochemical Cycles*, 22, n/a–n/a, <https://doi.org/10.1029/2007GB003153>, <http://dx.doi.org/10.1029/2007GB003153>, gB3018, 5 2008.
- Pongratz, J., Reick, C. H., Raddatz, T., and Claussen, M.: Biogeophysical versus biogeochemical climate response to historical anthropogenic land cover change, *Geophysical Research Letters*, 37, n/a–n/a, <https://doi.org/10.1029/2010GL043010>, <http://dx.doi.org/10.1029/2010GL043010>, 108702, 2010.
- Pongratz, J., Reick, C., Raddatz, T., and Claussen, M.: A reconstruction of global agricultural areas and land cover for the last millennium, *Global Biogeochemical Cycles*, 22, GB3018, doi: 10.1029/2007GB003153., 2008.
- Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: croplands from 1700 to 1992, *Global Biogeochemical Cycles*, 13(4), 997–1027, 1992.
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., J.Pongratz, Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium, *Geosci. Model Dev.*, 4, 33–45, 2011.
- Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., J.Pongratz, Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium, *Geosci. Model Dev.*, 5, 185–191, 2012.
- Smerdon, J. E.: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments., *WIREs Clim. Change*, 3, 63–77, 20 doi: 10.1002/wcc.149, 2012.
- Smerdon, J. E. and Stieglitz, M.: Simulating heat transport of harmonic temperature signals in the Earth’s shallow subsurface: lower-boundary sensitivities, *Geophys. Res. Lett.*, 33, L14402, doi: 10.1029/2006GL026816, 2006.
- Smerdon, J. E., Pollack, H. N., Enz, J. W., and Lewis, M. J.: Conduction-dominated heat transport of the annual temperature signal in soil, *J. Geophys. Res.*, 108, 2431, doi: 10.1029/2002JB002351, 2003.
- 25 Smerdon, J. E., Pollack, H. N., Cermak, V., Enz, J. W., Kresl, M., Safanda, J., and Wehmiller, J. F.: Air-ground temperature coupling and subsurface propagation of annual temperature signals, *J. Geophys. Res.*, 109, D21107, doi: 10.1029/2004JD0050506, 2004.
- Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., Dukowicz, J., Eden, C., Fox-Kemper, B., Gent, P., Hecht, M., Jayne, S., M. Jochum, W. Large, K. L., Maltrud, M., Norton, N., Peacock, S., Vertenstein, M., and Yeager, S.: The Parallel Ocean Program (POP) Reference Manual, Tech. Rep. LAUR-10-01853, Los Alamos National Laboratory, <http://tinyurl.com/SJBBD10>, 2010.
- 30 Solow, A. R.: Testing for Climate Change: An Application of the Two-Phase Regression Model, *Journal of Climate and Applied Meteorology*, 26, 1401–1405, [https://doi.org/10.1175/1520-0450\(1987\)026<1401:TFCCAA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1987)026<1401:TFCCAA>2.0.CO;2), [http://dx.doi.org/10.1175/1520-0450\(1987\)026<1401:TFCCAA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1987)026<1401:TFCCAA>2.0.CO;2), 1987.
- Stieglitz, M., Dery, S. J., Romanovsky, V. E., and Ostercamp, T. E.: The role of snow cover in the warming of the arctic permafrost, *Geophys. Res. Lett.*, 30, 1721, doi:10.1029/2003GL017337, 2003.
- 35 Tao, W., Chen, J., Li, Z., Wang, C., and Zhang, C.: Impact of aerosols on convective clouds and precipitation, *Reviews of Geophysics*, 50, <https://doi.org/10.1029/2011RG000369>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011RG000369>, 2012.
- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, *Bulletin of the American Meteorological Society*, 93, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>, <https://doi.org/10.1175/BAMS-D-11-00094.1>, 2012.

- Trenberth, K. E.: Some effects of finite sample size and persistence on meteorological statistics. Part I: autocorrelations, *Mon. Wea. Rev.*, 112, 2359–2368, 1984.
- Vieira, A., Solanki, S. K., Krivova, N. A., and Usoskin, I.: Evolution of the solar irradiance during the Holocene, *Astronomy & Astrophysics*, 531, A6, <https://doi.org/10.1051/0004-6361/201015843>, 2011.
- 5 von Randow, C., Manzi, A. O., Kruijt, B., de Oliveira, P. J., Zanchi, F. B., Silva, R. L., Hodnett, M. G., Gash, J. H. C., Elbers, J. A., Waterloo, M. J., Cardoso, F. L., and Kabat, P.: Comparative measurements and seasonal variations in energy and carbon exchange over forest and pasture in South West Amazonia, *Theoretical and Applied Climatology*, 78, 5–26, <https://doi.org/10.1007/s00704-004-0041-z>, <https://doi.org/10.1007/s00704-004-0041-z>, 2004.
- Zhao, K. and Jackson, R. B.: Biophysical forcings of land-use changes from potential forestry activities in North America, *Ecological*
- 10 *Monographs*, 84, 329–353, <http://www.jstor.org/stable/43187893>, 2014.

Table 1. Soil layers and node depths in the CLM4. Note that the node depth, that are the depth at which the thermal properties are defined for soil layers (Oleson et al., 2010), do not necessarily coincide with the center of the layer depth.

Layer	Layer depth (m)	Node depth (m)
L1	0.017	0.007
L2	0.045	0.027
L3	0.090	0.062
L4	0.165	0.118
L5	0.289	0.212
L6	0.492	0.366
L7	0.828	0.619
L8	1.382	1.038
L9	2.296	1.727
L10	3.801	2.864
L11	6.284	4.739
L12	10.377	7.829
L13	17.125	12.925
L14	28.252	21.326
L15	42.103	35.177

Table 2. Simulations and LM external forcing reconstructions used in CESM-LME. The single forcing simulations cover from 850 to 2005 CE except those of anthropogenic ozone and aerosols that span the period 1850 to 2005 CE. Legend for external forcing: SOL, changes in total solar irradiance; VOLC, volcanic activity; GHG, concentrations of the well-mixed greenhouse gases CO₂, CH₄, and N₂O; LULC, land use land cover changes; (ORB) orbital variations; and OZ/AER, anthropogenic ozone and aerosols

Forcing	No. of simulations	Reference
ALL-F	10	-
SOL	4	Vieira et al. (2011)
VOLC	5	Gao et al. (2008)
GHG	3	MacFarling Meure et al. (2006)
LULC	3	Pongratz et al. (2008) dataset, spliced to Hurtt et al. (2009) at 1500 CE. The only plant functional types (PFTs) that are changed are those for crops and pasture; all other PFTs remain at their 1850 control prescriptions
ORB	3	The CESM model adjusts yearly orbital position (eccentricity, obliquity and precession) following Berger et al. (1993)
OZ/AER	2	Fixed at the 1850 control simulation values until 1850 and then include the evolving anthropogenic changes to 2005. Stratospheric aerosols are prescribed in the model as a fixed single-size distribution in the three layers in the lower stratosphere above the tropopause. The ozone forcing is from the Whole Atmosphere Community Climate Model (WACCM)

Table 3. Simulations used in this study from the CESM-LME. The first and second columns present the Acronyms used in this manuscript for ensembles and ensemble members respectively. The id of the original experiment files is provided in column 3.

Ensemble acronym	Ensemble member	Simulation id
ALL-F	ALL-F _i	b.e11.BLMTRC5CN.f19_g16.0i
	i=1,...,10	i=01,...,10
GHG-only	GHG _i	b.e11.BLMTRC5CN.f19_g16.GHG.00i
	i=1,2,3	i=1,2,3
LULC-only	LULC _i	b.e11.BLMTRC5CN.f19_g16.LULC_HurttPogratz.00i
	i=1,2,3	i=1,2,3
OZ/AER-only	OZ/AER _i	b.e11.BLMTRC5CN.f19_g16.OZONE_AER.00i
	i=1,2	i=1,2

Table 4. Temporal correlation coefficients between last millennium SAT-GST, SAT-ST_{L8} and SAT-ST_{L15} anomalies relative to the 1850-2005 mean for the experiment shown in Fig. 1. The left side indicates the correlation at yearly resolution whereas the right side shows 31-yr low-pass filter outputs of SAT, GST, ST_{L8} and ST_{L15} for annual, DJF and JJA. Coefficients highlighted in bold are statistically significant at $p < 0.05$

	Yearly			31-yr filtered		
	Annual	DJF	JJA	Annual	DJF	JJA
SAT-GST	0.96	0.85	0.99	0.98	0.94	0.99
SAT-ST_{L8}	0.95	0.81	0.94	0.98	0.93	0.96
SAT-ST_{L15}	0.28	0.27	0.14	0.86	0.85	0.65

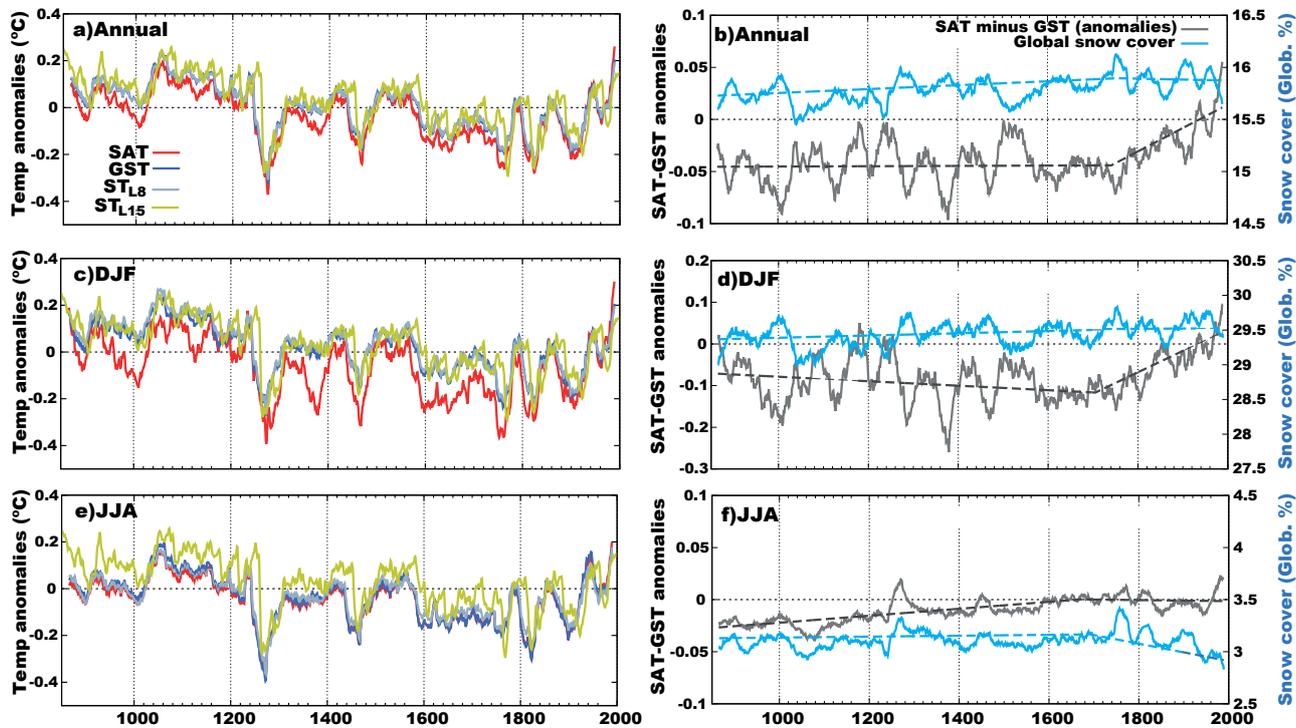


Figure 1. Left. (a) Annual, (c) DJF and (e) JJA global LM evolution of SAT, GST, ST_{L8} and ST_{L15} anomalies relative to 1850-2005 CE for the ALL- F_2 ensemble member (Table 3; Otto-Bliesner et al., 2016). Right. (b) Annual, (d) DJF and (f) JJA global LM evolution of SAT-GST offset (SAT minus GST anomalies) and global percentage of snow cover. Dashed lines are the result of a two-phase regression, indicating the linear trend that represents the best fit to the data before and after the estimated point of change. For snow cover, dashed lines represent linear fits to the data using the change points found for SAT-GST. All series are 31-yr moving average filter outputs except for ST_{L15} which is represented as the direct model output.

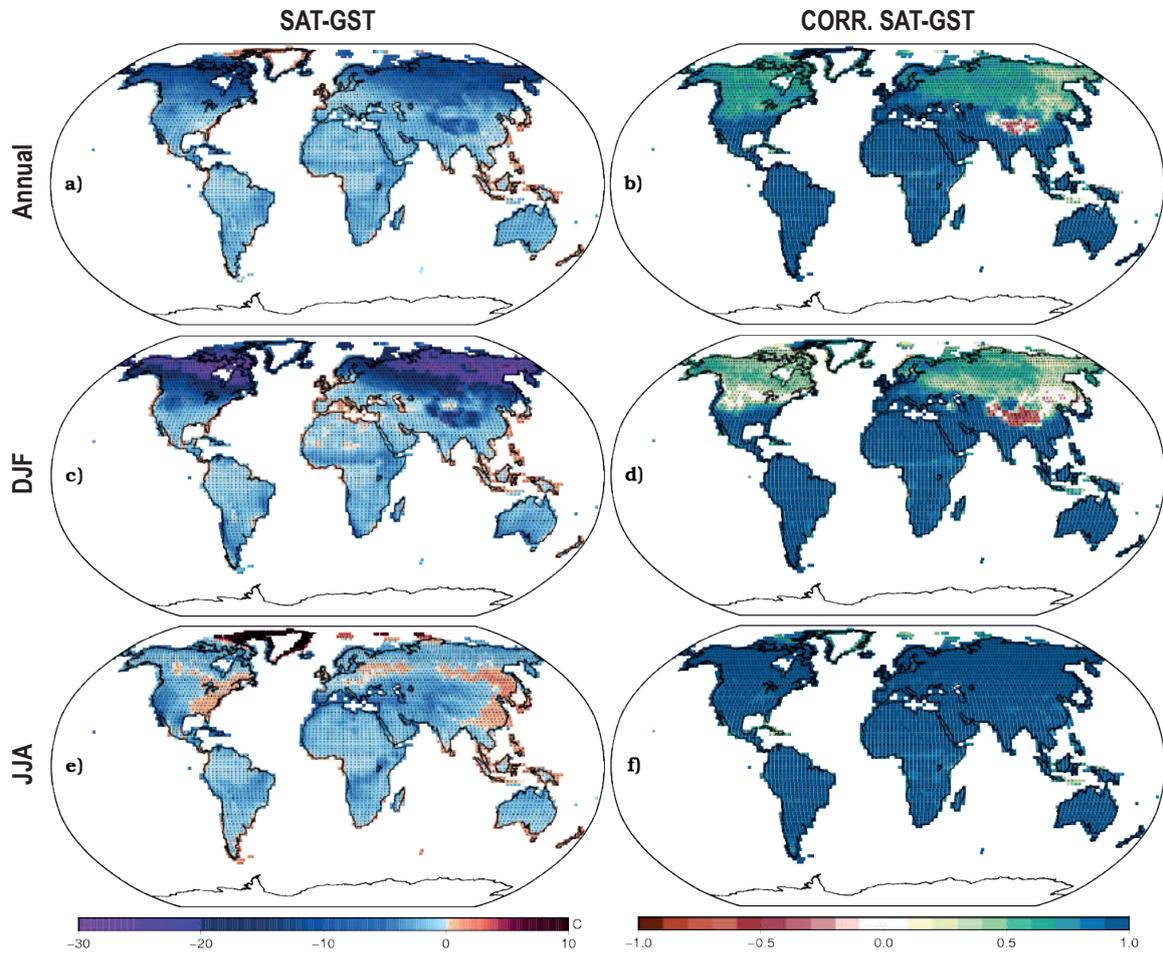


Figure 2. Left. (a) Annual, (c) DJF, (e) JJA spatial distribution of SAT minus GST differences during the 850-2005 CE period. Right. (b) Annual, (d) DJF, (f) JJA spatial distribution of the correlation coefficients between SAT and GST for the same period. Black dots indicate that 80% of the members within the ALL-F ensemble deliver statistical significance ($p < 0.05$).

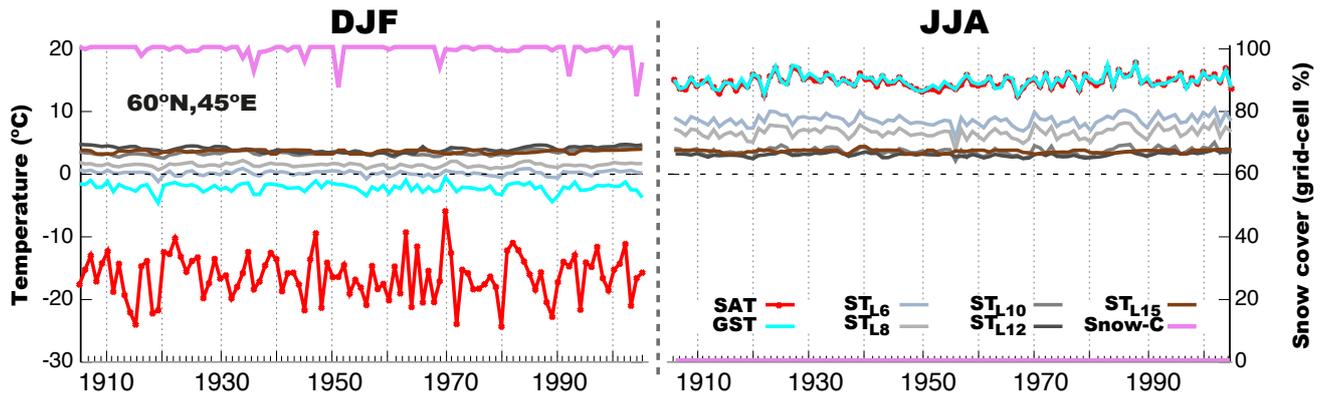


Figure 3. Evolution of DJF and JJA SAT, GST, ST_{L6} , ST_{L8} , ST_{L10} , ST_{L12} and ST_{L15} and percentage of snow cover during the period 1900-2005 CE for a grid-point (north-east of Russia; 60 °N, 45 °E) from the ALL-F₂ simulation (Table 3) where snow cover is characteristic during the cold season. Dashed lines indicate 0 °C.

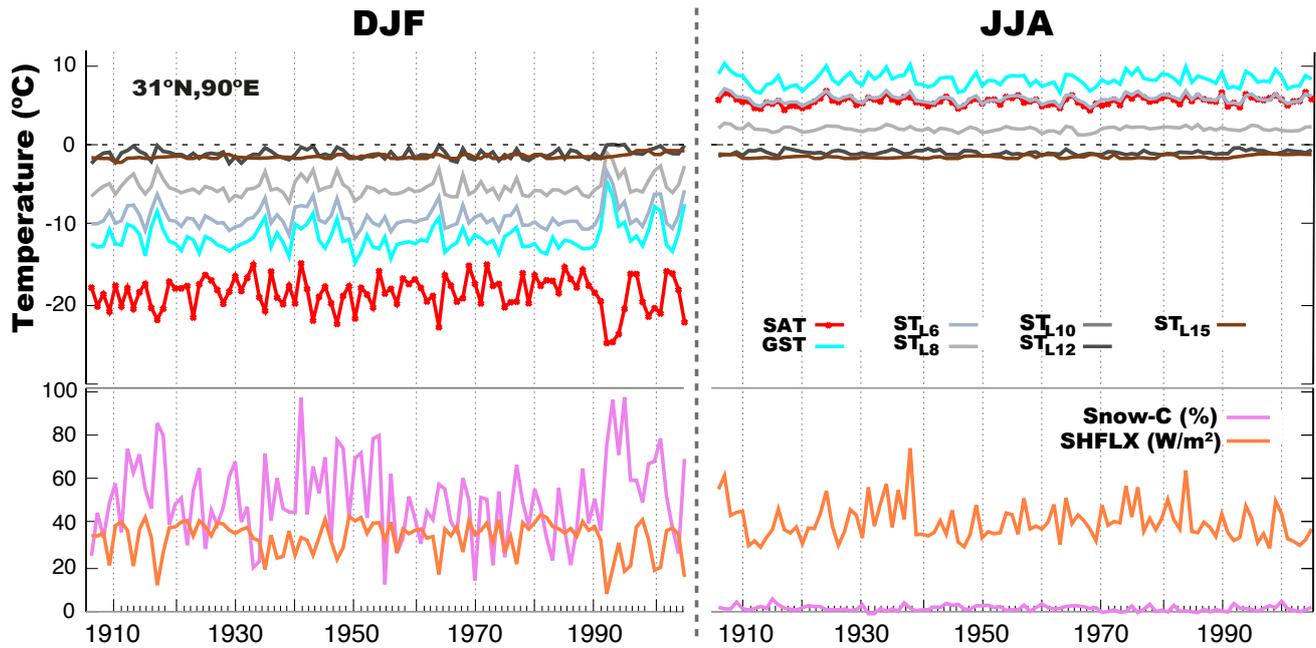


Figure 4. Top. DJF and JJA evolution of SAT, GST, ST_{L6} , ST_{L8} , ST_{L10} , ST_{L12} and ST_{L15} for a grid-point (Tibetan Plateau region; 31 °N, 90 °E) from the ALL-F₂ simulation (Table 3) during the period 1900–2005 CE. Dashed lines indicate 0 °C. Bottom. Percentage of snow cover (Snow-C) and surface sensible heat flux (SHFLX). Note that both Snow-C and SHFLX are indicated in the left axis. The units are shown within brackets in the figure legend.

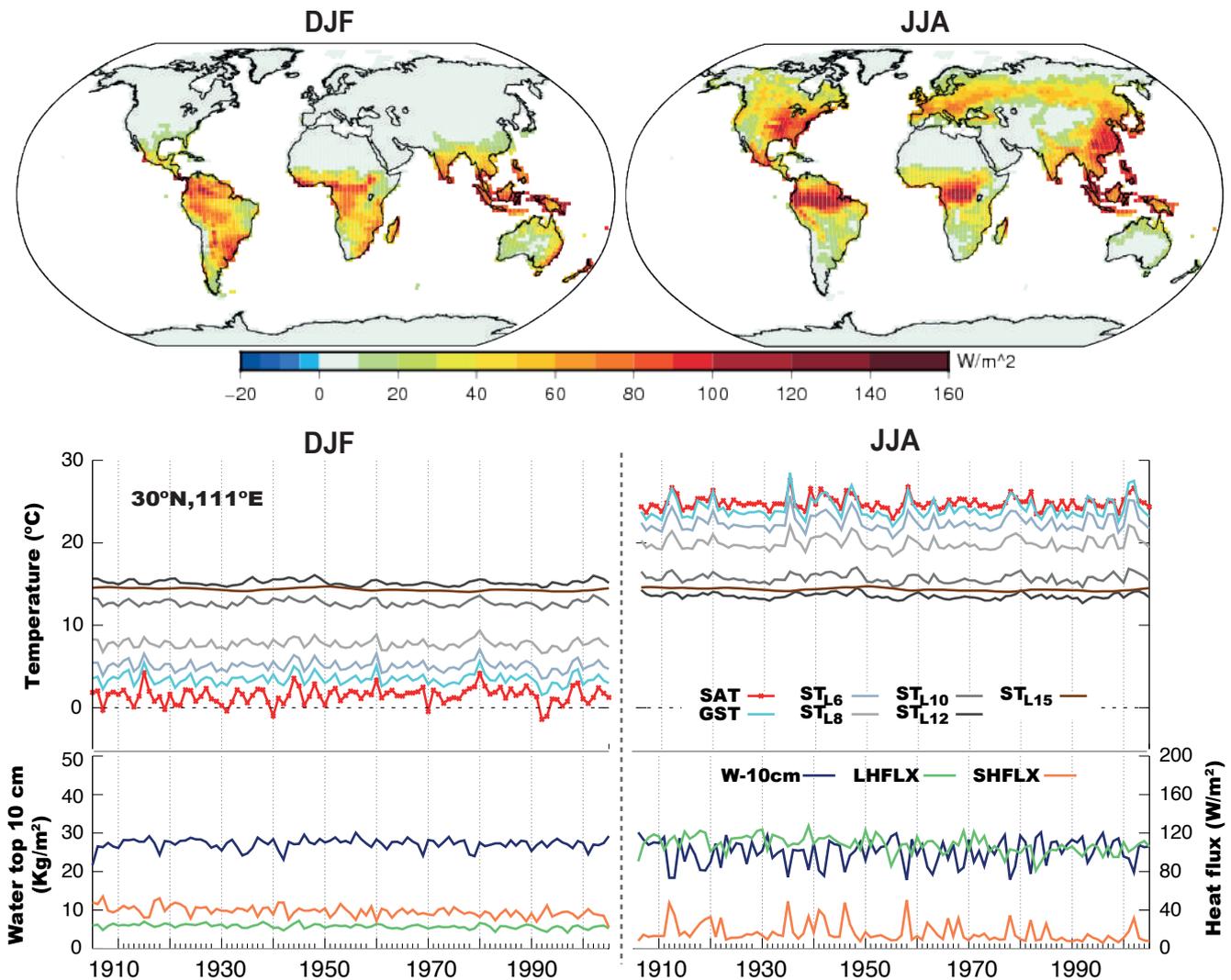


Figure 5. Top. DJF and JJA average surface latent heat fluxes (LHFLX) over the 850-2005 CE period. Middle. DJF and JJA time evolution of SAT, GST, ST_{L6} , ST_{L8} , ST_{L10} , ST_{L12} and ST_{L15} during the period 1900-2005 CE for a grid-point (southeastern China; $30^{\circ}N$, $111^{\circ}E$) from the ALL-F₂ simulation (Table 3). Bottom. Same as middle for SHFLX, LHFLX and water content in the top 10 cm of the soil (W-10cm).

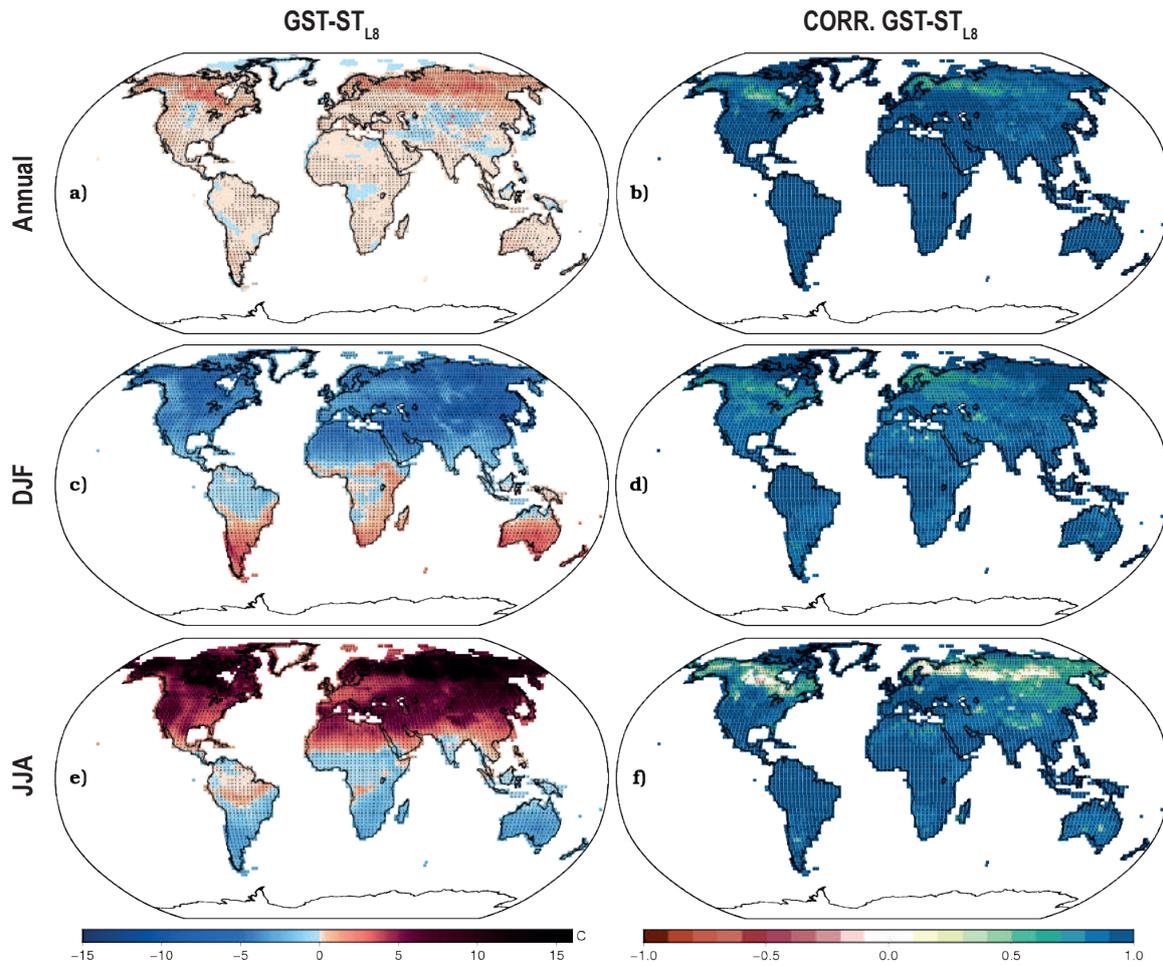


Figure 6. Left. (a) Annual, (c) DJF, (e) JJA spatial distribution of GST minus ST_{L8} differences during the 850-2005 CE period. Right. (b) Annual, (d) DJF, (f) JJA spatial distribution of the correlation coefficients between GST and ST_{L8} for the same period. Black dots indicate that 80% of the members within the ALL-F ensemble deliver statistical significance ($p < 0.05$).

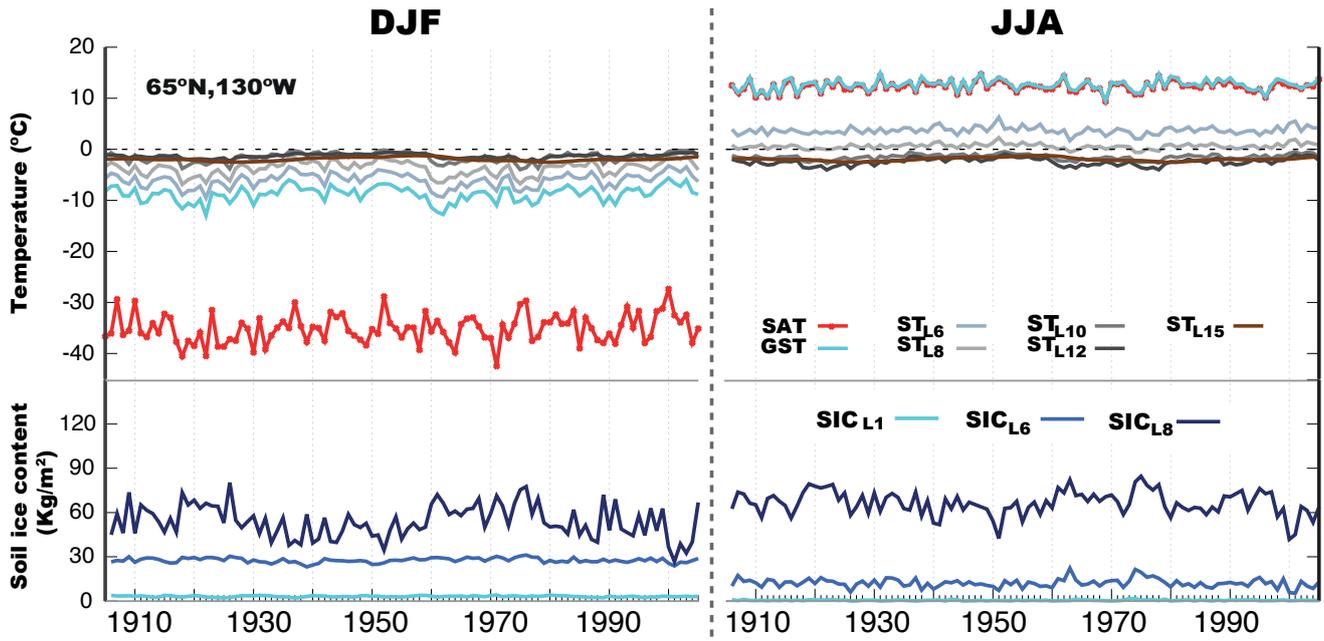


Figure 7. Top. DJF and JJA evolution of SAT, GST, ST_{L6} , ST_{L8} , ST_{L10} , ST_{L12} and ST_{L15} during the period 1900-2005 CE for a grid-point located at the north of Canada (65°N , 130°W) from the ALL-F₂ simulation (Table 3), an area with permanent frozen soils. Dashed lines indicate 0°C . Bottom. Soil ice mass content (SIC) in the L1, L6 and L8 soil layers.

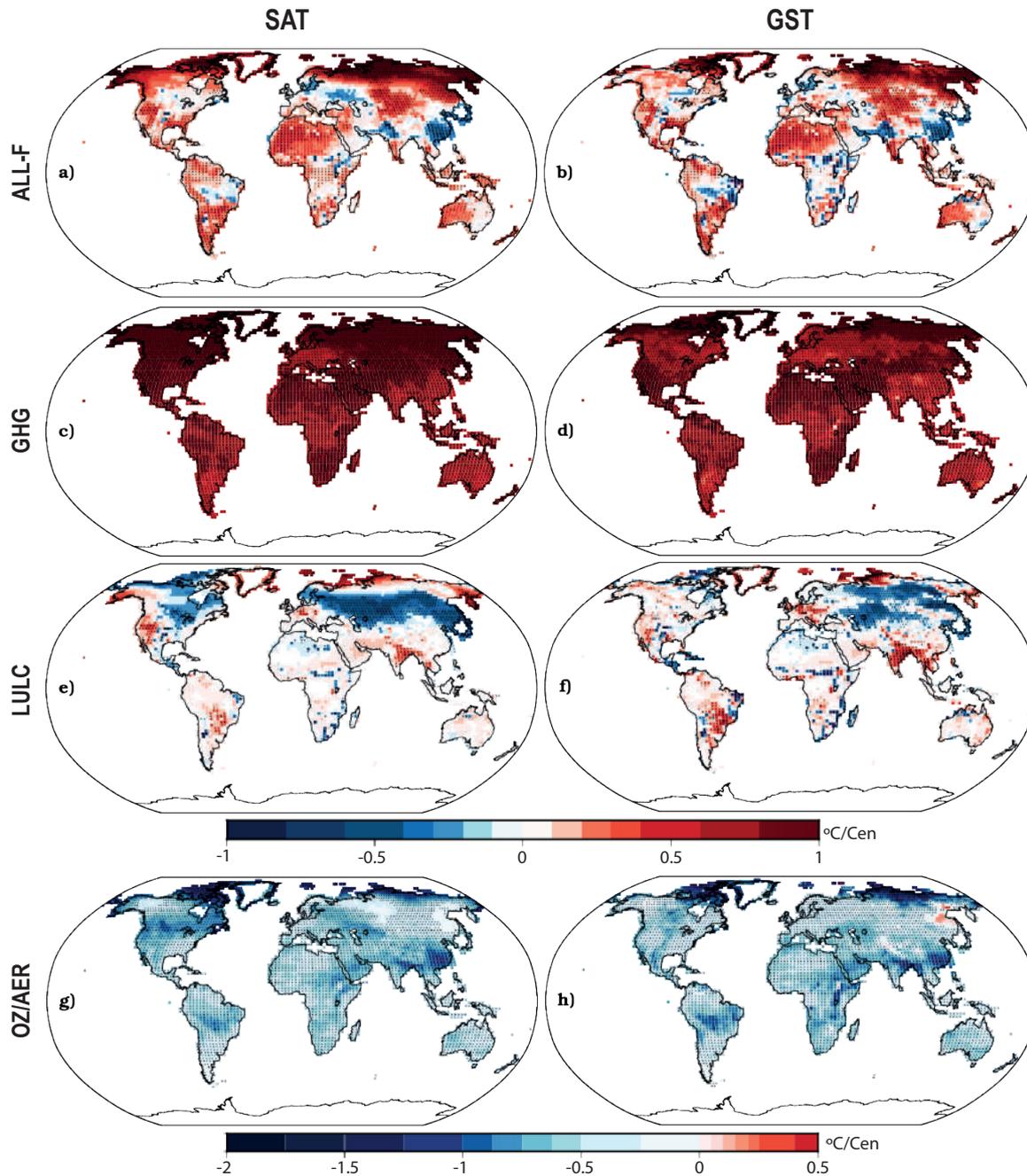


Figure 8. Spatial distribution of the linear trends in the industrial period for the ALL-F (a,b), GHG-only (c,d), LULC-only (e,f) and OZ/AER-only (g,h) ensembles for SAT (left) and GST (right). Trends are indicated in Celsius/century. Dots indicate agreement in 80% of the ensemble members in delivering significant trends at a grid-point in the case of the ALL-F ensemble. For the GHG-only and LULC-only ensembles, dots indicate that at least two of the three ensemble members agree in delivering significant trends for a grid-point whereas for the OZ/AER-only dots indicate that both of the ensemble members deliver significant trends. Note the different scale for OZ/AER-only

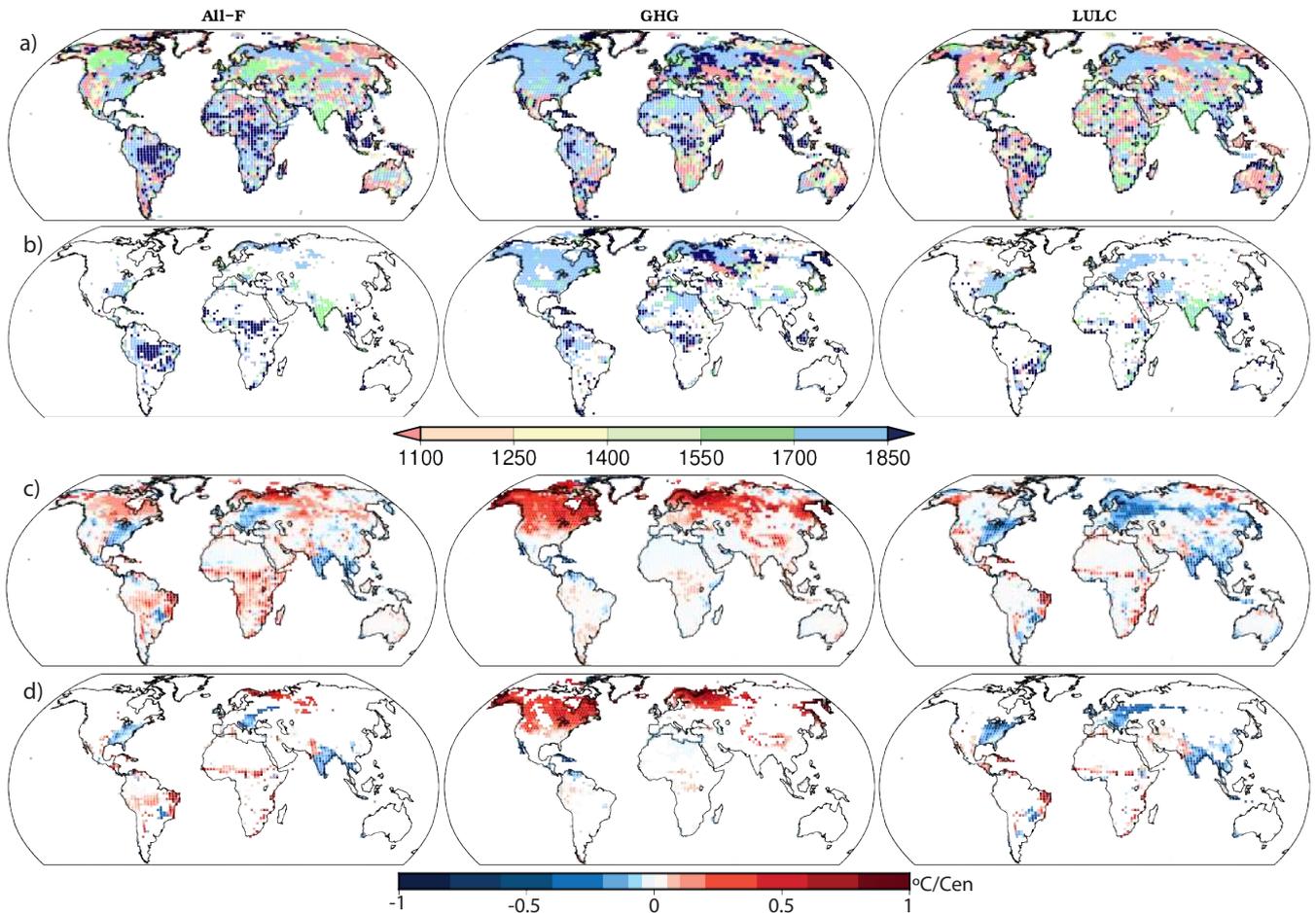


Figure 9. Spatial distribution of the two-phase regression results for SAT minus GST in the ALL-F₂, LULC₁ and GHG₁ ensembles as examples: (a) dates (years) of change; and (c) trends after the year-of-change. (b,d) are the same as (a,c) but only for areas where at least 80% of the ensemble members show significant changes. Trends before the year-of-change (not shown) are not significant in any of the ensembles. Significance ($p < 0.05$) is obtained based on a F-test (T-test) for year-of-change (trends) following Solow (1987). Significance also accounts for autocorrelation (Trenberth, 1984). Note that the spatial window has been modified to enhance visualization of land areas. Note that the spatial window has been modified to enhanced visualization over land areas.

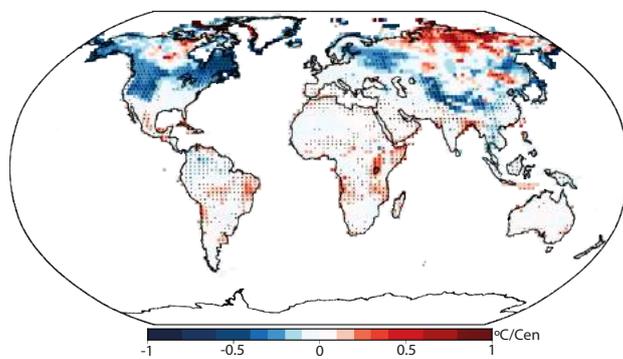


Figure 10. Spatial distribution of the linear trends for SAT minus GST in the OZ/AER-only ensemble. Trends are indicated in Celsius/century. Dots indicate agreement in both of the ensemble members

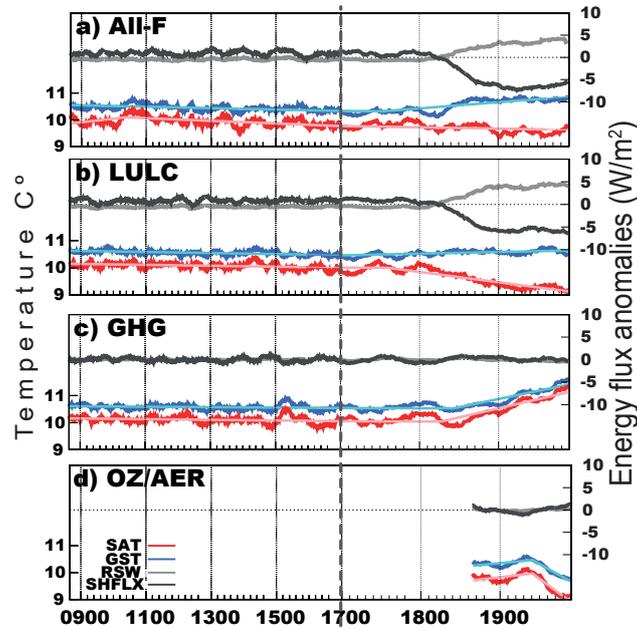


Figure 11. LM evolution of SAT, GST, reflected shortwave radiation (RSW) and surface sensible heat flux (SHFLX) for a grid-point located at 40 °N, 82 °W at the south of the Great Lakes, U.S. in the ALL-F₂, LULC₁, GHG₁ and OZ/AER₁ simulations (Table 3). The left axis correspond to SAT and GST while the right axis to the energy fluxes at the surface. Note that for the energy fluxes the anomalies with respect to 850-2005 CE are shown whereas for temperature absolute values are presented. All series are 31-yr moving average filter outputs. For SAT and GST the result of the two-phase regression model is displayed with thin solid lines. Note the change in timing of the x axis after 1700 CE.

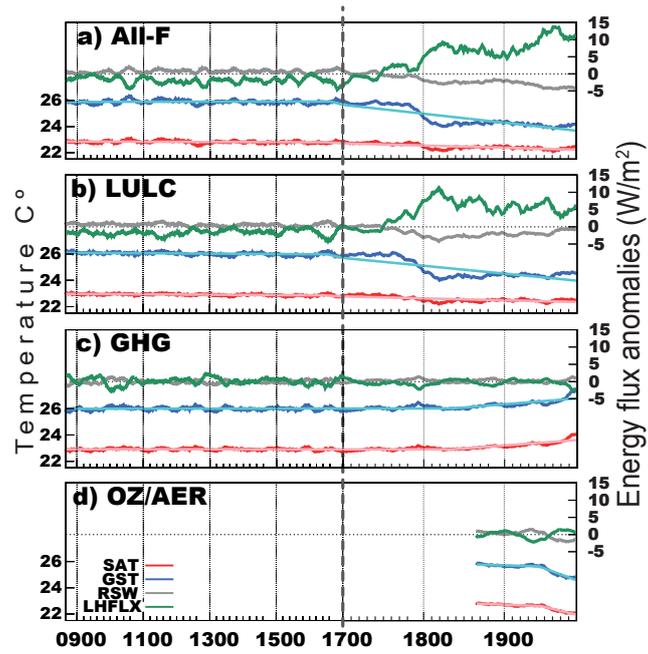


Figure 12. Same as in Fig. 11 but for a grid-point located at 12 °S, 40 °W at the northeast of Brazil and LHFLX is represented instead of SHFLX.

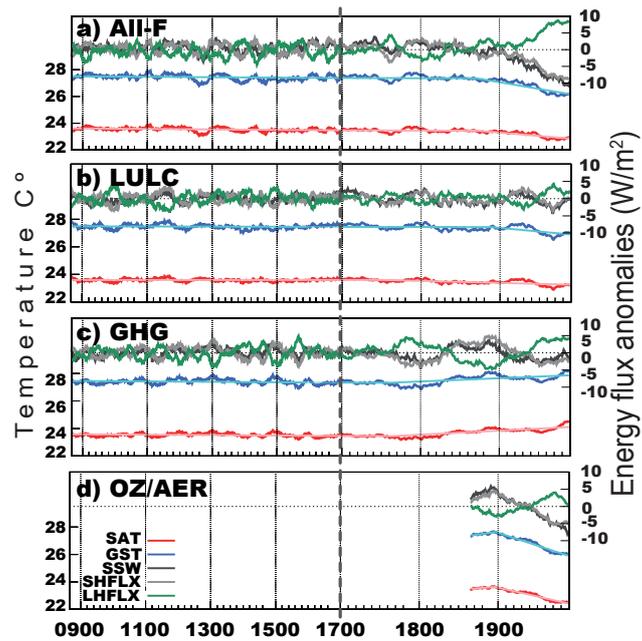


Figure 13. Same as in Fig. 11 but for a grid-point located at 2.5 °N, 33 °E over Uganda. Note that incoming short wave radiation at the surface (SSW) is represented instead of RSW and both LHFLX and SHFLX are shown.

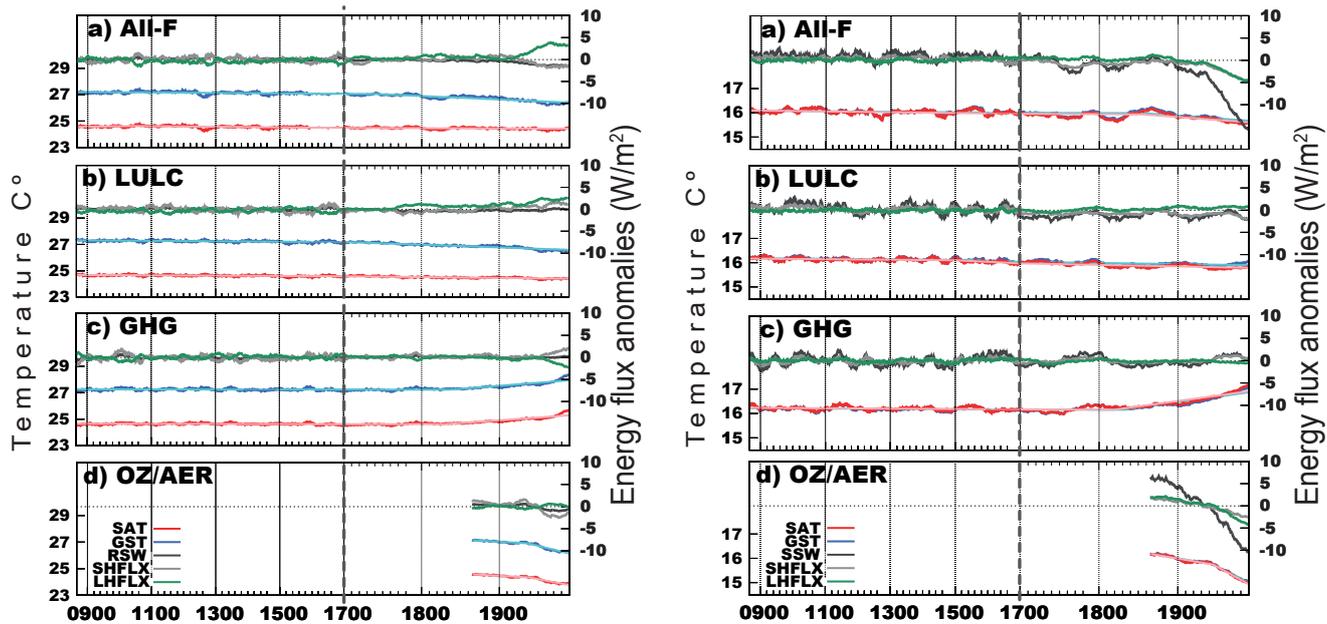


Figure 14. Same as in Fig. 11 but for two different regions: northeast of Brazil between 1 °S - 11 °S and 47 °W - 35 °W (left) and southeast of China between 22 °N - 32 °N and 103 °E - 122 °E (right). For the area of northeast Brazil RSW is represented whereas for the area of southeast China SSW is shown instead. LHFLX and SHFLX are shown in both cases.

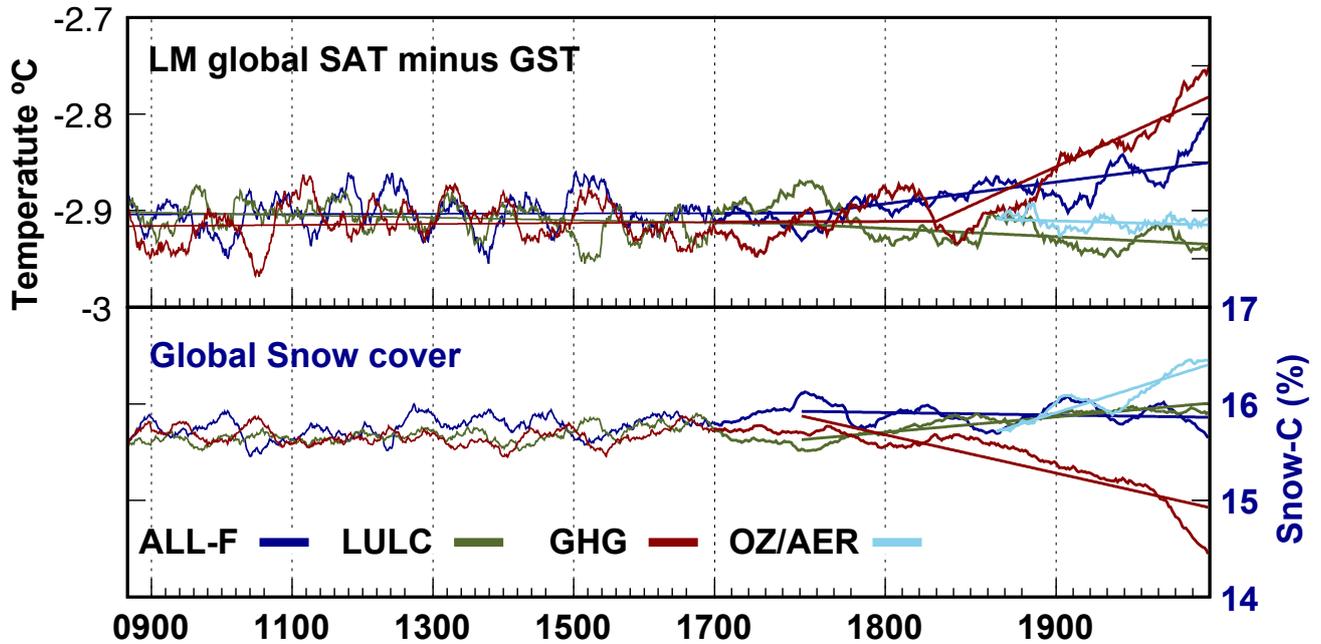


Figure 15. Top. LM evolution of the global SAT-GST offset for the ALL-F₂, LULC₁, GHG₁ and OZ/AER₁ simulations (Table 3). Straight lines indicate the long-term trend during the LM from a two-phase regression analysis except for the OZ/AER₁ that indicate linear trends in the industrial period. Bottom panel illustrates the LM evolution of global snow cover percentage for the same ensemble members. Straight lines indicate the long-term trend within the 1750-2005 period except in the case of OZ/AER that covers only the 1850-2005 period. The period from 1750 to 2005 was selected in order to match the timespan when the SAT-GST offset experiences the variation in the ALL-F₂ simulation. All series are 31-yr moving average filter outputs. Note the change in timing of the x axis after 1700 CE.