Comments to the Author:

Thank you for your comprehensive responses to the reviewers. the reviews were generally rather positive, but did suggest quite a few changes, many of which you have agreed to make. I would like to pass your revised manuscript by at least one of the reviewers again just to make sure they are content, and for that reason I have classed the revisions as "major", although the ones you propose should be adequate.

Dear Editor,

We would like to thank you for editing our manuscript. We have updated it according to the comments of the reviewers as suggested in our responses. You can find below the point by point response to each reviewer's comment as well as the final version of our manuscript with track changes.

On behalf of all co-authors,

Zhiqiang Lyu

Answer to referee 1

The referee's comments are shown in black and our answers in blue:

The article provides the first systematic model-data comparison based on borehole temperature-depth profiles in Antarctica. They elaborate two techniques (depth and time domains) to compare these profiles and their reconstructions from four sites with climate model output. They conclude by outlining some useful metrics for future model data comparison and highlight the importance of internal variability on the observed tendencies.

We would like to thank the reviewer for the careful evaluation of our work and the very useful comments that will be addressed in the revised version.

Specific comments

1. L51: "Since the variable measured in the borehole is the temperature itself,. . ." In most cases, resistivity is measured, which is easily converted to temperature. Is this true of your measurement techniques?

   Yes, we agree with the referee that resistivity is generally measured in the field, using thermistor, and then it is converted to temperature by Steinhart-Hart equation. But this is true of every modern sensor (an electronic signal is converted to a meaningful variable). We propose to replace the following sentence:

   "Since the variable measured in the borehole is the temperature itself,. . ."

   by :

   "The most significant advantage of borehole paleothermometry is that temperature is directly measured with a thermistor calibrated in the laboratory. Thus, the calibration is independent of the climate at the measurement site."

2. L53: "the surface temperature history makes the reconstruction mathematically undetermined." The equation of temperature at depth usually results in a system of linear equations which is mathematically under and overdetermined. Varying mathematical inversion techniques are then utilized to reconstruct the ground surface temperature history. Please clarify. This should also be clarified on L.267.

   Yes, we propose to replace the following sentence:

   "the surface temperature history makes the reconstruction mathematically undetermined."

   by:

   "Nevertheless, the characteristics of heat conduction that blurs the surface temperature history make the reconstruction mathematically undetermined: several temperature histories can result in the same borehole temperature profile, because diffusion will smooth out high frequency temperature variations. Consequently, the temperature history cannot be determined unequivocally."

   Line 267 reads "However, as stated above, borehole temperature reconstructions are "underdetermined", which means that there are many possible temperature histories that can fit the data." The word "underdetermined" will be between quote with an implicit reference to the explanation given above in line 53.

3. L.92: “Previous studies using forward models driven by climate model outputs were focused on ground temperature and not to borehole. . .” Please provide a couple examples (references) here.

   According to referee’s suggestion, we will add some references, Beltrami et al., 2005; García-García et al., 2016; González-Rouco et al., 2003, 2006.

4. In Section 2.1, the borehole measurements and reconstructions are briefly explained. Since they come from four different publications, how can differences in inversion/reconstruction techniques affect the results presented in Figure 1?

   The reviewer raises an important point but comparing the different inversion/reconstruction techniques is out of the scope of our study. The temperature reconstructions are sensitive to the technique used. Notably, because the problem is underdetermined, several temperature histories are equally probable, and the final result will depend on some parameters used to calculate the inversion. We can illustrate this for instance by driving the borehole temperature model selected in this study by the published reconstructed temperature history and compare it to the observed borehole temperature. Difference have been found that are likely attributed to the different methodology and hypothesis but they are relatively small, suggesting that they do not have a major impact on our conclusions. Nevertheless, a more substantial analyses would be required to formally prove this.
   We will add in the revised version a cautionary note mentioning the potential influence of the application of those different techniques.

5. L. 136: “CESM1-CAM5 and MPI-ESM-P are not continuous in 1850.” What is meant by this? Please clarify.

   This will be clarified:

   “CESM1-CAM5 and MPI-ESM-P simulations do not cover the entire millennium. Historical simulations covering 1851–2005 C.E. were launched independently of simulations covering 850–1850 C.E. (referred to as the past1000 experiment in CMIP/PMIP nomenclature). In order to obtain results over the full millennium, we adopt the approach from Klein and Goosse (2018) and merge the first ensemble members (r1i1p1) of the past1000 experiment with the corresponding ensemble members of the historical experiment. Although not continuous, there is no large discrepancy in 1850 C.E. between the two merged simulations (e.g., Klein and Goosse, 2018).”

6. In Section 2.2, for all models, excluding the CESM ensemble, which realization (r1i1p1) is used? Is it the only realization available? If not, why was this one selected? Please clarify.

   Yes, until now only one simulation for CCSM4, GISSE2-R, IPSL-CM5A-LR, MPI-ESM-P and BCC-CSM1-1 is publicly available. This will be clarified in the revised version of the manuscript:

   “For CESM1, an ensemble of simulations is available, providing an estimate of the internal variability as simulated by this model, but for CCSM4, GISSE2-R, IPSL-CM5A-LR, MPI-ESM-P and BCC-CSM1-1, there is only one simulation available.”

7. L. 160: “: : :for Mill Island, the heat flux is set to zero: : :” How realistic is this? Furthermore, this is a different technique than the other sites. Could this influence the results? If the heat flux is set to zero, how is the steady-state temperature calculated?

In the case of Mill Island, the hole is shallow (120 m), but the ice sheet is very deep at the site. At sites with such a deep ice sheet, and with a high accumulation rate, the conditions at the base are not impacting more than roughly the bottom 1000 m, so it is perfectly reasonable to model the top of the ice sheet only, with a zero heat flux at the bottom. The validity of this assumptions is discussed in detail in the original paper. Here is a quote: "The optimal surface temperature history was found to be essentially independent of the location of this bottom boundary condition for depths in excess of 180 m below the surface" (Roberts et al., 2013). The steady state temperature is calculated in the same way as the other models, and the only difference is the bottom boundary condition. This will be specified in the revised version.

8. L. 161: "For WAIS, a vertical step of 1 m for the upper 500 m and up to 25 m for the deepest part, and for other sites where the depth of borehole is close or less than 500 m, the step is set to 1 m for overall depth." Why are various techniques used again? What is the benefit of this?

   WAIS is the only very deep borehole, and we use a coarser model resolution for the deepest part to save some computer time as in Orsi et al 2012. This is not required for the shallower cores for which the computation time is lower and it is the reason why we keep a fine resolution for all the depth of the core. This will be specified in the revised version.

9. L. 183: "At WAIS-Divide, the spread of the sensitivity tests is lower than the spread if the different scenarios." What is meant by scenarios? Is it the different models being analyzed?

   The different scenarios mean the different simulated borehole profiles driven by different climate model results. This will be clarified in the revised version of the manuscript:

   "At WAIS-Divide, the spread of the sensitivity tests is lower than the spread in the simulated borehole profiles driven by different climate model results (solid lines in color in Figure 2 (a) and (b))."

10. L.196: ": : :but the deviation in the top 100 m show that there is climate information stored in the upper part of the profile, and that this profile cannot be fully determined by boundary conditions." Climate is not the sole reason why the top 100 m would show deviation. How can you be sure it is climatic information?

    We totally agree that the surface temperature change is not the sole reason why the top 100 m would show deviation. For instance, we can find that the initial and basal temperature have some impacts on the shape of simulated borehole temperature in the top 100 m shown in the Figure 2 (e). Meanwhile, we expect that some climate information is stored in the top 100 m from the comparison between the simulated borehole temperature profiles (solid lines in the Figure 2) driven by different GCMs with the stationary temperature profile (thick dash-dot line). This paragraph will be rephrased in the revised version of the manuscript:

    "At Styx, the boundary conditions are adjusted to reproduce the slope of the temperature profile in the deeper part (100-200 m). Compared with stationary temperature profile, the simulated borehole profiles driven by GCMs (solid lines in the Figure 4 (e)) show a deviation in the top 100 m, which suggests that there is climate information stored in the upper part of the profile. Meanwhile, at the depth shallower than 50m, the effect of boundary conditions is weaker than the differences in the temperature histories from the different model, which means the borehole temperature data can be used to discriminate between temperature histories provided by the different models."

11. At the start of the paragraph at L.198, it is stated that internal climate variability and the different characteristics of the climate models are the main sources of differences. The results from the CESM ensemble have not been discussed in this section. To strengthen this point, I recommend adding in a discussion of it.

As suggested by the reviewer, a discussion of the internal variability in the CESM ensemble will be added at the start of the paraph at L.198 in the revised version:

"The internal variability also has significant impact on the shape of the simulated borehole profiles. At these four sites, the range of simulations driven by CESM ensemble is much larger than range of the different sensitivity tests in the top of 50 m (shown as the shaded area in Figure 4 b, d, f, h), which conforms that the dominant source of uncertainty in a model–data comparison, at least in the top 50 m, is from the internal variability."

Furthermore, the statement that internal climate variability and different characteristics of the climate models being the main source of differences does not hold true for Mill Island. In Figure 1, only different depths of the zero heat flux are considered. More tests must be added to conclude the importance of the influence of internal variability and different model characteristics to the differences at this site.

Yes, in Figure 2(e), only the upper 50 m shows that there is a noticeable spread between the colored lines, illustrating that different climate model scenarios result in different temperature profiles, and that this difference is larger than the spread between the dashed lines (the sensitivity to model parameters). We will clarify it and include more sensitivity tests for Mill Island, as requested in the revised version.

12. In Figure 3, why are different smoothing techniques used? Can it influence the results?

The reason why we use different smoothing is to facilitate a comparison between the reconstruction and climate model results. The reconstruction provides a smoothed history of the past surface temperature changes, but the smoothing itself depends on the time and the characteristics of the site. We have tried to mimic this as much as possible by using variable smoothing in the plot. As the reconstructions at WAIS and Styx preserve mainly the centennial and multi-centennial variabilities, we applied longer smoothing (50-year) to the climate model result at WAIS and Styx. Similarly, at Larissa and Mill Island, the reconstructions show the multi-decadal and decadal variabilities, so we choose 10- year smoothing at Larissa and 3-year smoothing at Mill Island.
Using these different smoothing techniques is thus justified and not influencing significantly our conclusions, because our goal here is to perform a visual model-data comparison in the time domain in order to see if the reconstruction is within the range provided by the ensemble. Since the reconstructions have much wider ranges than those ones from the climate model results, the basic compatibility between model and data will not be changed.

13. The reconstructions from the climate models presented in Figure 3 are calculated using what technique? Their errors bounds are also not presented. How does this influence the results? Does the reconstruction from the climate model always lie within the error bounds? Please clarify in the manuscript.

The temperatures displayed in Figure 3 come directly from the surface temperature calculated by the climate model, based on its own dynamics and the forcing applied as discussed in section 2.2. Single time series are available for each model experiment without error bounds but providing an ensemble of experiments gives a range of current state-of-the-art models. This range provides a kind of uncertainty associated with model results but relating this to a precise estimate of the error is unfortunately a complex issues as models are for instance not independent of each other, sharing similar parametrizations, and may have common biases, due in particular to the relatively coarse resolution of climate models (Abramowitz et al., 2019;Knutti et al., 2017;Sanderson et al., 2015).

14. L.215: "In order to remove the bias on the mean state for each climate model, anomalies are shown using the total period covered by each reconstruction as reference." Which figure is being referred to? The paragraph starts discussing Figure 3 but these are not anomalies.

Here, we show an example to explain the methodology. In Figure R1, the original climate model result is shown as the red curve. Its mean over the period 850-2000 C.E. is different from the reconstruction. To remove this bias, we applied a very simple bias correction to climate model results, ensuring that after the adjustment the climate models have the same mean over the reference period as the reconstruction (Yellow curve in the Figure R1). This will be clarified in the revised version of the manuscript as follows:

"In order to ensure that the climate model results have the same mean over the reference period as the reconstruction, we applied a very simple, constant correction to remove the mean bias of the climate model results as shown on the Figure 3."

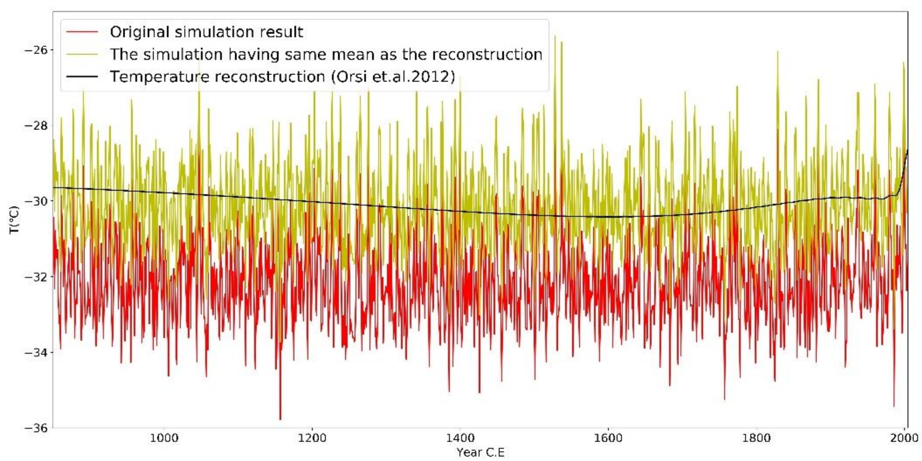

Figure R1. Comparison between reconstructed surface temperature series at WAIS and the climate model outputs at the grid cell-point closest to WAIS.

15. Since the temperature variability increases as you go back in time, there is less confidence with respect to the timing of events. Timing of events varies within climate models. Could this further explain any discrepancies of the timing of events? Please discuss.

The timing of the events differs indeed between the simulations if those events are related to internal variability and not caused by a specific forcing. This influence of internal variability can be estimated from the difference between the CESM members as discussed in Section 4. In addition, from the Figure R2, the range of CESM members does not increase back in time. This suggests that the temperature variability between the different members of CESM, and thus the associated uncertainty, does not change a lot over the time.

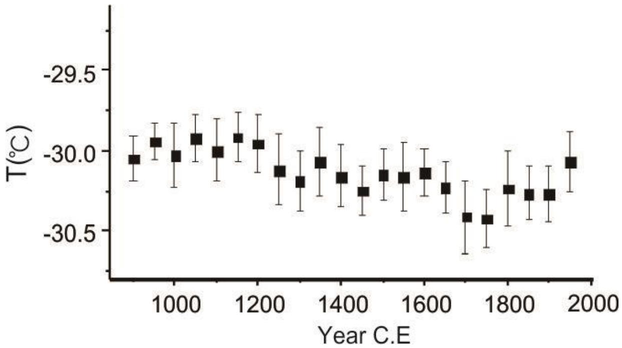

Figure R2. Temperature variability of the CESM ensemble at the grid cell of WAIS. The black square represents the mean of the CESM ensemble in the corresponding time. Their error bound present the 1 standard deviation (1σ) ranges of the CESM ensemble.

16. What causes the decrease in temperature at 1980 and 2000 in Styx and Larissa (Figure 3)? Is it climatic in origin or an artifact of the reconstruction technique?

    The four borehole reconstructions in the manuscript are from the original papers. In the papers related to the Styx (Yang et al., 2018) and Larissa (Zagorodnov et al., 2012), the authors have shown that the reconstructions from borehole are consistent with the weather stations, and ice core isotope-derived records. Consequently, the decrease in temperature in 1980 C.E. and 2000 C.E. at Styx and Larissa likely reflect climatic signals.

17. L.390: "Fig. 8 shows the spatial correlation in the Antarctica Peninsula (AP)." Do you mean the spatial correlation of the gridcells? Please clarify

    As suggested, we rewrite this sentence as follows:

    "Figure 8 shows the spatial correlation between the temperature from 1825 to 1925 C.E. at Larissa and other grids cells for each climate model."

18. L390-393: "Despite the correlation coefficient decreasing as the grid getting far away from the Larissa, the values, at least around Larissa for each model, are higher than 0.6, showing that this metric is representative of the whole peninsula region, and not extremely site-specific." A correlation coefficient of 0.6 means that it only explains 36% of the variance. How can you conclude that it is representative of the entire peninsula?

    Yes, we totally agree with the referee that a correlation coefficient of 0.6 is not that high and our view was probably a bit too optimistic. We have changed this sentence following the suggestion:

    "Despite the correlation coefficient decreasing with the distance from the Larissa, the values, at least around Larissa for each model, are higher than 0.6, showing that this metric is representative of part of the AP region, and not extremely site-specific."

19. Why are the CESM ensemble members not presented in Figure 8? How is this metric influenced by internal variability?

    Figure 8 shows the correlation between the temperature from 1825 to 1925 C.E at Larissa and other grids cells in the Whole Antarctica for each climate model. Since there are no significant differences between each member in CESM ensemble, we just show one member (CESM1) as an example in the manuscript.

This will be mentioned explicitly in the revised version of the manuscript and we will add the figure shown below (Figure S1) in the supplement:

"As there are no significant differences between each member in CESM emsemble (see in the Figure S1), only one member CESM1 and other GCMs are present in the Figure 8."

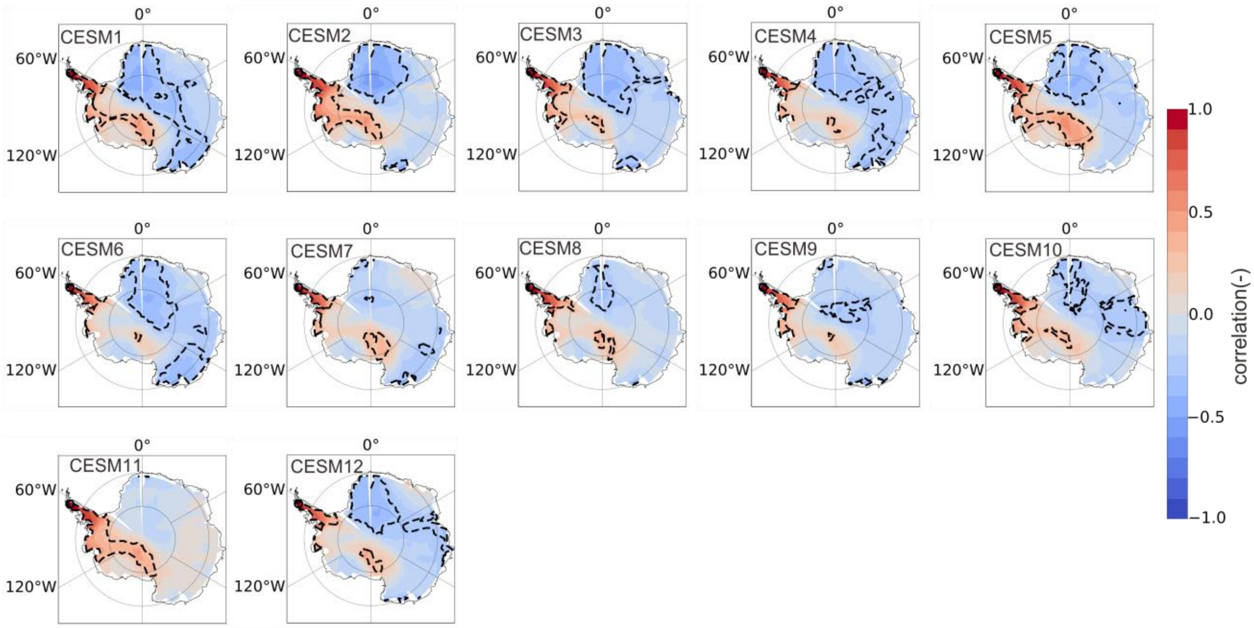

Figure S1. The correlation map (blue-red shading area) showing the relationship between the temperature from 1825 C.E. to 1925 C.E at Larissa and other grid cells in Antarctica for each CESM member. The black dotted contour lines show a significant correlation at the 99 % significant level.

20. L.426-428: "A model that responds clearly to the Ozone forcing, and has a strong SAM signature should exhibit this dipole pattern, and it is interesting that some models do not show it, indicating that the Ozone forcing is not dominating over internal variability." The CESM ensemble members are not seen in Figure 10 and 11 nor discussed. How can this be concluded?

    We agree with the reviewer that we jump a bit too quickly on the conclusions and a dedicated analysis should be performed to prove this. In particular, the correlation pattern can be also strongly influenced by the spatial response of each model to the ozone forcing itself. To avoid a long discussion on a point not central to our analysis, we have preferred to remove the part of the sentence relating to the respective role of ozone forcing and internal variability.

21. Borehole temperature profiles and their ground surface temperature histories are compared with those from climate models. They ability of the climate models to reconstruct the ground surface temperature was evaluated and three distinct metrics were created. From all of this, how do you think climate models could improve? From your analyses, what are their areas of weaknesses? It would be beneficial to add a section outlining this to the conclusions.

    We agree with the reviewer that this is a very interesting issue. Our goal was to provide a test to estimate the performance of climate models. This is of course the basis for model improvements but the link between a bias in one diagnostic and a model improvement is not straightforward and can be very different for different models. Making specific suggestions would require many additional diagnostics, comparisons and tests. Without that, we fear that any additional material we could add on this subject would be too general or speculative to be really informative for the reader. This is the reason we do not plan to develop this in the revised version.

Technical Points: There are many grammatical errors throughout the article impeding the reader's comprehension. They are not all outlined below but should be addressed in the revised manuscript.

Thank you for noticing all the following errors that will be corrected. We will check all the details in the grammar and improve it in the revised manuscript.

1. L36: Please define acronym AP

   Thanks, we will add the definition of AP (Antarctic Peninsula).

2. It would facilitate comprehension if the depths of each borehole were added to Table 1. This would help explain the various time periods for the reconstructions found in Figure 1.

   We will add a column for depth in the Table 1.

3. Please add the units of elevation to the map in Figure 1.

   We will add them.

4. In Figure 2, in the boxes below Figures 2 a,b and 2c,d, there is a typo in the word accumulation. The thermal diffusion used along with its units should be included. In the caption, " 2) sensitivity tests using the temperature history of once CESM member. . .", do you mean one CESM member? Also "The shade area represents the simulated subsurface temperature ensemble driven by CESM" should read "The shaded area. . ."

   As suggested, we will add the unit for the thermal diffusivity.
   We will modify "once CESM member" to "one CESM member" and also correct the phraseology of "shade" to shaded.

5. L.192: "At Mill island,. . ." Should read Mill Island to be consistent throughout the text.

   As suggested, we will correct it.

6. L.192: "..the ice thickness is much deeper. . ." Ice thickness cannot be deeper. Should read thicker.

   This will be modified.

7. In Figure 4, the y-axis of 4a,e and f are crowded. Either decrease the amount to y-ticks or increase the figure size. Some of the symbols, in particular the yellow triangle of CCSM4, are difficult to see. I would recommend increase the size of the markers for the climate models and the reconstruction. Also, the labels on 4c and d are cut off by the below figures. Please fix.

   We will fix these problems in the Figure 4, and it will be updated with a clearer figure in the revised version.

8. L.254: "Larissa shows a temperature minimum in 1940's. . ." should read . . .1940s.

   This will be modified as suggested.

9. L.270 a period is missing at the end of the sentence.

This will be modified as suggested.

10. In Figure 5, please use a different colour for the observations.

    We will modify the Figure 5 as suggested.

11. For consistency, use CCSM or CCSM4

    As suggested, we will check and replace CCSM by CCSM4.

12. For the techniques/metrics elaborated in Section 4, please be consistent with the use of grid, grid-point, and gridcell. Since you are comparing with data from the gridcell, I'd recommend the use of that word to facilitate the reader's comprehension.

    Thanks for your recommendation. We have replaced grid and grid-point by grid cell.

13. L.372: "For most of the models,. . ." It would be best to include a number or percentage of models to really illustrate your point.

    As suggested, we will add a percentage in the corresponding sentence : "75% models show WAIS displays a larger cooling from 1000 to 1600 C.E. than other locations in Antarctica (shown in blue) but with magnitude similar to other grid cells in West Antarctica.".

14. In Figures 6,9,12, please add the units to the colour bar as these are surface temperatures tendencies

    We will add the units to the colour bar in Figure 6, 9, 12.

15. In Figures 6,7,9, the circle illustrating the location of the observations is not clear. Maybe make it bolder or another colour.

    Done.

16. L.394: "Figure 9 shows the same temperature trend (1825-1925) for all models." Do you mean surface temperature since Figure 9 shows varying trends.

    Thanks for your remark. The previous sentence has been modified: no model is able to capture the observed temperature trend from 1825 C.E. to 1925 C.E..

17. L.395-396: "A majority of the CESM members(CESM1, 7, 8, and 9). . ." Do you mean minority? 4/12 is not a majority.

    Yes, we made a mistake here. Thank you for that. We propose to replace "A majority of the CESM members(CESM1, 7, 8, and 9)" by "Only four member of CESM (CESM1, 7, 8, and 9) show a cooling trend over AP, but the magnitudes of them are still less than the observed one".

18. L.403: There is a typo in the word overestimation.

Corrected.

19. In Figure 8, the dotted contour line is not clear to the reader. Also, indicate that the colour bar represents the correlation coefficient.

    Done.

20. L.427 please define acronym SAM.

    We will add the definition of SAM (Southern Annular Mode).

21. Figure 10, some of the numbers in the colour bar appear to be cut-off. The red dashed line is not visible to the reader. Please correct.

    We will fix these problems in the Figure 10, and it will be updated with a clearer figure.

22. Figure 11, the y-label of d is overlapping with c. Please clarify that it is the linear trends of surface temperature in the caption.

    We will fix these problems in the Figure 11.

23. Figure 12 is not referenced in the text

    Yes, we made a mistake here. We will remove it.

Reference:

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing, Earth Syst. Dynam., https://10, 91-105, 10.5194/esd-10-91-2019, 2019.

Klein, F. and Goosse, H.: Reconstructing East African rainfall and Indian Ocean sea surface temperatures over the last centuries using data assimilation, Climate Dynamics, 50, 3909–3929, https://doi.org/10.1007/s00382-017-3853-0, 2018.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V.: A climate model projection weighting scheme accounting for performance and interdependence, Geophys. Res. Lett., https://10.1002/2016gl072012, 2017.

Orsi, A. J., Cornuelle, B. D. and Severinghaus, J. P.: Little Ice Age cold interval in West Antarctica: Evidence from borehole temperature at the West Antarctic Ice Sheet (WAIS) Divide, Geophys Res Lett, 39(9), 1–7, doi:10.1029/2012GL051260, 2012.

Sanderson, B. M., Knutti, R., and Caldwell, P.: A Representative Democracy to Reduce Interdependency in a Multimodel Ensemble, J. Clim., 28, 5171-5194, https://10.1175/jcli-d-14-00362.1, 2015.

Yang, J. W., Han, Y., Orsi, A. J., Kim, S. J., Han, H., Ryu, Y., Jang, Y., Moon, J., Choi, T., Hur, S. Do and Ahn, J.: Surface Temperature in Twentieth Century at the Styx Glacier, Northern Victoria Land, Antarctica, From Borehole Thermometry, Geophys Res Lett, 45(18), 9834–9842, doi:10.1029/2018GL078770, 2018.

Zagorodnov, V., Nagornov, O., Scambos, T. A., Muto, A., Mosley-Thompson, E., Pettit, E. C. and Tyuflin, S.: Borehole temperatures reveal details of 20 th century warming at Bruce Plateau, Antarctic Peninsula, Cryosphere, 6(3), 675–686, doi:10.5194/tc-6-675-2012, 2012.

Answer to referee 2

The referee's comments are shown in black and our answers in blue :

The paper compares temperature observations from 4 Antarctic boreholes with climate model surface temperatures over the last millennium or so. The standard approach to do this is to reconstruct the temperature record from the borehole temperature using a model and compare it with climatic models. The main difficulty is that the thermal diffusivity of ice damps the temperature variations with time and details of the signal are lost: The farther back in time or deeper in the borehole we go, more details are lost. To help the analysis, the authors suggest comparing borehole temperatures with simulated borehole temperatures driven by the climate models using a thermal model. The authors identify a set of key features in the temperature records from the data: Cooling at WAIS over the last millennium, nineteenth century cooling at Antarctica Peninsula and Antarctic warming over the last 50 years or so. Interestingly, the existing climate models do not reproduce these results. They propose to use these key observations as a metric to test the next generation of climate models. The paper is in general clear and well written. I have some suggestions for the authors below but I can anticipate that any paper that encourages climate modelers to use data has my full support.

We would like to thank the reviewer for the positive evaluation of our manuscript and the very useful comments that will be addressed in the revised version as detailed below.

General comments

The borehole distribution is scarce. I know it will always be but I wonder how representative these 4 borehole records are. Inspecting Figure 1, I miss data in the interior of East Antarctica, perhaps Dronning Maud Land; and the coastal area of West Antarctica, Amundsen and Weddle Seas. My view is that a few more sites could improve considerably the benchmark for models. A detailed description of the climate models, thermal model and borehole temperature data is in other papers. This is understandable but a few short descriptions here and there will improve the clarity of the manuscript considerably. This is of particular importance as the methods used in the manuscript are taping on different scientific areas. To me, for example, Section 2.2 says nothing as I don't know what PMIP3-CMIP5 experiments are, or why they are discontinuous in 1850. I have several suggestions below in the specific comments.

a) How representative these 4 borehole records are?

When we propose the metrics of Antarctic climate for model validation in Section 4, we display the correlation maps showing the relationship between the temperature at each borehole site and other grid cells. The results illustrate that the metrics are able to be representative of a large spatial area, although they are calculated at a specific site.

b) Inspecting Figure 1, I miss data in the interior of East Antarctica, perhaps Dronning Maud Land; and the coastal area of West Antarctica, Amundsen and Weddle Seas. A few more sites could improve considerably the benchmark for models

We totally agree with the reviewer that a few more sites could improve considerably the benchmark for models. However, the sparsity of the dataset forbids us to evaluate the skill of the climate model results in other parts of Antarctica. We will insist on that point in the revised version of the manuscript.

c) A detailed description of the climate models, thermal model and borehole temperature data is in other papers. This is understandable but a few short descriptions here and there will improve the clarity of the manuscript considerably.

Our goal here is to simulate borehole temperature profiles by driving the forward model with the results from climate models. In order to make our results more robust, we need to consider the uncertainties in the parameters of the forward model. For the boreholes used in the manuscript, the values of the parameters are derived from the original papers describing the data. Those original studies describe the parameters that have the largest impacts on the surface air temperature reconstruction, a reasonable range for those parameters and the associated uncertainties on temperatures.

Detailed information is thus provided in those papers but as reviewer suggested, we will add some sentences in the revised version to further introduce how the forward model work. We propose to replace the following sentence:

"The term on the left side represents the change in heat content and the right terms are the rate of temperature change due to conduction, advection and heat production, respectively."

by:

"In the Equation 1, the term on the left side represents the change in heat content. On the right side, the first term corresponds to the rate of temperature change due to conduction based on the Fourier's law. Ice moving vertically (z-direction) with downward velocity, w , carries a heat flux $\rho c_p w$T across a plane of unit area, oriented perpendicular to z, which is accounted for in the heat transfer by advection shown as the second term. The third term, Q, consists of two part: (1) ice deformation (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.30), (2) firn compaction (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.33)."

Besides, we will also suppress the introduction of the parametrization of the forward model in the manuscript, and add the more specific details of the forward model as the supplementary in the revised version the as followed:

"According to the original publications, we applied different methods to fit the density data for each borehole in the model. For WAIS and Styx, the density profiles, $\rho(z)$, were obtained by a quadratic fit to measured bulk density data following Severinghaus et al. (2010). For Larissa, the density profile was approximated following Salamatin (2000). For Mill Island, due to the similarity between the density profiles at Mill Island and Law Dome (van Ommen et al., 1999), the fitting to the density data is described by a piecewise exponential plus linear or dual exponential according to the analysis on the Law Dome ice core density profile (van Ommen et al., 1999). The density is considered to be in a steady state.

For the other parameters in the forward model, the specific heat capacity c_p is calculated by c_p = 152.5+7.122T (J kg^{-1} K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.1, T means the temperature). The thermal conductivity in ice is taken from $K_{ice} = 9.828 \exp{(-5.7 \times 10^{-3} T)}$ (Wm^{-1}K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.2), and the thermal conductivity of the firn is calculated by Schwerdtfeger formula (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.4). The vertical velocity at the surface is simply the accumulation rate and decreases with depth as the integral of the densification process (compaction) and the strain due to ice flow divergence. The vertical velocity profile is determined by the method of Alley et al. (1990) and Cuffey et al. (1994) with a constant strain rate. For the accumulation rate, we use a constant value derived from their original publication, which is specified in the Table 3 of the main text. The bottom boundary condition is the basal heat flux and basal temperature. The heat flux is determined by matching the slope of the temperature increase in the bottom section of the record. At Mill Island, this was not possible, because the data do not extend very deep with respect to the total ice thickness. A zero heat flux boundary condition was chosen instead. The validity of this hypothesis is demonstrated in the original study of Roberts et al. (2013). The basal temperature is determined using the lower "undisturbed" sections of the measured borehole temperature extrapolated to the bottom.

In order to save computation time, the vertical discretization of the model is not homogenous. For WAIS, which is the only very deep borehole, a vertical step of 1 m for the upper 500 m and up to 25 m for the deepest part, and for other sites where the depth of borehole is close or less than 500 m, the step is set to 1 m for the whole depth.

Before the forward model is driven by the climate model results, it is initialized with a stationary profile, which is generated after a 20000-year model run with a constant climate history and a realistic seasonal cycle. Seasonal-scale variations are "undetectable below a depth of 20m" (Cuffey and Paterson, 2010), and its does not change throughout the run. At WAIS and Styx, the seasonal cycles are determined from weather station data; at Larissa and Mill Island, since the original studies do not give the seasonal cycle, we use a seasonal cycle amplitude of 10 °C similar to WAIS (Eq. S1). At WAIS, it includes a periodic function with annual and semi-annual components, fitted to 3 years of weather station data from WAIS Divide and Byrd station (AMRC, SSEC, UW-Madison) as follows (Orsi et al., 2012):

$$T(t) = 10(\cos(2\pi t) + 0.3\cos(4\pi t)) \text{ (in K)} \qquad \text{(S1)}$$

At Styx, the seasonal cycle is determined by fitting a sinusoidal function to the automated weather station data as follows (Yang et al., 2018):

$$T(t) = 10(\cos(2\pi t) + 0.35\cos(4\pi t)) \text{ (in K)} \qquad \text{(S2)}$$

Where t is time, T is the temperature.

Equations S1 and S2 for WAIS and STYX are nearly identical, so we presume the seasonal cycle is also similar at Larissa and Mill Island, where no seasonal data is available. Including a seasonal cycle wave is important because the heat capacity and thermal conductivity depend on temperature, and temperature changes a lot in the top 15m, but below that, it is of negligible effect.

"

Specific Comments

Title: This is minor point but I think that title is very specific and not easy to digest. What about something like 'Comparing temperature reconstructions from climate models with observed borehole temperature in Antarctica over the last millennium'?

The title is currently "Comparison of observed borehole temperatures in Antarctica with simulations using a forward model driven by climate model outputs covering the past millennium". We consider that one of the main originality of our study is to use a forward model so we prefer to keep this information it in the title.

L12-13 In this paper there are two types of 'models': climate and temperature models. I found data-model confusing here as often papers will compare borehole temperature with modelled temperature. The novelty of this paper is that is comparing 'climate models' temperature with observations. Figure 1. The Temperature vs depth plots don't show the full temperature profile, from surface to bedrock. I assume that the authors are only showing the fraction for the depth that affects the time of interests in the study. This should be made clear.

We use indeed two types of model and this may be confusing. For the revised version, we will check each time 'model' is used and ensure that the meaning is clear.

There is no full temperature profile from surface to bedrock in the Figure 1 because the borehole temperature measurements were made on shallow boreholes that did not always extend through the entire ice sheet. This will be specified in the revised version. Since it is a model-data comparison, we adjust the plot to the depth where there is data. Although it is the fraction of the total depth of the ice sheet, it is adequate to be used to reconstruct surface air temperature, as done in the literature, and to compare with the simulated borehole temperature profiles obtained by the forward model driven by climate model results at each site.

L129 The Tikhonov regularization is a regularization not an inversion method. It doesn't make sense to compare it with the least squares algorithm in Orsi et al 2012.

Yes, we agree that the Tikhonov regularization belongs to the larger family of least squares regressions. On Figure 1, we reproduce the published datasets and temperature reconstructions. We cannot avoid the fact that records were published with different methods. We don't intend to compare the methods, but we will add in the revised version a cautionary note mentioning the potential influence of the application of those different techniques.

Section 2.2 I am not a climate modeller, I simply don't understand this paragraph. What are all these acronyms? What is PMIP3-CMIP5 and why are you using the output? What is the discontinuity in 1850? A gentler introduction to the models used in the paper would be welcomed for CP readers

We are sorry that we forgot to explicit the acronyms. The full names of PMIP3 and CMIP5 are the third phase of the Past Model Intercomparison Project (PMIP3; Otto-Bliesner et al., 2009) and the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012). The Paleoclimate Modelling Intercomparison Project (PMIP) is a long - standing initiative that provides coordinated paleoclimate modeling and data collection activities to facilitate advances on the study of the mechanisms of climate change (Otto-Bliesner et al., 2009). The fifth phase of the Coupled Model Intercomparison Project (CMIP5) produced a state-of-the- art multimodel dataset designed to advance our knowledge of climate variability and climate change (Taylor et al., 2012). CMIP and PMIP are major sources of information for the assessment reports of the Intergovernmental Panel on Climate Change (IPCC).

This has been clarified as follows:

"The simulated surface air temperature used in this study is extracted from climate model simulations covering the past millennium performed in the framework of the third phase of the Past Model Intercomparison Project (PMIP3; Otto-Bliesner et al., 2009) and the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012)."

"CESM1-CAM5 and MPI-ESM-P simulations do not cover the entire millennium. Historical simulations covering 1851–2005 C.E. were launched independently of simulations covering 850 - 1850 C.E. (referred to as the past1000 experiment in CMIP/PMIP nomenclature). In order to obtain results over the full millennium, we adopt the approach from Klein and Goosse (2018) and merge the first ensemble members (r1i1p1) of the past1000 experiment with the corresponding ensemble members of the historical experiment. Although not continuous, there is no large discrepancy in 1850 C.E. between the two merged simulations (e.g., Klein and Goosse, 2018)."

Section 2.2. Do these models provide surface mass balance as well as temperature? Has the surface accumulation provided by the models been compared with the one observed and used in the temperature model?

Yes, we can obtain precipitation and sublimation/evaporation from models, to calculate the surface mass balance (SMB) as done for instance in Dalaiden et al (2020). Nevertheless, we did not use simulated SMB here because our focus is on temperature changes and we did not want that biases in the simulation of SMB influence our conclusions. Consequently, we use the original observed accumulation rate instead of the simulated one. A precise evaluation of the accuracy of the SMB in the climate models and of its impact is the topic of a paper in itself. We will add in the revised version a note mentioning the point as suggested by the reviewer:

"In addition, although we can obtain the simulated surface mass balance (SMB) from these models(e.g. Dalaiden et al., 2020), we do not use it here and keep the observed accumulation rate in the forward model since biases in the simulation of SMB may affect our conclusions and the focus here is on the simulated temperature evolution.

Equation 1 I may have missed this but I can't find a description of what is the vertical velocity that the authors are using. I imagine is connected to the surface accumulation but how? How does it vary with depth?

Yes, we agree with the reviewer that the vertical velocity is connected to the surface accumulation. Vertical velocity depends on the accumulation, the densification process (compaction), and finally the strain due to ice flow divergence. Vertical velocity at the surface is simply the accumulation rate, and it decreases to zero at the bottom, or a constant value equal to the melt rate if there is melting. The detailed vertical velocity profiles for the boreholes are shown in the papers describing the original data. We will add a sentence describing the vertical velocity parametrization in the revised paper.

L156 In addition to explaining how accumulation is used in the model, does it vary with time?

No, we use a constant accumulation rate derived from their original publication, which is specified in the Table 3 of the main text, because the model we use has a constant density profile, and only uses the accumulation rate in the calculation of the vertical velocity. Additionally, sensitivity studies made for WAIS-Divide (Orsi et al., 2012) and Styx (Yang et al., 2018) show that the variations of accumulation are small enough over the period considered that it does not appreciably change the results. We will add a sentence explaining this in the revised paper.

L158 The authors are working with shallow temperature, most likely in the firn area. I would like more explanation about how heat capacity and diffusivity depend of density and if density is assumed constant with time.

The specific heat capacity c_p is calculated by $c_p = 152.5 + 7.122T$ (J kg^{-1} K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.1, T is the temperature). The thermal conductivity in ice is taken from $K_{ice} = 9.828 \exp\left(-5.7 \times 10^{-3} T\right)$ (Wm^{-1}K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.2), and the thermal conductivity of the firn is calculated by Schwerdtfeger formula (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.4). The density profile is considered to be in steady state.

L158 Heating term in a heat equation is not specific enough. I assume that the authors refer to the internal or strain heating due to flow deformation. How is that calculated? I don't have access to Cuffey and Paterson but I assume that the term depends on the strain-rates. What components are the authors considering? I am assuming that the term is small but this point requires clarification.

The heating term Q consists of two part: (1) ice deformation (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.30), (2) firn compaction (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.33). We will add a section in the supplementary information with these details about the temperature diffusion model.

L160 The reasons to apply null heat gradient at Mill Island and explained later and this is confusing. I suggest a clear paragraph describing boundary conditions for Equation 1.

We suggest rephrasing this section to:

"The bottom boundary condition includes the basal heat flux and basal temperature. The heat flux is determined by matching the slope of the temperature increase in the bottom section of the record. At Mill Island, this was not possible, because the data do not extend very deep with respect to the total ice thickness. A zero heat flux boundary condition was chosen instead. The validity of this hypothesis is demonstrated in the original study of Roberts et al. (2013). The basal temperature is determined using the lower "undisturbed" sections of the measured borehole temperature extrapolated to the bottom."

L165-166 How recent is the 'recent annual average'? How does it compare with timesteps?

The "recent annual average" means the mean temperature that could be derived from weather station data, as described in the rest of the sentence. For instance, at WAIS, it is the average of 3 years of weather station data from WAIS Divide and Byrd station (AMRC, SSEC, UW-Madison). It is larger than the time step which is 200 time steps per year. We will clarify this point in the revised version.

Equation (2). Is 't' the time in years?

Yes, you are right, we will add the description of all the symbols paragraph in the revised version of the manuscript.

L179 It is not clear to me what this means. Is that 10 % variation of boundary and initial conditions? I am assuming that in Larissa the temperature gradient refers to the sensitivity to the bottom boundary condition. Why not in Styx or Mill Island, are they not also frozen to the bed with Neumman boundary conditions? Why some of the sites study more parameters than others?

(a) Is that 10 % variation of boundary and initial conditions?

For the boundary and initial conditions, we followed the tests proposed in the original papers. The 10% sensitivity is only applied for the thermal diffusivity ("*0.9" in the Figure 2a and 2b). Following Orsi et al (2012), we used the Schwerdtfeger formula (Cuffey and Patterson, 2010, Chapter 9), which depends on both temperature and density of the snow. It usually gives an upper estimate of the thermal diffusivity of snow. This is the reason why we decreased the thermal conductivity by 10%, and ran the optimization of the temperature again. Compared with the effect of the initial temperature on the shape of simulated borehole temperature, the thermal conductivity and accumulation rate have no significant effect on the result (Figure R1), which is in good agreement with the result in Orsi et al. (2012). Consequently, we prefer to remove the curves of sensitivity tests of thermal conductivity and accumulation rate in the revised version. This will also provide a simpler and clearer description of the remaining sensitivity experiments.

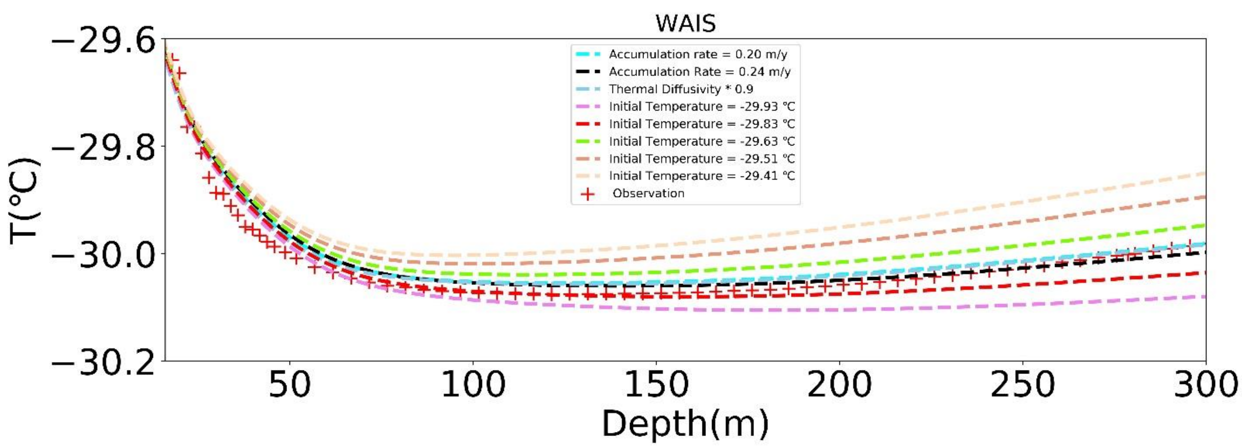

Figure R1. The sensitivity tests using the temperature history of one CESM member (dashed lines) at WAIS.

(b) I am assuming that in Larissa the temperature gradient refers to the sensitivity to the bottom boundary condition. Why not in Styx or Mill Island, are they not also frozen to the bed with Neumman boundary conditions? Why some of the sites study more parameters than others?

According to the original papers, the various parameters in the forward model have effects of different magnitude on the results. For instance, at Styx (Yang et al., 2018) and Larissa (Zagorodnov et al., 2012), the bottom temperature has significant influence while the bottom boundary conditions are of limited importance at WAIS (Orsi et al., 2012). Consequently, we test for some sites more parameters than for others. Nevertheless, in the revised version, we will add some sensitivity tests for the initial temperature for the Mill Island, and for the other sites, we will remove those curves which are of limited importance in order to make the Figure 2 clearer. We propose to replace the following sentence:

"In order to assess the uncertainty in the model-data comparison related to the parameters of the forward model, it is necessary to perform a series of sensitivity experiments as shown on Figure 2. We made different tests for the key parameters using the values proposed in the original publications (Table 1) and following the protocol of Orsi et al. (2012)."

by:

"According to the original papers, the various parameters in the forward model have effects of different magnitude on the results. Consequently, in order to assess the uncertainty in the model-data comparison related to the parameters of the forward model, we perform a series of sensitivity experiments on the parameters which have the largest effects on the simulated borehole profiles shown in the Figure 2."

L183 'if' should be 'in'

Corrected.

L264-266. I don't understand this paragraph. What is internal variability or a profound disagreement?

(a) What is internal variability?

Internal variability refers to the climate variability due to process internal to the climate system, in contrast to the forced variability that is a response to forcings like change in insolation, greenhouse gases, etc. Models that run for hundreds of years are not expected to reproduce the timing of the observed variations due to internal variability, in particular the exact phase of internal oscillations, such as the North Atlantic Oscillation, or ENSO. As a result, the same model, with the same physics and the same input forcings can produce different temperature when it is started with slightly different conditions. This is why "ensembles" are run: they are the outputs of the same model, with the same forcings, but slightly different initial conditions. If the results from different ensemble members are different, then this difference is attributable to internal variability rather than to the response to external forcings. By comparing the different temperature histories from the different CESM ensemble members, we can quantify the amplitude of internal variability. In many cases, the model-data difference for the other models is within this range of internal variability deduced from the CESM ensemble. As a result, we hypothesize that the discrepancy between an individual simulation with one model and data is likely due to the poor sampling of internal variability by only one simulation (reality is only one realization among all the possibilities), rather than a problem with model physics or inappropriate forcing factors. This potential role of internal variability will be explained in more detail by rewriting section which was in the submitted version at L264-L266.

(b) What is a profound disagreement?

The lines L264-L266 summarize the main source of model-data disagreement over the 20th century. For Styx (Figure 4(f)) and Mill Island (Figure 4(e)), the discrepancy between all the models and data is larger than the spread of the CESM ensemble. This suggests that the model-data difference cannot be simply attributed to uncertainties associated with the low number of ensemble members, but rather, with a systematic bias, which could come from model physics or input forcings.

This will be clarified in the revised version of the manuscript:

"Overall, for WAIS (Figure 4(b)) and Larissa (Figure 4(d)), the reconstructed trends lie in the CESM ensemble range, suggesting many apparent model disagreements for those sites can be due to internal variability. For Styx (Figure 4(f)) and Mill Island (Figure 4(e)), the reconstructed trends are larger than the spread of the CESM ensemble, which means the disagreements are not only due to internal climate variability but are related to a systematic climate model bias in this region."

Figure 6. I can't see the circles in most of the figures. Perhaps that is good but I would suggest a selection of figures, so that they are bigger or add an edge to the circle.

We will fix the problem in the Figure 6, and it will be updated with a clearer figure.

Reference

Cuffey, K., and W. Paterson, The Physics of Glaciers, Academic, Amsterdam, 2010.

Orsi, A. J., Cornuelle, B. D. and Severinghaus, J. P.: Little Ice Age cold interval in West Antarctica: Evidence from borehole temperature at the West Antarctic Ice Sheet (WAIS) Divide, Geophys Res Lett, 39(9), 1–7, doi:10.1029/2012GL051260, 2012.

Otto-Bliesner, B. L., Joussaume, S., Braconnot, P., Harrison, S. P., and Abe-Ouchi, A.: Modeling and Data Syntheses of Past Climates: Paleoclimate Modelling Intercomparison Project Phase II Workshop, Estes Park, Colorado, 15– 19 September 2008, Eos T. Am. Geophys. Un., 90, p. 93, https://doi.org/10.1029/2009EO110013, 2009

Roberts, J. L., Moy, A. D., Van Ommen, T. D., Curran, M. A. J., Worby, A. P., Goodwin, I. D. and Inoue, M.: Borehole temperatures reveal a changed energy budget at Mill Island, East Antarctica, over recent decades, Cryosphere, 7(1), 263–273, doi:10.5194/tc-7-263-2013, 2013.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Yang, J. W., Han, Y., Orsi, A. J., Kim, S. J., Han, H., Ryu, Y., Jang, Y., Moon, J., Choi, T., Hur, S. Do and Ahn, J.: Surface Temperature in Twentieth Century at the Styx Glacier, Northern Victoria Land, Antarctica, From Borehole Thermometry, Geophys Res Lett, 45(18), 9834–9842, doi:10.1029/2018GL078770, 2018.

Zagorodnov, V., Nagornov, O., Scambos, T. A., Muto, A., Mosley-Thompson, E., Pettit, E. C. and Tyuflin, S.: Borehole temperatures reveal details of 20 th century warming at Bruce Plateau, Antarctic Peninsula, Cryosphere, 6(3), 675–686, doi:10.5194/tc-6-675-2012, 2012.

Klein, F. and Goosse, H.: Reconstructing East African rainfall and Indian Ocean sea surface temperatures over the last centuries using data assimilation, Clim. Dyn., 50, 3909–3929, https://doi.org/10.1007/s00382-017-3853-0, 2018.

Comparison of observed borehole temperatures in Antarctica with simulations using a forward model driven by climate model outputs covering the past millennium

Zhiqiang Lyu[1], Anais J. Orsi[2], Hugues Goosse[1]

[1]Université catholique de Louvain (UCLouvain), Earth and Life Institute (ELI), Georges Lemaître Centre for Earth and Climate Research (TECLIM), Place Louis Pasteur, B-1348 Louvain-la-Neuve, Belgium
[2.]Laboratoire des Sciences du Climat et de l'Environnement (IPSL/CEA-CNRS-UVSQ UMR 8212), CEA Saclay, 91191 Gif-sur-Yvette CEDEX, France

Correspondence: *Zhiqiang Lyu (zhiqiang.lyu@student.uclouvain.be)*

Abstract. The reconstructed surface temperature series from boreholes in Antarctica have significantly contributed to our understanding of centennial and multi-decadal temperature changes and thus provide us a good way to evaluate the *ability of* climate model ~~ability~~ to reproduce low-frequency climate variability. However, up to now, there ~~were no~~*has not been any* systematic model-data comparisons based on temperature from boreholes at *a* regional or local scale in Antarctica. Here, we discuss two different ways to perform such a comparison using borehole measurements and the corresponding reconstructions of surface temperature at West Antarctic Ice Sheet (WAIS), Larissa, Mill Island and Styx *Glacier* in Antarctica. The standard approach is to *compare the surface temperature simulated by the climate model at the grid cell closest to each site with the reconstructions in the time domain derived from the borehole temperature observations.* ~~compare climate model outputs at the grid *cell*point closest to each site with the reconstructions in the time domain derived from the direct borehole temperature observations.~~ Although some characteristics of the reconstructions, for instance the non-uniform smoothing, limit to some extent the model-data comparison, several robust features can be evaluated. In addition, a more direct model-data comparison based on the temperature measured in the boreholes is conducted using a forward model that simulates explicitly the subsurface temperature profiles when driven with climate model outputs. *This comparison in the depth domain is generally consistent with observations made in the time domain, but also provides information that cannot easily be inferred from the comparison in the time domain*~~This comparison in the depth domain provides many consistent signals with those in the time domain, but also suggests some information that we cannot extract from the comparison in the time domain.~~ The major results from these comparisons are used to *derive metrics that can be applied for future model-data comparison. We also describe the spatial representativity of the sites chosen for the metrics.* ~~define some metrics derived from the borehole temperature data for future model-data comparison, and demonstrate the spatial representativity of the sites chosen for the metrics~~. The long term cooling trend in West Antarctica from 1000 to 1600 CE (-1.0 °C) is generally reproduced by the models, but often with a weaker amplitude. The 19th century cooling in the Antarctic Peninsula (-0.94°C) is not reproduced by any of the models, which tend to show warming instead. The trend over the last 50 years is generally well reproduced in West Antarctica and at Larissa (Antarctic Peninsula), but overestimated at other sites. The wide range of simulated trends indicates the importance of internal variability on the observed trends, and show*s* the value of model-data comparison to investigate the response to forcings.

1. Introduction

Although most of the world has been steadily warming over the last few decades, the temperature trend in Antarctica is not homogeneous (Jones et al., 2016). Several syntheses relying on instrumental air temperatures records have shown a large recent warming over the

Antarctic Peninsula (AP)- and parts of West Antarctica, but the trend for the other parts of the Antarctic continent remains less clear (Chapman and Walsh, 2007; Nicolas and Bromwich, 2014; Steig et al., 2009; Turner et al., 2005). *It remains difficult to characterize the large interannual to multi-decadal variability at high southern latitudes because instrumental data are sparse, and limited to the last 60 years, at best.* The sparse instrumental data and the series covering generally less than 60 years do not allow to characterize well the large interannual to multi-decadal variability at high southern latitudes. The mechanisms at the origin of recent changes are thus still uncertain (Goosse et al., 2012; Jones et al., 2016; Nicolas and Bromwich, 2014).

Proxy-based reconstructions offer the opportunity to place the recent temperature changes in a longer context. Thanks to their relatively good spatial coverage and their high resolution, the reconstructions based on water stable isotopes derived from ice core have provided important information on temperature variability during past two millennia over Antarctica. They indicate a significant cooling trend during the preindustrial period across all Antarctic regions and confirm the strong spatial heterogeneity of the recent warming (Goosse, 2012; Schneider et al., 2006; Stenni et al., 2017). However, the link between the isotope records and local climate is complicated, and this introduces significant uncertainties in the reconstructions (Stenni et al. 2017, Klein et al., 2019).

Borehole temperature observations provide another opportunity to reconstruct surface temperature and several studies have demonstrated their interest, particularly over Antarctica (i.e. Barrett et al., 2009; Muto et al., 2011; Orsi et al., 2012; Zagorodnov et al., 2012; Roberts et al., 2013; Yang et al., 2018). *The most significant advantage of borehole paleothermometry is that temperature is directly measured with a thermistor calibrated in the laboratory. Therefore, the calibration is independent of the climate at the measurement site.*Since the variable measured in the borehole is the temperature itself, i.e. the variable that is reconstructed, no calibration is required against independent climatologic data such as instrumental data. Nevertheless, the characteristics of heat conduction that blurs the surface temperature history makes the reconstruction mathematically undetermined. *: several temperature histories can result in the same borehole temperature profile, because diffusion will smooth out high frequency temperature variations. Consequently, the temperature history cannot be determined unequivocally.* Several approaches have been proposed to overcome the problem as synthesized in Orsi et al (2012), for instance the Bayesian Reversible Jump Markov chain Monte Carlo (Dahl-Jensen et al., 1999), the generalized least-squares inversion (Muto et al., 2011; Orsi et al., 2012; Yang et al., 2018), and the Tikhonov regularization method (Roberts et al., 2013). By applying these methods, the reconstructed temperature series have presented evidence of the existence of cold conditions corresponding to a Little Ice Age in West Antarctica from 1300 to 1800 CE (Orsi et al., 2012), as well as of a recent warming trend in West Antarctica (Barrett et al., 2009; Orsi et al., 2012; Yang et al., 2018), at some high elevation sites of the East Antarctica (Muto et al., 2011; Roberts et al., 2013) and over the Antarctica Peninsula (Zagorodnov et al., 2012), though the timing and magnitude vary between regions.

The reconstructed temperatures based on isotopic composition have been compared to results of climate models. Most models display a relatively large and homogenous warming over Antarctica since 1850 *CE*, which is inconsistent with the signal inferred from the isotope records (Goosse et al., 2012; Klein et al., 2019; Stenni et al., 2017, Abram et al. 2016). This disagreement may be due to the uncertainties in the reconstructions, or due to the *bias*uncertainties in the climate models that may overestimate the response to greenhouse gas forcing or underestimate the natural climate variability in the region (Jones et al., 2016; Neukom et al., 2018). However, a recent study assessing the link between isotope record from ice cores and regional climate over Antarctica using pseudoproxy and data assimilation experiments has not been able to identify any systematic bias in reconstructions on continental scale temperatures based on δ18O (Klein et al., 2019).

Up to now, there were no systematic model-data comparison for temperature reconstructed from boreholes at regional or local scale in Antarctica. This is, on the one hand, due to the characteristics of the inversion that imposes smoothing on a time window that increases as we go back in time and makes the comparison with the simulated surface temperature difficult (Beltrami et al., 2006; Harris and

Gosnold, 1999). Additionally, some reconstructions have an uncertainty range of the same magnitude as the full variability provided by the climate model results, which seriously limits the interest of data-model comparison.

As some of the difficulties in the comparison between the simulated surface temperature *from climate model results* and *the reconstructions from boreholes*~~the ones reconstructed from borehole~~ come from the inversion procedure, comparing directly the observed profile with the one obtained using a one-dimensional heat advection and diffusion forward model driven with climate model can provide new insight. This approach is an example of the application of Proxy System Models (PSM) that reproduce directly processes responsible for the signal recorded in the archive (Evans et al., 2013). PSMs have been applied recently for several proxies, such as tree ring width ~~(Evans et al., 2013)~~ or water-isotope in ice cores, corals, tree ring cellulose, and speleothem calcite (*Evans et al., 2013;* Dee et al., 2014). The ~~use~~ *application* of climate model outputs to drive a borehole temperature forward model has demonstrated the strong coupling between near-surface air and ground temperature changes over decades to centuries (e.g. Beltrami et al., 2005; García-García et al., 2016; González-Rouco et al., 2003, 2006), and *has also been used to validate climate model outputs* ~~validated climate model outputs~~ (e.g. Beltrami et al., 2006; Stevens et al., 2008).

Nevertheless, using a PSM ~~also~~ introduces some uncertainties that must be taken into account. A critical point for borehole temperature is the potential influence of long-term climate changes, such as glacial to inter-glacial cycles, ~~that~~*which* is difficult to estimate (Orsi et al., 2012, Rath et al., 2012). In addition, the simulated subsurface temperature profiles *in Antarctica* are sensitive to model parameters and inputs, such as snow accumulation, ice thickness, geothermal heat flow and the physical properties of ice or ground, which may have significant uncertainties.

Previous studies using forward models driven by climate model outputs were focused on ground temperature and not ~~t~~o*n* borehole obtained in the ice. Here, we will fill this gap by simulating directly subsurface temperature for the publicly available borehole profiles covering the past centuries in Antarctica, using the one-dimensional heat advection and diffusion forward model of Orsi et al.(2012). Our goal here is to provide a protocol for evaluating the climate model ability to reproduce observed low-frequency (multi-decadal to centennial scale) variability. We will analyze two model-data comparison methods to identify the potential advantage and drawbacks of each approach. The easiest way is ~~in~~ to directly compare the surface temperature reconstructed from the borehole measurements with the surface temperature time series simulated by the climate model at the grid ~~point~~ *cell* closest to each site. The second way is to compare the simulated subsurface borehole temperature with *the observation by driving the forward model with climate model outputs.* ~~the measurement by driving the forward model with climate model outputs.~~ In this case, we analyze the temperature at a fixed time (the one when observation where taken) as a function of depth. For simplicity, we will later refer to those two methods as a comparison in the time domain and depth domain, respectively.

This study is organized as follows. The borehole temperature observations, the climate model results, the forward model and the sensitivity of the results to key parameters of the forward model are briefly described in Section 2. Section 3 presents the comparison of simulated and reconstructed surface air temperatures, and the comparisons of simulated and observed subsurface temperature profiles. Some metrics of Antarctic climate for model validation are proposed and discussed in Section 4. Conclusions are given in Section 5.

2. Data and Methods

2.1 Borehole temperature observations and reconstructed surface temperature

The data used in this study includes measured temperature in four boreholes in Antarctica. We refer to them as 'WAIS', 'Larissa', 'Mill Island', and 'Styx' respectively. Figure 1 and Table 1 provide their locations and corresponding references. The borehole temperature profiles were sampled in January 2008 and January 2009 (WAIS), December-February 2009/10 (Larissa), the summer of 2009/10 (Mill Island), and the summer of 2014/15 (Styx). As shown in Fig. 1 (in red rectangles), the borehole temperature~~s~~ *is*~~are~~ affected by the seasonal cycle in the upper 15 meters (Bodri, et al., 2011, Chap. 1), which is not adequate for the reconstruction of annual mean surface temperature. Consequently, only the depth under 15 meters is used to reconstruct the surface temperature history and to compare with simulated subsurface temperature profiles.

Table 1. Location of the four boreholes. Elevation is in meters above sea level (m a.s.l.). *Depth is in meters (m).*

Region	Referenced Name	Latitude	Longitude	*Depth (m)*	Elevation (m a.s.l)	Reference
West Antarctica	WAIS	79°28′S	112°05′W	*3400*	1766	Orsi et al., 2012
Antarctic Peninsula	Larissa	66°02′S	64°04′W	*447.73*	1975.5	Zagorodnov et al., 2012
East Antarctic	Mill Island	65°33′25.84″S	100°47′11.44″E	*500*	503	Roberts et al., 2013
Western Coast of the Ross Sea	Styx	73°51.10′S	163°41.22″E	*550*	1623	Yang et al., 2018

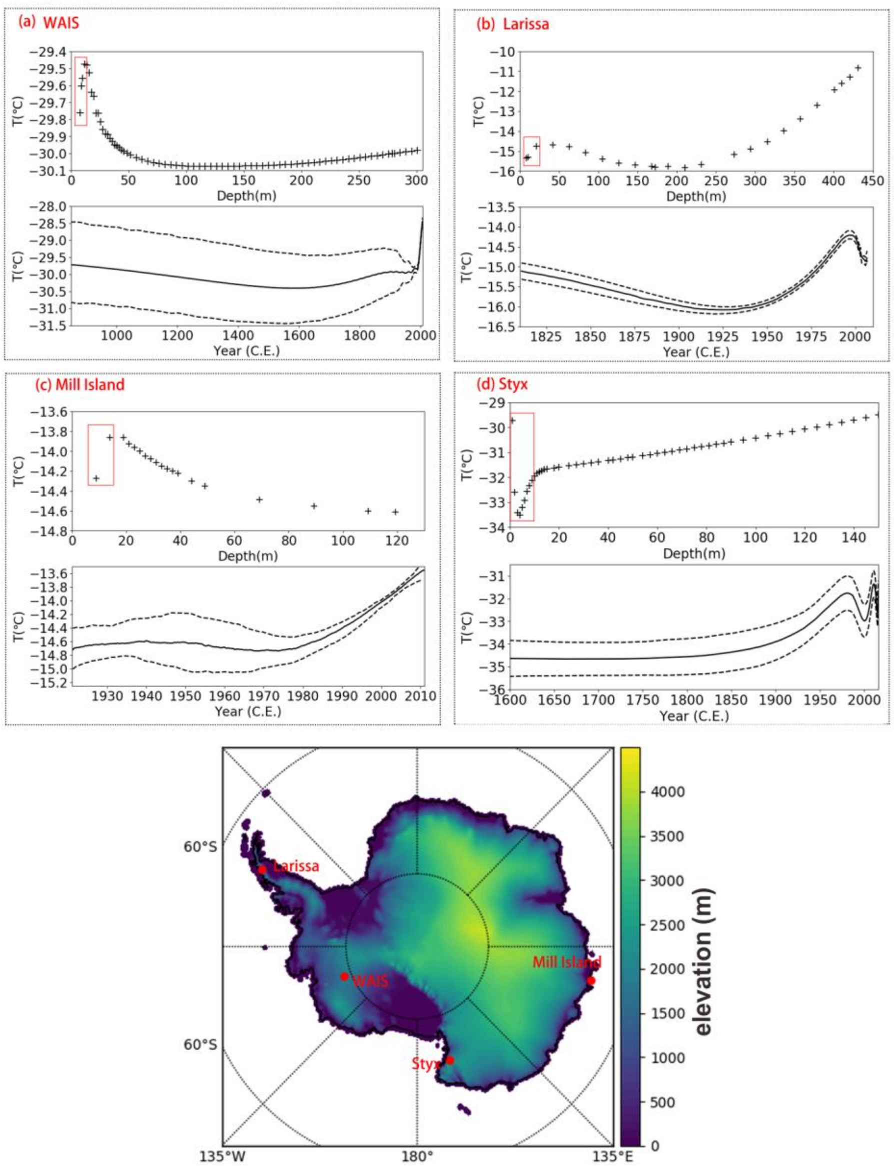

Figure 1. *The observed borehole profiles and corresponding surface temperature reconstructions at the four sites in Antarctica*~~The observed borehole profiles comparision corresponding reconstructed surface temperature and the location at~~ four sites in Antarctica. The symbols (+) show the measured borehole temperature. The dashed lines represent the reconstructed uncertainty and the thick black lines are the mean reconstructed

temperature. In (a), (b), (c), and (d), the red rectangles represent the borehole temperatures that are influenced by the seasonal cycle. The bottom panel shows the location of these four boreholes and their corresponding elevation over Antarctica.

The temperature reconstructions and *their* uncertainty estimates for the four boreholes are shown on Fig. 1. For WAIS, Styx, and Mill island, the reconstructed surface temperature series (Fig. 1 a, c, d) are computed using a generalized least-squares algorithm (e.g. Orsi et al., 2012). For Larissa, the surface temperature is recovered by the Tikhonov regularization algorithm (Zagorodnov et al., 2012). This method has been proved to be valid for inverse problems such as the reconstructions based on borehole temperature observations, and the details of this method are explained in (Nagornov et al., 2001, 2006). *Since the temperature reconstructions are sensitive to the technique used, when we drive the borehole temperature model selected in this study by the published reconstructed temperature histories and compare them to the observed borehole temperature, differences are found. They are likely attributed to the different methodology and hypothesis. However, they are relatively small (Fig. S6), suggesting that they do not have a major impact on the final conclusions.*

2.2 Climate model simulations

The *simulated* surface air temperature used in this study*(Table 2)* is extracted from *general climate model (GCM) simulations covering the past millennium performed in the framework of the third phase of the Past Model Intercomparison Project (PMIP3; Otto-Bliesner et al., 2009) and the fifth phase of the Coupled Model Intercomparison Project (CMIP5; Taylor et al., 2012).*~~PMIP3-CMIP5 experiments (Braconnot et al,2012, http://pmip3.lsce.ipsl.fr/; Taylor et al., 2012, http://cmip-pcmdi. llnl.gov/cmip5/). Table 2 shows the characteristics and the corresponding references.~~ *These simulations cover the period 850-1850 CE (referred to as the past1000 experiment in CMIP/PMIP nomenclature) and the years 1850-2005 CE (historical period). For the majority of the models, the simulations start thus in 850 CE and finishes in 2005 CE. However, for two of the models, CESM1-CAM5 and MPI-ESM-P, the historical simulations covering 1851-2005 CE were performed independently of the simulations covering 850-1850 CE. In order to obtain results over the full millennium, we adopt the approach from Klein and Goosse (2018) and merge the first ensemble members (r1i1p1) of the past1000 experiment with the corresponding ensemble members of the historical experiment. Although not continuous, there is no large discrepancy in 1850 CE between the two merged simulations (e.g., Klein and Goosse, 2018).* ~~These simulations both cover the past 1000 (850-1850 AD) and the historical period (1850-2005 AD). CESM1-CAM5 and MPI-ESM-P are not continuous in 1850. Such discontinuity for the variables employed in 1850 falls within the range of variability of the simulated climate, thus merging it with the historical period have limited effect on the results (Klein et al. 2016).~~

These simulations are driven by natural (orbital, solar irradiance, volcanic) and the anthropogenic (well-mixed greenhouse gases, ozone, aerosols, land use/land cover) forcings (Schmidt et al., 2011, 2012). Note that, BCC-CSM1-1 and IPSL-CM5A-LR ignore the impact of land use/land cover, and IPSL-CM5A-LR does not consider any variations in aerosols and tropospheric ozone. Further description of the simulations and the forcing can be found for instance in Klein et al., (2016). *For CESM1-CAM5, it produces 12 different simulations with the same physics and same input forcings but slightly different initial conditions in the model. Therefore, the differences between ensemble members attributable to the process internal to climate system, provide an estimate of the internal variability. For CCSM4, GISSE2-R, IPSL-CM5A-LR, MPI-ESM-P and BCC-CSM1-1, there is only one simulation available. In addition, although we can obtain the simulated surface mass balance (SMB) from these models (e.g. Dalaiden et al., 2020), we do not use it here and keep the observed accumulation rate in the forward model since biases in the simulation of SMB may affect our conclusions and the focus here is on the simulated temperature evolution.*~~For CESM1, an ensemble of simulations is available, providing an estimate of the internal variability as simulated by this model.~~

Table 2 Climate model simulations used to drive the forward model.

Name	Model resolution (lat × lon)	Number of simulations for 850-1850	Number of simulations for 1850-2005	Reference
CESM1-CAM5	96 × 144	12	12	Otto-Bliesner et al., (2016)
GISS-E2-R	90 × 144	1	1	Schmidt et al., (2014)
IPSL-CM5A-LR	96 × 96	1	1	Dufresne et al. (2013)
MPI-ESM-P	96 × 192	1	1	Stevens et al., (2013)
CCSM4	192 × 288	1	1	Gent et al. (2011)
BCC-CSM1-1	64 × 128	1	1	Wu et al., (2014)

2.3 The Forward Model Description

The forward model used herein to simulate the propagation of the signal coming from the surface temperature history into the subsurface is based on the one-dimensional heat and ice flow equation (Alley and Koci, 1990):

$$\rho c_p \frac{\partial T}{\partial t} = \frac{\partial}{\partial z}\left(k \frac{\partial T}{\partial z}\right) - \rho c_p w \frac{\partial T}{\partial z} + Q \qquad (1)$$

where T is the temperature, t is the time, c_p is the heat capacity, ρ is the density of firn/ice, z is the depth, w is the downward velocity of the firn/ice, Q is the heat production term. *In the Eq. 1,* The term on the left side represents the change in heat content. *On the right side, the first term corresponds to the rate of temperature change due to conduction based on the Fourier's law. Ice moving vertically (z-direction) with downward velocity, w, conveys a heat flux $\rho c_p w T$ across a plane of unit area, oriented perpendicular to z, which is accounted for in the heat transfer by advection shown as the second term.*~~and the right terms are the rate of temperature change due to conduction, advection and heat production, respectively. Important model parameters are summarized in Table 3.~~ *The third term, Q, consists of two part: (1) ice deformation (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.30), (2) firn compaction (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.33). Important model parameters are obtained from the references given in the Table 1, and they are summarized in Table 3. A detailed description of the model is available in the supplement material.*

Table 3 Optimal parameters used to simulate subsurface temperature profile in the forward model driven by the reconstruction for each site: (a) WAIS; (b) Larissa; (c) Mill Island; (d) Styx.

Site	**Surface temperature for steady state (℃)**	**Accumulation rate (m/second)**	**Temperature (T) at bottom (℃)**	**T gradient at bottom (℃/m)**	**Ice thickness (m)**
WAIS	-29.73	6.97×10^{-9}	-4.685	0.0256	3400
Larissa	-16	4.147×10^{-8}	-10.2	-0.04	447.73
Mill Island	-14.6	4.53×10^{-8}	-14.6	0	500
Styx	-32.5	2.6985×10^{-9}	-20.5	0.022	550

~~In the model, the density profile, ice thickness and accumulation rates are derived from onsite measurements according to the descriptions in the original studies while some parameters, such as heat capacity c_p, thermal diffusivity k and heating term Q, are obtained using classical formulations (Cuffey and Paterson, 2010, Chap. 9). The basal temperature and heat flux for WAIS, Larissa, and Styx are determined using the lower "undisturbed" sections of the measured borehole temperature extrapolated to the bottom, and for Mill Island, the heat flux is set to zero, following the original publication. The vertical discretization of the model is not homogenous. For WAIS, a vertical step of 1 m for the upper 500 m and up to 25 m for the deepest part, and for other sites where the depth of borehole is close or less than 500 m, the step is set to 1 m for overall depth.~~

~~Before the forward model is driven by the climate model results, it is initialized with a stationary profile, which is generated after a 20000-year model run with a constant climate history and a realistic seasonal cycle. The mean surface temperature is set to the recent annual average temperature and the season cycle is determined by simplifying the average over weather station data following Eq. 2 (Orsi et al., 2012).~~

$$T(t) = 10(\cos(2\pi t) + 0.3\cos(4\pi t)) \qquad (2)$$

2.4 Sensitivity of subsurface temperature to model parameters

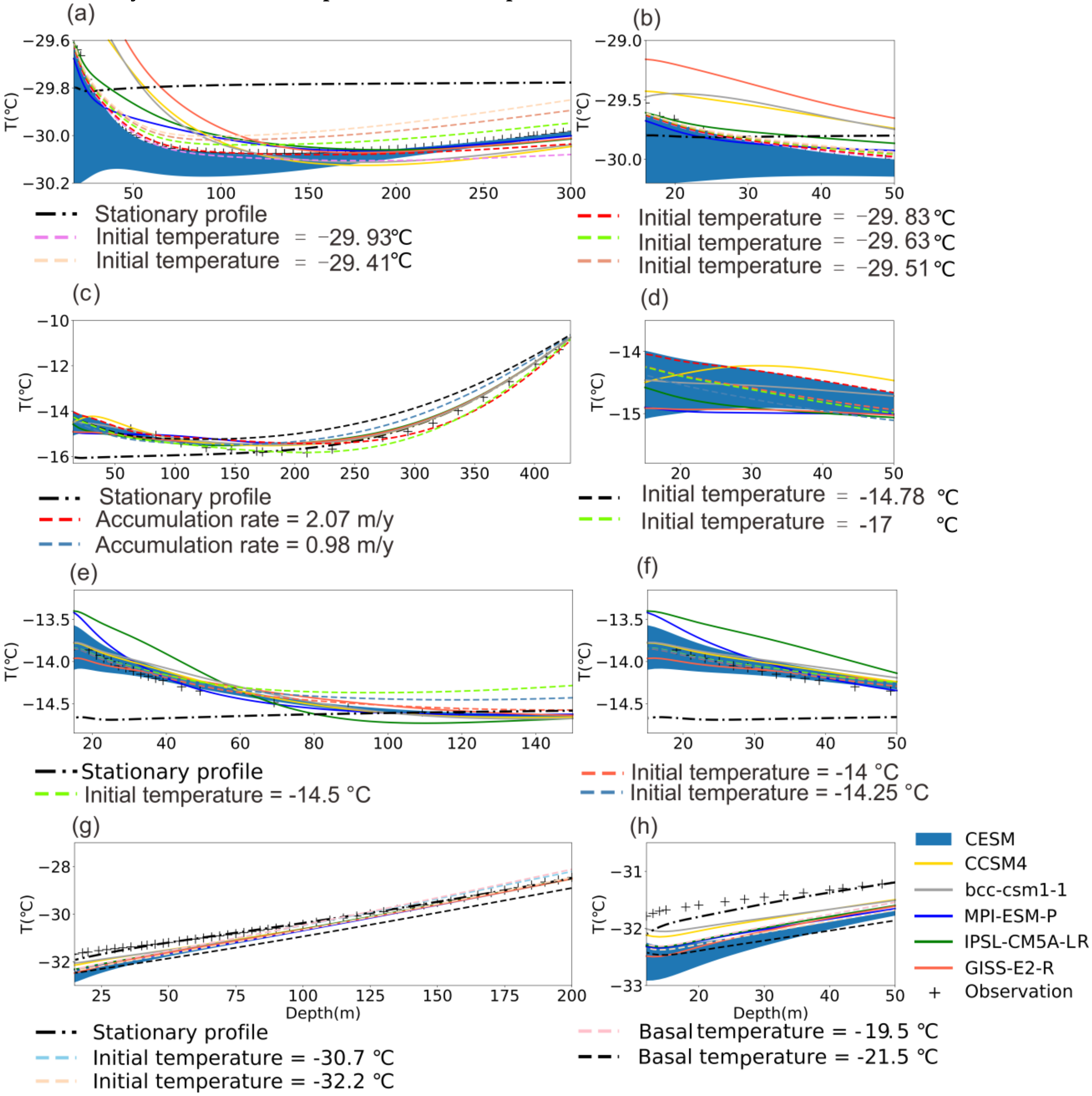

Figure 2. Comparison of borehole temperature profiles outputs for the forward model driven by GCMs surface temperature time series with optimal parameters (solid lines), and sensitivity tests using the temperature history of one CESM member (dashed lines) at each site. (a) WAIS: 15-300 m; (b) WAIS: 15-50 m; (c) Larissa: 15-430 m; (d) Larissa:15-50 m; (e) Mill Island:15-150 m; (f) Mill Island:15-50 m; (g) Styx:15-200 m; (h) Styx:15-50 m. The shaded area represents the simulated subsurface temperature ensemble driven by CESM using optimal parameters. The thick dash-dot line denotes the stationary profile at each site.

According to the original studies describing the records and the surface temperature reconstructions, the various parameters in the forward model have effects of different magnitude on the results for the different sites. Consequently, In order to assess the uncertainty in the model-data comparison related to the parameters of the forward model, *we perform a series of sensitivity experiments on the parameters which have been shown to have the largest effects on each of the borehole profiles shown in the Fig. 2.*~~it is necessary to perform a series of sensitivity experiments as shown on Fig. 2. We made different tests for the key parameters using the values proposed in the original publications (Table 1) and following the protocol of Orsi et al. (2012).~~

~~The range of tested model parameters in the forward model can influence significantly the shape of simulated subsurface temperature (Fig. 2), which is in good agreement with the previous studies at those sites.~~

At WAIS-Divide, the spread of the sensitivity tests is lower than the spread in the simulated borehole profiles driven by different climate model results (solid lines in colour in Fig. 2a and b).~~At WAIS-Divide, the spread of the sensitivity tests is lower than the spread if the different scenarios. An increase in the accumulation rate will reduce the temperature gradient in the borehole profile, but the effect is much weaker than the difference in temperature histories from the different models~~. *However, the initial temperature derived from a steady state profile* ~~A change in the initial temperature used to calculate a the starting steady state profile~~ has an influence on the slope of the profile in the deeper part and *on the* depth of the temperature minimum, contributing to the uncertainty in the intensity of the pre-1900 cooling trend and the timing of the temperature minimum.

At Larissa, the effect of the bottom boundary conditions is important in setting up the temperature gradient from the bottom to 300 m, and therefore, we will not interpret that segment of the data in terms of climate. It is also evident in Fig. 2c that the different temperature histories produce a very similar depth profile over that interval.

At Mill *I*~~i~~sland, *the borehole profile is shallow and covers only a fraction of the full thickness of the ice sheet. At sites with such a deep ice sheet and with a high accumulation rate, the optimal surface temperature history was found to be essentially independent of the location of the imposed bottom boundary condition for depths in excess of 180 m below the surface (Roberts et al., 2013).*~~although the borehole profile is shallow, the ice thickness is much *thicker*deeper, but unknown~~. Here we modeled *the temperature* ~~this~~ by assuming a zero heat flux bottom boundary ~~at various depth~~. *Although the initial temperature has an influence on the slope of the profile deeper than 120 m, this sensitivity is weak in the depth shallower than 80 m, and the borehole profile is dominated by the surface temperature history.*~~This sensitivity is weak over the data interval, and the borehole profile is dominated by the surface temperature history.~~

At Styx, the boundary condition~~s~~ *is*~~are~~ adjusted to reproduce the slope of the temperature profile in the deeper part (100-200m).~~,~~ *The simulated borehole profiles driven by GCMs (solid lines in the Fig. 4e) show the large deviation in the top 100 m compared with stationary temperature profile, which suggests that there is climate information stored in the upper part of the profile. Meanwhile, at the depth shallower than 50 m, the effect of boundary conditions is weaker than the differences in the temperature histories from the different models, which means the borehole temperature data can be used to discriminate between temperature histories provided by the different models.* ~~but the deviation in the top 100 m show that there is climate information stored in the upper part of the profile, and that this profile cannot be fully determined by boundary conditions.~~

The internal variability also has a significant impact on the shape of the simulated borehole profiles. At these four sites, the range of simulations driven by CESM ensemble is much larger than the range of the different sensitivity tests in the top of 50 m (shown as the shaded area in Fig. 4b, d, f, and h). This confirms that the internal ~~climate~~ variability *is* ~~and different characteristics of these climate models are the~~ *a* dominant source of the ~~differences~~ *uncertainty in a model-data comparison, at least in the top 50 m.* ~~and that the model-data comparison provides a robust evaluation of simulated temperature time series.~~ For the deeper part of WAIS ~~and Larissa~~, as the shape of subsurface temperature profiles is influenced by the parameters of the forward model, the evaluation of the long-term cooling trend is more uncertain.

3. Results

3.1 Comparison between the simulated temperature and reconstructions

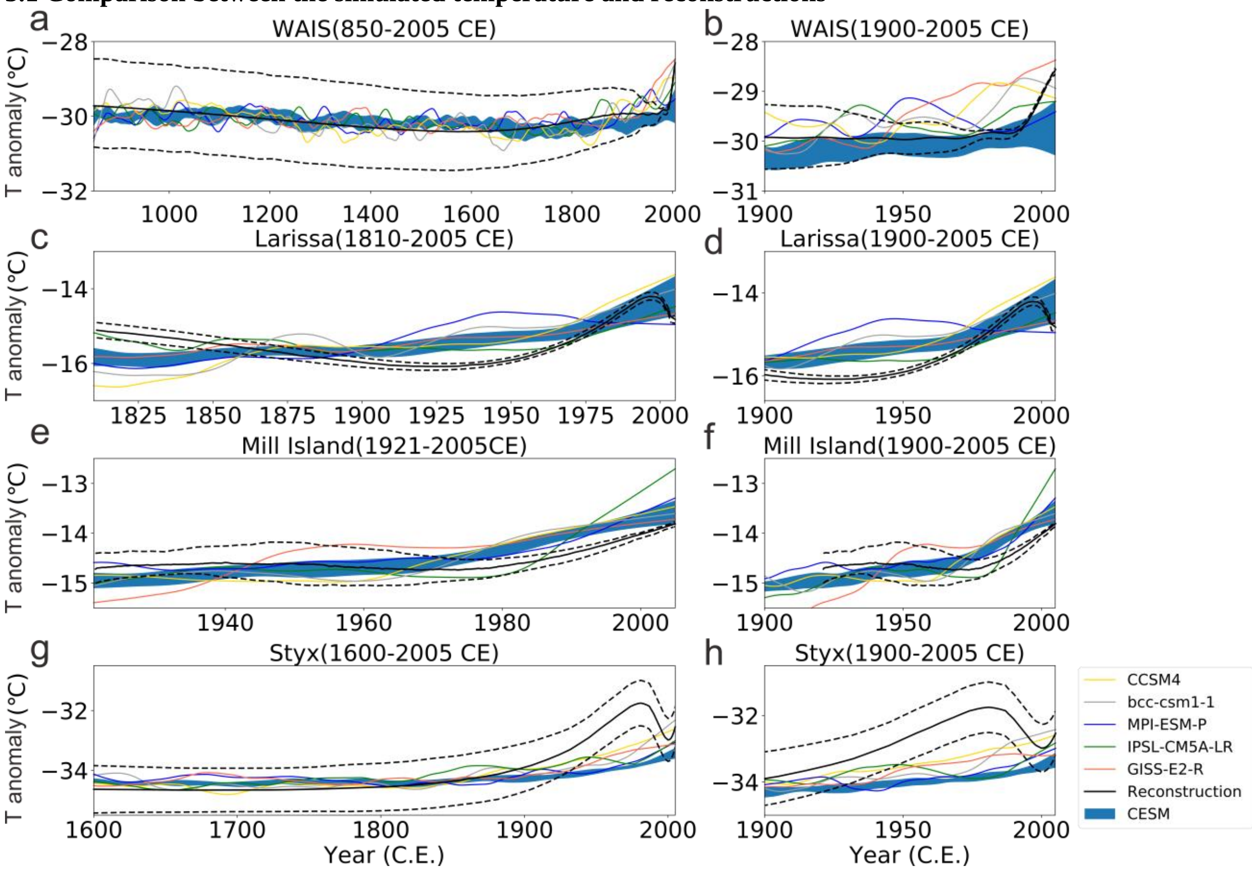

Figure 3. Comparison between reconstructed surface temperature series from boreholes and the climate model outputs at the grid cell closest to each borehole site. The borehole reconstructions are in black and their uncertainty ranges given by the dashed lines. Colour lines correspond to the climate model results. The shaded area represents the mean ±1 standard deviation of CESM model ensemble. For the left column, a 50-year Lowess smoothing has been applied for the WAIS and Styx time series; Larissa and Mill Island are smoothed using 10-year and 3-year windows respectively. The time series in the right column is smoothed using 3-year from 1900 to 2005 CE.

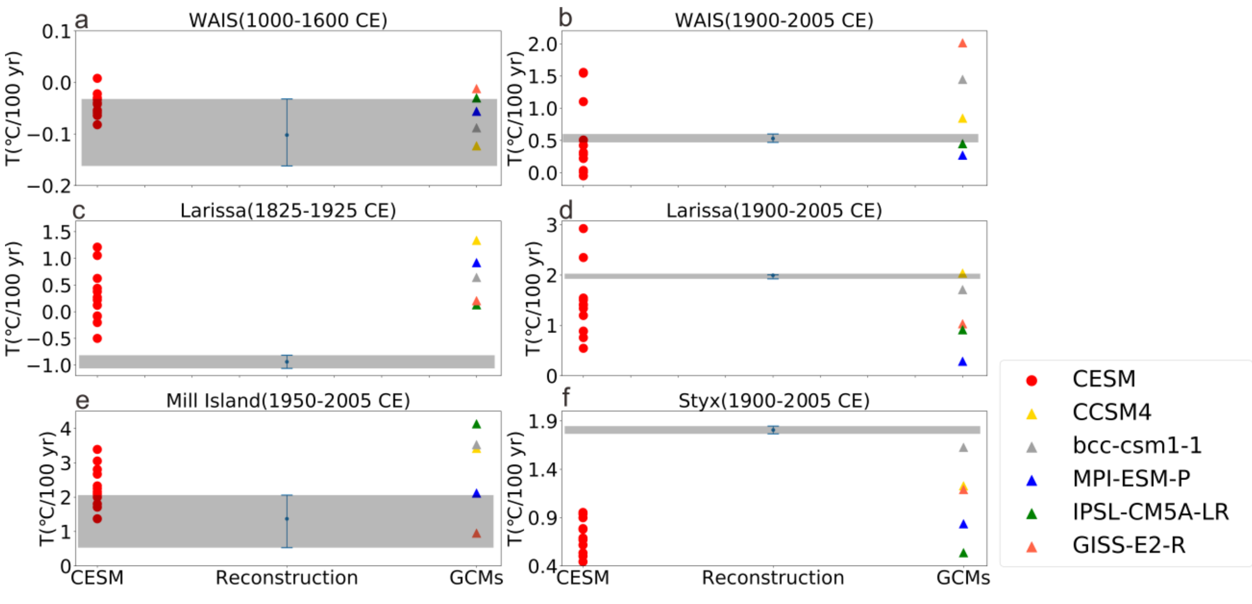

Figure 4. Linear trends for the four boreholes over different periods: (a) WAIS: 1000 to 1600 CE; (b) WAIS: 1900 to 2005 CE; (c) Larissa: 1825 to 1925 CE; (d) Larissa: 1900 to 2005 CE; (e) Mill Island: 1950 to 2005 CE; (f) Styx: 1900 to 2005 CE.

Figure 3 displays the comparisons between climate model results and temperature reconstructions from *the* borehole*s. The simulated temperatures displayed in Fig. 3 come directly from the surface temperature calculated by the climate model, based on its own dynamics and the forcing applied as discussed in Section 2.2. In order to ensure that the climate model results have the same mean over the reference period (which is the whole period derived from the reconstruction) as the reconstruction, we applied a very simple, constant correction to remove the mean bias of the climate model results as shown on the Fig. 3.*~~for the four selected sites. In order to remove the bias on the mean state for each climate model, anomalies are shown using the total period covered by each reconstruction as reference.~~ Due to the nature of physical diffusion, the heat propagation acts similarly as a low-pass filter. The reconstructions thus suffer from an attenuation of high frequency temperature variability that becomes stronger as times goes back (Beltrami et al., 2006; Harris and Gosnold, 1999). For instance, in the reconstructed surface temperature of Styx, the point corresponding to 1800 CE in the curve may represent an average temperature between around 1600 CE and 1900 CE while *in*~~at~~ 1900 CE it corresponds to an average over around 200 years. *This characteristic complicates the model-data comparison. Therefore, in order to facilitate the comparison between the reconstruction and climate model results, we use variable smoothings to mimic the characteristic as much as possible. Since the reconstructions have much wider ranges than those ones from the climate model results, the basic compatibility between model and data will not be changed due to various smoothing. Nevertheless, Fig. 3 must be interpreted carefully because of this inhomogeneous smoothing.* ~~This characteristic complicates the model-data comparison and trends as shown on Fig. 3 must be interpreted carefully because of this inhomogeneous smoothing.~~

Because of the internal variability of the system, a single simulation *without error bound* is not expected to reproduce well all the characteristics of the observed variations. The difference can be large, in particular at the local level (e.g. Goosse et al. 2005), but the observations should correspond to a credible member of an ensemble of simulations. Ensuring this compatibility can be achieved using various techniques but the first step is to simply check if the reconstruction is within the range provided by the ensemble (e.g. PAGES2k-PMIP 2015).

Considering the large uncertainty range in these reconstructions, the climate models are visually able to reproduce the general characteristics of reconstructed temperature variability, particularly ~~in~~ the long-term cooling during the last millennium and the recent warming (Fig. 3 and 4). Nevertheless, disagreements have also been identified.

The first major feature in the data is the long-term cooling trend, visible at the WAIS-Divide and Larissa sites. At Larissa ~~(Antarctic Peninsula)~~, the borehole temperature reconstruction *gives*~~finds~~ a cooling trend of -0.94 ± 0.12 °C/century from 1825 to 1925 (Zagorodnov et al., 2012). None of the models *is*~~are~~ able to reproduce this observation, and instead, they all show a warming trend of comparable magnitude (Fig. 3c and 4c). At WAIS-Divide, the borehole temperature inversion also shows a long-term cooling trend, from 1000 to about 1600 C.E., with a magnitude of -0.10 ± 0.07°C/century (Fig. 3a). The large uncertainty in the long term trend is principally due to the uncertainty in the initial surface temperature (Fig. 2a; Orsi et al., 2012, their Fig. 3). The quantitative comparison between the trend of reconstructions and climate model outputs (Fig. 4a) indicates that the simulations generally show a cooling trend over 1000-1600 CE, in agreement with previous studies (e.g., Goosse et al. 2012, Abram et al. 2016, Klein et al. 2019).~~,~~ *The amplitude of the trend is lower, particularly for GISS (-0.01 °C/century) and IPSL (-0.03 °C/century) models, but most remain within the lower end of the reconstructed uncertainty range.*~~but with a lower amplitude, particularly GISS (-0.01°C/century) and IPSL (-0.03°C/century), but most remain within the lower end of the reconstructed uncertainty range~~. This long-term cooling trend is a feature of the Antarctic climate that is visible in many other ice core records (Stenni et al., 2017). A recent compilation of *PAGES Antarctica2k*~~2K~~ datasets calculated a trend of -0.26 to -0.4°C/1000 years for the period 0-1900 AD for the Antarctic continental average (Stenni et al., 2017). In the high latitudes of the Southern Hemisphere, the origin to this millennial-scale cooling is currently not well understood, but an intermediate complexity model has shown a multi-millennial cooling in summer because of a delayed response to the decrease in local spring insolation (Renssen et al., 2005) with also a potential influence of volcanic forcing (Goosse et al. 2012, Abram et al. 2016, Stenni et al. 2017).

A second feature of the data is a warming trend in the twentieth century, which started at different times in the different records. Styx shows an early warming trend from 1900 to 1980 *CE*, and a *general* stabilization of the temperature afterwards (Fig. 3h). *This signal is consistent with the data form weather stations, and ice core isotope-derived records (Yang et al., 2018).* Models tend to show the opposite timing, with nearly no trend from 1900 to 1960, and a late warming trend that differs in amplitude between models. Overall, the warming of the 20th century is about half of what is observed (Fig. 4f), with bcc (1.63°C/century) and CCSM4 (1.23°C/century) having the largest trends, closest to the observations (1.81°C/ century).

Larissa shows a temperature minimum in 1940's, followed by a steady warming trend until around 1995 *CE*. The magnitude of the 20th century trend is 1.99°C/century. Most models reproduce the timing of the warming reasonably well, with the exception of MPI, which shows an early warming, but no trend in 1940-2005 *CE*, and GISS, which has a very muted trend. If the trend present in the other models is too low, *it seems rather due to a lack of cooling in the preceding century, than because of errors in the latest decades.* ~~it is rather because of the lack of cooling in the preceding century, than because of errors in the latest decades.~~

Mill Island shows a late warming trend starting in the 1980's. Models tend to overestimate this trend (Fig. 4e), in particular IPSL, bcc and CCSM4. Similarly to Mill Island, WAIS-Divide also shows a positive trend over the period 1900-present that intensifies after 1980. The amplitude of the 20th century warming (0.53 °C/century) is well simulated, but the start of the trend is often too early, with the exception of CESM, bcc and IPSL, which show a late warming trend (Fig. 3b).

Overall, *for WAIS (Fig. 4b) and Larissa (Fig. 4d), the reconstructed trends lie in the CESM ensemble range, suggesting many apparent model disagreements for those sites can be due to internal variability. For Styx (Fig. 4f) and Mill Island (Fig. 4e), the reconstructed trends are larger than the spread of the CESM ensemble, which means the disagreements are not only due to internal climate variability but are related to a systematic climate model bias in this region.* ~~the large variability of the trends over the 20th century within the CESM ensemble for WAIS and Larissa suggests that many apparent model disagreements for those sites can be due to internal variability while the disagreement may be more profound for Styx and Mill Island.~~

However, as stated above, borehole temperature reconstructions are "underdetermined", which means that there are many possible temperature histories that can fit the data *(more detailed explanation of "underdetermined" is given in the Introduction)*. The next step is to determine if the differences between simulated and reconstructed time series can be discriminated when analyzing observed and simulated temperature profile.

3.2 Comparison of the simulated subsurface temperature with observation

The simulated subsurface temperature profile is the results of the superposition of two components: (1) the initial temperature profile that incorporates *the effects of basal heat flux, and vertical advection due to ice accumulation*~~the effects of basal heat flux, vertical advection due to ice accumulation and initial temperature~~; (2) the subsurface temperature deviations arising from the surface temperature variability. Since the initial temperature profile for each borehole is obtained by driving the forward model with the optimal parameters obtained from the original publications ~~describing the reconstructions~~ (see Section 2.4), the differences among the simulated borehole profiles for each location are caused only by the changes in the upper boundary, i.e. in the climate model outputs. The simulated subsurface temperature profiles for each borehole are displayed in Fig. 5.

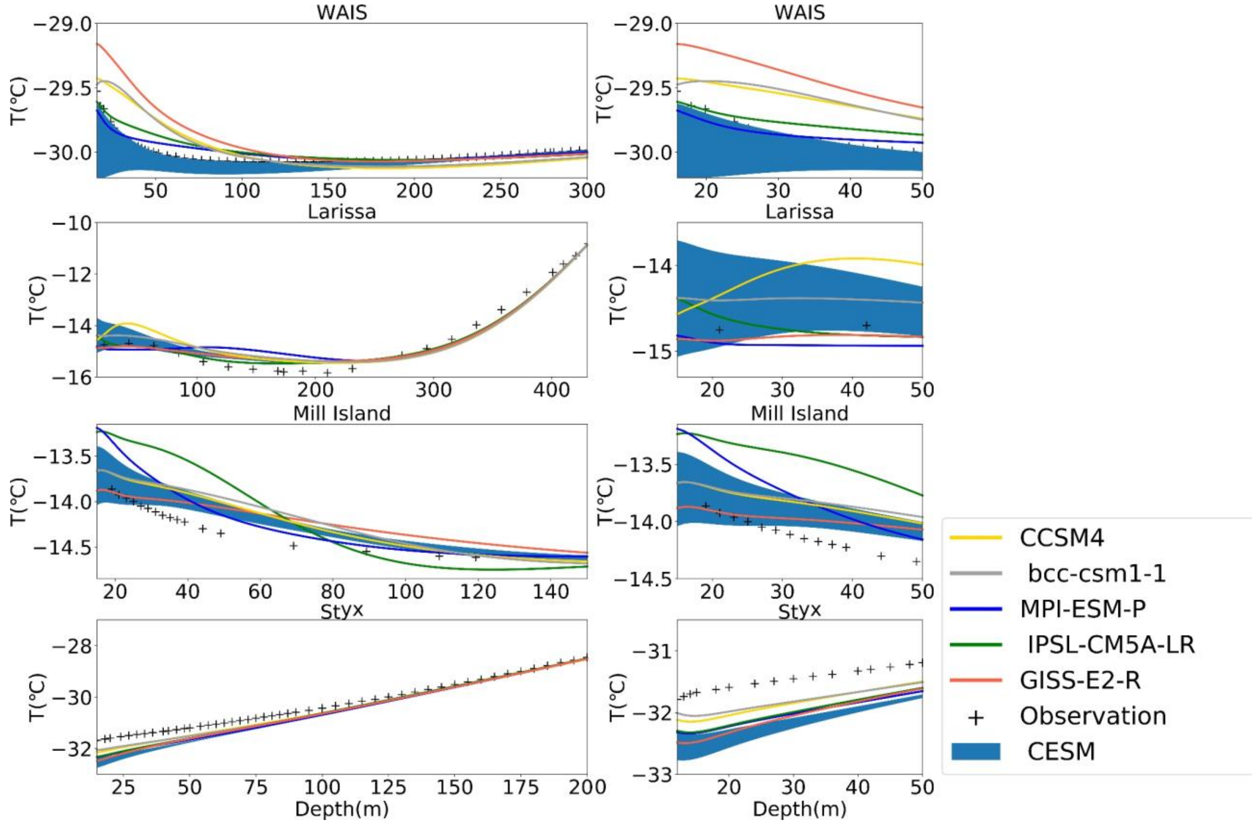

Figure 5. Comparisons between simulated subsurface temperature and measurements for : (a) WAIS: 15-300 m; (b) WAIS: 15-50 m; (c) Larissa: 15-430 m; (d) Larissa:15-50 m; (e) Mill Island:15-150 m; (f) Mill Island:15-50 m; (g) Styx:15-200 m; (h) Styx:15-50 m. The shaded area represents the simulated subsurface temperature ensemble driven by CESM ensemble. The right panel is a zoom over the upper 50 m for each borehole**.**

As previous studies shown (Bodri, Louise, and Vladimir Cermak, 2011, Chap. 2), a 'U' shape subsurface temperature profile is a direct evidence for the past climate change with a minimum that separates the deeper warming trend due to geothermal heating and shallower warming trend related to a recent temperature increase (Orsi et al., 2012; Stevens et al., 2008). Among these four sites, WAIS and Larissa have such characteristics of 'U' shape~~ curve~~. For Mill Island, this is less clear but a significant breaking point in each simulated subsurface temperature profile reflects the surface temperature warming over recent decades*. For Styx such break does not seem to be present at all and the slope does increase with depth.*~~while for Styx such break does not seem to be present at all, but the slope does increase with depth.~~

Aided by these key properties, we can identify a link between the interpretation in the depth domain and in the time domain. The analysis of the simulated and observed temperature profile confirm*s* the main conclusion obtained in section 3.1, in particular *the agreement between model and data on the general tendencies, characterized by a long-term cooling trend over last millennium and the recent warming.*~~with an agreement between model and data on the general tendencies, characterized by a long-term cooling trend over last millennium and the recent warming.~~ For the deeper part of the profile, the *simulated* ~~-~~temperature *profiles* ~~simulated in experiments~~ driven by MPI, IPSL, GISS at WAIS almost coincide~~s~~ with the corresponding observations, but they fail to reproduce the depth of the temperature minimum around 120 m in the data. This is consistent with the fact that IPSL and MPI are at the edge of the reconstructed cooling trend of the last millennium and GISS present*s* a significant underestimate (Fig. 4a). On the other hand, the CESM ensemble follows the borehole temperature profile (shaded area on Fig. 5a), and ~~could~~ *can* also reproduce the magnitude of the cooling trend for some of the members (Fig. 4a). Specifically, the minima in the simulated profiles driven by MPI, IPSL, GISS and CESM shows ~~the~~

value of -30.06 °C, -30.06 °C, -30.07 °C and a range of -30.8 °C to -30.17 °C respectively, which is very close to the minimum of -30.08 °C in the observations.

At Larissa, the bottom (270-450 m) of the profile is controlled by boundary conditions (Fig. 2c), and contains no climate information, as demonstrated by the fact that all curves are on top of each other on Fig. 5c. Additionally, no simulation has a pronounced inflection point around the 170 m as in the observation. These characteristics are perfectly consistent with the lack of a cooling trend from mid~19th century to the early 20th century in the simulations (Fig. 3c). We conclude from this that the cooling trend of 1825-1925 is a robust feature in the data that can be used to benchmark climate models.

For the recent warming, we see some significant discrepancies among the simulated subsurface temperature *profiles*s driven by different climate models at the four boreholes in the depth domain that are consistent with the signal analyzed in the time domain. For WAIS, in the uppermost part, the simulated subsurface temperature profiles driven by GISS, CCSM*4* and bcc display significantly higher temperatures than in the observations, while IPSL and MPI-simulated profiles are close to the measurements (Fig. 5b). This is in perfect agreement with the too high temperatures in models compared to the reconstructions in the second half of the 20th century (Fig. 3b).

For Larissa, all simulated profiles display an increasing temperature toward the surface as in observations but with different magnitude and shape (Fig. 5c). The temperature in the simulation driven by the MPI displays a relatively rapid increase until around 100 m and then is constant, which is consistent with the near constant temperature from 1940-2005 (Fig. 3d). For the ones driven by CCSM*4* and bcc, they are warmer than the observations between the depth 15m to 50m, which reflects the consistently warmer temperature shown in Fig. 3d. IPSL-simulated subsurface temperature profile displays the largest similarity to the observations, whilst the simulations performed with CESM can cover almost all the observation in the shallow zone.

For Mill Island, the simulated subsurface temperature profiles are warmer than observations above 50 m, confirming the too large warming trend deduced from the analysis of surface temperature. In particular, the IPSL model has the largest warming trend (Fig. 3 e *and*, f) and also has the warmest temperature profile (Fig. 5 e, f), followed by MPI. *For Styx,* The borehole data thus is providing constraints to evaluate the different simulations. For Styx, the main discrepancies occur over the shallow depths, between of 15 m to 60 m, where all the simulations depict colder condition compared with observations (Fig. 5 g, h), as for the surface temperature over the recent decades on Fig. 3.

Nevertheless, we also find in the depth domain some signals that are not obvious in the time domain. In particular, for WAIS, one of the CESM runs matches the warming trend of the top 100 m, while in time domain the CESM ensemble wa*i*s significantly colder than reconstruction over recent decades. The CESM outputs generally follow the data in the deeper part of the profile (200-300 m), and have an even steeper slope between 100 and 200 m (Fig. 5), while in the time domain, the cooling trend was underestimated (Fig. 4a). In addition, for WAIS, the simulated subsurface temperature driven by CCSM4 and bcc over the deeper part of the profile are colder than observations, but *the warming* start*s* the warming trend deeper, at about 200 m against 120 m in the observations. This seems puzzling because, in the time domain, the cooling trend continues until 1800 *CE* for CCSM4 (Fig. 2a, yellow)*. However,* but the larger warming in the last 100 years, is probably shifting the temperature minimum downwards. This example shows that it is difficult to pinpoint the date corresponding to a temperature minimum in the depth profile, because it depends on the respective speed of warming and cooling before and after. At Mill Island, in the deeper part (around from 140 m to 100 m) of the profile, the simulated subsurface temperature profile driven by IPSL, *is very different from the other ones* with a slightly decreasing temperature and a colder climate than observations., is very different from the other ones. However, in the time domain, the difference compared to other time series for IPSL was much less clear but the consistency between these two domains still exists, and especially the temperature minimum in 1980 *CE* might correspond to the deeper part in the depth domain.

The comparison between the analyses in the two domains appears thus complementary and instructive as it illustrates that the interpretation may be easier in one case or the other. It also shows that the different model runs produce different borehole temperature profile, and that the observations can help evaluate the models *by comparing different borehole temperature profiles driven by the different climate model results with the corresponding observation.*. In particular, the analysis of the simulated temperature profile confirms that CESM ensemble can reproduce the multi-decadal and centennial climate variability at WAIS.

4. Proposed metric of Antarctic climate for model validation

In this section, we use the results of the previous section to describe a few metrics that can be used easily to evaluate the next generation of climate model simulations (e.g. PMIP4-CMIP6, Jungclaus et al., 2017), and investigate the spatial representativity of the records.

4.1 Metric 1 : last millennium cooling at WAIS Divide

Of the four records presented here, WAIS-Divide has the longest retrievable history. We propose here to use the temperature trend of the period of 1000 to 1600 C.E. as a metric, with the magnitude of -0.102 ± 0.07 °C/century (Fig. 4a). The end of the cooling trend is not clearly defined by the data, due to the complex time-varying smoothing of the borehole temperature record, but 1600 C.E. seems to be safely in the cold interval (See Orsi et al., 2012, Fig. 4a for details). The start of the period is more open, and we chose 1000 C.E. to be compatible with last millennium simulations. External evidence from a compilation of water isotope records indicates that the cooling trend extended likely from 0 to 1900 C.E. in many parts of Antarctica (Stenni et al., 2017). It is a robust feature of the Antarctic climate of the last 2 ka, and the WAIS-Divide record is unique in providing a clear quantification of the temperature trend.

In Fig. 6, we show the 1000 to 1600 C.E. surface temperature trend at WAIS-Divide and at other sites in Antarctica from the models output. Visually, for most simulations, the cooling at the grid point *cell* of WAIS-Divide is similar to the one obtained at many location in West Antarctica. Only the first member of CESM shows a small warming trend in West Antarctica. The large spatial coherence of the trend indicates that, although we are making a single point comparison, it represents a signal common to a large part of the continent. It is also important to estimate the magnitude of the trend at WAIS compared to other regions. To do so, we calculate the ratio of the trend of surface temperature from 1000 to 1600 C.E. at any location with the one at WAIS-Divide (Fig. 7). Except the first member of CESM, if the value is greater than 1 (shown in red tones), it means the trend at the grid *cell* point is larger than that at WAIS-Divide; if the value lies between 0 and 1 (shown in blue tones), it means the trend at the grid *cell* is less than that observed at WAIS-Divide. Negative values (i.e., a trend of a different sign compared to WAIS) are not shown and the corresponding region left blank. Since the goal of Fig. 7 is to show the intensity of cooling at WAIS compared with other points in Antarctica, the first member of CESM 1, which shows a warming trend close to zero at WAIS, is not very meaningful but it is still included for completeness. *75% models show* For most of the models *(12/16)*, WAIS displays a larger cooling from 1000 to 1600 C.E. than other locations in Antarctica (shown in blue) but with magnitude similar to other grid points *cells* in West Antarctica., *This*which is consistent with the reconstruction of Stenni et al. (2017) that shows the largest cooling in this region over the period 0-1900 CE C.E (Stenni et al., 2017). The spatial patterns of the trends (Fig. 7) are different between models, but also within the CESM ensembles, showing that the changes in Antarctica are strongly influenced by internal variability, even at century timescale. Future work including more sites, or using water isotopes and the Antarctica-2K database will help constrain the spatial pattern of this trend.

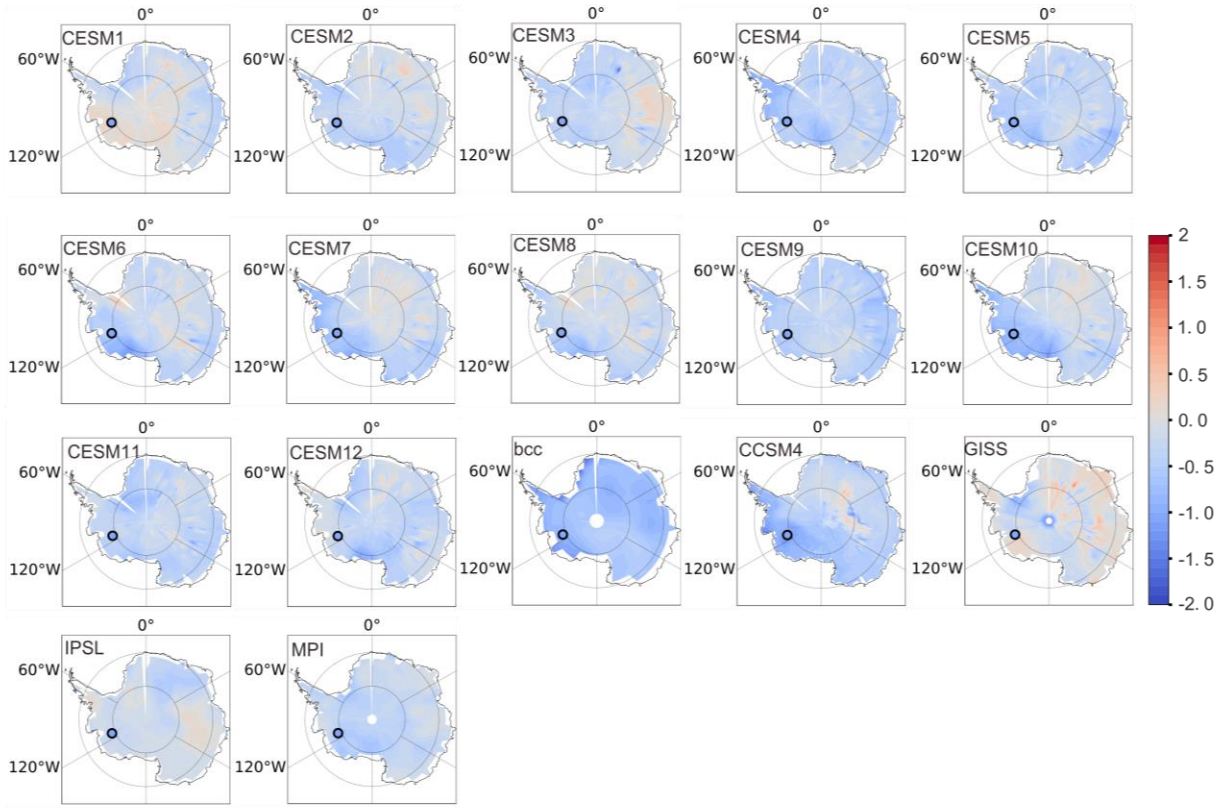

Figure 6. The simulated (blue-red shading area) and observed (circle) surface temperature trend from 1000 to 1600 C.E in Antarctica.

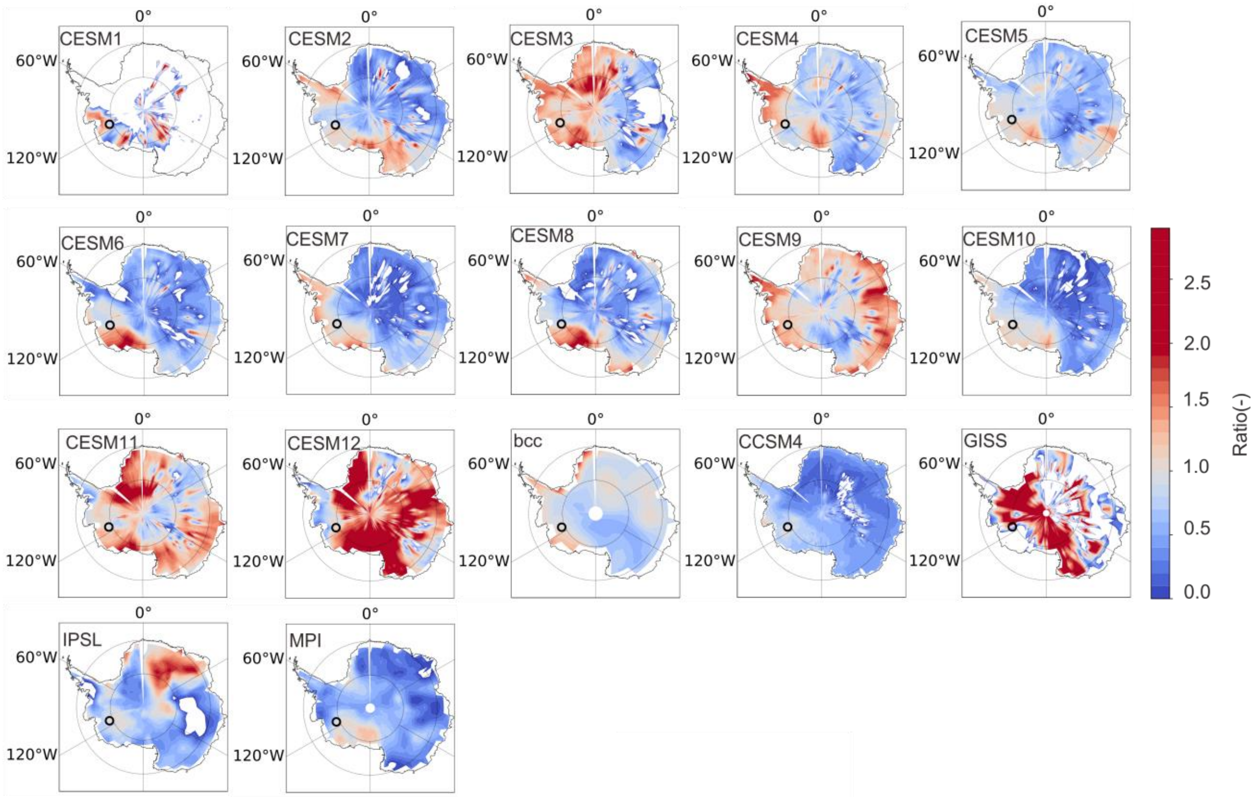

Figure 7. The ratio of the surface temperature trend (blue-red shading area) from 1000 to 1600 C.E between other grid~~s~~ *cells* in Antarctica and WAIS-Divide. The black circle denotes the location of the WAIS Divide.

4.2 Metric 2: nineteenth century cooling at Larissa

The second metric is the surface temperature trend over the period from 1825 C.E. to 1925 C.E. at Larissa, with the magnitude of -0.94 ± 0.12 °C/century. Fig. 8 shows the spatial correlation *from 1825 to 1925 CE at Larissa and other grid cells for each climate model.*~~in the Antarctica Peninsula (AP).~~ . *As there are no significant differences between each member in CESM ensemble (see in the Fig. S1), only one member CESM1 is presented in the Fig. 8.* Despite the correlation coefficient decreasing *with the distance from the Larissa*~~as the grid *cells* getting far away from the Larissa~~, the values, at least around Larissa for each model, are higher than 0.6, showing that this metric is representative of the ~~whole~~ *part of* peninsula region, and not extremely site-specific.

None of models is able to capture the observed temperature trend from 1825 to 1925 CE (Fig. 9). ~~Figure 9 shows the same temperature trend (1825-1925) for all models.~~ Overall, models are showing a warming trend (largest for CCSM*4*, MPI and BCC), contradicting the observations, as highlighted already in Fig. 4c. ~~A majority of the CESM members~~*Only four member of CESM* (CESM1, 7, 8, and 9) show a cooling trend over ~~Antarctica~~*AP*, *but the magnitudes of them are still less than the observed one*~~with CESM 1 and CESM 7 being able to capture the observed trend~~.

The 19~~-~~th century is a time period when the Northern Hemisphere has started warming, whereas Southern Hemisphere records (Neukom et al., 2014), and specifically Antarctica, show no general warming trend (Stenni et al., 2017). Models tend to over~~-~~estimate the interhemispheric synchroneity (Neukom et al., 2014), and show a warming trend also in Antarctica , possibly in response to the anthropogenic forcing. This metric is thus an important tool for future research to evaluate whether the model data mismatch is due to internal variability (which will be investigated with more ensembles of the same model), or to an overstimated sensitivity to the antropogenic forcing.

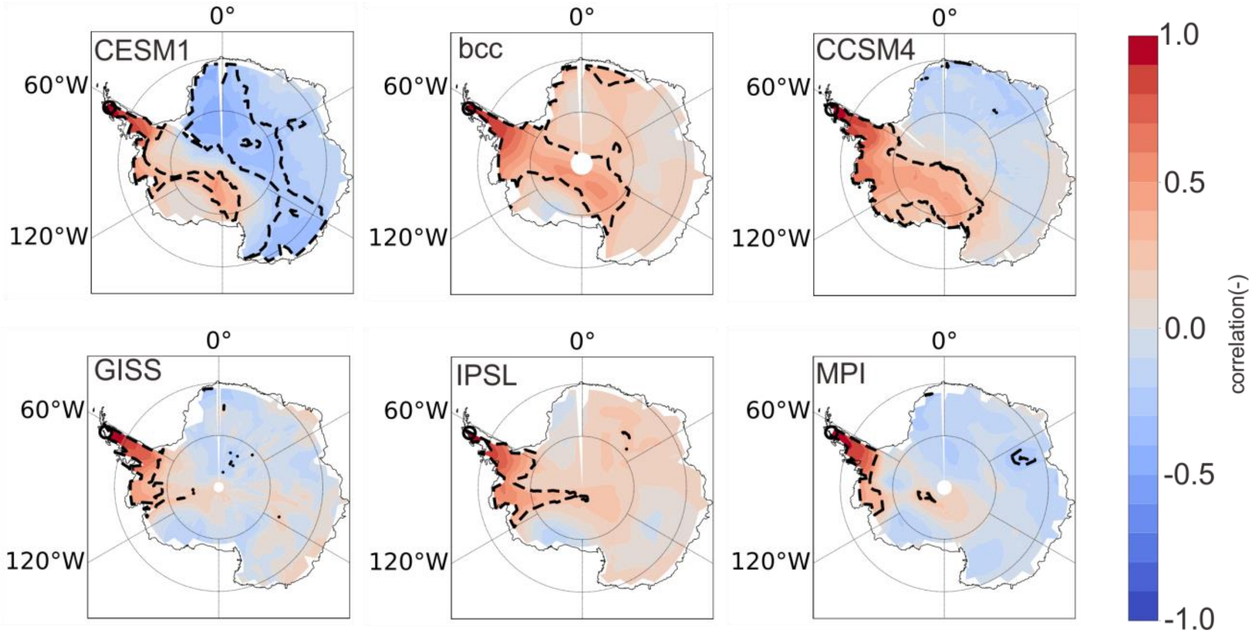

Figure 8. The correlation map (blue-red shading area) showing the relationship between the temperature from 1825 C.E. to 1925 C.E at Larissa and other grid~~s~~ *cells* in AP for each climate model~~s~~. The black dotted contour lines show a significant correlation at the 99 % significant level.

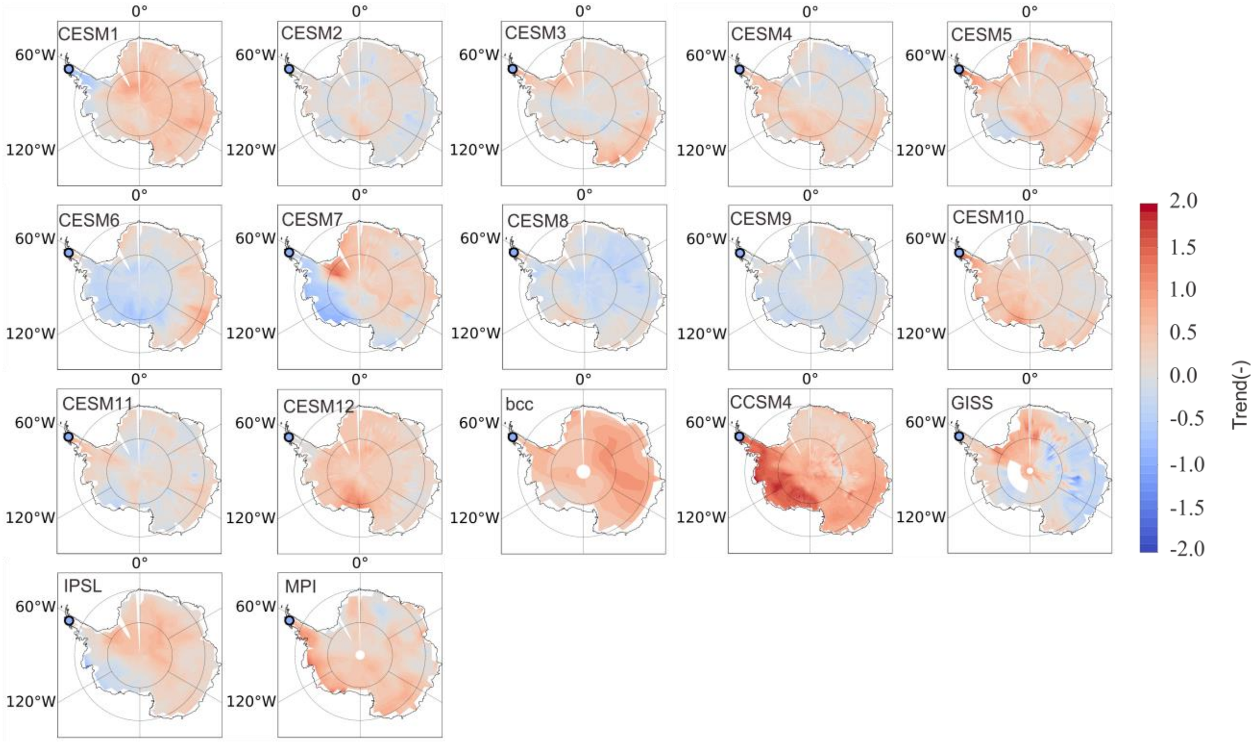

Figure 9. The simulated (blue-red shading area) and observed (circle) surface temperature trend from 1825 to 1925 C.E.

4.3 Metric 3: recent warming trend

The warming trend of the last 50 years is one of the clearest features of the observations.. The intensification of the Southern Annular mode, in response to the Ozone hole is expected to produce a strong warming in the Antarctic Peninsula (incl. Larissa), and cooling on the Antarctic Coast (incl. Styx and Mill Island). WAIS has also been warming significantly over the past decades, and this trend is attributed to variability in the strength and position of the Amundsen Sea low pressure system (Jones et al., 2016). Here we propose a metric of the warming trend from 1950 to 2005 at each of the four sites, to investigate whether model can reproduce these features.

First we look at the spatial correlation of the temperature between each site and other grid points *cells for all GCMs* (Fig. 10). *Only one member of CESM1 is presented in the Fig. 10 since no significant difference is observed between each member in CESM ensemble (see in the Fig. S2-S5).* The correlation is calculated on annual data for 1950 to 2005 C.E.. It is clear that each of our borehole temperature sites gives information about different sectors of Antarctica. Generally speaking, WAIS is representative of the West-Antarctic continent, with a more pronounced dipole between WAIS and the Weddell sea section in MPI, and to a lesser extent CESM and GISS. Larissa is representative of the Antarctic Peninsula as a whole, and from this resolution of climate model runs, there is no evidence of a dipole between either *both* side of the Transantarctic mountains. Similar to WAIS, MPI has the strongest expression of a dipole between the Antarctic Peninsula and East Antarctica, a feature that is weaker but also present in GISS. A model that responds clearly to the Ozone forcing, and has a strong SAM signature should exhibit this dipole pattern, and it is interesting that some models do not show it, indicating that the Ozone forcing is not dominating over internal variability. Mill Island is generally representative of the Wilkes Land sector of East Antarctica, with the largest spatial homogeneity for BCC and IPSL (Fig. 10c). Finally, for Styx, the models with the largest spatial homogeneity (BCC and IPSL) show a strong correlation between Victoria Land and the rest of East Antarctica.

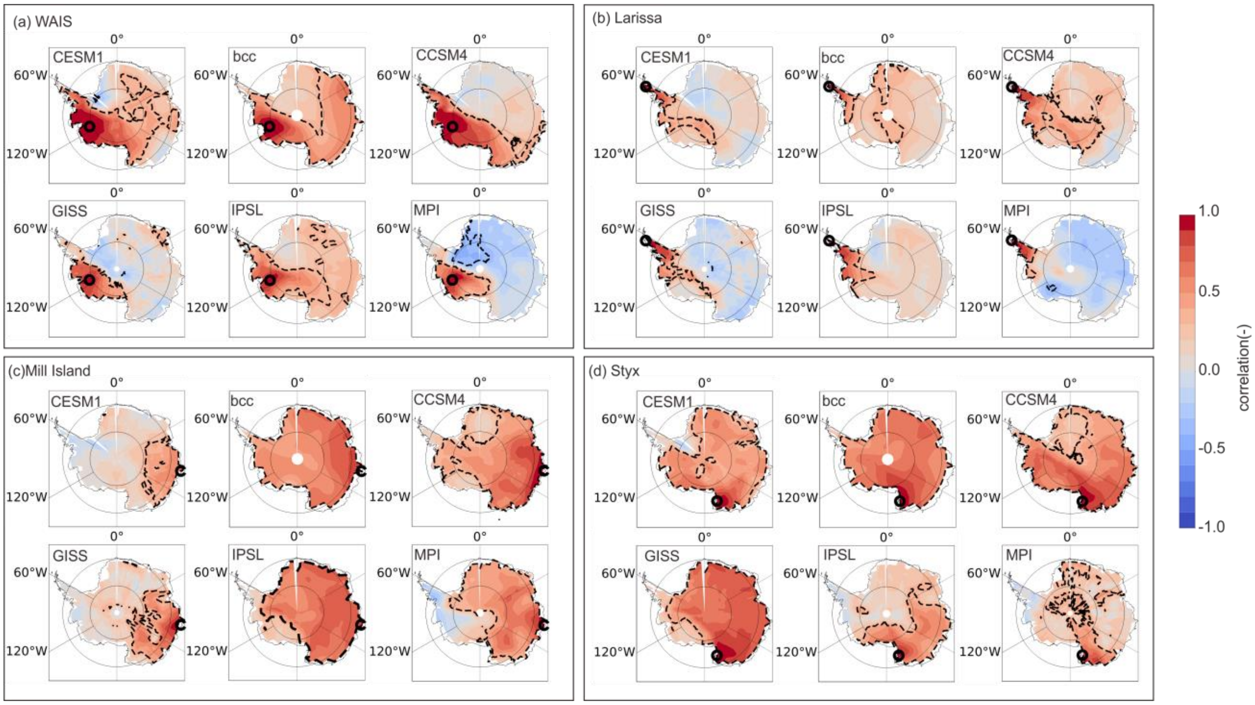

Figure 10. The correlation map showing the relationship between the temperature from 1950 C.E. to 2005 C.E at WAIS (a), Larissa (b), Mill Island (c), Styx (d) and other grid~~s~~ *cells* for each climate models. The red dashed contour lines show a significant correlation at the 99 % significant level.

Figure 11 shows the surface temperature trend from 1950 to 2005 C.E. The strong warming trend at Larissa is underestimated in *all the models except CESM ensemble (Fig. 11b).*~~most models (Fig. 11 (b))~~. ~~MPI, which shows a clear dipole between the Peninsula and East Antarctica (Fig. 10) surprisingly does not show a warming trend at Larissa. This suggests that further work is needed to diagnose the changes in SAM in those models, and the response of SAM to ozone and greenhouse gas forcing~~. Additionally, three out of twelve CESM simulations indicate cooling in West Antarctica, which is coherent with the hypothesis that the *part of of* observed warming is due to unforced variability and that models are not expected to match this trend perfectly. The warming at Mill Island is relatively well reproduced. However, none of the models can reproduce the *muted recent warming*~~weak cooling~~ seen at Styx. The lower spatial representativity of this site (Fig. 10) lead*s* us to interpret this as local processes missing in low resolution GCMs, such as the correct topography to account for the katabatic wind forcing, rather than a general failure of models to represent reality.

~~To sum up, the 1950 to 2005 trend at Larissa of 0.29°C/10-years is a useful benchmark for climate models to test their response to the Ozone forcing and the temperature pattern associated with the SAM index and other modes of variability influencing the Peninsula. The trend at Mill Island of 0.14°C/10-years is a useful target to ensure that Antarctica is not warming too much in response to Greenhouse gas forcing.~~

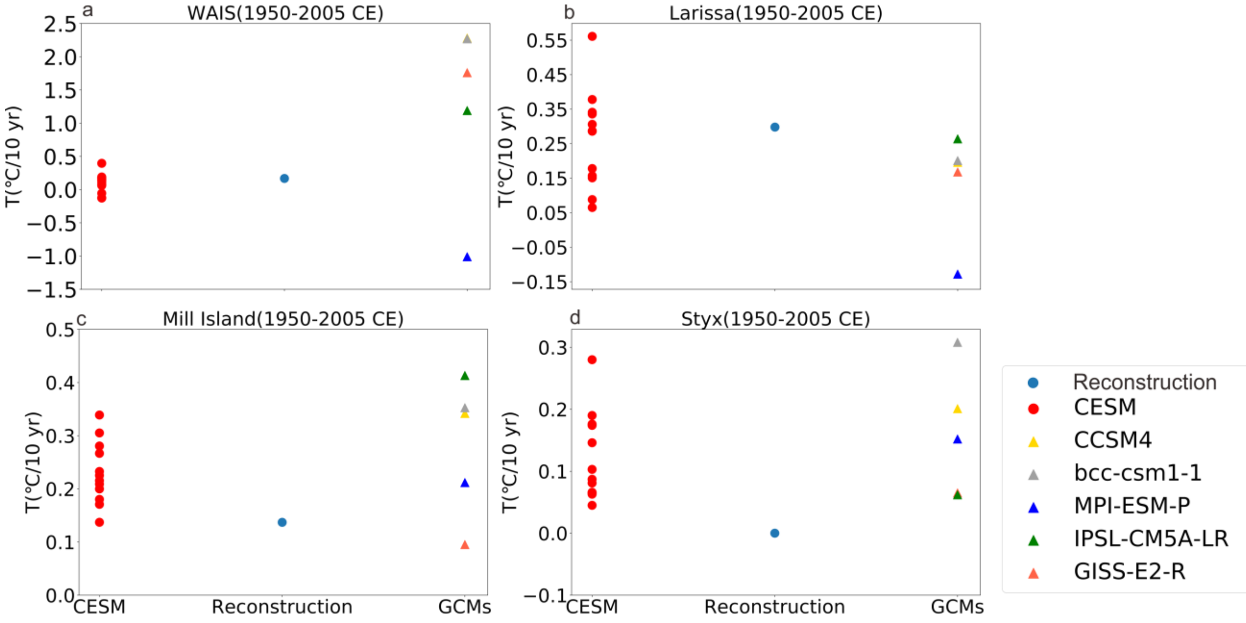

Figure 11. Linear trends for the four boreholes over 1950 to 2005 C.E.: (a) WAIS; (b) Larissa; (c) Mill Island; (d) Styx

5. Conclusion

In this study, we test two complementary ways to evaluate the climate model performance using borehole temperature observations. The standard way is to compare the reconstruction of surface temperature with simulated values in the time domain. The successful application here of a forward model driven with climate model results provides an additional way to analyze jointly model results and borehole temperature measurements. Compared to the model-data comparison in the time domain, the forward model allows us to reproduce the subsurface temperature profiles and to compare them directly with measured borehole temperature profiles.

The comparison of the surface temperature time series is simpler and more straightforward but it is limited by the different resolutions of the reconstructions and climate model results. Nevertheless, some robust conclusions can be derived from this model-data comparison that ~~are~~ *is* confirmed by the direct analyses of the temperature profiles as a function of the depth. For instance, the long-term cooling trend over last millennium observed at WAIS is relatively well reproduced in all models but with a weaker amplitude, which means the model maybe miss some feedbacks or low-frequency internal variability. Most simulations agree with data on a recent warming but the magnitude and timing vary a lot between models for the four sites. The large variability of the trends over the 20th century within the CESM ensemble for WAIS and Larissa suggests that many apparent model disagreements for those sites can be due to internal variability while the disagreement for Styx and Mill Island may be related to local processes not captured by global models.

The comparison of the model output and data in the depth domain is useful because the borehole temperature inversion is an under-determined problem, and many different temperature histories could fit the data equally well. The comparison of the temperature profiles confirms the conclusions found in the time domain, and validates the significance of some of the differences found. Some features are however difficult to interpret, such as the depth of the temperature minimum at the WAIS Divide site, which is not in the same order (deeper = older) as the timing of the temperature minimum between simulations. This points to the complexity of the interpretation of the borehole profiles, and the complementary use of the analyses in the depth and time domain.

Finally, some metrics derived from the corresponding reconstructions are proposed to be used more widely in model evaluation. The metrics used are demonstrated to be representative of a large spatial area, although they are calculated at a specific site. The results confirm that no models can reproduce the cooling during 19th over the AP and *a stabilization of the temperature over last 50-years in* *northern Victoria Land*~~the weak warming over last 50-years in northern Victoria Land.~~ Nevertheless, these models can capture the larger long-term cooling from 1000 to 1600 C.E. in West Antarctica, and the recent 50 years warming in West Antarctica and AP. This work brings quantitative tools to evaluate models and better simulate the Antarctic climate and its response to forcings.

Data availability The PMIP3/CMIP5 model results can be downloaded online from the Program for Climate Model Diagnosis and Intercomparison (PCMDI; http://pcmdi9.llnl.gov, last access: 20 April 2018). The Forward Model is available by request to Anais Orsi (anais.orsi@lsce.ipsl.fr).

Competing interests The authors declare that there is no conflict of interest.

Acknowledgements We acknowledge the World Climate Research Programme Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model outputs. This work was supported by the ~~–~~Chinese Government Scholarships (grant no. 201806040211) and the Belgian Research Action through Interdisciplinary Networks (BRAIN-be) from Belgian Science Policy Office in the framework of the project "East Antarctic surface mass balance in the Anthropocene: observations and multiscale modelling (Mass2Ant)" (Contrat n◦ 15 BR/165/A2/Mass2Ant). Hugues Goosse is the research director within the F.R.S.-FNRS. A. Orsi was supported by the french national programme LEFE/INSU ABN2K.

References

Abram N.J., H. V. McGregor, J. E. Tierney, M.N. Evans, N.P. McKay, D.S. Kaufman and the PAGES 2k Consortium, Early onset of industrial-era warming across the oceans and continents. Nature, 536, doi: https://doi.org/10.1038/nature19082, 411-418, 2016

Alley, R. B. and Koci, B. R.: Recent warming in central Greenland?, Ann Glaciol, 14(1977), 6–8, doi:https://doi.org/10.3189/S0260305500008144, 1990.

Barrett, B. E., Nicholls, K. W., Murray, T., Smith, A. M. and Vaughan, D. G.: Rapid recent warming on Rutford Ice Stream, West Antarctica, from borehole thermometry, Geophys Res Lett, 36(2), 1–5, doi:10.1029/2008GL036369, 2009.

Beltrami, H., Ferguson, G. and Harris, R. N.: Long-term tracking of climate change by underground temperatures, Geophys Res Lett, 32(19), 1–4, doi:10.1029/2005GL023714, 2005.

Beltrami, H., González-Rouco, J. F. and Stevens, M. B.: Subsurface temperatures during the last millennium: Model and observation, Geophys Res Lett, 33(9), 2–5, doi:10.1029/2006GL026050, 2006.

Bodri, Louise, and Vladimir Cermak. Borehole climatology: a new method how to reconstruct climate. Elsevier, 2011.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B. and Zhao, Y.: Evaluation of climate models using palaeoclimatic data, Nat Clim Chang, 2(6), 417–424, doi:10.1038/nclimate1456, 2012

Chapman, W. L. and Walsh, J. E.: A synthesis of Antarctic temperatures, J Clim, 20(16), 4096–4117, doi:10.1175/JCLI4236.1, 2007.

Cuffey, K., and W. Paterson, The Physics of Glaciers, Academic, Amsterdam, 2010.

Dahl-Jensen, D., Morgan, V. I. and Elcheikh, A.: Monte Carlo inverse modelling of the Law Dome (Antarctica) temperature profile, Ann Glaciol, 29(1), 145–150, doi: https://doi.org/10.3189/172756499781821102, 1999.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J. and Thompson: PRYSM: An open-source framework for Proxy System Modeling, with applications to oxygen-isotope systems, J Adv Model Earth Syst, 6, 513–526, doi:10.1002/2013MS000282.Received, 2014.

Dufresne, J. L., Foujols, M., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Jous- saume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahel- lec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, Clim. Dyn., 40, 2123–2165, https://doi.org/10.1007/s00382-012- 1636-1, 2013.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M. and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, Quat Sci Rev, 76, 16–28, doi: 10.1016/j.quascirev. 2013.

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, Environ Res Lett, 11(4), doi:10.1088/1748-9326/11/4/044022, 2016.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z.-L., and Zhang, M.: The Community Climate System Model Version 4, J. Clim., 24, 4973–4991, https://doi.org/10.1175/2011JCLI4083.1, 2011.

González-Rouco, F., Von Storch, H. and Zorita, E.: Deep soil temperature as proxy for surface airtemperature in a coupled model simulation of the last thousand years, Geophys Res Lett, 30(21), 1–4, doi:10.1029/2003GL018264, 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(1), 2–5, doi:10.1029/2005GL024693, 2006.

Goosse, H., Braida, M., Crosta, X., Mairesse, A., Masson-Delmotte, V., Mathiot, P., Neukom, R., Oerter, H., Philippon, G., Renssen, H., Stenni, B., van Ommen, T. and Verleyen, E.: Antarctic temperature changes during the last millennium: evaluation of simulations and reconstructions, Quat Sci Rev, 55, 75–90, doi:10.1016/j.quascirev.2012.09.003, 2012.

Goosse, H., Crowley, T.J., Zorita, E., Ammann, C.M., Renssen, H. and Driesschaert, E.: Modelling the climate of the last millennium: What causes the differences between simulations?. Geophysical Research Letters, 32(6), https:// doi.org/10.1029/2005GL022368, 2005.

Harris, R. N. and Gosnold, W. D.: Comparisons of borehole temperature – depth profiles and surface air temperatures in the northern plains of the USA, Geophys J Int, 138(2), 541–548, doi: doi.org/10.1046/j.1365-246X. 1999.

Jones, J. M., Gille, S. T., Goosse, H., Abram, N. J., Canziani, P. O., Charman, D. J., Clem, K. R., Crosta, X., De Lavergne, C., Eisenman, I., England, M. H., Fogt, R. L., Frankcombe, L. M., Marshall, G. J., Masson-Delmotte, V., Morrison, A. K., Orsi, A. J., Raphael, M. N., Renwick, J. A., Schneider, D. P., Simpkins, G. R., Steig, E. J., Stenni, B., Swingedouw, D. and Vance, T. R.: Assessing recent trends in high-latitude Southern Hemisphere surface climate, Nat Clim Chang, 6(10), 917–926, doi:10.1038/nclimate3103, 2016.

Klein, F., Abram, N. J., Curran, M. A. J., Goosse, H., Goursaud, S., Masson-delmotte, V., Moy, A., Neukom, R., Orsi, A., Sjolte, J., Steiger, N., Stenni, B. and Werner, M.: Assessing the robustness of Antarctic temperature reconstructions over the past two millennia , Clim Past, 661–684, doi: doi.org/10.5194/cp-15-661-2019, 2019.

Klein, F., Goosse, H., Graham, N. E. and Verschuren, D.: Comparison of simulated and reconstructed variations in East African hydroclimate over the last millennium, Clim Past, 12(7), 1499–1518, doi:10.5194/cp-12-1499-2016, 2016.

Ljungqvist, F. C., Zhang, Q., Brattström, G., Krusic, P. J., Seim, A., Li, Q., ... & Moberg, A.. Centennial-Scale Temperature Change in Last Millennium Simulations and Proxy-Based Reconstructions. Journal of Climate, 32(9), 2441-2482. https://doi.org/10.1175/JCLI-D-18-0525.1, 2019.

Muto, A., Scambos, T. A., Steffen, K., Slater, A. G. and Clow, G. D.: Recent surface temperature trends in the interior of East Antarctica from borehole firn temperature measurements and geophysical inverse methods, Geophys Res Lett, 38(15), 6-11, doi: 10.1029/2011GL048086, 2011.

Nagornov, O. V., Konovalov, Y. V. and Tchijov, V.: Temperature reconstruction for Arctic glaciers, Palaeogeogr Palaeoclimatol Palaeoecol, 236(1–2), 125-134, doi: 10.1016/j.palaeo.2005.11.035, 2006.

Nagornov, O. V., Konovalov, Y. V., Zagorodnov, V. S., & Thompson, L. G. Reconstruction of the surface temperature of Arctic glaciers from the data of temperature measurements in wells. Journal of engineering physics and thermophysics, 74(2), 253-265, doi: doi.org/10.1023/A:101666161, 2001.

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., González-Rouco, F., Linsley, B. K., Moy, A. D., Mundo, I., Raible, C. C., Steig, E. J., Van Ommen, T., Vance, T., Villalba, R., Zinke, J. and Frank, D.: Inter-hemispheric temperature variability over the past millennium, Nat Clim Chang, 4(5), 362–367, doi:10.1038/nclimate2174, 2014.

Nicolas, J. P. and Bromwich, D. H.: New reconstruction of Antarctic near-surface temperatures: Multidecadal trends and reliability of global reanalyses, J Clim, 27(21), 8070–8093, doi:10.1175/JCLI-D-13-00733.1, 2014.

Orsi, A. J., Cornuelle, B. D. and Severinghaus, J. P.: Little Ice Age cold interval in West Antarctica: Evidence from borehole temperature at the West Antarctic Ice Sheet (WAIS) Divide, Geophys Res Lett, 39(9), 1–7, doi:10.1029/2012GL051260, 2012.

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 C.E. An ensemble approach with the Community Earth System Model (CESM), B. Am. Meteor. Soc., 97, 735–754, https://doi.org/10.1175/BAMS- D-14-00233.1, 2016.

Rath, V., Rouco, J. F. G. and Goosse, H.: Impact of postglacial warming on borehole reconstructions of last millennium temperatures, Clim Past, 8(3), 1059–1066, doi:10.5194/cp-8-1059-2012, 2012.

Renssen, H., Goosse, H., Fichefeti, T., Masson-delmotte, V., Ko, N., Lemaftre, G., Louvain, D. and Cyclotron, C.: Holocene climate evolution in the high-latitude Southern Hemisphere simulated by a coupled atmosphere-sea ice-ocean-vegetation model, , 7, 951–964, 2005.

Roberts, J. L., Moy, A. D., Van Ommen, T. D., Curran, M. A. J., Worby, A. P., Goodwin, I. D. and Inoue, M.: Borehole temperatures reveal a changed energy budget at Mill Island, East Antarctica, over recent decades, Cryosphere, 7(1), 263–273, doi:10.5194/tc-7-263-2013, 2013.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Bra- connot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Cli- mate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), Geosci. Model Dev., 5, 185–191, https://doi.org/10.5194/gmd-5-185-2012, 2012.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P. C. T. J. D. G., Crowley, T. J., ... & Otto-Bliesner, B. L. Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1. 0). Geoscientific Model Development, 4(1), 33-45, doi: 10.5194/gmd-4-33-2011, 2011.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck, R., Canuto, V., Chen, Y., Cheng, Y., Clune, T. L., Genio, A. D., Fainchtein, R. D., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A., Legrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Put- man,W. M., Rund, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S., and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, J. Advan. in Mode. Earth Syst., 6, 141–184, https://doi.org/10.1002/2013MS000265, 2014.

Schneider, D. P., Steig, E. J., van Ommen, T. D., Dixon, D. A., Mayewski, P. A., Jones, J. M., & Bitz, C. M. Antarctic temperatures over the past two centuries from ice cores. Geophysical Research Letters, 33(16). doi:10.1029/2006GL027057, 2006.

Smith, K. L. and Polvani, L. M.: Spatial patterns of recent Antarctic surface temperature trends and the importance of natural variability: lessons from multiple reconstructions and the CMIP5 models, Clim Dyn, 48(7–8), 2653–2670, doi:10.1007/s00382-016-3230-4, 2017.

Smith, K. L., & Polvani, L. M. Spatial patterns of recent Antarctic surface temperature trends and the importance of natural variability: lessons from multiple reconstructions and the CMIP5 models. Climate dynamics, 48(7-8), 2653-2670. https://doi.org/10.1007/s00382-016-3230-4, 2017.

Steig, E. J., Schneider, D. P., Rutherford, S. D., Mann, M. E., Comiso, J. C. and Shindell, D. T.: Warming of the Antarctic ice-sheet surface since the 1957 International Geophysical Year, Nature, 457(7228), 459–462, doi:10.1038/nature07669, 2009.

Stenni, B., Curran, M. A. J., Abram, N. J., Orsi, A., Goursaud, S., Masson-Delmotte, V., Neukom, R., Goosse, H., Divine, D., Van Ommen, T., Steig, E. J., Dixon, D. A., Thomas, E. R., Bertler, N. A. N., Isaksson, E., Ekaykin, A., Werner, M. and Frezzotti, M.: Antarctic climate variability on regional and continental scales over the last 2000 years, Clim Past, 13(11), 1609–1634, doi:10.5194/cp-13-1609-2017, 2017.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: The atmospheric component of the MPI-M earth system model: ECHAM6, J. Adv. Model. Earth Syst., 5, 1–27, https://doi.org/10.1002/jame.20015, 2013.

Stevens, M. B., Gonza, J. F. and Beltrami, H.: North American climate of the last millennium: Underground temperatures and model comparison, 113, 1–15, doi:10.1029/2006JF000705, 2008.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Turner, J., Colwell, S. R., Marshall, G. J., Lachlan-Cope, T. A., Carleton, A. M., Jones, P. D., Lagun, V., Reid, P. A. and Iagovkina, S.: Antarctic climate change during the last 50 years, Int J Climatol, 25(3), 279–294, doi:10.1002/joc.1130, 2005.

Wu, T., Song, L., Li, W., Wang, Z., Zhang, H., Xin, X., Zhang, Y., Zhang, L., Li, J., Wu, F., Liu, Y., Zhang, F., Shi, X., Chu, M., Zhang, J., Fang, Y., Wang, F., Lu, Y., Liu, X., Wei, M., Liu, Q., Zhou, W., Dong, M., Zhao, Q., Ji, J., Li, L., and Zhou, M.: An overview of BCC climate system model development and application for climate change studies, J. Met. Res., 28, 34–56, https://doi.org/10.1007/s13351-014-3041-7, 2014.

Yang, J. W., Han, Y., Orsi, A. J., Kim, S. J., Han, H., Ryu, Y., Jang, Y., Moon, J., Choi, T., Hur, S. Do and Ahn, J.: Surface Temperature in Twentieth Century at the Styx Glacier, Northern Victoria Land, Antarctica, From Borehole Thermometry, Geophys Res Lett, 45(18), 9834–9842, doi:10.1029/2018GL078770, 2018.

Zagorodnov, V., Nagornov, O., Scambos, T. A., Muto, A., Mosley-Thompson, E., Pettit, E. C. and Tyuflin, S.: Borehole temperatures reveal details of 20th century warming at Bruce Plateau, Antarctic Peninsula, Cryosphere, 6(3), 675–686, doi:10.5194/tc-6-675-2012, 2012.

Supplementary material

S1: Forward model description

The equation ruling subsurface temperature evolution in the forward model is given in Eq. 1. According to the original publications, we applied different methods to determine the density profile for each borehole in the model. For WAIS and Styx, the density profiles, $\rho(z)$*, were obtained by a quadratic fit to the measured bulk density data following Severinghaus et al. (2010). For Larissa, the density profile was approximated following Salamatin (2000). For Mill Island, because of the similarity between the density profiles at Mill*

Island and Law Dome (van Ommen et al., 1999), the density is described by a piecewise exponential plus linear or dual exponential according to the analysis on the Law Dome ice core density profile (van Ommen et al., 1999). The density is considered to be constant in time in the model.

For the other parameters in the forward model, the specific heat capacity c_p is calculated by c_p = 152.5+7.122T (J kg^{-1} K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.1 where T is the temperature). The thermal conductivity in ice is taken from $K_{ice} = 9.828 \exp(-5.7 \times 10^{-3} T)$ (Wm^{-1}K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.2), and the thermal conductivity of the firn is calculated by Schwerdtfeger formula (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.4). The vertical velocity at the surface is simply the accumulation rate and decreases with depth as the integral of the densification process (compaction) and the strain due to ice flow divergence. The vertical velocity profile is determined by the method of Alley et al. (1990) and Cuffey et al. (1994) with a constant strain rate. For the accumulation rate, we use a constant value derived from their original publication, which is specified in the Table 3 of the main text. The bottom boundary condition is given by the basal heat flux and the basal temperature. The heat flux is determined by matching the slope of the temperature increase in the bottom section of the record. At Mill Island, this was not possible, because the data do not extend very deep with respect to the total ice thickness. A zero heat flux boundary condition was chosen instead. The validity of this hypothesis is demonstrated in the original study of Roberts et al. (2013). The basal temperature is determined using the lower "undisturbed" sections of the measured borehole temperature extrapolated to the bottom.

In order to save computation time, the vertical discretization of the model is not homogenous. For WAIS, which is the only very deep borehole, the vertical step is of 1 m for the upper 500 m and up to 25 m for the deepest part. For other sites where the depth of borehole is close or less than 500 m, the step is set to 1 m for the whole depth range.

Before the forward model is driven by the climate model results, it is initialized with a stationary profile, which is generated after a 20000-year model run with a constant climate history and a realistic seasonal cycle. Seasonal-scale variations are "undetectable below a depth of 20m" (Cuffey and Paterson, 2010), and its does not change throughout the run. At WAIS and Styx, the seasonal cycles are determined from weather station data; at Larissa and Mill Island, since the original studies do not give the seasonal cycle, we use a seasonal cycle amplitude of 10 °C similar to WAIS (Eq. S1). At WAIS, it includes a periodic function with annual and semi-annual components, fitted to 3 years of weather station data from WAIS Divide and Byrd station (AMRC, SSEC, UW-Madison) as follows (Orsi et al., 2012):

$$T(t) = 10(\cos(2\pi t) + 0.3\cos(4\pi t)) \text{ (in K)} \qquad \text{(S1)}$$

At Styx, the seasonal cycle is determined by fitting a sinusoidal function to the automated weather station data as follows (Yang et al., 2018):

$$T(t) = 10(\cos(2\pi t) + 0.35\cos(4\pi t)) \text{ (in K)} \qquad \text{(S2)}$$

Where t is time, T is the temperature.

Equations S1 and S2 for WAIS and STYX are nearly identical, so we presume the seasonal cycle is also similar at Larissa and Mill Island, where no seasonal data is available. Including a seasonal cycle wave is important because the heat capacity and thermal conductivity depend on temperature, and temperature changes a lot in the top 15m, but below that, it is of negligible effect.

S2: Supplementary figures

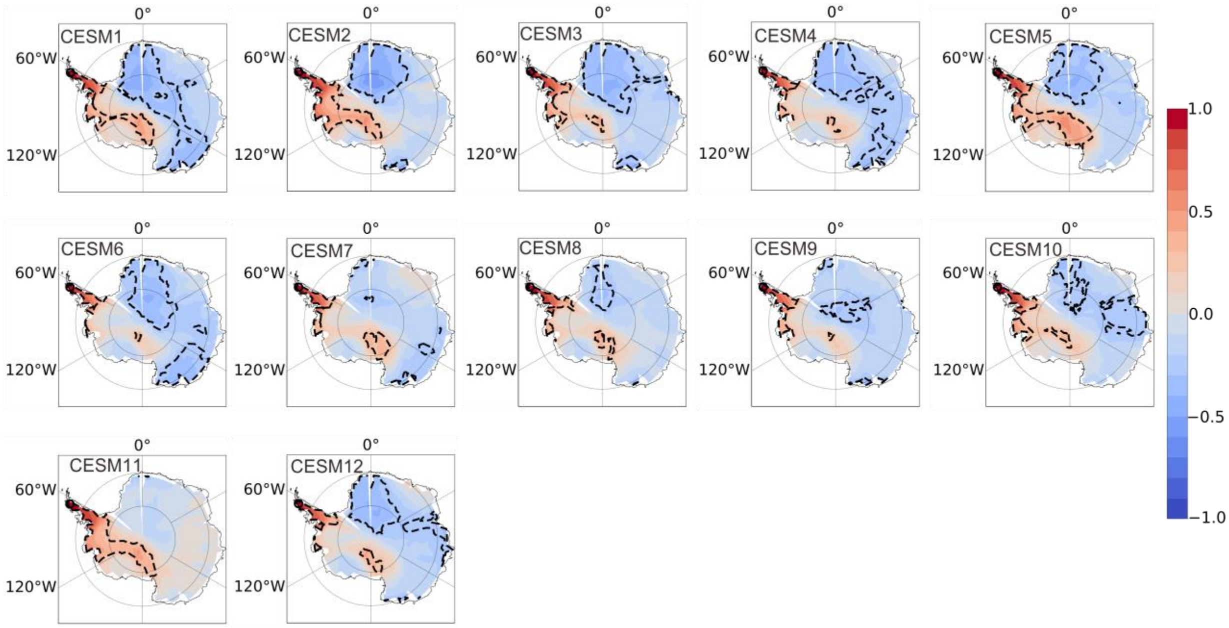

Figure S1. *The correlation map (blue-red shading area) showing the relationship between the temperature from 1825 to 1925 CE at Larissa and other grid cells in Antarctica for each CESM member. The black dotted contour lines show a significant correlation at the 99 % significant level.*

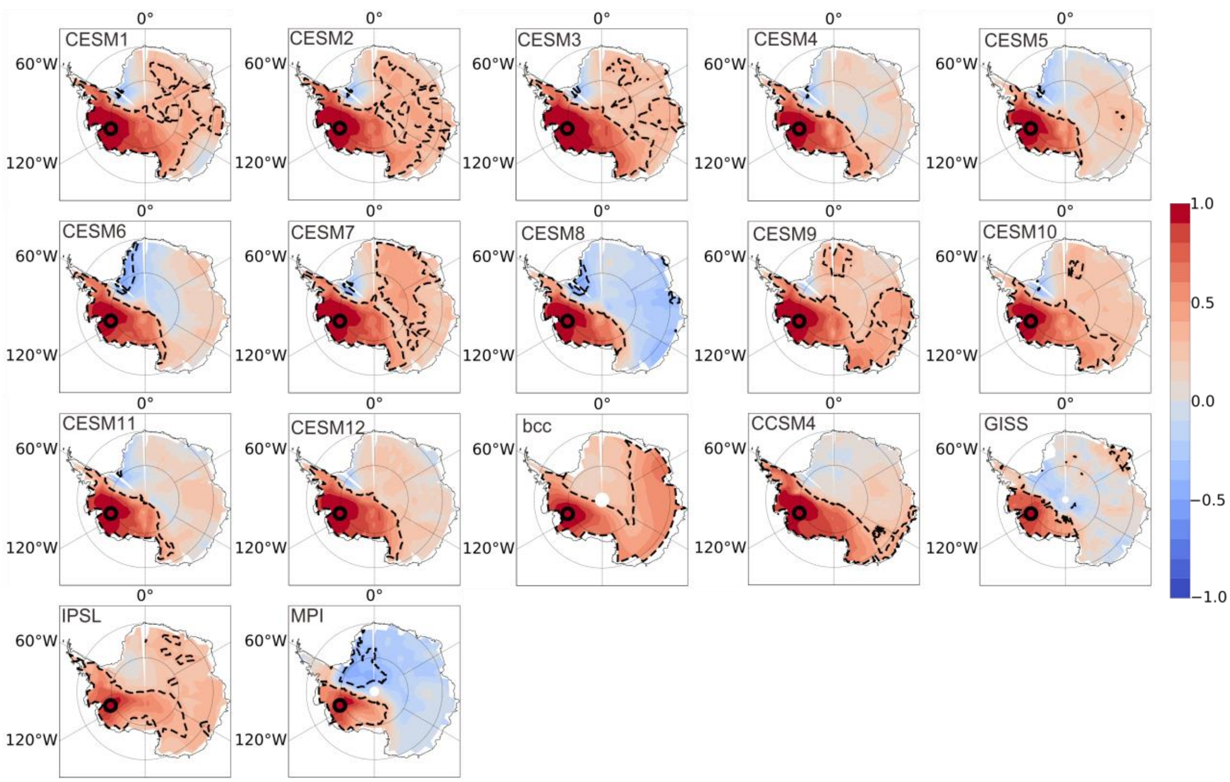

Figure S2. *The correlation map showing the relationship between the temperature from 1950 to 2005 CE at WAIS and other grid cells for each climate models. The red dashed contour lines show a significant correlation at the 99% significant level.*

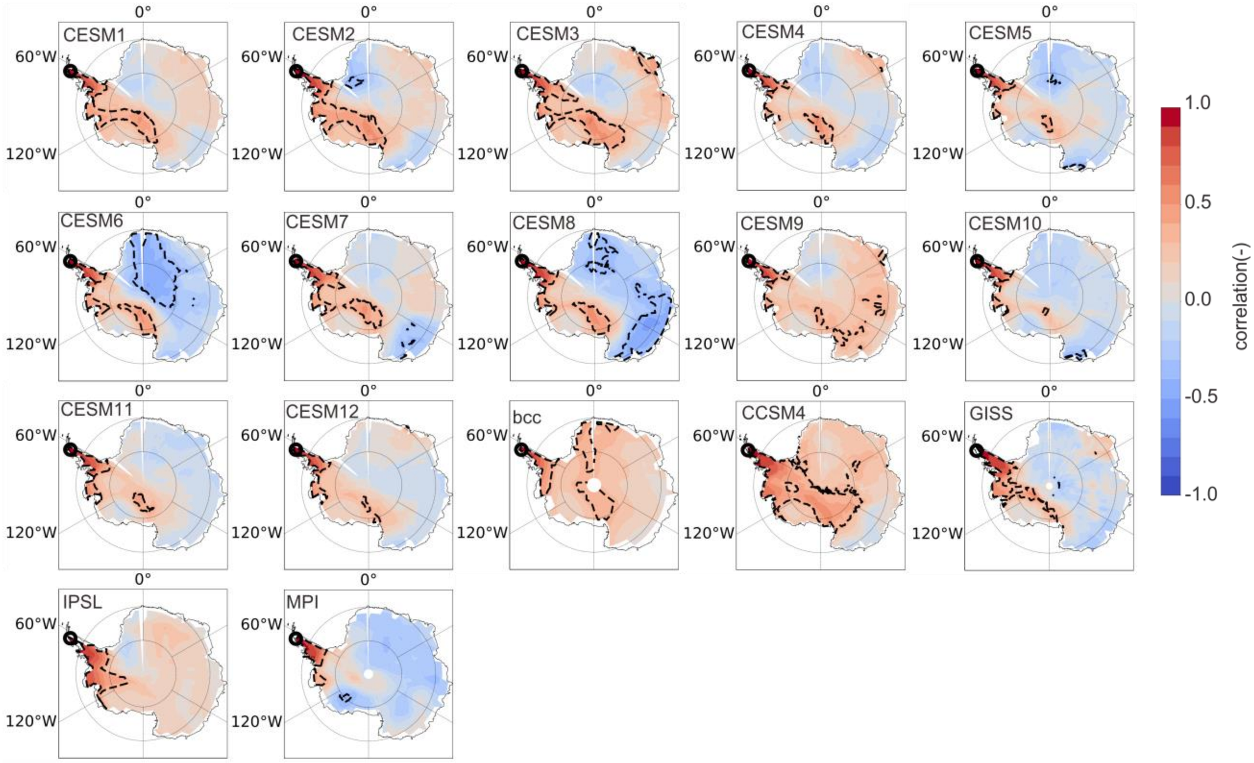

Figure S3. *The correlation map showing the relationship between the temperature from 1950 to 2005 CE at Larissa and other grid cells for each climate models. The red dashed contour lines show a significant correlation at the 99% significant level.*

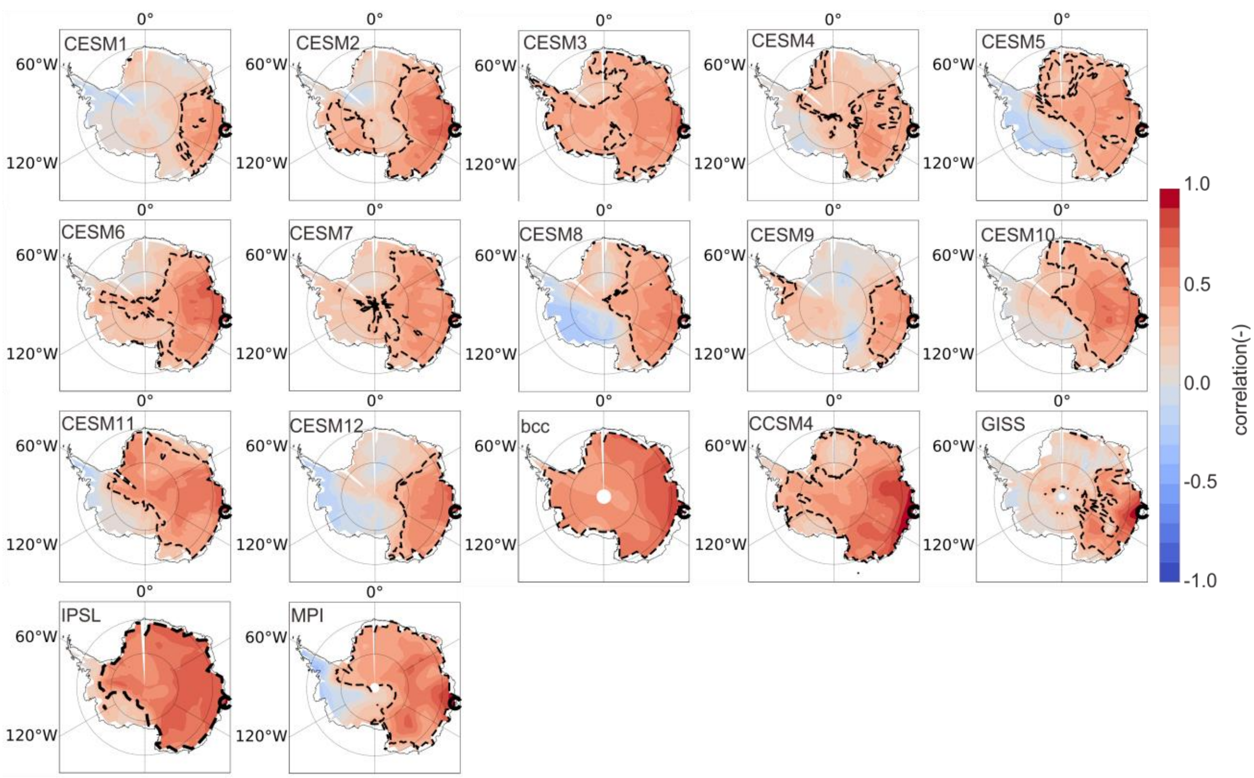

Figure S4. *The correlation map showing the relationship between the temperature from 1950 to 2005 CE at Mill Island and other grid cells for each climate models. The red dashed contour lines show a significant correlation at the 99% significant level.*

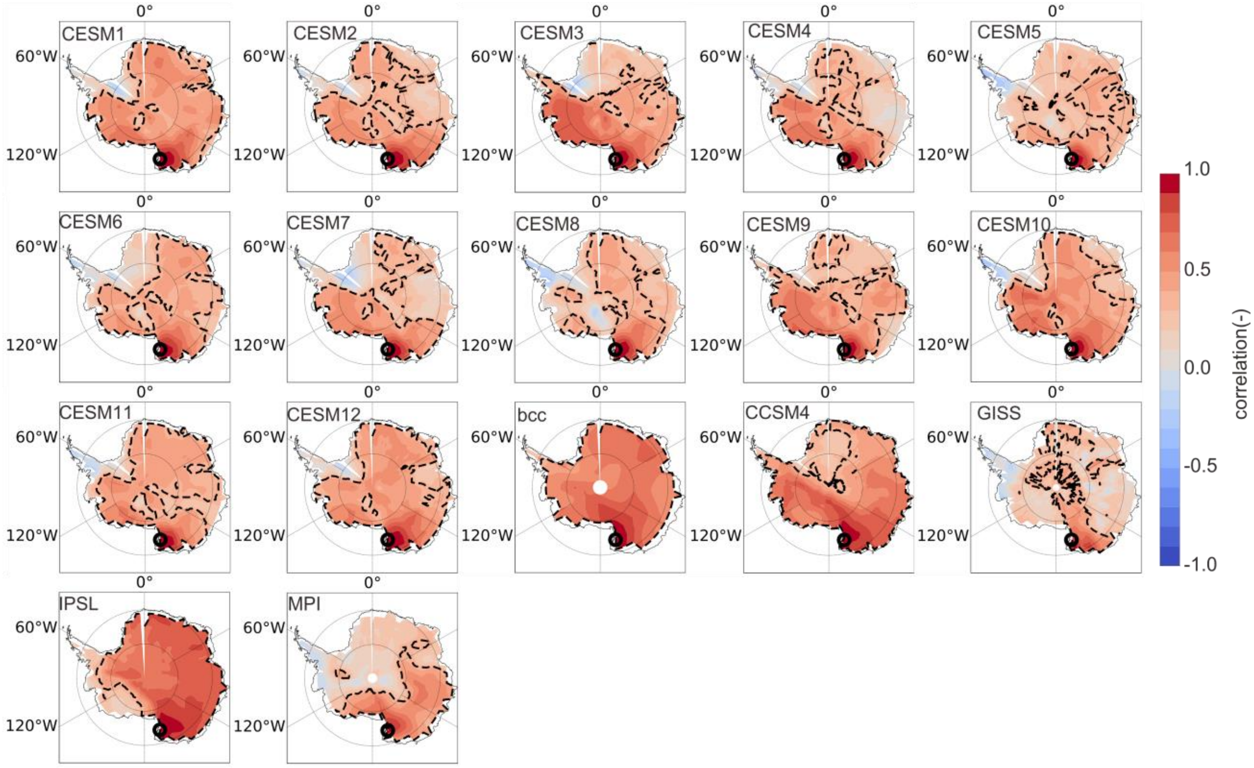

Figure S5. The correlation map showing the relationship between the temperature from 1950 to 2005 CE at Styx and other grid cells for each climate models. The red dashed contour lines show a significant correlation at the 99% significant level.

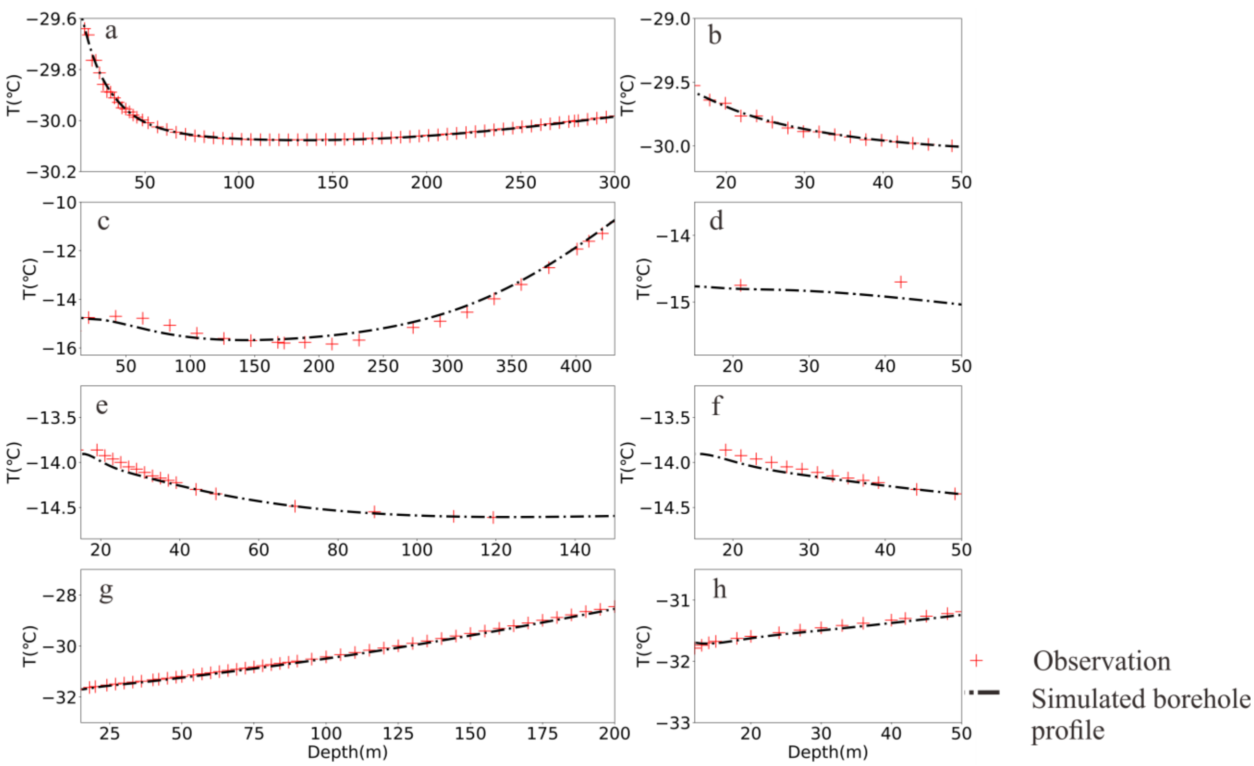

Figure S6. Comparison of borehole temperature profile outputs for the forward model driven by the corresponding reconstruction with the observation at each site. (a) WAIS: 15-300 m; (b) WAIS: 15-50 m; (c) Larissa: 15-430 m; (d) Larissa:15-50 m; (e) Mill Island:15-150 m; (f) Mill Island:15-50 m; (g) Styx:15-200 m; (h) Styx:15-50 m. The thick dash-dot line denotes the simulated borehole profile at each site, and red across represent the observation.

S3: Observed borehole temperature distribution at WAIS.

Depth(m)	*Temperature (°C)*
7.96	-29.7594
9.19	-29.604
9.95	-29.5588
11.94	-29.4741
13.93	-29.4806
15.92	-29.5267
17.91	-29.6397
19.9	-29.6645
21.89	-29.7642
23.89	-29.7642
25.88	-29.8132
27.87	-29.8587
29.86	-29.8871
31.85	-29.8891
33.84	-29.9115
35.83	-29.9285
37.82	-29.9503
39.81	-29.9549
41.8	-29.9663
43.79	-29.9785
45.78	-29.9872
48.77	-29.998
51.75	-30.0089
56.73	-30.0253
61.7	-30.0352
66.68	-30.046
71.65	-30.0538
76.63	-30.0603
81.6	-30.0635
86.58	-30.0682
91.55	-30.0701
96.53	-30.0724
101.5	-30.0738
106.48	-30.0746
111.55	-30.0753
116.43	-30.0755
121.4	-30.0757
126.38	-30.0756
131.35	-30.0753
136.33	-30.0752
141.3	-30.0748
146.28	-30.0743

151.25	*-30.0736*
156.23	*-30.0734*
161.2	*-30.0722*
166.18	*-30.0715*
171.15	*-30.0698*
176.13	*-30.0686*
181.1	*-30.0672*
186.08	*-30.0653*
191.05	*-30.0632*
196.03	*-30.0608*
201	*-30.0584*
205.98	*-30.0564*
210.95	*-30.0532*
215.93	*-30.0502*
220.9	*-30.0471*
225.88	*-30.0436*
230.85	*-30.0404*
235.83	*-30.0365*
240.8	*-30.0329*
245.78	*-30.0299*
250.75	*-30.0248*
255.73	*-30.022*
260.7	*-30.0165*
265.68	*-30.0129*
270.65	*-30.0078*
275.63	*-30.0042*
278.61	*-30.0017*
280.6	*-30.0003*
285.58	*-29.9954*
290.55	*-29.991*
295.53	*-29.9863*
300.5	*-29.9821*

S4: Observed borehole temperature distribution at LARISSA.

Depth(m)	***Temperature (°C)***
8.41	*-15.35*
10.51	*-15.3*
21.02	*-14.75*
42.04	*-14.7*
63.06	*-14.78*
84.09	*-15.07*

105.11	*-15.4*
126.13	*-15.61*
147.15	*-15.7*
168.17	*-15.77*
173	*-15.8*
189.19	*-15.77*
210.21	*-15.84*
231.24	*-15.68*
273.28	*-15.16*
294.3	*-14.9*
315.32	*-14.53*
336.34	*-13.98*
357.36	*-13.39*
378.86	*-12.7*
400.77	*-11.93*
409.92	*-11.61*
420.43	*-11.29*
430.94	*-10.82*

S5: Observed borehole temperature distribution at Mill Island.

Depth(m)	***Temperature (°C)***
9.05	*-14.275*
14.06	*-13.8625*
19.07	*-13.8625*
21.07	*-13.925*
23.07	*-13.9625*
25.07	*-14*
27.07	*-14.05*
29.07	*-14.075*
31.09	*-14.1125*
33.09	*-14.15*
35.11	*-14.175*
37.11	*-14.2*
39.11	*-14.225*
44.125	*-14.3*
49.14	*-14.35*
69.17	*-14.4875*
89.24	*-14.55*
109.3	*-14.6*
119.31	*-14.6125*

S6: Observed borehole temperature distribution at Styx.

Depth(m)	***Temperature (°C)***
1	*-29.7244*

2	*-32.6116*
3	*-33.4131*
4	*-33.522*
5	*-33.2036*
6	*-32.9317*
7	*-32.5716*
8	*-32.3349*
9	*-32.1212*
10	*-31.9562*
11	*-31.8512*
12	*-31.7853*
13	*-31.7379*
14	*-31.6974*
15	*-31.6752*
18	*-31.6255*
20	*-31.5921*
24	*-31.5332*
27	*-31.4905*
30	*-31.452*
33	*-31.4144*
36	*-31.3781*
40	*-31.3275*
42	*-31.3006*
45	*-31.2628*
48	*-31.2209*
50	*-31.1898*
54	*-31.1366*
57	*-31.0925*
60	*-31.0493*
63	*-31.0032*
66	*-30.9585*
69	*-30.9144*
72	*-30.8683*
75	*-30.82*
78	*-30.7722*
81	*-30.7296*
84	*-30.6835*
87	*-30.6367*
90	*-30.5892*
95	*-30.5113*
100	*-30.4264*
105	*-30.3411*
110	*-30.253*
115	*-30.1648*
120	*-30.081*
125	*-29.9906*
130	*-29.8973*
135	*-29.803*

140	*-29.7075*
145	*-29.6079*
150	*-29.5078*
155	*-29.412*
160	*-29.3098*
165	*-29.2065*
170	*-29.0968*
175	*-28.9922*
180	*-28.883*
185	*-28.7783*
190	*-28.6518*
195	*-28.5633*
200	*-28.4535*
205	*-28.3431*
210	*-28.2515*

Reference

Alley, R. B., and Koci, B. R.: Recent Warming in Central Greenland?, Ann. Glaciol., 14, 6-8, https://doi.org/10.3189/s0260305500008144, 1990.

Cuffey, K. M., Alley, R. B., Grootes, P. M., Bolzan, J. M., and Anandakrishnan, S.: Calibration of the $\delta^{18}O$ isotopic paleothermometer for central Greenland, using borehole temperatures, Journal of Glaciology, 40, 341-349, 10.3189/s0022143000007425, 1994.

Cuffey, K., and W. Paterson, The Physics of Glaciers, Academic, Amsterdam, 2010.

Orsi, A. J., Cornuelle, B. D., and Severinghaus, J. P.: Little Ice Age cold interval in West Antarctica: Evidence from borehole temperature at the West Antarctic Ice Sheet (WAIS) Divide, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012gl051260, 2012.

Roberts, J. L., Moy, A. D., van Ommen, T. D., Curran, M. A. J., Worby, A. P., Goodwin, I. D., and Inoue, M.: Borehole temperatures reveal a changed energy budget at Mill Island, East Antarctica, over recent decades, Cryosphere, 7, 263-273, https://doi.org/10.5194/tc-7-263-2013, 2013.

Salamatin A. N.: Paleoclimatic reconstructions based on borehole temperature measurements in ice sheets. Possibilities and limitations, in: Physics of Ice Core Records, edited by: Hondoh, T., Hokkaido University Press, Sapporo, 243–282, 2000.

Severinghaus, J.P., Albert, M.R., Courville, Z.R., Fahnestock, M.A., Kawamura, K., Montzka, S.A., Mühle, J., Scambos, T.A., Shields, E., Shuman, C.A. and Suwa, M.: Deep air convection in the firn at a zero-accumulation site, central Antarctica, Earth Planet. Sci. Lett., 293(3-4), 359-367, https://doi.org/10.1016/j.epsl.2010.03.003, 2010.

van Ommen, T., Morgan, V., Jacka, T., Woon, S., and Elcheikh, A.: Near-surface temperatures in the Dome Summit South (Law Dome, East Antarctica) borehole, Ann. Glaciol., 29, 141–144, https://doi.org/10.3189/172756499781821382 ,1999.

Yang, J.-W., Han, Y., Orsi, A. J., Kim, S.-J., Han, H., Ryu, Y., Jang, Y., Moon, J., Choi, T., Hur, S. D., and Ahn, J.: Surface Temperature in Twentieth Century at the Styx Glacier, Northern Victoria Land, Antarctica, From Borehole Thermometry, Geophys. Res. Lett., 45, 9834-9842, https://doi.org/10.1029/2018gl078770, 2018.