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 $K_{ice} = 9.828\exp(-5.7\times10^{-3}T)$ (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 $K_{schwerdtfeger} = \frac{2K_{ice}\rho}{3\rho_{ice}-\rho}$ (Wm^{-1}K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.4, ρ_{ice} = 916.5-0.14 T (kg/m^3), ρ is the density of firn/ice). 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. It is determined from weather station data. 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):

$T(t) = 10(\cos(2\pi t) + 0.3\cos(4\pi t))$ (in K)   (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):

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

”

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 $K_{ice} = 9.828 \exp(-5.7 \times 10^{-3} T)$ (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 $K_{schwerdtfeger} = \frac{2K_{ice}\rho}{3\rho_{ice}-\rho}$ (Wm^{-1}K^{-1}) (Cuffey and Paterson, 2010, Chap. 9, Eq. 9.4, ρ_{ice} = $916.5 - 0.14\,T$ (kg/m^3), ρ is the density of firn/ice). 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.

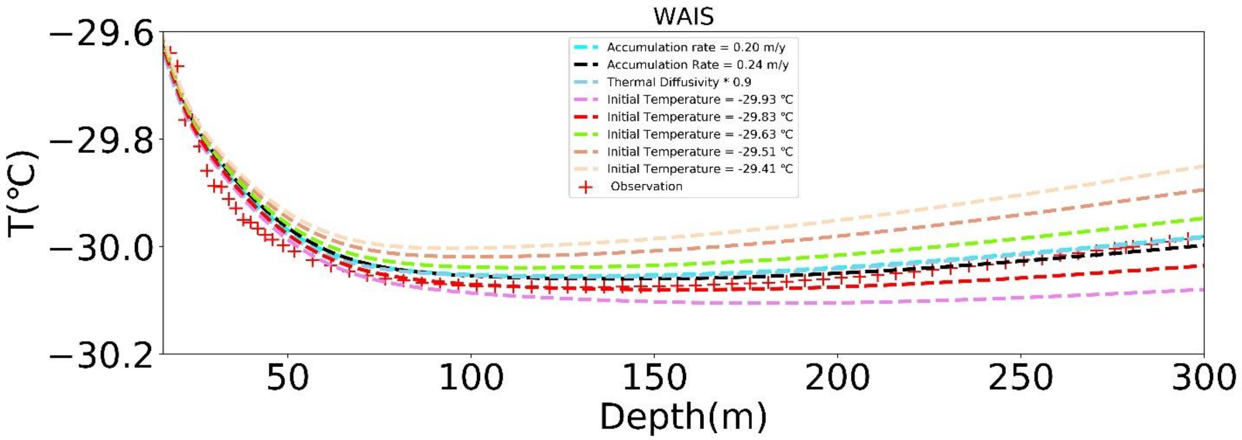

Figure R1. The sensitivity tests using the temperature history of one CESM member (dashed lines) at WAIS.

(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.