Included in this document:

1) Reply to Referee Comments (pages 1-43)

2) Summary of Changes made to the Manuscript (page 44)

3) Marked-up Manuscript (starts after page 44) *Note that the page numbers start over in this section.

1) Reply to Referee Comments

We thank all three referees for their time and insights. In the following, we address each of the referees' comments (in black) and, where applicable, include the edited section of the paper. Our replies are inline in blue and edited sections are indented. Please note that some edits were influenced by multiple referee comments.

Referee #1

Referee #1 General Comment

The authors present a new reconstruction of temperature and precipitation over Greenland covering the past 20 000 years using for the first time over such a long period a data assimilation technique successfully applied recently over the past millennia. The paper is very clear, justify nearly all the choices in a very rigorous way and provides comprehensive estimates of the uncertainties. I have thus no doubt that, in addition to the new reconstruction that can be used for instance to drive ice sheet models, this study opens new fields of application of data assimilation of multi-millennial timescales.

Thank you.

However, I consider that the impact of the choice of the prior is not enough discussed and this issue must be addressed before publication. If I understand well, the prior ensemble is made of 100 states obtained by averaging 50 years of model data. Those states are selected randomly over the full length of the simulation (line 175). This method is reasonable if the climate variations are weak, such as during the past millennia, but is it valid for very large changes as observed during the glacial interglacial periods? I may have missed something but, if I am right, a state obtained in the model in the late Holocene can be used to reconstruct the last glacial climate, which may be hard to justify. For instance, the authors argue that it is important to take into account the changes in seasonality of precipitation (e.g. line 233) but I wonder how this could be achieved by selecting model states that are coming from very different periods. I would suggest using as prior only years that are close to the period that is reconstructed so that only glacial states are used to reconstruct glacial climate for instance.

The reviewer's general comment broadly concerns our selection of the prior ensemble from the entire TraCE-21ka simulation rather than from specific time periods that align with the reconstruction time. This is an excellent point, which we have thought about carefully, but had not elaborated upon in the paper.

The reviewer states that, "If I understand well, the prior ensemble is made of 100 states obtained by averaging 50 years of model data. Those states are selected randomly over the full length of the simulation (line 175)." If we are understanding each other correctly, then yes, one state in the 100-member prior ensemble is an average over 50 years of the model data; these 100 states are selected randomly from the full length of the simulation. This implies that both glacial and Holocene states are likely to be contained within the same prior ensemble that is used to reconstruct all time steps over the last 20,000 years. To be clear, a prior ensemble could in principle contain only Holocene states; however, this is not the case for any of the ten prior ensembles we use in the paper. Thus, in reconstructing a time step in the glacial, for example, both glacial and Holocene states are part of the prior ensemble.

We agree with the reviewer that conditionally chosen prior ensembles would be preferable, but for the timescale under consideration this is not yet feasible. To explain the reasoning behind our choices in the paper, we elaborate on the pros and cons of four prior ensemble options we considered before deciding on the one that we use in this study (#4). In the revised supplementary information, we have included the following as Sect. S1:

S1 Prior ensemble considerations

Here we elaborate on the pros and cons of four prior ensemble options we considered before deciding on the one that we use in this study (#4).

1. For offline data assimilation (i.e., no information passed between assimilation time steps), a justifiable method for choosing the prior ensemble would be to use a 100-member ensemble of 20,000-year climate simulations. These climate simulations would be TraCE-21ka-like (i.e., results from fully-coupled GCMs at T31 resolution or higher), and have varied initial conditions, boundary conditions, and model physics. The prior ensemble for any assimilation time step would be taken from the same time step in the climate simulations, which would lead to a prior ensemble that varies smoothly in time and is a justifiable initial guess for the climate evolution over the past 20,000 years. Though this option is simple, it is not feasible because the computational cost of running even one TraCE-21ka-like simulation remains near computational limits.

2. Given that there is only one TraCE-21ka-like simulation, another method would be to select states from TraCE-21ka that are closest in time to the reconstruction time step. For example, if we were reconstructing the 50-year average centered on the year 5,000 CE, then we would select the 100 states from TraCE-21ka that are closest in time to 5,000 CE. Given that we are working with 50-year averages, this means we would select all the states between 7,500 and 2,500 CE. This method, which we call the "running-window" method, provides a prior that varies smoothly in time and is a justifiable initial estimate for the climate evolution.

For the running-window method, the variance of the prior ensemble would tend to be small. A prior with small variance would lead to underweighting of the proxy records during assimilation. To avoid this issue, we could use the well-accepted approach of inflating the prior variance (Anderson and Anderson, 1999). However, the use of inflation adds an additional tunable parameter; in this case, it would add an additional parameter per time step. Although inflation can, in principle, be constrained using the ensemble calibration ratio (computed for excluded proxies), we have too few proxy records to meaningfully constrain this parameter without overfitting.

In addition to estimating numerous inflation factors, the running-window method limits us to one estimate of the spatial covariance structure per time step. Thus, we have no way to quantify the uncertainty associated with the prior covariance structure. This could be fixed by expanding the running window and randomly selecting multiple prior ensembles; however, if the running window is expanded enough to create meaningfully different prior ensembles, then Holocene states will leak into glacial prior ensembles (and *vice versa*) and the method essentially becomes the method we use in the paper.

3. To reduce the number of inflation parameters, we could split TraCE-21ka into several distinct time periods. From these time periods, we would randomly select prior ensembles that are only used for the reconstruction of associated assimilation time steps. For example, if we split TraCE-21ka into glacial, transitional, and Holocene periods, then we'd make a glacial prior ensemble that is only used to reconstruct the glacial, a transitional prior ensemble that is only used to reconstruct the transition, and a Holocene prior ensemble that is only used to reconstruct the Holocene. This reduces the number of inflation factors we must estimate to a total of three. A disadvantage, however, is that this makes the prior discontinuous in time, which frequently leads to a discontinuous reconstruction. To adjust the reconstruction and make it continuous requires another source of information. Such post-processing adds an extra layer of complexity.

4. The method used in our study ensures a continuous reconstruction and removes the need for inflation factors. This method uses the same prior ensemble for all time steps (thus it is continuous) and the includes both glacial and Holocene states, which provides enough variance to appropriately weight the proxy records (thus no inflation is needed). Though the time-invariant prior is a poor estimate of the climate evolution over the last 20,000 years, the proxy records are given enough weight to result in a posterior that captures the large climate changes. In addition, we can quantify the uncertainty associated with the spatial covariance pattern by producing multiple posterior ensembles that each stem from a different prior ensemble. In the paper, we use ten different prior ensembles to quantify this uncertainty. Overall, this method is both feasible and simple, thus providing a first step in developing paleoclimate data assimilation for applications on glacial-interglacial timescales.

In this study, one state in a 100-member prior ensemble is an average over 50 years of the model data; these 100 states are selected randomly from the full length of the simulation. This implies that both glacial and Holocene states are likely to be contained within the same prior ensemble that is used to reconstruct all time steps over the last 20,000 years. A prior ensemble could in principle contain only Holocene states; however, this is not the case for any of the ten prior ensembles we use in the paper. Thus, in reconstructing a time step in the glacial, for example, both glacial and Holocene states are part of the prior ensemble.

We would also like to specifically address the following comment: "the authors argue that it is important to take into account the changes in seasonality of precipitation (e.g. line 233) but I wonder how this could be achieved by selecting model states that are coming from very different periods." We agree with the reviewer that the best way to account for changes in precipitation seasonality is to do it in a time-varying manner, but we use a time-invariant prior for reasons given above. This means that the mean precipitation seasonality is constant in time and determined by the states in our prior ensemble. What is new in our paper is that for each reconstruction, we use ten different prior ensembles, which gives us ten different estimates of the mean precipitation seasonality. In addition, our approach accounts for spatial variations in precipitation seasonality.

We do not wish to mislead readers into thinking that we account for time-varying precipitation seasonality. Instead, we account for a mean precipitation seasonality, which is determined by the states in our prior ensemble. To clarify this, we have edited lines 242-244 in the original paper (lines 293-302 in the revised paper) to say the following:

With T^*_{site} in our PSM, we find that the δ^{18}O-T_{site} slope is spatially variable (e.g., Fig. S5), ranging between 0.42 and 0.66 ‰$^{\circ}$C^{-1} at the ice-core sites (Table S1), and tending to be less than the modern spatial relationship of 0.67 ‰$^{\circ}$C^{-1} at most locations around Greenland. Due to the data assimilation method outlined above, these slopes vary both in space and across iterations, the latter being due to the varying prior ensembles. These slopes do not vary in time in the prior, but they do in the posterior (note that δ^{18}O-T_{site} slopes mentioned throughout this paper refer to the prior ensemble). By using ten different prior ensembles, we capture the uncertainty in the δ^{18}O-temperature relationship from variations in the precipitation seasonality. These TraCE-21ka-derived estimates lie within the range of slopes estimated for sites around Greenland for a variety of time periods (Table S1). Differences seen in Table S1 reflect both the different methods used and the time period considered. Some estimates, such as Guillevic et al. (2013) and Buizert et al. (2014) are for abrupt transitions, such as Dansgaard-Oeschger events, while others find mean slopes over longer periods of time, such as Kindler et al. (2014) and this investigation.

Referee #1 Specific Comment #1

More specifically, still related to the prior, the authors explain (line 135) that 'For paleoclimate data assimilation, it is important that the climate simulation capture a range of possible climate states over the time period of interest.' They should thus first discuss the results of the TraCE-21ka simulation as it seems from Figure 12 that it underestimates the magnitude of the changes. More generally, the authors do not discuss at all the biases of the climate model. They correct for biases in the modern state by using anomalies compared to 1850-2000 (line 146) but this seems to be a small change compared to the signal during the whole simulation (line 366). Besides, the response to forcing is very different between different models as illustrated by

the Paleoclimate Model Intercomparison Project. How this model behavior, which can also bias results for distant past, is influencing the results? Another way to phrase this point is that the model biases are not constant over time while the proposed correction assumes the stationarity of the biases.

The reviewer makes a good point that we should expand our discussion of the TraCE-21ka simulation, especially with respect to model biases.

Model-bias corrections rely on observations. In the modern, there are numerous observations from ground-based and satellite systems. From the past, there are relatively few observations, which are from proxy records. Given an assumption of stationary model bias (unchanging in time), we can use the modern observations to compute the bias correction; however, given an assumption of non-stationary model bias, we must rely on paleoclimate proxy records as well. In our paper, we assume a stationary model bias and apply the delta-change method (Teutschbein and Seibert, 2012). This leaves the proxy records available for data assimilation. Ideally, we would subsample the proxy records and use one subsample for data assimilation, another to correct for model bias *a priori*, and another to assess the influence of model bias on our reconstruction *a posteriori*. This, however, is not possible with a small number of proxy records. Therefore, we have chosen to reserve all proxy records for data assimilation and to assume a stationary bias correction.

In addition to a mean bias in the model, there may also be biases in the variance. The reviewer specifically points out this issue with TraCE-21ka: it has a small glacial-Holocene climate change relative to the ice-core records. This is a bias that is best addressed with information from proxy records, which we have reserved as independent observations for data assimilation.

As the reviewer alluded to, another opportunity to assess the influence of model bias is to simply select our prior ensemble from a different model. By examining how the results are affected by a variety of different model simulations, we could assess the sensitivity of our results to different models (and thus model biases). Doing this analysis in a rigorous manner is not yet possible because TraCE-21ka is the only-available continuous 20,000-year simulation completed with a fully-couple GCM at a T31 resolution or higher.

We have included this discussion as a paragraph in section 2.2 (lines 147-155 in the revised paper):

> From TraCE-21ka, we use two-meter air temperature for temperature (T) and the sum of large-scale stable precipitation and convective precipitation for precipitation (P). To correct for model bias in TraCE-21ka, we assume that the bias is stationary in time and apply the delta-change method (Teutschbein and Seibert, 2012) by taking the anomaly of temperature and the fraction of precipitation relative to the mean of our reference period (1850-2000 CE). An assumption of a stationary model bias is required because, with a small number of proxy records, we cannot afford to subsample them for the purposes of bias correction, data assimilation, and evaluation. After the bias correction, we average the TraCE-21ka variables (which originially have monthly resolution) to 50-year resolution, as we did for the ice-core records. In this process, we average 50 consecutive years (600 months) such that no year (or month) is used in more than one 50-year average. This averaging results in 440 time steps spaced 50 years apart.

Referee #1 Specific Comment #2

Estimating the skill of the reconstruction compared to a constant prior (line 204) is a too low target for me. If the reconstruction was only showing a warming between the glacial period and the Holocene, it would already be skillful compared to this initial estimate and this does not require a very sophisticated technique. The skill of the reconstruction should be evaluated against the transient TraCE-21ka simulation to see if the data assimilation brings some skill compared to the simulation not constrained by data.

We agree with the reviewer that we should compare our reconstruction skill against that of other 20,000-year reconstructions or simulations. The skill metrics are most comparable if there is either a δ^{18}O or T* variable available in the reconstruction

or simulation. Thus, this comparison is straightforward with TraCE-21ka, but not for other reconstructions like that of Buizert et al. (2018). We have computed the correlation coefficient, coefficient of efficiency (CE), and root mean square error (RMSE) for TraCE-21ka, and have added this information both as text (in Sect. 3.2 of the revised paper) and figures (Figs. S8-S11). These figures are also included in this document.

We have added the following text (to line 209 of the original paper and line 258 of the revised paper) to introduce our evaluation against TraCE-21ka:

> For further comparison, we additionally compute the correlation coefficient, CE, and RMSE between the TraCE-21ka simulation and the proxy records.

We have also edited section 4.1 of the original paper (section 3.2 of the revised paper), which discusses the results of the evaluation. Please see our reply to specific comment #6 for the edited text (lines 325 to 375 of this document).

Referee #1 Specific Comment #3

The authors explain at the end of the conclusion (line 485) that using a model that directly simulates isotopes would likely improve their results. It would be interesting to discuss that earlier because, for instance, they mention a different relationship between reconstructed precipitation and temperature at different time scales (line 352) but what is the potential role of a different relationship between temperature and δ^{18}O on this conclusion?

We agree with the reviewer that the isotope-temperature relationship is an interesting one to explore, a topic which we touch on with our sensitivity experiments described in section 2.3.1 of the paper (lines 245-255 of the original paper and lines 303-313 of the revised paper). We tested the sensitivity of our results to different isotope-temperature relationships, primarily the magnitude of the slope in a linear relationship, but also the spatial pattern of that slope. The reconstructions resulting from these experiments are discussed in Sect. 4.2 and Fig. 10 of the original paper (Sect. 3.3 and Fig. 9 of the revised paper).

An isotope-enabled model, as we mention in the conclusion, would provide another estimate of the magnitude and spatial pattern of the isotope-temperature relationship. The advantage of an isotope-enabled model is that it incorporates the variety of processes that can affect water isotope ratios, whereas, with our experiments, we focus on one primary process: precipitation seasonality. Isotope-enabled models may have biases, and the experiments we perform are important for assessing the sensitivity of reconstructions to the assumed isotope-temperature relationships. As the reviewer suggests, we have brought up this discussion earlier in the paper by adding the following sentences to the end of section 2.3.1 (lines 314-316 of the revised paper):

> These sensitivity tests are equivalent to testing different assumptions about the δ^{18}O-temperature relationship. The availability of a 20,000 year-long isotope-enabled climate simulation would allow us to determine this relationship from model physics, which incorporate a variety of processes that can affect water isotopes, including precipitation seasonality.

The reviewer also astutely suggests that we test the effect of the isotope-temperature relationship on the precipitation-temperature relationship. We have analyzed the temperature-precipitation relationship between all possible combinations of our main reanalysis and sensitivity reanalyses and have added the following paragraph to the paper (in section 4.1, lines 508-516 of the revised paper). Table S3 is included in this document and has also been added to the Supplementary Information.

> We examine how the sensitivity experiments (Figs. 9 and 10) affect the scaling factor (β) in the precipitation-temperature relationship. We pair the five temperature reconstructions (main, S1-S4) and three precipitation reconstructions (low, moderate, and high) into fifteen possible combinations and conduct the same analysis as described above. Across these fifteen combinations, we find that the spatial pattern of β is robust (Fig. 11a). The exact magnitude depends primarily on the temperature reconstruction and how cold it is in the glacial, with colder temperatures giving lower β values.

To a lesser degree, the magnitude also depends on the precipitation reconstruction, with wetter scenarios giving lower β values. As previously, we find that the low-pass filtered datasets have the same or nearly the same β value as the unfiltered dataset, while the high-pass filtered datasets have consistently lower β values. As an example of this, Table S3 shows the β value found for the Kangerlussuaq region for all fifteen combinations and three filtering options.

Referee #1 Specific Comment #4

If I am right, when the technique described in section 2.3 is applied for the past millennia, records related to both the temperature and hydrology are assimilated together, as the covariance between the variables can bring interesting information and reduces the uncertainties. Here, it is claimed that having independent temperature and precipitation reconstructions is an advantage. This also means that precipitation and temperature changes could not be dynamically consistent in the proposed reconstruction? Maybe the authors do not want to rely on the covariance between those two variables as simulated by the climate model but they should explain why and, in that case, explain in a bit more detail the added value brought by the assimilation using this model results as prior.

The reviewer is correct that previous work (e.g., Hakim et al., 2016; Tardif et al., 2019) has used a similar method to assimilate both temperature and precipitation-sensitive proxy records into a reconstruction of multiple climate variables over the last millennium. We agree with the reviewer that our choice to separate the proxy records and reconstruct temperature and precipitation independently requires more explanation in the paper.

We have chosen to independently reconstruct temperature and precipitation because the relationship between these two variables is highly non-linear and non-stationary over the last 20,000 years. Precipitation generally follows the expected thermodynamic relationship on glacial-interglacial timescales (Robin, 1977); however, there have been times in the last 20,000 years, as shown by ice-core records, that precipitation has deviated significantly (even having opposite sign) from the thermodynamic expectation (Cuffey and Clow, 1997). For this reason, we choose not to impose the climate-model-derived mean temperature-precipitation relationship on our reconstruction.

We would also like to address the reviewer's concern that the temperature and precipitation may not be dynamically consistent given our method of reconstructing them independently. The reviewer is correct that our reconstructions may not achieve dynamic consistency through the modeled relationships. However, the reconstructions are dynamically consistent insofar as the empirical δ^{18}O and accumulation records from the ice-core records are dynamically consistent, as they must be.

We have made the following edits to the paragraph at lines 175-183 of the original paper (lines 213-232 of the revised paper) to reflect this discussion.

> The prior ensemble is an initial estimate of possible climate states, which we form using 100 randomly-chosen 50-year averages from the TraCE-21ka simulation. States from both the glacial and the Holocene make up a prior ensemble. The same prior is used for all time steps in the reconstruction, leading to a prior that is constant in time. For a longer discussion on the reasoning behind our choice to use a constant prior, please refer to the Supplementary Information, Sect. S1. Proxy records are assimilated into the prior using Eq. 1, which produces the posterior ensemble, a new estimate of possible climate states. We assimilate δ^{18}O to reconstruct temperature and separately assimilate accumulation to reconstruct precipitation. This approach maintains independence between temperature and precipitation, which avoids imposing linearity and stationarity on the relationship between these two variables. As Cuffey and Clow (1997) show, not only is this relationship non-linear on long timescales but it is also not well-approximated by simple thermodynamic expectations. Separating these variables ensures that the relationship between temperature and precipitation is consistent with the empirical relationship between δ^{18}O and accumulation from ice cores, rather than being derived exclusively from the climate model.

We repeat the data assimilation process over multiple iterations, with each iteration using one of ten different 100-member prior ensembles and excluding one proxy record. Each of the ten prior ensembles is made up of a different random selection of 50-year averages from TraCE-21ka. Thus, each prior ensemble has a different variance and spatial covariance structure. Each proxy record is excluded from a total of ten iterations, where each of these iterations uses a different one of the ten prior options. Every iteration is uniquely identifiable by which prior ensemble is used and which proxy record is excluded. For a reanalysis, the total number of iterations is thus ten times the number of proxy records, such that for temperature we have 80 iterations and for precipitation we have 50 iterations. A reanalysis is a compilation of the 100-member posterior ensembles from these iterations, resulting in a temperature reanalysis having 8,000 ensemble members and a precipitation reanalysis having 5,000 ensemble members.

Referee #1 Specific Comment #5

Line 153, it is said that 'The offline method is appropriate when characteristic memory in the system is significantly shorter than the time step' (here 50 years). Is this valid here, for Bølling-Allerød and Younger Dryas events for instance?

The reviewer raises an excellent point that periods with strong forcing, such as the Bølling-Allerød and Younger Dryas, have a longer characteristic memory than periods with weaker forcing, such as the late Holocene. In general, models have little predictive skill on decadal or longer timescales, except perhaps for areas strongly influenced by the thermohaline circulation (Latif and Keenlyside, 2011, and references therein). During periods of strong forcing, model predictive skill may increase if both the forcing and the response are appropriately represented by the model; however, the predictive skill may also decrease if model uncertainty is large (Hawkins and Sutton, 2009). Rather than assuming our method is appropriate for all times in the last 20,000 years, we would ideally make use of an ensemble of long climate simulations or online data assimilation; however, both alternative approaches are not feasible at this time due to computational cost.

To clarify this point in the paper, we have edited the first paragraph of section 2.3 (lines 149-153 of the original paper, lines 179-190 of the revised paper) to say the following:

> To combine the ice-core and climate-model data, we use an offline data assimilation method similar to that described in Hakim et al. (2016). If no covariance localization is used, as in this study, this method can be summed up as a linear combination of randomly-selected model states that are weighted according to new information provided by the proxy records. "Offline" refers to the absence of a forecast model that evolves the climate state between assimilation time steps. The offline method is appropriate when model predictive-skill is small given the assimilation time step (Hakim et al., 2016, and references therein). Model predictive-skill is generally poor on decadal to longer timescales (Latif and Keenlyside, 2011, and references therein) except possibly during times of strong forcing, such as the Bølling-Allerød (14.7-12.7 ka) and the Younger Dryas (12.7-11.7 ka) (Hawkins and Sutton, 2009). Because each of our time steps is an average over 50 years, as dictated by the resolution to which we average the proxy records, the offline method is appropriate except possibly during these large-forcing events. For these events, an online method may be appropriate (assuming that the models correctly capture both the forcing and the response); however, online ensemble data assimilation over glacial-interglacial cycles using a fully-coupled earth system model is impractical due to the computational cost and is beyond the scope of this study.

Referee #1 Specific Comment #6

Line 375. The reanalysis skill over the full period is clear compared to a constant climate but this would be informative to quantify it more precisely for the two selected 5000-year periods. Stating that it is lower than for the full period is not enough I think. From the figures 7 and 8, it seems that the CE is negative for nearly all the points. Stating line 377 that 'the reanalysis shows overall improvement over the prior ensemble' is also a weak conclusion as discussed in point 2.

We agree with the reviewer that the evaluation over the two 5,000-year periods warrants more discussion. We have included this discussion as well as the comparison of our reconstruction skill to that of TraCE-21ka (in reply to specific comment #2, lines 182 to 196).

3.2 Independent proxy evaluation

Here we evaluate our results against proxy records that are excluded from ten of the iterations that make up the temperature and precipitation reanalyses. For this evaluation, we use the raw results (without a mean-bias correction). We find, however, that the mean biases are small relative to climate changes over the last 20,000 years; there is little difference between our bias-corrected and uncorrected results and it is unlikely that the mean bias has a large effect on our evaluation.

Evaluation against independent proxy data shows that our reanalysis captures the timing and magnitude of low-frequency climate changes (Figs. 6 and 7) and is an improvement over both the prior ensemble and TraCE-21ka (Figs. S6-S7 and S8-S9). Evaluation over the full 20,000 years of the temperature and precipitation results shows high, positive correlation coefficients (ranging from a minimum of 0.97 to maximum of 0.99), which indicate that the reanalysis captures both the timing and sign of climate events, while high CE (0.87-0.98) and low RMSE values (0.62-1.2 ‰ for δ^{18}O and 0.04-0.08 for accumulation) indicate that the reanalysis captures the magnitude of these events. Our skill during this longest evaluation period is primarily due to the presence of low-frequency climate changes, which tend to be coherent across Greenland, such that evaluation over this full 20,000-year period shows more skill than evaluation over the full Holocene, which shows more skill than evaluation over just 5,000 years in the Holocene (or 5,000 years in the glacial) (Figs. 6 and 7).

Our posterior ensemble consistently shows improvement over the uninformed, constant prior ensemble during the 20,000-year evaluation period (Figs. S6 and S7). The TraCE-21ka simulation is also uninformed by the ice-core data, but it is transient and generally captures glacial to Holocene changes. Over the 20,000-year evaluation period, relative to the reconstructions, we find that TraCE-21ka has consistently lower correlation coefficients (0.86-0.96), lower CE values (0.50-0.86), and higher RMSE values (1.9-2.8 ‰ for δ^{18}O and 0.11-0.15 for accumulation) (Figs. S8 and S9). This comparison suggests that our reconstruction captures the timing and magnitude of the glacial to Holocene transition better than TraCE-21ka.

Even for the shorter evaluation periods, which are dominated by high-frequency, spatially-incoherent noise, the reanalysis shows improvement over both the prior ensemble and TraCE-21ka (Figs. S6-S7 and S8-S9); however, the improvement is not as consistent as for the 20,000-year evaluation period. For our reconstruction, correlation is positive except for three locations in the temperature reconstruction, with Holocene precipitation showing the largest correlation values (up to 0.60) and the total range being -0.30 to 0.60. The prior correlation is zero for these shorter evaluation periods and locations; however, TraCE-21ka shows correlation values ranging from -0.29 to 0.69, with eight negative correlations (more than we find for our reconstruction), but generally higher correlations in the Holocene than our reconstruction.

For the shorter evaluation periods, the reconstruction CE is generally negative (ranging from -82 to 0.17) with a few exceptions; however, the reconstruction may still be skillful (e.g., Cook et al., 1999). The skill of the reconstruction is better measured by the difference between prior and posterior CE due to the strong influence that bias can have on CE (Hakim et al., 2016). There is consistent improvement of the posterior CE over that of the prior (ranging from an increase of 4.7 to 3200) and over that of TraCE-21ka (ranging from an decrease of -6.9 to an increase of 230). RMSE is the most consistent of the skill metrics for these shorter evaluation periods, with our reconstruction showing improvement over both the prior and TraCE-21ka, the one exception being that TraCE-21ka has greater skill at the Agassiz ice-core site in the Holocene. For the reconstruction, RMSE values range from 0.24 to 1.8 ‰^{2} for temperature and 0.025 to 0.10 for precipitation.

For all evaluation periods and both variables, the ensemble calibration ratio (ECR) for the prior is skewed towards values greater than unity (0.66-8.7), which suggests that the prior ensemble tends to have too little spread. Conversely, for the posterior, the ECR is generally less than unity (0.10-1.7) (Figs. 6 and 7), suggesting that the posterior ensemble has more spread than the error in the ensemble mean (as compared to the proxy records). This result indicates that the reanalysis ensemble encompasses the climate as recorded by the proxy records for most times and locations over the last 20,000 years.

Referee #2

The authors present new reconstructions of temperature and precipitation over the last 20,000 years over Greenland. For this, they apply a data assimilation technique on δ^{18}O and accumulation records from Greenland ice cores and use the temperature and precipitation outputs from the TraCE-21ka simulation to extend the information to all the continent. The paper is in general clear enough for that people not having skills in data assimilation can read and understand quite easily the methodology presented in this manuscript. In my knowledge, the technique presented here is innovative for such a long period, and the different assumptions are presented and tested in a very rigorous way. This manuscript is worthy for publication in Climate of the Past, after having considered the comments below.

Thank you.

Referee #2 General Comment #1

Compared to the ice cores part, which is well discussed in the Methods section and in the Supplementary Material, the results of the TraCE-21ka simulation are not discussed enough in my opinion. More details about the similarities/differences with other PMIP simulations and/or climate reconstructions for PI and LGM could be discussed for example. How do the last 1000/2000 years fit well with last millennium simulations or reconstructions from isotopic proxies? Moreover, the rapid climate transitions are not so well captured by the TraCE-21ka, especially the Younger Dryas. This point should be discussed in terms of potential consequences for the reconstructions by data assimilation. The spatial resolution (T31 and 26 atmospheric vertical levels for the atmosphere if I am not wrong) should be clearly stated, and the uncertainties related to this aspect could be discussed (even if it is mentioned later). For example, is the limited number of grid points over Greenland a problem for the paleo DA technique? The ice sheet boundary conditions are also of major importance in this type of simulation. The expected differences if a more recent ice-sheet reconstruction would be prescribed could be discussed. Last point: what about the precipitation seasonality in TraCE-21ka over Greenland? Is it consistent with observations? Is the seasonality different for Holocene and glacial periods? I guess it could have an important impact on the δ18O PSM and the final temperature reconstruction...

We agree with the referee that our discussion of the TraCE-21ka simulation is limited compared to our discussion of the ice-core records. This is primarily because TraCE-21ka is described extensively in the literature (e.g., Liu et al., 2009, 2012; He, 2011; He et al., 2013), whereas a number of aspects of the ice-core network, particularly the accumulation records, are novel or have been discussed little in previous work. In section 2.2 of the paper, we do discuss attributes of TraCE-21ka that are especially relevant to our method, including, the glacial to Holocene mean state changes, temperature and precipitation seasonality, and the ice-sheet boundary conditions. We agree with the referee that we are missing a statement about the spatial resolution of TraCE-21ka (indeed it is T31) and more details concerning seasonality and the ice sheets. We have included these details in the revised paper.

2.2 Climate-model simulation

We use TraCE-21ka, a simulation of the last 22,000 years of climate (22 ka to -.04 ka), which was run using the fully-coupled CCSM3 at T31 resolution (approximately 3.75 degrees horizontally) with transient ice-sheet, orbital, greenhouse gas, and meltwater flux forcings (Liu et al., 2009, 2012; He et al., 2013). For paleoclimate data assimilation, it is im-

portant that the climate simulation capture a range of possible climate states over the time period of interest. By design, TraCE-21ka captures the major glacial-to-Holocene temperature changes, as well as some of the short-term, rapid climate changes, such as the Bølling-Allerød transition (Liu et al., 2009). Many higher-resolution climate simulations are transient only over the last millennium (e.g., Bothe et al., 2015) or provide a snapshot of a certain time, such as last glacial maximum or the mid Holocene (e.g., Harrison et al., 2014). Individually, these millennial-length simulations have too little variability to capture the range of climate states across the glacial-interglacial transition. If combined, the biases in each simulation would need to be individually addressed, which is beyond the scope of this study.

From TraCE-21ka, we use two-meter air temperature for temperature (T) and the sum of large-scale stable precipitation and convective precipitation for precipitation (P). To correct for model bias in TraCE-21ka, we assume that the bias is stationary in time and apply the delta-change method (Teutschbein and Seibert, 2012) by taking the anomaly of temperature and the fraction of precipitation relative to the mean of our reference period (1850-2000 CE). An assumption of a stationary model bias is required because, with a small number of proxy records, we cannot afford to subsample them for the purposes of bias correction, data assimilation, and evaluation. After the bias correction, we average the TraCE-21ka variables (which originially have monthly resolution) to 50-year resolution, as we did for the ice-core records. In this process, we average 50 consecutive years (600 months) such that no year (or month) is used in more than one 50-year average. This averaging results in 440 time steps spaced 50 years apart.

TraCE-21ka includes changes in orbital forcing, which contribute to changes in the seasonality of temperature and precipitation. The strength of these seasonal cycles influences the mean-annual relationship between δ^{18}O and temperature (Steig et al., 1994; Werner et al., 2000; Krinner and Werner, 2003). TraCE-21ka consistently shows stronger temperature and precipitation seasonality in the glacial (20 to 15 ka) than in the Holocene (5 to 0 ka) at each of the eight ice-core sites considered in this study (Fig. S3). Relative to the annual mean, the glacial had warmer and wetter summers and colder and drier winters. The Holocene also shows such a seasonal cycle; however, there is a smaller difference between the summers and winters. Any seasonal signal with wetter summers than winters will bias the δ^{18}O towards summer values. According to TraCE-21ka, this effect is amplified in the glacial. As we discuss in Sect. 2.3.1, the particularly strong summer bias in the glacial affects the mean-annual δ^{18}O-temperature relationship in ways that are consistent with findings from borehole thermometry at the GISP2, GRIP, and Dye3 ice-core sites (Cuffey and Clow, 1997; Jouzel et al., 1997).

In addition to a change in the strength of the seasonal cycle, TraCE-21ka shows a temporal shift, with summer temperature and precipitation peaking earlier in the glacial (around June and July) than in the Holocene (from July to September) (Fig. S3). In the glacial, both variables peak around June and July, with only two exceptions: precipitation peaks in August at the Renland and Dye3 ice-core sites. In contrast, Holocene temperature peaks slightly later, in July, while precipitation peaks even later, in August and occasionally September. Both variables and both time periods show winter minimums in February, again with the two exceptions of Renland and Dye3, which show later precipitation minimums. In this study, the timing of the seasonal peaks is relevant because it affects the precipitation-weighted temperature, defined in Eq. 7.

TraCE-21ka also includes prescribed transient ice sheets as a boundary condition. The transience is important for capturing the influence of elevation change on the ice-core records; however, the low horizontal resolution of TraCE-21ka leads to difficulties in capturing elevation changes at the edges of the ice sheet, in southern Greenland, and at coastal ice caps. In addition, the ice sheets in TraCE-21ka are independent of climate, updated only every 500 years during the simulation, and taken from ICE-5G (Peltier, 2004), a now outdated ice-sheet reconstruction (Roy and Peltier, 2018).

We want to emphasize that our paper represents the first attempt to use the ensemble Kalman filter approach to reconstruct climate over glacial-interglacial timescales. Ideally, there would be more 20,000-year (or longer) TraCE-21ka-like simulations (i.e., from fully-coupled GCMs at T31 resolution or higher). With other simulations from different models, we could address the influence of model bias, spatial resolution, boundary conditions, and initial conditions (for a discussion of model bias, see

our reply to Referee #1's specific comment #1, lines 141-174 of this document). If we had other simulations, then we agree with the referee that it would warrant a comparison between simulations and a discussion of how the differences affect our results. With only one TraCE-21ka-like simulation available, we cannot conduct a meaningful comparison with how other climate-model simulations would affect our results. Our focus instead is on demonstrating that the method is feasible and skillful. Thus, we do not see it as productive to elaborate on how TraCE-21ka compares with PMIP or other simulations. In addition, the previous literature has extensively interrogated the results of the TraCE-21ka simulation (e.g., He, 2011; Buizert et al., 2014; Pedro et al., 2016; Zhang et al., 2017, 2018; Marsicek et al., 2018). We will add new text to refer the reader to this previous work.

The referee additionally suggests that we include a discussion of how well TraCE-21ka captures rapid climate transitions, such as the Younger Dryas. We agree that this would be necessary if we were to choose our prior ensemble exclusively from time periods that are close to our reconstruction time (for examples of this, see our reply to Referee #1's general comment, lines 29-129 of this document); however, our prior ensemble is time-invariant and thus our method is insensitive to the temporal evolution of TraCE-21ka (as long as the variance and spatial covariance structures of TraCE-21ka are preserved).

Referee #2 General Comment #2

The way how the prior ensemble is made should be clarified. To avoid misunderstanding, the authors should state at line 175 that the prior is constant in time (and not later at line 206). If I understand clearly, 10 different 100-member prior ensembles are made (it should be said directly at the beginning). To form a prior ensemble, do you then randomly pick up 100 snapshots from the resampled TraCE-21ka temperature and precipitation outputs (see my minor comment for the line 146)? Or do you take randomly from the yearly TraCE-21ka outputs 50 consecutive years of data that you average in time for a member, other 50 consecutive years of model outputs for another time period for the second member, and so on. . .? Or another way? Anyway, it needs clarification (in the way of the section 2.3 of Hakim et al. 2016 for example). I have the same remark as the first referee: what would be the difference if you would use, for instance, a "glacial prior" to reconstruct climate variables from glacial period and a "Holocene prior" for the warmer period instead of a constant prior for all the 20,000 years? In link with my first major remark, what would be the impact on the seasonality of precipitation, that influences the reconstructions of the authors?

We thank the referee for calling our attention to these points of confusion. We now state earlier in the paper that the prior is constant in time (line 215 of the revised paper). We have also clarified how we average TraCE-21ka by making the following edit (around line 146 of the original paper and lines 152-155 of the revised paper):

> After the bias correction, we average the TraCE-21ka variables (which originially have monthly resolution) to 50-year resolution, as we did for the ice-core records. In this process, we average 50 consecutive years (600 months) such that no year (or month) is used in more than one 50-year average. This averaging results in 440 time steps spaced 50 years apart.

We have also reorganized and edited a paragraph in Sect. 2.3 (starting at line 175 of the original paper and line 213 of the revised paper) to clarify how we form our ten different 100-member prior ensembles. Please see lines 264-285 in this document for the edits.

The referee additionally raises concerns over our random selection of the prior ensemble from the full TraCE-21ka simulation and the effect this has on the seasonality of precipitation and our reconstructions. For this we refer to our reply to the general comments from Referee #1 (lines 29 to 129 in this document).

Referee #2 Minor Comments

We thank the referee for these specific edits and the effort to make the paper clearer and more concise.

Line 14: "requires understanding its sensitivity to changes. . ."
Thank you. This edit has been be made in the revised paper.

Line 18: I would put into brackets the terms "and arid" and "and wet".
We think that it is important to equally emphasize the thermal and hydrologic differences between glacial and interglacial periods. For this reason, we have not make the suggested revision.

Line 31: the spatial resolution, especially for paleoclimate simulations, brings also uncertainties.
We have edited this to say: "In contrast, climate-model simulations are spatially-complete estimates of past climate, but they are subject to uncertainty due to model dynamics, boundary conditions, and spatial resolution."

Line 37 and passim: I think this is TraCE-21ka and not TraCE21ka.
We have made this change throughout the paper.

Paragraph lines 106-116: Does the matching of δ^{18}O from Dye3 to the δ^{18}O record from NGRIP bring a dependency when evaluating the posterior against Dye3 δ^{18}O record?
While dating uncertainty is non-zero, it is clear from multiple independent lines of evidence that the major δ^{18}O variations in all Greenland cores are nearly synchronous. Most important is the variance, rather than the timing, of δ^{18}O, and this is not affected by the minor changes to the timescales imposed here. (Note that our matching of the δ^{18}O from Dye3 to the δ^{18}O record from NGRIP is similar to the methods used to create the GICC05 depth-age scale and apply it to other ice cores.)

Section 2.2: I understand when you use the term "transient ice-sheet" that the prescribed ice-sheet is changed over time. But some people can misunderstand and think that it is done dynamically with a coupled ice-sheet model. I would use an expression like "prescribed transient ice-sheet boundary conditions" for example.
We have changed the sentence to say: "TraCE-21ka also includes prescribed transient ice-sheets as a boundary condition."

Line 146: What is the initial temporal resolution of TraCE-21ka outputs? Monthly mean? When you talk about "average of 50-year resolution", do you mean "resampling" every 50 model years? If you take the last 20,000 years, it makes something like 400 time steps, right?
We thank the referee for calling our attention to this point of confusion. We have edited section 2.2 to clarify (see lines 427 to 430 of this document). The referee is correct that we end up with about 440 time steps after averaging. Our method is to take 50 consecutive years (600 months) of TraCE-21ka and to average them. No year (or month) is used in more than one 50-year average.

Lines 230-232: Other model studies like Gierz et al. 2017 (JAMES) for the LIG and Cauquoin et al. 2019 (CP) for 6k-PI climates have shown that the seasonality of precipitation affects the δ^{18}O-temperature relationship over Greenland.
We thank the referee for this comment. We did not mean to imply that Werner et al. (2000) is the only modeling study that has shown that precipitation seasonality affects the δ^{18}O-temperature relationship in Greenland. We have added the recommended citations.

Line 286: "that the proxy y and prior estimate of the proxy H(xb)".
Thank you. This edit has been made in the revised paper.

Lines 311-316: What does it give compared to the TraCE-21ka results?
We agree with the referee that it's important to compare our results to that of TraCE-21ka; however, we have made a point of saving comparisons with TraCE-21ka and Buizert et al. (2018) for the discussion (Sect. 4.3 of the revised paper).

Line 326: "from nearly +2 °C in northern. . ."
Thank you. This edit has been made in the revised paper.

Line 367: "has a large effect on our evaluation."
Thank you. This edit has been made in the revised paper.

Line 380: "the ECR is. . ."
Thank you. This edit has been made in the revised paper.

Line 404: the slower warming trends are hard to see. Make a zoom in the figure or give numbers.
Thank you for the suggestion. We have edited the figure (Fig. 10 in the original paper and Fig. 9 in the revised paper) to show a zoom-in on the Younger Dryas to Holocene transition. The edited version is included in this document.

Line 428: For S4 and high P cases, say clearly that it refers to the "sensitivity" curves on figure 12.
Thank you, we have clarified that "sensitivity" in the legend label refers to S4 and high P (Fig. 12 of the original paper and Fig. 13 of the revised paper).

Line 486: you can add the reference Okazaki and Yoshimura 2017 (CP).
We agree that this is a relevant citation and have added it to the revised paper.

Line 488: Add the references Cauquoin et al. 2019 (CP) and Okazaki and Yoshimura 2019 (JGR Atmos).
We agree that these are relevant citations and have added them to the revised paper.

Figure 4: add maybe contours for more clarity. And change the scale for the precipitation fraction at the peak warmth in the Holocene (panel b).
Thank you for the suggestion. We have added contours for clarification and changed the color scale in panel (b). The edited version is included in this document.

Figures 7, 8, S3, and S4: quite normal that the correlation is improved for the full period compared to the constant prior climate state.
Yes, we agree and have stated this in the paper (lines 205-209 of the original paper and lines 254-259 of the revised paper). If the referee thinks it is necessary, we can state it again later in the text or in the figure captions. We have additionally be compared the skill of our results to TraCE-21ka, as suggested by Referee #1.

Referee #3

Review of Badgeley et al. 2020 on Greenland paleo data assimilation

Badgeley et al. present temperature and precipitation fields for the last 20,000 years over Greenland generated using a paleo data-assimilation technique. This is an interesting and potentially very valuable new approach to investigating past climates. The paper is well written and clearly illustrated, and I am generally enthusiastic about the work.

Thank you.

While the methodology represents a big step forward, the paper is also a step backwards in other regards as it assumes a constant linear scaling of d18O to site temperature for all sites and periods based on the spatial d18O-T relationship. This assumption has been disproven in the last 2 decades through careful work in the ice core community (including some of the papers cited here). This assumption will dominate all the spatial and temporal patterns in the temperature reconstructions,

and deserves more careful consideration than it is given here. The authors suggest that this problem is alleviated by using the precipitation weighted temperature, but they do not demonstrate this. Below I recommend some comparisons that should be done before the paper is suitable for publication.

My main concern is the use of a single linear d18O-T scaling based on the spatial d18O-T pattern at all sites and locations. While water isotopes are a valuable proxy, its temperature interpretation has proved very difficult. Borehole thermometry and d15N gas thermometry are the most reliable methods to get absolute (calibrated) temperature changes, and both suggest a d18O slope that is around half of the spatial relationship (0.67 permil/K) used here (as the authors acknowledge).

We agree with the referee that it would be overly-simplistic to assume "a constant linear scaling of δ^{18}O to site temperature for all sites and periods based on the spatial δ^{18}O-T relationship". This is not, however, what we do; we allow the δ^{18}O-temperature relationship to vary spatially by relying on precipitation-weighted temperature (T*) from TraCE-21ka. As we write in our paper (lines 229-232 of the original paper, lines 279-283 of the revised paper), "Numerous studies have suggested that precipitation seasonality is the largest source of nonlinearity in the δ^{18}O-T_{site} relationship (e.g., Steig et al., 1994; Pausata and Löfverström, 2015); changes in precipitation seasonality are thought to be the primary reason that the effective δ^{18}O-T_{site} relationship for the glacial-interglacial transition has such a low slope (Werner et al., 2000; Gierz et al., 2017; Cauquoin et al., 2019)." We convert T* to δ^{18}O using the equation $0.67\text{T}^* = \delta^{18}\text{O}$, and then compute the best-fit slope between δ^{18}O and temperature to find their relationship. This effectively assumes a constant linear scaling between the *instantaneous* δ^{18}O and site temperature, while allowing for changes in precipitation to affect the time-averaged relationship. These TraCE-21ka-derived slopes vary between 0.42 and 0.66 ‰ °C^{-1} at the core sites (Table S1), and are less than the modern spatial relationship of 0.67 ‰ °C^{-1} at most locations around Greenland (e.g., Fig. S5). The table and figure have been included in the supplementary information.

The TraCE-21ka-derived slopes vary both in space and across prior ensembles (Fig. S5 and Table S1). By using ten different prior ensembles, we capture the uncertainty in the δ^{18}O-temperature relationship, as determined by TraCE-21ka. We also examine a wider range of slope estimates in our sensitivity experiments. The slopes for S4 are shown in Table S1, and the slopes for S1, S2, and S3 are 0.67, 0.5, and 0.335 ‰ °C^{-1}, respectively. As described in our paper (lines 389-405 of the original paper and lines 446-464 of the revised paper), the magnitude of the slope affects the magnitude of the anomalies (Fig. 10 in the original paper and Fig. 9 in the revised paper). The spatial pattern of the slope also has an effect. For example, in the early Holocene, the spatially-variable slopes result in a reconstruction that warms more slowly. In addition, the reconstructions with spatially-varying slopes show stronger north-south gradients than those with spatially-constant slopes. This north-south gradient shows up especially in the abrupt transitions, with larger changes in the north relative to the south (Table S2). In the paper, we had not discussed the impact of the spatially-varying slopes on the spatial pattern of the reconstructions; we will include this discussion in the revised paper by making the following revisions (Sect. 4.2, lines 400-405 of the original paper and Sect. 3.3, lines 457-464 of the revised paper).

> The temperature results are also sensitive to the spatial pattern of the δ^{18}O-T relationship. We find this by comparing the results from the S1-S3 scenarios that assume a spatially-uniform relationship to results from the main reanalysis and S4 scenario that assume a spatially-variable relationship. The S1-S3 scenarios have a characteristic shape to their time series (Fig. 9), and, although the main reanalysis and S4 scenario generally fit this characteristic shape in the glacial and middle-late Holocene, in the early Holocene the main reanalysis and S4 diverge and show slower warming trends than the S1-S3 scenarios (Fig. 9b). In addition, the reconstructions with spatially-varying δ^{18}O-T relationships show stronger north-south gradients during times of abrupt temperature change than those with spatially-constant relationships (Table S2). These findings indicate that there is new information added by using a PSM that includes spatial variability in precipitation seasonality.

The reviewer is correct that we effectively assume that the δ^{18}O-temperature relationship is constant in time. We could in principle account for temporal changes in the δ^{18}O-temperature relationship by using a time-varying prior; however, this

method is complicated by discontinuities and/or extra assimilation parameters. We use the same prior ensemble for all time steps of the reconstruction, which avoids discontinuities in the reconstruction and does not require us to constrain extra assimilation parameters with so few proxy records. A consequence of this method is that the δ^{18}O-temperature relationship, which is derived from the prior ensemble, is constant in time. Unfortunately, we are restricted in our ability to both use a time-varying prior and avoid the complications stated above until more long, fully-coupled climate simulations become available (see our reply to Reviewer #1, lines 29 to 129 in this document, for more details).

As the referee mentioned, borehole thermometry and δ^{15}N gas thermometry are reliable methods to getting at the δ^{18}O-temperature relationship; however, as we say on lines 233-237 of the original paper (lines 283-287 of the revised paper), we do not rely on these methods because borehole thermometry and δ^{15}N gas thermometry are not available at all sites. Instead, we rely on the TraCE-21ka-derived relationships described above, which we compare to relationships found previously using borehole and δ^{15}N gas thermometry (Table S1). It is known that the δ^{18}O-temperature relationship varies temporally depending on the length of time considered and the date (Jouzel et al., 1997). Table S1 shows that indeed, different investigations have estimated a variety of slopes for a variety of time periods. Differences in the estimated slopes are likely to result from the method used in a particular investigation and the time period considered. Some estimates, such as Guillevic et al. (2013) and Buizert et al. (2014) are for abrupt transitions, such as Dansgaard-Oeschger events, while others find mean slopes over longer periods of time, such as Kindler et al. (2014) and our own paper. The slopes that we derive from TraCE-21ka mostly fall within the ranges found by previous studies, even though our slopes are estimated for a time period that is not addressed in the other investigations, the last 20,000 years.

We have included a fuller description and discussion of the slopes in revisions of Sect. 2.3.1:

> With T^*_{site} in our PSM, we find that the δ^{18}O-T_{site} slope is spatially variable (e.g., Fig. S5), ranging between 0.42 and 0.66 ‰$^{\circ}$C^{-1} at the ice-core sites (Table S1), and tending to be less than the modern spatial relationship of 0.67 ‰$^{\circ}$C^{-1} at most locations around Greenland. Due to the data assimilation method outlined above, these slopes vary both in space and across iterations, the latter being due to the varying prior ensembles. These slopes do not vary in time in the prior, but they do in the posterior (note that δ^{18}O-T_{site} slopes mentioned throughout this paper refer to the prior ensemble). By using ten different prior ensembles, we capture the uncertainty in the δ^{18}O-temperature relationship from variations in the precipitation seasonality. These TraCE-21ka-derived estimates lie within the range of slopes estimated for sites around Greenland for a variety of time periods (Table S1). Differences seen in Table S1 reflect both the different methods used and the time period considered. Some estimates, such as Guillevic et al. (2013) and Buizert et al. (2014) are for abrupt transitions, such as Dansgaard-Oeschger events, while others find mean slopes over longer periods of time, such as Kindler et al. (2014) and this investigation.

I suspect this assumption will lead to underestimated temperature variability in the posterior. The authors should check this for the abrupt transitions at the three sites (GISP2, NEEM, NGRIP) where d15N-based temperature changes are known (Buizert et al. 2014).

We thank the referee for this suggestion. In Table S4 we compare the magnitudes of three abrupt transitions (warming into the Bølling-Allerød, cooling into the Younger Dryas, and warming out of the Younger Dryas) against the δ^{15}N-derived temperature estimates from Buizert et al. (2014) at three locations. We note that these three sites are all in central and northern Greenland and are not necessarily representative of southern Greenland or the coastal ice caps (e.g., Agassiz and Renland). The comparison shows that our approach does not underestimate temperature variability as compared to Buizert et al. (2014), with some of our results showing larger changes and some showing smaller changes. The specifics of the comparison are dependent on the location and the time period. For example, our reconstruction shows greater variability than the δ^{15}N-derived reconstruction at NEEM, but less variability at NGRIP and GISP2. In addition, the two reconstructions agree better during the Younger Dryas cooling at GISP2, but are in better agreement during the Younger Dryas and Bølling-Allerød warmings at

NEEM and NGRIP.

However, there is also a clear spatial gradient, as first noted by [Guillevic et al., 2013], a paper that should be cited and discussed. Guillevic observes that d18O changes are largest towards the north (i.e. NEEM), and smaller towards the south (i.e. Summit). However, the actual temperature changes have the opposite gradient – smallest in the north and largest in the south. This means that the d18O-T relationship has an enormous spatial gradient, from ∼0.6 at NEEM to ∼0.3 at Summit. The Guillevic temperature gradient is seen in many (all?) climate model simulations and should thus be considered very robust.

Thank you for the suggestion. We agree that the Guillevic et al. (2013) paper should be cited. We are aware of the discrepancy between the spatial gradient in the δ^{18}O and in the δ^{15}N-derived temperature; however, we do not agree that this discrepancy has been resolved. The Guillevic temperature gradient (larger temperature changes in the south than in the north) is based on four central to northern ice core records (GISP2 and GRIP, however, are at essentially the same site near Summit). Guillevic et al. (2013) find that the temperature changes at NGRIP and the Summit cores are statistically indistinguishable; thus most of their spatial pattern is driven by the difference between NEEM and NGRIP/GISP2/GRIP. We cannot know without more constraints whether this north-south pattern that appears between NEEM and these three other cores holds for other parts of Greenland, such as southern Greenland. As we show in Sect. S3 (Sect. S4 of the revised paper) of our supplementary information, a southern data point is key to reconstructing southern Greenland climate.

The referee notes the reproduction of Guillevic et al.'s north-south pattern by climate models as evidence for the interpretation that this pattern extends to southern Greenland. We acknowledge that this has been found in some climate simulations, for example, Buizert et al. (2014) show that the TraCE-21ka simulation has this pattern during the Bølling-Allerød transition. The model simulations that find this pattern have a particular forcing, such as abrupt changes in freshwater in the North Atlantic (e.g., Liu et al., 2009) or specific sea ice configurations (e.g., Li et al., 2010). Other models show that abrupt climate transitions can be forced by other mechanisms, such as gradual changes in freshwater forcing (e.g., Obase and Abe-Ouchi, 2019), which lead to different spatial patterns. We have added a discussion of the spatial pattern of abrupt temperature changes to the revised paper (see lines 746 to 797 in this document).

These patterns are such that when using a single constant slope (as the authors do), the larger temperature changes would appear to be in the north, as is indeed the case in their reconstructions (Fig 4a, 4c). However, the Guillevic result would actually suggest the opposite pattern in temperature. The authors need to plot the magnitude of abrupt climate warming in their reanalysis (either the 14.7 ka or 11.6 ka transition), and compare it to the d15N-based values. My hunch is that they will find the opposite pattern from the Guillevic result.

As noted above, the reviewer is not correct that we use a single slope for all ice-core sites. We use a spatially-varying slope (for an example, see Fig. S5). We will be sure to clarify this in the revised paper.

We have reviewed the spatial pattern of three abrupt transitions in our main reconstruction: warming into the Bølling-Allerød, cooling into the Younger Dryas, and warming out of the Younger Dryas, respectively. These results are shown in Table S4 for the eight ice core locations and are compared to results from Buizert et al. (2014). Our results show increasing variability to the north accross seven of the eight cores, while the Guillevic et al. (2013) and Buizert et al. (2014) results show increasing variability to the south across just three cores. So yes, the referee is correct that our results show the opposite pattern from the Guillevic et al. (2013) results; however, the reason for this difference is not due to a spatially-constant δ^{18}O-tempreature relationship. Instead, our spatial pattern is a result of how the temporal information from the proxy records is spread throughout the domain via the spatial covariance pattern between temperature and precipitation-weighted temperature in the prior ensemble.

To show this, we have performed three additional experiments, O8, N3O5, and N3O5_BA, in which we have assimilated both δ^{18}O and δ^{15}N-derived temperature from Buizert et al. (2014) and tested a time-period specific prior ensemble. The experiments and their results are given in the following section that has been added to the paper. New figures that are referenced

in this section are included in this document.

4.2 Spatial patterns during abrupt climate change events

Our paleoclimate data assimilation framework is not limited to the assimilation of δ^{18}O and accumulation rate records. In this section we examine how our reanalysis compares to reconstructions driven by another type of proxy, δ^{15}N of N_2. In particular, we focus on abrupt temperature events, for which there is previous work using δ^{15}N (e.g., Severinghaus et al., 1998; Guillevic et al., 2013; Buizert et al., 2014). Abrupt temperature events increase the thermal gradient in the firn – the upper, porous portion of the ice column – which leads to fractionation of the stable isotopes of N_2 (Severinghaus et al., 1998). Using a firn compaction model, temperature can be derived from δ^{15}N measurements with inverse methods (e.g., Severinghaus et al., 1998; Severinghaus and Brook, 1999; Guillevic et al., 2013; Kindler et al., 2014; Buizert et al., 2014; Kobashi et al., 2017). We assimilate temperatures derived from δ^{15}N data collected from the GISP2, NGRIP, and NEEM ice cores (Buizert et al., 2014).

Our reanalysis and the Buizert et al. (2014) records cover three abrupt temperature events – the Bølling-Allerød warming (14.6 ka), cooling into the Younger Dryas (12.9 ka), and warming at the end of the Younger Dyras (11.5 ka). For these three abrupt temperature events, the δ^{15}N-derived temperature records show larger temperature changes at GISP2 in central Greenland (Buizert et al., 2014), while the δ^{18}O changes are larger at NEEM in northwest Greenland (Fig. 2). We perform three sets of experiments to investigate how these two sets of proxy records affect the mean spatial pattern indicated by our reanalysis during abrupt temperature events. The first experiment, "O8", assimilates all eight δ^{18}O records with one 100-member prior ensemble from TraCE-21ka. This results in a 100-member reanalysis ensemble. The second experiment, "N3O5", assimilates all three of the δ^{15}N-derived temperature records and the five remaining δ^{18}O records (those that do not overlap with the δ^{15}N sites), using the same 100-member prior ensemble as used in the O8 experiment. Finally, we perform a modified experiment, "N3O5_BA", with both δ^{18}O and δ^{15}N records, but using a prior ensemble selected from the 1,000 years surrounding the Bølling-Allerød warming. To maintain a 100-member prior ensemble, we use decadal rather than 50-year averages of TraCE-21ka for this experiment. This adjustment does not affect the comparison (results not shown). Detailed methods for these experiments are given in Sect. S6.

For both the “O8” and “N3O5” experiments, the spatial pattern for the abrupt climate change events are similar to our main reanalysis, with the largest magnitude temperature changes in northern and northwestern Greenland, decreasing magnitude with decreasing latitude, and slightly larger change in the central east and southeast than corresponding western regions. For example, the spatial pattern of the Younger Dryas cooling is nearly the same regardless of which grouping of records is assimilated (Figs. 12a and b). This overall finding is robust to different combinations of these proxy records; for example, if we assimilate just the three δ^{15}N-derived temperature records and no δ^{18}O records (results not shown) the pattern is not substantially changed. This pattern of temperature change differs from spatial patterns inferred previously from δ^{15}N for various abrupt climate events (Guillevic et al., 2013; Buizert et al., 2014); however, the O8 and N3O5 experiments show that these differences are not due to the assimilation of δ^{18}O rather than δ^{15}N-derived temperatures.

For our third experiment, N3O5_BA, in which we restrict the prior ensemble to the Bølling-Allerød warming, the spatial patterns obtained by our experiment are similar to those reported by Buizert et al. (2014). In TraCE-21ka, the Bølling-Allerød warming is forced by a sudden termination of freshwater forcing in the North Atlantic (Liu et al., 2009). This forcing leads to large temperature fluctuations in southern Greenland that decrease with increasing latitude. With this covariance pattern dominating the prior ensemble, the N3O5_BA reconstruction indeed shows the largest temperature changes in southern Greenland, followed by central and then northern Greenland (Figs. 12c, S12c, and S13c).

Importantly, these results depend on the spatial patterns in the prior ensemble, which themselves are a consequence of the particular forcing applied to the climate simulation. Some simulations suggest that the rate and timing of meltwater forcing imposed by TraCE-21ka may not be necessary to explain the abrupt climate change events and show that

different spatial patterns result from different forcing (e.g., Obase and Abe-Ouchi, 2019). Prior ensembles selected from such simulations may result in different spatial patterns that are also consistent, within uncertainty, with the proxy data. We suggest that current ice-core records are insufficient to place a strong constraint on the spatial pattern of abrupt climate events, and additional data, especially from southern Greenland (Sect. S4), would be beneficial. Future work to explore this question in more detail will also require model simulations that sample a greater range of forcing uncertainty.

It has also been documented that the d18O-T slope is strongly variable in time, changing by almost a factor of 2 [Kindler et al., 2014].

Please refer to lines 647-654 of this document for an explanation of why we use a method that has a constant δ^{18}O-T relationship in time in the prior ensemble. We note, however, at the δ^{18}O-T relationship is free to vary in time in the posterior ensemble.

It would be unreasonable to ask the authors to redo all the work abandoning a key assumption; rather I think they should do a careful comparison to d15N-based estimates of abrupt climate change to assess how well their method captures both the magnitude and spatial pattern of abrupt temperature changes – and the implications this may have for the LGM and Holocene optimum patterns shown in Fig 4a and 4c. Perhaps they can provide some suggestions for future work on ways to assimilate the d15N-based climate constraints directly.

We agree with the referee that it would be ideal to assimilate more than just δ^{18}O records. Though including a large variety of proxy records is beyond the scope of this paper, we have done three new reconstructions to show how our results are affected by assimilating δ^{15}N-derived temperature records from Buizert et al. (2014). Please see lines 746-797 of this document for more information on these new reconstructions.

If the reconstructed N-S temperature gradient during abrupt change is indeed opposite to the Guillevic gradient, this should be clearly stated in the abstract.

We will edit the relevant part of our abstract to say the following. Major edits are in *italics*. Note that some of the edits are in response to later comments (see lines 985-987).

> Reconstructions of past temperature and precipitation are fundamental to modeling the Greenland Ice Sheet and assessing its sensitivity to climate. Paleoclimate information is sourced from proxy records and climate-model simulations; however, the former are spatially incomplete while the latter are sensitive to model dynamics and boundary conditions. Efforts to combine these sources of information to reconstruct spatial patterns of Greenland climate over glacial-interglacial cycles have been limited by assumptions of fixed spatial patterns and a restricted use of proxy data. We avoid these limitations by using paleoclimate data assimilation to create independent reconstructions of temperature and precipitation for the last 20,000 years. Our method uses *oxygen-isotope ratios of ice and accumulation rates* from long ice-core records and extends *this information* to all locations across Greenland using spatial relationships derived from a transient climate-model simulation. *Standard evaluation metrics for this method show that our results capture climate at locations without ice-core records. Our results differ from previous work in the reconstructed spatial pattern of temperature change during abrupt climate transitions; this indicates a need for additional proxy data and additional transient climate-model simulations.* We investigate the relationship between precipitation and temperature, finding that it is frequency dependent and spatially variable, suggesting that thermodynamic scaling methods commonly used in ice-sheet modeling are overly simplistic. Our results demonstrate that paleoclimate data assimilation is a useful tool for reconstructing the spatial and temporal patterns of past climate on timescales relevant to ice sheets.

The authors suggest that using precipitation-weighted temperatures alleviates the problems associated with using a linear d18O-T scaling. To validate this claim, at the very least they should show a comparison of the 21ka histories of TraCE 2m temperature and TraCE precipitation-weighted temperature at a key site (e.g. Summit), to show how different these two really

are. Ideally, they would show more clearly how this impacts the reconstructed magnitude of the abrupt climate change events (that are most strongly constrained by the d15N data).

We have included a figure in the Supplementary Information as the referee suggests (Fig. S4), which shows that there is a difference between the temperature and precipitation-weighted temperature. The figure is only for the grid cell closest to Summit, but this is the case for all locations around Greenland, though the magnitude of the difference varies by location.

Through our temperature sensitivity experiments, we have tested the impact that precipitation-weighted temperature has on our reconstructions (see lines 389-405 of the original paper or line 446-464 of the revised paper for a discussion of the results of our sensitivity experiments). We thank the reviewer for their suggestion that we discuss this in more detail. Previously in this reply (lines 625-645 of this document), we discussed the impact that a spatially-varying δ^{18}O-temperature relationship has on our results, and provided a revised section of the paper. This is relevant because the spatial variations in the δ^{18}O-temperature relationship are a result of using precipitation-weighted temperature.

Referee #3 General Comments

Please describe the data assimilation method in more general terms understandable to the non-initiated, so the reader won't have to track down the Hakim reference. Can we think of the posterior as a cleverly weighted sum of the randomly selected model timesteps put into the prior? Is there some relationship between the posterior and the 21ka climate simulation – for example, is the posterior solution for the LGM very similar to the TraCE simulation of the LGM? Is the posterior LGM solution strongly weighted towards LGM model years randomly selelcted in the prior?

The referee is correct that this data assimilation method can be thought of as "a cleverly weighted sum of the randomly selected model timesteps put into the prior" because the ensemble mean is indeed a weighted sum of the ensemble members *if* there is no covariance localization, as is the case in our paper. The referee is also correct that it can be determined whether the posterior mean for a time step in the glacial is more strongly weighted towards ensemble members that were selected from the glacial period in TraCE-21ka; however, these weights are not a direct output of our method.

Thank you for the suggestion to describe data assimilation in more general terms. We agree that it's a good idea to provide a summary of the method. We will make the following edit to the first paragraph in Sect. 2.3, which describes the data assimilation method. The edit is in *italics*.

> To combine the ice-core and climate-model data, we use an offline data assimilation method similar to that described in Hakim et al. (2016). *If no covariance localization is used, as in this study, this method can be summed up as a linear combination of randomly-selected model states that are weighted according to new information provided by the proxy records.*

The TraCE simulation has quite a coarse grid I imagine? Please specify the exact resolution. I imagine it may even put multiple of the ice core sites in a single grid box. Perhaps the grid box resolution could be drawn onto figure 1? It seems that the spatial fields in Fig 4 are much smoother than the model would be. Did you apply smoothing or some other technique?

Yes, we agree that we should specify the exact resolution of TraCE-21ka. We have added this to Sect. 2.2 of the paper. The spatial resolution is T-31, or about 3.75 degrees, and the temporal resolution is monthly, but we average it to 50-year resolution (resulting in about 440 time steps).

We thank the reviewer for calling our attention to the fact that we forgot to say in the figure captions that the spatial fields are smoothed. We agree that it could be helpful to see the original resolution, so we have converted our plots back to the original

resolution.

How meaningful is it to use global climate simulations and constrain them only in Greenland? From a global perspective, Greenland is essentially a single location and the global climate field is not at all constrained. How well-behaved is the far-field response in the reanalysis? And does this somehow impact the reconstruction? I think doing this with global proxy databases (such as [Shakun et al., 2012] would be a great next step (beyond the scope of this paper of course).

Global reconstructions are beyond the scope of this paper, which focuses on reconstructing Greenland climate using proxy records from Greenland. We agree that a next step in applying paleoclimate data assimilation to galcial-interglacial timescales is to use a global proxy database to reconstruct global climate variables.

Seasonality is very briefly addressed, but it deserves more attention as it is an important climate parameter. Please specifically address seasonality in both the prior and posteriors. Will the reconstructions made available online have T and/or P seasonality in them, and if so, describe how this seasonality is derived. I imagine the seasonality of the posterior can be derived via the assimilation method?

We agree that seasonality deserves more attention. As we note in our reply to Referee #2, we have included a description of TraCE-21ka's seasonality in Sect. 2.2 of the paper. Seasonality is only used to compute T^*; it is in no other way included in the data assimilation process or results. Our results are mean-annual 50-year averages, which we state in the paper (line 369-370 of the revised paper). The referee is correct that climate variables for specific seasons or the seasonal cycle itself can theoretically be reconstructed using this data assimilation method.

The authors find an unusually late timing of the Holocene optimum around 5ka – much later than other ice-core based estimate from both d18O and melt layers. Looking at Fig 2, it appears that Camp Century (and perhaps Dye 3) are the only cores that suggest such timing, and since the temperature reanalysis is fully determined by ice core d18O, it follows that those two cores must be responsible for this timing (do you agree with this assessment?). However, as pointed out by [Vinther et al., 2009], these sites experience strong thinning in the first half of the Holocene, which will shift their apparent climatic optimum towards a later age (as early Holocene climatic warmth is masked by a cooler site temperature at higher elevation). Could the late (5ka) timing of the climatic optimum in your reanalysis be an artifact of the thinning history of the Greenland ice sheet? Please discuss briefly in the text.

We agree with the referee that the timing of the Holocene thermal maximum (HTM) in our reconstructions tends to be on the later end of the range of previous findings (see lines 317-325 of the original paper or lines 379-387 of the revised paper). We also agree that this signal appears to result from the Camp Century, Dye3, and perhaps the NEEM δ^{18}O records. The referee mentions how thinning in the early Holocene (Vinther et al., 2009) would affect our results. As we state in lines 127 and 460 of the original paper (lines 131 and 615 of the revised paper), our reconstructions are for climate at the ice-sheet surface, meaning that they include the lapse-rate effect of changing surface elevation. This is opposed to climate reconstructions at a reference elevation. Thus, our reconstruction of a later HTM is consistent with previous findings of an earlier HTM (given a fixed reference elevation) (e.g., McFarlin et al., 2018) and early Holocene thinning (Vinther et al., 2009). This is something we have thought about in great detail, but decided not to include in the paper; however, as suggested by the referee, we will make the following edits to the discussion section(Sect. 5, lines 455-465 of the original paper and Sect. 4.3, lines 610-627 of the revised paper). New text is in *italics*.

> An important distinction among various different paleoclimate reconstructions for Greenland is in the treatment of elevation changes. Any paleoclimate reconstruction from ice-core records is complicated by ice-sheet elevation changes. In Vinther et al. (2009), it is assumed that the climate history is the same at all locations around Greenland, and that any differences among the ice core paleotemperature records is a result of that elevation change. In B18, past elevation changes are assumed to be negligible. In our reconstruction, the impact of elevation change on the spatial covariances of temperature and precipitation is implicitly accounted for as part of the data assimilation methodology. Formally,

our reconstruction is of surface climate, not climate at a fixed elevation. *Consequently, our reanalysis may not be directly comparable to other paleoclimate reconstructions. For example, the HTM is commonly reconstructed as an early Holocene event in records that are at a fixed or nearly-fixed elevation. In our reanalysis, the early Holocene is cooler than the mid Holocene. Changes in the ice-surface elevation could account for this apparent discrepancy. Thinning in the early Holocene (Vinther et al., 2009) would result in a lowering of the ice surface and an apparent warming at the ice surface due to lapse rate effects. This warming signal would be captured in ice-core records. If the warming trend due to surface lowering occurs at the same time as an overall climate cooling, then the climate signal would be dampened or possibly reversed.*

Our method depends on the accuracy of the climate-elevation relationships in our prior – i.e. in the TraCE-21ka climate model simulation, which probably does not capture such relationships with particularly high fidelity since the model resolution is low and the climate and ice-sheet models are not coupled. Future work could take advantage of the probabilistic relationships among accumulation, temperature, and surface elevation as simulated in fine-scale regional climate models (Edwards et al., 2014).

The data assimilation is fully dependent upon the accuracy of the TraCE-21 climate model simulation in capturing Greenland climate. Therefore, the paper needs a short evaluation of how well this model actually simulates Greenland T and P in the modern day. The TraCE T and P fields should be compared to modern-day Greenland reconstructions thereof; I would recommend the works by Box et al. on this topic [Box, 2013; Box and Colgan, 2013; Box et al., 2009; Box et al., 2013], but general reanalysis products such as NCEP or ERA5 are suitable also.

The statement that "data assimilation is fully dependent upon the accuracy of the TraCE-21ka climate model simulation in capturing Greenland climate" is incorrect. Our results are dependent only on the spatial covariance patterns of the temperature anomalies and fractional precipitation in TraCE-21ka (referenced to 1850-2000 CE) and the proxy records. For our use of TraCE-21ka, what matters is how well TraCE-21ka captures the spatial pattern and variance of past climate anomalies, which is unknown except by comparison with the ice-core data (which our method does implicitly). The fidelity of the modern-day climatology is not particularly relevant.

All the figures show relative temperature changes and accumulation changes (relative to the reference period, which is not defined as far as I can tell). But when forcing ice sheet models absolute values are needed. Are these absolute values taken from the last time-slice of the TraCE simulations, or is something better used?

Before we use TraCE-21ka in our assimilation, we subtract (for temperature) or divide (for precipitation) by the mean of TraCE-21ka over the period 1850-2000 CE. Before we assimilate each proxy record, we subtract (for δ^{18}O) or divide (for accumulation rate) by the mean of that record over the same period, 1850-2000 CE. Thus, the reconstructions are of temperature anomalies and fractional precipitation, which are referenced to the 1850-2000 CE mean climate as recorded by the ice-core records. We thank the referee for pointing out that this was not entirely clear. We have stated this more clearly and earlier on in the revised paper (lines 133, 150, and each of the relevant figure captions in the revised paper).

Now that our reconstructions are complete, they may be turned into absolute values for applications such as ice-sheet modeling. This can be done by adding a 1850-2000 CE temperature climatology to the temperature fields and multiplying the precipitation fields by a 1850-2000 CE precipitation climatology (e.g., from Box, 2013). We do not do this in our paper.

Referee #3 Minor Comments

L8: What are “independent ice core records”? d18O? Again, I think the reconstructions should be compared during the abrupt temperature transitions at NEEM NGRIP and GISP2, which is where d15N-N2 provides a very robust estimate of the magni-

tude of change. Those are the truly independent ice core records to compare to.

Our use of the term "independent ice-core records" refers to the comparison of our results against δ^{18}O records that are excluded from that reconstruction iteration. This is explained later in the text around line 184 of the original paper and line 233 of the revised paper, so we will edit the abstract to clarify what is meant. See lines 823 to 837 of this document for these edits.

In our results and discussion we compare to findings from previous work; however, the referee is correct that we do not directly compare to temperature reconstructions derived from δ^{15}N of N_2. We have now done this comparison, discussed it previously in this reply, and revised the paper by including a new section, Sect. 4.2 (see lines 746-797 of this document for the revisions).

L24: This is somewhat misleading, because you'll always need to do such precip corrections unless you are doing a fully coupled ice-climate simulation. As the ice elevation in the ice sheet simulation evolves, it differs from the reference elevation at which the climate field is defined; this needs to be corrected for via clausius-clapeyron or similar. So also with your forcing the ice sheet models will need to apply thermodynamic precip corrections.

We agree that ice-sheet models need to correct precipitation fields for changing surface elevation using some assumption about the relationship between precipitation and elevation. Our point, however, is about ice-sheet simulations that base their precipitation histories entirely on their temperature histories using a thermodynamic relation. Thus, the time series and spatial pattern the precipitation anomalies perfectly match those from the temperature fields, which we know to be false from ice-core records. We have tried to clarify this distinction in the revised paper:

> Despite such evidence, paleo ice-sheet models typically assume precipitation fields that are parameterized in time using a thermodynamic relationship that is constant for all locations and timescales (e.g., Huybrechts, 2002; Greve et al., 2011).

L28: Many more d15N studies to cite here: [Guillevic et al., 2013; Kindler et al., 2014; Severinghaus and Brook, 1999; Severinghaus et al., 1998]

Thank you. We appreciate the recommended citations. We have included them throughout the revised paper where appropriate.

L38: "restricted to a single climate model realization"; wouldn't this critique apply to your study as well? It appears that both use the exact same climate model run.

Yes, the referee is correct, and we have removed the quoted language from the revised paper. Our current results rely on a single climate simulation; however, our method is easily generalizable to any climate simulation of the past 20,000 years. The method in Buizert et al. (2018) is also generalizable to other climate simulations as long as these simulations are accompanied by single-fociring experients.

L70: "measured layer thickness" is not really true. For several cores you use volcanic ties, in which case the layer thicknesses are not measured but inferred

We have changed the phrase to "the layer thickness".

L136: "captures the. . .." This is in the eye of the beholder. With the exception of the Bolling warming itself the TraCE run matches the abrupt transitions poorly – there is no YD to speak of.

We have reworded this sentence as follows: "By design, TraCE-21ka captures the major glacial-to-modern temperature change, as well as some of the short-term, rapid climate changes, such as the Bølling-Allerød transition (Liu et al., 2009)."

L145: why not use P-E? is evaporation negligible?

As we write later in the paper (lines 257-262 of the original paper and lines 318-323 of the revised paper), "accumulation is closely related to total precipitation at our ice-core sites". We demonstrate this with high-resolution model results (Langen et al., 2015, 2017) that show at most of the ice-core sites in Greenland, precipitation and accumulation are nearly equivalent, suggesting that evaporation is negligible. Dye3 is the only site where there is a difference between accumulation and precipitation, and this is due primarily to melt, not evaporation. This is why we extract P from TraCE-21ka, and use it, rather than P-E, for our proxy system model (described in Sect. 2.3.2).

L217: "highly correlated" is a strong statement. Do you have a reference? Normally d18O and site temperature are not highly correlated at most sites on observational time scales (< 0.5).

The referee is correct that the correlations are low on interannual timescales, but it is well established to be high on longer timescales (Jouzel et al., 1997).

L224: Based on the recent literature, I think that post-depositional alternation may be the largest complication in interpreting the d18O record. Please mention.

We disagree. As we note in the paper, diffusion in the firn column is irrelevant for our timescales of 50 years (Cuffey and Steig, 1998). The reviewer may be thinking of post-depositional processes involving water exchange between the snowpack and the atmosphere (e.g., Steen-Larsen et al., 2011), but it is not established that this has any significant effect on long-term relationships. Indeed, if anything, such processes improve the relationship between temperature and δ^{18}O as they tend to reduce the bias caused by the fact that snow does not accumulated continuously, but as discrete events.

L241: Can you plot T_{site} and T^*_{site} together for the last 21ka at a key site (e.g. Summit). That will let the reader judge the impact of using T^* instead of T.

Thank you for the suggestion. We have done this in Fig. S4, which we have included in the supplementary information.

How is the seasonality of the posterior linked to the seasonality of the prior?

As we stated previously, we only use seasonality to compute T^*; it is in no other way included in the data assimilation process or results. Both our prior ensemble and our reconstructions are made up of mean-annual 50-year averages.

L261: "grid-cell closest to site" is this also done for T, or do you use 2D linear interpolation or similar? Are there cases where multiple cores share a closest grid cell?

Yes, this is also done for selecting which T^* value to use in the δ^{18}O PSM. We have clarified that in the revised paper. At the resolution of TraCE-21ka, only the GISP2 and GRIP ice cores have the same closest grid cell.

L295: maybe a sentence on how this was estimated?

Thank you for bringing our attention to this point of confusion. We have clarifed this in the revised paper. We use the same method as was explained for δ^{18}O.

L313-314: But [Dahl-Jensen et al., 1998] estimates it a lot colder at GRIP, more like -22K cooling at the LGM (25ka). This should be mentioned.

Dahl-Jensen et al. (1998) found a colder temperature at GRIP for the LGM, but the time period we discuss is 20-15ka, which is five to ten thousand years later, when the Dahl-Jensen et al. (1998) estimate shows that it has significantly warmed.

L340: Maybe reference [Buchardt et al., 2012] who did very detailed analyses of this.

Thank you. We are aware of this study, and have included it in our discussion of the temperature-precipitation relationship on shorter timescales (lines 484-494 of the revised paper).

L416: are other d18O records really independent? They suffer the same biases from seasonality, source effects, etc. For true independence, compare to d15N-N2.

We agree with the referee that it is important to compare our results to other types of proxy records. In this reply we have compared our temperature reconstructions to those from Buizert et al. (2014) and revised the paper.

L438: TraCE has no HTM anywhere! (one of its many problems. . .)

We have changed the wording to say, "TraCE-21ka has no obvious HTM in this location or any location around Greenland."

L476: This is more of a discussion than a conclusion item. Consider moving it. Also, see my comment above, the 5ka timing could be an artifact of ice sheet elevation changes.

We have moved this paragraph to the discussion (Sect. 4.3 of the revised paper). We have also included a discussion of elevation effects (see lines 919-950 of this document).

Figure 4: please add panels (e) and (f) with the T and P change over an abrupt transition (e.g. the Bolling onset). In panel (c), only show the cores that actually constrain the LGM (so not Agassiz, camp century and Renland). Why are the field so much smoother than the TraCE CCSM3 model resolution? Baffin bay has a large temp response with no cores to constrain it – can we trust this?

Thank you for these suggestions. We have added spatial plots of the abrupt temperature transitions (Figs. 12, S12, and S13). We have also included only the ice core locations that contribute to the reconstruction of each time period. As we mentioned previously in this reply, we have restored the plots to their original resolution because we agree with the reviewer that this is helpful information. We had originally smoothed the plots due to rendering issues, but we have fixed the previous problem.

The goal in using a method like data assimilation is to spread the information from point proxy locations to locations without proxy records. This allows us to reconstruct spatially-complete climate fields. With few proxies and many locations, the problem is underconstrained; however, we have shown that the method is skillful for some locations where proxies are not assimilated (see Sect. 4.1, lines 364-382 of the original paper and Sect. 3.2, lines 401-439 of the revised paper).

Fig 5: the “noise” in T (i.e. high frequency signals) at all core sites seem strongly correlated. How come? Could it be that the posterior is more or less reflecting the mean d18O of the various sites?

Each core has some influence on the reconstruction at every location. Thus, the time series at each location is a weighted sum of the time series of each core. This would indeed make the higher-frequency noise correlated at different locations.

Fig 6: The largest features in the plot are not directly constrained by any cores. Do you trust these?

For our reply, please see lines 1115-1118 in this document.

Figs 7 and 8 are very technical and could be moved to the supplement.

These figures demonstrate how well our reconstruction performs at locations without any assimilated information. We think this is an important aspect of our paper, and we wish to keep these figures in the main text.

References

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Monthly Weather Review, 127, 2741–2758, 1999.

Bothe, O., Evans, M., Donado, L., Bustamante, E., Gergis, J., Gonzalez-Ruoco, J., Goosse, H., Hegerl, G., Hind, A., Jungclaus, J., et al.: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past, 11, 1673–1699, 2015.

Box, J. E.: Greenland ice sheet mass balance reconstruction. Part II: Surface mass balance (1840–2010), Journal of Climate, 26, 6974–6989, 2013.

Buizert, C., Gkinis, V., Severinghaus, J. P., He, F., Lecavalier, B. S., Kindler, P., Leuenberger, M., Carlson, A. E., Vinther, B., Masson-Delmotte, V., et al.: Greenland temperature response to climate forcing during the last deglaciation, Science, 345, 1177–1180, 2014.

Buizert, C., Keisling, B., Box, J., He, F., Carlson, A., Sinclair, G., and DeConto, R.: Greenland-Wide Seasonal Temperatures During the Last Deglaciation, Geophysical Research Letters, 45, 1905–1914, 2018.

Cauquoin, A., Werner, M., and Lohmann, G.: Water isotopes-climate relationships for the mid-Holocene and preindustrial period simulated with an isotope-enabled version of MPI-ESM, Climate of the Past, 15, 1913–1937, 2019.

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought reconstructions for the continental United States, Journal of Climate, 12, 1145–1162, 1999.

Cuffey, K. M. and Clow, G. D.: Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition, Journal of Geophysical Research: Oceans, 102, 26 383–26 396, 1997.

Cuffey, K. M. and Steig, E. J.: Isotopic diffusion in polar firn: implications for interpretation of seasonal climate parameters in ice-core records, with emphasis on central Greenland, Journal of Glaciology, 44, 273–284, 1998.

Dahl-Jensen, D., Mosegaard, K., Gundestrup, N., Clow, G. D., Johnsen, S. J., Hansen, A. W., and Balling, N.: Past temperatures directly from the Greenland ice sheet, Science, 282, 268–271, 1998.

Edwards, T., Fettweis, X., Gagliardini, O., Gillet-Chaulet, F., Goelzer, H., Gregory, J., Hoffman, M., Huybrechts, P., Payne, A., Perego, M., et al.: Probabilistic parameterisation of the surface mass balance–elevation feedback in regional climate model simulations of the Greenland ice sheet, The Cryosphere, 8, 181–194, 2014.

Gierz, P., Werner, M., and Lohmann, G.: Simulating climate and stable water isotopes during the L ast I nterglacial using a coupled climate-isotope model, Journal of Advances in Modeling Earth Systems, 9, 2027–2045, 2017.

Greve, R., Saito, F., and Abe-Ouchi, A.: Initial results of the SeaRISE numerical experiments with the models SICOPOLIS and IcIES for the Greenland ice sheet, Annals of Glaciology, 52, 23–30, 2011.

Guillevic, M., Bazin, L., Landais, A., Kindler, P., Orsi, A., Masson-Delmotte, V., Blunier, T., Buchardt, S. L., Capron, E., Leuenberger, M., Martinerie, P., Prié, F., and Vinther, B. M.: Spatial gradients of temperature, accumulation and δ 18 O-ice in Greenland over a series of Dansgaard-Oeschger events., Climate of the Past, 9, https://doi.org/10.5194/cp-9-1029-2013, 2013.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, 2016.

Harrison, S., Bartlein, P., Brewer, S., Prentice, I., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Climate Dynamics, 43, 671–688, 2014.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, 2009.

He, F.: Simulating transient climate evolution of the last deglaciation with CCSM 3, vol. 72, 2011.

He, F., Shakun, J. D., Clark, P. U., Carlson, A. E., Liu, Z., Otto-Bliesner, B. L., and Kutzbach, J. E.: Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation, Nature, 494, 81, 2013.

Huybrechts, P.: Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles, Quaternary Science Reviews, 21, 203–231, 2002.

Jouzel, J., Alley, R. B., Cuffey, K., Dansgaard, W., Grootes, P., Hoffmann, G., Johnsen, S. J., Koster, R., Peel, D., Shuman, C., et al.: Validity of the temperature reconstruction from water isotopes in ice cores, Journal of Geophysical Research: Oceans, 102, 26 471–26 487, 1997.

Kindler, P., Guillevic, M., Baumgartner, M. F., Schwander, J., Landais, A., and Leuenberger, M.: Temperature reconstruction from 10 to 120 kyr b2k from the NGRIP ice core, Climate of the Past, 10, 887–902, 2014.

Kobashi, T., Menviel, L., Jeltsch-Thömmes, A., Vinther, B. M., Box, J. E., Muscheler, R., Nakaegawa, T., Pfister, P. L., Döring, M., Leuenberger, M., et al.: Volcanic influence on centennial to millennial Holocene Greenland temperature change, Scientific reports, 7, 1441, 2017.

Krinner, G. and Werner, M.: Impact of precipitation seasonality changes on isotopic signals in polar ice cores: a multi-model analysis, Earth and Planetary Science Letters, 216, 525–538, 2003.

Langen, P., Mottram, R., Christensen, J., Boberg, F., Rodehacke, C., Stendel, M., Van As, D., Ahlstrøm, A., Mortensen, J., Rysgaard, S., et al.: Quantifying energy and mass fluxes controlling Godthåbsfjord freshwater input in a 5-km simulation (1991–2012), Journal of Climate, 28, 3694–3713, 2015.

Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H., and Box, J. E.: Liquid water flow and retention on the Greenland ice sheet in the regional climate model HIRHAM5: Local and large-scale impacts, Frontiers in Earth Science, 4, 110, 2017.

Latif, M. and Keenlyside, N. S.: A perspective on decadal climate variability and predictability, Deep Sea Research Part II: Topical Studies in Oceanography, 58, 1880–1894, 2011.

Li, C., Battisti, D. S., and Bitz, C. M.: Can North Atlantic sea ice anomalies account for Dansgaard–Oeschger climate signals?, Journal of climate, 23, 5457–5475, 2010.

Liu, Z., Otto-Bliesner, B., He, F., Brady, E., Tomas, R., Clark, P., Carlson, A., Lynch-Stieglitz, J., Curry, W., Brook, E., et al.: Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming, Science, 325, 310–314, 2009.

Liu, Z., Carlson, A. E., He, F., Brady, E. C., Otto-Bliesner, B. L., Briegleb, B. P., Wehrenberg, M., Clark, P. U., Wu, S., Cheng, J., et al.: Younger Dryas cooling and the Greenland climate response to CO2, Proceedings of the National Academy of Sciences, 109, 11 101–11 104, 2012.

Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L., and Brewer, S.: Reconciling divergent trends and millennial variations in Holocene temperatures, Nature, 554, 92–96, 2018.

McFarlin, J. M., Axford, Y., Osburn, M. R., Kelly, M. A., Osterberg, E. C., and Farnsworth, L. B.: Pronounced summer warming in northwest Greenland during the Holocene and Last Interglacial, Proceedings of the National Academy of Sciences, 115, 6357–6362, 2018.

Obase, T. and Abe-Ouchi, A.: Abrupt Bølling-Allerød Warming Simulated under Gradual Forcing of the Last Deglaciation, Geophysical Research Letters, 46, 11 397–11 405, 2019.

Pausata, F. S. and Löfverström, M.: On the enigmatic similarity in Greenland δ18O between the Oldest and Younger Dryas, Geophysical Research Letters, 42, 10–470, 2015.

Pedro, J. B., Bostock, H. C., Bitz, C. M., He, F., Vandergoes, M. J., Steig, E. J., Chase, B. M., Krause, C. E., Rasmussen, S. O., Markle, B. R., et al.: The spatial extent and dynamics of the Antarctic Cold Reversal, Nature Geoscience, 9, 51–55, 2016.

Peltier, W.: Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE, Annu. Rev. Earth Planet. Sci., 32, 111–149, 2004.

Robin, G. d. Q.: Ice cores and climatic change, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 280, 143–168, 1977.

Roy, K. and Peltier, W.: Relative sea level in the Western Mediterranean basin: A regional test of the ICE-7G_NA (VM7) model and a constraint on Late Holocene Antarctic deglaciation, Quaternary Science Reviews, 183, 76–87, 2018.

Severinghaus, J. P. and Brook, E. J.: Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice, Science, 286, 930–934, 1999.

Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B., and Bender, M. L.: Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice, Nature, 391, 141–146, 1998.

Steen-Larsen, H. C., Masson-Delmotte, V., Sjolte, J., Johnsen, S. J., Vinther, B. M., Bréon, F.-M., Clausen, H., Dahl-Jensen, D., Falourd, S., Fettweis, X., et al.: Understanding the climatic signal in the water stable isotope records from the NEEM shallow firn/ice cores in northwest Greenland, Journal of Geophysical Research: Atmospheres, 116, 2011.

Steig, E. J., Grootes, P. M., and Stuiver, M.: Seasonal precipitation timing and ice core records, Science, 266, 1885–1887, 1994.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last millennium reanalysis with an expanded proxy database and seasonal proxy modeling, Climate of the Past, 15, 1251–1273, 2019.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of hydrology, 456, 12–29, 2012.

Vinther, B. M., Buchardt, S. L., Clausen, H. B., Dahl-Jensen, D., Johnsen, S. J., Fisher, D., Koerner, R., Raynaud, D., Lipenkov, V., Andersen, K. K., et al.: Holocene thinning of the Greenland ice sheet, Nature, 461, 385, 2009.

Werner, M., Mikolajewicz, U., Heimann, M., and Hoffmann, G.: Borehole versus isotope temperatures on Greenland: Seasonality does matter, Geophysical Research Letters, 27, 723–726, 2000.

Zhang, Y., Renssen, H., Seppä, H., and Valdes, P. J.: Holocene temperature evolution in the Northern Hemisphere high latitudes–Model-data comparisons, Quaternary Science Reviews, 173, 101–113, 2017.

Zhang, Y., Renssen, H., Seppä, H., and Valdes, P. J.: Holocene temperature trends in the extratropical Northern Hemisphere based on inter-model comparisons, Journal of Quaternary Science, 33, 464–476, 2018.

Figures

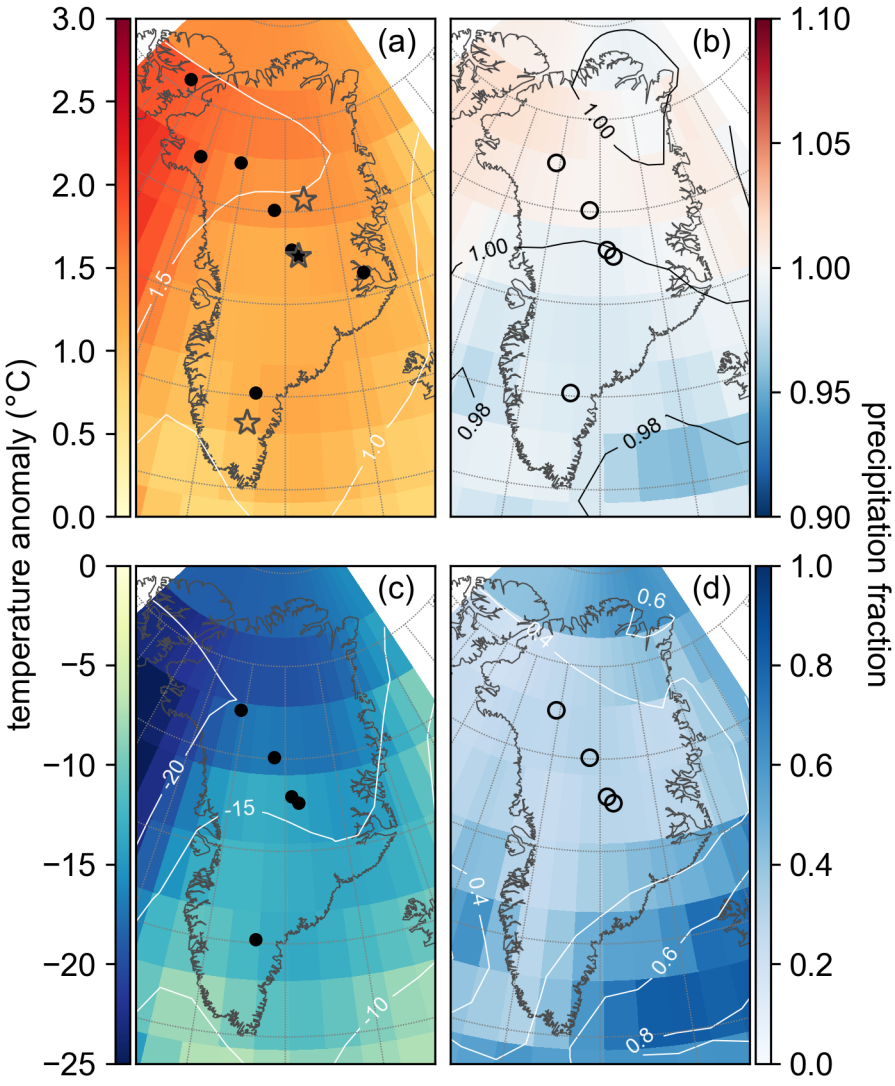

Figure 4. Spatial pattern of the reanalysis mean for temperature (panels (a), (c)) and precipitation (panels (b), (d)) with contours for clarity. (a) and (b) are averaged over 1,000 years around the peak warmth in the Holocene, 5.5-4.5 ka, while (c) and (d) are averaged over 5,000 years in the late glacial, 20-15 ka. Anomalies and fractions are with respect to the mean of 1850-2000 CE. Points show ice-core locations used for each reanalysis with closed circles indicating δ^{18}O records and open circles indicating accumulation records. Grey stars show the locations of the EGRIP ice-core site, Summit, and South Dome, which are referenced in Figs. 5 and 11.

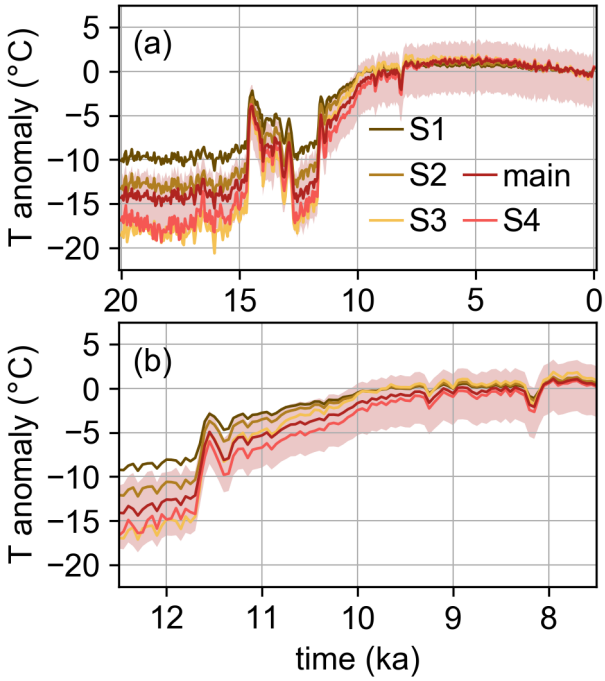

Figure 9. The main temperature (T) reanalysis (ensemble mean and 5^{th} to 95^{th} percentile shading) and ensemble mean for four sensitivity scenarios, S1-S4. Each sensitivity scenario reflects a different assumption about precipitation seasonality, with S1-S3 assuming a spatially-uniform seasonality and S3-S4 assuming stronger seasonality than the main reanalysis. Anomalies are with respect to the mean of 1850-2000 CE. These time series are for the location closest to Summit, which is representative of the results around Greenland. Panel (a) shows the full time period, and panel (b) the Younger-Dryas through the Holocene, showing that S1, S2, and S3 warm more quickly than the main reconstruction and S4.

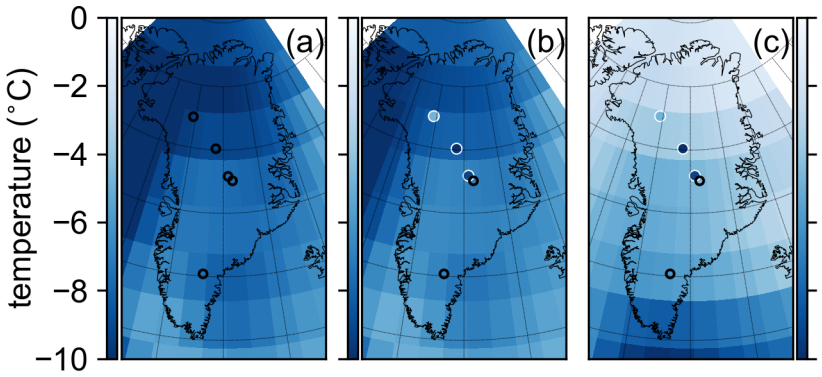

Figure 12. Spatial pattern of the abrupt cooling event into the Younger Dryas. Panel (a) shows results from experiment O8, assimilating all eight δ^{18}O records, panel (b) shows results from experiment N3O5, assimilating all three δ^{15}N-derived temperature records and the remaining five δ^{18}O records (those that do not overlap with the δ^{15}N sites), and panel (c) shows results from experiment N3O5_BA, which is similar to the N3O5 experiment except the prior ensemble is selected from the 1,000 years surrounding the Bølling-Allerød warming. Unfilled black circles show locations of assimilated δ^{18}O records, while filled circles with white outlines show locations of assimilated δ^{15}N-derived temperature records. Filled circles in panels (b) and (c) show the δ^{15}N-derived temperature values as reported by Buizert et al. (2014) on the same color scale as the rest of the panel. The temporal definition of this event is the same as defined in Buizert et al. (2014).

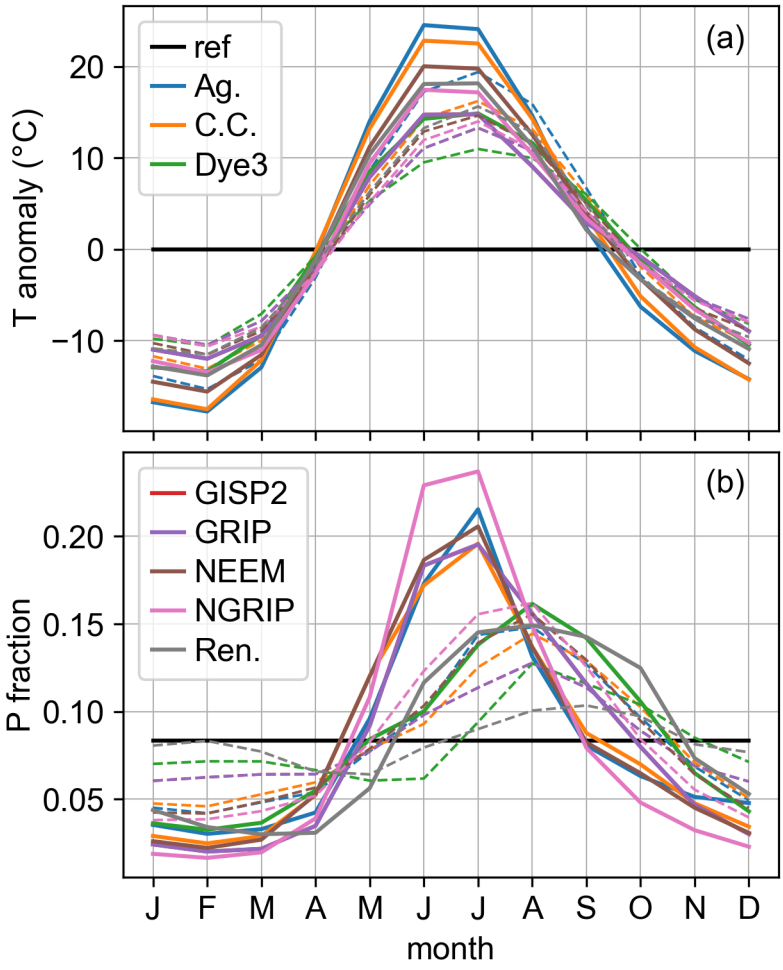

Figure S3. A comparison of TraCE-21ka glacial and Holocene seasonality at each ice-core site considered in this study. The mean glacial (20 to 15 ka) seasonality is shown as solid lines and the mean Holocene (5 to 0 ka) seasonality is shown as dashed lines. Panel (a) shows the monthly temperature anomaly referenced to the annual mean. Panel (b) shows the fraction of total annual temperature that fell each month. For both panels, the reference line (black) shows no seasonal cycle. Both the magnitude and the timing of the seasonal cycle change between the glacial and the Holocene, with the glacial generally showing a stronger seasonal cycle and an earlier summer peak. ref = reference line, Ag. = Agassiz, C.C. = Camp Century, and Ren. = Renland.

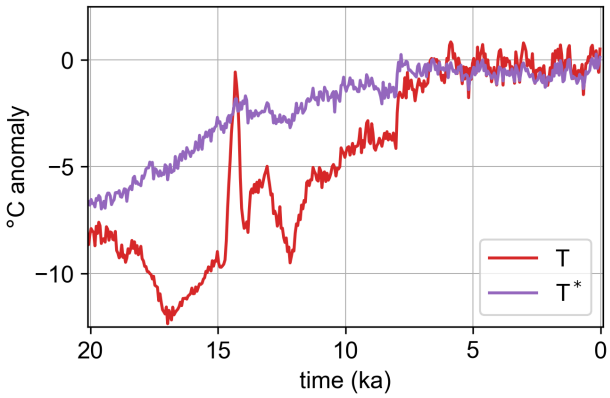

Figure S4. Temperature (T) and temperature weighted by monthly precipitation (T^*) from TraCE-21ka at Summit, Greenland. Both variables are shown as anomalies with respect to 1850-2000 CE and have been averaged to 50-year resolution. T^* was computed before the anomaly was taken.

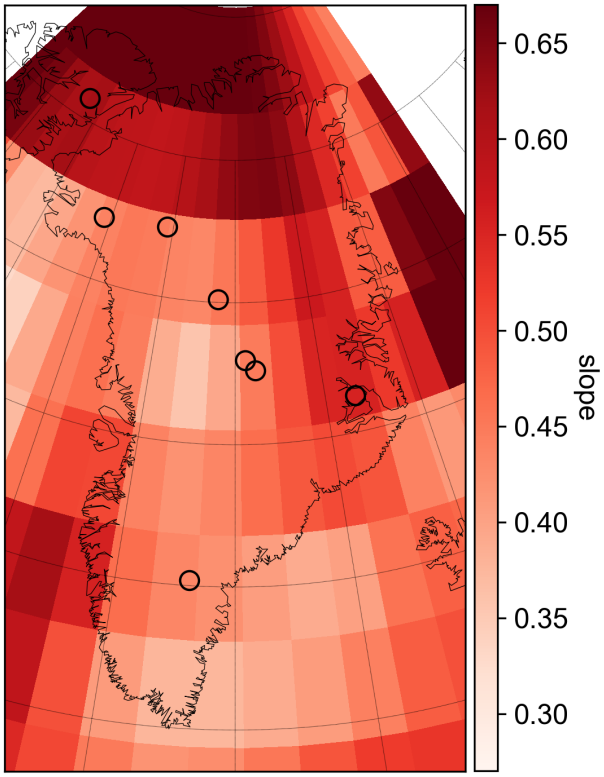

Figure S5. Slope (°C $‰^{-1}$) of the linear δ^{18}O-temperature relationship for each grid cell. The δ^{18}O-temperature relationship between grid cells is not shown. This is an example from one of the prior ensembles. For reference, open circles show the locations of ice-core records used in this study.

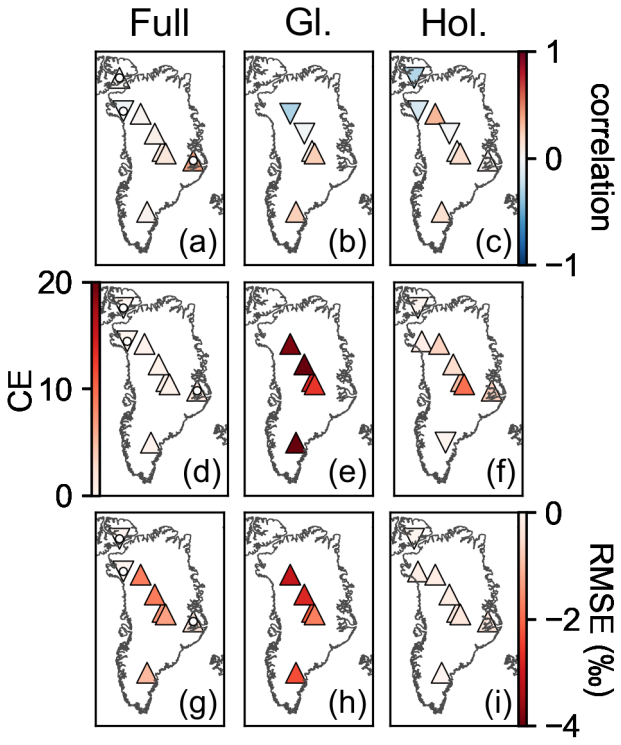

Figure S8. Change in skill metrics from TraCE-21ka to posterior ensemble averaged over iterations and time for the temperature reanalysis. The first column (panels (a), (d), (g), and (j)) shows the skill-metric change for the full overlap (Full) between the proxy record and reanalysis. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill-metric change for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the change in correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble evaluates better than the prior ensemble for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

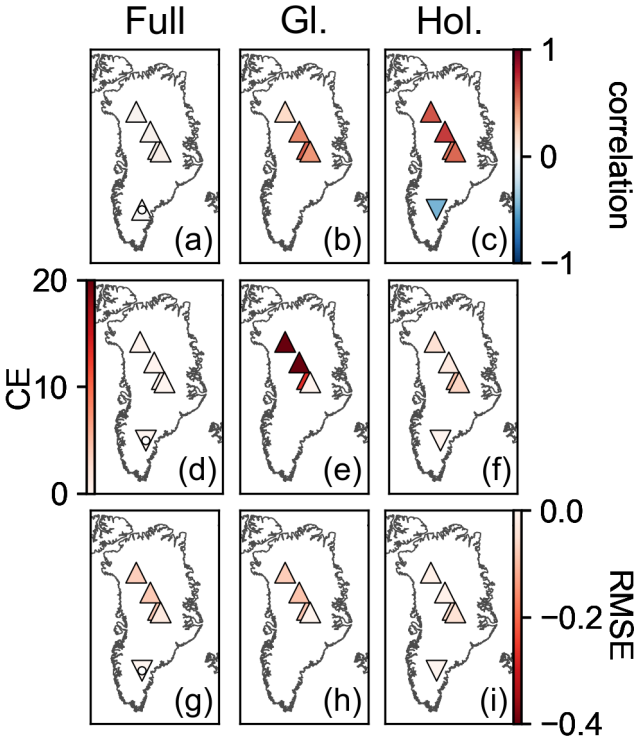

Figure S9. Change in skill metrics from TraCE-21ka to posterior ensemble averaged over iterations and time for the precipitation reanalysis. The first column (panels (a), (d), (g), and (j)) shows the skill-metric change for the full overlap (Full) between the proxy record and reanalysis. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill-metric change for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the change in correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble evaluates better than the prior ensemble for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

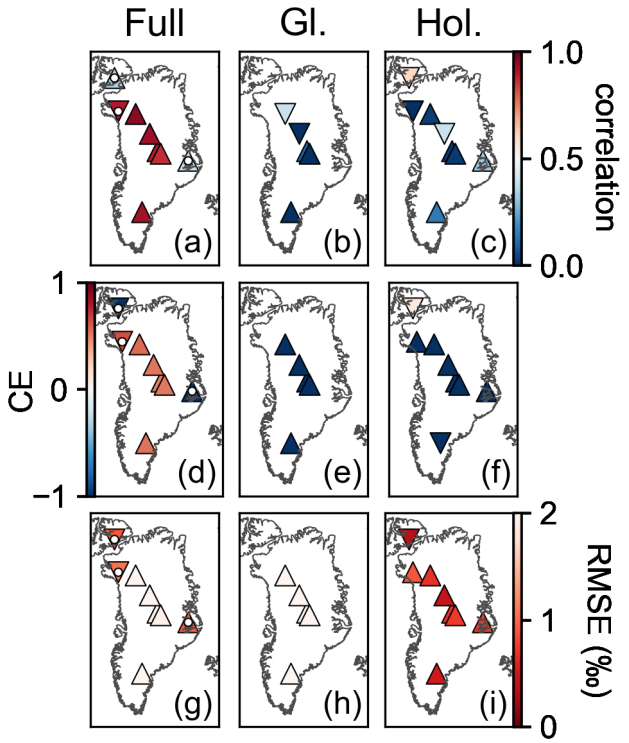

Figure S10. TTemperature skill metrics for the TraCE-21ka simulation. The first column (panels (a), (d), (g), and (j)) shows the skill metrics for the full overlap (Full) between the proxy record and TraCE-21ka. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill metrics for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble of our main reanalysis evaluates better than TraCE-21ka for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

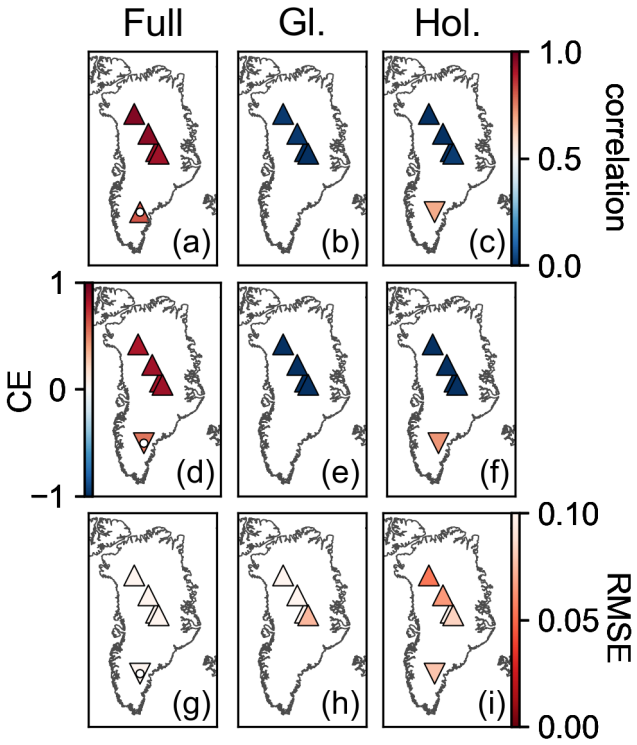

Figure S11. Precipitation skill metrics for the TraCE-21ka simulation. The first column (panels (a), (d), (g), and (j)) shows the skill metrics for the full overlap (Full) between the proxy record and TraCE-21ka. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill metrics for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble of our main reanalysis evaluates better than TraCE-21ka for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

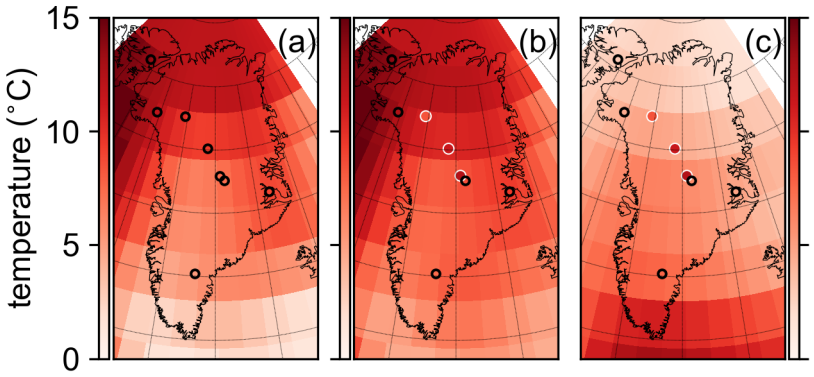

Figure S12. Spatial pattern of the abrupt warming event out of the Younger Dryas. Panel (a) shows results from experiment O8, assimilating all eight δ^{18}O records, panel (b) shows results from experiment N3O5, assimilating all three δ^{15}N-derived temperature records and the remaining five δ^{18}O records (those that do not overlap with the δ^{15}N sites), and panel (c) shows results from experiment N3O5_BA, which is similar to the N3O5 experiment except the prior ensemble is selected from the 1,000 years surrounding the Bølling-Allerød warming. Unfilled black circles show locations of assimilated δ^{18}O records, while filled circles with white outlines show locations of assimilated δ^{15}N-derived temperature records. Filled circles in panels (b) and (c) show the δ^{15}N-derived temperature values as reported by Buizert et al. (2014) on the same color scale as the rest of the panel. The temporal definition of this event is the same as defined in Buizert et al. (2014).

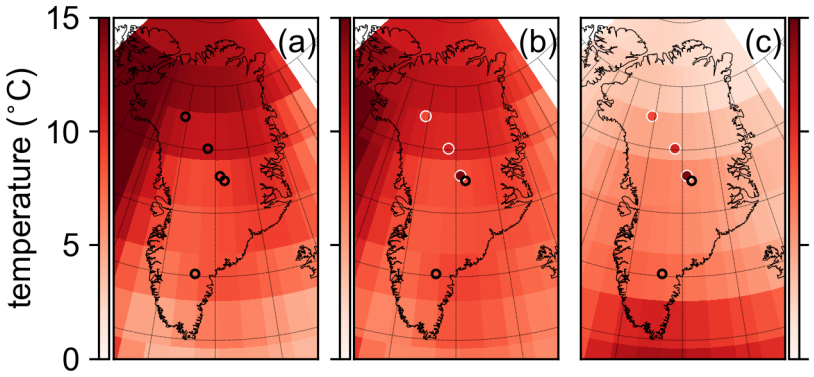

Figure S13. Spatial pattern of the abrupt warming event into the Bølling-Allerød. Panel (a) shows results from experiment O8, assimilating all eight δ^{18}O records, panel (b) shows results from experiment N3O5, assimilating all three δ^{15}N-derived temperature records and the remaining five δ^{18}O records (those that do not overlap with the δ^{15}N sites), and panel (c) shows results from experiment N3O5_BA, which is similar to the N3O5 experiment except the prior ensemble is selected from the 1,000 years surrounding the Bølling-Allerød warming. Unfilled black circles show locations of assimilated δ^{18}O records, while filled circles with white outlines show locations of assimilated δ^{15}N-derived temperature records. Filled circles in panels (b) and (c) show the δ^{15}N-derived temperature values as reported by Buizert et al. (2014) on the same color scale as the rest of the panel. The temporal definition of this event is the same as defined in Buizert et al. (2014).

Tables

Table S1. The mean slope for the linear δ^{18}O-temperature relationship used for the main reconstruction in this study (black, "main") and the mean slope for the relationship used in the S4 sensitivity experiment in this study (red, "S4"). We also include estimates from previous work, inlcuding, slopes found by Buizert et al. (2014) (purple, "B14") as seen in their Fig. 3 for five discontinuous time periods between 20 and 10 ka; slopes found by Guillevic et al. (2013) (green, "G13") from their Table 3 for Dansgaard-Oeschger events 8, 9, and 10; slopes found by Kindler et al. (2014) (blue, "K14") from their Fig. 5 and estimated from their Fig. 6 for 120 to 10 ka; and slopes found by Cuffey and Clow (1997) (orange, "C97") for time periods between 50 and 0.5 ka. The cores are arraged from North (top) to South (bottom).

Core Name	**Slope Range (°C ‰$^{-1}$)**	**Slope Average (°C ‰$^{-1}$)**
Agassiz	main: 0.618 - 0.656 S4: 0.412 - 0.437	0.640 0.425
Camp Century	main: 0.439 - 0.468 S4: 0.293 - 0.312	0.456 0.304
NEEM	main: 0.450 - 0.480 S4: 0.300 - 0.320 B14: 0.25 - 0.76 G13: 0.51 - 0.63	0.465 0.310 0.50 0.57
NGRIP	main: 0.454 - 0.489 S4: 0.303 - 0.326 B14: 0.29 - 0.41 G13: 0.34 - 0.47 K14: 0.3 - 0.57	0.470 0.313 0.36 0.42 0.52
GISP2	main: 0.442 - 0.493 S4: 0.294 - 0.329 B14: 0.11 - 0.30 G13: 0.38 C97: 0.251 - 0.465	0.467 0.311 0.25 0.324
GRIP	main: 0.442 - 0.493 S4: 0.294 - 0.329 G13: 0.49	0.467 0.311
Renland	main: 0.546 - 0.595 S4: 0.364 - 0.397	0.571 0.381
Dye3	main: 0.424 - 0.475 S4: 0.283 - 0.317	0.444 0.296

Table S2. Comparison of abrupt climate transitions in our main reconstruction and sensitivity reconstructions, S1, S2, S3, and S4. We use the same time definitions as in Buizert et al. (2014). Note that the main reconstruction and S4 do not warm as rapidly out of the Younger Dryas as S1, S2, and S3. Thus, our use of a single time definition may not allow us capture the full transition for all of these reconstructions.

Core Name	Reconstruction Name	Bølling-Allerød warming	Younger Dryas cooling	Younger Dryas warming
Agassiz	Main (δ^{18}O= $0.67T^*$)	12.83	-9.10	11.64
	S1 (δ^{18}O= $0.67T$)	8.67	-5.16	7.78
	S2 (δ^{18}O= $0.5T$)	11.20	-6.96	10.50
	S3 (δ^{18}O= $0.335T$)	15.41	-9.99	14.84
	S4 (δ^{18}O= $0.67T^*$, stronger P seasonality)	15.49	-11.05	13.06
Camp Century	Main	11.71	-8.51	10.53
	S1	8.04	-4.94	7.13
	S2	10.40	-6.60	9.65
	S3	14.34	-9.41	13.71
	S4	14.07	-10.36	11.46
NEEM	Main	10.17	-7.53	9.10
	S1	7.11	-4.51	6.34
	S2	9.22	-5.99	8.33
	S3	12.74	-8.47	12.16
	S4	12.18	-9.20	9.71
NGRIP	Main	9.62	-7.17	8.57
	S1	6.76	-4.33	6.03
	S2	8.77	-5.73	8.14
	S3	12.13	-8.09	11.57
	S4	11.52	-8.77	9.07
GISP2	Main	7.78	-6.03	6.88
	S1	5.66	-3.88	5.18
	S2	7.39	-5.06	6.92
	S3	10.26	-7.03	9.81
	S4	9.27	-7.41	7.03
GRIP	Main	7.78	-6.03	6.88
	S1	5.66	-3.88	5.18
	S2	7.39	-5.06	6.92
	S3	10.26	-7.03	9.81
	S4	9.27	-7.41	7.03
Renland	Main	8.51	-6.27	7.77
	S1	5.93	-3.78	5.39
	S2	7.70	-5.02	7.22
	S3	10.64	-7.08	10.22
	S4	10.15	-7.62	8.53
Dye3	Main	6.45	-5.64	5.41
	S1	5.33	-4.34	5.26
	S2	7.05	-5.45	6.84
	S3	9.92	-7.29	9.60
	S4	7.62	-7.10	4.79

Table S3. Scaling factors (β) for the temperature-precipitation relationship in the Kangerlussuaq region. The results for the main reanalysis are in bold.

Temperature scenario	**Precipitation scenario**	**No Filtering**	**Low-Pass (5,000 years^{-1})**	**High-Pass (5,000 years^{-1})**
Main	Low	0.09	0.09	0.04
Main	**Moderate**	**0.08**	**0.08**	**0.04**
Main	High	0.08	0.08	0.05
S1	Low	0.12	0.11	0.05
S1	Moderate	0.11	0.11	0.06
S1	High	0.11	0.11	0.07
S2	Low	0.09	0.09	0.04
S2	Moderate	0.09	0.08	0.04
S2	High	0.09	0.08	0.05
S3	Low	0.06	0.06	0.03
S3	Moderate	0.06	0.06	0.03
S3	High	0.06	0.06	0.04
S4	Low	0.07	0.07	0.03
S4	Moderate	0.07	0.07	0.03
S4	High	0.07	0.07	0.04

Table S4. Comparison of abrupt climate transitions in our main reconstruction (black) and from Buizert et al. (2014) (purple) values in the second row for NEEM, NGRIP, and GISP2). Uncertainties are given as standard deviations. We use the published values and uncertainties from Buizert et al. (2014). We use the same time definitions as in Buizert et al. (2014).

Core Name	**Bølling-Allerød warming**	**Younger Dryas cooling**	**Younger Dryas warming**
Agassiz	12.83 ± 3.49	-9.10 ± 3.65	11.64 ± 3.46
Camp Century	11.71 ± 3.23	-8.51 ± 3.42	10.53 ± 3.24
NEEM	10.17 ± 3.01 8.9 ± 1.2	-7.53 ± 3.2 -4.8 ± 0.6	9.10 ± 3.05 8.4 ± 1.5
NGRIP	9.62 ± 2.90 10.8 ± 1.0	-7.17 ± 3.09 -10.9 ± 1.0	8.57 ± 2.95 11.4 ± 1.6
GISP2	7.78 ± 2.77 14.4 ± 0.95	-6.03 ± 2.96 -9.2 ± 0.45	6.88 ± 2.84 12.4 ± 1.7
GRIP	7.78 ± 2.77	-6.03 ± 2.96	6.88 ± 2.84
Renland	8.51 ± 2.55	-6.27 ± 2.70	7.77 ± 2.57
Dye3	6.45 ± 3.84	-5.64 ± 4.10	5.41 ± 3.96

2) Summary of Changes Made to the Manuscript

Here we summarize the major changes made to the manuscript and Supplementary Information. All manuscript edits are shown in the marked-up version below.

1. In the Supplementary Information we added Sect. S1, which is a discussion of the pros and cons of four different prior ensemble options. The point of the section is to clarify our choice of a time-constant prior ensemble over time-period specific prior ensembles.

2. We expanded Sect. 2.2, "Climate-model simulation", to include a more thorough description of the TraCE-21ka simulation, especially on the topics of seasonality and prescribed ice sheet evolution. This change was made to provide more context for a major component of the data assimilation method: the climate simulation from which the prior ensemble is drawn.

3. In Sect. 2.3.1, we expanded our discussion of the δ^{18}O-temperature slope as derived from the prior ensemble. Two new figures and one new table were added for visual clarification (Figs. S4 and S5; Table S1). These changes were made to clarify that the slopes we use are spatially-variable and that this is due to the inclusion of precipitation-weighted temperature (T^*) in the δ^{18}O proxy system model (PSM).

4. We rearranged the results (Sect. 3) and discussion (previously Sect. 5, now Sect. 4) to increase clarity and flow of the paper. The evaluation section (previously Sect. 4) was combined with the results section (now Sects. 3.2 and 3.3). The discussion of the relationship between temperature and precipitation was moved from the results to the discussion section (it is now Sect. 4.1).

5. We added a new section (Sect. 4.2) describing three new experiments to show the impact of including δ^{15}N-derived temperature records and a time-period specific prior ensemble from the Bølling-Allerød. Three new figures and one new table accompany this new section (Figs. 12, S12, and S13; Table S4). A new Supplementary Information section (Sect. S6) was also added to detail the methods of these new experiments.

6. Numerous minor edits were made in response to all three referees. These are described in the replies above.

3) Marked-up Manuscript

Greenland temperature and precipitation over the last 20,000 years using data assimilation

Jessica A. Badgeley[1], Eric J. Steig[1,2], Gregory J. Hakim[2], and Tyler J. Fudge[1]

[1]Department of Earth and Space Sciences, University of Washington
[2]Department of Atmospheric Sciences, University of Washington

Correspondence: Jessica Badgeley (badgeley@uw.edu)

Abstract. Reconstructions of past temperature and precipitation are fundamental to modeling the Greenland Ice Sheet and assessing its sensitivity to climate. Paleoclimate information is sourced from proxy records and climate-model simulations; however, the former are spatially incomplete while the latter are sensitive to model dynamics and boundary conditions. Efforts to combine these sources of information to reconstruct spatial patterns of Greenland climate over glacial-interglacial cycles have been limited by assumptions of fixed spatial patterns and a restricted use of proxy data. We avoid these limitations by using paleoclimate data assimilation to create independent reconstructions of temperature and precipitation for the last 20,000 years. Our method uses ~~information~~ oxygen-isotope ratios of ice and accumulation rates from long ice-core records and extends ~~it~~ this information to all locations across Greenland using spatial relationships derived from a transient climate-model simulation. ~~Our reconstructions evaluate well against independent~~ Standard evaluation metrics for this method show that our results capture climate at locations without ice-core records. ~~In addition, we find that the~~ Our results differ from previous work in the reconstructed spatial pattern of temperature change during abrupt climate transitions; this indicates a need for additional proxy data and additional transient climate-model simulations. We investigate the relationship between precipitation and temperature, finding that it is frequency dependent and spatially variable, suggesting that thermodynamic scaling methods commonly used in ice-sheet modeling are overly simplistic. Our results demonstrate that paleoclimate data assimilation is a useful tool for reconstructing the spatial and temporal patterns of past climate on timescales relevant to ice sheets.

1 Introduction

Predicting the future behavior of the Greenland Ice Sheet requires understanding ~~the sensitivity of the ice sheet~~ its sensitivity to changes in temperature and precipitation (Bindschadler et al., 2013). One important constraint on this sensitivity is the response of the paleo ice-sheet to changing climate in the past~~(Alley et al., 2010)~~. On glacial-interglacial timescales, temperature, not precipitation, appears to be the dominant control on the size of the Greenland Ice Sheet (Alley et al., 2010), as evidenced by the fact that the ice sheet is largest during cold and arid glacial periods and smallest during warm and wet interglacials. On these timescales, precipitation over the Greenland Ice Sheet scales positively with temperature (Robin, 1977), as anticipated by the Clausius-Clapeyron relation between temperature and saturation vapor pressure. ~~However, ice-core records~~Ice-core records, however, show that this thermodynamic relation is a poor approximation on annual to multi-millennial timescales

(Kapsner et al., 1995; Fudge et al., 2016). For example, the GISP2 ice core from central Greenland shows that cooling coincided with increased snowfall between the early Holocene and present (Cuffey and Clow, 1997). Despite such evidence, paleo ice-sheet ~~modeling experiments typically assume that precipitation consistently follows the thermodynamic , Clausius-Clapeyron relationship (e.g., Huybrechts et al., 1991; Greve et al., 2011; Pollard and DeConto, 2012)~~models typically assume precipitation fields that are parameterized in time using a thermodynamic relationship that is constant for all locations and timescales (e.g., Huybrechts, 2002; Greve et al., 2011).

Ice-core records provide the best empirical estimates of climate history over the Greenland Ice Sheet. For temperature, important proxies include oxygen and hydrogen isotopes of ice (e.g., Jouzel et al., 1997), nitrogen isotope ratios of gas trapped in ice ~~(e.g., Buizert et al., 2014; Kobashi et al., 2017)~~(e.g., Severinghaus et al., 1998; Severinghaus and Brook, 1999), and borehole thermometry (e.g., Cuffey et al., 1995; Dahl-Jensen et al., 1998). For precipitation, the thickness of annual layers of accumulated ice, corrected for thinning, is used (e.g., Dahl-Jensen et al., 1993). Ice-core records, however, are limited in their spatial coverage. In contrast, climate-model simulations are spatially-complete estimates of past climate, but they are subject to uncertainty ~~in model dynamicsand boundary conditions~~due to model dynamics, boundary conditions, and spatial resolution.

Efforts to combine information from proxy data and climate models have long been a part of ice-sheet modeling. The most common approach is to scale the modern spatial pattern of temperature and precipitation using data from a single ice core (e.g, Huybrechts et al., 1991; Huybrechts, 2002; Greve, 1997; Greve et al., 2011; Nielsen et al., 2018). This assumes a fixed spatial pattern through time, which is unlikely to be valid. Recently, Buizert et al. (2018) used the average of the three best-understood Greenland ice-core records to adjust the results of a transient climate-model simulation ~~(the transient climate evolution experiment, TraCE21ka; Liu et al., 2009; He et al., 2013)~~(the transient climate evolution experiment, TraC This approach allows for possible changes in spatial relationships, but ~~is restricted to a single climate-model realization, makes relatively little use of the available data from ice cores outside northern Greenland ,~~ focuses on ice cores in central and northern Greenland and provides no information on precipitation. Other attempts to incorporate more proxy data have been limited to short time periods (e.g., Simpson et al., 2009; Lecavalier et al., 2014).

In this study we apply ~~a novel method,~~ paleoclimate data assimilation ~~(paleo DA),~~ to obtain a new, spatially-complete reconstruction of Greenland temperature and precipitation. We focus on the last 20,000 years, which includes the end of the last glacial period, the glacial to interglacial transition, and the Holocene thermal maximum (HTM), when temperatures at the Greenland Ice Sheet summit reached 1-2 °C warmer than present ~~(Cuffey and Clow, 1997; Dahl-Jensen et al. , 1998).~~ (Cuffey and Clow, 1997; Dahl-Jensen et al., 1998). The climate history, and the corresponding changes in the size of the ice sheet, are well-documented over this time period ~~(Kaufman et al. , 2004; Young and Briner, 2015).~~ (e.g., Kaufman et al., 2004; Young and B

~~Paleo DA~~ Paleoclimate data assimilation combines spatial information from a climate-model simulation and temporal data from proxy records to produce a climate "reanalysis", where the term is taken from the modern climate reanalysis methods on which ~~paleo DA~~ the data assimilation method is based (e.g., Kalnay et al., 1996). We adapt the ~~paleo DA~~ paleoclimate data assimilation framework developed by Hakim et al. (2016), who reconstructed annual two-meter air temperature and 500 hPa geopotential height over the last millennium using a global network of temperature and precipitation-sensitive proxy

records. Here, we use oxygen isotopes of ice and layer thickness from ice cores to reconstruct temperature and precipitation, respectively. We choose these proxies for their high temporal resolution, direct relationships to climate over the ice sheet, and availability from multiple ice cores. For the climate-model simulation, we use ~~TraCE21k~~ TraCE-21ka (Liu et al., 2009; He et al., 2013), which was run with the Community Climate System Model version 3 (CCSM3) to simulate the last ~~21~~22,000 years. We compare the resulting reanalysis to previously-published climate reconstructions (Sects. ~~3 and 5~~3.1, 4.2, and 4.3), and assess the precipitation-temperature relationship (Sect. ~~3~~4.1). We evaluate the reanalysis with independent proxy records and sensitivity tests (~~Sect. 4~~Sects. 3.2 and 3.3).

2 Methods

Our paleoclimate reconstruction method assimilates oxygen isotope ratios and accumulation from ice cores with a transient climate-model simulation to reconstruct the last 20,000 years of Greenland temperature and precipitation. In the following subsections we discuss the ice-core data, the climate-model simulation, and the details of our paleoclimate data assimilation approach.

2.1 Ice-core data

We use proxy records from eight ice cores from the Greenland Ice Sheet and nearby ice caps (Fig. 1, Table 1). As a proxy for temperature, we use previously-published measurements of oxygen isotope ratios from the ice, which we discuss using the conventional δ^{18}O nomenclature (Dansgaard, 1964). We note that while other proxies (such as borehole thermometry or δ^{15}N of N_2) have been used to produce temperature estimates ~~(e.g., Cuffey et al., 1995; Dahl-Jensen et al., 1998; Buizert et al., 2014; Kobashi et al. this has not been done at all our~~ (e.g., Cuffey et al., 1995; Dahl-Jensen et al., 1998; Severinghaus et al., 1998; Severinghaus and Brook, 199 they are not available for many of the core locations; we instead rely on such data to obtain independent estimates of error in the δ^{18}O-temperature relationship (see Sect. 2.3.3) and as comparisons to our resulting reanalysis (see Sect. 3). The accumulation history has been estimated for five of these cores from ~~measured~~ layer thickness corrected for vertical ice-thinning due to dynamical strain in the ice sheet. We rely on previously-published accumulation histories for the GISP2 and NEEM cores (Cuffey and Clow, 1997; Rasmussen et al., 2013), and we estimate accumulation for the Dye3, GRIP, and NGRIP cores using available layer-thickness data and simple ice-thinning calculations (see below and Sect. ~~S1~~S2). We do not use accumulation records from the Agassiz, Camp Century, or Renland cores because the ice-thinning history at these sites is not adequately constrained. Most of the ice-core data are available at 50-year or higher resolution and have been synchronized to a common depth-age scale (the Greenland ice core chronology 2005, GICC05; Andersen et al., 2006; Rasmussen et al., 2006; Svensson et al., 2006; Vinther et al., 2006, 2008). All of these ice-core records extend from the beginning of the Holocene to the present. Five δ^{18}O and four accumulation records also include the last glacial period. To evaluate the impact of the differing lengths of these records, we produce a sensitivity reanalysis for which we assimilate just the fixed proxy-network (i.e., only those data that span the full reanalysis time period, the last 20,000 years).

To extract the accumulation signal from measured layer-thickness, the layers must be destrained using assumptions about the history of ice flow. For the Dye3, GRIP, and NGRIP cores, we use a one-dimensional ice-flow model (Dansgaard and Johnsen, 1969) to calculate the cumulative vertical strain the layers have experienced at each core site. The Dansgaard-Johnsen model requires specifying the vertical velocity at the surface and a kink height which determines the shape of the vertical velocity profile. The velocity profile below the kink height approximates the influence of greater deformation rates in deeper ice due to increased deviatoric stress and warmer ice temperature. For sites at the pressure-melting point at the bed, such as NGRIP, we also implement the basal melt-rate (e.g., Dahl-Jensen et al., 2003). Previous work on many of the Greenland ice cores has estimated cumulative vertical thinning assuming that the accumulation history scales with δ^{18}O and then found optimal parameter values by comparing the modeled and measured depth-age relationships (Dahl-Jensen et al., 1993, 2003; Rasmussen et al., 2014). Here, we wish to maintain independent determinations of the δ^{18}O and accumulation proxy records. To do this, we select reasonable ice-flow parameters independently, based on the glaciological setting of each site; specifically, we use kink-height values of 0.1-0.2 for flank flow and 0.4 for ice flow near ice divides where the deviatoric stress is low (Raymond, 1983; Conway et al., 1999). Where available, we use published values or kink-height values that result in a good match to published accumulation records (Dahl-Jensen et al., 2003; Gkinis et al., 2014). Based on this range of plausible ice-flow parameters, we develop three scenarios for each site: "low", "moderate", and "high", where the names reflect the relative magnitude of accumulation in the glacial and early Holocene. We assimilate the intermediate-value ("moderate") accumulation records to produce our main precipitation reanalysis, while we assimilate the high and low accumulation records into high and low sensitivity scenarios, respectively, to provide a conservative estimate of uncertainty. Descriptions of the rationale for the parameter choices at each site are given in Sect. ~~S1~~S2. Our method to estimate accumulation should be most accurate for the interior ice cores (i.e., GISP2, GRIP, NEEM, NGRIP); these sites are thicker and have lower accumulation rates such that layers of the same age have experienced less cumulative strain than for the more coastal cores (i.e., Agassiz, Camp Century, and Renland). We do not attempt to reconstruct accumulation from these coastal cores because the layers cannot be destrained with sufficient accuracy. Dye3 suffers from the same challenges; however, it is the only ice core with long-term climate data south of 70°N (Fig. 1). Thus, we include the Holocene Dye3 accumulation rates despite the greater uncertainty relative to the interior cores.

Because records from the Dye3 ice core are our only source of information in southern Greenland, we take the following steps to increase the data available for assimilation. The Dye3 record has not been previously assigned a depth-age scale beyond 11.7 ka (throughout this paper, “ka” refers to thousands of years before 1950 CE). We extend the depth-age scale to 20 ka to take advantage of the glacial portion of the δ^{18}O record. To do this, we match the δ^{18}O record from Dye3 to the δ^{18}O record from NGRIP using the cross-correlation maximization procedure from Huybers and Wunsch (2004) (Sect. ~~S2~~S3). We interpolate the glacial δ^{18}O record from Dye3 (which, as measured, has an average resolution of only 85 years) to the same 50-year resolution used for our other ice-core records. Extension of the Dye3 depth-age scale also provides a layer-thickness record from 20 ka to present; however, we do not use accumulation data from Dye3 for the period 20-11.7 ka because the low resolution impedes our ability to estimate accumulation variations from layer thickness. Using this depth-age scale extension

for Dye3 may introduce error that is difficult to quantify; however, we find that including Dye3 has an important impact on the resulting reanalysis of southern Greenland climate (Sect. ~~S3~~S4).

Where possible, we account for non-local effects on the ice-core records. The global-mean δ^{18}O of seawater fluctuates with global ice-volume, while on the regional scale, horizontal advection brings ice from other elevations and latitudes. We correct for changes in the oxygen-isotope composition of seawater following the methods of Stenni et al. (2010), using the benthic foraminifera dataset from Bintanja et al. (2005). For ice cores in regions of high horizontal advection, we correct for elevation and latitude differences between the site of snow deposition and the ice-core site. Following the methods from Dahl-Jensen et al. (2013), we apply corrections for advection-caused elevation changes in the δ^{18}O records from Camp Century, Dye3, and NEEM and for advection-caused latitude changes in the δ^{18}O record from NEEM (Vinther et al., 2009; Dahl-Jensen et al., 2013). We do not correct the accumulation records for advection from upstream because the elevation-accumulation relationship is complicated by the prevailing wind direction (Roe and Lindzen, 2001) and the thinning function uncertainties are likely to be larger than the effects of ice advection. We also do not correct δ^{18}O or accumulation for changing elevation at the ice-core site; our goal is to reconstruct conditions at the surface, rather than at a constant reference elevation. We take the anomaly of each corrected δ^{18}O record and the ratio of each accumulation record relative to the mean of all data in the record that falls within the time period 1850-2000 CE. We then average each record to 50-year resolution, the lowest resolution in these records (with the exception of δ^{18}O from the glacial period in the Dye3 core). It is these corrected, averaged records that we use in the data assimilation (Figs. 2 and 3).

2.2 Climate-model simulation

We use ~~TraCE21ka~~TraCE-21ka, a simulation of the last ~~21~~22,000 years of climate (22 ka to -.04 ka), which was run ~~with~~ using the fully-coupled CCSM3 ~~and~~ at T31 resolution (approximately 3.75 degrees horizontally) with transient ice-sheet, orbital, greenhouse gas, and meltwater flux forcings (Liu et al., 2009, 2012; He et al., 2013). For paleoclimate data assimilation, it is important that the climate simulation capture a range of possible climate states over the time period of interest. By design, ~~TraCE21ka~~ TraCE-21ka captures the major glacial-to-Holocene temperature changes, as well as some of the short-term, rapid climate changes~~(i.e., the transitions into and out of the~~ , such as the Bølling-Allerød ~~and Younger Dryas events). In contrast, many other transient climate simulations cover only~~ transition (Liu et al., 2009). Many higher-resolution climate simulations are transient only over the last millennium ~~and~~ (e.g., Bothe et al., 2015) or provide a snapshot of a certain time, such as last glacial maximum or the mid Holocene (e.g., Harrison et al., 2014). Individually, these millennial-length simulations have too little variability to capture the range of climate states across the glacial-interglacial transition. ~~Furthermore, TraCE21ka~~ If combined, the biases in each simulation would need to be individually addressed, which is beyond the scope of this study.

From TraCE-21ka, we use two-meter air temperature for temperature (T) and the sum of large-scale stable precipitation and convective precipitation for precipitation (P). To correct for model bias in TraCE-21ka, we assume that the bias is stationary in time and apply the delta-change method (Teutschbein and Seibert, 2012) by taking the anomaly of temperature and the fraction of precipitation relative to the mean of our reference period (1850-2000 CE). An assumption of a stationary model bias is required because, with a small number of proxy records, we cannot afford to subsample them for the purposes of bias

correction, data assimilation, and evaluation. After the bias correction, we average the TraCE-21ka variables (which originially have monthly resolution) to 50-year resolution, as we did for the ice-core records. In this process, we average 50 consecutive years (600 months) such that no year (or month) is used in more than one 50-year average. This averaging results in 440 time steps spaced 50 years apart.

TraCE-21ka includes changes in orbital forcing~~and therefore~~ , which contribute to changes in the seasonality of temperature and precipitation~~, which strongly influence the~~ . The strength of these seasonal cycles influences the mean-annual relationship between δ^{18}O and temperature ~~(Steig et al., 1994; Werner et al., 2000). TraCE21ka also includes~~ (Steig et al., 1994; Werner et al., 2000; Kr TraCE-21ka consistently shows stronger temperature and precipitation seasonality in the glacial (20 to 15 ka) than in the Holocene (5 to 0 ka) at each of the eight ice-core sites considered in this study (Fig. S3). Relative to the annual mean, the glacial had warmer and wetter summers and colder and drier winters. The Holocene also shows such a seasonal cycle; however, there is a smaller difference between the summers and winters. Any seasonal signal with wetter summers than winters will bias the δ^{18}O towards summer values. According to TraCE-21ka, this effect is amplified in the glacial. As we discuss in Sect. 2.3.1, the particularly strong summer bias in the glacial affects the mean-annual δ^{18}O-temperature relationship in ways that are consistent with findings from borehole thermometry at the GISP2, GRIP, and Dye3 ice-core sites (Cuffey and Clow, 1997; Jouzel et al., 1997).

In addition to a change in the strength of the seasonal cycle, TraCE-21ka shows a temporal shift, with summer temperature and precipitation peaking earlier in the glacial (around June and July) than in the Holocene (from July to September) (Fig. S3). In the glacial, both variables peak around June and July, with only two exceptions: precipitation peaks in August at the Renland and Dye3 ice-core sites. In contrast, Holocene temperature peaks slightly later, in July, while precipitation peaks even later, in August and occasionally September. Both variables and both time periods show winter minimums in February, again with the two exceptions of Renland and Dye3, which show later precipitation minimums. In this study, the timing of the seasonal peaks is relevant because it affects the precipitation-weighted temperature, defined in Eq. 7.

TraCE-21ka also includes prescribed transient ice sheets ~~, which~~ as a boundary condition. The transience is important for capturing the influence of elevation change on the ice-core records~~. We note, however, that the ice sheets in TraCE21ka~~ ; however, the low horizontal resolution of TraCE-21ka leads to difficulties in capturing elevation changes at the edges of the ice sheet, in southern Greenland, and at coastal ice caps. In addition, the ice sheets in TraCE-21ka are independent of climate, updated only every 500 years during the simulation, and taken from ICE-5G (Peltier, 2004), a now outdated ice-sheet reconstruction (Roy and Peltier, 2018). ~~From TraCE21ka, we use two-meter air temperature for temperature (T) and the sum of large-scale stable precipitation and convective precipitation for precipitation (P). We take the anomaly of temperature and the fraction of precipitation relative to the mean for 1850-2000 CE. We then average to 50-year resolution, as for the ice-core records.~~

2.3 Paleoclimate data assimilation

To combine the ice-core ~~data~~ and climate-model data, we use an offline data assimilation method similar to that described in Hakim et al. (2016). If no covariance localization is used, as in this study, this method can be summed up as a linear combination

of randomly-selected model states that are weighted according to new information provided by the proxy records. "Offline" refers to the absence of a forecast model that evolves the climate state between ~~time steps, such that in offline data assimilation the same initial climate state is used for every time step~~assimilation time steps. The offline method is appropriate when ~~characteristic memory in the system is significantly shorter than the time step (Hakim et al., 2016, and references therein). In our case, each time step~~ model predictive-skill is small given the assimilation time step (Hakim et al., 2016, and references therein). Model predictive-skill is generally poor on decadal to longer timescales (Latif and Keenlyside, 2011, and references therein) except possibly during times of strong forcing, such as the Bølling-Allerød (14.7-12.7 ka) and the Younger Dryas (12.7-11.7 ka) (Hawkins and Sutton, 2009). Because each of our time steps is an average over 50 years, as dictated by the resolution to which we average the proxy records~~.~~, the offline method is appropriate except possibly during these large-forcing events. For these events, an online method may be appropriate (assuming that the models correctly capture both the forcing and the response); however, online ensemble data assimilation over glacial-interglacial cycles using a fully-coupled earth system model is impractical due to the computational cost and is beyond the scope of this study.

Our paleoclimate data assimilation framework uses ensembles for the initial (prior) and final (posterior) estimates of the climate state, providing a probabilistic framework for interpreting and evaluating the results. To compute the posterior ensemble, we apply the Kalman update equation (Whitaker and Hamill, 2002), which spreads new information gained from proxy records to all locations and variables in the prior ensemble:

$$\boldsymbol{x_a} = \boldsymbol{x_b} + \mathbf{K}(\boldsymbol{y} - \mathcal{H}(\boldsymbol{x_b})) \tag{1}$$

~~where bold~~ Bold lowercase letters are vectors, bold capital letters are matrices, and script capital letters are mapping functions. The posterior ensemble is $\boldsymbol{x_a}$, $\boldsymbol{x_b}$ is the prior ensemble, $\boldsymbol{y}$ is the proxy data, and $\mathcal{H}$ is the function that maps from the prior variables to the proxy variables. $\mathbf{K}$ is the Kalman gain matrix:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{2}$$

where ^{T} indicates a matrix transpose, $\mathbf{B}$ is the covariance matrix computed from the prior ensemble, $\mathbf{H}$ is the linearization of $\mathcal{H}$ about the mean value of the prior, and $\mathbf{R}$ is a diagonal matrix containing the error variance for each proxy record, the use of which requires an assumption that error covariances between proxy records are negligible.

To compute the new information gained from the proxy records, the prior ensemble must first be mapped into proxy space to get prior estimates of the proxy ($\mathcal{H}(\boldsymbol{x_b})$). This mapping ($\mathcal{H}$) is the proxy system model (PSM). Our PSM for the δ^{18}O-temperature relationship is a linear function and for accumulation is a direct comparison between ice-core-derived accumulation and precipitation from the prior~~, which is~~. For these PSMs, both temperature and precipitation are interpolated from the climate-model grid to the geographic location of the ice core. These PSMs are detailed in Sects. 2.3.1 and 2.3.2. Comparing the prior estimates of the proxy ($\mathcal{H}(\boldsymbol{x_b})$) to the proxy data ($\boldsymbol{y}$) yields the innovation ($\boldsymbol{y} - \mathcal{H}(\boldsymbol{x_b})$), the new information gained from the proxy records.

The Kalman gain ($\mathbf{K}$, Eq. 2) weights the innovation by the relative magnitude of the ensemble covariance of the prior estimates of the proxy ($\mathbf{H}\mathbf{B}\mathbf{H}^T$), and the error covariance of the proxy records ($\mathbf{R}$). The Kalman gain spreads the weighted innovation to all locations and variables in the prior ensemble, using the covariance structure ($\mathbf{B}\mathbf{H}^T$) from the prior ensemble.

The prior ensemble is an initial estimate of possible climate states, which we form using 100 randomly-chosen 50-year averages from the ~~TraCE21ka simulation.~~ TraCE-21ka simulation. States from both the glacial and the Holocene make up a prior ensemble. The same prior is used for all time steps in the reconstruction, leading to a prior that is constant in time. For a longer discussion on the reasoning behind our choice to use a constant prior, please refer to the Supplementary Information, Sect. S1. Proxy records are assimilated into the prior using ~~the Kalman update (~~Eq. 1~~)~~, which produces the posterior ensemble, a new estimate of possible climate states. We assimilate δ^{18}O to reconstruct temperature and separately assimilate accumulation to reconstruct precipitation. This ~~process is repeated~~ approach maintains independence between temperature and precipitation, which avoids imposing linearity and stationarity on the relationship between these two variables. As Cuffey and Clow (1997) show, not only is this relationship non-linear on long timescales but it is also not well-approximated by simple thermodynamic expectations. Separating these variables ensures that the relationship between temperature and precipitation is consistent with the empirical relationship between δ^{18}O and accumulation from ice cores, rather than being derived exclusively from the climate model.

We repeat the data assimilation process over multiple iterations, with each iteration using one of ten different 100-member prior ensembles and excluding one proxy record. Each of the ten prior ensembles is made up of a different random selection of 50-year averages from TraCE-21ka. Thus, each prior ensemble has a different variance and spatial covariance structure. Each proxy record is excluded from a total of ten iterations, where each of these iterations uses a different one of the ten prior options. Every iteration is uniquely identifiable by which prior ensemble is used and which proxy record is excluded. For a reanalysis, the total number of iterations is thus ten times the number of proxy records, such that for temperature we have 80 iterations and for precipitation we have 50 iterations. A reanalysis is a compilation of the 100-member posterior ensembles from these iterations, resulting in a temperature reanalysis having 8,000 ensemble members and a precipitation reanalysis having 5,000 ensemble members.

The proxy record that is excluded from an iteration is independent of that iteration's posterior ensemble, such that we can evaluate the posterior against this record. With our PSMs, we convert the posterior ensemble into predictions of the independent record using the mapping $\mathcal{H}$ and compare these predictions to the record along the time axis. We use four skill metrics to evaluate different aspects of the predictions. The correlation coefficient (corr; Eq. 3) measures the relative timing of signals in the predictions and the proxy record:

$$corr = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{(y_i-\bar{y})(v_i-\bar{v})}{\sigma_y\sigma_v}\right) \tag{3}$$

where v is the ensemble mean of the predicted values, y is the proxy record value, an overbar indicates a time mean, n is the number of time steps, and σ is the standard deviation of the variable in the subscript. The coefficient of efficiency (CE; Eq. 4) (Nash and Sutcliffe, 1970) is affected by signal timing as well as signal amplitude and mean bias:

$$CE = 1 - \frac{\Sigma_{i=1}^{n}(v_i-y_i)^2}{\Sigma_{i=1}^{n}(v_i-\bar{v})^2} \tag{4}$$

The root mean squared error (RMSE; Eq. 5) gives an intuitive sense for the magnitude of the differences between the predictions and proxy record:

$$RMSE = \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - v_i)^2 \right)^{1/2} \tag{5}$$

The ensemble calibration ratio (ECR; Eq. 6) indicates whether the ensemble has enough spread (uncertainty) given the error in the ensemble mean (e.g. Houtekamer et al., 2005):

$$ECR = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{(y_i - v_i)^2}{var(\boldsymbol{v_i}) + r} \right) \tag{6}$$

where $\boldsymbol{v}$ is a vector of the ensemble members of the predicted values, r is the error variance for the proxy record (y), and var indicates the variance. Accordingly, if the ensemble variance is appropriate for the amount of error, then the ensemble calibration ratio is near unity.

We compute all four skill metrics for both the posterior and prior ensembles, which shows whether assimilation of proxy records results in an improved estimate of the climate state over our initial estimate. We define improvement as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ensemble calibration ratio closer to 1. We anticipate that our reanalysis will show improvement over the prior because the prior is constant in time and contains no information about temporal climate variations; however, improvement is not gauranteed, especially if proxy records contain highly-localized climate signals or if the prior covariance structure is unable to appropriately spread information from the proxy sites to other locations. For further comparison, we additionally compute the correlation coefficient, CE, and RMSE between the TraCE-21ka simulation and the proxy records.

We evaluate results over three time periods: 1) the full overlap between the reanalysis time period and the proxy record, 2) a time representative of the glacial, 20-15 ka, and 3) a time representative of the Holocene, 8-3 ka. Some proxy records overlap the full reanalysis period, 20-0 ka, while others overlap just the Holocene, 11.7-0 ka. The skill metrics computed for these two groups should be considered separately.

To remove mean bias from temperature, we subtract out the reference-period mean. For precipitation, we divide by the reference-period mean. It is these bias-corrected results that are referred to unless noted otherwise.

2.3.1 Proxy system model: δ^{18}O

The isotopic composition of precipitation, as recorded in ice cores, is highly correlated with temperature at the time and location of deposition, but is also sensitive to conditions at the moisture source region (i.e. sea surface temperature and relative humidity at the ocean surface; e.g., Johnsen et al., 1989). Moisture-source conditions primarily affect the deuterium excess, which we do not use here, and the influence on δ^{18}O is comparatively weak (e.g., Armengaud et al., 1998). For our δ^{18}O PSM, we use a linear relationship with temperature at the ice-core drill site (T_{site}) that has a slope of 0.67 $\pm$ 0.02 ‰ °C^{-1}, which was calibrated using modern, spatial data (Johnsen et al., 1989). It is well known that this modern, spatially-derived slope does not necessarily apply to temporal δ^{18}O-T_{site} relationships, which have effective slopes that are time, frequency, and location

dependent. Temporal changes in precipitation seasonality, inversion-layer thickness, and source region conditions introduce nonlinearity into the effective δ^{18}O-T_{site} relationship (e.g., Jouzel et al., 1997; Pausata and Löfverström, 2015). Diffusion in the firn column also affects this relationship, but it is negligible for annual and longer timescales at the locations of the ice cores we use (Cuffey and Steig, 1998). Borehole thermometry at the GISP2 and GRIP sites show that for the low-frequency changes associated with the last glacial-interglacial transition, the temporal slope is less than half the modern, spatial slope (Cuffey and Clow, 1997; Jouzel et al., 1997). Numerous studies have suggested that precipitation seasonality is the largest source of nonlinearity in the δ^{18}O-T_{site} relationship (e.g., Steig et al., 1994; Pausata and Löfverström, 2015); changes in precipitation seasonality are thought to be the primary reason that the effective δ^{18}O-T_{site} relationship for the glacial-interglacial transition has such a low slope ~~(Werner et al., 2000)~~(Werner et al., 2000; Gierz et al., 2017; Cauquoin et al., 2019).

We rely on ~~TraCE21ka~~ TraCE-21ka to estimate the site-specific effects of precipitation seasonality on the δ^{18}O-T_{site} relationship. Site-specific effects can also be estimated using independent temperature reconstructions, e.g. from borehole thermometry or δ^{15}N of N_2 measurements; however, such independent reconstructions for the last 20,000 years exist only for a few of the long ice-core records, GISP2, GRIP, and NGRIP (Buizert et al., 2018; Dahl-Jensen et al., 1998; Gkinis et al., 2014), limiting the utility of such records to capture ~~spatial variability~~the spatial variability of the δ^{18}O-T_{site} relationship across Greenland .

To incorporate estimates of the site-specific effects of precipitation seasonality from ~~TraCE21ka~~TraCE-21ka, we adjust the linear δ^{18}O PSM by replacing T_{site} with T^*_{site}, the precipitation-weighted temperature at the model grid-cell closest to the ice-core drill site. We compute T^* across Greenland using T and P from ~~TraCE21ka~~TraCE-21ka (see Fig. S4 for a visual comparison of T and T^* at the GISP2 ice-core site):

$$T^* = \sum_{i=1}^{n=12} \left(T_{mon} \frac{P_{mon}}{P_{ann}} \right) \tag{7}$$

With ~~T^*_{site}~~ T^*_{site} in our PSM, we find that the δ^{18}~~O-T^*_{site}~~ O-T_{site} slope is spatially variable (e.g., Fig. S5), ranging between 0.42 and 0.66 ‰°C^{-1} at the ice-core sites (Table S1), and tending to be less than the modern spatial relationship of 0.67 ‰°C^{-1} at most locations around Greenland. Due to the data assimilation method outlined above, these slopes vary both in space and across iterations, the latter being due to the varying prior ensembles. These slopes do not vary in time in the prior, but they do in the posterior (note that δ^{18}O-T_{site} slopes mentioned throughout this paper refer to the prior ensemble). By using ten different prior ensembles, ~~but on average it is about 75% of modern . Thus, in using TraCE21ka,~~ we capture the ~~mean of the modern (i. e. high-frequency)relationship and that of the glacial-interglacial (i. e. low-frequency), effective temporal-slope~~uncertainty in the δ^{18}O-temperature relationship from variations in the precipitation seasonality. These TraCE-21ka-derived estimates lie within the range of slopes estimated for sites around Greenland for a variety of time periods (Table S1). Differences seen in Table S1 reflect both the different methods used and the time period considered. Some estimates, such as Guillevic et al. (2013) and Buizert et al. (2014) are for abrupt transitions, such as Dansgaard-Oeschger events, while others find mean slopes over longer periods of time, such as Kindler et al. (2014) and this investigation.

To evaluate the sensitivity of our results to the choice of PSM, we produce four other reconstructions (S1-S4). The S1 scenario uses the PSM, $\delta^{18}\text{O} = 0.67 T_{site}$, which is the modern (high-frequency) relationship and does not account for precip-

itation seasonality. The S2 scenario uses $\delta^{18}\text{O} = 0.5 T_{site}$, the mean of the ~~modern and glacial-interglacial~~ high-frequency and low-frequency temporal slopes. The S3 scenario uses $\delta^{18}\text{O} = 0.335 T_{site}$, which is similar to published estimates of the glacial-interglacial temporal slope (half the ~~modern~~ high-frequency slope) (Cuffey and Clow, 1997; Jouzel et al., 1997). By lowering the slope in the S2 and S3 scenarios, we implicitly account for precipitation seasonality; however, in these scenarios and S1, the $\delta^{18}\text{O}$-T_{site} relationship is spatially uniform, whereas it is spatially variable in the main reanalysis because we include the spatial pattern of precipitation seasonality. The S4 scenario uses the same PSM as in the main reanalysis, $\delta^{18}\text{O} = 0.67 T^{*}_{site}$, but we adjust the strength of the precipitation seasonality in ~~TraCE21ka~~ TraCE-21ka such that the average $\delta^{18}\text{O}$-T_{site} slope around Greenland is approximately 0.335 ‰ $^{\circ}\text{C}^{-1}$. The S4 scenario thus has the same spatially-variable $\delta^{18}\text{O}$-T_{site} relationship as in the main reanalysis, but ~~with~~ a greater influence of precipitation seasonality.

These sensitivity tests are equivalent to testing different assumptions about the $\delta^{18}\text{O}$-T_{site} relationship. The availability of a 20,000 year-long isotope-enabled climate simulation would allow us to determine this relationship from model physics, which incorporate a variety of processes that can affect water isotopes, including precipitation seasonality.

2.3.2 Proxy system model: Accumulation

Accumulation is closely related to total precipitation at our ice-core sites, which have limited surface melting and evaporation. Simulations from the regional climate model HIRHAM5 show that for modern climate at the GISP2, GRIP, NEEM, and NGRIP ice-core sites, surface mass balance, snowfall, and precipitation are all within 1.6 cm water equivalent (w.e.) when averaged over the years 1989 to 2012 CE (Langen et al., 2015, 2017). For this reason, and because ~~TraCE21ka~~ TraCE-21ka lacks process-based ablation variables, our PSM is a direct-comparison between ice-core accumulation and simulated precipitation at the model grid-cell closest to the ice-core site.

This direct-comparison PSM may be an incomplete model of the accumulation-precipitation relationship at the Dye3 ice-core site; regional climate simulations show that modern surface mass balance is lower than precipitation due to melt rates that average 84 cm w.e. year^{-1}. Significant melt rates would cause the spatial covariance structure of accumulation across these sites to differ from that of precipitation; however, both models and observations lack the necessary variables or duration to show the extent to which this difference exists for 50-year timescales through the last glacial-interglacial cycle. We emphasize that we use relative, rather than ~~aboslute~~ absolute changes in the data assimilation, to account for the mean bias between precipitation and accumulation.

2.3.3 Proxy error variance estimation

In the Kalman filter (Eq. 1), the diagonal elements of $\mathbf{R}$ contain the error variance of each proxy record, which includes how we model the proxy (the PSM). We compute representative error variances for $\delta^{18}\text{O}$ and accumulation, and apply them to all records and time slices. We do not include error associated with corrections applied to the ice-core data (Sect. 2.1) or associated with the accumulation PSM (Sect. 2.3.2) because we cannot characterize the statistical properties of these errors.

A universal, but typically small, error source is from the measurement of proxies. For $\delta^{18}\text{O}$, measurement error is equivalent to laboratory precision. We compute a representative measurement error from the GISP2 ice core, for which a single mea-

surement of δ^{18}O has a laboratory precision (variance) of 0.024 $‰^2$ (Stuiver and Grootes, 2000). Assuming independent error and annual measurements, the 50-year average error variance reduces to 0.0034 $‰^2$, which is insignificant compared to other sources of error. For accumulation, the measurement error is from the measurement of layer thickness, which is related to the error in annual-layer counts per unit depth. Again, we assume GISP2 is a representative core and estimate the layer-thickness error from Table 3 in Alley et al. (1997), which provides repeat annual-layer counts for several depth intervals. From this table, we find the standard deviation of counted years in each depth interval, divide by the average number of years, average across all depth intervals, and square the result. This computation results in a layer-thickness error variance of 0.0015, a unitless number due to our use of fractional accumulation records.

Another source of error is the extent to which a model grid-cell may misrepresent a point proxy-measurement. In the innovation, there is an implicit assumption that the proxy ($\boldsymbol{y}$) and the prior estimate of the proxy ($\mathcal{H}(\boldsymbol{x_b})$) are representative of the same processes. However, an ice core is about 100 cm^2, an area that is affected by processes at all scales, from regional change to local, sub-meter-scale topography, while a model grid-cell in ~~TraCE21ka~~ TraCE-21ka can cover tens of thousands of square kilometers, an area that is affected by only the largest scales, from global to regional. Thus, there is an inherent inability of the prior to represent local processes at the ice-core site, which we refer to as the spatial representation error. To estimate this error, we compute the variance of the local noise (e.g., Reeh and Fisher, 1983) using the GISP2 and GRIP ice-core records, which are located about 30 km apart within the same model grid cell. For δ^{18}O, our estimate is 0.21 $‰^2$ which is about half the value determined by Fisher et al. (1985) at several locations around Greenland. Our estimate is relatively conservative, considering that we are using 50-year averages rather than annual averages as in Fisher et al. (1985). For accumulation, we estimate a spatial representation error variance of 0.0023 using the same method as for δ^{18}O.

A third source of error is the extent to which the δ^{18}O PSM may be an inaccurate model of the δ^{18}O-T_{site} relationship. The less accurate the PSM, the less weight that should be given to the innovation. We estimate PSM error variance by calculating the mean squared error (MSE) between a δ^{18}O record and an independent temperature record mapped to δ^{18}O using the PSM, $\delta^{18}\mathrm{O} = 0.67 T_{site}$. We use independent datasets taken from the GISP2 ice core: the δ^{18}O record and three δ^{15}N-derived temperature estimates for the Holocene, a mean estimate and the two-standard-deviation uncertainty bounds (Kobashi et al., 2017). From these datasets, we estimate three PSM error variances that range from 0.56 to 1.1 $‰^2$, from which we choose the largest.

For the assimilation of each δ^{18}O record, we use an estimated total error variance of 1.3 $‰^2$, which is a sum of the measurement, spatial representation, and PSM error variances. For the assimilation of each accumulation record, we use an estimated total error variance of 0.0038, which is a sum of the measurement and spatial representation error variances.

3 Results

3.1 Overview

Through the assimilation of ice-core data with a prior ensemble that is constant in time, we produce a spatially-complete, mean-annual Greenland temperature and precipitation reanalysis at 50-year resolution (Figs. 4 and 5). Here we focus on results

~~relevant to the evolution and sensitivity of the Greenland Ice Sheet, including the late glacial anomaly,~~ for the late glacial and the Holocene thermal maximum (HTM)~~, and the relationship between temperature and precipitation~~.

In our reanalysis, late glacial (20-15 ka) mean-temperature anomalies range from about -20 °C in northern Greenland to less than -10 °C in southern Greenland (Fig. 4c). At the GRIP and GISP2 ice-core sites, the reanalysis has a -14 °C anomaly with a standard deviation of 2 °C. This is in excellent agreement with the mean-temperature anomaly of -14 °C for the same period at the GISP2 site, which was derived from δ^{18}O calibrated with borehole thermometry (Cuffey et al., 1995; Cuffey and Clow, 1997). Also in agreement with borehole thermometry, this period is warmer than the last glacial maximum (Dahl-Jensen et al., 1993; Cuffey and Clow, 1997). Average late-glacial precipitation in the reanalysis ranges from a third to half of modern with the highest values on the coasts around southern Greenland (Fig. 4d).

Our reanalysis shows warmest temperatures occurred across Greenland between 7 and 3 ka, reaching a temperature maximum around 5 ka (Fig. 5). Although this timing tends to be later than many estimates of the HTM, it lies within the ranges reported in the literature; for example, a summary of proxy records from around Greenland shows peak warmth usually occurring around 9-5 ka, but also as early as 10.8 ka and as late as 3 ka (Kaufman et al., 2004). Borehole thermometry shows that temperatures peaked around 6-7.7 ka at Summit and 4.5 ka at the Dye3 ice-core site (Cuffey and Clow, 1997; Dahl-Jensen et al., 1998; Kaufman et al., 2004). Temperature estimates from δ^{15}N of N_2 show an earlier HTM peak at Summit around 8 ka (Kobashi et al., 2017). In northwest Greenland, δ^{18}O measurements from lake sediments show the HTM starting before 7.7 ka and ending around 6 ka (Lasher et al., 2017), while chironomid assemblages show peak warmth around 10-8 ka (McFarlin et al., 2018).

Mean-annual HTM temperature anomalies in our reanalysis range from nearly +2 °C in northern Greenland to about +1 °C in southern Greenland (Fig. 4a). Similar to the late-glacial temperature anomalies, the pattern of the HTM is dominated by a north-south trend that has the greater temperature changes to the north, especially in northwest Greenland. While this spatial pattern agrees well with previous studies which have noted especially warm Holocene temperature anomalies in northwest Greenland (Lasher et al., 2017; Lecavalier et al., 2017; McFarlin et al., 2018), many estimates of HTM anomalies around Greenland are higher than our reanalysis indicates. Our low temperature estimates (compared to previous work) may be in part due to our reconstructing the annual mean rather than the summer mean; the greatest temperature anomalies in the HTM are thought to have occurred in the summer months and many proxies are more sensitive to summer than annual temperature (Kaufman et al., 2004). Importantly, in our reanalysis, the higher HTM temperatures do not translate to a marked increase in precipitation as would be expected from a thermodynamically-scaled relationship between temperature and precipitation. Instead, we find fractional precipitation within 2% of modern values (1.0 ± 0.02) during the HTM (Fig. 4b), with slightly higher-than-modern precipitation in central Greenland and slightly lower-than-modern precipitation in northwestern Greenland.

~~Our results allow us to investigate the relationship between temperature and precipitation. To facilitate comparison with thermodynamic scalings widely used by ice-sheet models, we define the relationship as exponential and find the best fit for our reanalysis:~~

$$P_{fraction} = \frac{P_{past}}{P_{modern}} = e^{\beta(T_{past} - T_{modern})}$$

where P is the precipitation rate, T is the temperature, and β is a scaling factor (Greve et al., 2011). For a given temperature change, a higher value of β results in a larger change in precipitation (orange in Fig. 11a). In ice-sheet models that use this scaling, it is commonly applied with a uniform β value for all locations (e.g., Huybrechts et al., 1991; Huybrechts, 2002; Greve, 1997; Grev We find that our best-fit scaling-factors (β) center on the Greve et al. (2011) value of 0.07 for locations around Greenland, but our scaling factors tend to be larger ($\beta >$0.10) where late-glacial precipitation is lowest and smaller ($\beta <$0.05) where late-glacial precipitation is highest (Figs. 4d and 11a).

The precipitation-temperature relationship in our reanalysis is driven by the assimilated ice-core records, though the spatial pattern of this relationship is also influenced by the spatial covariance structures of the prior ensembles. Previous work with ice-core records has found that the relationship between temperature and precipitation is frequency dependent, with a stronger relationship at lower frequencies (Cuffey and Clow, 1997); as expected, there is a similar frequency-dependence in our reanalysis. We find that an exponential scaling captures the low-frequency glacial to Holocene precipitation change; however, this fails at higher frequencies (Fig. 11b-e). To evaluate this frequency-dependence, we filtered our results using 6th order, low-pass and high-pass Butterworth filters with 5,000 year^{-1} cutoff frequencies. The low-pass filtered dataset shows the same precipitation-temperature relationship as the unfiltered dataset (Fig. 11c-d), while the high-pass filtered dataset shows that precipitation is less sensitive to changes in temperature (i.e., the value of β is lower) at these higher frequencies (Fig. 11e). This decoupling of temperature and precipitation is apparent in the amplitude difference of high-frequency signals in the glacial and the Holocene. In our temperature reanalysis, we find that high frequencies in the glacial have a greater amplitude than those in the Holocene, while in our precipitation reanalysis, we find the opposite. A single scaling, as is typically used in ice-sheet modeling, cannot capture this difference.

4 Evaluation

3.1 Independent proxy evaluation

Here we evaluate our results against proxy records that are excluded from ten of the iterations that make up the temperature and precipitation reanalyses. For this evaluation, we use the raw results (without a mean-bias correction). We find, however, that the mean biases are small relative to climate changes over the last 20,000 years; there is little difference between our bias-corrected and uncorrected results and it is unlikely that the mean bias has a large affect effect on our evaluation.

Evaluation against independent proxy data shows that our reanalysis captures the timing and magnitude of low-frequency climate changes (Figs. 6 and 7) and is an improvement over both the prior ensemble and TraCE-21ka (Figs. S3 and S4S6-S7 and S8-S9). Evaluation over the full 20,000 years of the temperature and precipitation results shows high, positive correlation coefficients (ranging from a minimum of 0.97 to maximum of 0.99), which indicate that the reanalysis captures both the timing and sign of climate events, while high CE (0.87-0.98) and low RMSE values (0.62-1.2 ‰ for δ^{18}O and 0.04-0.08 for accumulation) indicate that the reanalysis captures the magnitude of these events. Our skill during this longest evaluation period is primarily due to the presence of low-frequency climate changes, which tend to be coherent across Greenland, such

that evaluation over this full 20,000-year period shows more skill than evaluation over the full Holocene, which shows more skill than evaluation over just 5,000 years in the Holocene (or 5,000 years in the glacial) (Figs. 6 and 7).

Our posterior ensemble consistently shows improvement over the uninformed, constant prior ensemble during the 20,000-year evaluation period (Figs. S6 and S7). The TraCE-21ka simulation is also uninformed by the ice-core data, but it is transient and generally captures glacial to Holocene changes. Over the 20,000-year evaluation period, relative to the reconstructions, we find that TraCE-21ka has consistently lower correlation coefficients (0.86-0.96), lower CE values (0.50-0.86), and higher RMSE values (1.9-2.8 ‰ for δ^{18}O and 0.11-0.15 for accumulation) (Figs. S8 and S9). This comparison suggests that our reconstruction captures the timing and magnitude of the glacial to Holocene transition better than TraCE-21ka.

Even for the ~~these~~ shorter evaluation periods, which are dominated by high-frequency, spatially-incoherent noise, the reanalysis shows ~~overall improvement over~~ improvement over both the prior ensemble and TraCE-21ka (Figs. ~~S3 and S4).~~ S6-S7 and S8-S9); however, the improvement is not as consistent as for the 20,000-year evaluation period. For our reconstruction, correlation is positive except for three locations in the temperature reconstruction, with Holocene precipitation showing the largest correlation values (up to 0.60) and the total range being -0.30 to 0.60. The prior correlation is zero for these shorter evaluation periods and locations; however, TraCE-21ka shows correlation values ranging from -0.29 to 0.69, with eight negative correlations (more than we find for our reconstruction), but generally higher correlations in the Holocene than our reconstruction.

For the shorter evaluation periods, the reconstruction CE is generally negative (ranging from -82 to 0.17) with a few exceptions; however, the reconstruction may still be skillful (e.g., Cook et al., 1999). The skill of the reconstruction is better measured by the difference between prior and posterior CE due to the strong influence that bias can have on CE (Hakim et al., 2016). There is consistent improvement of the posterior CE over that of the prior (ranging from an increase of 4.7 to 3200) and over that of TraCE-21ka (ranging from an decrease of -6.9 to an increase of 230). RMSE is the most consistent of the skill metrics for these shorter evaluation periods, with our reconstruction showing improvement over both the prior and TraCE-21ka, the one exception being that TraCE-21ka has greater skill at the Agassiz ice-core site in the Holocene. For the reconstruction, RMSE values range from 0.24 to 1.8 ‰^{2} for temperature and 0.025 to 0.10 for precipitation.

For all evaluation periods and both variables, the ensemble calibration ratio (ECR) for the prior is skewed towards values greater than unity (0.66-8.7), which suggests that the prior ensemble tends to have too little spread. Conversely, for the posterior, the ~~ERC~~ ECR is generally less than unity (0.10-1.7) (Figs. 6 and 7), suggesting that the posterior ensemble has more ~~than enough spread given~~ spread than the error in the ensemble mean (as compared to the proxy records). This result indicates that the reanalysis ensemble encompasses the climate as recorded by the proxy records for most times and locations over the last 20,000 years.

3.2 Sensitivity evaluation

Proxy networks that change in time, such as ours, can introduce artificial discontinuities into the reanalysis, especially if the number of proxies is low or the proxy uncertainty values are inappropriate. We produce another reconstruction with a fixed proxy network, in which all assimilated proxy records participate in every time step in the reconstruction (see Sect. 2.3). A

comparison of these results with our main reanalysis shows no apparent discontinuities for the ensemble mean and 5th to 95th percentiles at Summit (Fig. 8) or other locations around Greenland.

Our results are sensitive to the δ^{18}O-T relationship. To test this, we compare the main reanalysis to the four scenarios (S1-S4, as described in Sect. 2.3.1) that use different δ^{18}O PSMs, each of which assumes a different slope and spatial pattern of the δ^{18}O-T relationship. We show this comparison for Summit as an example (Fig. 9), but the findings are applicable for all locations. As discussed previously, scenarios S1-S3 all assume that the δ^{18}O-T relationship has a uniform spatial pattern, but they each assume a different influence of precipitation seasonality. From these scenarios, we find that the temperature reconstruction is sensitive to the assumed precipitation seasonality, especially in the glacial where a stronger seasonality results in a greater glacial temperature anomaly. At Summit, this difference is nearly 10 °C between S1, which assumes no influence, and S3, which assumes the most influence of precipitation seasonality (Fig. 9). Similarly, the main reanalysis and S4 scenario both assume that the δ^{18}O-T relationship has a spatially-variable pattern, but S4 assumes a greater influence of precipitation seasonality. Again we find that the results are sensitive to assumed seasonality, with the greatest impact on the glacial-to-interglacial change.

~~We also find that the~~ The temperature results are also sensitive to the spatial pattern of the δ^{18}O-T relationship. ~~To test this , we compare~~ We find this by comparing the results from the S1-S3 scenarios that assume a spatially-uniform ~~pattern~~ relationship to results from the main reanalysis and S4 scenario that assume a spatially-variable ~~pattern~~relationship. The S1-S3 scenarios have a characteristic shape to their time series (Fig. 9), and, although the main reanalysis and S4 scenario generally fit this characteristic shape in the glacial and middle-late Holocene, in the early Holocene the main reanalysis and S4 diverge and show slower warming trends than the S1-S3 scenarios ~~. This indicates~~ (Fig. 9b). In addition, the reconstructions with spatially-varying δ^{18}O-T relationships show stronger north-south gradients during times of abrupt temperature change than those with spatially-constant relationships (Table S2). These findings indicate that there is new information added by using a PSM that ~~accounts for~~ includes spatial variability in precipitation seasonality.

For precipitation, we find that the results are sensitive to which accumulation record is assimilated at each ice-core site. As explained in Sect. 2.1 and S1, we use a low, moderate, and high accumulation record for most of the ice-core sites to produce the low, main, and high precipitation scenarios, respectively (Fig. 10). The largest spread among the scenarios is in the earliest part of the reconstruction, i.e. the last glacial through the early Holocene, since uncertainties in the ice thinning history have the greatest impact at greater depths (and hence, greater ages). There is also a larger spread among the scenarios at more southern locations because the accumulation record at Dye3 has both the most influence on southern Greenland (Sect. ~~S3~~S4) and the largest uncertainty in the ice thinning history (Fig. 3, Sect. ~~S1~~S2).

4 Discussion

~~Using paleoclimate data assimilation to combine ice-core records and the TraCE21ka climate-model simulation, we have obtained a Greenland climate reanalysis for the last 20,000 years that both captures the timing and magnitude of major climate events and shows good evaluation against independent proxy records.~~

4.1 Precipitation-temperature relationship

Our results allow us to investigate the relationship between temperature and precipitation. To facilitate comparison with thermodynamic scalings widely used by ice-sheet models, we define the relationship as exponential and find the best fit for our reanalysis:

$$P_{fraction} = \frac{P_{past}}{P_{modern}} = e^{\beta(T_{past} - T_{modern})} \tag{8}$$

where P is the precipitation rate, T is the temperature, and β is a scaling factor (Greve et al., 2011). For a given temperature change, a higher value of β results in a larger change in precipitation (orange in Fig. 11a). In ice-sheet models that use this scaling, it is commonly applied with a uniform β value for all locations (e.g., Huybrechts et al., 1991; Huybrechts, 2002; Greve, 1997; Grev
The best-fit scaling-factors (β) center on the Greve et al. (2011) value of 0.07 for locations around Greenland, but the scaling factors tend to be larger ($\beta >$0.10) where late-glacial precipitation is lowest and smaller ($\beta <$0.05) where late-glacial precipitation is highest (Figs. 4d and 11a).

Spatial varibility in the temperature-precipitation relationship was also found by a previous study, which looked at decadal averages over a recent 110 year period (Buchardt et al., 2012). Also using an exponential to represent the relationship, Buchardt et al. (2012)
found weaker relationships in northern than in central and southern Greenland (compare their Fig. 3 to our Fig. 11). In southern Greenland, they found a negative β value in southwest and a positive value in southeast Greenland, which they attribute to the Foehn effect. We find the opposite pattern, with larger, positive β values in southwest Greenland as compared to southeast Greenland. These differences may result from the time periods, timescales, and spatial resolutions considered in each investigation and from methodology. The spatial pattern found by (Buchardt et al., 2012) is limited to the ice-core locations
they considered, such that they have little information about the relationship near the ice-sheet edges. The spatial pattern in our reanalysis is influenced by both the ice-core records and the spatial covariance structures of the prior ensembles. With only one long ice core in southern Greenland, the east-west gradient in β values is highly influenced by the spatial patten inherited from TraCE-21ka.

The precipitation-temperature relationship in our reanalysis is driven by the assimilated ice-core records, though, as mentioned
above, the spatial pattern of this relationship is also influenced by the spatial covariance structures of the prior ensembles. Previous work with ice-core records has found that the relationship between temperature and precipitation is frequency dependent, with a stronger relationship at lower frequencies (Cuffey and Clow, 1997); as expected, there is a similar frequency-dependence in our reanalysis. We find that an exponential scaling captures the low-frequency glacial to Holocene precipitation change; however, this fails at higher frequencies (Fig. 11b-e). To evaluate this frequency-dependence, we filtered our results using 6th
order, low-pass and high-pass Butterworth filters with 5,000 year^{-1} cutoff frequencies. The low-pass filtered dataset shows the same precipitation-temperature relationship as the unfiltered dataset (Fig. 11c-d), while the high-pass filtered dataset shows that precipitation is less sensitive to changes in temperature (i.e., the value of β is lower) at these higher frequencies (Fig. 11e). This decoupling of temperature and precipitation is apparent in the amplitude difference of high-frequency signals in the glacial and the Holocene. In our temperature reanalysis, we find that high frequencies in the glacial have a greater amplitude

than those in the Holocene, while in our precipitation reanalysis, we find the opposite. A single scaling, as is typically used in ice-sheet modeling, cannot capture this difference.

We examine how the sensitivity experiments (Figs. 9 and 10) affect the scaling factor (β) in the precipitation-temperature relationship. We pair the five temperature reconstructions (main, S1-S4) and three precipitation reconstructions (low, moderate, and high) into fifteen possible combinations and conduct the same analysis as described above. Across these fifteen combinations, we find that the spatial pattern of β is robust (Fig. 11a). The exact magnitude depends primarily on the temperature reconstruction and how cold it is in the glacial, with colder temperatures giving lower β values. To a lesser degree, the magnitude also depends on the precipitation reconstruction, with wetter scenarios giving lower β values. As previously, we find that the low-pass filtered datasets have the same or nearly the same β value as the unfiltered dataset, while the high-pass filtered datasets have consistently lower β values. As an example of this, Table S3 shows the β value found for the Kangerlussuaq region for all fifteen combinations and three filtering options.

4.2 Spatial patterns during abrupt climate change events

Our paleoclimate data assimilation framework is not limited to the assimilation of δ^{18}O and accumulation rate records. In this section we examine how our reanalysis compares to reconstructions driven by another type of proxy, δ^{15}N of N_2. In particular, we focus on abrupt temperature events, for which there is previous work using δ^{15}N (e.g., Severinghaus et al., 1998; Guillevic et al., 2013; E Abrupt temperature events increase the thermal gradient in the firn – the upper, porous portion of the ice column – which leads to fractionation of the stable isotopes of N_2 (Severinghaus et al., 1998). Using a firn compaction model, temperature can be derived from δ^{15}N measurements with inverse methods (e.g., Severinghaus et al., 1998; Severinghaus and Brook, 1999; Guillevic et al., 20 We assimilate temperatures derived from δ^{15}N data collected from the GISP2, NGRIP, and NEEM ice cores (Buizert et al., 2014).

Our reanalysis and the Buizert et al. (2014) records cover three abrupt temperature events – the Bølling-Allerød warming (14.6 ka), cooling into the Younger Dryas (12.9 ka), and warming at the end of the Younger Dyras (11.5 ka). For these three abrupt temperature events, the δ^{15}N-derived temperature records show larger temperature changes at GISP2 in central Greenland (Buizert et al., 2014), while the δ^{18}O changes are larger at NEEM in northwest Greenland (Fig. 2). We perform three sets of experiments to investigate how these two sets of proxy records affect the mean spatial pattern indicated by our reanalysis during abrupt temperature events. The first experiment, "O8", assimilates all eight δ^{18}O records with one 100-member prior ensemble from TraCE-21ka. This results in a 100-member reanalysis ensemble. The second experiment, "N3O5", assimilates all three of the δ^{15}N-derived temperature records and the five remaining δ^{18}O records (those that do not overlap with the δ^{15}N sites), using the same 100-member prior ensemble as used in the O8 experiment. Finally, we perform a modified experiment, "N3O5_BA", with both δ^{18}O and δ^{15}N records, but using a prior ensemble selected from the 1,000 years surrounding the Bølling-Allerød warming. To maintain a 100-member prior ensemble, we use decadal rather than 50-year averages of TraCE-21ka for this experiment. This adjustment does not affect the comparison (results not shown). Detailed methods for these experiments are given in Sect. S6.

For both the "O8" and "N3O5" experiments, the spatial pattern for the abrupt climate change events are similar to our main reanalysis, with the largest magnitude temperature changes in northern and northwestern Greenland, decreasing magnitude with decreasing latitude, and slightly larger change in the central east and southeast than corresponding western regions. For example, the spatial pattern of the Younger Dryas cooling is nearly the same regardless of which grouping of records is assimilated (Figs. 12a and b). This overall finding is robust to different combinations of these proxy records; for example, if we assimilate just the three δ^{15}N-derived temperature records and no δ^{18}O records (results not shown) the pattern is not substantially changed. This pattern of temperature change differs from spatial patterns inferred previously from δ^{15}N for various abrupt climate events (Guillevic et al., 2013; Buizert et al., 2014); however, the O8 and N3O5 experiments show that these differences are not due to the assimilation of δ^{18}O rather than δ^{15}N-derived temperatures.

For our third experiment, N3O5_BA, in which we restrict the prior ensemble to the Bølling-Allerød warming, the spatial patterns obtained by our experiment are similar to those reported by Buizert et al. (2014). In TraCE-21ka, the Bølling-Allerød warming is forced by a sudden termination of freshwater forcing in the North Atlantic (Liu et al., 2009). This forcing leads to large temperature fluctuations in southern Greenland that decrease with increasing latitude. With this covariance pattern dominating the prior ensemble, the N3O5_BA reconstruction indeed shows the largest temperature changes in southern Greenland, followed by central and then northern Greenland (Figs. 12c, S12c, and S13c).

Importantly, these results depend on the spatial patterns in the prior ensemble, which themselves are a consequence of the particular forcing applied to the climate simulation. Some simulations suggest that the rate and timing of meltwater forcing imposed by TraCE-21ka may not be necessary to explain the abrupt climate change events and show that different spatial patterns result from different forcing (e.g., Obase and Abe-Ouchi, 2019). Prior ensembles selected from such simulations may result in different spatial patterns that are also consistent, within uncertainty, with the proxy data. We suggest that current ice-core records are insufficient to place a strong constraint on the spatial pattern of abrupt climate events, and additional data, especially from southern Greenland (Sect. S4), would be beneficial. Future work to explore this question in more detail will also require model simulations that sample a greater range of forcing uncertainty.

4.3 Climate and the ice sheet: A case study of Southwest Greenland

Our main reanalysis is one of very few spatially-complete time series of Greenland climate over the last 20,000 years. Here, we compare our reanalysis with other Greenland climate histories, and suggest that, together, they should be treated as an ensemble of climate boundary conditions that can be used to produce ensembles of ice-sheet model simulations. These climate histories can also be further evaluated using a combination ice-sheet models and independent constraints from the glacial-geologic record of past ice-sheet configurations.

We compare our results with the recent reconstruction of Buizert et al. (2018), hereafter referred to as the B18 reconstruction. The B18 temperature reconstruction was produced by adjusting a part of the ~~TraCE21ka~~ TraCE-21ka temperature field that is affected by changes in the Atlantic meridional overturning circulation such that the full temperature field provides a good match to an average of three δ^{15}N-derived temperature records recovered from ice cores. The B18 snow-accumulation reconstruction is simply a reference climatology scaled to accumulation rates from the GISP2 ice core. We treat this as a

precipitation reconstruction, but note that accumulation may be less than precipitation at some locations around Greenland, especially near the coast. It is also informative to compare these results with our S4 temperature and high precipitation reconstructions ~~, (~~hereafter referred to as S4 and high P or simply 'sensitivity' as in Fig. 13), as well as ~~with the TraCE21ka~~ the TraCE-21ka simulation itself (i.e., the climate model output, unconstrained by data). For brevity, we focus on the area around Kangerlussuaq in southwest Greenland, but the comparisons are generally applicable to any region of Greenland. Southwest Greenland is of interest because the ice-sheet behavior here is primarily a response to changes in surface forcing (i.e., temperature and precipitation) because there are few tidewater glaciers (Cuzzone et al., 2019). Furthermore, the Kangerlussuaq region has a particularly well-documented ice-sheet retreat history through the Holocene (Young and Briner, 2015; Lesnek and Briner, 2018).

In the Kangerlussuaq region, the B18 reconstruction shows more extreme temperature changes than our reconstructions, with late-glacial (20-15 ka) anomalies of about -20 °C and peak HTM temperature anomalies of about $+2$ °C at 9 ka (Fig. 13). B18 also shows a faster rate of transition between the glacial and Holocene, reaching temperatures close to modern by 10 ka. In constrast, ~~TraCE21ka~~ TraCE-21ka shows more moderate temperature anomalies and a slower transition, with late-glacial anomalies of about -8.6 °C and near-modern temperatures that first appear around 7 ka. ~~TraCE21ka~~ TraCE-21ka has no obvious HTM in this location or any location in Greenland. Our main reanalysis and the S4 version of our temperature reanalysis both lie between B18 and ~~TraCE21ka~~TraCE-21ka, with late-glacial anomalies of about -12 and -14 °C, respectively, Holocene peak temperature anomalies of $+1$ °C around 5 ka, and temperatures close to modern first appearing around 8 ka.

For precipitation, the B18 reconstruction again tends to show the largest fluctuations and fastest transition, with a late-glacial precipitation fraction of about 0.26 and precipitation rates close to modern first appearing just after 10 ka. ~~TraCE21ka~~ TraCE-21ka again shows the most moderate fluctuations and a slower transition, with a late-glacial fraction of about 0.38 and rates close to modern not appearing until around 5 ka. Our main reanalysis and high P lie in the middle during the late-glacial, with fractions of about 0.32 and 0.36, respectively; however, our main reanalysis has a slow transition into the Holocene, similar to ~~TraCE21ka~~TraCE-21ka, while high P has a fast transition similar to B18. In the Holocene, high P shows the most elevated precipitation out of all the reconstructions, with 10-15% more precipitation than modern occurring around 7-3 ka. B18 shows precipitation values similar to modern for the last 10,000 years of the Holocene, while ~~TraCE21ka~~ TraCE-21ka and our main reanalysis show lower-than-modern precipitation throughout most of the Holocene.

All of these paleoclimate reconstructions – our main reanalysis, the sensitivity scenarios, and B18 – are plausible histories of temperature and precipitation over Greenland. Given any past change in the ice-sheet, each of these histories has a different implication for ice-sheet sensitivity to climate, the veracity of which could be tested by using them to force an ice-sheet model and comparing this ensemble of results to the geologic record.

Our results have potentially important implications for the response of the Greenland Ice Sheet to climate change. In particular, we find maximum Holocene temperatures were reached around 5 ka, which is between 500 years earlier and 4,000 years later than most previous estimates. Moreover, there is little corresponding change in precipitation in our main reanalysis and the low sensitivity scenario. If these findings are correct, they imply a relatively rapid response to temperature forcing for sections of the ice sheet margin that retreated less than a century later (Young and Briner, 2015). A caveat is that proxy

data remains very sparse, particularly in southern Greenland, where the poorly-resolved Dye3 core is the only long record. Future work to obtain improved measurements on the Dye3 core, or to gather new data from southern Greenland, would help to alleviate this limitation, as would the incorporation of data from off the ice sheet, such as from lake and ocean sediment cores.

An important distinction among various different paleoclimate reconstructions for Greenland is in the treatment of elevation changes. Any paleoclimate reconstruction from ice-core records is complicated by ice-sheet elevation changes. In Vinther et al. (2009), it is assumed that the climate history is the same at all locations around Greenland, and that any differences among the ice core paleotemperature records is a result of that elevation change. In B18, past elevation changes are assumed to be negligible. In our reconstruction, the impact of elevation change on the spatial covariances of temperature and precipitation is implicitly accounted for as part of the data assimilation methodology. Formally, our reconstruction is of surface climate, not climate at a fixed elevation~~, but it~~ . Consequently, our reanalysis may not be directly comparable to other paleoclimate reconstructions. For example, the HTM is commonly reconstructed as an early Holocene event in records that are at a fixed or nearly-fixed elevation. In our reanalysis, the early Holocene is cooler than the mid Holocene. Changes in the ice-surface elevation could account for this apparent discrepancy. Thinning in the early Holocene (Vinther et al., 2009) would result in a lowering of the ice surface and an apparent warming at the ice surface due to lapse rate effects. This warming signal would be captured in ice-core records. If the warming trend due to surface lowering occurs at the same time as an overall climate cooling, then the climate signal would be dampened or possibly reversed.

Our method depends on the accuracy of the climate-elevation relationships in our prior – i.e. in the ~~TraCE21ka~~ TraCE-21ka climate model simulation, which probably does not capture such relationships with particularly high fidelity since the model resolution is low and the climate and ice-sheet models are not coupled. Future work could take advantage of the probabilistic relationships among accumulation, temperature, and surface elevation as simulated in fine-scale regional climate models (Edwards et al., 2014).

5 Conclusions

Paleoclimate data assimilation is a novel method for reconstructing climate fields over the Greenland Ice Sheet. Our approach, combining ice-core records with a climate-model simulation, provides complete spatial reanalyses of both temperature and precipitation covering the last 20,000 years. Evaluation against independent proxy records shows that this methodology leads to significant and meaningful improvement over the prior ensemble (drawn from a climate simulation) ~~. The results also quantify uncertainty in all aspects, which provides~~ and TraCE-21ka. Between the posterior ensemble and sensitivity experiments, our results provide a range of climate scenarios for ice-sheet modeling. Moreover, independently reconstructing both precipitation and temperature allows the assumption of purely thermodynamic control on precipitation to be relaxed, and an examination of the relationship between these quantities over a range of timescales. Specifically, we find that the Clausius-Claypeyron scaling is a good approximation over glacial-interglacial cycles, but not for shorter timescales where precipitation variability partially decouples from temperature.

~~Our results have potentially important implications for the response of the Greenland ice sheet to climate change. In particular, we find maximum Holocene temperatures were reached around 5 ka, which is between 500 years earlier and 4,000 years later than previous estimates. Moreover, there is little corresponding change in precipitation in our main reanalysis and one of the sensitivity scenarios. If these findings are correct, they imply a relatively rapid response to temperature forcing for sections of the ice sheet margin that retreated less than a century later (Young and Briner, 2015). A caveat is that proxy data remains very sparse, particularly in southern Greenland, where the poorly-resolved Dye3 core is the only long record. Future work to obtain improved measurements on the Dye3 core, or gather other new data from southern Greenland, would help to alleviate this limitation, as would the incorporation of data from off the ice sheet, such as from lake and ocean sediment cores.~~

~~Finally, we note that our paleoclimate data assimilation approach would~~ Paleoclimate reconstructions would benefit from a larger selection of long climate-model simulations at higher resolution. Particularly valuable would be transient simulations that include water isotopes as prognostic variables, which allows for direct assimilation of water isotope ratios ~~(Steiger et al., 2017)~~(Steiger et al. rather than the use of an explicit proxy system model between temperature and δ^{18}O. Recent work shows significant improvements to the realism of water-isotope enabled models in the polar regions ~~(Nusbaumer et al., 2017; Dütsch et al., 2019)~~(Nusbaumer et al., 2 and longer simulations, once available, should allow us to further improve upon the results we have presented here. In principle, our method could also be applied to climate-model simulations that include a fully-coupled Greenland Ice Sheet. At present, fully-coupled simulations of Greenland over thousands of years are prohibitively expensive except at low resolution, and the limited work that has been done shows significant biases (Vizcaino et al., 2015). Nevertheless, incorporating data assimilation into such models would provide the groundwork for more-complete data-constrained simulations as computing power becomes less of a limiting factor in the future.

Code and data availability. The paleoclimate reconstructions in this paper made use of code from the Last Millennium Reanalysis project, which is publicly available at https://github.com/modons/LMR (Hakim, 2019). The reconstructions, along with the new accumulation histories for Dye3, GRIP, and NGRIP, will be made publicly available at www.pangaea.de at the time of publication.

Author contributions. JAB and EJS conceived the idea for the study. JAB wrote code improvements necessary for this work, completed the paleoclimate reconstructions with guidance from EJS and GJH, conducted all analyses of the results, and wrote the first draft of the paper. GJH provided expert advice on data assimilation methodology and code development. TJF made the calculations of ice flow used to model accumulation for the Dye3, GRIP, and NGRIP cores. All authors contributed to the final version of the manuscript.

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

Disclaimer. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements. Funding for this study was provided by the National Science Foundation Grant ARCSS no. 1503281 awarded to the University of Washington. In addition, this material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant no. DGE-1256082 awarded to JAB. GJH also acknowledges support from the Heising Simons Foundation through grant 2016-14 and the NSF through grant AGS-1602223. We thank Robert Tardif for help with code development, and we thank Joshua Anderson, Bo Vinther, Christo Buizert for their help compiling the ice-core data. We also thank the Snow on Ice project members for their discussions and support, especially Joshua Cuzzone, Jason Briner, and Elizabeth Thomas.

References

Alley, R. B., Shuman, C., Meese, D., Gow, A., Taylor, K., Cuffey, K., Fitzpatrick, J., Grootes, P., Zielinski, G., Ram, M., et al.: Visual-stratigraphic dating of the GISP2 ice core: Basis, reproducibility, and application, Journal of Geophysical Research: Oceans, 102, 26 367–26 381, 1997.

Alley, R. B., Andrews, J. T., Brigham-Grette, J., Clarke, G., Cuffey, K. M., Fitzpatrick, J., Funder, S., Marshall, S., Miller, G., Mitrovica, J., et al.: History of the Greenland Ice Sheet: paleoclimatic insights, Quaternary Science Reviews, 29, 1728–1756, 2010.

Andersen, K. K., Azuma, N., Barnola, J.-M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., et al.: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147, 2004.

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Röthlisberger, R., Ruth, U., Siggaard-Andersen, M.-L., Steffensen, J. P., Dahl-Jensen, D., et al.: The Greenland ice core chronology 2005, 15–42 ka. Part 1: constructing the time scale, Quaternary Science Reviews, 25, 3246–3257, 2006.

Armengaud, A., Koster, R. D., Jouzel, J., and Ciais, P.: Deuterium excess in Greenland snow: Analysis with simple and complex models, Journal of Geophysical Research: Atmospheres, 103, 8947–8953, 1998.

Bindschadler, R. A., Nowicki, S., Abe-Ouchi, A., Aschwanden, A., Choi, H., Fastook, J., Granzow, G., Greve, R., Gutowski, G., Herzfeld, U., et al.: Ice-sheet model sensitivities to environmental forcing and their use in projecting future sea level (the SeaRISE project), Journal of Glaciology, 59, 195, 2013.

Bintanja, R., van de Wal, R. S., and Oerlemans, J.: Modelled atmospheric temperatures and global sea levels over the past million years, Nature, 437, 125, 2005.

Bothe, O., Evans, M., Donado, L., Bustamante, E., Gergis, J., Gonzalez-Ruoco, J., Goosse, H., Hegerl, G., Hind, A., Jungclaus, J., et al.: Continental-scale temperature variability in PMIP3 simulations and PAGES 2k regional temperature reconstructions over the past millennium, Climate of the Past, 11, 1673–1699, 2015.

Buchardt, S. L., Clausen, H. B., Vinther, B. M., and Dahl-Jensen, D.: Investigating the past and recent [delta] 18O-accumulation relationship seen in Greenland ice cores, Climate of the Past, 8, 2053, 2012.

Buizert, C., Gkinis, V., Severinghaus, J. P., He, F., Lecavalier, B. S., Kindler, P., Leuenberger, M., Carlson, A. E., Vinther, B., Masson-Delmotte, V., et al.: Greenland temperature response to climate forcing during the last deglaciation, Science, 345, 1177–1180, 2014.

Buizert, C., Keisling, B., Box, J., He, F., Carlson, A., Sinclair, G., and DeConto, R.: Greenland-Wide Seasonal Temperatures During the Last Deglaciation, Geophysical Research Letters, 45, 1905–1914, 2018.

Cauquoin, A., Werner, M., and Lohmann, G.: Water isotopes-climate relationships for the mid-Holocene and preindustrial period simulated with an isotope-enabled version of MPI-ESM, Climate of the Past, 15, 1913–1937, 2019.

Conway, H., Hall, B., Denton, G., Gades, A., and Waddington, E.: Past and future grounding-line retreat of the West Antarctic Ice Sheet, Science, 286, 280–283, 1999.

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K.: Drought reconstructions for the continental United States, Journal of Climate, 12, 1145–1162, 1999.

Cuffey, K. M. and Clow, G. D.: Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition, Journal of Geophysical Research: Oceans, 102, 26 383–26 396, 1997.

Cuffey, K. M. and Steig, E. J.: Isotopic diffusion in polar firn: implications for interpretation of seasonal climate parameters in ice-core records, with emphasis on central Greenland, Journal of Glaciology, 44, 273–284, 1998.

Cuffey, K. M., Clow, G. D., Alley, R. B., Stuiver, M., Waddington, E. D., and Saltus, R. W.: Large arctic temperature change at the Wisconsin-Holocene glacial transition, Science, 270, 455–458, 1995.

Cuzzone, J. K., Schlegel, N.-J., Morlighem, M., Larour, E., Briner, J. P., Seroussi, H., and Caron, L.: The impact of model resolution on the simulated Holocene retreat of the southwestern Greenland ice sheet using the Ice Sheet System Model (ISSM), The Cryosphere, 13, 879–893, 2019.

Dahl-Jensen, D., Johnsen, S., Hammer, C., Clausen, H., and Jouzel, J.: Past accumulation rates derived from observed annual layers in the GRIP ice core from Summit, Central Greenland, in: Ice in the climate system, pp. 517–532, Springer, 1993.

Dahl-Jensen, D., Mosegaard, K., Gundestrup, N., Clow, G. D., Johnsen, S. J., Hansen, A. W., and Balling, N.: Past temperatures directly from the Greenland ice sheet, Science, 282, 268–271, 1998.

Dahl-Jensen, D., Gundestrup, N., Gogineni, S. P., and Miller, H.: Basal melt at NorthGRIP modeled from borehole, ice-core and radio-echo sounder observations, Annals of Glaciology, 37, 207–212, 2003.

Dahl-Jensen, D., Albert, M., Aldahan, A., Azuma, N., Balslev-Clausen, D., Baumgartner, M., Berggren, A.-M., Bigler, M., Binder, T., Blunier, T., et al.: Eemian interglacial reconstructed from a Greenland folded ice core, Nature, 493, 489, 2013.

Dansgaard, W.: Stable isotopes in precipitation, Tellus, 16, 436–468, 1964.

Dansgaard, W. and Johnsen, S.: A flow model and a time scale for the ice core from Camp Century, Greenland, Journal of Glaciology, 8, 215–223, 1969.

Dansgaard, W., Clausen, H., Gundestrup, N., Hammer, C., Johnsen, S., Kristinsdottir, P., and Reeh, N.: A new Greenland deep ice core, Science, 218, 1273–1277, 1982.

Dütsch, M., Blossey, P. N., Steig, E. J., and Nusbaumer, J. M.: Non-equilibrium fractionation during ice cloud formation in iCAM5: evaluating the common parameterization of supersaturation as a linear function of temperature, Journal of Advances in Modeling Earth Systems, 2019.

Edwards, T., Fettweis, X., Gagliardini, O., Gillet-Chaulet, F., Goelzer, H., Gregory, J., Hoffman, M., Huybrechts, P., Payne, A., Perego, M., et al.: Probabilistic parameterisation of the surface mass balance–elevation feedback in regional climate model simulations of the Greenland ice sheet, The Cryosphere, 8, 181–194, 2014.

Fisher, D. A., Reeh, N., and Clausen, H.: Stratigraphic noise in time series derived from ice cores, Annals of Glaciology, 7, 76–83, 1985.

Fudge, T., Markle, B. R., Cuffey, K. M., Buizert, C., Taylor, K. C., Steig, E. J., Waddington, E. D., Conway, H., and Koutnik, M.: Variable relationship between accumulation and temperature in West Antarctica for the past 31,000 years, Geophysical Research Letters, 43, 3795–3803, 2016.

Gierz, P., Werner, M., and Lohmann, G.: Simulating climate and stable water isotopes during the L ast I nterglacial using a coupled climate-isotope model, Journal of Advances in Modeling Earth Systems, 9, 2027–2045, 2017.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J., and Vinther, B. M.: Water isotope diffusion rates from the NorthGRIP ice core for the last 16,000 years–Glaciological and paleoclimatic implications, Earth and Planetary Science Letters, 405, 132–141, 2014.

Greve, R.: Application of a polythermal three-dimensional ice sheet model to the Greenland ice sheet: response to steady-state and transient climate scenarios, Journal of Climate, 10, 901–918, 1997.

Greve, R., Saito, F., and Abe-Ouchi, A.: Initial results of the SeaRISE numerical experiments with the models SICOPOLIS and IcIES for the Greenland ice sheet, Annals of Glaciology, 52, 23–30, 2011.

Grootes, P. and Stuiver, M.: Oxygen 18/16 variability in Greenland snow and ice with 10- 3-to 105-year time resolution, Journal of Geophysical Research: Oceans, 102, 26 455–26 470, 1997.

Guillevic, M., Bazin, L., Landais, A., Kindler, P., Orsi, A., Masson-Delmotte, V., Blunier, T., Buchardt, S. L., Capron, E., Leuenberger, M., Martinerie, P., Prié, F., and Vinther, B. M.: Spatial gradients of temperature, accumulation and δ 18 O-ice in Greenland over a series of Dansgaard-Oeschger events., Climate of the Past, 9, https://doi.org/10.5194/cp-9-1029-2013, 2013.

Hakim, G. J.: Source code for the Last Millennium Reanalysis (LMR) project, available at: https://github.com/modons/LMR, 2019.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, 2016.

Harrison, S., Bartlein, P., Brewer, S., Prentice, I., Boyd, M., Hessler, I., Holmgren, K., Izumi, K., and Willis, K.: Climate model benchmarking with glacial and mid-Holocene climates, Climate Dynamics, 43, 671–688, 2014.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1108, 2009.

He, F., Shakun, J. D., Clark, P. U., Carlson, A. E., Liu, Z., Otto-Bliesner, B. L., and Kutzbach, J. E.: Northern Hemisphere forcing of Southern Hemisphere climate during the last deglaciation, Nature, 494, 81, 2013.

Huybers, P. and Wunsch, C.: A depth-derived Pleistocene age model: Uncertainty estimates, sedimentation variability, and nonlinear climate change, Paleoceanography, 19, 2004.

Huybrechts, P.: Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles, Quaternary Science Reviews, 21, 203–231, 2002.

Huybrechts, P., Letreguilly, A., and Reeh, N.: The Greenland ice sheet and greenhouse warming, Palaeogeography, Palaeoclimatology, Palaeoecology, 89, 399–412, 1991.

Johnsen, S., Dansgaard, W., and White, J.: The origin of Arctic precipitation under present and glacial conditions, Tellus B: Chemical and Physical Meteorology, 41, 452–468, 1989.

Johnsen, S. J., Clausen, H. B., Dansgaard, W., Gundestrup, N. S., Hammer, C. U., Andersen, U., Andersen, K. K., Hvidberg, C. S., Dahl-Jensen, D., Steffensen, J. P., et al.: The δ18O record along the Greenland Ice Core Project deep ice core and the problem of possible Eemian climatic instability, Journal of Geophysical Research: Oceans, 102, 26 397–26 410, 1997.

Jouzel, J., Alley, R. B., Cuffey, K., Dansgaard, W., Grootes, P., Hoffmann, G., Johnsen, S. J., Koster, R., Peel, D., Shuman, C., et al.: Validity of the temperature reconstruction from water isotopes in ice cores, Journal of Geophysical Research: Oceans, 102, 26 471–26 487, 1997.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–472, 1996.

Kapsner, W., Alley, R. B., Shuman, C., Anandakrishnan, S., and Grootes, P.: Dominant influence of atmospheric circulation on snow accumulation in Greenland over the past 18,000 years, Nature, 373, 52, 1995.

Kaufman, D. S., Ager, T. A., Anderson, N. J., Anderson, P. M., Andrews, J. T., Bartlein, P. J., Brubaker, L. B., Coats, L. L., Cwynar, L. C., Duvall, M. L., et al.: Holocene thermal maximum in the western Arctic (0–180 W), Quaternary Science Reviews, 23, 529–560, 2004.

Kindler, P., Guillevic, M., Baumgartner, M. F., Schwander, J., Landais, A., and Leuenberger, M.: Temperature reconstruction from 10 to 120 kyr b2k from the NGRIP ice core, Climate of the Past, 10, 887–902, 2014.

Kobashi, T., Menviel, L., Jeltsch-Thömmes, A., Vinther, B. M., Box, J. E., Muscheler, R., Nakaegawa, T., Pfister, P. L., Döring, M., Leuenberger, M., et al.: Volcanic influence on centennial to millennial Holocene Greenland temperature change, Scientific reports, 7, 1441, 2017.

Krinner, G. and Werner, M.: Impact of precipitation seasonality changes on isotopic signals in polar ice cores: a multi-model analysis, Earth and Planetary Science Letters, 216, 525–538, 2003.

Langen, P., Mottram, R., Christensen, J., Boberg, F., Rodehacke, C., Stendel, M., Van As, D., Ahlstrøm, A., Mortensen, J., Rysgaard, S., et al.: Quantifying energy and mass fluxes controlling Godthåbsfjord freshwater input in a 5-km simulation (1991–2012), Journal of Climate, 28, 3694–3713, 2015.

Langen, P. L., Fausto, R. S., Vandecrux, B., Mottram, R. H., and Box, J. E.: Liquid water flow and retention on the Greenland ice sheet in the regional climate model HIRHAM5: Local and large-scale impacts, Frontiers in Earth Science, 4, 110, 2017.

Lasher, G. E., Axford, Y., McFarlin, J. M., Kelly, M. A., Osterberg, E. C., and Berkelhammer, M. B.: Holocene temperatures and isotopes of precipitation in Northwest Greenland recorded in lacustrine organic materials, Quaternary Science Reviews, 170, 45–55, 2017.

Latif, M. and Keenlyside, N. S.: A perspective on decadal climate variability and predictability, Deep Sea Research Part II: Topical Studies in Oceanography, 58, 1880–1894, 2011.

Lecavalier, B. S., Milne, G. A., Simpson, M. J., Wake, L., Huybrechts, P., Tarasov, L., Kjeldsen, K. K., Funder, S., Long, A. J., Woodroffe, S., et al.: A model of Greenland ice sheet deglaciation constrained by observations of relative sea level and ice extent, Quaternary Science Reviews, 102, 54–84, 2014.

Lecavalier, B. S., Fisher, D. A., Milne, G. A., Vinther, B. M., Tarasov, L., Huybrechts, P., Lacelle, D., Main, B., Zheng, J., Bourgeois, J., et al.: High Arctic Holocene temperature record from the Agassiz ice cap and Greenland ice sheet evolution, Proceedings of the National Academy of Sciences, 114, 5952–5957, 2017.

Lesnek, A. J. and Briner, J. P.: Response of a land-terminating sector of the western Greenland Ice Sheet to early Holocene climate change: Evidence from 10Be dating in the Søndre Isortoq region, Quaternary Science Reviews, 180, 145–156, 2018.

Liu, Z., Otto-Bliesner, B., He, F., Brady, E., Tomas, R., Clark, P., Carlson, A., Lynch-Stieglitz, J., Curry, W., Brook, E., et al.: Transient simulation of last deglaciation with a new mechanism for Bølling-Allerød warming, Science, 325, 310–314, 2009.

Liu, Z., Carlson, A. E., He, F., Brady, E. C., Otto-Bliesner, B. L., Briegleb, B. P., Wehrenberg, M., Clark, P. U., Wu, S., Cheng, J., et al.: Younger Dryas cooling and the Greenland climate response to CO2, Proceedings of the National Academy of Sciences, 109, 11 101–11 104, 2012.

McFarlin, J. M., Axford, Y., Osburn, M. R., Kelly, M. A., Osterberg, E. C., and Farnsworth, L. B.: Pronounced summer warming in northwest Greenland during the Holocene and Last Interglacial, Proceedings of the National Academy of Sciences, 115, 6357–6362, 2018.

Nash, J. and Sutcliffe, J.: River forcasting using conceptual models, 1. A discussion of principles, J. Hydrol, 10, 280–290, 1970.

Nielsen, L. T., Aðalgeirsdóttir, G., Gkinis, V., Nuterman, R., and Hvidberg, C. S.: The effect of a Holocene climatic optimum on the evolution of the Greenland ice sheet during the last 10 kyr, Journal of Glaciology, 64, 477–488, 2018.

Nusbaumer, J., Wong, T. E., Bardeen, C., and Noone, D.: Evaluating hydrological processes in the C ommunity A tmosphere M odel V ersion 5 (C AM5) using stable isotope ratios of water, Journal of Advances in Modeling Earth Systems, 9, 949–977, 2017.

Obase, T. and Abe-Ouchi, A.: Abrupt Bølling-Allerød Warming Simulated under Gradual Forcing of the Last Deglaciation, Geophysical Research Letters, 46, 11 397–11 405, 2019.

Okazaki, A. and Yoshimura, K.: Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction., Climate of the Past, 13, 2017.

Okazaki, A. and Yoshimura, K.: Global evaluation of proxy system models for stable water isotopes with realistic atmospheric forcing, Journal of Geophysical Research: Atmospheres, 124, 8972–8993, 2019.

Pausata, F. S. and Löfverström, M.: On the enigmatic similarity in Greenland δ18O between the Oldest and Younger Dryas, Geophysical Research Letters, 42, 10–470, 2015.

Peltier, W.: Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE, Annu. Rev. Earth Planet. Sci., 32, 111–149, 2004.

Pollard, D. and DeConto, R.: Description of a hybrid ice sheet-shelf model, and application to Antarctica, Geoscientific Model Development, 5, 1273–1295, 2012.

Rasmussen, S. O., Andersen, K. K., Svensson, A., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., et al.: A new Greenland ice core chronology for the last glacial termination, Journal of Geophysical Research: Atmospheres, 111, 2006.

Rasmussen, S. O., Abbott, P., Blunier, T., Bourne, A., Brook, E. J., Buchardt, S. L., Buizert, C., Chappellaz, J., Clausen, H. B., Cook, E., et al.: A first chronology for the North Greenland Eemian Ice Drilling (NEEM) ice core, Climate of the Past, 9, 2013.

Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., et al.: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy, Quaternary Science Reviews, 106, 14–28, 2014.

Raymond, C. F.: Deformation in the vicinity of ice divides, Journal of Glaciology, 29, 357–373, 1983.

Reeh, N. and Fisher, D.: Noise in accumulation rate and δ (18O) time series as determined from comparison of adjacent Greenland and Devon Island ice cap cores, 1983.

Robin, G. d. Q.: Ice cores and climatic change, Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 280, 143–168, 1977.

Roe, G. H. and Lindzen, R. S.: The mutual interaction between continental-scale ice sheets and atmospheric stationary waves, Journal of Climate, 14, 1450–1465, 2001.

Roy, K. and Peltier, W.: Relative sea level in the Western Mediterranean basin: A regional test of the ICE-7G_NA (VM7) model and a constraint on Late Holocene Antarctic deglaciation, Quaternary Science Reviews, 183, 76–87, 2018.

Schüpbach, S., Fischer, H., Bigler, M., Erhardt, T., Gfeller, G., Leuenberger, D., Mini, O., Mulvaney, R., Abram, N. J., Fleet, L., et al.: Greenland records of aerosol source and atmospheric lifetime changes from the Eemian to the Holocene, Nature communications, 9, 1476, 2018.

Severinghaus, J. P. and Brook, E. J.: Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice, Science, 286, 930–934, 1999.

Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B., and Bender, M. L.: Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice, Nature, 391, 141–146, 1998.

Simpson, M. J., Milne, G. A., Huybrechts, P., and Long, A. J.: Calibrating a glaciological model of the Greenland ice sheet from the Last Glacial Maximum to present-day using field observations of relative sea level and ice extent, Quaternary Science Reviews, 28, 1631–1657, 2009.

Steig, E. J., Grootes, P. M., and Stuiver, M.: Seasonal precipitation timing and ice core records, Science, 266, 1885–1887, 1994.

Steiger, N. J., Steig, E. J., Dee, S. G., Roe, G. H., and Hakim, G. J.: Climate reconstruction using data assimilation of water isotope ratios from ice cores, Journal of Geophysical Research: Atmospheres, 122, 1545–1568, 2017.

Stenni, B., Masson-Delmotte, V., Selmo, E., Oerter, H., Meyer, H., Röthlisberger, R., Jouzel, J., Cattani, O., Falourd, S., Fischer, H., et al.: The deuterium excess records of EPICA Dome C and Dronning Maud Land ice cores (East Antarctica), Quaternary Science Reviews, 29, 146–159, 2010.

Stuiver, M. and Grootes, P. M.: GISP2 oxygen isotope ratios, Quaternary Research, 53, 277–284, 2000.

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Davies, S. M., Johnsen, S. J., Muscheler, R., Rasmussen, S. O., Röthlisberger, R., et al.: The Greenland ice core chronology 2005, 15–42 ka. Part 2: comparison to other records, Quaternary Science Reviews, 25, 3258–3267, 2006.

Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, Journal of hydrology, 456, 12–29, 2012.

Vinther, B. M., Clausen, H. B., Johnsen, S. J., Rasmussen, S. O., Andersen, K. K., Buchardt, S. L., Dahl-Jensen, D., Seierstad, I. K., Siggaard-Andersen, M.-L., Steffensen, J. P., et al.: A synchronized dating of three Greenland ice cores throughout the Holocene, Journal of Geophysical Research: Atmospheres, 111, 2006.

Vinther, B. M., Clausen, H. B., Fisher, D., Koerner, R., Johnsen, S. J., Andersen, K. K., Dahl-Jensen, D., Rasmussen, S. O., Steffensen, J. P., and Svensson, A.: Synchronizing ice cores from the Renland and Agassiz ice caps to the Greenland Ice Core Chronology, Journal of Geophysical Research: Atmospheres, 113, 2008.

Vinther, B. M., Buchardt, S. L., Clausen, H. B., Dahl-Jensen, D., Johnsen, S. J., Fisher, D., Koerner, R., Raynaud, D., Lipenkov, V., Andersen, K. K., et al.: Holocene thinning of the Greenland ice sheet, Nature, 461, 385, 2009.

Vizcaino, M., Mikolajewicz, U., Ziemen, F., Rodehacke, C. B., Greve, R., and Van Den Broeke, M. R.: Coupled simulations of Greenland Ice Sheet and climate change up to AD 2300, Geophysical Research Letters, 42, 3927–3935, 2015.

Werner, M., Mikolajewicz, U., Heimann, M., and Hoffmann, G.: Borehole versus isotope temperatures on Greenland: Seasonality does matter, Geophysical Research Letters, 27, 723–726, 2000.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Monthly Weather Review, 130, 1913–1924, 2002.

Young, N. E. and Briner, J. P.: Holocene evolution of the western Greenland Ice Sheet: Assessing geophysical ice-sheet models with geological reconstructions of ice-margin change, Quaternary Science Reviews, 114, 1–17, 2015.

Figures

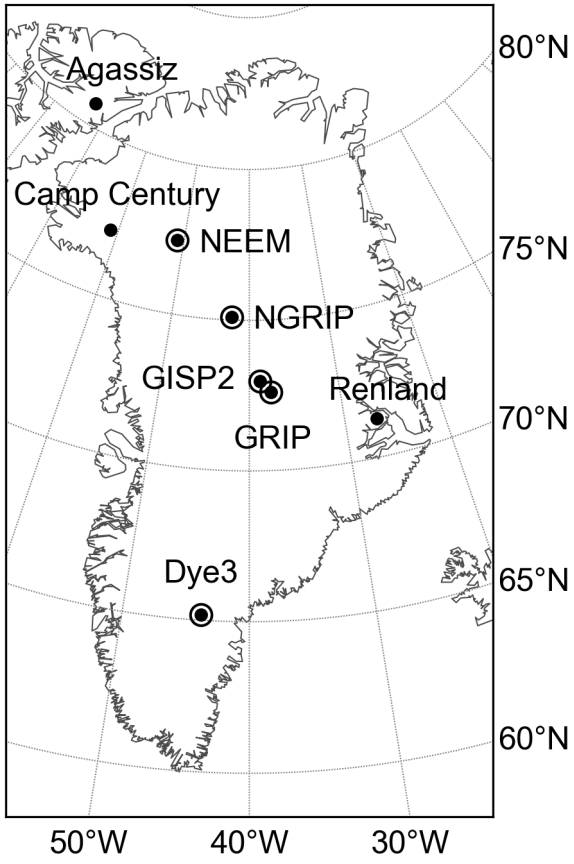

Figure 1. Locations of the ice-core sites referenced in this study. We use oxygen isotope (δ^{18}O) records from all eight sites and accumulation records from the five circled sites.

Table 1. Metadata for the water isotope (δ^{18}O) and accumulation (accum) records referenced in this study. "NBI" refers to the Niels Bohr Institute data access site (http://www.iceandclimate.nbi.ku.dk/data/) and "Pangaea" refers to the Pangaea data access site (https://www.pangaea.de/). Latitude and longitude are in units of decimal degrees (dd) and dates are in thousands of years before 1950 CE (ka).

Ice core name	Latitude (dd)	Longitude (dd)	Variables	Oldest (ka)	Youngest (ka)	Source	Citations
Agassiz	80.7	286.9	δ^{18}O	11.64	−0.02	NBI	1
Camp Century	77.18	298.88	δ^{18}O	11.64	−0.02	NBI	1
NEEM	77.45	308.94	δ^{18}O	>20	−0.0108	NBI	2, 3, 4
			accum	>20	−0.04	NBI	5
NGRIP	75.1	317.7	δ^{18}O	>20	−0.04	NBI	6
			accum	>20	−0.02	this study	7, 8, 9, 10
GISP2	72.97	321.2	δ^{18}O	>20	−0.04	NBI	11, 12
			accum	>20	−0.0375	Pangaea	13
GRIP	72.6	322.4	δ^{18}O	>20	−0.02	NBI	14
			accum	>20	−0.02	this study	7, 8, 9, 10
Renland	71.27	333.27	δ^{18}O	11.64	−0.02	NBI	1
Dye3	65.18	316.18	δ^{18}O	>20	−0.02	NBI	1, 15
			accum	11.640	0	this study	16, 17

[1] Vinther et al. (2009), [2] Dahl-Jensen et al. (2013), [3] Schüpbach et al. (2018), [4] personal comm. Bo Vinther, [5] Rasmussen et al. (2013), [6] Andersen et al. (2004), [7] Vinther et al. (2006), [8] Rasmussen et al. (2006), [9] Andersen et al. (2006), [10] Svensson et al. (2006), [11] Grootes and Stuiver (1997), [12] Stuiver and Grootes (2000), [13] Cuffey and Clow (1997), [14] Johnsen et al. (1997), [15] Dansgaard et al. (1982), [16] Vinther et al. (2009), [17] this study

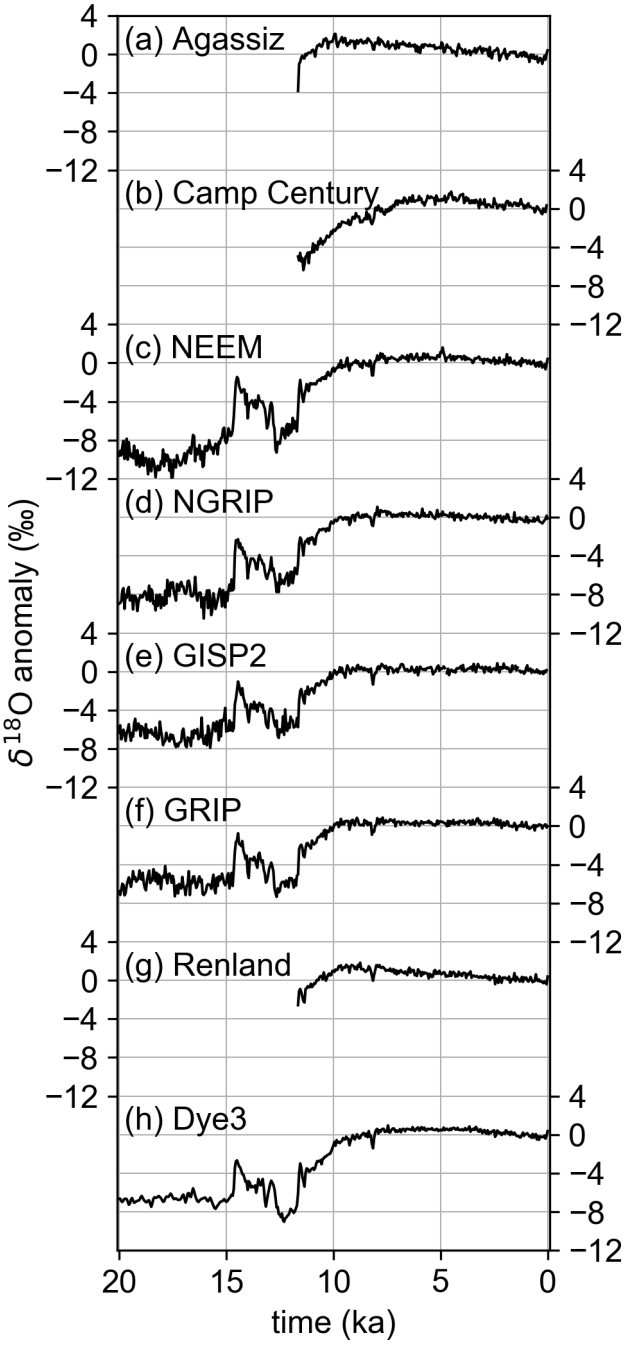

Figure 2. δ^{18}O records assimilated into the temperature reconstruction. Records are shown as anomalies relative to the mean of 1850-2000 CE and are ordered top to bottom from northernmost to southernmost. Ice-core site names are given above each record.

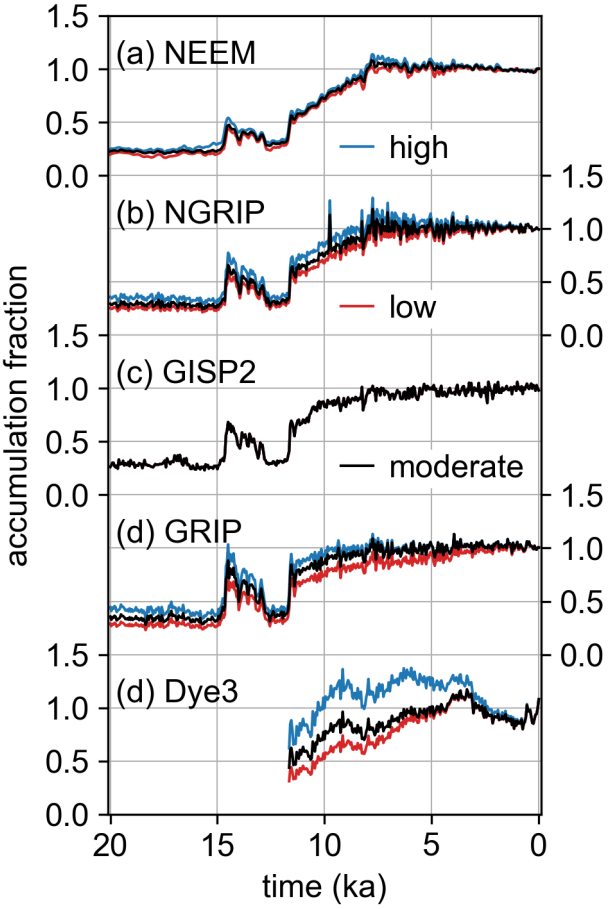

Figure 3. Accumulation records assimilated to reconstruct precipitation for the main reanalysis and two sensitivity scenarios. Records are shown as fractions relative to the mean of 1850-2000 CE and are ordered top to bottom from northernmost to southernmost. Black lines are the moderate records which are included in the main precipitation reanalysis, red lines are the low records which are included in the low sensitivity scenario, and blue lines are the high records which are included in the high sensitivity scenario. Note that we use the same GISP2 accumulation record for the main, high, and low scenarios. Ice-core site names are given above each set of records.

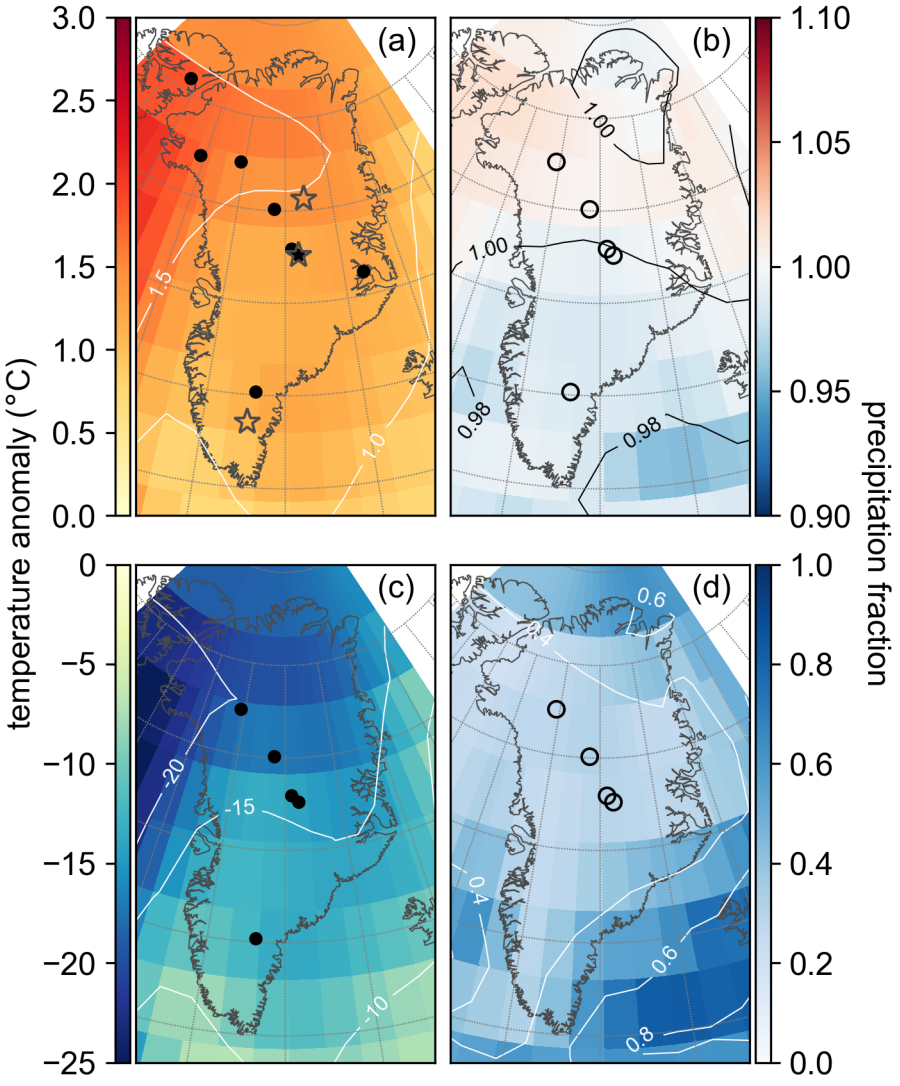

Figure 4. Spatial pattern of the reanalysis mean for temperature (panels (a), (c)) and precipitation (panels (b), (d)). (a) and (b) are averaged over 1,000 years around the peak warmth in the Holocene, 5.5-4.5 ka, while (c) and (d) are averaged over 5,000 years in the late glacial, 20-15 ka. Anomalies and fractions are with respect to the mean of 1850-2000 CE. Points show ice-core locations used for each reanalysis with closed circles indicating δ^{18}O records and open circles indicating accumulation records. Grey stars show the locations of the EGRIP ice-core site, Summit, and South Dome, which are referenced in Figs. 5 and 10.

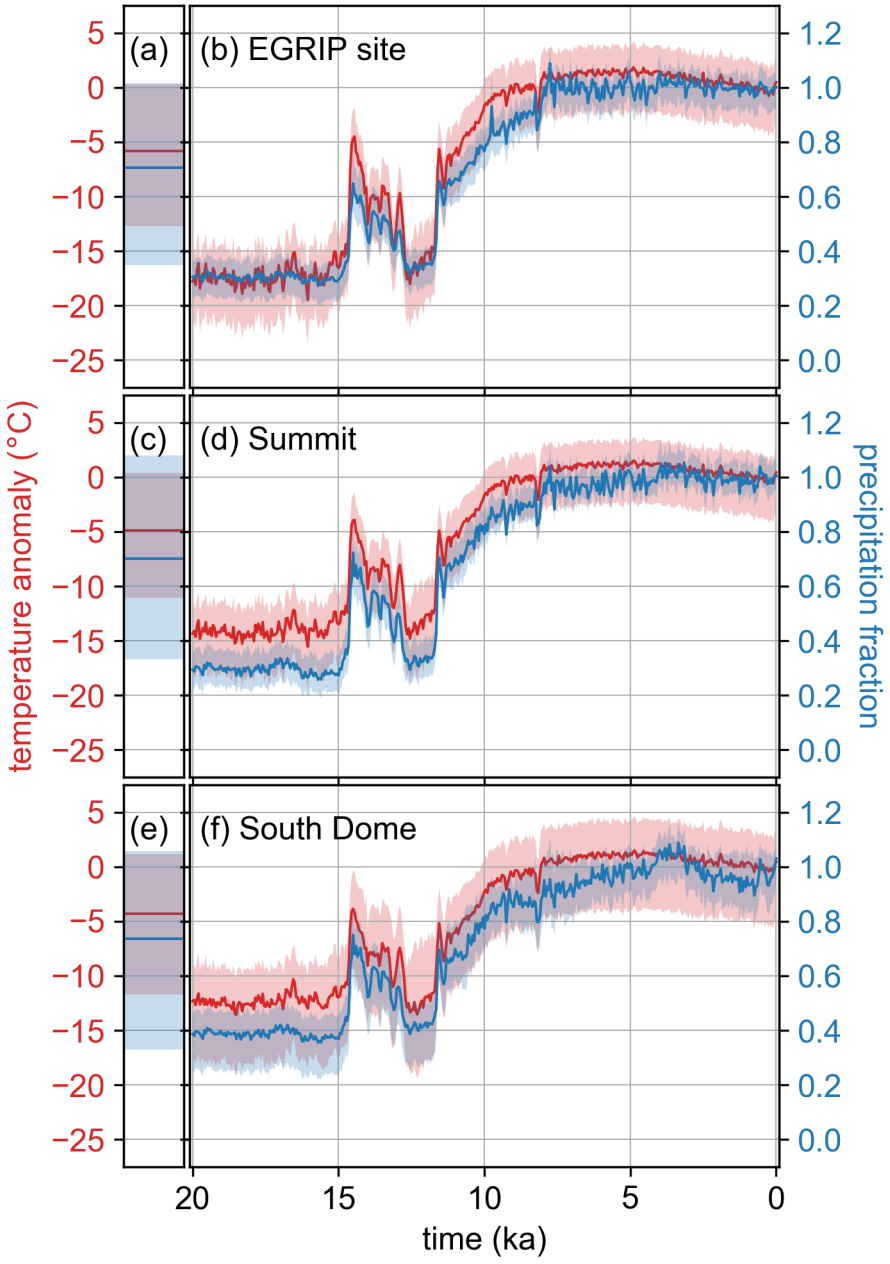

Figure 5. Time series of the prior (panels (a), (c) and (e)) and reanalysis (panels (b), (d) and (f)) ensemble mean and 5^{th} to 95^{th} percentile shading for temperature (red) and precipitation (blue) at three locations. Anomalies and fractions are with respect to the mean of 1850-2000 CE. (a) and (b) show these time series for the location closest to the EGRIP ice-core site, (c) and (d) show for the location closest to Summit, and (e) and (f) show for the location closest to South Dome. These locations are ordered from northernmost (top) to southernmost (bottom) and are shown on a map in Fig. 4.

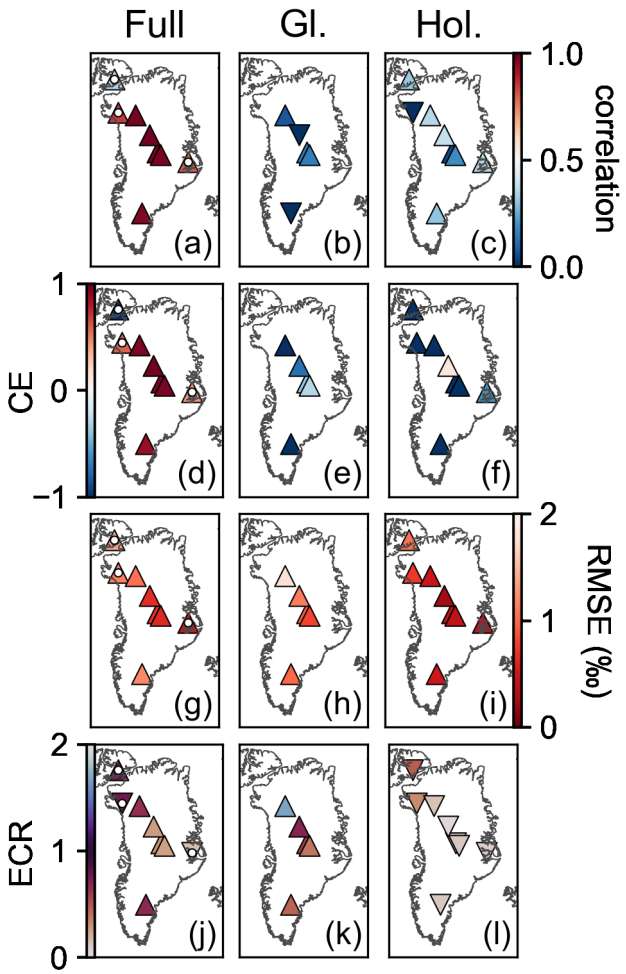

Figure 6. Skill metrics averaged over iterations and time for the temperature reanalysis. The first column (panels (a), (d), (g), and (j)) shows the skill metrics for the full overlap (Full) between the proxy record and reanalysis. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill metrics for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble evaluates better than the prior ensemble for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

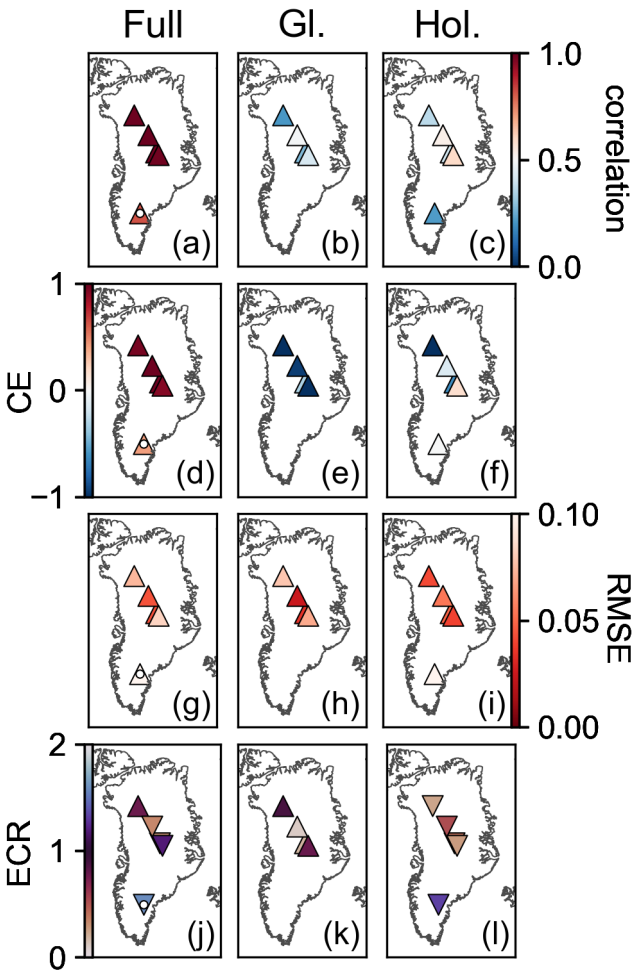

Figure 7. Skill metrics averaged over iterations and time for the precipitation reanalysis. The first column (panels (a), (d), (g), and (j)) shows the skill metrics for the full overlap (Full) between the proxy record and reanalysis. A white dot indicates evaluation against proxy records that overlap only the Holocene (11.7-0 ka). The middle column (panels (b), (e), (h), and (k)) shows the skill metrics for a period in the glacial (Gl.) (20-15 ka), while the right column (panels (c), (f), (i), and (l)) is for a period in the Holocene (Hol.) (8-3 ka). The first row (panels (a)-(c)) reports the correlation coefficient, the second row (panels (d)-(f)) the coefficient of efficiency (CE), the third (panels (g)-(i)) the root mean square error (RMSE), and the fourth row (panels (j)-(l)) the ensemble calibration ratio (ECR). Triangle symbols pointing up indicate that the posterior ensemble evaluates better than the prior ensemble for that location and statistic. Triangle symbols pointing down indicate the opposite. We define better evaluation as correlation coefficient closer to 1, CE closer to 1, RMSE closer to 0, and ECR closer to 1.

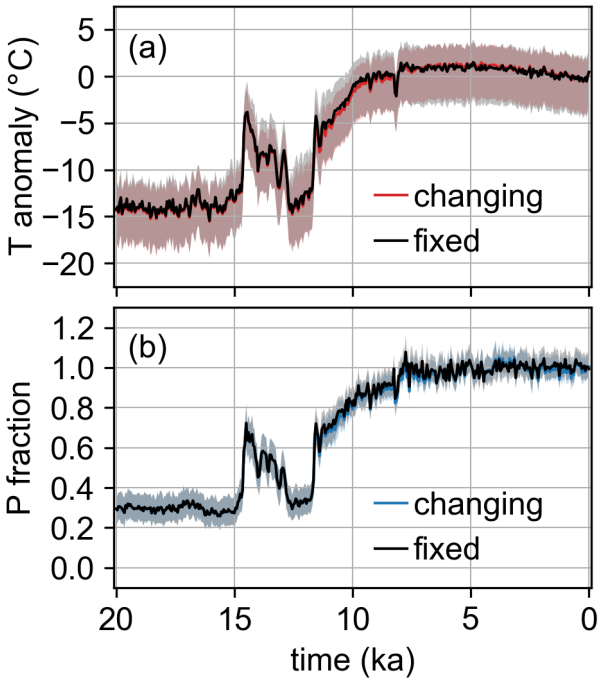

Figure 8. Changing (red and blue) vs. fixed (black) proxy-network for the (a) temperature (T) and (b) precipitation (P) reanalysis mean and 5^{th} to 95^{th} percentile shading. Anomalies and fractions are with respect to the mean of 1850-2000 CE. These time series are for the location closest to Summit, which is representative of the results around Greenland.

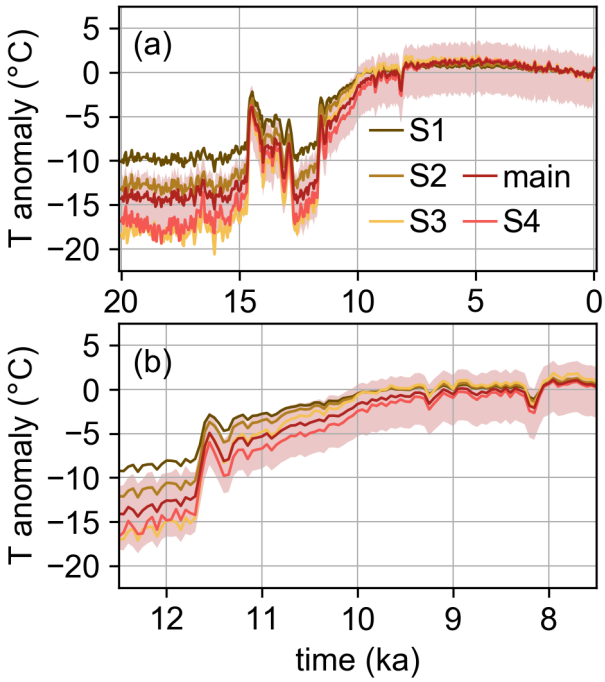

Figure 9. The main temperature (T) reanalysis (ensemble mean and 5^{th} to 95^{th} percentile shading) and ensemble mean for four sensitivity scenarios, S1-S4. Panel (a) shows the full 20,000 year reconstruction, while panel (b) shows a the Younger Dryas to early Holocene period (13 to 7 ka). Each sensitivity scenario reflects a different assumption about precipitation seasonality, with S1-S3 assuming a spatially-uniform seasonality and S3-S4 assuming stronger seasonality than the main reanalysis. Anomalies are with respect to the mean of 1850-2000 CE. These time series are for the location closest to Summit, which is representative of the results around Greenland.

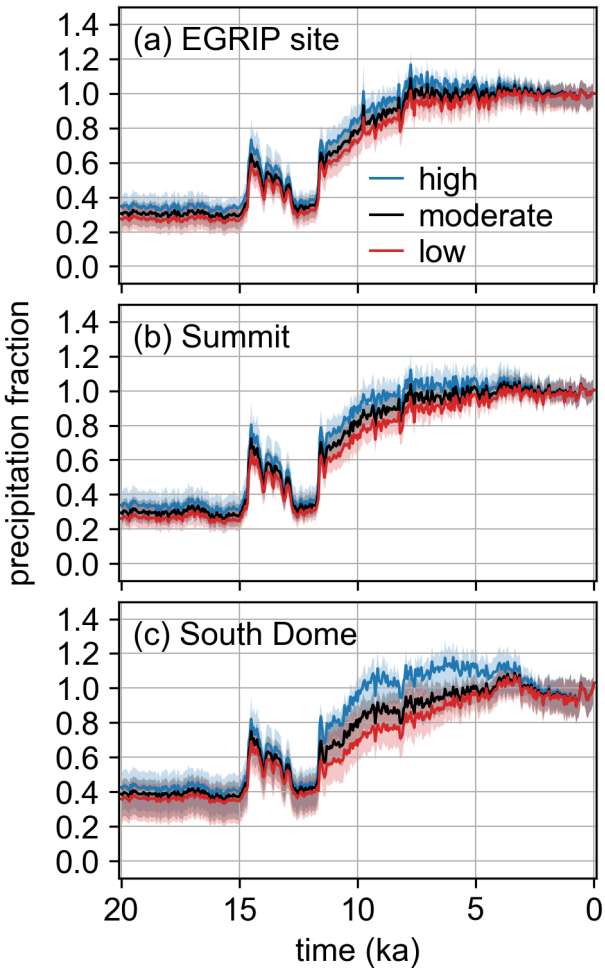

Figure 10. Ensemble mean and 5^{th} to 95^{th} percentile shading for the main precipitation reanalysis (black), high sensitivity scenario (blue), and low sensitivity scenario (red). Fractions are with respect to the mean of 1850-2000 CE. (a) is the time series for the location closest to the EGRIP ice-core site, (b) is closest to Summit, and (c) is closest to South Dome, which are representative of northern, central, and southern Greenland and are shown on a map in Fig. 4.

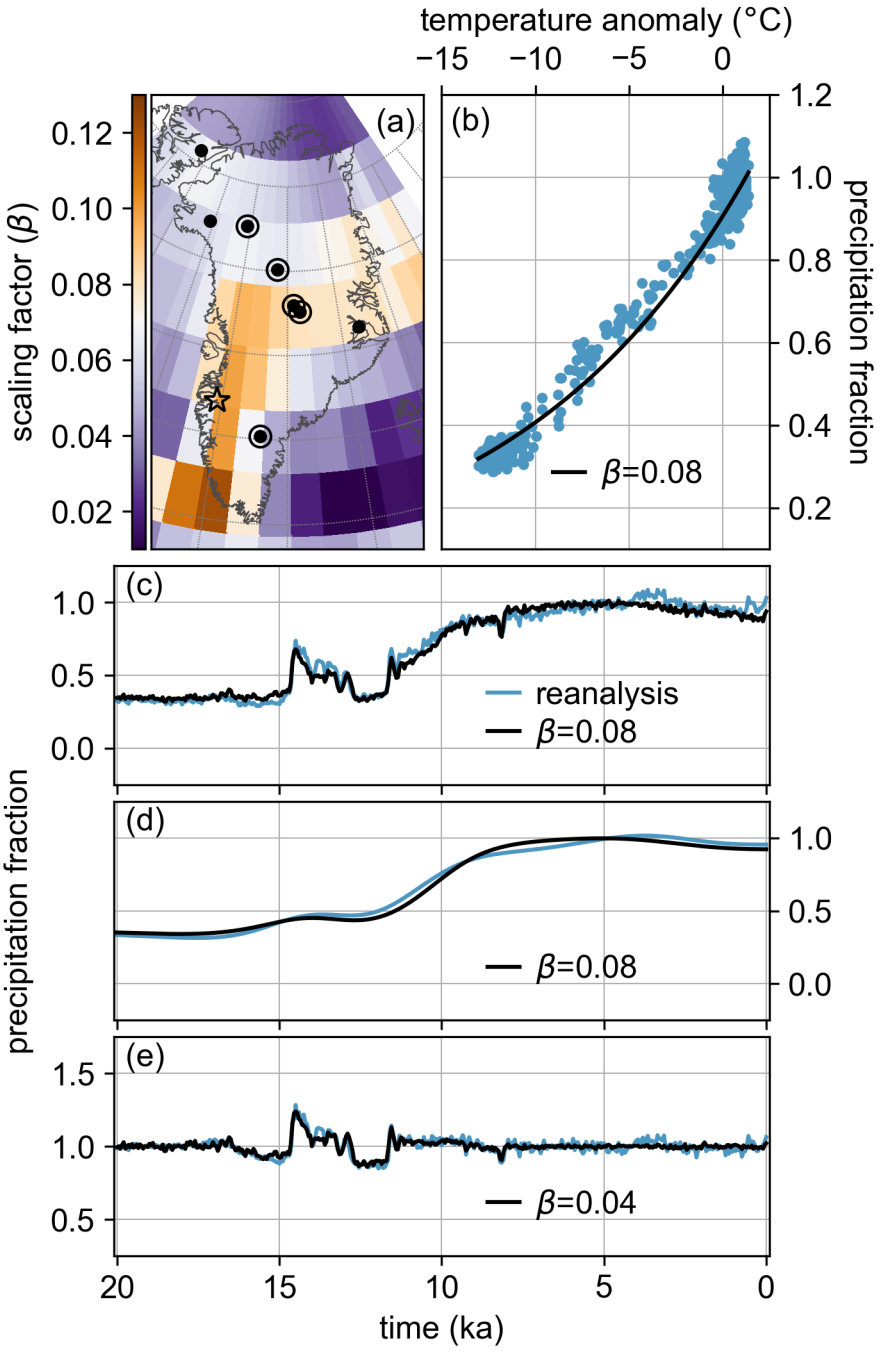

Figure 11. The precipitation-temperature relationship in our reanalysis. (a) shows the spatial pattern of the scaling factor (β) for the best-fit thermodynamic scaling. The colorbar is centered on 0.07, the value used by Greve et al. (2011). Points indicate ice-core locations used for each reanalysis with closed circles indicating δ^{18}O records and open circles indicating accumulation records. The star is at the center of the area used in panels (b)-(e) (65°N to 68.7°N and 48.5°W to 52.5°W). (b) is a scatter plot of temperature anomaly vs. precipitation fraction from the reanalysis (blue points). The black line shows the best-fit exponential scaling. (c) shows the time series of the precipitation reanalysis (blue line) and precipitation scaled from temperature using the best-fit scaling (black line). (d) and (e) are the same as (c) except low-pass and high-pass filtered, respectively, with a cutoff frequency of 5,000 year^{-1}.

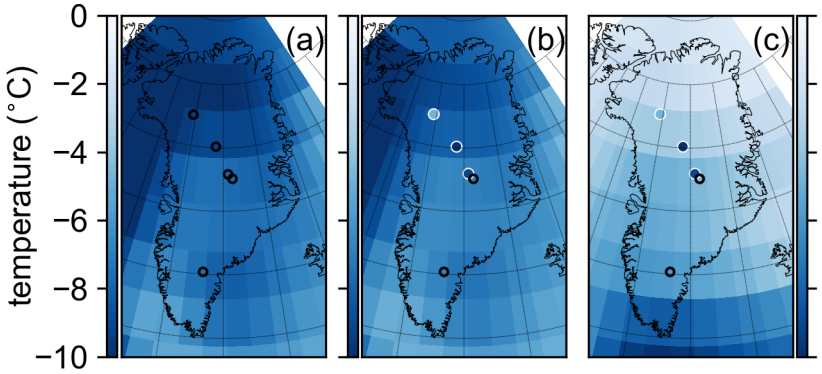

Figure 12. Spatial pattern of the abrupt cooling event into the Younger Dryas. Panel (a) shows results from experiment O8, assimilating all eight δ^{18}O records, panel (b) shows results from experiment N3O5, assimilating all three δ^{15}N-derived temperature records and the remaining five δ^{18}O records (those that do not overlap with the δ^{15}N sites), and panel (c) shows results from experiment N3O5_BA, which is similar to the N3O5 experiment except the prior ensemble is selected from the 1,000 years surrounding the Bølling-Allerød warming. Unfilled black circles show locations of assimilated δ^{18}O records, while filled circles with white outlines show locations of assimilated δ^{15}N-derived temperature records. Filled circles in panels (b) and (c) show the δ^{15}N-derived temperature values as reported by Buizert et al. (2014) on the same color scale as the rest of the panel. The temporal definition of this event is the same as defined in Buizert et al. (2014).

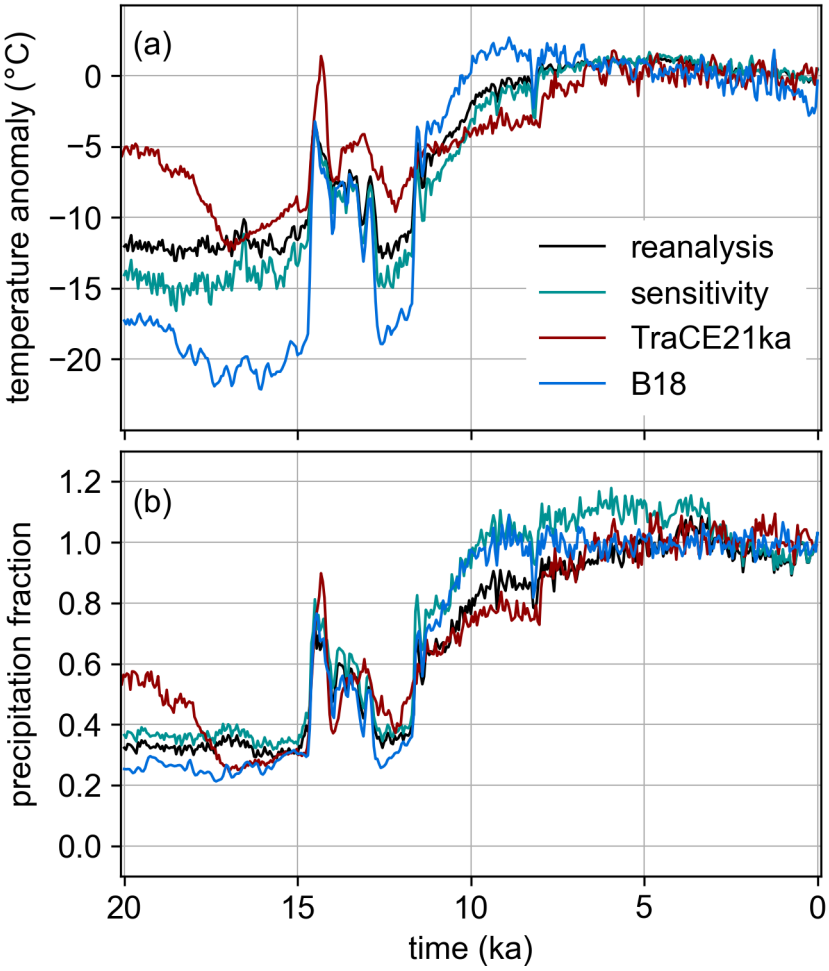

Figure 13. Temperature (a) and precipitation (b) reconstructions from our main reanalysis (black), our sensitivity scenarios S4 and high P (green), ~~TraCE21ka~~ TraCE-21ka (red), and B18 (blue) (Buizert et al., 2018). Each reconstruction is averaged to a 50-year time resolution and averaged over a spatial domain in the Kangerlussuaq region, defined by the latitude-longitude box 65°N to 68.7°N and 48.5°W to 52.5°W, the center of which is located at the star in Fig. 11a. Temperature anomalies and precipitation fraction are defined with reference to the mean of 1850-2000 CE.