Reply to Referee #1. Author comments in blue italics.

Anonymous Referee #1

Received and published: 25 December 2019

The manuscript presents several reconstructions of climate fields in the North Atlantic over the last centuries. The reconstructions are based on proxy data on the one hand (oxygen isotope concentrations in Greenland ice cores and on European dendroclimatological data) and on climate simulations with a isotope enabled climate model. Both types of data are combined applying the analog method. The manuscript is mainly focused on testing the sensitivity of the reconstructions on the number of ice cores, on the definitions of the target seasons. The manuscript is less focused on the physical results and the implications of these reconstructions. It is therefore rather technical manuscript. There are almost no results or dicussion on the physical mechanisms or past climate variations. This is in principle fine, but the readers should be made aware of this early on in the abstract and in the introduction. In my opinion, the manuscript addresses interesting issues. For instance, the use isotopic data from ice cores in combination with climate simulations to reconstruct the atmospheric circulation is an interesting idea. However, being this a more technical manuscript, the description of the methods applied leaves many open questions for the reader. This description should be considerably improved. At some points it is so unclear that I had doubts about the correct application of the method, although I hope that it is in the end correct.

- *We are grateful for the very detailed and helpful comments by the reviewer, which has helped us to produce a much improved revised manuscript. The method used in this manuscript was first published in Sjolte et al. (2018). Therefore some aspects of the description of the methodology is less detailed. This is clearly not the right way to do it and we have worked on clarifying technical details and the motivation of our approach. We also agree the this is mainly a method paper. This will be stated clearly in the revised manuscript.*

Main points:

1. One main concern is the limited methodological description. The authors apply the analog method after a pre-filtering by Principal Component Analysis, but it is not cleat how this pre-filtering is actually conducted. Important questions that may impact the results :

1.1 Are the PCs calculated from the covariance of correlation matrix?

- *The PCs are calculated from the covariance matrix. This is specified in the revised manuscript*

1.2 in Equation 1, are the PCS normalized to unit variance or has each PC the platitude representing the corresponding explained variance. This is important because this point represents two options: all PCs are equally weighted for the calculation of the distance of the analogues, or each PCs is weighted according to the varianve it represents. The selected analogues are different depending on the option chosen.

- *The PCs are normalized to unit variance. This is specified in the revised manuscript. The approach of fitting the PCs is motivated by that we want to capture as much of the regional signal from ice cores as possible, but not overfit noise at individual sites. Fitting normalized PCs allows us to capture the 'average' signal at all the ice core sites, while constraining this signal with the regional variability between the cores.*

1.3 More importantly, for a correct application of the method, the EOFs patterns (derived from ice-core records and from model grid cells) associated to each PC in equation 1 must be the same . Otherwise, the PC-coordinates PC_{model} and PC_{icecores}) would not be linked to the same EOFs and therefore would represent coordinates in different subspaces. The calculated distances would not be meaningful. It may happen that the patterns of EOF_{model} and EOF_{icecores} are very similar, in which case this problem would be minor. But this is not guaranteed. A way to ensure that the PCs in equation 1 refer to the same EOF patterns is to use, for instance, EOF_{model} for both data sets and calculate PC_{icecores} by projecting the ice core anomalies onto EOF_{model} . (Or viceversa, use EOF_{icecores} for both data sets). This is the point that most strongly worries me. If EOF_{model} and EOF_{icecore} patterns are really different the whole application of the method is not correct, and all results should be re-calculated.

- *The EOFs of the model output and ice core data are indeed very similar. The map in Figure 1 is to some extent showing this, although it can be hard to determine how close from a map. As shown in Sjolte et al. (2018) we actually use an additional step to test if the selected model analogues fit the original ice core data (so, not the PCs), where we extract the reconstructed ensemble mean model d18O and correlate it to the ice core data. This step tests if the matching of the PCs works. We include a description of this step of the reconstruction method in the revision, and illustrate it with a figure similar to Figure S1 from Sjolte et al. (2018), but for all seasons (see below).*

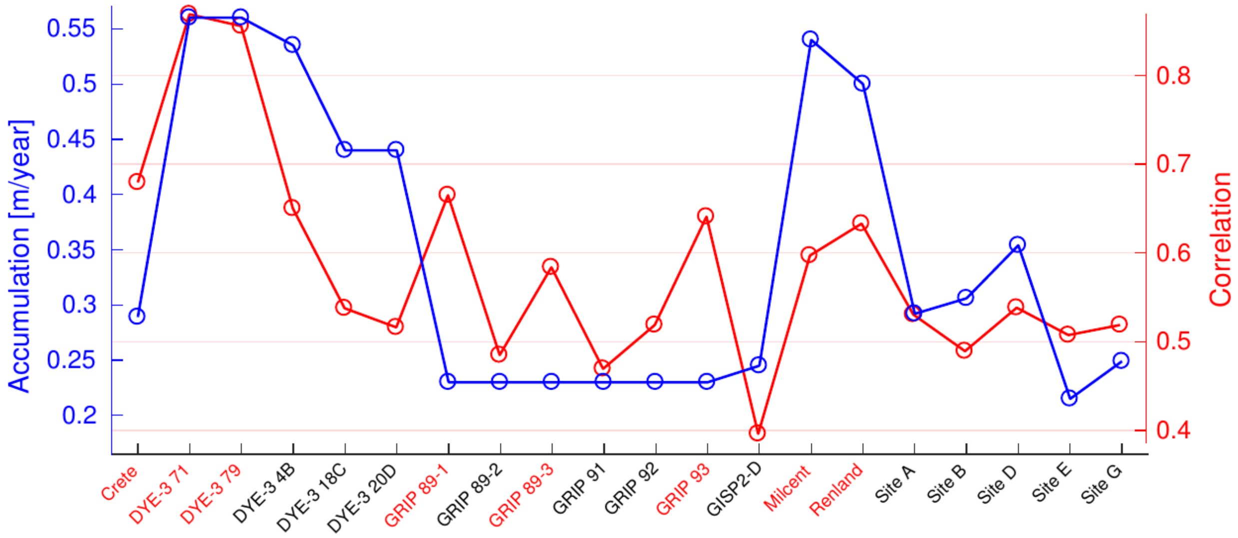

Figure S1. Accumulation at ice core sites (blue) and correlation between reconstructed winter δ^{18}O and ice core winter δ^{18}O (Vinther et al., 2010) 1778-1970 (red). The names of the ice core sites used in the reconstruction are written in red.

2. Another unclear methodological point is how the dendro data are included for the reconstruction. Here, I cannot make any useful suggestion because the authors approach remains unclear to me. This needs to be much better explained:
2.1 Are the analogues searched using ice-core and dendro data simultaneously, i.e. a 12-month-long model analog have to be close to the icecore data in the target season and close to the dendro-reconstructed summer temperatures in the summer season?. If yes, how is the EOF filtering implemented here? How many ’temperature’ EOFs are used.
or
2.2 Do the selected analogues (using ice core data) undergo a secondary selection procedure targeting the dendro-reconstructed temperatures ?

- *It is a secondary selection as stated in L145 “we sort the 39 existing ensemble members based on the ice core selection”. The test with the tree-ring data is to see if we can constrain* *the temperature further. This is not meant as a final reconstruction, but as a test of the common signal between Greenland and European proxy data. As the results show there is quite some common signal for ~20 of the analogues out of the 39 pre-selected analogues based on ice core data. The purpose and technical details connected to the use of tree-ring data will be clarified both in terms of methods and motivation.*

2.3 In both cases, are the distances to the dendro-data and the distances to the ice core data equally weighted ? How is this implemented if the number of EOFs for each type of data sets is presumably not the same.

- *See reply above.*

3. The validation of the results is essentially made by calculating the correlation between reconstructions and 20CR reanalysis. However, the amplitude of the reconstructions is not validated. This may be important because the amplitude of reconstructed variability may depend on the number of analogs selected: best-analog-selection (just one analogue) will roughly produce the same amplitude of variations, although the validation correlation will be lower; in contrast, using the mean of a larger number of analogues subdues the variability, and this effect can be substantial when using 39 (?) analogues. There is an unavoidable trade-off between better correlations and more realistic amplitudes, as shown in Gomez-Navarro et al. Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era, Clim. Past, 13, 629–648, https://doi.org/10.5194/cp-13-629-2017, 2017).

- *This point is discussed by Sjolte et al. (2018) in terms of the amplitude of d18O for the method used here. There is a trade off, and some of the amplitude of the signal is smoothed out when using the ensemble approach. Comparing to the d18O amplitude in the ice cores data this is a minor effect, and some of the amplitude in the ice core data is in fact depositional noise, which we don't want to fit the model data to.*
- *We do show the amplitude compared to SST data, where it is quite realistic for winter and annual variability. The underestimation of the SST amplitude for summer has more to do with the model biases and seasonal climate/proxy differences already discussed in the manuscript.*
- *The above points will be included in the revised version.*

Particular points:

4. We test a range of climate reconstructs varying the definition of the seasons
climate reconstructions

- *Corrected*

5. The abstract does not mention the reconstruction method at all, despite the manuscript being essentially methodological in nature.

- *Corrected*

6. best captured when defining the season December-February
the season as December-February

- *Corrected*

7. line 10 best captured when defining the season December-February due to the dominance of large scale patterns, while for summer the weaker, albeit more strongly auto-correlated, variability is better captured using a longer season of May-
This sentence becomes clearer later in the manuscript. Here, I would suggest to improve its clarity, for instance, using 'more persistent in time' instead of autocorrelated.

- *Corrected*

8. One point that sets the study by Sjolte et al. (2018) apart from the other studies mentioned in this section, is the use of
delete comma after section

- *Corrected*

9. line 145 For the summer reconstructions also using tree-ring data we sort the 39 existing ensemble members
I am rather confused by this sentence. The number 39 is mentioned here for the first time, if I am not mistaken. What are these ensemble members? are they the analogs previously selected targeting the ice-core data ?

- *See reply to major point 2.2*

10. line 155 In this study we follow the convention of using the term PCs for the time series of the main modes of variability, while using the term EOFs for the spatial patterns of the modes. The method of Ebisuzaki (1997) is used to calculate the significance when
this sentence should appear before equation 1, at the very least

- *Corrected*

11. line 163 A key factor in how well seasonal climate reconstructions can represent climate itself, is the auto-correlation structure of atmospheric
climate itself ? I guess the authors mean to what extent can seasonal proxy data represent annual means ?

- *Corrected. It should have been "A key factor in how well seasonal proxy data can represent climate variability … " this applies to both seasonal and annual variability.*

12. line 167 Figure 2 shows the monthly auto-correlation of each month of the PC-based NAO calculated from the 20CR. These figures show that during the cold season the
Which is the PC that represents the NAO ? Here, it is assumed that, for each month the leading PC is the NAO. In summer this is not always the case, and it depends on the geographical region selected to conduct the PCA.

- *We checked the patterns and the leading PC is NAO for all months. However, as we write there is little consistency between the months of the secondary PCs.*

13. line 180
circulation modes. We do this by performing monthly reconstructions for pressure and evaluating the resulting main modes of circulation against the modes of the 20CR. This is done using the same method as for the seasonal reconstructions,
what does 'evaluating the modes' mean here ?. The spatial patterns (EOFs) resulting from an analog reconstructions can be very realistic irrespective of the skill of the analog method. The analog method is just a resampling from a data pool. A random resampling of SLP from the 20CR reanalysis or from a model run will produce the same EOFs as the orginal data set, so even if the analog is wrongly implemented, the resulting EOFs may look correct. This is different for a temporal validation, e.g. correlation between reconstructions and observations, where the skill of the analog selection is critical. The authors should be here more specific.

- *The resampling changes the major modes, so the modes do not simply come from the model, and the modes are also reshuffled when the variability is resampled. We write this in L273-276 of the original manuscript. We have further clarified this section in the revision.*