Response to the reviewers

We thank both reviewers for their constructive comments. Please find below detailed answers.

Anonymous Referee #1

1. The authors should also indicate how the data will be made available after manuscript has been published.

After publication the data will be available from the data base PANGAEA http://dx.doi.org/10.1594/PANGAEA.849161.

2. Section 2: Given that two co-authors were also co-author on the recent paper presenting the stacked ice core record from the NEEM shallow ice paper (Masson-Delmotte et al. 2015) I find it odd not to include this nearby isotope record in this study. Given the high signal-to-noise ratio of that stacked core it might make sense to compare the individual NGT cores with this ice core.

We decided not to include the NEEM data, because of the covered time period. The NEEM (shallow) record covers the most recent (300) years. Our focus is on the last millennium. The core B26 was drilled only view kilometers from the NEEM site and covers a much longer time period.
However, of course we include the NEEM record in the discussion of our results (e.g. in the section "3.3 The northern Greenland δ18O-stack and its paleoclimatic significance")

3. Page 2346 L 15-19: It is unclear to me how the authors have taken into account layer thinning when assigning the depth-age model. Please explain in details the model used.

Layer thinning was not taken into account as the cores were drilled only in the upper part of the ice sheet. We add this information to the manuscript.
For the ice core dating we only picked the depth of the time-marker horizons. Between two horizons we assume the accumulation rate to be constant.
Compared to densification thinning effect is of second order at the top 100 m.
E.g. if the Greenlandic ice sheet is about 3000 m 100 m is about one thirtieth. In about 1000 years the mistake in the accumulation rate is about 2 % which means that we have 102 instead of 100 mm.

4. P2347 L3: What is the dating uncertainty for the period between 1100-**1**500 years? Later (p. 2349) the authors decide apply a running mean of 5 years – it is unclear to me why 5 years were chosen and not 10 years given that the uncertainty is estimated to be 10 years. I suggest that the authors discuss and quantify the effect of this dating uncertainty on the correlation coefficients discussed on pager 2349. Please consider to use a 11-year running mean to filter out the solar cycle instead.

There is no error for our Depth-Age Model (well dated volcanic horizons). But there is a maximum of 10 years difference between annual layer counted NGT records and our dating assumption (See table 2).
We will show 11-years running means instead of 5 years to better take into account the potential dating uncertainty.
For 11-year running means about 50% of the records have correlation coefficients higher 0.3 (5-year running mean 25%).

5. P. 2347 L15-24: Please explain the model used to account for thinning.

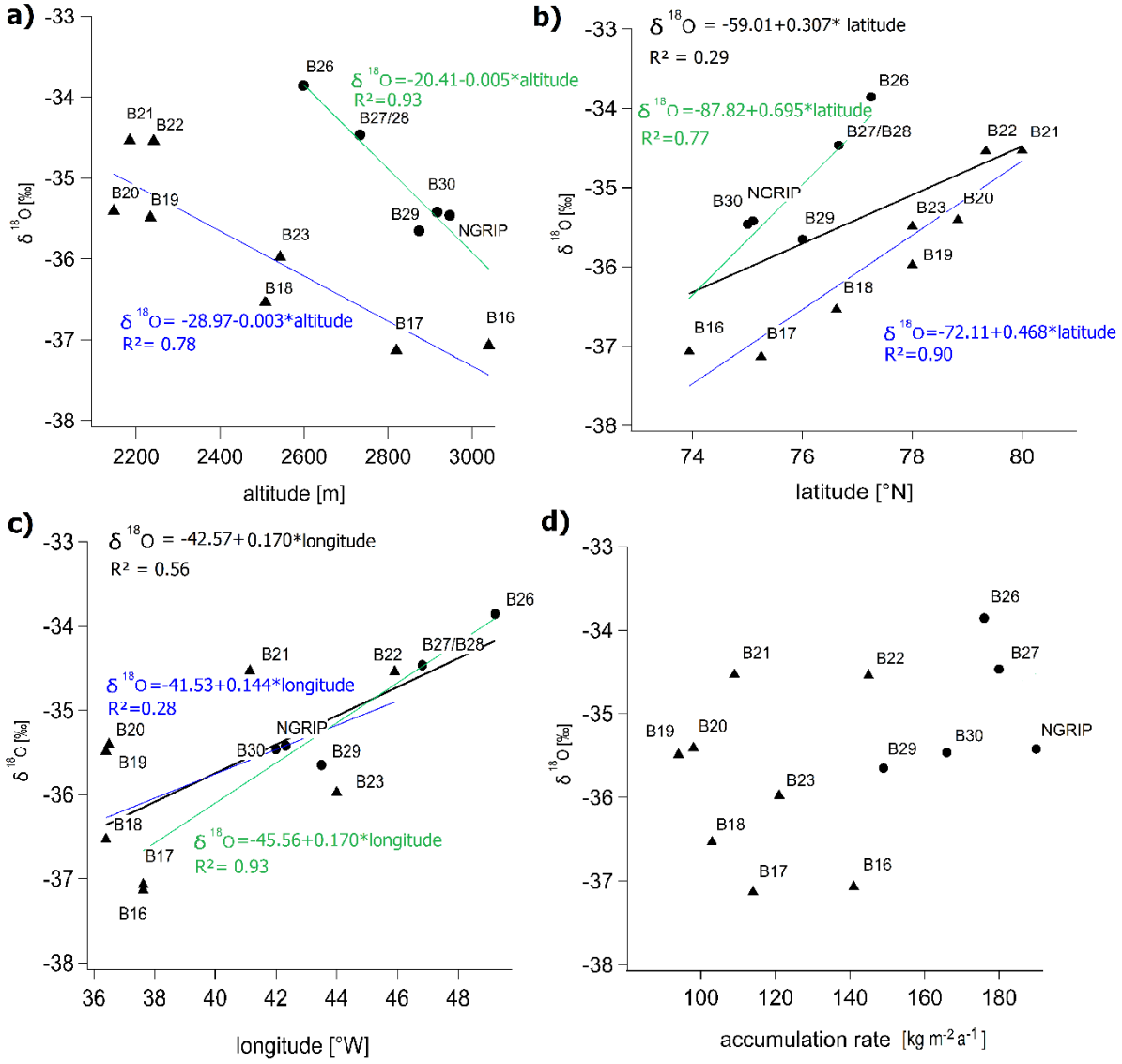
Thinning was not taken into account. Please see point 3

6. P. 2348 L3-6: Given that you are using mean annual isotope values from ice cores with very low accumulation you will need to argue why diffusion does not alter the mean annual isotopic value. I would suggest that the best approach to do this would be to calculate the maximum diffusion length found at each site and compare this with the mean annual layer thickness.

Following Johnsen et al. 2000 the maximal diffusion length is 9 cm, which is slightly smaller than the annual layer thickness. Diffusion might be affecting the absolute difference in isotope content of neighboring years but the mean over 11 or 30 years will not be affected. We add this information to the manuscript in the end of section "3.1 Depth-age models and snow accumulation rates".

7. P2349 L3-11: You need to remove the linear regressions from the figures, which are not statistical significant. Showing these fits does not make sense if they are not statistical significant. I suggest that you carry out a multivariable linear regression as well. I don't believe that showing the linear regression between the isotopic value and the longitude reveal any useful information. Instead I suggest that you use a relevant metric instead of the longitude. For example you could use the distance between the ice core site and the ice divide as a relevant metric, which is more physical appropriate than the longitude.

We agree to remove the insignificant regression lines. Instead we add statistically significant fits for group I ("east") and group II ("divide") to show the differences on and east of the main ice divide. We think it is necessary to show all these plots to make clear, that it is not an easy situation in northern Greenland to see the relationships because of the canceling out of different effects. If you go northward (lighter d18O values) you also go downward (heavier d18O values). Multilinear regression was performed as well and the results are discussed in the manuscript.



8. P 2350 L5-9: I'm pretty sure that the pattern, which you show in Figure 5 is called the EOF while the temporal component of the EOF/PC analysis is called the PC. Under all circumstances you should give error bars on the to show that the first two EOF/PC components are independent of each other. Secondly you should also show the temporal component of the first 2 PC's

Because we do not see any extra information and to shorten the manuscript for better readability we will remove the figure showing the EOFs and most parts in the text dealing with the PCA. At the end

of the section "Regional variability of δ18O in northern Greenland" we only say that there are two EOFs above the noise level while EOF1 has homogenous and EOF 2 and bipolar (east-west) pattern.

9. P2351 L3-6: As argued above it would make sense to also compare the your NG-stack with the NEEM stack.

We found a strong correlation of the NEEM-stack to our NG-stack (r = 0.8 for 30-year running mean). And added this to the discussion in the manuscript.

10. P2352 L1-14: What is the argument for not using the isotope-temperature relationship derived by Masson-Delmotte et al. 2015 for present day conditions, but instead use isotope-temperature relationship for glacial-interglacial conditions?

We do not use gradients derived from deep borehole temperatures at one site (glacial-interglacial conditions) the gradient we use (0.67) was derived from special isotope-temperature surface relationship studies for present day conditions which therefore can be used for our Holocene dataset. If using the NEEM gradient of 1.1±0.2 ‰ per °C the temperature range is: -1.3°C to 2.28 °C which is slightly smaller than the results with the Dansgaard/Johnsen gradient. The gradient derived by Masson-Delmotte et al. 2015 can provide a more correct gradient in time while the gradient, which was valid for 2007-1979 AD, is also usable for other periods of strong warming, but not in space.
We will show these results additionally to the other calculated results in section "3.3 The northern Greenland δ18O-stack and its paleoclimatic significance".

11. P2355 L6-12: Please consider the fact that for the cores located to the east you likely have higher fraction of winter accumulation compared to summer accumulation. This likely means that variations in amount of summer accumulation would increase the noise, and thereby decrease the inter-core correlation. You could also consider using an 11-year running mean, as you have indicated earlier that uncertainties for some part of the cores were 10 years.

Yes, a changing seasonal distribution can be an additional source of noise. We therefore focus the interpretation on the running mean data. However, it is not possible to assess the seasonal accumulation from our data.

12. P2356 L 3-8: You need to support your statement about the spatial pattern of temporal variability with some kind of analysis.

We calculated the dominant frequency in 11-year running mean records of the individual cores. B18-B21 but also B29 show longer periods (117-248 a) than B16-B17 and B22-B30 (besides B29) (81-39 a)

13. P2356 L24-27: It is not clear to me how you based on Figure 4D, which does not show a significant correlation between d18O and accumulation can draw the conclusion that your data are consistent with the hypothesis that the foehn effect causes an anti-correlation between d18O and the accumulation rate.

We will remove the regression line between accumulation and d18O in figure 4 and also the discussion about foehn effect in the manuscript.

14. P 2357 L 12-21: You need to explain the model parameters, which you tune and how you set up the model to simulate your data. Otherwise I suggest removing this section.

We removed most parts of this section. We will only give the information that multilinear regression is necessary to clarify the correlation to the d18O values because due to the topography there are several processes that are cancel out each other in northern Greenland e.g. the effect of altitude and latitude.

15. P2357 L28-P2358 L2: It is hardly surprising that you have a high correlation between PC1 and the NG stack as you use the same dataset to produce both records. You might also want to clarify that you refer to the temporal component of EOF/PC1 and not the spatial component shown in the figure. As mentioned above I strongly suggest that you show both the spatial and temporal component of the EOF/PC analysis.

Because we shortened the PCA part in the manuscript e.g. removed the EOF plot and cannot attribute the PCs to any forcing factor we will not show the individual PCs.

16. P2358 L1-2: I understand where your 22% comes from – however I do not think that

it is correct to claim that only 22% of the NG-stack variability is caused by a regional climate signal. One could make the claim that by making the NG-stack you smooth out most of the influence from deposition noise hence all the variability in the NG-stack is caused by climate one way or the other. When reading your sentence I am left with the answer 'what is the remaining 78% variability in the NG stack?'

See point 8.

17. P2358 L11-13: This is a circular argument – you calculate the PC's based on the same cores, which make up the stack.

We removed this argument.

18. P2359 L 1-4: What is the correlation for SON and DJF? What are the p-values?

The r values for SON and DJF are smaller (see table 5). DJF is only significant ($p<0.05$) for Illulisat and the merged southern record. For SON we calculate r = 0.5-0.31. We added the information to the text of the manuscript.

19. P2359 L10-13: Please quantify the similarities with the stack of Masson-Delmotte et al. 2015.

We find similarities in the identification of the depleted or most enriched events as described p.2359. We correlate both records and find a strong and statistically significant correlation with r= 0.54 (annual values) and r= 0.82 (30 year running mean). We added this information in the discussion part.

20. P2359 L22-23: Wouldn't it be possible to reach this conclusion base on the PC1 and PC2 instead of having to separate the cores into subjectively chosen groups?

See point 8.

21. P2361 L10-13: To support this hypothesis you might want to consider comparing your Stack I (the cores to the east of the divide) with the LOMO-core, as you have argued for these cores to be dominated by a winter precipitation signal.

The correlation for 30 years running means between the Lomo-core and the stack I ("east") shows a correlation coefficient of r = 0.2 whereas stack "divide" has negative correlation (r = -0.12). This can be a first hint that there is more winter snow in the northeastern part of northern Greenland.
For the other arctic records we also compare to the stacks "east" and "divide".
Dye3 and AN show comparable r values as for the NG-stack, the correlation between Agassiz core and stack "east" (r= 0.53) is slightly better than stack "divide" (r= 0.47). However the NG-stack has strongest correlation at each site. Only for Lomo we found differences in the strength of correlation. We add this information to the manuscript in section 3.4.

22. P2363 L1-17: The analysis of the response from volcanic eruptions on the isotopic signal presented in this section need to be supported by a statistical analysis otherwise I suggest to remove this section.

We removed this part for a clearer structure and to shorten the length of the paper.

23. P2365 L13-19: to support your hypothesis that the year 1420-event is related to sea ice decline you will need to argue why you do not have a positive anomaly in the d18O record for year _1590 and year _1700 where sea ice is at a significantly lower level than at 1420.

We change the wording in the text to weaken the statement.
As the shown sea-ice record is an arctic-wide reconstructed record and the used data older than in-situ observations, the interpretation has to be seen with the necessary respect.

24. P2366 L23-25: This sentence needs to be restructured. You need to point out that you are referring to mean annual d18O values (I presume this is the case). You need to point out that you mean that 12% of the spatial variability is due to ice sheet topography. As argued above it is hard to understand how 78% of a stacked ice core record can be due to random noise not related to the climate. As shown by White et al. 1997 a significant part of a single core compare to a stacked record is noise – however they estimate _40% common variability between the individual cores and the stacked record. It does seem that you are mixing temporal and spatial variability together here. Please correct appropriately.

We make clearer that we are discussing spatial variability. However, we shortened the length of the PCA section drastically to shorten the papers length and to achieve a better readability.

Minor revision

P2343 L15: Add reference Merlivat and Jouzel 1979 **-> done**
L16: Add reference Jouzel and Merlivat 1984**-> done**
L21: Add reference Fisher et al. 1985**-> done**
L22: Add reference Pinzer et al. 2012, Steen-Larsen et al. 2014, Johnsen et al. 2000**-> done**
L24: Add reference White et al. 1997, Masson-Delmotte et al. 2015**-> done**
P2345 L21: Please provide detailed information about which cores and which part of the core were samples with what resolution.
Samples were cut with 1-5 cm depth-resolution as given in the manuscript. Most of the ice was sampled with 2-2.5 cm resolution. Only at the upper most parts of the core samples were cut with lower depth-resolution (up to 5 cm). For some meters of special interest a resolution of 1 cm was used. We added this more detailed information to the manuscript.

P2348 L14: Please explain that the calculated SD is the standard deviation of the mean annual isotope values (I presume this is the case) **-> done**

P 2349 L17-20: You might want to compare the inter-core correlation coefficients with those derived by White et al. 1997 and Masson-Delmotte et al. 2015
The cores in the surroundings of the deep cores at NEEM and GRIP had been drilled in closer distance to each other than the NGT cores. Therefore, they find higher correlations with r~0.54 (Delmotte et al. 2015) and r= 0.41-0.55 (White et al. 1997).
Added the given information to the manuscript.

P 2357 L 15: Change enormous to significant**-> done**
P 2358 L 28: may be - > maybe**-> done**
P2364 L 11: Change ‘the temperature records’ to ‘the isotope records’**-> done**

Anonymous Referee #2

1. As already mentioned the paper suffers from a poor structure and as a consequence it is far too long. There is a lot of unnecessary repetition. One way of avoiding this would be to combine “Results and Discussion”. I also suggest removing Fig 6 since it is included in Fig 8. I also wonder if it is necessary to show all the plots? For instance, can Fig 7 be left out?

We add figure 8 (sub-stacks) to 6 and left out figure 7 (residuals).
We shortened the paper length by combining “discussion” and “results to avoid unnecessary repetitions.

2. I lack meteorological information which is the fundamental background for interpreting d18O so a section on this would be an important part of improving this paper. Some of this information is included in other parts of the paper (mainly in the discussion part) but I think that the paper would benefit from having it all collected in one section in the beginning.

We add information in the introduction to a give short overview about the most important meteorological facts for the d18O values in northern Greenland. Furthermore we add the results of 15m firn temperature measurements to table 4.

3. I also lack a proper “background and previous work”.

We add information on previous d18O NGT work and describe the scientific activity in northern Greenland in the introduction

For instance there are several comments in the paper to how northern Greenland is different from Southern Greenland but few (if any?) specific examples. Please include some of these with proper references.

In the introduction part we already described that northern Greenland differs significantly from the south in terms of lower air temperatures and lower snow accumulation rates. We describe the meteorological characteristics of northern Greenland now more in detail.

For people not working with Greenland even terms like “southern Greenland” and “northern Greenland” need to be specified.

We separate the Greenlandic ice sheet at summit in north and south and will specify that now in the text.

4. My main concern with this paper is that the accumulation records are not given the space and consideration for the results that I think it should have. I understand that some of the accumulation data has been published before but I think that these data has to be viewed in connection with the d18O data.

For results on the mean accumulation rate data see Weißbach et al. 2015
DOI 10.1007/978-3-319-13865-7_21
The accumulation rate at the different drill sites was over the last 1000 years rather constant. We observe no stat. significant correlation between the mean accumulation rate and the d18O values (see figure 4). On annual resolution we would expect that the accumulation rates are following the d18O characteristics.

Some of the questions that come to my mind that I do not find an answer for in the paper are: - What is the seasonal distribution of snowfall in Greenland?

With the given data we cannot answer this question. Due to the dating assumptions and sampling resolution we do not have seasonal resolution.

In Table 5 the seasonal impact of temperature on the d18O is investigated but the seasonal distribution of accumulation is not really discussed anywhere. - What does the long term (stacked) accumulation record look like for this area? Maybe some insight could be revealed from combing this with the d18O record for instance in Fig 8. I also wonder what the accumulation records from these cores look like over time?

Because of the dating assumptions we do not have “real” annual resolution. For that reason we do not focus on the annual accumulation rates or their influence to $\delta^{18}O$.

Is there was any GPR data collected in connection with the cores that would indeed provide information to the questions asked; how representative single ice cores can be for a specific area. I

suggest to look at the literature from similar studies done in Antarctica where there are exactly the same concerns with low accumulation as in Northern Greenland.

GPR data is a totally different topic. Because we do not focus on the accumulation rates we decided not to include this data to our paper.
There was a traverse from NGRIP to NEEM with GPR. The results are given in the Diploma thesis of Radzic 2008 (unpublished).

5. The paper claims to provide a "new climate record" but it is not so clear what if there was a previous climate record- and if it was what it looked like. Like many things in this paper it might be buried in the text but it is not easy to find.

There is no "old" stacked record for northern Greenland. Only single NGT records are published (Schwager 2000, Fischer et al. 1998). The collection of NGT cores and the generation of the stack(s) offer us for the first time the possibility to have a representative stack for northern Greenland. Before, there were stacks for smaller areas at NEEM e.g. Masson-Delmotte et al. 2015 or Summit e.g. White et al.1997.

6. Results absolutely need some kind of error analysis both on the dating and the stacked record.

There is no dating error due to our dating assumptions. If there was a prior dating with annual layer counting we give the maximum difference between both dating methods in table 2.

7. Regarding about the AMO impact in Greenland I suggest looking at this paper in addition to what is presented: Chylek, P., C. Folland, L. Frankcombe, H. Dijkstra, G. Lesins, and M. Dubey (2012), Greenland ice core evidence for spatial and temporal variability of the Atlantic Multidecadal Oscillation, Geophys. Res. Lett., 39, L09705, doi:10.1029/2012GL051241.

We add information about AMO studied from d18O in ice cores and its temporal and spatial relevance in Greenland. However the AMO influence is reduced in the north we see the AMO as possible explanation for the periodic oscillation between 1100 and 1600 AD.

Some minor specific comments
Abstract: The first paragraph is including many unnecessary numbers which are not informative. I suggest to rewrite the abstract. **We remove some of the given numbers and shortened the abstract.**
Material and methods: l. 20 please add reference to this method. **Done. Kaufmann et.al 2008**
The expression "very low accumulation" is used on many places in the text without defining it. **Very low means 100 mm/a or less.**
The language it generally acceptable but there are a number of typos that eventually needs to be fixed; as an example p 2352, l. 13: should be "firm" instead of "firn". **Done**