Reviewer 1

This manuscript examines the impact of summertime NAO events on European weather in subsequent summers. The paper is framed as being the impacts of North Atlantic freshwater events on European weather but the index that is used to depict these "freshwater events" is actually the summer NAO. The paper demonstrates some interesting connections between the summer NAO and the following summer weather. At this point, I'm unsure as to whether this manuscript is acceptable for publication. I have a number of comments on the analysis as outlined below. Overall, I'm giving a recommendation of major revisions to allow the authors to respond to these. My major concerns are that the direct link between the summer NAO and the freshwater events is unclear to me. This may be because I'm not an oceanographer and I haven't read the authors previous papers, so I hope that one of the other reviewers will be able to assess this aspect. I have some other concerns about the statistical methods used and the choices made for the scatter plots as outlined in my comments below.

We strongly thank the reviewer for reviewing the manuscript and providing many detailed comments and suggestions! The major concern is that the link between the summer NAO, the SST and freshwater is unclear. We will simplify the mass balance analysis, add further explanations, a schematic, and an example to make it easier for the reader to follow the analysis and understand how this link is derived. We will also add a section to motivate the approach of using a mass balance analysis to infer the variability of surface freshwater.

Currently available data products for the sea surface salinity have large biases, short time spans or poor spatial and temporal resolutions. For instance, satellite products have biases of up to ~1 g kg^{-1} in the polar regions and are only available since 2011 (Bao et al., 2019).

To overcome these challenges associated with freshwater analyses, we take advantage of the influence of freshwater on the SST in order to infer its variability. Specifically, we select conditions that are associated with pronounced cold anomalies in the subpolar region, and then estimate the extent to which freshwater has contributed these cold anomalies. In the following, we outline the main steps, which will also explained in the main manuscript, rather than only the appendix:

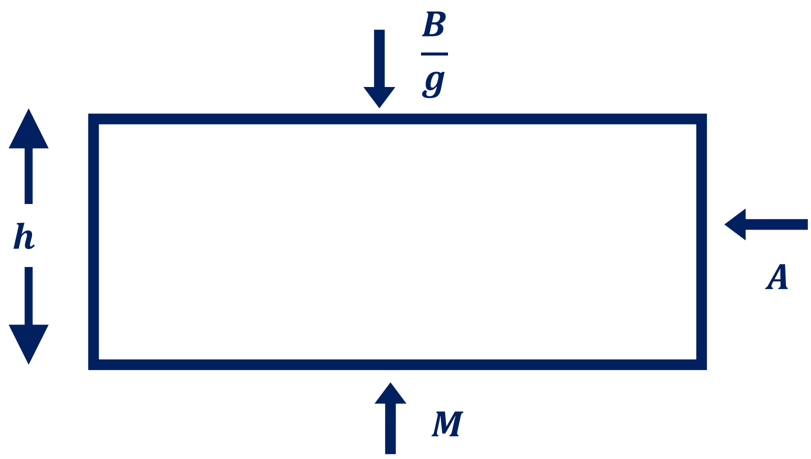

Figure 1: Mass budget for a mixed layer of depth h in the cold anomaly region. $\boldsymbol{A}$ corresponds to horizontal advection, $\boldsymbol{M}$ is an anomalous density flux from beneath the mixed layer, and $\frac{\boldsymbol{B}}{\boldsymbol{g}}$ refers to the density contribution from the surface buoyancy flux.

(1) We start with conservation of mass: $\frac{\partial}{\partial t}\int_{-h(t)}^{0}\rho dz = -\frac{B}{g} + M + A$, where B is the buoyancy flux through the surface, g is the gravitational acceleration, ρ is density, M is the mass flux through the base of the mixed layer, and A is horizontal advection (e.g. Griffies and Greatbatch, 2012).

(2) Next, we discretise the mass equation and integrate over the winter, using a variable mixed layer depth evolution from h_0 to h_n:

$$\rho_n = \frac{h_0}{h_n}\rho_0 + \left(-\frac{B_n}{g} + M_n + A_n\right)\cdot\frac{\Delta t}{h_n}$$

Here, the subscript $n \in 1..N$ refers to the n'th winter of an arbitrary subset of N winters. Before the winter, the mixed layer (h_0) is several tens of metres deep while during the winter, it reaches several hundred metres. Therefore, the density anomaly in the initial shallow mixed layer is distributed over a much larger depth range and the first-term on the right-hand side is negligible compared to the other terms. Any density anomalies beneath the initial, shallow mixed layer are included in M_n.

(3) We then linearise the equation of state: $\rho_n \approx \rho_m[1 - \alpha(T - T_m) + \beta(S - S_m)]$, where T is the temperature, S is the salinity, and α and β are the thermal and haline expansion coefficients. The subscript m refers to an arbitrary reference state, which, for simplicity, is chosen to be the mean over the subset.

(4) In order to infer the salinity, we select indices that are well-correlated with a cold anomaly in the subpolar North Atlantic but not with the potential drivers of density anomalies on the right-hand side of the equation. As an educated guess, we start with indices that we expect to be well-correlated with freshwater in the North Atlantic. Thus, we find that the freshwater indices F_M and F_C are associated with pronounced cold anomalies.

(5) After evaluating each term in the mass balance for the cold anomaly region, we find that the density increase, resulting from the cold anomaly, is more than one order of magnitude larger than any of the potential drivers on the right-hand side of the equation.

This implies that, for the selected indices, there is no anomalous density increase in the cold anomaly region but that the density increase implied by the cold anomaly must be balanced by a density decrease associated with a fresh anomaly $\alpha(T - T_m) \approx \beta(S - S_m)$. The revised manuscript will go through the involved calculations in more detail.

To demonstrate that the surface mass balance yields a good approximation of the salinity anomaly, we consider the last two decades, when sufficient Argo float observations are available to test the results. During the strongest observed cold anomalies over this period, which occurred in the winters 2015 and 2016, the correlation between the temperature and salinity anomalies has a p-value of $p = \sim 5.0 \cdot 10^{-242}$, with the regression of salinity on the temperature closely matching the regression predicted by the mass balance analysis (Fig. 2). The approximation, obtained with this method has a root mean square error of ~ 0.09 g kg^{-1} and is thus more accurate than any currently available data product for the sea surface salinity. In addition, it provides longer time series, higher resolution, and better coverage than the available in-situ data.

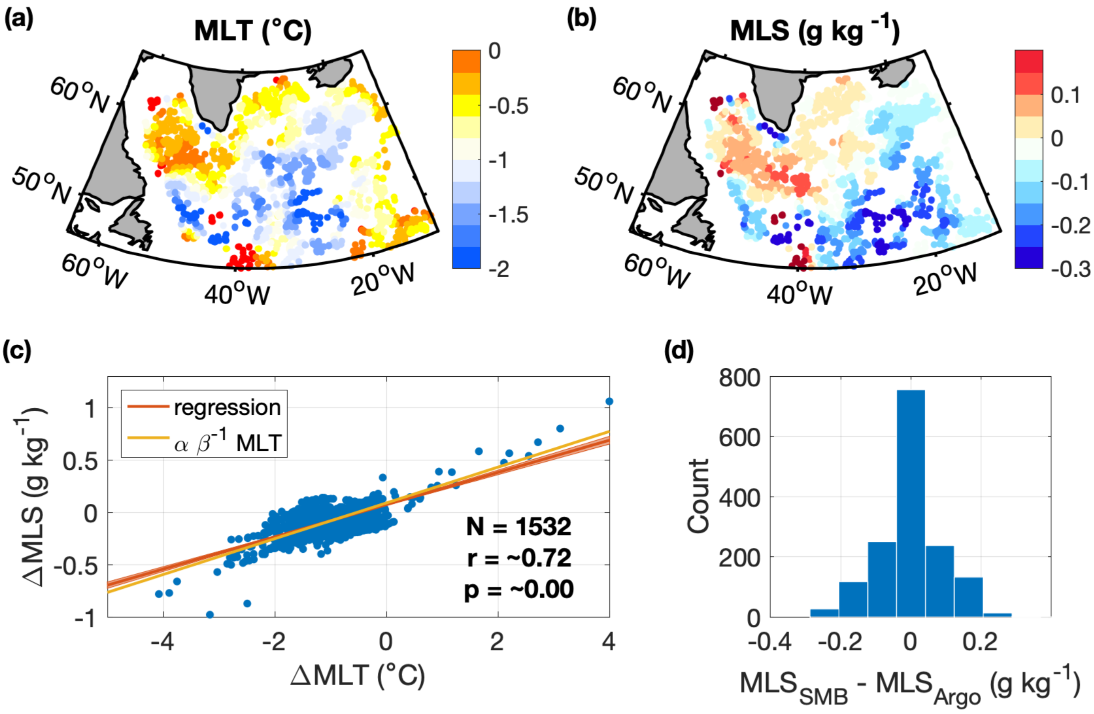

Figure 2: Demonstration of the surface mass balance. (a) MLT and (b) MLS are the mixed layer temperature and salinity anomalies during the winters 2015 and 2016, obtained from Argo float profiles (Holte et al., 2017). (c) The red line corresponds to the regression of the mixed layer salinity anomalies on the mixed layer temperature anomalies, while the yellow line corresponds to the approximation obtained from the mass balance analysis. (d) Histogram of the error of the mass balance analysis, corresponding to the difference between the calculated and observed salinity anomalies. The associated root mean square error is ~0.09 g kg^{-1}.

Bao, S., Wang, H., Zhang, R., Yan, H., & Chen, J. (2019). Comparison of satellite-derived sea surface salinity products from SMOS, Aquarius, and SMAP. Journal of Geophysical Research: Oceans, 124(3), 1932-1944.

Griffies, S. M., & Greatbatch, R. J. (2012). Physical processes that impact the evolution of global mean sea level in ocean climate models. Ocean Modelling, 51, 37-72.

Holte, J., Talley, L. D., Gilson, J., & Roemmich, D. (2017). An Argo mixed layer climatology and database. Geophysical Research Letters, 44(11), 5618-5626.

General comments:

(1) The link to freshwater anomalies and the role of low frequency North Atlantic ocean variability. I am not an oceanographer, so I hope that one of the other reviewers will have the expertise to comment on this. The link between the summer NAO index and the freshwater anomalies was a bit lost on me. My understanding of lines 84-92 is that the authors are assuming that the temperature anomalies associated with the summer NAO are due to freshwater anomalies because they find that the cooling is not strongly related to surface fluxes, wind driven Ekman transports, Ekman pumping and re-emergence of SST anomalies from previous years, so by a process of elimination they conclude that it's freshwater anomalies. But I don't see how the role for other ocean circulation anomalies such as the AMOC or advective heat convergence due to circulation anomalies produced by things other

than the wind driven Ekman transports has been eliminated. The atlantic ocean circulation exhibits variability on long timescales which can be a driver of the NAO and vice-versa (e.g., Zhang et al 2019, Review of Geophysics, 10.1029/2019RG000644 and references therin). It's not clear to me (a) whether it can really be concluded that the SST anomalies are related to freshwater inputs and (b) whether such low frequency variability in the ocean circulation has been appropriately taken into account. The NAO index is being described as a "freshwater index" (l99) but I'm not sure how appropriate this is and I'm not sure that much would be lost by instead referring to it as the NAO index and focussing on the impact of the summertime NAO on the climate in subsequent years.

We thank the reviewer for pointing out that the derivation of the advective transports and the role of low-frequency variability was unclear. The first question is why we exclude that advective anomalies contribute to the cooling. Away from the boundaries, advective transports must be forced either mechanically through winds or by density gradients. As the reviewer mentions, Ekman transports, driven by the winds, were shown to be negligible.

Buoyancy-driven flows, moreover, are the response to existing density gradients in the ocean but cannot create strong gradients on their own (e.g. Wunsch and Ferrari, 2004). In the absence of any external buoyancy forcing, the ocean cannot create density gradients by itself. In the analysis, we show that the surface buoyancy flux is too weak, not significantly correlated with the freshwater indices, and inconsistent to drive the cold anomaly (Figs. A1c and A2c). Thus, it will not create an anomalous buoyancy-driven flow. We conclude that neither wind- nor buoyancy-driven flows can explain the cold anomaly.

The second question is whether low-frequency variability has been taken into account. On long timescales, the external freshwater forcing can lead to a reduction in the buoyancy-driven overturning circulation (e.g. Stommel, 1968). In this case, the density decrease associated with the freshening even exceeds the density increase associated with the cold anomaly $\beta \Delta S < \alpha \Delta T$. However, the resulting reduction in the overturning circulation would lead to a positive surface heat flux anomaly in the subpolar region, such that the ocean loses less heat to the atmosphere (e.g. Gulev et al., 2013). The reduction in ocean heat losses can also be understood by considering that freshwater increases the stratification and thus reduces the amount of heat that is available to the atmosphere.

As shown in the mass budget, we do not find a significant heat and, in turn, buoyancy flux anomaly (Figs. A1c and d and A2c and d). For instance, the buoyancy flux anomaly associated with F_M events results in a mass decrease of $\sim$7 kg m^{-2}, while the cold anomaly implies a mass increase of $\sim$204 kg m^{-2}. Thus, the freshwater anomaly implied by the cold anomaly is more than one order of magnitude larger than the freshwater anomaly implied by a potential slowdown of the overturning circulation. In turn, this means that the freshwater increase resulting from a slowdown of the overturning circulation is negligible on the timescales considered.

In the revised manuscript, we will clarify that the buoyancy-driven flows are a response, not a driver of density anomalies (Tziperman, 1986; Wunsch and Ferrari, 2004), and we will explicitly state that the freshening, associated with a potential slowdown of the overturning is more than one order of magnitude smaller than the freshening, implied by the cold anomaly. We will further clarify that, on the spatial scales and interannual timescales considered, by far the strongest advective transports of heat and freshwater in the interior subpolar region result from geostrophic flows, both within eddies and as part of the subpolar gyre circulation. However, they do not contribute to the mass budget since geostrophic flows are along density contours.

Gulev, S. K., Latif, M., Keenlyside, N., Park, W., & Koltermann, K. P. (2013). North Atlantic Ocean control on surface heat flux on multidecadal timescales. Nature, 499(7459), 464-467.

Stommel, H., & Rooth, C. (1968, April). On the interaction of gravitational and dynamic forcing in simple circulation models. In Deep sea research and oceanographic abstracts (Vol. 15, No. 2, pp. 165-170). Elsevier.

Tziperman, E. (1986). On the role of interior mixing and air-sea fluxes in determining the stratification and circulation of the oceans. Journal of Physical Oceanography, 16(4), 680-693.

Wunsch, C., & Ferrari, R. (2004). Vertical mixing, energy, and the general circulation of the oceans. Annu. Rev. Fluid Mech., 36, 281-314.

(2) Detrending: It's stated at line 53-55 that regionally averaged trends were subtracted from the air temperatures to remove the greenhouse gas effects. It doesn't really seem appropriate to me to remove the linear trend from one field but not others. The NAO index that is used clearly has a linear trend in it (Figure 1a). I'd suggest detrending everything or detrending nothing. I'm not arguing that the NAO trend seen in Figure 1 is greenhouse gas forced or that this trend should necessarily be removed, but it just doesn't seem appropriate to me to remove the trend in one field and not in the others. Is the detrending also done on the SSTs? It doesn't make much sense to me to remove the trend from the surface air temperature but not the SSTs.

We thank the reviewer for making us aware that the removal of the trend was insufficiently explained. Since freshwater has a trend (Fig. 3), trends are part of the dynamic signal we are interested in. For instance, a trend in the freshening would lead to a trend in the cold anomaly, and in turn, a trend in the jet stream shift and so on.

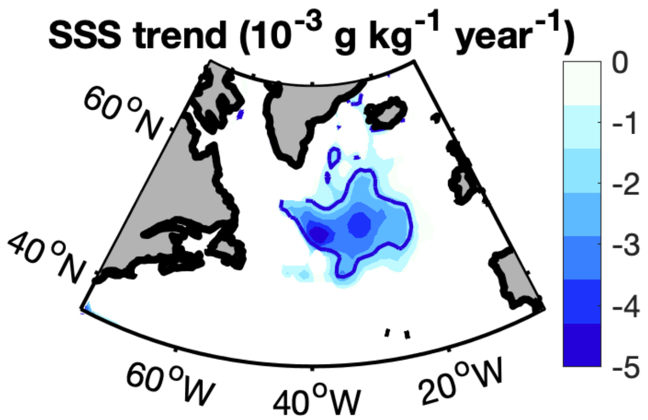

Figure 3: Average trend in the sea surface salinity over the last 70 years, inferred from a surface mass balance using Hadley SST data and the reanalysis ERA5.

For the air temperature, however, there is an additional, large trend due to increased greenhouse gas concentrations. In contrast to the temperature trend that results from the jet stream shift, the warming trend due to increased greenhouse gases is distributed relatively uniformly. Thus, it can be separated from the warming trend due to the trend in the jet stream shift by averaging the temperature over a sufficiently large area before removing the trend.

The removal of the trend in the air temperature is thus based on the assumption that any potential warming trend associated with an SST anomaly over the North Atlantic must be balanced by a cooling trend over the ocean, if the warming and cooling are linked to the same atmospheric instability.

We agree with the reviewer that, for consistency, this trend should be removed from all variables. However, we found that neither the SST nor the other atmospheric variables have a significant trend when they are averaged over a large area. For instance, P-E in summer has a trend of $\sim -2.9 \cdot 10^{-4} \pm 3.0 \cdot 10^{-4}$ *m year^{-1} when it is averaged over the same area, which is not significant. Likewise, the SST has a trend of -0.0046 ± 0.0062 °C year^{-1} when it is averaged over the North Atlantic (from 0 to 65 °N), which is also not significant. Removing these trends prior to the analysis does not lead to any notable differences in the results.*

However, upon checking again, we found that the absolute dynamic topography (ADT) also has a significant positive trend when averaged over the North Atlantic. The identified increase is likely due to the long-term ice loss of glaciers, and thermal expansion (Church et al., 2001). In the revised version, we have therefore also removed the ADT trend. This did not appreciably affect the results.

In the revised version, we will explain the removal of the trend in more detail in the method section. We will also specify the region used for the averaging and point out that the results are not sensitive to the choice of the region as long as the region is sufficiently large. In addition, we will clarify that none of the other responses (apart from the air temperature and the ADT) has a significant trend when it is averaged. We thank the reviewer for making us check all potential trends again to ensure consistency in the analysis.

With regard to the summer NAO, we find that it has a weak trend of 0.01 year^{-1}. However, for the reasons stated above, we think that removing this trend would not be meaningful. It is used as an indicator for freshwater. Thus, any manipulation of this time series would affect its representation of the freshwater anomaly and therefore be counterproductive.

It is worth noting that the precipitation minus evaporation anomalies and the temperature anomalies show very similar patterns, and that these are consistent with the jet stream shift and, in turn, the cold anomaly. The consistency of the observed patterns across all investigated variables, where only one of them is affected by the direct warming due to increased greenhouse gases (and therefore has a trend when it is averaged), provides compelling evidence that the warming due increased greenhouse gas forcing has successfully been removed.

Church, J. A., Gregory, J. M., Huybrechts, P., Kuhn, M., Lambeck, K., Nhuan, M. T., ... & Woodworth, P. L. (2001). Changes in sea level. In , in: JT Houghton, Y. Ding, DJ Griggs, M. Noguer, PJ Van der Linden, X. Dai, K. Maskell, and CA Johnson (eds.): Climate Change 2001: The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel (pp. 639-694).

(3) For the scatter plots, the regions where the correlation is significant at the 95% level is used for the spatial averaging. This seems like cherry picking to me. Of course, the correlations look good because you've chosen them to be that way. It would make more sense to choose a physically motivated region or it would seem to be less cherry picking if a regular spatial region such as a rectangle were chosen. The result is that in Figure 2b there is a correlation of 0.98, which seems quite unbelievable to me, but maybe it isn't if you are just averaging over regions where the correlation is high.

The significance of the identified relationships is assessed by the thick contours in all regressions that are shown. However, due to the low number of degrees of freedom in the F_M events, the scatter plots are an important addition. This is because the obtained relationships could potentially result from individual outliers or clusters of values. In order to rule out that the significance of the relationships is due to outliers, it is important to show the individual values of the involved timeseries for which the significance is shown.

In theory, we could show the scatter plots for all variables but the information would be redundant since we always use the same regressor (F_M), and the first and last variables already show that there are no outliers or clusters in the distribution of F_M. Instead, the points are evenly distributed. Therefore, it is sufficient to show the values only for the first step in the chain of events. We also show them for the last step (precipitation minus evaporation), to point out the steep regression slopes and thus, the high sensitivity of the P-E response to small variations in the freshwater forcing.

Since the scatter plots are needed to confirm the significance of the identified relationships, they must, by definition, include the variables at the locations where a significant link is found. An analysis of the significance outside these regions would not be meaningful.

(4) Has autocorrelation been accounted for when calculating the significance levels? If not, I think it should be. Clearly each year is not independent and there is some low frequency variability and autocorrelation, as apparent in the NAO index (Figure 2a).

High auto-correlations across the events can reflect a potential redundancy in the events. Thus, we ensured that the auto-correlations of the freshwater indices are negligible, and we also show them in the appendix. Since we are comparing freshwater events of the same type with each other, rather than different types of freshwater events, we need to consider the autocorrelations across the events for consistency with the analyses. The results are in good agreement with the scatter diagrams which showed that there are not outliers or clusters. To avoid that this information is missed, we will move it to the beginning, where the indices are introduced for the first time, and we will better explain it.

(5) It is argued that this work reveals new potential to enhance the predictability of European summer weather, but I think for the impacts on European summer weather the results have only been presented in the form of regression coefficients. To make this more relevant for predictability, it might be worth showing the variance explained.

We thank the reviewer for this excellent suggestion. In the revised manuscript, will make sure to include an additional figure showing the explained variance of European summer weather.

Comments by line number:

Figure 2: It seems like it would be interesting to have the regression maps for F_C as well as F_M. You use the regions based on the regression onto F_M for both F_M and F_C, so it would be good to see whether the regression map for F_C has a similar spatial pattern to that for F_M or not.

Thank you for letting us know that the figure raises the question about the corresponding regression maps for F_C. In the new version, we will show both regression maps in a single figure. We will also provide the correlation between the two timeseries obtained from each pattern.

Figure 2 caption: F_M and F_C are only defined in this figure caption. Given their central importance, I think they should also be defined in the text. Furthermore, it would be worthwhile making clear the motivations for this naming convention. It's not very intuitive where the choice of "F_M" and F_C" comes from and I think it would help readers to follow if you make that clear. In the end, I realized that this corresponds to "melt-driven' and "circulation-driven" events and I'm overall just very confused about how this distinction can be made just on the basis of the NAO index, which relates to my general comments above. I think this needs to be made clearer throughout the manuscript.

Thank you for pointing out that the motivation for the naming was not clear. We will explain the naming more clearly and have included additional details on the cause of the freshwater anomalies associated with F_C and F_M.

Freshwater anomalies associated with F_C are characterised by an enhanced offshore advection of fresh, polar water into the subpolar region. We have now shifted the associated figure, showing the circulation anomaly, to the beginning of the analysis, when we introduce the freshwater indices. This circulation anomaly results from the increased windstress curl in the subpolar region during positive NAO years (e.g. Häkkinen et al., 2011).

Freshwater anomalies associated with F_M, on the other hand, are characterised by more freshwater inside the currents, rather than a change in the currents themselves. This additional freshening is quantified by the negative summer NAO (without sub-sampling) and was derived in Oltmanns et al. following a similar mass balance analysis but for the density change $\frac{\partial \rho}{\partial t}$ from summer to winter rather than absolute anomalies because the focus was on a shallow surface layer of 30 m, allowing to evaluate seasonal differences. The freshening, represented by the negative summer NAO corresponds to the anomalous seasonal freshwater that is added to the subpolar North Atlantic during autumn.

The anomalous seasonal freshening associated with a more negative summer NAO applies to all years, without sub-sampling them. However, smaller freshwater anomalies are mixed down before a significant fresh and cold anomaly in winter develops. Therefore, we cannot use the negative summer NAO to obtain absolute anomalies (unless we sub-sample it and only use strong events).

When the seasonal surface freshening is too small, the final freshwater anomaly in winter is dominated by changes in the circulation (and thus F_C events), which have the opposite atmospheric driver. This is why we need two freshwater indices. Using both allows to separate the different drivers of fresh, and hence cold, anomalies.

The additional seasonal freshening, associated with the negative summer NAO, must be due to runoff, melting or precipitation minus evaporation. After comparing the associated runoff and precipitation anomalies, we find that P-E is anti-correlated with the negative summer NAO, while runoff and melting are correlated (e.g. Hanna et al. 2013). For instance, the correlation between the negative summer NAO and runoff from Greenland and Canada over the last 40 years is $r = \sim 0.63$ with a p-value of $p = \sim 1.5 \cdot 10^{-5}$, obtained from the Greenland climate model MAR (Fettweis et al., 2013). However, we do not differentiate between these sources. We only refer to them as F_M events and do not specify whether it is melt from sea ice or glacial ice or runoff, and where the melting originally occurred. We will clarify this in the revised manuscript.

So, by (1) showing the change in the ocean circulation, we can link the freshwater anomalies associated with F_C events to a change in the subpolar gyre circulation. By (2) showing the seasonal freshening associated with the negative summer NAO, we can link the freshwater anomalies associated with F_M events to increased seasonal freshwater input into the currents. By (3) comparing the likely roles of *runoff and melting with precipitation anomalies, we can link the increased freshwater inside the currents to more runoff and melt from sea ice and glacial ice.*

We will add a section in the manuscript to derive these links more clearly and thus motivate the naming of the indices. In this section, we will also include further references on the cause of freshwater anomalies, supporting the naming, and the seasonality of freshwater export into the subpolar region during autumn (e.g. Fratantoni and McCartney, 2010; Schmidt and Send, 2007; and references therein). We will also clarify that the additional freshwater during F_M events has multiple origins related to enhanced seasonal runoff, sea ice and glacial melting, but not precipitation.

Fettweis, X., Franco, B., Tedesco, M., Van Angelen, J. H., Lenaerts, J. T., van den Broeke, M. R., & Gallée, H. (2013). Estimating the Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR. The Cryosphere, 7(2), 469-489.

Fratantoni, P. S., & McCartney, M. S. (2010). Freshwater export from the Labrador Current to the North Atlantic Current at the Tail of the Grand Banks of Newfoundland. Deep Sea Research Part I: Oceanographic Research Papers, 57(2), 258-283.

Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2011). Warm and saline events embedded in the meridional circulation of the northern North Atlantic. Journal of Geophysical Research: Oceans, 116(C3).

Hanna, E., Jones, J. M., Cappelen, J., Mernild, S. H., Wood, L., Steffen, K., & Huybrechts, P. (2013). The influence of North Atlantic atmospheric and oceanic forcing effects on 1900–2010 Greenland summer climate and ice melt/runoff. International Journal of Climatology, 33(4), 862-880.

Schmidt, S., & Send, U. (2007). Origin and composition of seasonal Labrador Sea freshwater. Journal of Physical Oceanography, 37(6), 1445-1454.

Figure 3 caption: Maybe explain a bit more what the "absolute dynamic topography" is. Is it just sea surface height?

Thank you for pointing out that this was unclear. In the revised version, we will clarify that the absolute dynamic topography is the sea level anomaly with respect to the geoid. Absolute dynamic topography thus also allows to show the mean location of the ocean currents, which would be averaged out in the sea level anomalies. However, since we only showed the absolute dynamic topography in regressions, using sea level anomalies would have led to the same results.

l138: "expansion of the cold anomaly" - perhaps be clear about what this "expansion' is relative to? Is it relative to the previous summer?

This and the following comments include very helpful suggestions. We thank the reviewer for providing all these suggestions and we will follow all of them to clarify the manuscript.

Section 4.5: It might be worth making it clear at the beginning here that this is now back to looking at the observations, since in the previous section the focus was on model simulations.

Figure 11 caption: The referencing to the panel labels is messed up in the caption.
Typo's/wording suggestions:
l62: "this index" --> "the NAO index"

l68: suggest "smaller values" --> "more negative values" because the magnitude of the NAO index isn't smaller.

l84: "Fig. 2d" --> "Fig. 2c" (I think d is showing salinity, not temperature)

Figure 4 caption: "The thick contours show the 95% confidence levels" --> "The thick contours encompass regions that are significant at the 95% confidence levels"

l158: "SST-forced" is a bit unclear. Suggest "Simulations performed with prescribed observation-based SSTs".

We again thank the reviewer for the detailed review, helping us to improve this manuscript!