Response to Reviewer # 1

We thank anonymous Reviewer # 1 for evaluating our manuscript. Below, we list our responses to each comment (in blue). We have updated Figure 1 (and others), and have added a new regression analysis, which better illustrates the role of aerosols via shortwave radiation perturbations (and the role of clouds) in driving multidecadal AMOC variability.

Reviewer # 1

General Comments:

This manuscript examines the multi-decadal variability of the AMOC and a number of atmospheric variables in the North Atlantic in the historical observations and simulations in the CMIP6 models. The main finding is the implication of anthropogenic aerosols in strengthening the AMOC during the period 1950-1990, while a reduction in aerosols afterwards lead to GHGs having a larger effect and weakening the AMOC. I think the methods and results are reasonable, and I particularly appreciated the well-balanced considerations on the roles of the anthropogenic drivers in the AMOC, the possible errors in the models, and the discussion of observational evidence, including AMOC proxies derived from other ocean variables (e.g., SST) and radiation estimates of aerosol forcing. I recommend acceptance pending minor revisions.

Specific Comments:

In fig. 1 please add the correlation coefficient of all the panels with panel a (AMOC) in the figures (maybe upper right?). Since in the text it is repeatedly discussed how these timeseries are related, it is important to have a quantitative reference. I am aware that this information is present in fig. 2, but it would improve the readability of the text until we get to fig. 2.

Thank you for the comment. We have added the 1950-2020 correlation coefficient (r) between the time series of each variable in Figure 1 and the AMOC. These correlation coefficients are shown in the upper right-hand side of each panel.

We have also included three more panels in Figure 1, including March mixed layer depth (MMLD), sea surface density (SSD) and the haline component of the surface density flux (HSDF). These new variables are also added to the corresponding figure that shows time series based on CMIP6 AA simulations.

Throughout the manuscript it is unclear to me what is the role of SW radiation since it is related to a number of factors: clouds, aerosols, sea ice. Please elaborate on what you are looking at when you discuss SW, and consider using regressions (or kernels) to be more specific about the radiative contribution of each variable you are interested in examining.

The multi-decadal evolution of subpolar North Atlantic net surface shortwave radiation from 1950-2020 is largely driven by changes in anthropogenic aerosol emissions. Several figures support this claim, including close correspondence between subpolar North Atlantic SW radiation ensemble mean time series in CMIP6 ALL and CMIP6 AA (the latter being driven by changes in AA only). Our lead-lag plots also show that subpolar North Atlantic SW and aerosol optical thickness (AOT) are significantly correlated at -0.89 (no lag). Two figures in the supplement also show close correspondence between AOT and SW spatial trends.

We also note that subpolar North Atlantic SW is in phase with sea surface temperature (SST) with a high correlation of 0.91. Both SW and SST lead the AMOC by ~12 years, with a maximum correlation of -0.84 and -0.85, respectively. Thus, the lead-lag correlation analysis shows that AOT, SW and SST are highly correlated, in phase, and lead the AMOC by ~decade. This suggests AOT, via perturbations to SW and SST, is driving the AMOC changes. We have elaborated upon this in the revision.

Additional statistical analyses (via regressions) are now included to further show the role of net surface shortwave radiation (and other variables). This new analysis is in subsection 3.1.3 "Regression Decomposition into Aerosol-Forced and AMOC Feedback Components", and new figures have been added to the revision.

We decomposed the North Atlantic climate response into an anthropogenic aerosol-forced component and a subsequent AMOC-related feedback. We use the negative of net downward surface shortwave radiation (-1xSW) as a proxy for the change in anthropogenic aerosols (similar results are obtained if we use AOT). The forced response is obtained by regressing -1xSW onto different fields such as sea surface temperature (SST), surface wind speed (SFWD), sea level pressure (PSL), etc. The regression coefficients are based on linear least-squares regression analysis applied to the CMIP6 ensemble annual mean. We subsequently remove this anthropogenic aerosol related variability to isolate the AMOC related feedback, by regressing the AMOC time series onto the new field (with aerosol-related variability removed). This regression method is described in section, 2.4 "Regression Analysis".

Our new regression analysis shows that SSTs have a significant negative sensitivity (and the temperature component of sea surface density, SSD, has a significant positive sensitivity) to changes in subpolar North Atlantic -1xSW, meaning that an increase in aerosols is associated with SST cooling (and increased SSD) and vice versa (Figure R1 below). On the other hand, the SST-AMOC feedback component shows the opposite sensitivity. That is, an increase in aerosols initially cools SSTs and increases SSD, which promotes deep convection in the subpolar North Atlantic and eventually an enhanced AMOC; the enhanced AMOC (the feedback) acts to counter the cooler SSTs/enhanced SSD, due to enhanced poleward oceanic heat transport by a stronger AMOC. This is consistent with prior studies that have suggested AMOC weakening is associate with the "warming hole" in the subpolar North Atlantic.

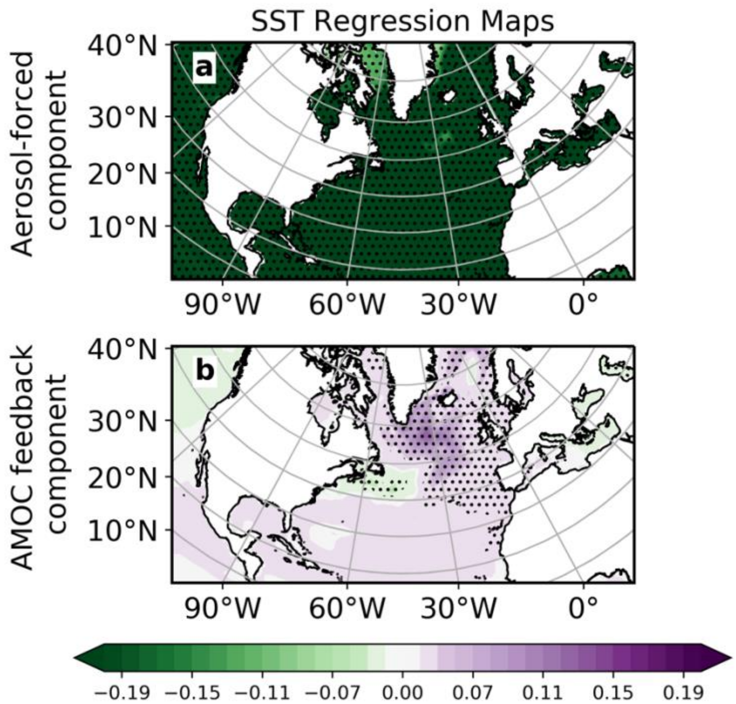

Figure R1. *Ensemble mean annual mean CMIP6 all forcing regression analysis. Decomposition of sea surface temperature (SST) into (top panel) aerosol forced and (bottom panel) AMOC feedback components. The forced response is obtained by regressing the subpolar North Atlantic -1xSW time series (a proxy for anthropogenic aerosols) onto each field. The AMOC-related feedback is obtained by removing the variability associated with the forced response, and then regressing the AMOC time series onto this new field. The feedback field is converted to the same units as the aerosol-forced field by multiplying the feedback field by the regression slope between the AMOC and -1xSW subpolar North Atlantic time series (δ(AMOC)/δ(-1xSW) = 0.32 Sv/(W m^{-2}), significant at the 95% confidence level). The units for the SST regression maps are K/(W m^{-2}). Symbols denote regression significance at the 95% confidence level.*

We have also conducted similar regression analyses for other climate variables, including total cloud cover (CLT; Figure R2 below). The aerosol forced response show positive subpolar North Atlantic sensitivities, which is consistent with aerosols increasing cloud cover. This would act to reinforce cooling of subpolar North Atlantic SSTs, strengthening the forced component of aerosols on the AMOC. In contrast, the AMOC feedback shows opposite signed CLT sensitivities, which is likely related to the aforementioned AMOC feedback on SST and subsequent impacts on lower-tropospheric atmospheric stability. This, and additional information, is elaborated upon in the revision.

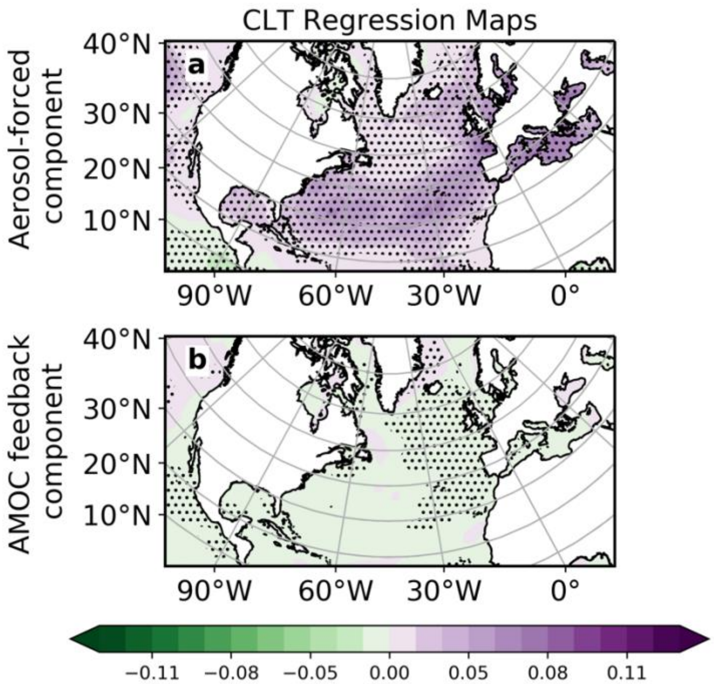

Figure R2. *Ensemble mean annual mean CMIP6 all forcing regression analysis. Decomposition of total cloud cover (CLT) into (top panel) aerosol forced and (bottom panel) AMOC feedback components. The forced response is obtained by regressing the subpolar North Atlantic -1xSW time series (a proxy for anthropogenic aerosols) onto each field. The AMOC-related feedback is obtained by removing the variability associated with the forced response, and then regressing the AMOC time series onto this new field. The feedback field is converted to the same units as the aerosol-forced field by multiplying the feedback field by the regression slope between the AMOC and -1xSW subpolar North Atlantic time series (δ(AMOC)/δ(-1xSW) = 0.32 Sv/(W m^{-2}), significant at the 95% confidence level). The units for the CLT regression maps are fraction/(W m^{-2}). Symbols denote regression significance at the 95% confidence level.*

<u>Technical Comments:</u>

L50: should be 'remains'

Fixed.

L230: should be 'significant'

This paragraph has been removed and replaced with a new regression analysis.