Author's response to Referee #3

egusphere-2026-2588

Recommendation: Major Revision. The manuscript uses the 100-member CESM2 Large Ensemble (CESM2-LE) to investigate large multidecadal shifts in density-space overturning strength at 55∘N during the historical period (1850–2014) and near-term future (2015–2030). A change-point detection algorithm is applied to individual ensemble member time series, and shifts are classified into three categories: strengthening events, pre-1985 weakening events, and post-1985 weakening events. Composite analysis is used to characterise the hydrographic evolution preceding and following each shift category. The headline finding is a transition from a salinity-dominated, internally driven variability regime to a thermodynamically forced, temperature-dominated regime around 1985, which the authors attribute to the historical transition from aerosol-dominated to greenhouse gas-dominated radiative forcing.

The dataset is large and the composite methodology is internally consistent. However, as detailed below, the statistical foundations of the three-category classification are inadequate, the central attribution claim is unsupported by the analysis, and the paper's scientific contribution does not yet substantially exceed what has already been established in the literature.

We thank the referee for the thorough review and valuable suggestions. We will add significance testing and bootstrap confidence intervals throughout, and revise the framing of the aerosol/greenhouse-gas forcing discussion to state that this aligns with, rather than independently establishes, prior forcing-attribution studies. Below, we respond to each point individually: referee comments are shown in italics, followed by our reply.

Comment RC3.1

Asymmetric detection thresholds and the 12-event pre-1985 category *The three-category framework that organises the entire paper rests on an asymmetric detection design that is never formally justified. Strengthening shifts and post-1985 weakening shifts are detected at a 3σ threshold; pre-1985 weakening shifts are detected at a 2σ threshold because — by the authors' own account — they are absent at the 3σ level. The consequence is that the three categories do not represent homogeneous tiers of a single classification system: the pre-1985 category is structurally weaker in amplitude than the other two. This matters because the paper's central scientific argument depends on the pre-1985 category being a physically distinct mechanism and not merely the small-amplitude tail of the same weakening process that operates more vigorously after 1985. The pre-1985 composite is based on 12 events drawn from a 100-member ensemble spanning 1850–2014. For context, the post-1985 composite contains 195 events. With $n = 12$, the composite mean is imprecisely estimated. The standard error of the composite mean for a variable with standard deviation $\hat{\sigma}$*

is $\hat{\sigma}/\sqrt{12} \approx 0.29\hat{\sigma}$, meaning that even the ensemble-mean signal is uncertain at the ∼30% level before accounting for event-to-event heterogeneity. No confidence intervals, bootstrap distributions, or significance tests are presented for any composite, in any category, anywhere in the paper. The authors therefore cannot demonstrate that the patterns shown in Figures 6–7 and the corresponding panels of Figures 5 and 10 are distinguishable from sampling noise. This is not a presentational deficiency. It is a logical gap: the paper claims a third, mechanistically distinct circulation category whose entire evidential basis is a 12-event composite with no significance assessment. If a permutation test or bootstrap analysis were to reveal that the pre-1985 composite anomalies are not statistically distinguishable from random draws of 12 events from the strengthening or post-1985 categories, the three-category framework collapses into two. The paper must demonstrate that this is not the case before any conclusions about pre-1985 mechanisms can be drawn.

Author's reply:

We share the concern the referee raises here. We will ensure that significance testing is presented throughout the manuscript, and for this specific case, we propose the following.

As suggested, we will (1) bootstrap the pre-1985 composite, resampling the 12 events with replacement, to obtain confidence intervals on the composite pattern, and (2) run a permutation test drawing repeated random samples of 12 events from the pooled strengthening and post-1985 weakening categories, to test whether the pre-1985 composite differs from a null built from other real events.

We will complement this with a third test: a permutation test drawing repeated random samples of 12 non-event periods, matched in sample size to the real composite, taken from ensemble members and time windows in the pre-1985 era that did not register a detected shift. This tests more directly whether these are genuine events rather than sampling noise.

We have also investigated whether the same mechanism appears in the CESM2 piControl run, where no external forcing is present. Multidecadal variability is inherently difficult to characterise statistically, particularly in forced simulations where the historical record limits how many qualifying events can occur; the pre-1985 category is drawn from the full 100-member CESM2-LE ensemble, and 12 is the number of events this ensemble contains within that period. Applying the same change-point detection algorithm to the piControl run, using both a 40/20-year and a 20/20-year window configuration, we find 2 and 7 decrease events respectively with characteristics consistent with the pre-1985 category, providing independent physical support for this mode of variability despite this constraint.

Comment RC3.2

Absence of significance testing throughout. *The absence of statistical significance assessment is not confined to the pre-1985 category. No figure in the paper indicates where composite anomalies exceed a significance threshold. No time series panel shows confidence intervals. The normalised anomaly time series in Figure 11 are presented without uncertainty bounds. This omission is standard practice that would be unacceptable at this journal. For spatial composite maps, significance stippling (e.g. regions where the fraction of events exceeding a threshold exceeds 95% under a sign test, or where a t-test against zero is significant at* $p < 0.05$ *with appropriate multiple-comparison correction) is required. For time series, shaded uncertainty bounds derived from bootstrap or ensemble spread are expected. The current* *presentation makes it impossible for a reader to distinguish robust features from artefacts.*

Author's reply:

We agree and will add significance stippling to all spatial composite maps (using a sign test and/or t-test against zero with appropriate multiple-comparison correction) and shaded confidence intervals derived from bootstrap resampling to all timeseries panels, including Figure 11.

Comment RC3.3

The change-point methodology and composite averaging conflate heterogeneous events. *The composites pool events from the full historical period (1850–2030). For the strengthening category, this includes clusters near 1895 and 1955 (Figure 2b). These events occur against very different background states: the late-nineteenth-century North Atlantic has a different mean density, stratification, sea-ice extent, and external forcing than the mid-twentieth century. Composite averaging across these epochs implicitly assumes stationarity — that the same mechanism operates identically in all events and that the pre-shift reference state is interchangeable across the ensemble. This assumption is not tested.*

The authors should at minimum check whether the 1895 and 1955 strengthening sub-composites yield consistent anomaly patterns. If they do not, the single strengthening composite obscures mechanistic heterogeneity. Similarly, the post-1985 weakening events span 1970–2030 (Figure 2b), during which the forced AMOC trend is itself non-stationary. Compositing over this span mixes events occurring at different phases of the forced trend, which may distort the composite amplitude and temporal structure.

Author's reply:

We have checked whether the 1895 and 1955 strengthening sub-composites yield consistent anomaly patterns, and separately whether the post-1985 weakening composite is sensitive to splitting by phase of the forced trend across its 1970–2030 span. In both cases we found no significant difference between the sub-composites. We will add these as supplementary figures in the appendix.

Comment RC3.4

Forcing attribution without forcing-separated experiments

The paper's headline result — that the pre-1985 regime reflects aerosol-driven dynamics and the post-1985 regime reflects greenhouse gas-driven dynamics — is stated in the abstract, results, and discussion as a finding of this study. It is not. No forcing-separated simulations (e.g. CMIP6 hist-aer, hist-GHG, hist-nat, or analogous LESFMIP experiments) are analysed. The attribution is made entirely by cross-referencing to prior studies — principally Menary et al. (2020) and Robson et al. (2022) — who conducted the relevant experiments. This is not original attribution; it is post-hoc narrative alignment.

The logical structure of the paper's argument is as follows: prior studies showed (using forced experiments) that aerosol forcing drives AMOC strengthening before

∼1985 and greenhouse gas forcing drives weakening thereafter; this study identifies two kinds of weakening separated by ∼1985; therefore, the two kinds of weakening reflect aerosol and greenhouse gas forcing respectively. This inference is not deductively valid. Multiple mechanisms can produce circulation shifts at the same epoch, and without forcing-separated runs it is impossible to determine what fraction of the variance in the detected shifts is attributable to a specific forcing agent versus internal variability, which is acknowledged in the literature to be substantial on multidecadal timescales (Latif et al., 2022).

Stripping away the unsubstantiated attribution, the remaining scientific content of the paper can be summarised as follows: density-space overturning at 55°N in CESM2-LE exhibits two distinct weakening mechanisms, one dominant before ∼1985 and one after. This observation is broadly consistent with what has already been established:

- *The transition from AMOC strengthening to weakening around the 1980s, and its attribution to aerosol and greenhouse gas forcing in CMIP6, is the central finding of Menary et al. (2020) (Geophysical Research Letters) and Robson et al. (2022) (Journal of Climate), who used proper forcing-attribution experiments.*
- *The positive feedback between surface heat loss, convection, and horizontal SPG transport — identified here as the "increase shift" mechanism — was described by Born and Mignot (2012) and is confirmed by Falkena et al. (2025) as a feature specific to CESM2.*
- *The eastward shift of deep convection from the Labrador Sea to the Irminger Sea in the twenty-first century — identified here as characteristic of post-1985 decreases — was reported by Rühs et al. (2021).*
- *The role of density anomaly propagation through the SPG as a conduit for multidecadal variability is established in Eden and Willebrand (2001), Tulloch and Marshall (2012), and Kwon and Frankignoul (2014).*

The methodological innovations — the density-space index at 55°N, the change-point detection framework, and the four-region depth–time decomposition — are real contributions but are instrumental rather than conceptual. They are being used to rediscover, with less statistical rigour, results that prior studies established with better-controlled experiments.

Author's reply:

We regret that this impression arose; it was not our intention. Throughout the manuscript, the forcing regimes discussed here are attributed to Menary et al. (2020) and Robson et al. (2022), consistent with how other findings drawn from prior studies are treated in the text. Two summary sentences, in the abstract and in the conclusions, state this more strongly than the rest of the manuscript intends. We will revise both, and review the manuscript as a whole, to ensure the aerosol/greenhouse-gas framing is presented as consistent with prior forcing-attribution studies rather than as a finding of this analysis.

Comment RC3.5

Conflation of SPG overturning at 55°N with basin-scale AMOC

The primary circulation metric, $\psi_\sigma^{\max}(55°N)$, is the maximum of the overturning streamfunction in potential density space at 55°N. This quantity characterises water *mass transformation in the subpolar gyre and is physically distinct from the basin-scale AMOC conventionally measured at 26°N ($\psi_d^{\max}(26°N)$), which is the index monitored by the RAPID array and the metric most relevant to the climate impacts invoked in the introduction (heat transport, sea level, carbon uptake). The two indices are shown together in Figures 4, 7, and 9 as normalised time series, but their correlation structure, typical lag relationship, and the fraction of 26°N variance explained by changes at 55°N are never quantified across the ensemble.*

This conflation has consequences for the paper's framing. The discussion invokes AMOC collapse risk and tipping-point concerns — which relate to the basin-scale overturning — but the analysis characterises SPG-scale density overturning. A weakening of $\psi_\sigma^{\max}(55°N)$ does not necessarily imply a corresponding change at 26°N on the same timescale, particularly given that southward propagation of density anomalies from the subpolar gyre to the subtropics involves considerable transit time and mixing. The paper should quantify the relationship between the two indices, or restrict its conclusions and framing to the SPG overturning specifically.

Author's reply:

We will clarify in the text that $\psi_\sigma^{\max}(55°\text{N})$ characterises SPG-scale water mass transformation and is physically distinct from the subtropical AMOC at 26°N, and we will adjust our conclusions and framing accordingly. We will examine the lead-lag relationship between these two metrics in more detail, performing this analysis both on the composites and at the level of individual ensemble members within each composite, to better understand this link.

Comment RC3.6

Bistability framing is not supported by the analysis

The discussion invokes bistability (Born et al., 2013; Li and Born, 2019; Swingedouw et al., 2021) to suggest that the circulation shifts identified here imply a risk of future irreversible change. This framing is physically incorrect for the events shown. Bistability requires the existence of two stable equilibrium states separated by a potential energy barrier, with transitions between them exhibiting hysteresis: the system does not return to the original state upon small reversal of the forcing parameter. The shifts identified in this paper are transient variability events. In all three categories, the composite overturning strength either recovers toward the pre-shift baseline or continues on a forced trend within the composite window. There is no evidence in the paper of hysteresis, of threshold sensitivity to the forcing magnitude, or of irreversibility in individual ensemble members.

Citing bistability strategies in connection with transient variability events risks misrepresentation of both the results of those studies and the findings of the present paper.

Author's reply:

The manuscript does not claim to demonstrate bistability. The mechanisms we describe, for instance the SPG positive feedback in the increase shifts, share dynamical similarities with those discussed in Born et al. (2013), Li and Born (2019), and Swingedouw et al. (2021), and we would like to retain these citations for that reason. However, the shifts themselves are transient variability events rather than transitions between stable equilibria, and we do not find evidence of hysteresis or irreversibility in our results. We will state this distinction explicitly.

Comment RC3.7

> ***Duplicate phrase (page 15, lines 280–281).*** *The phrase "across the SPNA" appears twice in immediate succession at the end of the paragraph beginning "This warming is coupled...". One instance should be deleted.*

Author's reply:

Thank you for catching this, we will correct it.

Comment RC3.8

> ***Reference period for Figure 11.*** *Anomalies in Figure 11 are computed relative to the 1980–2025 mean — a period spanning the AMOC peak, the aerosol-to-GHG transition, and the onset of the strongest forced AMOC decline. Normalising a 175-year record against this interval artificially distorts both the early-historical baseline (which appears anomalously different from modern conditions) and the interpretation of the thermal and haline contributions. A preindustrial reference period (e.g. 1850–1900) would be physically better motivated and more consistent with the pre-shift reference used in the composites.*

Author's reply:

We thank the referee for pointing this out. The caption contains an error: anomalies in Figure 11 are in fact computed relative to the 1850–2025 mean, not the 1980–2025 mean as stated. In addition, we agree that a preindustrial reference period (1850–1900) is physically better motivated and more consistent with the pre-shift reference used in the composites, and will adopt this reference period in the revised figure.

Comment RC3.9

> ***Freshwater budget and sea ice dismissal.*** *Section 2 states that sea ice area "did not provide further insight for the present analysis" and was excluded. Given that the post-1985 weakening events feature persistent SPNA freshening that prevents overturning recovery, and that sea-ice area in the Labrador and Irminger Seas is directly linked to surface buoyancy fluxes, this dismissal requires quantitative justification. A brief Supplementary figure showing the sea-ice composite anomaly for post-1985 events, with an explanation of why it does not affect the conclusions, would be appropriate.*

Author's reply:

We agree and will add a supplementary figure showing the sea-ice composite anomaly for post-1985 events, along with a discussion of why it does not affect our conclusions.

References

Born, A. and Mignot, J.: Dynamics of decadal variability in the Atlantic subpolar gyre: a stochastically forced oscillator, Clim. Dynam., 39, 461–474, https://doi.org/10.1007/s00382-011-1180-4, 2012.

Born, A., Stocker, T. F., Raible, C. C., and Levermann, A.: Is the Atlantic subpolar gyre bistable in comprehensive coupled climate models?, Clim. Dynam., 40, 2993–3007, https://doi.org/10.1007/s00382-012-1525-7, 2013.

Eden, C. and Willebrand, J.: Mechanism of interannual to decadal variability of the North Atlantic circulation, J. Climate, 14, 2266–2280, https://doi.org/10.1175/1520-0442(2001)014¡2266:MOITDV¿2.0.CO;2, 2001.

Falkena, S. K. J., Dijkstra, H. A., and Von der Heydt, A. S.: Causal mechanisms of subpolar gyre variability in CMIP6 models, Earth Syst. Dynam., 16, 1833–1844, https://doi.org/10.5194/esd-16-1833-2025, 2025.

Kwon, Y.-O. and Frankignoul, C.: Mechanisms of multidecadal Atlantic Meridional Overturning Circulation variability diagnosed in depth versus density space, J. Climate, 27, 9359–9376, https://doi.org/10.1175/JCLI-D-14-00228.1, 2014.

Latif, M., Sun, J., Visbeck, M., and Hadi Bordbar, M.: Natural variability has dominated Atlantic Meridional Overturning Circulation since 1900, Nat. Clim. Chang., 12, 455–460, https://doi.org/10.1038/s41558-022-01342-4, 2022.

Li, C. and Born, A.: Coupled atmosphere-ice-ocean dynamics in Dansgaard–Oeschger events, Quat. Sci. Rev., 203, 1–20, https://doi.org/10.1016/j.quascirev.2018.10.031, 2019.

Menary, M. B., Robson, J., Allan, R. P., et al.: Aerosol-forced AMOC changes in CMIP6 historical simulations, Geophys. Res. Lett., 47, e2020GL088166, https://doi.org/10.1029/2020GL088166, 2020.

Robson, J., Menary, M. B., Sutton, R. T., et al.: The role of anthropogenic aerosol forcing in the 1850–1985 strengthening of the AMOC in CMIP6 historical simulations, J. Climate, 35, 3243–3263, https://doi.org/10.1175/JCLI-D-22-0124.1, 2022.

Rühs, S., Oliver, E. C. J., Biastoch, A., et al.: Changing spatial patterns of deep convection in the subpolar North Atlantic, J. Geophys. Res. Oceans, 126, e2021JC017245, https://doi.org/10.1029/2021JC017245, 2021.

Swingedouw, D., Bily, A., Esquerdo, C., et al.: On the risk of abrupt changes in the North Atlantic subpolar gyre in CMIP6 models, Ann. N.Y. Acad. Sci., 1504, 187–201, https://doi.org/10.1111/nyas.14659, 2021.

Tulloch, R. and Marshall, J.: Exploring mechanisms of variability and predictability of Atlantic Meridional Overturning Circulation in two coupled climate models, J. Climate, 25, 4067–4080, https://doi.org/10.1175/JCLI-D-11-00460.1, 2012.