Author's response to Referee #2

egusphere-2026-2588

This study investigates the processes associated with multidecadal shifts toward strengthened and weakened AMOC phases using a 100-member ensemble from CESM2 simulations. Based on composites of hydrographic properties, circulation, and air–sea heat flux changes, the authors identify one class of strengthening events and two classes of weakening events. The interpretation of strengthening events—positive feedback between surface density and overturning, triggered by initial surface heat loss—is plausible. However, I am not so convinced by some arguments. Overall, the analysis is logically organized and offers useful insights into mechanisms driving multidecadal AMOC variability. Several conclusions would benefit from stronger supporting evidence and clearer explanations. My main comments are below.

We thank the referee for this careful and constructive review, which raises several important points that will strengthen the manuscript's supporting evidence and clarity. We take the referee's reservations about some of our arguments for the weakening mechanisms seriously, and address these below with additional metrics and clarifications. Below, we respond to each point individually: referee comments are shown in italics, followed by our reply.

Comment RC2.1

Fig. 2b: Circulation-shift occurrences are labeled every 5 years. How are "years" defined? Where does the 20-year mean window start, and why are labels shown every 5 years?

Author's reply:

We will clarify in the caption that "years" refers to the model year in which each shift is detected, with occurrences grouped into 5-year bins for readability, and that the 20-year window starts at that detected shift year.

Comment RC2.2

Equations (1)–(3): Several points need clarification:

- *Eq. (1): The upper density limit appears to correspond to the densest class. Please either adjust the sign convention or clarify that this is the density at which the streamfunction reaches its maximum.*
- *Eq. (2): The overturning streamfunction depth range appears to be integrated from bottom to surface. Is this intentional or a typo?*

- *Eq. (3): The barotropic streamfunction is defined via zonal accumulation of meridional transport (west to east). How should this be interpreted for the SPNA, where flow is not simply bounded west-to-east? How is the spatial map (e.g., Fig. 3c) constructed? If this is effectively a vertically integrated transport at each grid point, please state so explicitly. Also define what is meant by the "minimum barotropic streamfunction within the SPNA box" (l. 119).*

Author's reply:

We clarify each point below.

- **Eq. (1):** The upper limit $\sigma_{\max}$ denotes the densest density class included in the integration range, not the density at which the streamfunction reaches its maximum.
- **Eq. (2):** The bottom-to-surface integration is intentional. We exclude the upper 100 m of the water column from this integral specifically to remove surface-trapped overturning cells. We will state this explicitly in the revised Methods.
- **Eq. (3):** The barotropic streamfunction is computed at each grid point by vertically integrating the meridional velocity v over depth, then accumulating this vertically integrated transport along the x-direction of the model's native grid, from the western boundary eastward. This produces the two-dimensional spatial field shown in Fig. 3c. We use the SPNA box, defined in Fig. 1, as a spatial mask over this field: at each time step, we take the minimum value of the streamfunction within the SPNA box to obtain the scalar metric $\psi_b^{\min}$(SPNA). We will state all of this explicitly in the revised text.

Comment RC2.3

Do you use March alone to represent winter conditions? Please justify this choice and indicate whether results differ when using a seasonal mean (e.g., December–March).

Author's reply:

We use the December–March mean throughout, not March alone. We will revise the text to state explicitly that the winter mean is computed over December to March, and that March is highlighted separately only because the climatological mixed layer depth in the Labrador Sea peaks during this month.

Comment RC2.4

Please clarify whether potential density is averaged over the upper 500 m or over the full water column (l. 114–116, l. 124, Fig. 4).

Author's reply:

Potential density is treated with two different depth conventions in the manuscript, and this was not made sufficiently clear. In the composite anomaly maps (Figs. 3, 6, 8), density is averaged over the upper 500 m. In the SPNA-box timeseries (Fig. 4 and related figures), density and its thermal and haline contributions are integrated over the full water column. We will state both conventions explicitly alongside the relevant figures to avoid this ambiguity.

Comment RC2.5

> *If the AMOC metric is defined in density space at 55°N, why is northward heat transport at 45°N used for process analysis? Using consistent latitudes may improve interpretability.*

Author's reply:

Northward heat transport at 45°N is used because this latitude lies at the boundary between the subpolar and subtropical gyres, making it well suited to characterise inter-gyre exchange, which is central to the mechanisms described in this study. We have also examined northward salt transport at 45°N and find it behaves consistently with the heat transport in the cases analysed here. We will clarify this rationale in the revised methodology.

Comment RC2.6

> *The pre-1985 weakening group includes only 12 cases (l. 128), fewer than the other groups, and the weakening signal appears weak (Fig. 11). Please discuss robustness and whether including this group is physically justified.*

Author's reply:

We share the concern the referee raises here. We plan to add several statistical tests to both assess the pre-1985 weakening group's robustness and more clearly define the limit of what can be concluded based on our limited dataset. We propose the following.

As suggested by referee 3, 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 test speaks more directly to whether these are genuine events rather than sampling noise.

Additionally, we have 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 RC2.7

> *AMOC at 26°N is introduced without sufficient explanation. Does it vary in phase with 55°N, and what is its role in the argument?*

Author's reply:

The maximum overturning streamfunction in depth space at 26°N, $\psi_d^{\max}(26°\text{N})$, is included alongside our primary metric, $\psi_\sigma^{\max}(55°\text{N})$, in Figs. 4, 7, and 9 to relate our SPG-scale density overturning to the subtropical AMOC conventionally measured at 26°N. 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 RC2.8

Shifts toward increased overturning

- *Phase relation: The analysis assumes AMOC strength is controlled by density anomalies. However, density and AMOC do not appear fully in phase (Fig. 4; l. 150–152; l. 161–162). How do you explain continued AMOC strengthening after density anomalies stop increasing?*
- *Stratification evidence: In Fig. 5, positive density anomalies are widespread in the upper ocean, which does not necessarily imply reduced stratification. It would be more convincing to show buoyancy frequency (N^2) or vertical density anomaly differences (surface minus subsurface) as a direct stratification metric. Likewise, reduced full-column density anomaly alone is not sufficient evidence for weakened convection.*
- *Dynamical linkage: Please elaborate on the mechanism linking deep convection, baroclinic flow anomalies, and strengthened horizontal circulation (l. 144–145).*

Author's reply:

Phase relation: We will discuss this further in the revised text; a lag between density changes and the circulation response is a plausible contributor, but we will examine this more closely before finalising our explanation.

Stratification evidence: We will examine buoyancy frequency (N^2) and vertical density anomaly differences as direct stratification metrics, alongside the full-column density anomaly already shown, and include it in the revised manuscript if it proves informative.

Dynamical linkage: We will include a more detailed discussion of the mechanism linking deep convection, baroclinic flow anomalies, and strengthened horizontal circulation.

Comment RC2.9

Shifts toward decreased overturning (pre-1985)

- *Meridional density gradients: For l. 174–175, please provide explicit evidence for reduced meridional density gradients, since this is presented as the key mechanism. Showing the relevant density anomalies earlier (currently in Fig. 10) may help.*
- *Initiation of heat-transport decline: At l. 177, if gradual decline in northward heat transport drives density and AMOC decreases, what initiates the heat-transport decline?*

Author's reply:

On the meridional density gradients: we will introduce an explicit metric for the meridional density gradient between the subpolar and subtropical gyres and show it earlier in the results, alongside the density anomalies currently shown only in Fig. 10.

On the initiation of the heat-transport decline: we attribute this to internal ocean variability, consistent with the mechanism described by te Raa and Dijkstra (2002), rather than to an externally forced trigger. We will state this explicitly in the revised text.

te Raa, L. A. and Dijkstra, H. A.: Instability of the thermohaline ocean circulation on interdecadal timescales, J. Phys. Oceanogr., 32, 138–160, 2002.

Comment RC2.10

> *Comparison of mechanisms Please make the comparison more explicit and process-based. It would help to clearly state feedback loops and their signs (surface heat loss/gain, horizontal transport strengthening/weakening, overturning response, and thermohaline contributions).*

Author's reply:

We will make the comparison more explicit and process-based in Section 3.4, stating each mechanism as a signed feedback chain.

Comment RC2.11

> *Potential inconsistency (l. 236–241) Fig. 11 suggests density anomalies are mostly salinity-dominated, which appears inconsistent with some composite interpretations (e.g., Figs. 4, 7). Please clarify this apparent discrepancy*

Author's reply:

This apparent discrepancy stems in part from an typo in the Figure 11 caption, which we will correct, and we will also change the reference period for Figure 11 to 1850–1900 to further clarify the figure. More substantively, Figure 11 tracks the ensemble-mean historical evolution of density and its thermal and haline contributions across the full 175-year record, including the periods associated with each shift category, whereas Figs. 4 and 7 show composite anomalies relative to the 40-year pre-shift baseline for individual strengthening and pre-1985 weakening events respectively. The salinity dominance in Fig. 11 during the strengthening periods is therefore consistent with, not contradictory to, the composite mechanism described for those same shifts: haline changes there also drive the density increase, but Fig. 11 shows this as a long-term ensemble-mean trend rather than an event-centred anomaly. We will clarify this distinction in the text.

Comment RC2.12

> *l. 101–102: "around 1895 and 1955" → "starting around 1895 and around 1955"; "from the 1970s onwards" → "from the 1980s onwards"*

Author's reply:

We will make these wording changes.

Comment RC2.13

> *l. 147: Please use consistent terminology throughout (e.g., "strengthening shifts" vs. "increased shifts"; see also l. 127 and l. 140).*

Author's reply:

We will use consistent terminology throughout the manuscript.

Comment RC2.14

> *Fig. 3: Please add latitude/longitude labels for readability, and clarify the year indexing (including what "year 25–35" means).*

Author's reply:

We will add latitude/longitude labels to Fig. 3 for readability, and clarify in the caption that "year" refers to time since the start of the 40-year pre-shift composite period, with the detected shift occurring at year 40.

Comment RC2.15

> *l. 197–198: The deepest anomaly appears first in the central SPNA and Labrador Sea, rather than in the Irminger Sea.*

Author's reply:

We agree the current text oversimplifies the picture, and will describe this in more detail.

Comment RC2.16

> *l. 199–200: The salinity-chain argument seems somewhat ad hoc; please clarify the physical linkage between convection and SPG strength.*

Author's reply:

We agree and will clarify the physical linkage between convection and SPG strength.