

I first thank Mayer et al. for the continued and constructive exchange on this topic. Methodological discussions like this one are exactly what is needed to make progress on a genuinely difficult question, and I value the chance to work through these points carefully together. I think it is worth saying upfront that we agree on most of this, and that this discussion, however detailed, appears to not change the overall picture in the end. If I am not wrong, neither our reconstructions from raw reanalysis fluxes, nor the adjusted fluxes from Sohail and Zika (2025), the inferred and adjusted fluxes from Pan et al. (2026) (their Fig. 9c), or the inferred fluxes from MERRA2 and ERA5 that I had seen in personal communication, appear to suggest an AMOC decline at the moment. The length and detail of this reply reflect how much I value the exchange and how seriously I take the points raised, not the degree of disagreement between us, which I think is actually rather small.

I thank Mayer et al. for highlighting the Pan et al. (2026) study, which does look very promising and exciting indeed. It is encouraging that the inferred flux apparently requires only a small correction, such that the choice between a uniform and a regional adjustment makes little difference for that particular product. The similarity between global and regional adjustment means that either global biases are much smaller in the inferred fluxes than in the raw fluxes from reanalysis products (Sohail and Zika, 2025) or that North Atlantic biases are much larger in the inferred fluxes. Both processes would lead to better agreement between global and regional adjustments. As Mayer et al. have written that inferred fluxes appear to be better constrained, this appears to suggest that inferred fluxes work much better globally, especially in the Southern Ocean and tropical oceans, where direct fluxes appear to be highly uncertain according to Sohail and Zika (2025). If I understood it correctly, the globally more accurate fluxes in Pan et al. (2026) are indeed a great and important advancement and it is great to see these fluxes used to provide AMOC reconstructions.

This similarity between a global and a regional adjustment for the inferred flux does not, however, in my view, justify applying a globally uniform adjustment to the direct reanalysis fluxes used in Terhaar et al. (2025) or in Mayer et al. (2022). It remains unclear to me why a globally uniform adjustment would work well for the direct fluxes from reanalysis products specifically, given that these fluxes appear to have much larger problems in the Southern Ocean and the tropics than in the North Atlantic. A correction designed to remove a global mean bias that is largely driven by errors concentrated in those other regions will, when applied uniformly everywhere, effectively import the Southern Ocean and tropical problems into the North Atlantic, where the actual bias appears to be considerably smaller. Thus, to apply a globally uniform adjustment, one would have to show that the necessary adjustment is indeed globally uniform. Even for the inferred flux adjustment, a similarity between globally uniform adjustments and regional adjustments can arise by chance and is not proof that this methodical choice is okay, although it may well be, as explained above.

If it is indeed the inferred and adjusted flux that should be more robust than raw reanalysis fluxes in the North Atlantic, I would also note that Pan et al.'s own Fig. 9c points toward a fairly stable AMOC with decadal variability, or at least a stable MHT, which is broadly consistent with our own reconstruction. Treating AMOC and MHT as broadly interchangeable here is plausible in part because the direct contribution of water mass warming, that is, of ocean heat content changes, to

MHT changes has historically been small in the North Atlantic, both in observations and as represented in CMIP6 models. It is thus not contradictory to what we conclude in our 2025 study. I am genuinely excited to see if this method proves accurate going further into the past (although I have doubts about the data quality, see comments below) and to see what reconstructions it will yield.

The point that OHC changes are small in the North Atlantic compared to air-sea heat flux changes, is also relevant to the question of which adjustment for the direct fluxes is correct. I agree that Sohail and Zika do not settle this, but they do show that a globally uniform adjustment is roughly two to three times larger than the regional one they find appropriate for the North Atlantic. This gap between globally uniform and regional adjustment appears to be far too large to be explained by uncertainty in their underlying OHC estimates alone. On Mayer et al.'s related concern, that changing subsurface data coverage before Argo could introduce a spurious trend into that OHC based adjustment: OHC changes are much smaller than air-sea heat flux changes, both in CMIP6 models and in observations, and while there is no clear trend in the observed air-sea heat fluxes themselves, their decadal variability is large compared to that of OHC. The observed OHC change is, in any case, generally increasing, as physically expected from ocean warming, so any spurious component would only change the magnitude of that increase and have to be considerably smaller than this real signal. Since the OHC signal is already small relative to the air-sea heat flux changes, even a spurious trend within it would make only a small contribution to a trend in the adjusted fluxes. Matching OHC changes regionally to air-sea heat fluxes, as Sohail and Zika do, therefore appears not to be much affected by potential spurious OHC trends and appears to be preferable to globally uniform adjustments that are 2 to 3 times larger (estimated from supplementary data from Sohail and Zika, 2025). Thus, a regionally informed adjustment along these lines, even an imperfect one, appears to remain preferable to a globally uniform one for the reasons just given.

On the comparison with RAPID, I think there is a logical issue worth spelling out. We do not include OHC changes in our reconstruction, and OHC changes are precisely what drives short term, interannual, AMOC variability. We would therefore not expect a high  $r^2$  against RAPID over a record this short, effectively under 20 years once the edges affected by the 3-year running mean are trimmed. What matters more is whether RAPID falls within the uncertainty range of our estimate, as this uncertainty is mainly representing the interannual variability from OHC that we do not cover in our reconstruction, and it does: the RAPID observations lie within plus or minus 1 sigma most of the time and always within plus or minus 2 sigma, exactly as expected given that these intervals are designed to cover 68 percent and 95 percent of the variability respectively. We have separately tested, within CMIP6 models, whether adding the models' own simulated OHC changes alongside their simulated air-sea heat fluxes allows for a much better annual reconstruction of the simulated AMOC, and we find that it does. This confirms that even perfect estimates of air-sea heat flux alone cannot capture short term AMOC variability in CMIP6 models, and suggests that the strong interannual skill of the inferred flux estimates against RAPID likely comes via an implicit or explicit OHC term in that approach, rather than from the surface flux itself being a fundamentally better proxy for short term variability.

I should clarify why we left OHC changes out of our own reconstruction. Our focus is mainly on the period since 1960, precisely because this is the period for which we do not have RAPID measurements. Our goal was not to produce the best possible annual reconstruction of the last 20 years, since for that period we already have direct observations from RAPID, but to extend understanding of AMOC variability consistently further back in time, into the period where no direct observations exist and where annual, observed OHC changes are potentially not good enough to use. To keep the time series methodologically consistent, we excluded OHC for the full record rather than only for the years before Argo. This does not mean that our reconstruction over the last 20 years is worse than other reconstructions, it simply means that it does not include the short term variability but the longer timescales, including the magnitude, averaged over 3 to 10 years, the longer timescale variability, and the overall trend, and that the AMOC should be within the range of the given uncertainty accounting for not representing the interannual variability; and it does. With hindsight, I agree it would have been valuable to also show a comparison to recent conditions including OHC changes, if only as an additional check, but we only realized the extent to which adding OHC improves the annual reconstruction in CMIP6 models after our paper was already submitted. I also want to be transparent and avoid any confusion here: the test showing that OHC plus air-sea heat flux can reconstruct the AMOC annually was carried out entirely within CMIP6 models, using simulated AMOC, simulated flux, and simulated OHC. It is not yet a demonstration that this works with observed OHC. Extending this approach to the observational record over the last two decades, and into the future, is something we are now actively pursuing and that I would have added to this discussion if I had a few more days time.

I still think the agreement between ERA5, JRA 55, and COREv2 in the North Atlantic carries real value, and importantly, this is not instead of agreement with independent observations but alongside it: our reconstruction also agrees well with the RAPID array over the overlapping period and with the hydrography based reconstruction of Fraser and Cunningham (2021), and it captures known features such as the Great Salinity Anomaly. If there were a genuine drift specific to the North Atlantic affecting these direct fluxes, I would expect it to differ across the three reanalysis products, given how differently they drift at the global scale. The fact that it does not, together with the agreement with independent observations just mentioned, argues against a globally uniform correction for direct reanalysis flux estimates and for confidence in the anomaly estimated from these air-sea heat fluxes, while still not forgetting the uncertainties involved.

I want to highlight again that I am highly impressed by the performance of the Pan et al. (2026) reconstruction over the RAPID period, though I do wonder whether it performs as well before that period. I want to be upfront that I have not yet had time to read Pan et al. (2026) as closely as it deserves, simply because the comment containing the reference reached me only last night, and the discussion period here is closing very soon. I do intend to read it carefully once there is time and may revise my view accordingly. That said, I want to flag a concern based on what I have seen so far. The inferred flux based MHT in Pan et al. shows its strongest positive anomalies between 1987 and 1997, right at the end of the Great Salinity Anomaly. The expectation of a weak AMOC during and shortly after that period is supported not only by the proxy reconstructions discussed here,

including our own, which agree well with one another for that interval, but also by the underlying physical mechanism: the large freshwater input and resulting low salinity anomaly associated with the Great Salinity Anomaly would have weakened deep convection and vertical mixing in the subpolar North Atlantic, which is a mechanism by which such an anomaly weakens the AMOC. I would be curious to understand how the high MHT values found in Pan et al. for that window can be reconciled with both the proxy evidence and this mechanism, and whether this points to the inferred flux approach being less well constrained further back in time, perhaps because changes in the observations needed to estimate these inferred fluxes could themselves affect the time series as one goes further into the past.

To close, I think it is important to highlight that the agreement here is substantially larger than the disagreement. While we do not agree on how certain or uncertain direct air-sea heat fluxes are, and have had a lively and constructive discussion about exactly why, I think we agree on essentially everything else.

It is worth recalling briefly why I remain reasonably confident in our own reconstruction specifically, since I think this may have gotten a little lost in the back and forth: the agreement we find between ERA5, JRA 55, and COREv2 in the North Atlantic, the further agreement with independent observations from RAPID and from the hydrography based reconstruction of Fraser and Cunningham (2021), the fact that our reconstruction captures known features such as the Great Salinity Anomaly, and the demonstrated decadal scale tracking of the AMOC by air-sea heat fluxes more generally (Terhaar et al., 2025), together give me confidence in the broad picture even while specific adjustment choices remain debated.

Improving these air-sea heat flux estimates further is especially important given their potential as a proxy for the AMOC. This appears especially important since SSTs remain limited by the strength of their calibration against AMOC in CMIP5 and CMIP6 models. This relationship that stays weak in both model ensembles even once the most aerosol sensitive models are excluded, as Menary et al. (2020) and our own work show. This is a structural limitation that cannot be removed and is unlikely to disappear with another model ensemble. For this reason, I think better constrained air-sea heat flux estimates, of the kind Mayer and colleagues are clearly pushing forward, look like the most promising way ahead for resolving this question, and I am genuinely impressed by the progress they have made on this over the past years, and by how much it has helped the wider community to improve various aspects of this field, and also allowed me to better understand these flux estimates.

Most importantly, though, I think the finish here matters more than any of the individual points of disagreement above: none of the time series discussed in this exchange, not ours and not Pan et al.'s, appears to show a decline in any robust way. What both show, instead, is mostly decadal variability, and whether the result reads as a decline or an increase depends quite sensitively on the choice of start and end points. I think this is worth stating plainly, since it directly supports the central conclusion of our own paper, Terhaar et al. (2025), that there is no robust evidence of an AMOC decline from air-sea heat fluxes, and at the same time it is a point in favour of using air-sea

heat fluxes to reconstruct the AMOC, both the inferred flux approach and direct approach, rather than against it: rather than undermining it, this convergence suggests that both lines of evidence, properly examined, are telling a broadly consistent story for now, even if the uncertainties on each side differ in their details. Uncertainties remain in both, and I am genuinely impressed by the effort going into improving these estimates further. As things stand, I do not think the evidence does not allow us to infer an AMOC decline, but absence of proof is not proof of absence, which is worth keeping in mind as this discussion continues.