This manuscript addresses a highly relevant and still unresolved question: whether the observed Atlantic Multidecadal Variability (AMV) since 1850 is primarily an internally generated mode of variability or largely an externally forced response. The paper argues that AMV-like variability is largely externally forced and that future North Atlantic SST evolution will therefore be driven mainly by anthropogenic forcing. A major strength of the paper is the use of several SMILEs together with a large CMIP6 ensemble, as well as the inclusion of observational uncertainty through different SST datasets. However, I find that the main conclusion is currently stronger than the supporting analysis. Overall, I think the manuscript has clear potential, but substantial revisions are needed. I therefore recommend major revisions. Specific comments are provided below:

Specific comments:

1) The separation of forced and internal variability is central to the paper. The manuscript appropriately acknowledges that AMV estimates are sensitive to methodological choices. However, the actual sensitivity analysis is limited to linear versus quadratic detrending, so the claim that the main conclusions are robust is currently demonstrated only within these two methods. The authors should either test at least one alternative estimate of the forced signal from a different methodological family, for example a regression-based approach using global-mean SST/ near-surface air temperature (SAT), or provide a stronger justification for why such alternatives are not considered appropriate here. At present, citing one paper regarding the limitations of regression-based methods is useful, but not enough on its own to establish methodological robustness.
2) The manuscript relies heavily on correlation-based comparisons. However, I could not find a significance framework. Given that the AMV indices are smoothed with a 10-year running mean, the effective number of independent samples is substantially reduced. The authors should therefore explicitly state how significance is assessed and account for serial autocorrelation when interpreting correlation coefficients. An effective-sample-size correction such as the Bretherton lag-1 autocorrelation adjustment (Bretherton et al., 1999) would be appropriate; alternatively, the robustness of the results could also be checked with a bootstrap or Monte Carlo approach that preserves low-frequency dependence. As it is now it is difficult to judge whether the reported differences in correlation across models and highlighted ensemble members are statistically significant.
3) The choice of the 1961–1990 climatology is not justified. The authors should justify this choice and discuss whether the results are sensitive to an alternative standard reference period (e.g., 1981 -2010).
4) The manuscript should clarify how much model resolution can realistically matter for the chosen diagnostic. Since all SST fields are remapped to a common 5° × 5° grid and then

reduced to a basin-mean NASST index, the potential advantages of higher resolution are not directly assessed.

5) A more complete evaluation of the AMV should include its spatial fingerprint, not only time-series correlations of indices. In the literature, the positive phase of AMV is associated with a characteristic Atlantic SST pattern, including warm anomalies across much of the North Atlantic and, in some studies, a North-Atlantic warm / South-Atlantic cool contrast, together with broader SAT responses over surrounding continents (e.g., Bellomo et al., 2018; Deser et al., 2010; Otterå et al., 2010; Ruprich-Robert et al., 2017). I suggest that the authors complement their index-based correlation. This could be done with pattern-correlations between regression maps of detrended Atlantic SST and/or SAT onto their AMV index. This quantification would allow a more physically grounded comparison and could substantially strengthen the paper.

Bellomo, K., Murphy, L. N., Cane, M. A., Clement, A. C., & Polvani, L. M. (2018). Historical forcings as main drivers of the Atlantic multidecadal variability in the CESM large ensemble. *Climate Dynamics*, *50*(9–10). https://doi.org/10.1007/s00382-017-3834-3

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., & Bladé, I. (1999). The effective number of spatial degrees of freedom of a time-varying field. *Journal of Climate*, *12*(7). https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2

Deser, C., Alexander, M. A., Xie, S.-P., & Phillips, A. S. (2010). Sea Surface Temperature Variability: Patterns and Mechanisms. *Annual Review of Marine Science*. https://doi.org/10.1146/annurev-marine-120408-151453

Otterå, O. H., Bentsen, M., Drange, H., & Suo, L. (2010). External forcing as a metronome for Atlantic multidecadal variability. *Nature Geoscience*, *3*(10). https://doi.org/10.1038/ngeo955

Ruprich-Robert, Y., Msadek, R., Castruccio, F., Yeager, S., Delworth, T., & Danabasoglu, G. (2017). Assessing the climate impacts of the observed atlantic multidecadal variability using the GFDL CM2.1 and NCAR CESM1 global coupled models. *Journal of Climate*, *30*(8). https://doi.org/10.1175/JCLI-D-16-0127.1