This short paper presents an inter-comparison of transports in correspondence with the GSR section from several state-of-the-art reanalyses and simulations. It also includes a more detailed analysis of the recent decrease in heat fluxes after 2017, with some explanations and a proposal for sea level-based indexes. As such, it is an interesting piece of work, useful for the oceanic community, and as a reference for future validation and process-oriented studies. I recommend publishing this work after the authors address a few general comments.

Thanks for the appreciative comments. Please see below for our responses in black. Before this, we would like to point out two changes that we plan to include in the revised manuscript:
1) We recently discovered an error in the computation of ORA-based transports. Correction of the error leads to generally improved correlations with observation. Correction of the error also leads to an increase of the mean Atlantic Water (AW) volume flux diagnosed from the lower resolution (GREP) products, i.e. their low bias in *total* AW influx is reduced. Nevertheless, the conclusions about higher resolution ORAs exhibiting a better *distribution* of inflow across the different branches still holds. We will revise the manuscript to reflect these changes.
2) Reviewer #2 criticized that differences between GLOB16 (the forced ocean run at 1/16 degree resolution without data assimilation) and lower-resolution ORAs were not necessarily attributable to resolution as we did not include a low-resolution control run for a clean comparison. To address this, we now include results from GLOB4 (a ¼ degree version of GLOB16). Yet, we will limit the discussion to their representation of volume fluxes (i.e. ocean circulation) to strengthen the conclusions about resolution-dependence of circulation in the ORAs. Heat fluxes from GLOB4 and GLOB16 will not be discussed as they exhibit temperature (and hence heat flux) biases (due to their nature of being non-assimilating runs) that detract from the main points of the study.

1) Table 2 and the figures are not consistent: the figures report all the GREP members while the Table only the GREP ensemble mean. I recommend the authors report the values of all GREP members and their ensemble mean. Furthermore, the table should contain more diagnostics and include those already presented in the text (temporal standard deviation, correlation values, etc.). This could be very useful for future quantitative studies.

We agree a more detailed display of metrics in table 2 could be useful as a reference for future studies. However, this expansion needs to be balanced with the space constraints for OSR contributions. We thus suggest to expand the table to display the averages for all GREP reanalyses separately, not only the GREP ensemble mean. We note that the correlation coefficients are provided in the time series plots and thus do not necessarily have to be replicated in the table.

2) The inclusion of GLOB16 is not sound. This is a free-running model experiment, at a higher resolution than the other products, and a corresponding assimilation experiment does not exist. In practice, its inclusion poses more questions and does not help to answer any question. Using an ensemble of reanalyses helps to identify the envelope of uncertainty; using GLORYS12 helps address the impact of spatial resolution. But it is not clear whether results from GLOB16 can be ascribed to different resolutions? different model configuration? lack of observational constraints? different initial conditions? I miss the point for including it.

To address this point, we assessed the ¼ deg counterpart of GLOB16, named GLOB4, which allows for a clear attribution of differences to resolution. Indeed, the higher resolution run exhibits a distribution of volume fluxes across the different branches crossing the GSR that are in better agreement with observations compared to GLOB4. We propose to only discuss

these runs in terms of volume fluxes / circulation, i.e. to strengthen the points regarding resolution of the ORAs. Heat transports from GLOB16 and GLOB4 will not be discussed, as their temperature biases lead to biased heat flux estimates.

3) Section 3.3 is not much conclusive either. One suggestion from the authors is that SPG strength induces a delayed response in the AW temperature. However, lagged correlations and similar tools could be used to provide a quantitative answer about the fitness of the SLA-based indexes for capturing the AW inflow variability.

Good suggestion. Cross-correlation analysis of AW temperature and the SPG strength seems to support our statement by showing maximum positive correlation when AW temperature lags by ~2-3 years, with maximum correlation coefficients between 0.3 and 0.5 (depending on the considered data set). However, these correlations are not statistically significant because of the high auto-correlation of the involved time series, which reduces the effective degrees of freedom. This result will be mentioned in the revised manuscript.

Minor point

In line 46: I suggest referring to "the so-called Arctic Mediterranean" as it might be non-obvious for a general readership

Agreed.