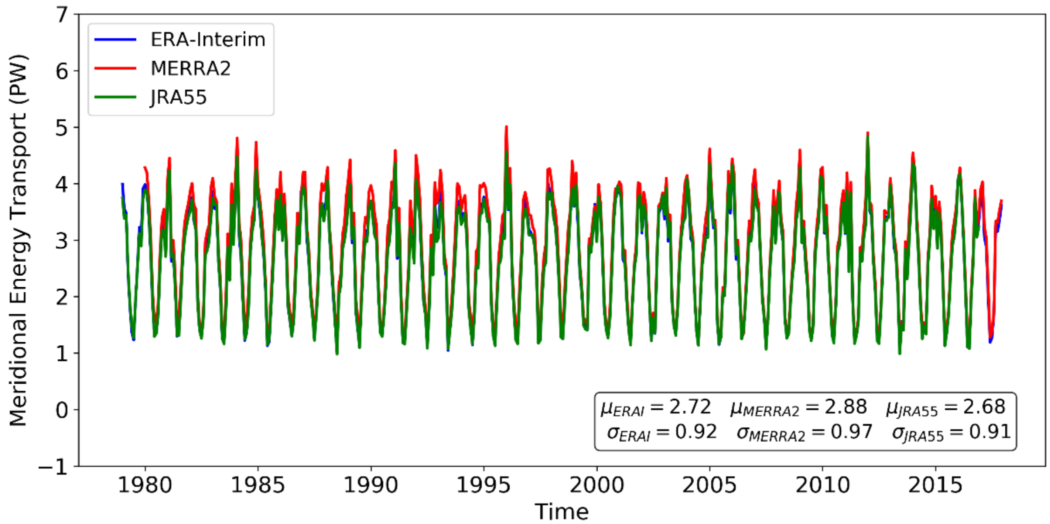
We are very glad to receive so many constructive comments and we would like to thank the reviewer for his help. We are able to address all points and will do our best to improve the quality of our paper accordingly. Below is our response to the reviewer's comments point by point:

(the *original comments* are given by *Italic gray text*, and each follows our response in plain text.)

**Response to the major comments**

*1) Figure 3a indicates that the annual cycle of AMET from ERA-Interim and JRA55 is very different, with yearly minima in JRA55 going down to ~0.8PW,while annual minima in ERA-Interim remain closer to 1.5PW, which implies a huge relative discrepancy (>50%) in the amplitude of the annual cycle of AMET from the two products. From our own computations using basically identical scripts for both products I can tell that their annual cycle of AMET actually agrees very well (within <10%). Also, low-pass-filtered variability of AMET looks quite different from our results compared to the authors' figure 3b. Thus, I presume there is something wrong in the authors' computation of AMET (at least for ERA-Interim and JRA55, I cannot judge the results from MERRA2), possibly in the mass adjustment that they apply. I recommend to thoroughly check the chain of computations. NCAR provides a quite detailed step-by-step instruction how to perform these computations: http://www.cgd.ucar.edu/cas/catalog/newbudgets/index.html*

After thoroughly checking our script and the whole computation procedure, we find that the discrepancy between AMET from ERA-Interim and JRA55 comes from the numerical scheme used in performing the mass correction. The barotropic mass correction method involves calculations of the divergence, the inverse Laplacian and a gradient (Trenberth, 1991). We used a finite difference (central scheme) method to compute these terms. However, after some tests we noticed that it is more accurate to bring the fields to the spectral domain and compute these terms via spherical harmonics. The adjustments to the barotropic wind as results of mass budget correction based on these two numerical methods differ much in the polar regions, in particular the amplitude. Since AMETs are very sensitive to the mass budget, the results with these two different numerical schemes can lead to very different results. After recomputing the AMET fluxes from ERA-Interim, MERRA2 and JRA55 with mass correction using spherical harmonics, now the difference between annual cycles of AMET from ERA-Interim and JRA55 is very small. The results are shown in the figure below. They are consistent with the reviewer's results. In our revised paper we will use the new numerical method for the diagnostics and update the results, figures and discussion.




The low frequency series of AMET from ERA-Interim, MERRA2 and JRA55 also changed due to the implementation of mass correction via computation with spherical harmonics. But as they are still very different, the major results in our paper remains the same. Changes are made to the regressions of different variables with AMET mainly.

Trenberth, K. E. (1991). Climate diagnostics from global analyses: Conservation of mass in ECMWF analyses. Journal of Climate, 4(7), 707-722.

*2) The authors note that there exist improved diagnostic equations for energy budget diagnostics (Mayer et al. 2017; Trenberth and Fasullo 2018), but do not use those. I strongly recommend to make use of these updated equations. There is no reason not to do so.*

Thanks for pointing out this. At the time the initial work for this manuscript was done these improved methods were not published yet. Now, together with the implementation of mass correction via spherical harmonics, we recomputed the AMET using the improved diagnostic equations for energy budget diagnostics (Equation 24 in Mayer et. al., 2017).

Mayer, M., Haimberger, L., Edwards, J. M., & Hyder, P. (2017). Toward consistent diagnostics of the coupled atmosphere and ocean energy budgets. Journal of Climate, 30(22), 9225-9246.

*3) Ocean energy budget (section 2.3.2) and discussion about reference temperature: The authors discuss the need for a reference temperature as long as the mass budget is not closed. This is indeed important and this issue has been extensively discussed in the oceanographic literature, most notably Schauer and Beszczynska-Möller (2009). However, the present study only considers oceanic transports in a zonally integrated sense. Full zonal cross-sections should have a net mass flux close to zero, making the use of a reference temperature unnecessary. The same applies for the statement about recirculation (p8 l13). In fact, for zonal integrals, there only is a small imbalance coming from P-E, leaving a small ambiguity, which in the same manner applies to the atmosphere. For that reason, the same reference temperature should be used for both atmosphere and ocean to obtain consistent results (Mayer et al. 2017). The discussion on these issues must be clarified.*

We agree with the reviewer that the reference temperature is not needed when the mass is perfectly balanced. In oceanographic literature it is common to use a reference temperature when calculating OMET in both observations and model diagnostics (e.g. Bryan, 1962; Hall and Bryden, 1982; Johns et. al., 2011; van der Linden et. al. 2019). As we cannot perform barotropic correction for the ocean, we take reference temperature as a compromise. In the ocean, with its strong boundary circulations even the smallest imbalance can lead to large errors in the heat flux. Besides, it is very hard to include all NEMO components derived from ORAS4 to close the budget. Thus, we think it is still safe to take a reference temperature. However, we agree with the reviewer that when taking the zonal integral the benefit from a reference temperature is not as large as when considering a single strait transports (Schauer and Beszczynska-Möller, 2009). To make the formulation much clear, we will modify our explanation about the usage of the reference temperature as well as the statement about recirculation. We will add Schauer and Beszczynska-Möller (2009) to our reference list and discuss the mass imbalance coming from P-E (Mayer et. al., 2017).

Bryan, K. (1962). Measurements of meridional heat transport by ocean currents. Journal of Geophysical Research, 67(9), 3403-3414.

Hall, M.M. and H. L. Bryden (1982) Direct estimates and mechanisms of ocean heat transport, Deep-Sea Research, Vol. 29, No. 3A, pp. 339 to 359

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., ... & Curry, R. (2011). Continuous, array-based estimates of Atlantic Ocean heat transport at 26.5 N. Journal of Climate, 24(10), 2429-2449.

Schauer, U., & Beszczynska-Möller, A. (2009). Problems with estimation and interpretation of oceanic heat transport–conceptual remarks for the case of Fram Strait in the Arctic Ocean. Ocean Science, 5(4), 487-494.

van der Linden, E. C., Le Bars, D., Bintanja, R., & Hazeleger, W. (2019). Oceanic heat transport into the Arctic under high and low CO2 forcing. Climate Dynamics, 1-18.

**Response to minor comments**

*Generally: the plural of “reanalysis” is “reanalyses”, while the plural of “reanalysis dataset” is “reanalysis data sets”. Please correct throughout the manuscript.*

We will correct these incorrect formulations.

*P2L16: Is this result based on models? Please clarify.*

Their study uses reanalyses data (Yang et. al., 2010). We will further specify the details of these studies that listed in this paragraph.

Yang, X. Y., Fyfe, J. C., & Flato, G. M. (2010). The role of poleward energy transport in Arctic temperature evolution. Geophysical Research Letters, 37(14).

*P2L24: In this context it might be worth mentioning that ocean reanalyses do not show a clear sign of Arctic amplification in Arctic OHC increases (Mayer et al. 2016; von Schuckmann et al. 2018)*

Thanks for the comment. We will include this point.

*P2L34: Please rewrite the sentence to something like: "These are representations of the historical state of the atmosphere and ocean optimally combining available observations and numerical simulations using data assimilation techniques."*

We will rewrite the sentence following your suggestion.

*P3L3: Please spell out this acronym (and all others).*

The acronym will be spelled out.

*P3L7: Please be more specific about the "model". Is this a forced ocean model run?*

Yes, this is an ocean model run forced by surface fields from the Drakkar Surface Forcing data set version 5.2.

*P3L23: "higher" than what?*

This is a bit confusing. Here we mean we choose reanalysis products with relatively high resolution (compared to old reanalyses) as the quantification of energy transport need this. We will change "higher" to "high", just to be accurate.

*P3L24: change "preferably" to "preferable"*

We will change it.

*P3L25: "For an inter-comparison purpose, they better not resemble each other". What is meant here exactly? Please reword.*

Thanks for pointing this out. The sentence is indeed very confusing. Here we want to emphasize that we try to include various reanalyses data sets to make the comparison more informative. We will delete this sentence.

*P4L7-8: Is there a reference for the statement about divergent winds? It might be worth checking Graversen et al. (2007)*

Berrisfold et. al. (2011) discussed the improvement of divergent winds in their work. We will include this reference now.

Berrisford, P., Kållberg, P., Kobayashi, S., Dee, D., Uppala, S., Simmons, A. J., ... & Sato, H. (2011). Atmospheric conservation properties in ERA‐Interim. Quarterly Journal of the Royal Meteorological Society, 137(659), 1381-1399.

*P4L14: add "scheme" after "assimilation".*

We will add it.

*P4L22: add "upper air" before "observations"*

We will add it.

*P4L26: oceans -> ocean's*

We will correct it.

*P4L30: data with 3D-Var assimilation -> analyses with a 3D-Var FGAT assimilation scheme*

We will change it.

*P5L9: Not quite right. The forcing is a combination of ERA-Interim fluxes (e.g. shortwave radiation) and bulk formulae using ERA-Interim near-surface parameters. Please correct.*

Thanks for the information. We will correct it.

*P5L16-17: "To be consistent with the other two reanalyses datasets assessed in this study, the SODA 3.4.1 is chosen since it applies surface forcing from ERA-Interim". This statement seems to be opposite of what you say above in P3L25*

We will delete the apparently confusing statement in P3L25.

*P5L18: r –> R*

We will correct it.

*P6L5: What is the "Drakkar forcing data" based on?*

Thanks for reminding. The NEMO ORCA run is forced by the Drakkar Surface Forcing data set version 5.2, which supplies surface air temperature, winds, humidity, surface radiative heat fluxes and precipitation, and a formulation that parameterizes the turbulent surface heat fluxes and is provided for the period 1958 to 2012 (Dussin et al., 2014; Brodeau et al., 2010).

Brodeau, L., B. Barnier, A. M. Treguier, T. Penduff, and S. Gulev (2010), An ERA40-based atmospheric forcing for global ocean circulationmodels, Ocean Modell., 31, 88–104, doi:10.1016/j.ocemod.2009.10.005.

Dussin, R., B. Barnier, and L. Brodeau (2014), The making of Drakkar forcing set DFS5, DRAKKAR/MyOcean Rep. 05-10-14, LGGE, Grenoble,France.

*P7L21: explain u_c and v_c*

These are the correction terms for zonal and meridional wind components as a result of barotropic mass budget correction. We will further explain this based on the explanation given by Trenberth (1991).

Trenberth, K. E. (1991). Climate diagnostics from global analyses: Conservation of mass in ECMWF analyses. Journal of Climate, 4(7), 707-722.

*P8L10: the equation gives OHC across a certain circle of latitude. Is this meaningful? How would that relate to the transports across that latitude? Did the authors mean OHC integrated across the area north of a given latitude?*

The equation here is a bit confusing. We agree with the reviewer that OHC integrated in the polar cap is likely closely related to OMET at 60N (the chosen reanalysis products are all forced by surface fields from ECMWF atmospheric reanalyses.). We will update the equation here and calculate the OHC in the polar cap.

*P8L11-12: I can reassure you that ocean reanalyses do have sources and sinks from temperature increments, but they do not suffer from mass inconsistencies as atmospheric reanalyses do. The divergence of ocean currents exactly balances the surface freshwater flux and local sea level variations, so there is no mass adjustment needed. Please see the NEMO documentation for details (Madec 2008).*

We update this and put an accurate formulation now. However, it is very hard to include all NEMO components derived from ORAS4 to close the budget. See also van den Berk et. al. (2019) for the freshwater budget.

van den Berk, J., Drijfhout, S. S., & Hazeleger, W. (2019). Atlantic salinity budget in response to Northern and Southern Hemisphere ice sheet discharge. Climate Dynamics, 52(9-10), 5249-5267.

*P9L3: Do you mean "decorrelation"?*

Yes, this is what we aim to do.

*P9L6: Do you mean "by a factor of 3"?*

Yes.

*P9L8: Why not do this the other way around? Apply filter first, and then estimate effective degrees of freedom.*

Statistically, it seems to be more strict to estimate the effective degrees of freedom after applying filter. We will apply the filter first and then estimate effective degrees of freedom.

*P9L10: I do not understand the statement about statistical significance*

Here we mean that the relatively short records of reanalyses do not have many samples at interannual time scales, compared to outputs from numerical climate models. We will rephrase this to make it more clear.

*P9L18: not only solar radiation, but also OLR. In that sense, transports balance NET radiation.*

Thanks for the comment. We will correct this.

*P9L26: How do you know this?*

Now we notice that the latitudinal variation is due to the mass correction based on finite difference scheme. We will correct this.

*P9L29: "ERA-Interim res" is not shown in the figure.*

This is a mistake. We will correct this now.

*P9L21: 1.21PW is probably too high (see major comment #1)*

We will update this according to the results with improved methods. Now it is of the same scale as ERA-Interim.

*P10L10: Here it is important to know whether you are using monthly or sub-monthly data, which should be stated in the methods section. If you use monthly means, you will miss eddy transports and consequently underestimate ocean heat transports.*

With ORAS4 and GLORYS2V3, we use monthly data since only the data at monthly scales are available. For SODA3, we use sub-monthly data (5 day averaged) (which includes the eddy components). We will further explain this now after including your comment.

*P10L12: What is plotted in Fig4? The poleward component of OMET? How is this obtained, if OMET is computed on the tripolar NEMO grids?*

It is the poleward component of energy transport in the ocean. We regrid from curvilinear grid to lat-lon grid as we only want to emphasize the resolution and its potential influence on our results. But it seems to be a bit confusing, thus we decided to remove it. Thanks for asking.

*P10L19: Is "hindcast" the appropriate term? Isn't it a forced model run?*

The model is forced by (near) surface fields from historical data only. It is common to call this a hindcast in oceanographic literature (indeed, this differs from atmospheric literature). But we agree that it is also "a forced model run". We have added this description in the method section.

*P11L1: correlation or regression?*

We check the correlation between the total AMET and each component. So it is "correlation".

*P11L5: r=0.07 seems very small (again, see major comment #1)*

We will update this according to the new results.

*P11L20: With these diagnostics, you are running into problems with reference temperature, as the point-wise transports obviously do not have a zero mass flux. Figure 7d (showing T_mean*delta_v) will have much larger values when you use K instead of C, leading to a different conclusion. Which one would be correct? I suggest to remove 7c and 7d and discuss a and b in more detail. Instead of 7c and d, it would be interesting to show the difference sections also for JRA55.*

Thanks for the point. We agree that we don't have zero mass flux in this point-wise set-up. We will keep only Figure 7a and 7b.

The reason that we only show this for ERA-Interim and MERRA2 is that these two reanalyses have points at the same location in terms of pressure levels, lat and lon on the close-to-native grids that they are produced with (those grids are slightly interpolated grids compared with native grids, e.g. for ERA-Interim TL255 is roughly equal to 0.75 x 0.75 deg, and MERRA2 natively 0.5 x 0.625 deg). We only choose the points at exactly the same location. However, to include JRA55 (TL319 roughly equal to 0.5625 x 0.5625 deg) a further interpolation is inevitable, which can introduce errors. So we only make such point-wise comparison between MERRA2 and ERA-Interim, as a compromise.

*P12L5: Please use a different name than "NEMO", as all your ocean products are based on NEMO.*

We will change the name to “OGCM hindcast”.

*P12L15ff: Similar to my above comment: I do not understand why you look at bands of OHC and not at OHC north of 60N.*

The explanation is given above. But we will switch to the OHC north of 60N.

*P13L9: Better use an independent SIC product. SIC from ERA-Interim is of questionable quality, which can be seen e.g. from the “disc” around the North Pole. Also, ORAS4 does not have an active sea ice model. Would you expect a correlation between OMET and SIC then?*

We wish to obtain a consistent picture by regressing OMET generated with models forced by ERA-Interim fields on sea ice also from ERA-Interim. Indeed, there is no active sea ice model, but for SIC the temperature criteria are assumed to be reasonable. Of course this is not the case for thickness or volume.

*P13L18: no -> not*

Corrected.

*P13L32: Fig 13: I am not sure what is shown here. Are these instantaneous regressions? The legend says 1-month lag - does this make sense when using 12-monthly smoothed data? At which lag do you get highest correlation?*

In this regression OMET leads the sea ice by one month. We observe the highest correlation between OMET and sea ice when OMET leads by one month. But the difference between correlations are very small, as long as OMET is leading.

*P14L1: I think one has to be careful with the timescales here. What timescales are the cited studies looking at?*

Thanks for the reminding. These studies focus on decadal to inter-decadal scales. We will add the timescales here.

*P14L25: How can you see this from the time series?*

The low frequency anomalies of OMET (Figure 5b) shows that GLORYS2V3 differs a lot to ORAS4 and SODA3. By saying this we want to emphasize that the differences in OMET are reflected in the regressions on sea ice. But this seems a bit confusing. We will rephrase this.

*P14L27: Be cautious: A lot of heat transported across 60N is stored in the North Atlantic or released from there through air-sea fluxes and will never reach sea-ice covered regions.*

Thanks for reminding. We will rephrase this part and mention that the relation is “indirect”.

*P14L28: “patterns” of what?*

“Horse shoe” or “dipole” patterns of the link between OMET and SST over the Atlantic. We will add this to the text.

*P14L31: remove “the” before “tropical”*

We will correct this.

*P15L8: "less consistent" than what?*

Sorry for the typo. We mean they are not consistent with each other. We change "less consistent" to "not consistent".

*P15L24-26: Hard to understand. Please rewrite the sentence.*

We mean the sink and source in reanalyses cause large uncertainties in the computation of energy transport. We will rephrase this part.

Again, we'd like to express our gratitude to the time that the reviewer spent on it. Thank you very much!

With best regards

Yang Liu, Jisk Attema, Ben Moat, and Wilco Hazeleger