Answer to Reviewer 3

We thank the reviewer for thorough reading and thoughtful comments and suggestions. A detailed discussion of the changes that we made in response to the reviewer's comments is given below. In what follows, we state the reviewer's comment in boldface, and describe our response in plain text. Text in the manuscript is represented in italics. The text that has been modified/included in the new version has been highlighted in red.

Review for esd-2020-54 "Water transport among the world ocean basins within the water".
In general, this study is interesting to me, with the global ocean basins to study water mass transport based on GRACE and ERA5. The conclusions are generally supported by the data analyses in this study, but more validations/evaluations are necessary to improve the reliability of the results. At least, a careful intercomparsion between this study and previous studies/literatures can be discussed to enhance our understanding.

We have included several improvements in the manuscript:

1. **Comparison with similar studies:** As far as we are applying a new methodology, there are not many studies to compare with. The only two studies, up to our knowledge, doing something similar are discussed in the third paragraph of the section "Discussion and Conclusions":

   "The results presented here are consistent with the well-known salinity asymmetry between the Pacific and Atlantic Oceans (Reid, 1953; Warren, 1983; Broecker et al., 1985; Zaucker et al., 1994; Rahmstorf, 1996; Emile-Geay et al., 2003; Lagerloef et al., 2008; Czaja, 2009; Reul, 2014). However, they are in contrast to previous GRACE-based studies where a simple seesaw WT between the Pacific and the Atlantic/Indian oceans was reported [Chambers and Willis, 2009; Wouters et al., 2014]. In those studies, the P−E+R term in Equation 2 in both Pacific and Atlantic/Indian Oceans was approximated by that from the global ocean mean. However, the mean freshwater flux in the Pacific (1261 Gt/month) quite mismatches that in the Atlantic/Indian Oceans (−1837 Gt/month), meaning that the approximation was quite poor and hence the N term was not properly estimated in these studies (see Appendix for further discussion)."

   As stated in the text, in the Appendix we explain in details why the proposed methodology overcomes some important limitations of previous approaches which will always show a seesaw of water transport, even if it does not exist.

2. **Other datasets:**

   The *P*, *E*, *P*−*E*, and *R* components are auxiliar in this study. However, we have added some more references for comparison purposes in the last paragraph of Section 3.1:

"Corresponding analyses have been performed for the Atlantic, Indian, and Arctic Oceans separately. The time evolution of the WT components in Eqs. 1 and 2 are shown in Figure 4, and a diagram of the water-mass fluxes is shown in Figure 5. On average, the Atlantic Ocean receives 926 Gt/month (CI_{95}=[876, 980]; or 0.36 Sv) of salty water, and loses to the atmosphere 879 Gt/month (CI_{95}=[828, 930]) via P−E+R. The latter is equivalent to a freshwater deficit of 0.34 Sv, which increases the near-surface salt concentration and enables water to sink in North Atlantic producing deep water. These values are close to the 0.13-0.32 Sv estimated from ocean models, as needed to keep salinity balance in the Atlantic Ocean (Zaucker et al., 1994). Similarly, the Indian Ocean loses 957 Gt/month (CI_{95}=[894, 1022]) of freshwater that is restored by 991 Gt/month (CI_{95}=[907, 1073]) of salty water. The freshwater lost via P−E+R by the Atlantic and Indian Oceans goes to the Pacific (1261 Gt/month, CI_{95}=[1171, 1347]) and Arctic (730 Gt/month, CI_{95}=[712, 747]) Oceans, which return 1194 (CI_{95}=[1096, 1291]) and 723 (CI_{95}=[708, 739]) Gt/month of salty water through the ocean, respectively. Then, the Pacific presents a surplus of freshwater that reduces near-surface salt concentration, which prevents the formation of deep water. Together, the Pacific and Arctic Oceans supply 1917 Gt/month (CI_{95}=[1812, 2021]) of water to the Atlantic and Indian Oceans, where it is reincorporated into the water cycle via net E−P. As in previous studies (see Craig et al., 2017 for a synthesis), the freshwater lost in the Indian Ocean is similar to that in the Atlantic Ocean. In those studies, P-E+R is close to zero in the Pacific Ocean, producing a difference of 0.4 Sv between Atlantic and Pacific Oceans. In this study, P-E+R is 1261 Gt/month in the Pacific Ocean and the difference with the Atlantic increases to ~0.8 Sv. Some of these differences would be expected as far as the ocean basins are not defined in exactly the same way. On the other hand, the global R is 3781 Gt/month (or 3781 $\times$ 12 = 45368 km^3/year), close to the 41867 km^3/year reported by the Global Runoff Data Centre (GRDC, 2014). At basin scale, R is 16834 km^3/year in the Pacific, greater than the 11826 km^3/year reported by GRDC. In the Atlantic, Indian, and Arctic, R is 18228, 4479, and 5827 km^3/year, respectively, which is closer to the GRDC values: 20772, 5238, and 4080 km^3/year. Finally, according to the diagram in Figure 5, the water content in the atmosphere decreases 178 Gt/month (and it is gained by Earth's surface), but this amount is not realistic as discussed in Section 2 since it should increase a few Gt/month [Nilsson and Elgered, 2008]. This value differs from the 188 Gt/month mentioned in Section 2 because the endorheic regions are not accounted here."

More importantly, we have extended our analysis to other datasets. The objective is to show that our main results concerning the *N component*, are not an artifact of CSR GRACE and ERA5 datasets. As a result, there is new section entitled "Comparison with other datasets":

"Equations 1 and 2 are applied to estimate the Pacific outflow using different datasets:

(1)CSR GRACE mascon solution is replaced by the JPL GRACE mascon solution provided by the Jet Propulsion Laboratory/NASA (Watkins et al., 2015; Wiese et al., 2019). Similarly to CSR data, JPL are corrected for GIA effects, C_{20} Stoke coefficients

are replaced by a solution from SLR, and data are reduced to 1º regular grids from 0.5º regular grids. Besides, we have applied the degree-0 Stoke coefficients correction. However, CSR and JPL mascon solutions are not directly comparable. The main reason is that an estimate of degree-1 coefficients has been added to JPL mascon solutions, and the GAD product has not been added back. The corrections applied by JPL are not supplied separately and we cannot do/undo any of the corrections to process JPL data as we did with CSR data. In particular, the GAD product is not available for JPL. In any case, the JPL solution is useful here since it provides an error estimate of the mascon solution that can be propagated to obtain confidence intervals of N, which are independent from those estimated with the bootstrap analysis. Table 2 shows the CI_{95} of the mean values of the N component for different ocean basin estimated from error propagation and bootstrap analysis. It is observed that in all cases the CI_{95} from error propagation are included in those from bootstrap analysis, meaning that the latter are a conservative estimate of the error. JPL propagated error can be expected to be similar to that propagated from CSR error estimates (which are not available), and then we can assume that the reported CI_{95} for N calculated from CSR data are a conservative estimate. Besides, comparing Tables 1 and 2, it is observed that the mean values of N are quite similar and that the CI_{95} largely overlap. Regarding to the time variability, the values of the N component from CSR and JPL mascon solutions show Pearson correlation coefficients greater than 0.85 (p-value < 10^{-3}), except for the Atlantic (0.70). Thus, despite the different processing of CSR and JPL data, the reported analysis for the N component is robust with respect to the choice of GRACE datasets.

Table 2. Mean net WT from JPL mascon for different ocean basins according to Equation 2 . CI_{95} are estimated as propagation of mascon errors provided by JPL, and from bootstrap analysis. Units are Gt/month.

		Mean (CI_{95} from error propagation)	Mean (CI_{95} from bootstrap)
Outflows	Pacific	1182 (1143, 1220)	1182 (1062, 1306)
	Arctic	735 (713, 757)	735 (711,761)
	Pacific + Arctic	1917 (1872, 1961)	1917 (1806, 2036)
Inflows	AIA	1183 (1092, 1274)	1183 (1077, 1282)
	Atlantic	919 (866, 972)	919 (845, 985)
	Indian	999 (980, 1018)	999 (928, 1067)
	Atlantic + Indian	1918 (1862, 1974)	1918 (1838, 2003)

(2) ERA5 P and E data are replaced by several datasets for comparison purposes. The objective is not to be exhaustive in the selection, but rather to show that the reported features of the N component are quite robust with respect to the choice of the P and E datasets. The data sets considered are:

(i) Continental P from GPCC (Schneider et al., 2011), GPCP (Adler et al., 2018), CMAP (Xie and Arkin, 1997), UDel (Willmott and Matsuura, 2001), and GLDAS/Noah (Rodell et al., 2004; Beaudoing and Rodell, 2016).

(ii) Ocean P from GPCP and CMAP.

(iii) Continental E from GLEAM (Miralles et al., 2011; Martens et al., 2017) and GLDAS/Noah.

(iv) Ocean E from OAFlux (Yu et al., 2008) and HOAPS/CM SAF (Schulz et al., 2009).

The Pacific outflow is estimated with the 162 possible combinations of P and E, including ERA5. The time period is 2003-2016, except for HOAPS/CM SAF and GPCP, which span from 2003 to 12/2014 and 10/2015, respectively. The degree-0 corrections in GRACE data is made for each combination. Note that only ERA5 includes P and E for both continents and oceans. All grids have been homogenized to 1 °regular grids. The main concern here is the heterogeneity of the spatial coverage among datasets. To make the results comparable among datasets, the computations are restricted to the common grid points, which do not cover the entire Earth (Figure 8a).

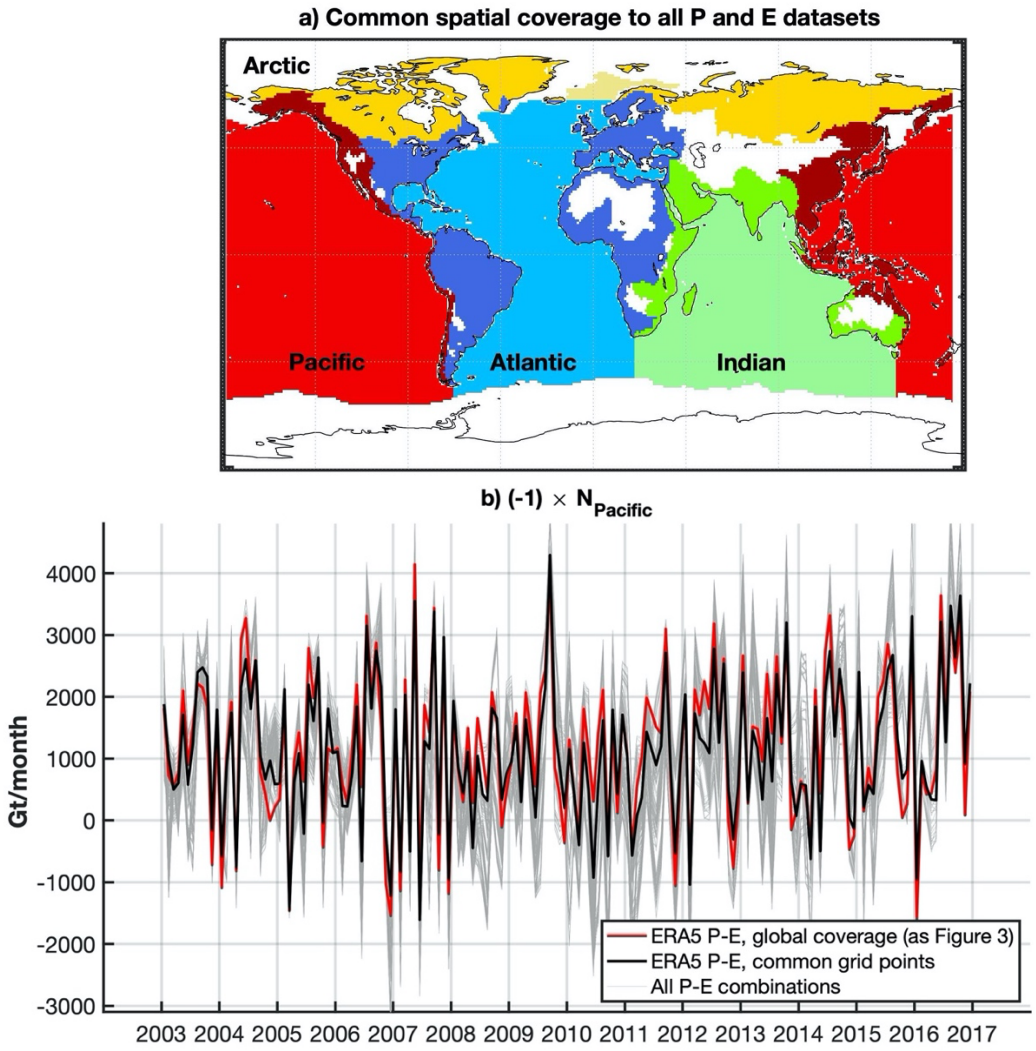

Figure 8. Monthly time series of (the opposite of) the Pacific outflow estimated from 162 combinations of *P* and *E* datasets. a) Spatial coverage common to all datasets. b) Pacific outflows: Gray thin curves are the 162 Pacific outflows estimated in the common grid points to all datasets (no global coverage); black and red curves are based on ERA5 *P* and *E* and are obtained using either only the grid points common to all datasets (black curve) or global coverage (red curve). Note that the red curve is the same as in Figure 3.

However, in spite of the fact that due to the partial coverage the principle of water mass conservation is not accomplished, the Pacific outflow obtained in the common grid points from ERA5 (black line in Figure 8b) is quite in agreement with the same signal obtained with global coverage (red line in Figure 3 which is also reported as red line in Figure 8b). The Pearson correlation coefficient between the two signals is 0.994 (p-values < 10^{-3}) with an average difference around 50 Gt/month. In general, the Pacific outflows estimated from all the P and E dataset combinations show qualitatively the same signal than the one reported in Figure 3. For each of the 162 estimates of the Pacific outflows corresponding to the possible P and E dataset combinations, we evaluated the average outflow (over the period of study), which is 968 Gt/month (STD: 489), and the correlation with the Pacific outflows in Figure 3, which is 0.82 (STD: 0.06; p-values < 10^{-3}).

These experiments show that the reported net WT are physically consistent among datasets, at least qualitatively."

3. **Bootstrap Vs Error propagation:** The confidence intervals estimated from bootstrap have been compared to those estimated from error propagation of the GRACE mascon. As CSR mascon solution does not provide such error estimates, we have used the JPL mascon solution for the comparison. An explanation of why bootstrap confidence intervals contains, as expected, the error propagation confidence interval has been also provided. In the description of the bootstrap method we have included the following text:

As an independent check of the bootstrap, confidence intervals for the mean value of N have been also evaluated by propagating the error estimate in GRACE data (using the JPL GRACE mascon solution for which error estimates are available). The resulting intervals were consistent with those of the bootstrap method. In particular (see Section 4 for details), we show that in all cases the bootstrap intervals contain the intervals obtained from error propagation. In this respect, the CI_{95} from bootstrap analysis can be considered a conservative estimate. This should be expected, since the residual component underlying the bootstrap approach includes measurement errors and other type of errors (related, for example, with the estimate of the trend and seasonal terms). As a result, the uncertainties in the transports estimated by the bootstrap should be larger than the corresponding uncertainties estimated by error propagation.

The details are shown in the new section "Comparison with other datasets", that can be found above.

4. About the lack of correlation. We have included a discussion on the lack of correlation between the inter-annual transports and the indices of ocean-atmosphere interaction. In particular we propose the two following explanations:

"To explore this lack of correlation, we have estimated the correlation coefficient between each climatic index and each WT component (Figure 7b).

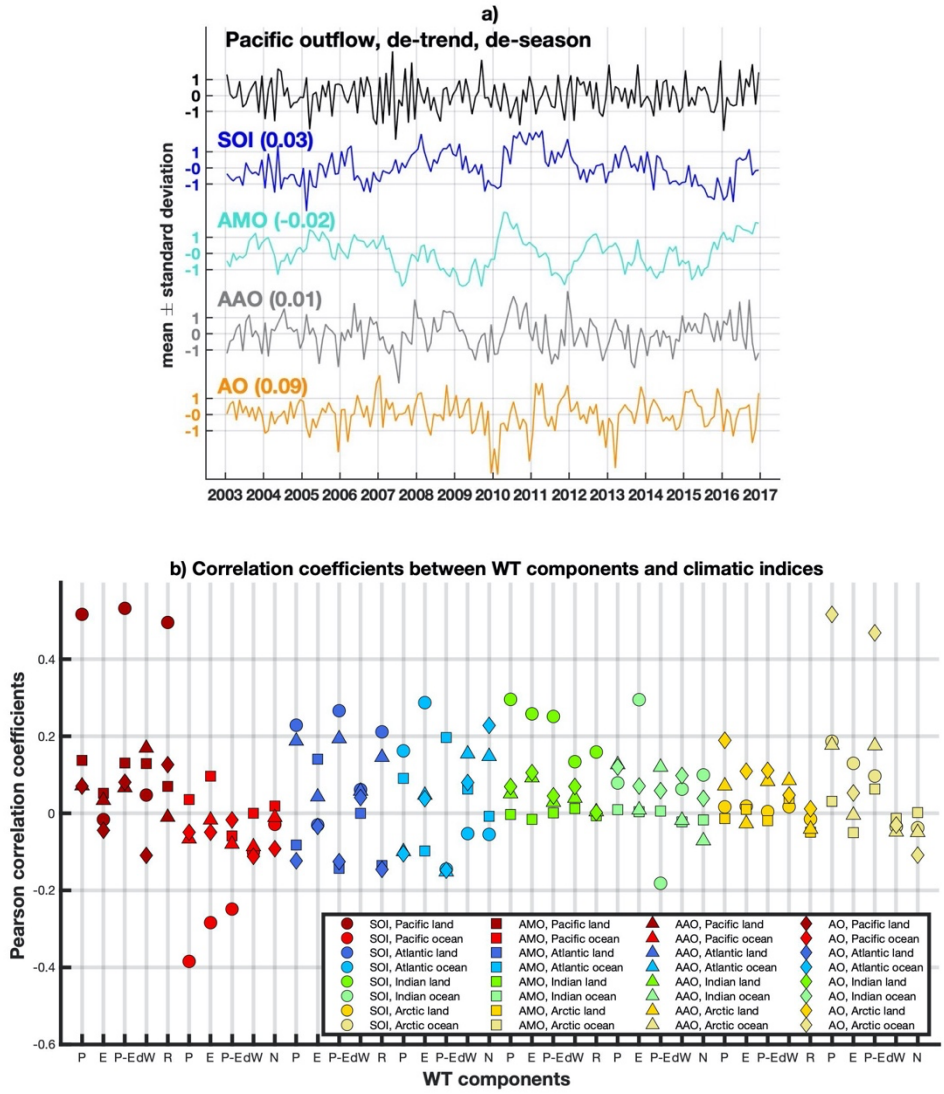

Figure 7. Pacific outflow and climatic indices for ENSO, AMO, AO, and AAO. a) Time series of Pacific outflow is de-trend and de-season. All time series are normalized to have unit variance. Values in the parenthesis are the correlation coefficient between the corresponding climatic index and the Pacific outflow. b) Correlation coefficients between de-trend and de-season WT components of different regions and the climatic indices.

All of them are lower than 0.3 except for 6 cases in 2 regions. In the Arctic, P and P–E in the drainage basins of the Arctic show a correlation of ~0.5 with the AO. This correlation is natural since that is the area of influence of the AO. The other region is the Pacific, where, as expected, the SOI shows a correlation around 0.5 with P, P–E, and R in the drainage basins, and around -0.4 with P in the ocean. However, this individual correlation does not extend to the Pacific outflow. In order to understand why this is the case, it is convenient to express the N component of the water transport as a function of (P-E) and dW. According to Equations 1 and 2 we have:

$$N = -(P-E)_{ocean} - R + dW_{ocean} = \underbrace{-(P-E)_{ocean}}_{X_1} \underbrace{-(P-E)_{land}}_{X_2} + \underbrace{dW_{land}}_{X_3} + \underbrace{dW_{ocean}}_{X_4}. \quad (3)$$

It can be shown that the correlation between N and a given index can be express as follows

$$corr(N, Index) = \sum_{i=1}^{4} corr(X_i, Index) \cdot \frac{std(X_i)}{std(N)}, \quad (4)$$

where corr denotes the correlation coefficient, and std stands for standard deviation. As shown in equation (4), the correlation between N and a given index is a linear combination of the correlation between each component and the index. The coefficients of the linear combination std(X_i)/std(N) are proportional to the standard deviation of each component. The components of equation (4) for the Pacific outflow and the SOI index are shown in Table 3. Despite the fact that some of the individual component exhibits significant correlation with SOI (in particular P-E in land and ocean) when combined with the corresponding coefficients their effects are canceled out yielding to a negligible correlation between water transport and SOI (below 0.03 in magnitude).

Another possible reason for the lack of correlation resides in the definition of the studied regions, for which the presence of subregions with positive and negative influence of an index results in an overall negligible/attenuated influence of the index in the overall region. For example, a positive phase of the AMO is related to an increase of P in western Europe (Sutton and Hodson, 2005), and the Sahel (Folland et al., 1986; Knight et al., 2006; Zhang and Delworth, 2006; Ting et al., 2009), but to a decrease of P in the U.S. (Enfield et al., 2001; Sutton and Hodson, 2005), and northeast Brazil (Knight et al., 2006; Zhang and Delworth, 2006). All these regions are included in the Atlantic drainage basin, and then the influence of a positive phase of the AMO is attenuated."

Table 3. Correlation coefficients between SOI and de-trend and de-season WT components involved to estimate the Pacific outflow according to Equations 3 and 4.

	$std(X_i)$ (Stand. Deviation)	$corr(X_i, SOI)$ (Correlation between X_i with SOI)	$\frac{std(X_i)}{std(N)}$ (Coefficients)	$corr(X_i, SOI) \cdot \frac{std(X_i)}{std(N)}$ (Correlation ·Coefficient)
$X_1 = -(P-E)_{ocean}$	605	0.25	0.57	0.14
$X_2 = -(P-E)_{land}$	212	-0.53	0.20	-0.11
$X_3 = dW_{land}$	96	0.048	0.09	0.004
$X_4 = dW_{ocean}$	711	-0.10	0.67	-0.07
		Corr(*N*,SOI)		-0.03

Note that table 3 provides also some insights about the causes of the interannual variability of Pacific Ocean outflow. The largest standard deviation of *P-E* and *dW* in the ocean suggests that these two components might drive the interannual variability of the Pacific Ocean outflow. This is confirmed by a correlation analysis. The correlation between *N* and the $(P\text{-}E)_{ocean}$ is -0.70. The correlation between *N* and the dW_{ocean} is 0.84. The correlation of *N* with the corresponding land components is below 0.18. In all cases, prior to the evaluation of the correlation the corresponding time series have been de-trend and de-season.

5. **Improvement in the visualization of main results**. We have included a new Figure with a diagram of the mean WT components to ease the reading:

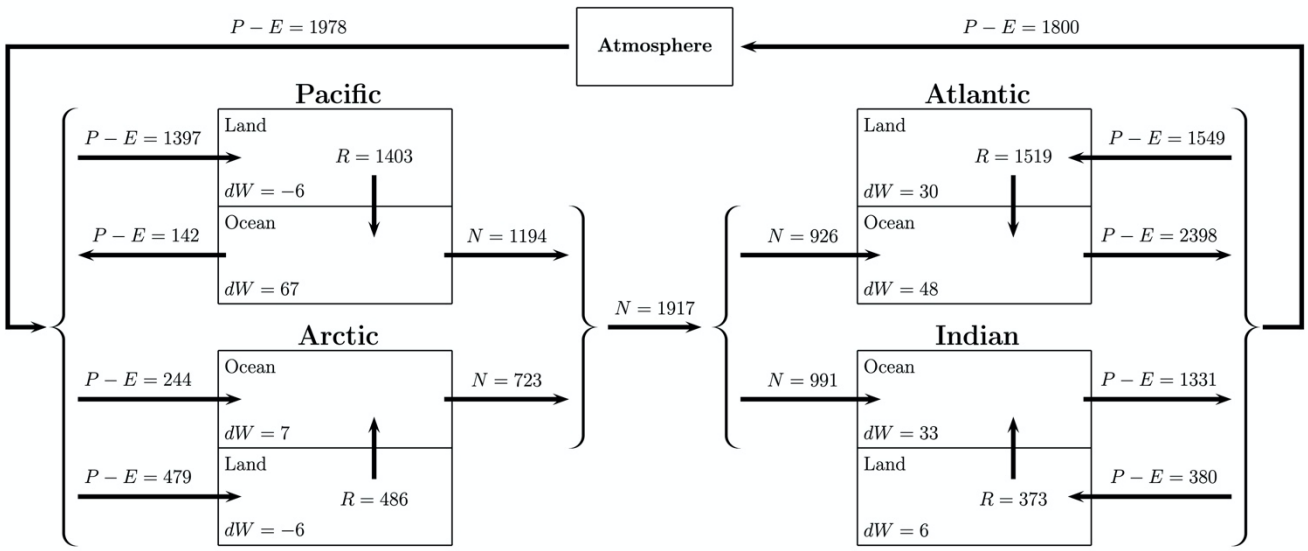

Figure 5. Diagram of the mean values of the WT of the studied regions. Units are Gt/month.

Other comments:
L44, change 'de' to 'the'.

Typo has been corrected.

In section 2: Methodology and Data: please use subtitles to re-organize the section, and improve the readability.

Section 2 is now divided in 4 subsections:
2.1 Methodology
2.2 Precipitation and Evaporation data
2.3 Time-variable GRACE data
2.4 Confidence intervals

Fig 1: suggest to add legends to indicate the locations of different ocean basins.

Figure 1 has been modified, now it includes the names of the basins in the figure itself.

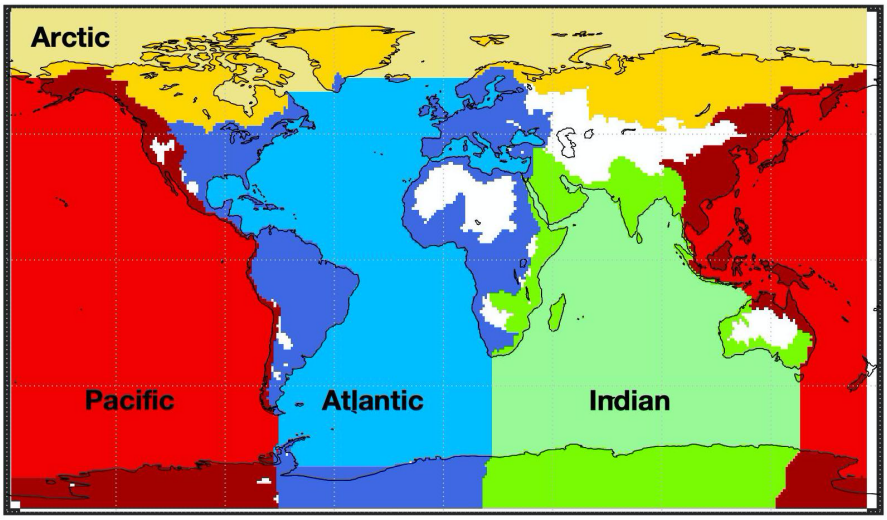

Figure 1. Pacific, Atlantic, Indian, and Arctic Ocean basins and their associated continental drainage basins according to the global continental runoff pathways scheme of Oki and Sud (1998). Within each basin, darker colour represents the continental basin, lighter colour the ocean basin. White regions represent endorheic basins.

Fig 2: are these values obtained from GRACE? or a combination of ERA5 and GRACE? please briefly clarify this in the figure caption.

The next sentence has been added to the caption:
"P, E, and P–E are from ERA5 dataset; dW is estimated from GRACE; R and N are estimated as a residual in equations 1 and 2, respectively."

Fig 3: black curve is the AIA inflow: ’if’ should be ’is’.

Typo has been corrected.

Fig 4: N component should be briefly explained in the caption.

The last sentence of the caption is now:
“Black lines are the WT N component, which are estimated as residuals in Equation 2.”

Fig 6: please indicate the correlation coefficient R between each climate index and the Pacific outflow, at each time series.

It has been included. See the new Figure 7b in the point 4 (About the lack of correlation) of this response.