RC1

The manuscript describes assessing the water mass balance (WMB) method for estimating river discharge into the Atlantic basin, in which change terrestrial water storage calculated from satellite. River discharge from variations on this method are compared with in situ observations. The impact of the different estimates of river discharge on ocean model simulations of salinity and currents are further investigated. The manuscript is well-written, and both the potential and the limitations of the method are reasonably communicated. The manuscript should be a useful contribution to the literature.

We thank the reviewer for their comments which contributed to improving the manuscript.

There are a couple of areas where the content could be expanded. Firstly, the authors remark how the method is sensitive to the reanalyses used - it would good to get an idea of the spread possible with a range of reanalyses. I don't suggest repeating the analysis multiple times, but it might be possible to get a picture of the spread by calculating one diagnostic (say net evaporation over the amazon) for different reanalyses.

The estimation of river discharge was performed using two different sets of atmospheric forcing: WMB1, based on ERA5 net precipitation, and WMB2, based on the difference between GPCC precipitation and ISBA-CTRIP evapotranspiration. The difference between the two approaches provides an indication of the sensitivity of the river discharge estimates to the atmospheric forcing. For the Amazon River basin, the difference in evapotranspiration between ERA5 and ISBA-CTRIP is consistent with the difference between ERA5 and GLEAM reported by Chen et al. (2020). Similarly, Lehmann et al. (2022) evaluated a wider range of precipitation, evapotranspiration, storage, and discharge products for closing the water budget and found that ERA5-Land generally performs well, except over regions such as the Amazon and eastern China. Although only three atmospheric flux sources are considered here (ERA5, GPCC, and ISBA-CTRIP), the fact that precipitation and evapotranspiration are estimated using substantially different methods—including atmospheric reanalyses, in situ observations, and hydrological models—likely provides a realistic representation of the dispersion among atmospheric flux estimates.

An explanation has been added at lines 275-281.

Secondly, it would be good for the authors to unpack the confusing finding that WMB1 is better than WMB2 at improving the representation of river discharge but worse at improving the representation of ocean salinity.

The performance of the WMB approach differs between the river-basin and ocean-basin scales. At the river-basin scale, the approach is sensitive to local and regional biases, which may compensate for each other when integrated over an ocean basin. Consequently, a reanalysis may provide a more accurate representation of net precipitation at the regional scale, while performing differently when its freshwater flux is evaluated in an ocean model against the Atlantic. In other words, the freshwater flux can be locally accurate while the integrated freshwater input to the Atlantic Ocean remains biased. Note, however, that differences in the impact on SSS between WMB1 and WMB2 are very small and not statistically significant.

An explanation has been added at lines 557-564.

Further specific and technical comments.

- Title: Suggest a title more like 'Quantifying freshwater fluxes from Major rivers and assessing their impact on the Atlantic Ocean'

The title has been modified

- Abstract

- Line 10. Suggest ‘riverine freshwater fluxes’.

Indeed, riverine fluxes are specifically studied, this should be stated in the first sentence of the abstract. The sentence has been modified.

- Methods

- Line 84. A little justification/ quantification required here. Which rivers were excluded? What was the discharge of the largest of these? What was the total annual discharge excluded (say as a percentage of the Amazon)?
  Seven rivers without in-situ observations. What uncertainties does this bring to the result. Is there enough data to estimate variability of in those rivers?

Additional explanations concerning the quantification of the river discharge excluded from this study have been added lines 86-87 and lines 93-95. It should be noted that in the ocean simulation, the discharge for the rivers which are not studied is the monthly climatology of Dai and Trenberth (2002), as compiled by Bourdalle-Badie and Tréguier (2006).
For the 7 rivers without in situ observation, no discharge data was available during the GRACE era, or there were too many gaps in the measurements, so that the annual and inter-annual variability could not be assessed. These rivers account for 8% of the mean discharge of all the studied rivers according to the Hydro Rivers dataset.

- Line 167-170. Consistency of fluxes. Are the ERA5 fluxes self-consistent? Specifically, globally does precipitation=evaporation on a reasonable time-scale? (I enquire because we know that most reanalyses are not close to producing a globally balanced net ocean heat flux).

ERA5 is indeed known not to be globally balanced. Its global moisture budget does not generally close (see e.g. Mayer et al., 2021), with spurious variability likely linked to the data assimilation system; however, at the basin scale relevant to the WMB, ERA5 residuals are smaller than in earlier reanalyses; any residual regional bias in P−ET propagates directly into our discharge estimates, which is likely the limitation we see for some basins with the WMB approach.

Mayer, J., M. Mayer, and L. Haimberger, 2021: Consistency and Homogeneity of Atmospheric Energy, Moisture, and Mass Budgets in ERA5. J. Climate, 34, 3955–3974, https://doi.org/10.1175/JCLI-D-20-0676.1.

We added a sentence on that at lines 168-172

- Tsujino et al (2018) use a catchment basin river routing model driven by JRA reanalysis. Shouldn’t the fact that this is driven by reanalysis mean the approach is not too dissimilar to the authors and that there might not be much quantitative difference?

The two datasets share an atmospheric reanalysis as input dataset but are constructed in fundamentally different ways. JRA55-do runoff is the output of the CaMa-Flood river routing model driven by JRA-55 land-surface runoff, whose input is rescaled basin by basin to fit Dai et al. (2009) (namely constrained towards the same climatology used in REF and contains no observational information on water storage). The WMB instead derives discharge from observed GRACE/GRACE-FO terrestrial water storage changes, with no land-surface or routing model and no posterior calibration.

We added a sentence on that after introducing the J5 experiment.

- Line 188. ‘Several experiments’. My understanding is the authors conducted 6 experiments (4 experiments and 2 ‘control’ simulations). Could the authors list these in a table, giving the control simulations an abbreviation?

We have added a new Table (Table 2) summarising the main characteristics of the experiments.

- Lines 200-202. I find this unclear. Firstly NEMO is an ocean model, so it does not have river basins as such, merely river outlets. It makes sense to regrid or reassign the outlets on to the NEMO grid, but I don’t understand regridding the basin. Secondly, I cannot understand the phrase ‘the local spread of runoff with the coastlines of pertinence of each river drainage basin, proportional in area to the monthly climatology’. Could the authors, break this down into smaller more easily comprehensible chunks?

We agree that "remapping the basins" is somehow misleading, as nothing is done with the drainage basins on the ocean grid. The WMB provides one discharge value per river mouth, and we simply reuse the spatial distribution of coastal ocean points over which the NEMO climatology already spreads each river, rescaling it in time so that the integrated flux matches the WMB value. We rewrote the paragraph to make it clearer.

- Line 213. The correlation is high because of the strong seasonal cycle. The authors should correlate the time series with the seasonal cycle removed.

Indeed, the strong seasonal cycle impacts the correlation scores. The correlation computed from the WMB1 and WMB2 anomalies have been added, still exhibiting good agreement with the reference. We have chosen to keep the annual cycle when computing the correlation coefficient in Table A1, as this definition is the one being used for the KGE’ score.

- Lines 217 onwards. Do you the authors understand the reason for the weaker agreement in the annual cycle they describe in WB2?

The differences between WMB1 and WMB2 arise from differences in the estimated net precipitation, primarily due to lower mean, higher noise and greater variability in ISBA evapotranspiration, including at annual timescales, when compared with ERA5.

An explanation has been added at lines 253-255 and a figure in Appendix A.

- Lines 230- 234. Suggest this paragraph is replaced by something like. ‘Linear trends of the lines in Fig 3a were calculated and found to be statistically insignificant.’

This paragraph has been modified.

- Line 254. ‘and highlight’ suggest ‘but highlight’.

The text has been modified.

- Line 296. ‘overestimated’. Just say it is larger in WMB2 by this amount. It cannot be described as an overestimate as WMB1 is not ground truth.

The text has been modified.

- Line 330. ‘overestimated’ suggest ‘overestimated relative to the in-situ observations.’

The text has been modified.

- Table 2/ 3 could the authors clarify the difference between Table 2 and Table 3.

Table 2 shows metrics corresponding to the two WMB approaches at the river outlets, these discharge estimates are being used for the A1, A2, S1 and S2 simulations. As there are no reference measurements available at the river outlets, the same metrics are computed for the WMB approaches at the gauging stations closest to the outlet, and compared to the in situ discharge in Table 3.
The corresponding table captions have been modified to make this point clearer. An additional comment has been added lines 348-350.

- Line 345. 'Fig. 6' should be 'Fig. 5', I think.

The text has been corrected.

- Line 368. 'Figure 7' should be 'Fig. 6'

The text has been corrected.

- If I understand correctly Fig. 6a shows the MAE of A1 relative to EN4 minus the MAE of ref relative to EN4. I think this could be stated with greater clarity in the caption

We rewrote Figure 6's caption to make it clearer.

- Line 373. I suggest the authors add 'not shown'. i.e. 'The REF experiment…West African coast (no shown). Alternatively, the authors could show The MAE of REF relative to EN4 in another panel.

We added a panel with the MAE of the REF experiment

- Line 376. 'Key Regions'. Because the MAE of Ref relative to EN4 is not shown we don't know what the key regions are.

We rewrote the sentence, because "key regions" seems quite arbitrary here; however, the MAE of REF is now shown.

- Line 384. Suggest 'into the ocean model'

Modified according to the Reviewer's suggestion

- Please comment on the larger impact of the WMB method around the Gulf of St. Lawrence and Congo.

Both are semi-enclosed regions where a given runoff change may produce a very large salinity response, and both are cases where our discharge differs most from the climatology (St. Lawrence: WMB1 mean is ~70% higher than the Dai and Trenberth value; Congo: the WMB1/WMB2 spread is the largest of all basins). We have added a sentence on this.

- Also on Table 3 it would be of great interest to add the Dai and Trenberth Values as 2 columns to Table 3 if possible and comment on them. This is highly relevant since the authors compare WB forced ocean simulations with those forced by Dai and Trenberth.

We agree that the comparison with the forcing used in the REF experiment is highly relevant. However, the reference climatology at the river outlets is obtained from the Dai and Trenberth discharge at the in-situ station through several processing steps and adjustments. Because of this, directly comparing the discharge datasets at the station in Table 3 would not provide an accurate reflection of the reference fluxes actually used in the simulation.

Instead, we have chosen to directly compare the forcing used in the REF experiment (located at the outlet), with the WMB estimates at the outlet.
Overall, the mean WMB fluxes are lower than the reference climatology, but no general pattern emerges regarding the comparison of seasonal discharge amplitudes.
Table 2 has been updated, and an additional comment has been added lines 327-330.

- Figure 7 and elsewhere. The authors do not include any caveats about the EN4 dataset. The data set is far from perfect and the key input from Argo floats are absent in the shelf regions around the coasts.

We added a sentence on the limitations of EN4, stressing that the main purpose is to assess the impact on large-scale patterns rather than small scale variations. See also reply to RC2.

- Line 422. Fig. 8 ERA5 WMB1 is better than WMB2 at improving river discharge but worse (or no differerent) at simulating ocean salinity. How can this be explained?

Please note that the difference is very small (−0.08 vs −0.07 psu in terms of RMSE improvement), and the ocean response to runoff is not linear, so better discharge skill at the gauges does not translate one-to-one into better SSS. Also, the discharge metrics are computed not for all 18 rivers, while the ocean sees the mouth discharge of all 18 rivers. We have added a sentence to justify the apparent inconsistency.