**Response to Anonymous Referee #1**

Thank you very much for your positive comments and constructive feedback, you addressed some important points. Your clarifications helped to make the manuscript clearer for the reader. Our responses are provided in blue together with your original comments in black.
We really appreciate your time and insight in reviewing our manuscript!

Kind regards,
Susanna (on behalf of all co-authors)

General/overarching comments:

*To answer your general comments, we revised the whole conclusion section (from line 555 onward) and marked adapted passages* ***bold****:*

**Summarizing we refined past Arctic water budget estimates (Serreze et al., 2006; Dickson et al., 2007, e.g.,) and their uncertainties by combining some of the most recent reanalyses datasets and observations, and by applying a variational optimization scheme. The variational adjustment worked very well on an annual scale and brought reliable estimates of the volume budget terms, requiring only moderate adjustments of less than 3% for each individual term. Adjustments are considered reliable if budget closure was achievable within the respective terms error bounds and if the terms were comparable to estimates from past studies.** With an annual value of 4379 ± 25 km3 (calculated over 1993-2018), our adjusted runoff estimate is slightly higher than estimates made by Shiklomanov and Lammers (2013) and Shiklomanov et al. (2021a) for the period 1936-2006. However, considering the different calculation periods and assuming a decadal rise of roughly 2%, the estimates come quite close. On a seasonal scale however, the adjustment process was not a full success, as the budget residuals of some months are too large to be eliminated within the a priori spreads of the individual terms. This is very likely caused by systematic errors being present in the datasets, or at least in their seasonal cycles, that are not considered in our a priori uncertainty estimates.
**Especially when calculating Pan-Arctic runoff, caution is needed. Our results show that seasonal peaks of river discharge are underestimated in almost all of the assessed reanalyses (ERA5, ERA5-Land, GloFASE5, GloFASE5L). The biggest anomalies are caused due to inhomogeneities in the data assimilation system (ERA5 and GloFASE5) which led to a great underestimation of runoff, especially at the latter half of the time series. However also reanalyses without data assimilation (ERA5-Land**

**andGloFASE5L) were not able to reproduce the seasonal cycle of river discharge accurately.** On the other hand, we find distinct improvements in the new GloFASERA5newproduct, especially when investigating seasonal cycles and long term means it features vast enhancements compared to its precursors. However, for interannual variability and trend analysis we would recommend the use of the VIWVD estimate, as it reproduces trends from gauge observations quite accurately. **When extrapolating observed river discharge to the whole Pan-Arctic area we found that the common method of hydrological analogy tends to underestimate the discharge peaks. We therefore advise to use river discharge observations where available and reliable runoff/discharge estimates from reanalyses (e.g., GloFASERA5new or ERA5-Land) to extrapolate discharge to the ungauged areas.**

A further possible reason for inconsistencies between runoff and ocean reanalyses is the usage of climatological river discharge data to specify land freshwater input in ocean reanalyses. Our analyses show that the seasonal cycle of the ORAS5 runoff climatology Bt06 fits quite well to our observation-based estimates (see Fig. 5), however the lack of inter-annual variability in the freshwater input alters i.a. oceanic volume transports. In their BRONCO (Benefits of dynamically modelled river discharge input for ocean and coupled systems) project Zuo et al. (in preparation) work on implementing a time-varying land freshwater input, derived from discharge data from GloFAS version 2.1, into ORAS5. This should further reduce the inconsistencies between runoff and oceanic fluxes from ocean reanalyses.

**To further refine the budget estimates, longer time scales of all budget terms would be needed. For example, one could repeat the analysis using the back extension of ERA5 which goes back to 1950. There's also a new bias corrected ERA5 data set (Cucchi et al., 2020, WFDE5), that could be examined in relation to the Arctic water budget. Further it would help to include a precipitation observation data set, preferably one that combines available satellite-based and gauge-based data sets. Concerning the biases in ocean reanalyses, one could refer to oceanographic data for comparison. Generally, comparison to oceanographic data is difficult, as observations are limited concerning their temporal and spatial coverage. Nevertheless, the unique form of the Arctic Ocean (as water leaves and enters only through a handful of gateways) allows relatively easy measurements of the in- and outgoing fluxes. This was done in the ARCGATE project, where data from arrays of moored instruments (like e.g., Acoustic Doppler Current Profilers, MicroCAT – CTD Sensors and Seagliders) were taken to estimate transports through the Arctic gateways. Our results showed that it is possible to estimate annual fluxes into and out of the Arctic Ocean quite accurately, however the moored instruments did not measure the**

**velocity field accurately enough to resolve the barotropic wave signal arising from temporally varying runoff (T. Tsubouchi, personal communication 2021) leading to differences in the seasonality of the fluxes. A longer measuring period with an even denser monitoring network could help with this aspect.**

1. Although river discharge data and land/ocean water storage data are used carefully, the paper doesn't use oceanographic data to estimate the marine water fluxes (it only uses reanalysis data). This isn't a major problem because the paper focuses on runoff and river discharge, but it should be mentioned and discussed somewhere (Conclusions?).

   We discussed it at the end of the revised conclusion above.

2. Related to point 1: What's the scope/opportunity for future improvements on the Arctic water budget analysis? What model and data assimilation improvements would help? What data are needed to refine the budget estimates? Again, this isn't a major omission, but it will help set the context for future work if this point is discussed somewhere (Conclusions?).

   What's the scope/opportunity for future improvements on the Arctic water budget analysis?
   See revised conclusion above.

   What model and data assimilation improvements would help?
   *We added the following paragraph in section 4.2.3:*
   "This discontinuity issue is not only limited to ERA5, but rather a general issue in reanalyses, as observation platforms are changing through time, making it practically very hard to make these products 100% homogeneous in time. Especially satellite data were not available in the early days and were introduced with the development and introduction of new instruments. If the redundancy is large enough, then any discontinuity impact should be less pronounced. However, specifically for snow there's only the IMS product that was introduced in 2004, and hence any inhomogeneity generates a larger impact. Thus, this impact could be reduced when using other data sets on top of the IMS product or instead of it, which ideally go further back in time."

   What data are needed to refine the budget estimates?
   See revised conclusion above. The part about longer time scales, the bias corrected ERA5 and precipitation observations.

Specific Comments:

1. Line 50: Cite where it says the ERA5 runoff features spurious trends.

We added a citation of Zsoter et al. (2020)

2. Figures 1 and 2: What is the source of catchment data in Figures 1 and 2?

We added the following catchment sources (shapefiles):

**Individual river catchment outlines** were taken from the CEO Water Mandate Interactive Database of the World's River Basins (http://riverbasins.wateractionhub.org/)
**Regional outlines** (CAA etc.): GRDC (2020): WMO Basins and Sub-Basins / Global Runoff Data Centre, GRDC. 3rd, rev. ext. ed. Koblenz, Germany: Federal Institute of Hydrology (BfG).

3. Section 2: Add a table containing information on the runoff and discharge sources (ERA5, ERA5-Land, GloFAS…, GREP etc.)

We added a table containing all runoff and river discharge sources and key information of the individual products.

| Product | Description | Variable | Period | |
|---|---|---|---|---|
| ERA5 | Fifth generation ECMWF reanalysis using IFS (+ HTESSEL) | Runoff [m/s] | 1979-2019 (back extension to 1950 available) | Hersbach et al., 2020 |
| ERA5-Land | Offline simulation of ERA5 without DA using HTESSEL | Runoff [m/s] | 1981-2019 (back extension to 1950 expected in 2021) | Muñoz-Sabater et al., 2021 |
| $GloFAS_{E5}$ | ERA5 runoff + simplified LISFLOOD | River discharge [$m^3/s$] | 1979-2019 | Harrigan et al., 2019 |
| $GloFAS_{E5L}$ | ERA5-Land runoff + simplified LISFLOOD | River discharge [$m^3/s$] | 1999-2018 | - |
| $GloFAS_{E5new}$ | Full configuration of LISFLOOD | River discharge [$m^3/s$] | 1979-2019 | - |
| Bt06 | Runoff climatology used in ORCA025 | River discharge [$m^3/s$] | Climatology | Bourdalle-Badie and Treguier, 2006 |

| | | | | |
|---|---|---|---|---|
| Observations | Measurements at gauging stations | River discharge [$m^3/s$] | - | - |

New citation:
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

4. Line 70: "river discharge" includes both liquid water and ice (presumably)?

   Yes, it does, we clarified it.

5. Line 73: For clarity, say that "associated domain" means the catchment area.

   We clarified that.

6. Line 104: It talks about "different bulk formulas and differences in the data assimilation..."  Different to what?  Be specific.

   *We specified some of the differences and referred to the individual documentations for further details. The following paragraph was added after line 103:*

   "While the GREP ensemble members use the same ocean model (NEMO, nucleus for European Modelling of the Ocean) and atmospheric forcing, there are differences in observational data and data assimilation techniques, as well as in the reanalysis initial states, NEMO versions, the sea-ice models, and physical parametrizations.
   The data assimilation methods differ in many points, including the deployed assimilation schemes which range from 3DVAR (three-dimensional variational data assimilation) to SEEK (Singular Evolutive Extended Kalman Filter). Furthermore, there are differences in the input observational dataset, in surface nudging, in the time-windows for assimilation and analysis as well as in the applied bias correction schemes. All those differences lead to an important dispersion between the reanalysis implementations and add up to the ensemble spread. (Storto et al., 2019)
   For further details we refer to Storto et al. (2019) as well as the individual data documentations, which are listed in the manuscript."

7. Line 105: The sentence starting "We also look into....ORCA025" appears out of place. Move up to line 72?

We moved it up.

8. Line 130: “additional area” needs to be clarified. Is this a catchment area?

   Yes, it’s a catchment area, we added the word “catchment”.

9. Section 3: Many math terms aren’t defined clearly. E.g., S_A, S_L, S_O, F. \sigma^2_k. Make sure all terms are carefully defined.

   We looked through them all and defined the terms that were missing.

10. Line 142: Justify the neglect of atmospheric liquid water and ice.

    *We added the following justification:*

    “Atmospheric liquid water and ice are neglected, as they represent only a very small fraction of water in comparison to atmospheric water vapor and lateral moisture fluxes – the liquid water portion stored in the atmosphere as well as liquid transports represent only about 1% compared to water vapor transports globally and even less in the northern latitudes (Hantel and Haimberger, 2016). Generally atmospheric water in liquid and solid phase are only significant in regions with high tropical cumulus clouds and over warm ocean currents (Serreze and Barry, 2014).”

    Hantel, Michael, and Haimberger, Leopold (2016). Grundkurs Klima. Berlin, Heidelberg: Springer Berlin Heidelberg.

    Serreze, M., and Barry, R. (2014). The Arctic Climate System (2nd ed., Cambridge Atmospheric and Space Science Series). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139583817

11. Line 151: What does A_total represent?  The Arctic Ocean?  The Arctic Ocean plus terrestrial catchments?

    A_total is the sum of the Arctic Ocean and terrestrial areas, hence the total Arctic area considered in this study. We specified it in the paper.

12. Line 158: What about groundwater contributions to the land water budget?  (And their changes in time).

    That’s a good point. As the ERA5 and ERA5-Land reanalyses do not contain groundwater storage and soil moisture is only given to the depth of 289cm. However, in the end we use GRACE satellite data to estimate land water storage changes, where groundwater changes are indeed included.
    **Hence, we added the change of groundwater in equation 2 and also**

**mentioned the lack of groundwater in the reanalysis products in the data section:**
"The representation of frozen land components is not ideal and also groundwater storage is not represented in the examined reanalyses" (see also general comment 2 of Anonymous Referee #2)

As a side note, ECMWF currently works on increasing the number of soil layers and introducing a groundwater storage using a flexible, modular system called ECLand (Boussetta et al., 2021, Muñoz-Sabater et al., 2021).

Also, LISFLOOD, the hydrological model used in the GloFAS river discharge reanalysis, includes a groundwater module. This module consists of two reservoirs that store and subsequently output the water into the river channel after a certain time delay (Harrigan et al. 2020). We added this aspect in the data section as well.

New citations:
Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Boussetta, S., Balsamo, G., Arduini, G., Dutra, E., McNorton, J., Choulga, M., Agustí-Panareda, A., Beljaars, A., Wedi, N., Muñoz Sabater, J., de Rosnay, P., Sandu, I., Hadade, I., Carver, G., Mazzetti, C., Prudhomme, C., Yamazaki, D., and Zsoter, E.: ECLand: The ECMWF Land Surface Modelling System, Atmosphere, 12, 723, https://doi.org/10.3390/atmos12060723, 2021.

13. Line 178: It says "we assume sea-ice to be transported by the ocean currents..." but sea ice moves (somewhat) independently from the surface ocean current. More explanation/justification is needed.

*That's also a very good point, that we investigated more thoroughly. In conclusion we changed our method of analysis as follows:*
*Instead of omitting sea ice transports we add them to our volume transports but at the same time remove the liquid water volume that is actually replaced by sea ice, which we call the equivalent liquid water flux.*

*Line 177 onwards now reads:*
"The liquid portion of F is calculated by integrating the cross-sectional velocity component along the side areas of the Arctic boundary. Additionally, we add ice transports, which are calculated analogical by integrating the cross-sectional ice velocity over the grid-point-average ice depth, multiplying it by the grid cells sea ice fraction and integrating it over the Arctic boundary. As

volume exchange between liquid ocean and sea-ice is conserved in the NEMO model, we additionally remove the liquid water volume that is actually replaced by sea ice, which we call the equivalent liquid water flux. The equivalent liquid water flux at a given grid point is calculated by integrating the liquid volume flux over the grid-point-average ice depth, multiplying it by the grid cells sea ice fraction and taking 90% of the result (as only ~90% of the icebergs are underwater). As ice velocities from the public CMEMS data portal are only available from two of the ocean reanalyses (ORAS5 and GLORYS2V4) we calculate the ice flux "correction" term for the GREP ensemble by taking the mean of only those two products. However, as the impact of the correction is quite similar for ORAS5 and GLORYS2V4 we believe that the correction is accurate enough for our purpose."

*This of course led to slight changes in figures 9-13 and tables 6 and 8, which have been updated in the revised manuscript. However, the results do not change substantially, and the main conclusions remain valid.*

14. Line 219: Equation (10) and the text about it are unclear.

    *We changed the equation and the text to the following:*

    "We calculate relative, decadal trends following Zsótér et al. (2020) and Stahl et al. (2012) by applying a linear regression to the annual mean time series:

    trend=10*slope/mean

    The *slope* of the time series is the annual trend obtained through the linear regression and the *mean* is the long-term annual mean of the timeseries. The ratio *slope/mean* is multiplied by the factor 10 to provide decadal trends. Hence a relative decadal trend of e.g. 0.1 is equal to an increase of 10% over a decade. All trends are calculated over the common period of the discharge datasets 1981-2019, except for GloFASE5L which is calculated over 1999-2018. We do not consider auto-correlation and determine significance using the Wald Test with a t-distribution, where p-values smaller than 0.05 are considered as significant."

15. Line 257: The sentence starting "Cuchi et al. (2020) run the hydrological model...." is out of place. Cut?

    Thanks for spotting this. We now cite the work of Cucchi et al. (2020) in the conclusion section. (see general comments)

16. Line 259: It says "Model runs with ERA5 forcing show similar river discharge seasonalities at the Lena catchment as GloFASERA5new." This isn't what I see in Figure 2 for the Lena. Check and cut or clarify.

This was referring to the work of Cucchi et al. (2020) in the sentence above. Nevertheless, we cut it.

17. Line 278: "Again this could be caused by delayed river ice breakup and backwater that is considered in GRACE, but not in ERA5" is a bit misleading. GRACE observes the natural system, which includes delayed river ice breakup. The ERA5 model excludes does not represent these processes. Instead, maybe end the sentence with "that is observed by GRACE, but not considered in ERA5".

    Thanks for this comment. We clarified this accordingly.

18. Line 309: Say that the "hydrological analogy" means extrapolation to the un-gauged rivers and streams.

    We added this as well.

19. Figure 3: The legend identifies "GloFAS_{ERA5}" and similar, but the legends in Figure 2 call it "Glo_{E5}" (also Table 3). Use consistent notation throughout.

    We checked all notations and adapted them to be consistent.

20. Line 341: The sentence "Additionally Greenland features a storage decline of -134 km3 per year, accounting for roughly 50% of the total storage change" is unclear. Clarify and cite.

    *Clarification: The two sentences above address land water storage change in GRACE over the Pan-Arctic area excluding Greenland. In addition to this, Greenland features a storage decline of -134 km3 per year, also taken from GRACE.*

    *Correction of the passage:*
    "And just as for the four major basins, also for the sum of all Eurasian and North American catchments (excluding Greenland) GRACE data show a major decline of land water storage over the past decades, reaching -132 km3 per year for our area of interest, while land storage from ERA5 shows considerably smaller declines of -34 km3 per year. The largest changes for GRACE water storage occur over the Canadian Arctic Archipelago and the mountainous areas of Mackenzie and Yukon basin, suggesting a tight linkage to glacial melting. Additionally, GRACE water storage shows a strong decline of -134 km3 per year over the Greenlandic ice cap (north of Fram and Davis Strait), raising the total Pan-Arctic storage change to 266 km3 per year."

21. Figure 5: Remind the reader that the dashed lines sum to the brown line.

We added it to the figure caption.

22. Table 4: Add a column with units (applies to other tables too, trends in particular have an unclear unit). The "m3 s−1 ∗ 10−3" unit contradicts Table 3.

    We corrected the units and added a column in the tables.

23. Section 4.2.3: The discussion on the ERA5 runoff glitches is useful. Speculate on how they could be fixed?

    *We added the following paragraph in section 4.2.3 (see also general questions):* "This discontinuity issue is not only limited to ERA5, but rather a general issue in reanalyses, as observation platforms are changing through time, making it practically very hard to make these products 100% homogeneous in time. Especially satellite data were not available in the early days and were introduced with the development and introduction of new instruments. If the redundancy is large enough, then any discontinuity impact should be less pronounced. However, specifically for snow there's only the IMS product that was introduced in 2004, and hence any inhomogeneity generates a larger impact. Thus, this impact could be reduced when using other data sets on top of the IMS product or instead of it, which ideally go further back in time."

24. Figure 9: Explain what the different lines in the right panel mean.

    Those are the different realizations of FW-input minus storage change using the various datasets described in the text. We clarified/repeated it in the figure caption.

25. Line 512: It says "both F and runoff R feature adjustments beyond their a priori spreads, demonstrating that the a priori uncertainties are larger than indicated as systematic biases are not incorporated". What are the likely systematic biases?

    *We added the following explanation after line 513:*

    "The state-of-the art reanalyses exhibit systematic errors concerning the runoff seasonality, where the seasonal runoff peaks in summer are too low in comparison to observations, while winter and spring values are too high. Due to the lack of reliable seasonal observations of the oceanic volume fluxes, it's hard to define systematic biases in the ocean reanalyses. However, all four ocean reanalyses feature quite low September volume fluxes, which are not found in their forcing components (see figure 9). Uotila et al. (2019) assess ten ocean reanalyses, including CGLORS, FOAM, GLORYS and ORAS5, specifically in the polar regions and find multiple systematic errors concerning sea ice thickness and extent, temperature profiles, mixed layers as well as ocean transports. Unfortunately, seasonal cycles of volume transports were not assessed either, however seasonal cycles of sea ice components and heat transports did exhibit systematical biases. Further analyses would be necessary to find certain, systematical errors in the seasonal volume transports."

New reference:
Uotila, P., Goosse, H., Haines, K. et al. An assessment of ten ocean reanalyses in the polar regions. Clim Dyn 52, 1613–1650 (2019). https://doi.org/10.1007/s00382-018-4242-z

26. Line 535: State briefly the origin of the ERA5-Land runoff declines of 5-6% and comment on their realism.

*We added the following explanation:*

"Those declines in ERA5-L runoff are caused by similar declines in P-E from ERA5 short-term forecasts, as P-E is used as a forcing in ERA5-L (see figure 6/table 4). As observations agree on an increase of river discharge, those declines are deemed unrealistic. An improvement may possibly be achieved when taking the divergence of moisture flux (VIWVD) as forcing, as VIWVD, which is computed from analysed fields rather than short-term forecasts, features similar trends as discharge observations. See J. Mayer et al. (2021) for an assessment of the ERA5 water budget and the different behavior of budget quantities taken from analyses or short-term model forecasts."

New citation:
Mayer, J., Mayer, M., & Haimberger, L. (2021). Consistency and Homogeneity of Atmospheric Energy, Moisture, and Mass Budgets in ERA5, Journal of Climate, 34(10), 3955-3974, https://doi.org/10.1175/JCLI-D-20-0676.1.

27. Line 552: It says "With oceanic and land storage declining…", yet many papers exist on the accumulation of freshwater in the western Arctic Ocean (e.g., see Proshutinsky et al., 2019, 10.1029/2019JC015281). Mention and comment on this issue.

*This is not necessarily contradictory to our studies as GRACE measures changes in **mass** not freshwater. Furthermore, considered timespan and areal extend have great effects, as storage is tightly linked with winds and circulation patterns. Concerning the spatial distribution of the trends, we examined trend maps and also found slight increases of mass in the Chukchi and Beaufort Sea (= western Arctic), however stronger declines in the eastern part of the Arctic, with the strongest changes in Baffin Bay.*
*We added the following paragraph after line 507:*

“Other papers mention the accumulation of freshwater in the western Arctic Ocean (e.g, Proshutinsky et al., 2019). This is not necessarily contradictory to our findings of a slight decline in ocean storage, as GRACE does not measure the accumulation/decline of freshwater but rather the change in mass. Furthermore, areal differences (we consider the whole Arctic Ocean) and differences in the considered timespan have great effects, as storage is tightly linked with winds and circulation patterns and exhibits strong nonseasonal fluctuations. Longer time series would be needed to determine whether the trends that we found are caused primarily by such fluctuations or indicate a true loss of mass.”

Technical Corrections:

1. Many places: Apostrophes are not used correctly and there are spelling and grammar errors.

   We revised the whole paper for correct grammar and to eliminate spelling errors.

2. Abstract, line 16: Reword “look into Greenlandic discharge”

   Reworded to:
   In addition we examine Greenlandic discharge,...

3. Abstract, line 22: Which “data-sets”? Be specific for clarity.

   As the systematic biases come from the reanalysis and ocean-reanalysis datasets we added those.

4. Line 169: “reference salinity” should read “reference density” I think.

   That’s correct, we changed it.

5. Figure 2: What are R and \mu in Figure 2 legends?

   R is runoff and \mu is the long-term mean. We added them to the figure caption.