Response to Anonymous Referee #2

Thank you very much for your constructive feedback and your positive comments, you addressed some important points. Your clarifications helped to make the manuscript clearer for the reader. Our responses are provided in green 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 comments:

The manuscript is well organized, but there are some spelling and grammatical errors that needs to be considered, including the use of commas and apostrophes. I also suggest to avoid using words such as "spurious", "huge", "clearly" etc., especially for the results and conclusions sections (see also specific comments).

We revised the whole paper and correct spelling and grammatical errors and we made sure to avoid words as "spurious", "huge", "clearly",...

Considering that previous studies focusing on the Arctic drainage basin have used different approaches and motivations for its geographical domain, I am missing a motivation for the chosen boundary of the Arctic Ocean drainage basin in this study, and why e.g., Hudson Bay, and James Bay was not included? (e.g., L302-305).

We added a clarification at line 121:
"Figure 1 presents the study domain. As there's no strict boundary to the south, the definition of the Arctics geographic extent varies between past studies and there's no general rule whether to include Greenland and the Hudson Bay or not. We chose our study domain to be consistent with Tsubouchi et al. (2012) as we wanted to compare the oceanic fluxes from ocean reanalysis with the observation-based estimates from the ARCGATE project. The Arctic Ocean is bounded by the position of hydro-graphic moorings in the main gateways. ..."

Clarify also in L313 that total drainage area refers to the area for this study.

We clarified this.

How do these reanalysis products take frozen components of the freshwater system into consideration, e.g., glaciers and permafrost, considering that many of the river basins in the study are underlain by permafrost? For example, lines

497-498 includes an interesting aspect that I would like to see more elaboration on.

We added an explanation on the representation of frozen land components in reanalyses in the data section (section 2):

"The representation of frozen land components is not ideal and also groundwater storage is not represented in the examined reanalyses. In HTESSEL, the land-surface model used in ERA5 and ERA5-Land, glaciers are represented as large amounts of snow which are kept fixed to 10 m of snow water equivalent. When melting conditions are reached, the snow produces a water influx to the soil and consequently contributes to the total runoff. However, the mass balance is not accounted for over glaciers as the snow is restocked to constantly stay at the fixed 10m level and hence changes in the glacial storage component can't be assessed properly. The soil water content includes liquid as well as frozen components and thus also includes permafrost. When the soil temperature reaches melting conditions, the soil water contributes to sub-surface runoff and the soil water storage declines. However, a recent study by Cao et al. (2020) concluded that ERA5L soil data are not optimal for permafrost research, due to a warm bias in soil temperature that leads to an overestimation of the active-layer thickness and an underestimation of the near-surface permafrost area. Therefore, we additionally use GRACE (Gravity Recovery & Climate Experiment) satellite data for our final analyses to accurately estimate the land water storage, as GRACE includes liquid water (including groundwater), glaciers and permafrost."

Additionally, we justified the use of ERA5 land water storage for certain purposes at lines 497-498:

"Oceanic transports out of the Arctic domain exceed the atmospheric moisture entering the Arctic (6295±121km3) by nearly 5%, indicating an annual loss of water volume of roughly 300km3. The bulk part of this loss is generated through terrestrial water mass losses. Even though the representation of frozen land components is not ideal in the considered reanalyses, the comparison of GRACE mass changes to the sum of ERA5 storage changes (snow and soil water) and glacial changes taken from literature agreed well. Therefore, we use land storage changes in ERA5 (excluding glaciers) to estimate which terrestrial sources contributed to what extend to the general storage decline in the Arctic. We found that approximately 50% of the 266 km3/year decline are generated through liquid and solid discharge from Greenland, while about 40% come from Arctic glaciers (excluding Greenland) and the remaining 10% are the result of a decline in land water storage due to permafrost and snow cover degradation."

New references:

Cao, B., Gruber, S., Zheng, D., and Li, X.: The ERA5-Land soil temperature bias in

permafrost regions, The Cryosphere, 14, 2581–2595, https://doi.org/10.5194/tc-14-2581-2020, 2020.

In the conclusions, I am missing a general discussion on implications for future studies and assessments of freshwater budgets of the Arctic Ocean.

This question is similar to the 2. general comment by Anonymous Referee #1. We revised the whole conclusion section. Some of the key points that were added are the following:

"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 data-sets 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."

"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."

"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."

"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."

Specific comments:

L12: I suggest to avoid the use of "spurious" and instead explain or reference to what you are referring to.

Rephrased to:
Seasonal river discharge peaks are underestimated in ERA5 and GloFAS v2.1 by up to 50%, due to pronounced declining runoff trends which are caused by two temporal inhomogeneities in ERA5.

L37: consider removing "remarkably"

We removed it.

L41-43: Consider rephrasing for clarity and also specify the part on climatological conditions.

Rephrased to:
In addition, significant portions of the rivers discharge may bypass the gauging stations through braided channels or as submarine groundwater. Furthermore, also climatological conditions pose a hindrance to gauge measurements, as the low temperatures in the northern latitudes often lead to river freeze up in late autumn and flooding in spring due to river-ice break up (Syed et al., 2007).

L45: avoid using "huge"

We changed it to “great”.

L47-48: Consider rephrasing for clarity. and
L48-49: This is not very clear, please explain what you mean by “spurious” (see also previous comment related to this).

We rephrased it to:
However data assimilation systems can introduce biases, as temporal changes in the observing system are inevitable and may lead to inhomogeneities in the time series. One known change is the introduction of the IMS (Interactive Multisensor Snow and Ice Mapping System) snow product in ERA5, which led to a negative shift in ERA5’s snowmelt and consequently also runoff. (Hersbach et al., 2020, Zsoter et al., 2020)

L92: Which 16 rivers were included in the study, and how was the shorter observational records treated for the analysis in comparison to the longer observational records?

We added the names of the 16 rivers:
Pur, Taz, Khatanga, Anabar, Olenek, Yana, Indigirka, Alazeya, Abadyr, Kobuk, Hayes, Tana, Tuloma, Ponoy, Onega, Mezen

And clarified the consideration of shorter vs longer records:
“We calculated an observation-based Pan-Arctic river discharge for the period of 1979 to 2019. Therefore, we calculated discharge for every time step (= every month) separately and used all river discharge measurements that were available at this certain timestamp. Analogous river discharge for the ungauged area was estimated for each individual timestamp over the area that was not gauged at this time, using two calculation methods (see section 4.2).”

L119-120: What about frozen storage components, such as glaciers?

We added permafrost and glaciers as those were considered too through GRACE. (See answer of general comments)

L134: I suggest to add references to earlier studies, and revise “popular” to “common” – if this is what you are referring to?

Yes, we changed the wording and added a couple references:
A common way to calculate the oceanic freshwater budget is through the assumption of a reference salinity (e.g., Serreze et al., 2006; Dickson et al., 2007; Curry et al., 2011; Haine et al., 2015).

I suggest to remove a, b, c in subheadings (e.g., L196, 205).

We removed those.

L244: consider removing “clearly”

We removed it.

Fig 4: Is this figure only considering the shorter time series of the 16 catchments, or for the full time period (1981-2019)? Same question for figure 6 and the observed Pan-Arctic river discharge data.

The black lines (observations) of Figure 4 consider the shorter time series of all catchments, while the reanalyses are taken over the full time period. We considered the difference as insignificant, as the purpose of the figure was just to show the differences between the different extrapolation methods.
We added all time periods into the figure captions.

L497-498: This is an interesting aspect that I would like to see more elaboration on.

See answer to general comments

L527: I suggest to include references here, and do you mean “common” rather than “popular”?

Yes, we again changed popular to common and added references (e.g., Shiklomanov and Shiklomanov, 2003).

L528: How does this result compare to other studies?

We a comparison to other studies:
“We estimate Pan-Arctic river discharge from gauge observations using monthly correction factors from GloFASERA5new, as the popular method of hydrological analogy tends to underestimate the high flow summer peaks (see Fig. 4) and obtain a long-term annual flux of 4031km3 ± 203 (excluding Greenland). To compare our results to past studies we adapted the time periods and areal extents accordingly and found reasonable accordance with Haine et al. (2015), who combined runoff from ERA-Interim with river discharge observations and obtained a total discharge about 5% higher than our estimate. An even better agreement was found with the estimates made by Shiklomanov et al. (2021a), as the total Pan-Arctic discharges (including Greenland) agree within 2%.”

L529: Runoff from ERA5 is substantially “too low” – do you mean “underestimated compared to observed discharge” or similar?

Yes, that’s what we meant. We rephrased it accordingly.

L531: Please consider rephrasing and describe the "unrealistic" aspects.

We rephrased it:
"Those strong declines are caused by two inhomogeneities (1992 and 2004) in ERA5's snow melt time-series and contradict the discharge increases found in gauge observations. Those inhomogeneities are caused by a loss of snow through changes in the data assimilation system."

L548: What is considered "trustworthy" here – please explain.

By trustworthy we mean the data products we have most confidence in after comparing them amongst each other and to observational datasets. We changed the sentence to the following:
"Comparing the estimates of freshwater input into the Arctic Ocean that we have most confidence in after the preceding analysis (listed in table 7), to volume transports from ocean reanalysis yields…"

L555: What is considered "reliable" – please explain

We added an explanation:
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.

L559: What would be a full success here, please elaborate.

We changed the sentence to the following:
On a seasonal scale however, the adjustment process was not a full success, as some of the adapted fluxes fell out of their a priori uncertainty range. A full success would include elimination of the budget residuals for every single month, while at the same time staying inside the respective 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 taken into account in our a priori uncertainty estimates.

L560: revise month to months

We changed it.

L571: Please specify what you refer to with "in most reanalyses".
L572: Please specify what you refer to with "spurious signals".

We changed it to:
"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 and GloFASE5L) were not able to reproduce the seasonal cycle of river discharge accurately.”