Point-by-point changes made in response to reviews

Referee 1 comments:

1. **Please label the river basins in the Figure 1.**

   Basins are now labelled in Figure 1.

2. **The boundary of the Amur River is not correct, which would make following results not right. Please check different maps to use the correct boundary.**

   A new mask has been made for the Amur based on Hydrobasin (Lehner, 2020). Figures 1 and 2 containing maps have been updated to include new mask. Figures 4 and 8 and appendix figures have been updated to include results obtained using the updated mask. Information in Table 2 and 3 have also been updated according to new mask.

3. **Please use appropriate font size and keep consistent in each figure. The font is too small to be readable.**

   All figures have been updated to use larger font size. Size 13 is used for axis labels and legends, and font size 12 for axis values. Previous font size used was python default size of 10.

4. **Lines 374-381. It seems these are methodology, and should not be placed in the results.**

   We have altered this text slightly but have mostly left these lines here as they describe the other products shown in Figure 6 and we believed they were relevant to explain this figure and did not include our own methods.

5. **Lines 419-427. This part is not well written. Each paragraph has only two or three lines. Please rearrange the text.**

   These paragraphs have been rewritten (now lines 455-465)

6. **Line 471. References are needed to support your statement.**

   Here we have referenced Abolafia-Rosenzweig et al., 2021 and Hobeichi et al., 2020 which are examples of studies where no long-term constraints have been used. Now line 509.

7. **When talking about optimization, we always cannot forget some popular optimization algorithms, such as SCE, DDS, GA, etc. What are the differences between your method and these popular ones?**

   We have added some text regarding choice of optimisation algorithm (lines 83-85)

8. **When reading paper, we always want to see the differences between your study and previous ones. Lines 457-470 stressed the similarities but not the differences. Please dig a little bit more to show the differences.**

   We have added text explaining what we are trying to emphasise through discussing similarities (lines 506-508).

9. **From the conclusion, I can see the main contribution from your study is that you introduced a sequential optimisation approach. Other than this, is there any new findings that different from other studies? There are a lot of optimization method can do the similar job. I want to see new findings that can advance our understanding of the hydrological processes.**

   We have added text in conclusion to highlight the methodological advancement we have made in fulling utilising the GRACE storage data on multiple timescales also re-emphasised the advancement of our coupled energy budget (lines 563- 570).

10. **A better judgment of the selections of the river basins should be given. Is it because the observation data in these rivers are better than others? Or other reasons. Some important river basins, such as the Mekong River, are not selected. No rivers in Western Europe are selected. I don't mean you have to select all the rivers, but an appropriate reason should be given.**

    Text has been added to discuss the selection of basins in section 2.7 (lines 229-232).

Referee 2 Comments:

1. **Then, I failed to understand what the usefulness of the "Our Optimised Storage" and associated figures 4 and 5 are. It seems quite circular to me that if you force it to match GRACE it matches GRACE better than other products. From a hydrologist perspective I would rather see the optimized P, Q and E compared to other products. Perhaps even an independent better regional precipitation or river discharge dataset for specific basins, to see whether the optimized fluxes match that better and which would then clearly demonstrate the strength of their method. I am not sure if this is going to require major changes to the paper, or just a clearer explanation of the objectives and results.**

   Text has been added in sections 4.2 and 4.3 explaining why we have chosen to show "our optimised storage" and what the information that this quantity reveals.

2. **A second major, but not difficult to solve issue, is that I find the authors somewhat sloppy regarding equations and symbology. Unfortunately, this set of guidelines has disappeared recently from the HESS manuscript preparations guidelines online: https://iahs.info/Publications - News/Otherpublications/Guidelines-for-the-use-of-units-symbols-and-equations-in-hydrology.do, but I personally still appreciate it if we all try to follow this as much as possible. Of serious concern are Eq. (1) and Eq. (5), which should obviously read dS/dt instead of dS as the fluxes are per unit of time. It would also be helpful if the equations would contain the dimensions, thus, e.g. length^3 per time [L3 T -1] for Eq. (1) so this becomes obvious. Even expressed per unit area [L T-1] would also be fine of course as long it clearly remains a flux and not a stock. Moreover, please use single italicized symbols, so something like Sfi instead of FIS, which makes it directly clear we are talking about a storage. I know many other papers invent funny acronyms as well, maybe it is even the rule rather than the exception, but in my opinion, it is simply not pleasant for any reader.**

   We have redefined 'FIS' to S_fi throughout text, and wrote dS/dt in places where dS was previously used. As well as some other alterations to the symbology.

3. **A third major question is why the authors chose the data they chose and whether it matters for the main point they are trying to make. For precipitation and evaporation, many more observation based products exist, so did they select the 'best' according to some previous studies or did they just select 'good' data and does it not matter a lot whether it is really the 'best'. I hope the authors can explain. Moreover, the runoff data is even dependent on precipitation and evaporation from GSWP3, which is a bit of a vague product in terms of how it was constructed and I think it may even rely partly on GPCP and FLUXCOM, making the estimates of P, E and Q not completely independent. Moreover, I fail to see why spatially varying runoff is necessary at all, as on the basin scale, the actual river discharge measurements at the river mouth would suffice for which, for example, the GSIM archive (Do et al., 2018) could have also been used**.

   We have added text in Section 2 discussing the choice of dataset (lines 121-126).

4. **L1-2: "improving climate and earth system models" I would say 'validating' or 'assessing the capability of' which is to be done first before anything can be improved.**

   This line has been changed to "assessing the capability of.." (lines 1-2).

5. **L6: "the corresponding turbulent heat fluxes ranges between ± 10 W m−2 " I suppose something should range between value x and value y, thus this sentence misses something.**

   This has been replaced to say the imbalances ranges between 1 and 12 Wm-2 (lines 6-7).

6. **L8: "This exposes mismatches in seasonal water storage" Mismatches between what and what exactly?**

   The new manuscript explicitly states that we mean the difference between GRACE and the storage suggested by the other flux observations (line 9).

7. **L14: "The FIS metrics" What are 'the FIS metrics'?**

   This has phrase has been removed.

8. **L23: "Water is a conservative quantity" Technically speaking this statement is incorrect. Water is used by plants for photosynthesis and released by decomposition or fire. Probably it is an order of magnitude lower than the errors made in the products of P, E and Q, but not entirely negligible.**

   This phrase has been replaced to say that the mass of water will remain constant (line 24).

9. **Table 1 "present" and general period statements It is rather irrelevant whether e.g. GRUN is available until 'present' or that it starts in 1902, what matters is which years you used for the analysis.**
   The time resolution column has been removed from Table 1 and text has been added to state the years we downloaded the data for the study.

10. **L91 “evaporation” I strongly support the use of evaporation over the ambiguous term evapotranspiration, see Miralles et al. (2020) for the arguments why that is, so perhaps you could simply use evaporation also elsewhere in the manuscript.**

    The word ‘evaporation’ has been used throughout text now instead of ‘evapotranspiration’.

11. **Equation 5 The integral is between what and what? What does the to the power 0 between brackets mean? Is this equation supposed to present a time series? Then it would be clearer if Sfi,w(t) was explicitly written.**

    Equation 5 has been rewritten. The power 0 is no longer used and limits have been added to the integrals.

12. **Technical corrections L93: “Earths” Earth’s**
    This correction has been made (line 97).

13. **L101: “Land” land**
    This correction has been made (line 104).

Referee 3 comments:

1. **The authors appear to claim that their optimization method works well by evaluating the results with GRACE - a product that was used in the optimization process. Please consider validation/evaluation with an independent product and/or different time periods.**

   In section 4.2 we now explain the purpose of comparing results to GRACE (e.g. line 393-394).

2. **The authors aim to present better water and energy data and methods. For the effort to be impactful and meaningful, please share the data and the scripts (the scripts were shared, but I could not find any content in the readme file).**

   The data has been shared under new doi referenced in the revised manuscript.

3. **Since the paper argues that the produced method constitutes an improvement upon current optimisation methods, it would be useful if the evaluation/comparison figures and results section could show a clearer distinction between comparisons with products that are “optimized” datasets and those that are not.**

   We have now stated in text in section 4.3 that the comparison made in this section is with optimised products. It now reads “Each of these three products contains optimised estimates..”

4. **Since the paper explicitly aims to improve optimization at all time scales (monthly, interannual, trend), it would be useful if the figures and results section could clearly and explicitly show the improvements at each of those time scales.**

   Text has been added to section 4.2 and 4.3 to highlight how Figures 4 and 5 show improvements over multiple timescales (e.g. lines 395-401).

5. **L53: “is these” should be “in these”.**

   This correction has been made (line 54).

6. **L106: Instead of “short and long time scales”, please consider being more precise (e.g., monthly, interannual, long-term trend).**

   This line has been updated to include specific timescales (lines 112-113).

7. **Other parts of the paper suggest that the aim is to both produce optimized estimates and an optimisation method/methodology. Please include all study aims in this “aim” paragraph.**

   This aims paragraph has been updated to include all aims (line 109).

8. **Table 1: “present” is ambiguous, it would be clearer if you simply state the years that were downloaded for use in this study. Also make sure that the capitalisation of the headings are consistent. “Parameter” should be “Variable”, I think. In addition, please consider adding a column describing the dataset type (e.g., satellite, in-situ measurements etc). For GRACE, should the variable be “water storage anomaly”?**

   The time resolution column has been removed from table 1 and the years of downloaded have been stated instead. A new column has been added to table 1 with heading ‘Dataset type’, ‘Parameter’ has been changed to ‘Variable’ and heading have been capitalised.

9. **Methods section: Please consider adding an overview figure of the methodological steps. For variable symbols, please consider using single-letter symbols rather than multi-letter symbols.**

   An overview figure has been added and the flux-inferred storage has been renamed to S_fi throughout manuscript.

10. **Figure 4 (and elsewhere), please check - “total water storage” or “total water storage anomaly”?**

    ‘Total water storage’ has been replaced with ‘Total water storage anomaly’ through out text and in all figures necessary.

11. **L350 First use of ITCZ, write out.**

    Intertropical convergence zone has been written out before first use (line 370).

12. **L461 Please consider providing the relative error in the unit of % for Amazon as well.**

    This value has been added as a percentage of precipitation (line 500).

13. **L468 Since the imbalances of the Amazon and Amur were explained by the lack of measurements, it seems odd that Congo is presented in this context as the basin with lowest imbalance without further explanation. Between the lines, the text seems to imply that the lack of measurements is not as much an issue in the Congo, which is not true. If**

any, the lack of measurements is even a bigger issue in this region. Please consider a revision of the paragraph.

Text has been added to explicitly state that low imbalance is not necessarily due to good measurement coverage to avoid this implication (lines 503-505). This section also no longer discusses Amur imbalance.

14. **Sect 5.1. Consider moving relevant parts to the Methods.**

    A new section named 'Goodness of fit' has been added in methods section 3.2.3. This contains relevant parts of section 5.1 (lines 335-339).

15. **L551. Could the authors also share the optimized results?**

    Optimised results have now been shared and are available under new doi referenced in the revised manuscript. The text now also states that data on additional basins can be made available upon request.

16. **I could not find any content in the readme.md file beside a single row stating "Water-and-energy- budgets". I have attempted to view it both by downloading it and opening it using a text editor, and by previewing it on GitHub. Please check.**

    A new doi has been included where the script and results should be accessible.