Dear Susanna,

Thank you very much for your valuable comments. We thoroughly revised the manuscript based on your suggestions.

Please see our answers below in red.

With kind regards,

Eva Boergens (on behalf of the authors)

General Comments

The study presents an analysis of long-term variations in GRACE total water storage variations (TWS) over the last 22 years for Africa and compares the data to surface water storage (SWS) variations in major lakes derived from satellite altimetry in central Africa. The authors compare the datasets also with meteorological/drought data via time series analysis and statistical methods. They discuss the influence of human and climate on the variability in TWS and surface SWS in Central Africa. They also provide some novel insight into the TWS dataset through a cluster analysis for the continent Africa. The authors have conducted a good work. It provides detailed information on data and methods used and provides very interesting insight into the water storage variations in the study area around Lake Victoria in Africa. I do think, however, that a more structured organization of the manuscript; a quantification of uncertainty of the surface water storage estimates; and, following from that, a more comprehensive discussion and concise conclusion of the results would make the work clearer and more significant.

On organization of the work: In Section 3-7, the authors cover certain topics. For Section 3-5 they combine respective methods, results and discussions into one section. For Section 6, some of the relevant methods are explained in Section 2, then some more method description is added after results and discussion in this section. The authors often jump between results presentations along various figures and corresponding piece-wise discussion and conclusions. It makes it harder for the reader to discriminate objective facts from opinion or suggestions by the authors. Some of this becomes especially a problem in Section 5 and even more so Section 7, which are presented the least clear. A clearer structure should be introduced to the entire manuscript, for example, for example, either the methods, or the discussions should be split off in some way. Also, some of the figure organization need some improvements, for example some legends are incomplete. Introducing panel letters might help to address results in figures with more than two panels more clearly. Some figures might be better suited for a supplement. Further suggestions are given below.

Following your advice, we restructured the manuscript as follows:

1. Introduction

2. Study Region
3. Data
    1. Terrestrial Water Storage Data
    2. Precipitation Data and Precipitation Indices
    3. Surface Water Storage Data
    4. Soil Moisture and Groundwater Storage Data
4. Methods:
    1. Time series analysis
    2. Clustering Algorithm
    3. Validation and Assessment Metrics
5. Results and Discussion
    1. Clustering of Interannual TWS Variations
    2. Comparison between TWS Signals and Precipitation
    3. Comparison between TWS Signals and Water Storage Compartments
    4. Dynamics of Lake Victoria and of Downstream Water Bodies in the Nile River Basin
6. Conclusion
7. Appendix A and B

The subsections of section 5 are each be subdivided into results and discussion (in the ordering of text, not with subsubsections). Following your suggestions, we revised the figures and decided which could be moved to the supplement or even omitted altogether. Section 5.3 has been renamed as further storage data sets have been incorporated (see comments below).

On uncertainty of the results: estimates of TWS from GRACE as well as SWS from altimetry and subsequent modeling includes several sources of uncertainty, e.g. measurement errors, parameter uncertainty. These uncertainties should be discussed. But since the authors are quantifying percentage of explained signal variance, a quantification is also suggested, especially for SWS. The conclusions need to be put into perspective of the uncertainties (see also next section).

Thank you very much for this remark, which was raised similarly by the other reviewers. We added uncertainty estimates to all data sets involved as far as possible and included these uncertainties in discussing the results.

Boergens et al. (2020, 2022) developed a covariance model for TWS data to assess the uncertainties of this study's used TWS data set. From these, the uncertainties of the STL-derived time series components are derived with the help of a Monte-Carlo simulation.

Although the altimetric water level time series come with an error, these are only formal errors from the Kalman filter estimation. They can only be used for an internal comparison between different time series but not as a measure of uncertainty of the water level observations. Based on literature data, we assume for all altimetric water level time series an uncertainty of 5cm, which is in line with validations against in-situ gauge data. We assume an uncertainty of 5% for the water area extend data. Based on these uncertainties, the uncertainties of the volume time series and SWS are gained through variance propagation.

The newly introduced data set of root zone soil moisture comes with an uncertainty layer, which is used as it is.

For the newly introduced data set of groundwater storage, we variance propagate the uncertainties from TWS, SWS, and RZSM.

The precipitation data set is provided without uncertainty assessments.

Boergens, Eva, et al. "Modelling spatial covariances for terrestrial water storage variations verified with synthetic GRACE-FO data." GEM-International Journal on Geomathematics 11.1 (2020): 24.

Boergens, Eva, et al. "Uncertainties of GRACE-Based Terrestrial Water Storage Anomalies for Arbitrary Averaging Regions." Journal of Geophysical Research: Solid Earth 127.2 (2022): e2021JB022081.

On the conclusions: The study's conclusion on the nature of the driver of TWS variations, i.e. whether it is either human or climate during certain temporal periods, is not fully supported by the results and analysis provided. First, this statement is mainly directly addressed in Section 7, where water levels of various lakes and river level are compared, and the impact of dam management is highlighted. There is no direct comparison with SWS and TWS variations provided. Second, a correlation of TWS to drought indicators is not an explanation or proof of climate dominance, as stated in the conclusion (L357-358), because human water use (e.g. of surface or groundwater) itself is typically also heavily influenced by drought conditions and might therefore similarly impact TWS. In addition, in the rest of the manuscript, the authors only analyze the SWS portion of TWS variation but no soil moisture or groundwater, hence, a large portion of TWS variation remains unexplained, and therefore a conclusion on human or climate dominance in TWS remains very speculative. I am also wondering if such a conclusion is even relevant to emphasize on the importance of the work, but rather may take away from the actual interesting quantitative and qualitative findings of the work on the importance of SWS in the region. This could be more highlighted by slightly altering the discussion of the findings.

Following your comment, as well as similar comments by the other reviewers, we decided to shift the focus of the manuscript away from the anthropogenic influence and towards the (geodetic) observations of the variability of the hydrological storage compartments and of TWS. We carefully revised the conclusion accordingly.

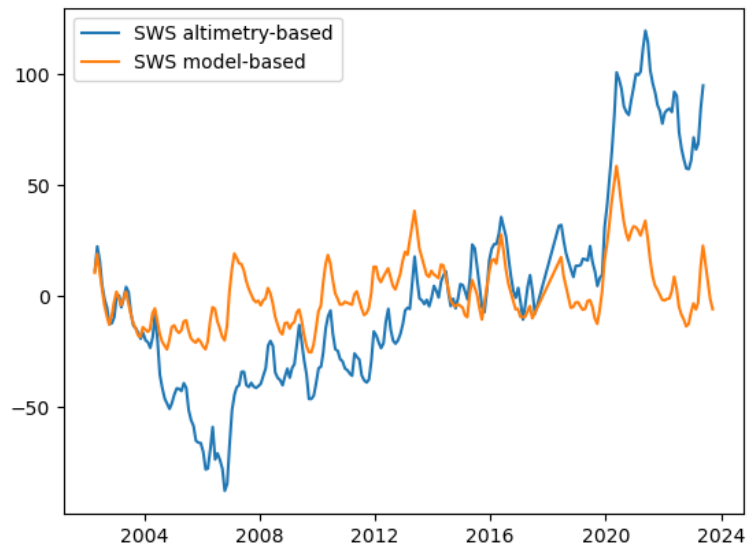
In addition, the authors do not comprehensively quantify and discuss why TWS may be rising overall in the Central Africa/Lake Victoria region over the last two decades (they did so only for specific sections of the TWS time series or in relation to P and SWS). It was shown that precipitation plays an important role. However, the P increase (or change in ET) does not indicate if and where the water is stored. (Here, the authors could also make the role of the hydrological processes - flux versus storage - more clear in the work.)

Also, based on a comment by Reviewer #3, we included an explanation about flux vs. storage in Section 5.2 of the manuscript.

Then, is the overall TWS increase mostly due to the accumulation of water in the lakes/reservoirs, or may other storages also play a role? The results the authors show, do suggest that quite some of the increase sources from the lakes. However, since up to 50% of annual variations occur only during very specific times, e.g., dry years (further comments below), and the size of the linear trend is quite different (Figure 10, additional numeric quantification of this overall increase might be helpful to compare SWS and TWS) a large part of the interannual increase is still unexplained by SWS. However, the correlation between the time series (TWS and SWS in Figure 10, bottom) is striking and the overall rise over the last decades very congruent, just the amplitudes are not matching. So, the question is, does the uncertainty of the SWS amplitudes (from sensors and model parameter) (or from TWS) play a role here? Or are maybe other storage components besides SWS equally important for explaining TWS rise in the region? Just as an example (no need to cite), Werth et al. (2017) have suggested groundwater storage increase may play a role for the storage increase in the Niger basin, and the argument was supported by reports of increasing groundwater levels in the region. Since the cluster for Niger and Lake Victoria have some similarity, maybe groundwater might be relevant in your study area as well. Such or similar thoughts could be included in the discussion and conclusions of the work.

Thank you for this very valid and valuable suggestion, which aligns with the other reviewers' comments. We decided to introduce further analyses of soil moisture and groundwater storage change in the manuscript (now Section 5.3 Comparison between TWS Signals and Water Storage Compartments).

The authors of this manuscript have been part of the international consortium of the "Global Gravity-based Groundwater Product" (G3P) project funded by the EU as a Horizon 2020 project (https://www.g3p.eu/), joining several leading experts in Europe for satellite-based remote sensing of soil moisture (W. Dorigo, TU Wien), glaciers (M. Zemp, Uni Zurich), snow (K. Luojus, FMI) and mass changes with GRACE (A. Güntner, F. Flechtner, GFZ, T. Mayer-Gürr, TU Graz, A. Jäggi, Uni Bern). G3P provides groundwater storage changes as the difference between TWS and surface water storage (SWS), root zone soil moisture (RZSM), snow, and ice. The latest data set version, including all individual storage compartments, is available until 09/2023 (Güntner et al., 2024). While RZSM is satellite-based, the SWS variations are based on simulation results of the hydrological model LISFLOOD (Van der Knijff et al., 2008). However, LISFLOOD simulations of surface water storage changes are considered unreliable in the study region (cf. Prudhomme et al., 2024). In particular, despite similar dynamics and shorter time scales, the modelled SWS does not show the distinct and strong interannual variability we see in the altimetry-derived SWS of the study (see the following figure, not included in the manuscript).



Thus, we computed groundwater storage variations for the present study based on our altimetry-based SWS results as GWS = TWS - RZSM - SWS(altimetry). Snow and ice can be neglected in the study region.

Side note: Our altimetry-based SWS does not include river storage variations. Based on the model's different SWS components of rivers, lakes and reservoirs, we estimate that river SWS explains roughly 10% of the seasonal SWS variations in the study area and does not show large interannual trends.

We now assess the year-wise storage change in TWS, SWS, and GWS. The TWS changes before 2006 are largely explainable with SWS while the changes since 2018 are equally explained by SWS and GWS changes. Most of the annual variation originates from RZSM. We also include uncertainties from all water storage data sets in these investigations. See details of the results in the manuscript.

In addition to the gravity-based groundwater storage estimation, we examined in-situ groundwater observations provided by the Global Groundwater Monitoring Network (GGNM, https://ggis.un-igrac.org/view/ggmn/) to assess how they could be employed in the study. This network has some groundwater time series available in the study region. However, their time frame is very limited. Thus, we decided against introducing them to the study.

Güntner, Andreas; Sharifi, Ehsan; Haas, Julian; Boergens, Eva; Dahle, Christoph; Dobslaw, Henryk; Dorigo, Wouter; Dussailant, Inés; Flechtner, Frank; Jäggi, Adrian; Kosmale, Miriam; Luojus, Kari; Mayer-Gürr, Torsten; Meyer, Ulrich; Preimesberger, Wolfgang; Ruz Vargas, Claudia; Zemp, Michael (2024): Global Gravity-based Groundwater Product (G3P). V. 1.12. GFZ Data Services. https://doi.org/10.5880/g3p.2024.001

Van Der Knijff, J. M., Younis, J., & De Roo, A. P. J. (2008). LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. International Journal of Geographical Information Science, 24(2), 189–212. https://doi.org/10.1080/13658810802549154

Prudhomme, Christel, et al. "Global hydrological reanalyses: The value of river discharge information for world-wide downstream applications–The example of the Global Flood Awareness System GloFAS." Meteorological Applications 31.2 (2024): e2192.

In addition, a few clarifications on the methods and discussions are requested in specific comments further below.

Specific Comments

Abstract: The authors state that the study's main objective "determine whether natural variability or human interventions caused these changes" in TWS variations. However, based on the presented results, the authors can only discuss this for SWS, not for TWS, since they do not analyze other storage components (see comment above).

The abstract has been thoroughly revised after incorporating the changes from the revision. The focus is now given as:

It aims to characterise and analyse the interannual TWS variations compared to meteorological observations and geodetic observations of the water storage compartments surface water, soil moisture, and groundwater.

Introduction: Clarify why were specifically the interannual variations analyzed and not (also) the seasonal variations?

Our interest in the study region was initially triggered by the strong long-term positive trend of TWS found by Kvas et al. (2023) for the Lake Victoria region. The time series decomposition of STL also allowed us to investigate changes in the amplitude of the seasonal signal. There, we found only minimal changes in the amplitude over time. Thus, we decided to focus on the interannual variations.

We added to the manuscript:

The TWS variations show a distinct and significant interannual variability but no substantial changes in the seasonal component. Thus, we focus on interannual variability in this study.

A Kvas, E Boergens, H Dobslaw, A Eicker, T Mayer-Guerr, A Güntner, Evaluating long-term water storage trends in small catchments and aquifers from a joint inversion of 20 years of GRACE/GRACE-

FO mission data, *Geophysical Journal International*, Volume 236, Issue 2, February 2024, Pages 1002–1012, https://doi.org/10.1093/gji/ggad468

L91: SPEI is typically labeled a drought index. On the data website they define it as follows: "The SPEI is a multiscalar drought index based on climatic data."

Changed

L130: Approach to estimate water area bases on optical data. How would the uncertainty of the water occurrence probability due to weather conditions affect the final SWS estimate of the study? Also, this drawback of visual light imagery has been solved by other studies that rely on radar data to detect surface water occurrence, with the advantage that they are not weather-dependent. The authors could include in the discussion, why they have not referred to such data instead, or how application of radar instead of visible light remote sensing images might enhance the accuracy of the method.

We agree that estimating water surface extents from radar imagery is an alternative method, especially during the rainy season in the tropics. However, deriving surface extents from SAR imagery is also challenging.
One limitation is the number of bands available. At least 6 different bands (red, green, blue, near-infrared and 2 shortwave infrared) can be used from optical imagery. SAR imagery provides only one band but with different polarisations (vertical/horizontal). The number of bands available from optical imagery affects the quality of land-water masks. The estimation of land-water masks from optical imagery is more accurate than that of radar images because the processing of SAR images requires speckle filtering of the speckle noise to reduce the noise. The applied thresholding based on a combination of different water indices (MNDWI, NDWI, AWEI, etc.) derived from different optical bands makes the result more accurate. The applied thresholding used for SAR imagery is based on a single band and is, therefore, not as robust as optical imagery.
It should be noted that the JRC water occurrence mask also uses optical imagery containing data gaps caused by clouds, ice cover, etc. However, using several hundred images (even with partial data gaps) since 1984 to estimate the water occurrence mask leads to robust results of the water occurrence mask, even if not all periods can be covered. The increase in temporal resolution over the last few decades has also improved the ability to monitor flood events in our study area.
Finally, we agree that radar imagery could improve the estimation of the land-water mask, especially in the rainy season, but to our knowledge, no datasets equivalent to the JRC dataset based on radar imagery exist. Furthermore, the processing of land-water masks from SAR imagery is beyond the scope of this paper.

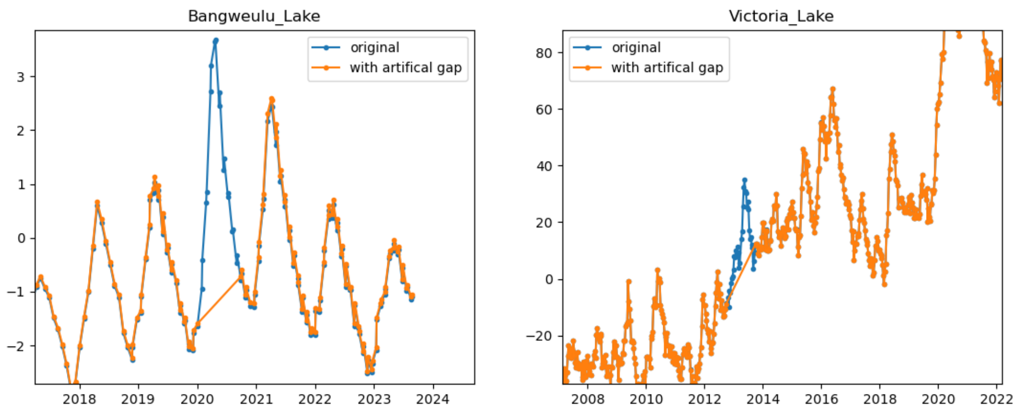
We do not think that discussing alternative water area estimates would benefit the manuscript.

L137: culuative > culmulative

Changed

L145: Add a statement to further spell out what your assumption on the lake profile shape for the volume estimation is, e.g. how steep is the pyramid wall inclined?

The pyramid formula assumes linear lake profiles between the heights of two observations. In most cases, the differences between consecutive height observations are as small as a few centimetres, where this assumption is reasonable. However, the differences can be as large as one metre due to data gaps or rapid changes in the water level observed only with sparse altimetry coverage. Here, the assumption is indeed questionable. To investigate the effect, we tested introducing large artificial data gaps in the water level time series and compared the resulting volume time series. The following figure shows both the original volume time series and one computed with the data gaps for Lake Victoria and Lake Bangweulu.



Lake Bangweulu has the most complicated profiles of all lakes in the study area and also the most significant surface extent variations (compared to size). But even for this lake, the effect is very small. For Lake Victoria, on the other end of the spectrum, no change at all is visible.

We added a discussion about this to the manuscript:

The pyramid formula assumes linear lake profiles between the two WL observations. In most cases, the differences between consecutive height observations are as small as a few centimetres, where this assumption is reasonable. Nevertheless, the differences can be as large as one metre due to data gaps or rapid changes in the water levels observed with temporally sparse altimetry. Thus, we tested the assumption by artificially removing WL observations. We found only very minor differences in the resulting volume time series with *and without artificial data gaps. Especially given the uncertainties of the WSA and WL data (see below), these differences are negligible.*

L30ff/L147: Please clarify, if all lakes in the region were included? Or to what percentage are smaller lakes neglected?

No, only those lakes were considered where surface elevation data from altimetry are available. However, this accounts for 94% of the lake area in the study region (according to the outlines of GLWD). This information has been added to the manuscript (Section 2 Study Region):

All lakes that are accessible with satellite altimetry are included in this study. They account for 94% of the surface water bodies, by area, of the region according to the Global Lake and Wetland Database (GLWD).

Equation 2) How representative is such a profile for the lakes? This approach probably has some uncertainty because the lake wall angle is likely heterogeneity inclined, for example, shallower near the shore. Can this introduce a significant error to the total surface water storage estimate? And how large is the uncertainty? It would help to provide a reasonable range of uncertainty for this.

See also the answer above. The method's assumption will introduce no significant error.

L151: I appreciate that the authors spatially filter the surface water data to mimic the sensitivity of the GRACE observations to water mass changes. The author's did not, however, clearly state if the applied gaussian filter width of 350 km is comparable to that applied during the GRACE data processing as conducted for the COST-G dataset. A different filter width can significantly alter the amplitude in storage variations. Since the GRACE dataset used is a unified from various datasets, this might be a bit more complex to evaluate. However, a discussion of it is missing. Optionally, this could be included as another source of uncertainty in the surface water storage time series.

The TWS data set is filtered with the time variable anisotropic VDK filter (Horvath et al. 2018). The different data sets in COST-G are combined on an L2 basis; thus, the combined L2 solution is subsequently filtered as one. For clarity, we added both pieces of information to the TWS data section.

In the meantime, for the G3P groundwater product mentioned above, we investigated which filter width of the Gaussian filter best fits the spatial resolution of GRACE-based TWS filtered with VDK. The related publication of Sharifi et al. is being prepared, and we hope to be able to cite it upon publication of this manuscript. So far, we can only provide a technical report as a reference (Güntner et al., 2023). According to this, a Gaussian filter width of 250km is best suited for data sets of water storage compartments to make them comparable in spatial resolution to VDK-filtered GRACE-based TWS data. We thus changed the SWS filtering to this value.

Of course, using one isotropic filter for all water storage compartments is a simplification. However, we consider it reasonable to allow comparability between the storages.

We added more information about the filter to the manuscript:

Second, we employ a spatial Gaussian filter with a half-width of 250km to mimic the spatial resolution of TWS. The half-width of 250km has been found by comparing the empirical spatial correlation function of TWS and other water storage compartments filtered with different Gaussian filters (Güntner et al., 2023). That results in the surface water storage (SWS) data set for this study.

Horvath, A., Murböck, M., Pail, R., Horwath, M., 2018. Decorrelation of GRACE time variable gravity field solutions using full covariance information. Geosciences 8, 323. https://doi.org/10.3390/geosciences8090323

Güntner, A., Sharifi, E., Haas, J., Ruz Vargas, C., and Kidd, R.: Deliverable 4.1 – G3P Product Report – Revision 1, 2023.

151: Please indicate how the filtering was conducted, e.g. in the spatial or frequency (spherical harmonic) domain.

See the answer above.

L154: the term “simple” is vague here. I assume you are referring an assumption for stationarity of the temporal components in the time series, as stated further below in L159? Different approaches available (e.g. fourier based, or others) are not more or less simple, but instead they are potentially better applicable to climate processes. Also, the non-stationarity of climate signals is not only present in seasonal components but also in the inter-annual/trend components, hence, why STL is better applicable for both. Please rephrase to make this clearer.

True, the term “simple” here does not fit. We removed it and added the possible change in the amplitude of the seasonal signal due to climate change:

A time series decomposition into deterministic periodic (e.g. annual and semiannual) signals and a linear trend is not well suited to characterise the temporal variations of TWS in Africa over the available 21 years. It cannot describe the substantial interannual variability and possible change in the seasonal amplitude due to climate change.

L160ff: how does the smoothing parameter affect the signal decomposition? What was the criteria for choosing them. I understand this is a trial and error approach, and requires some empirical decision making. However, it would be good to try to write down what you were aiming for, when choosing the parameter.

Please also see our answer to Reviewer #3 on this question. The choice of n_i, n_o, and n_l is straightforward. The most difficult choice is n_s (smoothing length of the seasonal signal), on which n_t (smoothing length of the trend/interannual signal) depends. We chose n_s such that the seasonal sub-cycle (collection of all values of a given month) no longer contains variations that can be interpreted as noise. N_s=35 was the smallest value where we considered the seasonal sub-cycles noise-free. Our reasoning for choosing the parameter values is included in the manuscript:

The results of the STL decomposition depend on several parameters that govern the smoothness of the interannual and the annual signal. Cleveland et al. (1990) provide guidelines for choosing the parameters, which we used together with empirical testing and visual inspection. That results in the following parameter values. n_p: Length of annual signal, 12 in our case; n_i and n_o: the number of passes through the inner and outer loop, set to 1 and 10, respectively; n_l: the width of the low-pass filter, to be set to the least odd integer larger than n_p, thus set to 13; and n_t and n_s: the trend and seasonal signal smoothing parameter, both set to 35. While the former four parameters are straightforward, n_s requires more considerations. We chose the value n_s=35 in such a way that we consider the interannual variability of the seasonal signal no longer governed by noise. n_t depends on the value of n_s. However, we found that the value provided by the rationale given in Cleveland et al. (1990) (n_t=19) produced a trend component containing still too many short-term variations. Finding the value n_t=35 was done with empirical testing and visual inspection.

L156/Section 3: Please indicate if the STL is loss-free or not.

STL is loss-free, and we added this to the text.

L171/Figure 4: If I understand this correctly, the black time series (original in a) is corresponding to the blue long-term signal in b (no-data gap)? I wonder if it makes sense to match the color (same in c and d)?

You are correct! We changed the colours in the figure.

Figure 5&6: The clusters are coded two ways, once by colors and once by numbers. It would be easier to if this is limited to either one. Or also add the colors in the titles, behind numbers in figure 5, e.g. cluster 5 (red) and add numbers to colored dendogram in Figure 6.

Following your very good suggestion, we added the colours to the subfigure captions in Fig. 5 for easier recognition. However, we like to keep the numbers as well, as they are very helpful in the text, and not every reader can distinguish colours in figures that well (although we used a colour-blind-friendly colormap). Fig. 6 is removed (see answer below).

Figure 6: I was wondering, if it would be sufficient to have this in a supplement. The additional information is minor, as the time series in Figure 5 already show degree of similarity.

L206: I suggest to add brief explanation: regions with overall positive trend are those located in Central Africa (including blue, yellow, dark green, pink).

L207ff: Here, the authors shift from a 7-cluster analysis to an 8-cluster analysis without a more detailed explanation. This should either be a new paragraph, to make that shift more clear. Alternatively, I am wondering if Figures 5-7 could be combined. For example, why is cluster 8 not also shown in Figure 5?

L207: if I understand it correctly, the sub-clusters in Figure 7 are also appearing in the cluster tree in Figure 6, as the authors emphasis on that here. However, in Figure 6 they are colored all light blue. I was wondering, if it makes sense to mark the purple cluster 8 also in Figure 6, to be more clear.

Answer to the four comments above: We realised that Fig 6 and Fig 7 and the corresponding text were more confusing to the reader than intended. Our original intention in this part of the manuscript was to discuss the choice of cluster number in more detail and interpret the dendrogram shown in Fig. 6 more. Thus, we removed both figures. At the same time, we added some more discussion about the meaning of the clusters and the choice of cluster numbers to the manuscript as follows:

The numbering of the clusters does not have any further meaning. However, by step-wise increasing the number of regions m the dissimilarity of a found cluster to the other regions can be investigated. If we assume only two clusters (m=2), the algorithm first separates Madagascar Island from mainland Africa due to its spatial disconnection. With m=3, Cluster 0 is already separated from all other regions in mainland Africa, indicating that the TWS signals are most distinct compared to all other TWS signals. The distinction between the TWS signals of the clusters can be measured by the Euclidean distance, on which the algorithm is based, between the mean time series. We found that Cluster 0 has the largest Euclidean distance to all other clusters.

The following two regions to be split off are Cluster 5 and Cluster 6, which both have distinct interannual variations, too. As m further increases, the split-off clusters become less distinct and have a larger signal spread within.

We decided on the final value for m based on the results, especially the size and shape of the regions. We sought the largest number of clusters while keeping them reasonable for GRACE data interpretation. Here, we found m=8 to be the optimal number.

The ninth cluster would be ring-shaped and only about 100km across in the narrowest place. Thus, such a region is no longer meaningfully interpretable with GRACE data.

L209: … has even larger TWS amplitudes than … > … has a larger TWS amplitude than ...

L210-211: change the word "marked" to "significant","distinct", or "fast"

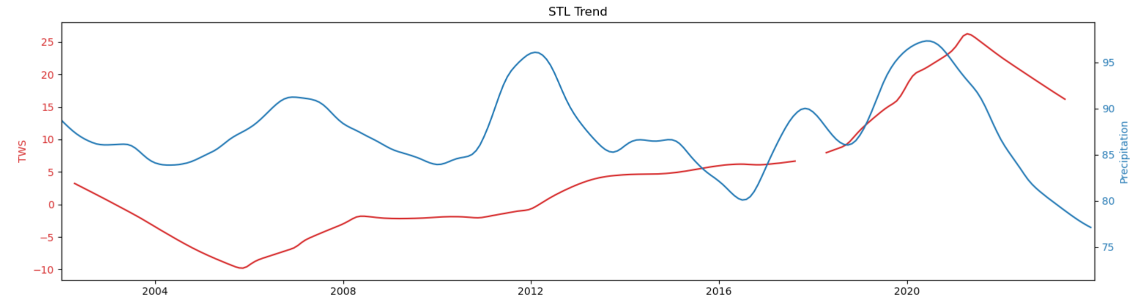
Both sentences are no longer in the manuscript.

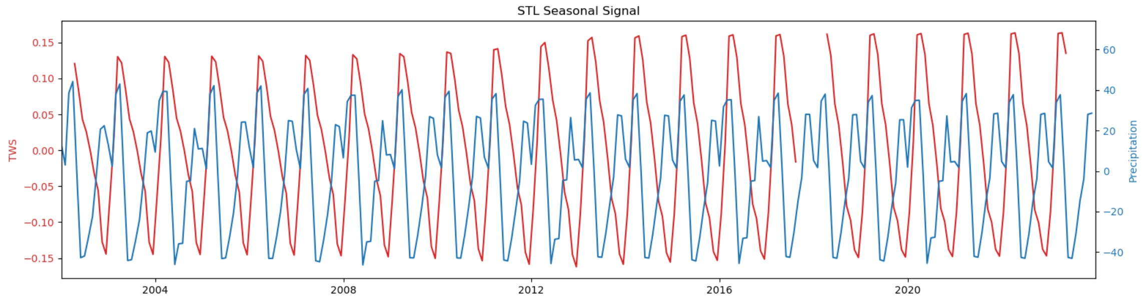
L214/Figure 9: The graph in Figure 9 does not look like the values are accumulated, but rather filtered with some kind of moving-window filter of certain width (or accumulated within a moving window). In case of only accumulating, you would have only values every n months, with n being the accumulation period. Please clarify.

Throughout the manuscript, we use accumulation with a moving window, which is standard for the published SPEI. We clarified this in the text.

L214-215: You compare accumulated precipitation with SST filtered TWS. The two time series are treated with different methods. Are they really comparable this way? Why do the authors not also apply an SST filter (using the same parameter as for TWS) to the precipitation data instead? This would also save them from estimating the correct filter-width for P.

We tested your very interesting suggestion and present the results for the seasonal and trend/interannual signals in the figure below.





The interannual component of the STL decomposition of precipitation (blue line in the trend plot above) is mainly similar to the accumulated precipitation shown in Fig 9 of the manuscript. TWS has only an annual seasonal signal at the seasonal scale, but precipitation also has a semi-annual seasonal signal (lower plot).

If we were only investigating TWS vs. precipitation, using the STL decomposed precipitation time series would indeed be a good idea for further investigation. However, we also investigated the drought index SPEI, where we had to decide on an accumulation period. Thus, we decided to keep processing the meteorological data sets (precipitation and SPEI) as in the original version of the manuscript, as the same procedure could not be applied to all of them.

Figure 8 might also be ok for a supplement, instead of the main manuscript?

L220-2029: I am wondering if this can be shortened, as P becomes less relevant given their concluding that E is missing to better compared to TWS. However, this conclusion is rather trivial from a hydrological perspective.

L218: Maybe add a sentence explaining the purpose of the violin plot. Does the change in width of the blue areas (violins) have any meaning?

Answers to the three comments above. Following the suggestion of Reviewer #3, we used the parameters chosen for the TWS STL decomposition as a guiding light for the correct accumulation period. We found that we need at least a 3-year period to estimate the annual and trend components reliably (see also our answer above to your question regarding the parameters). Thus, we should also choose at least an accumulation period of 3 years for the meteorological data sets. With the investigations shown in the manuscript (so far), we reached a similar conclusion. But as you correctly state, it is relatively trivial from the hydrological perspective and will be significantly shortened in the revised manuscript. Although we found a higher correlation between TWS and SPEI48, we decided to use 36 months accumulation period in the revised manuscript.

We changed the manuscript accordingly and remove Fig. 8.

L232-233: unclear formulations, please rephrase a bit simpler.

L233-234: unclear formulation, rephrase. “…longterm observation of ?”; also you do not put P-E in relation to TWS, but SPEI

Changed former L232-234 to: *For precipitation minus (potential) evapotranspiration (P-ET), we do not use direct observations but the Standardised Precipitation-Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010). This index relates current P-ET observations to the long-term observations since 1955.*

Figure 9: add precipitation to the legend.

Added

L243: do > does

Changed

L253: for the names > for their names

No longer in the manuscript

L256: I cannot see the 50% in Figure 10, the color bar is kind of vague. The top left Figure 10 colors seem saturated given the color bar. What are the maximum value in Figure 10 top row? It looks to me more like 30%, given the time series in Figure 10 bottom.

Figure 10: The red polygon shown in the upper three panels is neither labeled the legend, nor in the caption. I assume it is outline for cluster 7? Please add.

During the large and widespread revision of this section, we decided to remove Fig 10 from the manuscript. With the inclusion of more water storage compartments, namely root zone soil moisture and groundwater, part of the discussion of SWS was removed. We no longer investigate the spatial pattern of SWS-Lake Victoria separately.

L261: space missing

Removed from manuscript

Figure 11 caption: correct spelling of de-sesonalized

Figure 11: compares PEV and correlation for de-sesonalized SWS and TWS. It would be useful to show the deseasonalized time series somewhere, e.g. add to Figure 11 or Figure 10 bottom?

The caption was incorrect, as we already used the full time series for PEV and correlation before. The figure now also contains PEV and correlation to all WSCs. The new Fig 7 shows the mean time series for all WSCs.

L285: the 50% occur only for years with very low TWS, but not for wetter years. Hence, this feels like an overstatement (also in the abstract). Maybe it would be more representative to also estimate the median or mean of the explained percentage over the years? Or it would be more transparent to discriminate between dry and wet years (see also comment for abstract above)?

We changed the statement to a more moderate one. In median Lake Victoria explains 63% of the SWS change, thus dominating SWS. We did not find a clear relationship between the magnitude of annual SWS change and the explained percentage of Lake Victoria. Further, we also looked into the percentages for GWS and SWS vs TWS and again found no simple relationship (e.g. drying years have more SWS influence, and wetting years have more GWS influence). The discrimination between different years and their drivers is more complicated to discuss, which is now included in the manuscript.

L289: Victoria Nile > Nictoria Nile River

Changed

L291-295: this information might be better suited already in Section 2.3 to provide more detail on the surface water bodies in the region and how they are managed. It would already help for understanding previous sections.

After the introduction, we added a new section, "Study Region", which collects the information in one place.

L311: Can you provide a reference to support this statement?

L235: govern > governed

Changed

L335-335: sentence unclear, reformulate

Changed.

Figure 13: This is not compiled well to support the discussion in Section 7. Maybe presenting the time series in a single or stacked panels and/or in comparison to TWS and/or SWS time series would help the purpose more?

Following your advice, we stacked the time series plots above each other. Plotting them together in one plot (after shifting the time series to a common basis) made reading and understanding the figure even more complicated. We also took care, that the y-axis is in the same resolution, i.e., covering 5m WL for all four plots, to make the amplitude easier to compare.

L363: reformulate sentence, a lake cannot lead, rather results for the lake.

L360-362: I disagree, SWS does not fully explain the steady increase of TWS, as shown in Figure 10, only partially. The value of this multi-year TWS/SWS rise was also not quantified in the manuscript, maybe it would help to add this?

L366: The connection between dam discharge and TWS is not clearly shown in the manuscript.

We thoroughly revised the conclusion, taking the three comments above into account.

References

Werth, S., White, D., & Bliss, D. W. (2017). GRACE Detected Rise of Groundwater in the Sahelian Niger River Basin. Journal of Geophysical Research: Solid Earth, 122(12), 10,459-10,477. https://doi.org/10.1002/2017JB014845