Dear reviewer, dear Ehsan,

Many thanks for reviewing our manuscript and for providing feedback to improve the manuscript. Our responses to your four comments are given below in reddish color.

Dear Authors,

I have read your manuscript with great interest. The study presents a data-driven "residual" approach to estimate groundwater storage, where time-variable terrestrial water storage (TWS) changes derived from GRACE and GRACE-FO are interpreted as the vertically integrated signal of total land water storage. Estimates of other land water components are then obtained from independent satellite observations or models and subtracted from TWS. To enhance consistency among the different datasets, the manuscript introduces spatial filtering techniques.

While the resulting groundwater estimates are potentially valuable, I believe several important technical issues need to be addressed to strengthen the robustness and interpretability of the results. These points likely require additional analysis, validation, and deeper discussion.

1. Filtering strategy

The filtering approach adopted from Sharifi et al. (2025) is not entirely convincing from a physical perspective. Defining an "optimal" filter based on maximizing correlations between individual storage components and TWS lacks a clear physical justification. If the objective is to ensure spectral consistency, a more transparent approach would be to represent the storage components in the spherical harmonic domain and apply filters consistent with GRACE(-FO) processing. This would allow consistency to be enforced directly in the spectral domain rather than through empirical tuning.

The goal of the filtering strategy adopted in this manuscript following Sharifi et al. (2025) is to make global data sets of water storage variations of different storage compartments compatible for the data combination process in terms of their spatial resolution and smoothness to the TWS anomaly Level-3 data products derived from GRACE/GRACE-FO satellite gravimetry. As a proxy for characterizing the spatial smoothness of the GRACE-based data, Sharifi et al. (2025) suggested the spatial autocorrelation function of water storage anomalies as an empirical, though physically meaningful measure. This measure is then used to parametrize the strength of a (Gaussian) smoothing function applied to the storage compartment data sets. In line with the reviewer's suggestion, Sharifi et al. (2025) also analysed an alternative approach: transformation of the storage compartment data sets into the spherical domain in which GRACE Level-2 data are provided, and subsequent GRACE-like filtering with the DDK or VDK filter that is used for GRACE data processing to suppress anisotropic correlated noise. However, the results show that this method is not suitable for non-GRACE-based storage data sets as it introduces spurious longitudinal stripes at the global scale (see Figure 8 in Sharifi et al. 2025). This approach has thus been discarded for G3P processing. To point this out, we add a sentence to the revised manuscript in Chapter 2.3.3.

2. Consistency of storage magnitudes

Even if the filtering approach is accepted, there is no guarantee that independently derived storage components (from models or other satellite products) will match the magnitude of GRACE(-FO) TWS. Such inconsistencies are well documented and can be substantial. They are more appropriately addressed through data assimilation frameworks that explicitly account for uncertainties and model–data discrepancies. I encourage the authors to consider and discuss this aspect, for example as highlighted in Getirana et al. (2025, GRL, Inconsistencies in GRACE-Based Groundwater Storage Estimation—A Call for a Proper Use of Land Surface Models - Getirana - 2025 - https://doi.org/10.1029/2025GL119197).

In addition, several offline data-merging approaches have been proposed in the literature that attempt to address such inconsistencies more systematically. These include ensemble-based approaches, statistical decomposition methods, and Bayesian merging frameworks (e.g., https://doi.org/10.5194/hess-18-2955-2014; doi:10.1016/j.rse.2013.09.025; https://doi.org/10.1016/j.scitotenv.2020.143579). A discussion of how the present method relates to or differs from these established techniques would significantly strengthen the manuscript.

We acknowledge the reviewer's comment of possible inconsistencies between the sum of storage variations of independently derived water storage compartments (WSC) and GRACE/GRACE-FO-based TWS. In the original manuscript we addressed this already in the introduction by referring to the comprehensive study of Akl and Thomas (2024) on this topic (lines 113-115). In the revised manuscript, we add Getirana et al. (2025) as an additional reference that points to this issue, following the suggestion of both reviewers. In the original manuscript, we took up the topic again in the discussion chapter and devoted a full paragraph to it (lines 773-781). We see the reviewer's point that inconsistencies between TWS and the sum of individual WSCs might be suppressed by using a data assimilation framework in which the model's water balance is forced to be closed. Nevertheless, this approach does not necessarily guarantee that the separation of TWS variations into individual storage compartments such as groundwater is closer to reality. The separation depends, e.g., on the model structure, including the way if and how individual WSCs are represented in the model, and on the uncertainties of the individual WSC data that are assimilated into the model, determining the weight with which they constrain individual WSCs. In a similar way this can apply to offline data merging approaches that ingest model outputs, as those mentioned by the reviewer, with references partly added by us to the revised manuscript.

We stress that for the G3P approach presented here, we take an observation-based and constraint-free approach with consistent global coverage as far as possible, based on mature and well-documented satellite- and in situ-based data products of Copernicus services. We consider the uncertainties of TWS and of each WSC and propagate them to the residual groundwater data set. This approach can provide insight into the hydrological system of terrestrial water storage dynamics and indicate to which extent the balance closure of storage data sets that can be achieved with current observation capabilities. We consider the approach to be complementary to model-based approaches and offline data merging frameworks that are forced to converge. We believe that, ultimately, taking a holistic view and comparing the various approaches on an equal footing can advance the quality of quantifying storage dynamics. A plausibility testing framework such as suggested by Arifin et al. (2025) can then help sorting out physically plausible WSC combinations that are consistent with hydrological processes, and ultimately contribute to improving observation systems towards better water balance closure. We laid out this reasoning in the mentioned paragraph already in the original manuscript and slightly expand it in the revised manuscript.

3. Magnitude of groundwater estimates

The magnitude of the derived groundwater storage changes appears questionable, possibly due to over-filtering or the rescaling procedure. For instance, the spatial patterns and amplitudes shown in Figure 7 over Europe seem unrealistic. A comparison against independent estimates, such as outputs from global hydrological models (e.g., WGHM) or GRACE-based groundwater products (e.g., constrained estimates available via Figshare: https://doi.org/10.6084/m9.figshare.31073527), would be helpful to contextualize and validate the results.

We apologize that in a global map as the one shown in Figure 7, the colour scale needs to cover a wide range of values. Thus, recognizing spatial patterns at the coarse resolution becomes difficult. This also applies to Europe where seasonal amplitudes of GWSC in the G3P data set are overall small, as noted by the reviewer. Assessing whether these and other GWSC amplitudes worldwide are questionable or unrealistic requires a thorough validation by independent observation-based data. However, such data are not directly available at the large spatial scales represented by the low-resolution GRACE and G3P data because (1) in-situ observations of groundwater level variations in monitoring wells are often sparse in space or hardly available and (2) converting these water level variations into basin- or aquifer-average groundwater storage changes at the GRACE/G3P-scale is a challenging step by its own: it requires poorly available information on the storativity or specific yield of the aquifer in which the observation well is located, and a regionalization method of these values and of the hydrogeological units to the large spatial scale of interest. A comprehensive validation of G3P in this way is subject to ongoing work. Similarly, as we emphasis in the conclusions of our manuscript, we encourage validation feedback from data users who compare the G3P data for their regions of interest with suitable in-situ observations.

For Europe, we included one evaluation case in the original manuscript: the Seine basin aquifer system (see Figures 6 and 16). G3P dynamics fit well to the in-situ data, albeit with higher noise and somewhat smaller seasonal variations, but well capturing the inter-annual groundwater dynamics. Again, still lacking the conversion of the in-situ groundwater levels to aquifer-average groundwater storage variations, the real validation of the G3P-based GWSC is subject of future work. Furthermore, given the limitations of model-based groundwater storage variations as also mentioned above in the response to the reviewer's second question, we do not consider the comparison of G3P to a model-based GWSC as a comprehensive validation option. We test and illustrate this for the evaluation examples used in the manuscript with in-situ groundwater level data: the Paris basin, the High Plains Aquifer and the Guaraní Aquifer system. For these regions, simulated area-average GWSC was subset from the output of hydrological models that are part of ISIMIP3b (Gosling et al. 2025). ISIMIP models that represent GWSC for the study period are CWatM, H08 and WaterGAP. The results show (see Figure I below) that (i) simulated GWSC varies considerably between the individual models (in particular for the High Plains and Guarani aquifers) and (ii) model-based GWSC dynamics are largely different from the dynamics of in-situ-observation-based aquifer-average groundwater levels (compare Figure 16 in the original manuscript). A conclusive validation of G3P-based GWSC is thus not possible.

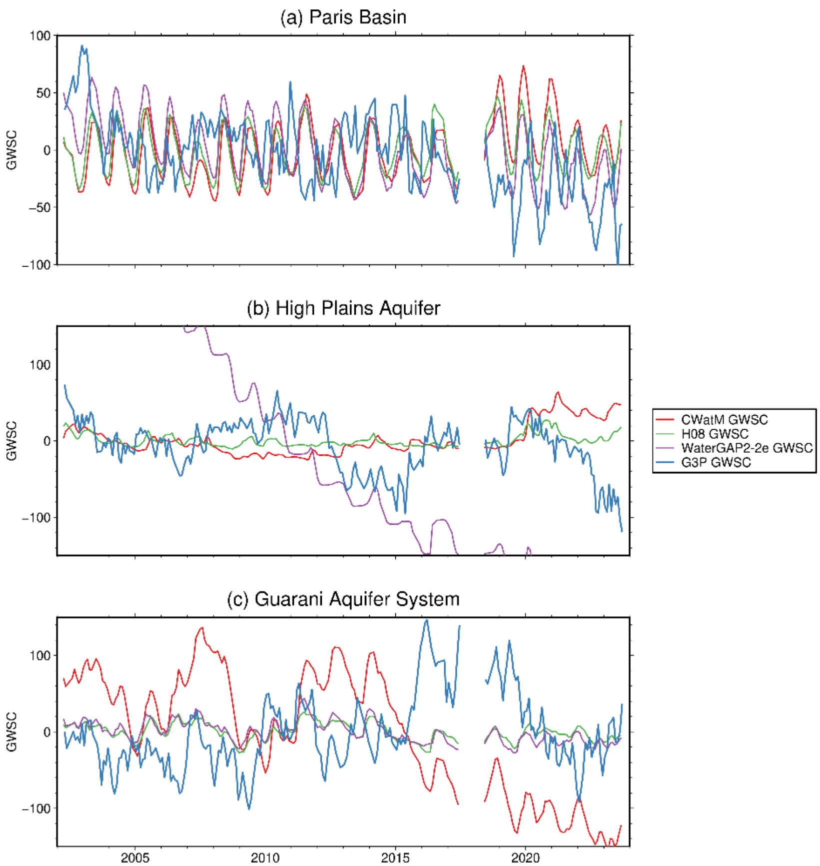

Figure I: Groundwater storage change (GWSC) of G3P in three aquifer systems, compared to simulated GWSC of three global hydrological models that are part of ISIMIP (all data are given as area-average anomaly values in mm water equivalent).

4. Rescaling procedure

The manuscript would also benefit from a clearer description of the rescaling methodology. While filtering is described, it is not sufficiently clear how the final amplitudes are restored or interpreted. Since end users are typically interested in physically meaningful magnitudes, a transparent explanation of how near-realistic amplitudes are ensured is essential.

For G3P, no rescaling is carried out, neither at the level of the GRACE/-FO-based TWSC data nor at the level of the final GWSC data. In the original manuscript, we devote paragraphs in the introduction, in the methods section (lines 550-565) and in the discussion part (lines 795-804) to the issue of leakage mitigation and rescaling. We point out that while several approaches for leakage mitigation or rescaling have been proposed in the literature in the past, a consolidated way of proper leakage treatment of mass change data derived from satellite gravimetry is still a topic of ongoing research and debate in the geodetic community such as in the International Association of Geodesy (IAG) working group JWG 2.3.1 on leakage mitigation (https://geodesy.science/com2/structure/jwg/jwg-2-3-1/).
To mention a few issues at this point: The degree of leakage in gravity-based TWSA or GWSC data – and thus the degree to which leakage mitigation or rescaling eventually needs to be applied - varies in space and in time, depending on the spatial and temporal patterns and correlations of the actual TWSA and GWSC signal. This signal, however, is poorly known from independent sources, i.e., in-situ data or hydrological models, in a comprehensive form (see our discussion above in response to the reviewer's comment 3 and in the manuscript on the many limitations of in-situ based and model-based TWSA or GWSC information). Thus, rescaling based on such information would imply a lot of uncertainty and it is not clear whether the resulting rescaled data would really come closer to 'near-realistic amplitudes' as suggested by the reviewer. Also, it should be noted that rescaling of amplitudes needs to cover all time scales, i.e., the amplitudes of seasonal, interannual and long-term (trend) storage variations, which makes a suitable rescaling procedure even more difficult to develop. Given these limitations and unknowns, we decided to provide the G3P data set without applying any leakage or rescaling procedure. We make this point clearer in the respective paragraph of the discussion chapter, and expand on the limitations that this may imply.

Overall, I believe the study addresses an important topic, and the proposed dataset could be valuable to the community. Addressing the points above, particularly with respect to methodological justification, consistency, and validation, would considerably improve the quality and reliability of the manuscript.

I hope these comments are helpful, and wish you all the best.

Ehsan Forootan

References:

Gosling, S.N., Müller Schmied, H., Bradley, A., Burek, P., Gedney, N., Grillakis, M., Guillaumot, L., Hanasaki, N., Ito, A., Kou-Giesbrecht, S., Koutroulis, A., Nishina, K., Otta,K., Qi, W., Sahu, R.-K., Satoh, Y., Schewe, J., 2025. ISIMIP3b Simulation Data from the Global Water Sector. https://doi.org/10.48364/ISIMIP.230418.4