Referee #1

1. Overview

The article proposes using Terrestrial Water Storage Anomalies (TWSA) from GRACE missions to independently evaluate precipitation products, specifically GPCC and GPCP, since traditional methods rely on gauge data used in the products' creation. By calculating Drought Recovery Times (DRT) with TWSA alone and combined with precipitation data, the authors aim to provide a more robust assessment of the alignment between precipitation products and TWSA.

Authors: We thank the reviewer for the invaluable comments. In this version, we have revised the text according to the suggestions. Our responses are given below in red while the reviewer's comments are given in black.

2. Major comments

Throughout the article, I encountered many interesting results, but the core scientific analysis was lacking. It was unclear what these findings reveal, how they answer the research question, or how the new method clarifies the strengths and weaknesses of the precipitation products. Specifically, there is no discussion of contexts in which one product outperforms the other, nor an exploration of why this might be.

Additionally, the research question itself is not well-defined. While the abstract claims the goal is to evaluate precipitation products using TWSA data to calculate DRT, the paper sometimes shifts focus to evaluating TWSA products or comparing which precipitation/TWSA combination best estimates DRT.

To address this, I recommend: 1) clearly defining the research question in the introduction and maintaining alignment throughout, and 2) adding a discussion section (included in or separate from the " results " section) to interpret the results in light of the research question.

Authors: Thank you for your helpful comments. In this revised version, we have substantially refined the research question and integrated a discussion section within the results. We have also addressed the other feedback received. To clarify the research question, we have updated the abstract introduction and added a discussion section as suggested as follows:

"This study aims to assess the accuracy of the Global Precipitation Climatology Center (GPCC) Full Data Monthly Product v2022 and Global Precipitation Climatology Project (GPCP) v3.2 Monthly Analysis Product by estimating hydrological drought recovery time (DRT) from precipitation and terrestrial water storage anomaly (TWSA) acquired from satellite gravimetry. This study also evaluates the performance of G3P and JPL mascon TWS monthly-solutions from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) satellite missions."

Following the above comment, we now revise the title of the manuscript to better reflect the focus of the study.

“Evaluation of Globally Gridded Precipitation Data and Satellite-Based Terrestrial Water Storage Products Using Hydrological Drought Recovery Time”

Additionally, to better highlight the goal of the study, now the relevant section in introduction is modified as follows:

“The current study aims to independently evaluate and compare frequently used global gridded precipitation products (i.e., from GPCC and GPCP) by using the GRACE/GRACE-FO TWS data (i.e., the JPL mascon and G3P products) in order to assess drought conditions. Also, this research evaluates the performance of the JPL mascon and G3P TWS products. Both evaluation is conducted by estimating DRT based on TWSA and required precipitation amount. Comparing the suitability of these precipitation and TWS products for global hydrological applications across various Köppen-Geiger climate zones enhances our understanding of the relationship between hydrological droughts and global precipitation and TWS products through DRT estimations.”

Again, following the reviewer’s comments, we have now added a new section called *3.4 Discussions* to interpret the results in light of the research question, as follows:

“Both precipitation products provided similar global mean DRT estimations (12.6 months), with a high consistency rate of 87.5%. The largest discrepancy in mean DRT estimation (0.1 months) was observed in the polar (E) zone, with no significant difference in the snow (D) zone. The consistency between the two precipitation products was less than 1% across all the climate zones.

For the TWS products, the global mean DRT estimation using JPL mascon (13.8 months) was 2.6 months higher than that of G3P (11.4 months). G3P exhibited 5.0% higher global consistency (90.0%) than did JPL mascon (85.0%). The largest difference in mean DRT estimation between G3P (13.2 months) and JPL mascon (18.9 months) occurred in the polar (E) zone, with the smallest difference between G3P (10.0 months) and JPL mascon (10.6 months) in the equatorial (A) zone. The greatest consistency disparity (7.0%) occurred in the polar (E) and arid (B) zones, with G3P (78.6% and 93.3%) outperforming JPL mascon values (71.6% and 86.3%) in the polar and arid zones, respectively. The smallest difference (1.3%) between G3P (98.5%) and JPL mascon (97.2%) was recorded in the equatorial (A) zone. In terms of consistency across all the climate zones, G3P outperformed JPL mascon.”

Finally, the second paragraph of section *4 Summary and Conclusions* have been substantially revised following this comment:

“GRACE/GRACE-FO directly provide water storage anomalies, offering a novel approach to characterize drought by assessing the storage deficits. The time required for drought recovery can be directly derived from the temporal evolution of these deficits (Singh et al., 2021), enabling the measurement of both drought duration and severity. Both GPCC and GPCP products exhibited not only similar mean DRT estimations but also comparable consistency in their DRT estimations, globally and across all the Köppen-Geiger climate zones. For the TWS products, the mean DRT estimations from JPL mascon were, on average, 2.6 months higher, than those from G3P, globally and across all the Köppen-Geiger climate zones. However, G3P

showed slightly higher consistency in the DRT estimations (5.0% difference) than did JPL mascon. Furthermore, G3P demonstrated greater consistency than JPL mascon across all the Köppen-Geiger climate zones."

3. Minor comments

Abstract: Consider simplifying the abstract by emphasizing the key findings, rather than delving into specific details. This will help focus the reader's attention on the main outcomes without overwhelming them with too much information.

Authors: Now we have simplified the abstract by emphasizing the key findings and also reduced the length from 362 words to 257 words. Revised abstract reads as follows:

"Accurate precipitation observations are crucial for understanding meteorological and hydrological processes. Most precipitation products rely on station-based observations, either directly or for bias-correcting satellite retrievals. To validate these station-based precipitation products, additional independent data sources are necessary. This study aims to assess the performance of the Global Precipitation Climatology Center (GPCC) Full Data Monthly Product v2022 and Global Precipitation Climatology Project (GPCP) v3.2 Monthly Analysis Product by estimating hydrological drought recovery time (DRT) from precipitation and terrestrial water storage anomaly (TWSA) acquired from satellite gravimetry. This study also evaluates the performance of G3P and JPL mascon TWS monthly-solutions from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) satellite missions. The current study employed two methods to estimate DRT and evaluated the consistency of DRT estimations by calculating the time difference between the two methods. No significant differences in the mean DRT estimations using GPCC and GPCP were found for both globally and across all climate zones, with comparable consistencies from GPCC and GPCP. For the TWS products, DRT estimation using JPL mascon was, on average, 2.6 months longer than that using G3P. However, the G3P showed approximately 5.0% higher consistency than the JPL mascon globally and across each climate zone. These results indicated a close agreement between GPCC and GPCP in DRT estimations. G3P showed greater consistency in DRT estimation with the precipitation products than JPL mascon. These findings provide valuable insights into the accuracy of precipitation and TWS anomaly products by utilizing hydrological drought characteristics, enhancing our understanding of meteorological and hydrological processes."

Line 28: You present the consistency results without explaining what they entail. Providing context here will help the reader understand the significance of these results.

Authors: To clarify the consistency result, we have added the following text in the manuscript (Abstract):

"The current study employed two methods to estimate DRT and evaluated the consistency of DRT estimations by calculating the time difference between the two methods."

Line 178: The detrending process would benefit from more explanation. Please consider adding details on the method used, along with a relevant reference.

Authors: To clarify the detrending processes, we have now revised the following text and added the relevant reference in the manuscript. Also, we have added the equation for the detrending process (Section 2.3):

"To isolate the impact of such long-term processes, we detrended the TWSA data for each grid by removing the linear trend of relevant grid (Singh et al., 2021)."

$$\text{“}dTWSA_{x,y,t} = sTWSA_{x,y,t} - trend(sTWSA_{x,y,t}), \qquad (2)$$

where $sTWSA_{x,y,t}$ *is the smoothed TWSA at x,y grid point and time t, and* $trend(sTWSA_{x,y,t})$*, is the trend of the smoothed TWSA at x,y grid point and time t."*

Line 287: You describe the correlations over Australia (0.55), South America (0.46), and South Africa ($\rho > 0.47$) as "high." Please clarify the criteria or thresholds you used to define these correlations as high.

Authors: All the r values ≤ 0.13 were not significant ($p > 0.05$; $n = 216$). To clarify the threshold which we used to define these correlations, we provide the following classification in the manuscript (Section 2.5):

"We classified the r values as follows: no or insignificant correlation (0.0–0.13), weak correlation (0.14–0.39), moderate correlation (0.40–0.69), and strong correlation (0.70–1.0)."

Also, we have revised the text in *Section 3.1* accordingly:

"All the r values ≤ 0.13 *were not significant* ($p > 0.05$; $n = 216$)*. Significant and moderate correlations were found over Australia (0.55), South America (0.46), and southern Africa (0.60), where not only are water storage variations substantial, but also in situ observing networks are dense. These correlations indicate substantial agreement in these areas."*

Line 298: The text references Figures 2c and 2d, but these should be Figures 2b and 2d. Please adjust for accuracy.

Authors: Thank you for your invaluable comment. The references should be Figures 2c and 2d. However, to improve the clarity of the relevant section, we have now revised the text in the manuscript (Section 3.1) as follows:

"GPCC (Fig. 2d, JPL mascon&GPCC) affected correlations to a larger extent than GPCP (Fig. 2c, JPL mascon&GPCP), in particular over places with less dense in situ networks. Given the standard deviation values of correlation differences (Figs. 2c and 2d) due to switching from GPCP to GPCC, the variability was higher in Fig. 2d (global average: 0.21) than in Fig. 2c (global average: 0.14)."