Referee #1

1. Overview

I appreciate the efforts made to better outline the objectives in the abstract, introduction, and conclusion, as well as the proposed discussion section. The objectives are more clearly defined, which makes it easier to follow the structure and flow of the article. However, the manuscript still lacks a robust discussion and does not sufficiently highlight the key insights of the study.

Authors: We thank the reviewer for the constructive comments. In this revised version, we have substantially strengthened the discussion and better highlighted the key findings of the study. Additionally, we have revised the text in accordance with the other comments.

2. Major Comments

Abstract:

The abstract would benefit from more information on the implications of the analysis, particularly which product performs better and in which contexts. Highlight the key insights more explicitly to provide a clearer summary of the manuscript's findings.

Authors: We thank the reviewer for this valuable feedback. In this revised version, we have expanded the abstract to explicitly highlight the key findings of the study. We have also included a clearer discussion of the implications of our analysis, specifying which product performs better and under which contexts. We have revised the abstract 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 corrected 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 the hydrological drought recovery time (DRT) from precipitation and the terrestrial water storage anomaly (TWSA) acquired from satellite gravimetry. This study also evaluates the drought monitoring performance of G3P and JPL mascon Total Water Storage (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 estimates by calculating the time difference in DRT values derived from the two methods. Globally and across all climate zones, GPCC and GPCP showed comparable performance in hydrological applications with no significant differences in the mean DRT estimates. For the TWS products, DRT estimates using JPL Mascon were, on average, 2.6 months longer than those using G3P. However, the G3P showed approximately 5.0% higher consistency than the JPL mascon globally and across each climate zone, suggesting its better suitability for more precise drought related analyses. These findings indicate that G3P outperforms JPL Mascon in aligning with precipitation products and offers better consistency in DRT estimation. These results provide valuable insight into accuracy of precipitation and TWSA products by utilizing hydrological drought characteristics, enhancing our understanding of meteorological and hydrological processes."

Discussion:

The section titled 3.4 "Discussion" has been included within section 3 "Results." A discussion cannot be a subsection of the results. I recommend renaming section 3 to "Results and Discussion."

The current 3.4 "Discussion" does not provide a proper discussion but rather a detailed description of the results. While this description is interesting and useful, it lacks a synthesis and a discussion of the main findings.

The discussion should address key questions such as for example:

- What do the analyses reveal overall?
- Which precipitation product performs better, and why?
- Which TWSA product performs better, and why?
    - For example, what are the implications of JPL Mascon showing, on average, a DRT that is 2.6 months longer than that derived using G3P?

Additionally, the discussion should address potential limitations of the study to provide a more balanced evaluation.

Authors: Following the recommendation of the reviewer, now we have renamed Section 3 to 'Results and Discussion' to better reflect the content. Additionally, we have substantially revised the discussion to provide a more comprehensive synthesis of the main findings. Specifically, we now explicitly address key questions regarding the overall implications of the analyses, the relative performance of the precipitation and TWSA products, and the reasons behind their differences. Furthermore, we have incorporated a discussion on the potential limitations of the study to ensure a more balanced evaluation. We have revised 3.4 "Discussion" as follows:

"Both precipitation products provided similar global mean DRT estimates (~12 months), with a high consistency rate of 87.5%, suggesting that GPCC and GPCP are both reliable for global hydrological applications. 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 of the two precipitation products differed by less than 1% across all the climate zones. These results for the precipitation products highlight the robustness of these products across the diverse 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%), suggesting that it is better suited for analysing hydrological drought characteristics, particularly in regions with extreme climate conditions, such as the polar (E) and arid (B) zones, where G3P outperformed JPL mascon by 7.0% in consistency. The significant disparity in mean DRT estimation in the polar zone (13.2 months for G3P versus 18.9 months for JPL mascon) highlights the challenges of accurately representing water storage dynamics in high-latitude regions, possibly due to differences in how the two products handle ice and snow storage variability. Conversely, the smallest differences in DRT estimates and consistency in the equatorial (A) zone suggest that both TWS *products perform effectively in regions with stable precipitation patterns. These findings reveal that while both precipitation products perform similarly, G3P outperforms JPL mascon in consistency and alignment with TWS and precipitation based DRT. This suggests that G3P may provide a more accurate representation of terrestrial water storage dynamics in diverse climate zones.*

Potential limitations of this study include the reliance on specific TWS and precipitation products, which may not fully capture uncertainties associated with data retrieval and processing. The study also does not account for external factors such as anthropogenic water use or land cover/use changes, which can influence DRT estimates. Future studies could benefit from integrating additional datasets and models to provide a more comprehensive understanding of these dynamics."

Conclusion:

The manuscript appears to provide an innovative and interesting method for evaluating precipitation products but neglects the main conclusions and their implications regarding the evaluation of GPCC, GPCP, G3P, and JPL Mascon. The conclusion should explicitly state which products show the best performance and in which contexts, along with the broader implications of these findings.

Authors: In this revised version, we have strengthened the conclusion to explicitly state which products demonstrate the best performance and under which contexts. Additionally, we have highlighted the broader implications of these findings to enhance the overall clarity and impact of the study. We have revised the conclusion section as follows:

"TWS changes from one time epoch to another as observed by satellite gravimetry are closely related to the precipitation amount occurring during that time interval. The novel observing concept realized by the GRACE and GRACE-FO missions thus provides a unique opportunity to evaluate the frequently used global precipitation products on monthly or longer timescales. GRACE/GRACE-FO directly provides water storage anomalies, offering a novel approach to characterize drought by assessing the storage deficit. 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.

Our assessments reveal that both GPCC and GPCP products exhibited not only similar DRT estimates but also comparable consistency globally and across all the Köppen-Geiger climate zones. However, as noted, GPCP's reliance on satellite data enhances its utility in data-sparse regions, making it a more versatile choice in such contexts. However, this advantage does not extend to regions with dense in situ networks, where the inclusion of satellite data does not improve its performance.

For TWS products, the mean DRT estimates 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 estimates (5.0% difference, globally) than did JPL mascon. Furthermore, G3P demonstrated greater consistency than JPL mascon across all the Köppen-Geiger climate zones. These findings highlight G3P's reliability for applications requiring precise water storage anomaly data, such as drought monitoring.

The results of our study underline the potential value of GRACE/GRACE-FO for hydrometeorological research due to the strong relationship between precipitation and TWS changes. Its global coverage (albeit its rather low spatial resolution) allows the testing of different precipitation products not only those derived from combinations of satellite and in situ observations (as done in this study) but also from numerical weather prediction models and global atmospheric reanalyses. Both NASA and the European Space Agency (ESA) are currently working on future satellite gravity missions with even more precise sensors and different orbital configurations to further enhance the quality of satellite gravimetry products for hydrological applications."

3. Minor comments

Line 24: The term "performance" is too vague. Please specify what aspect of performance is being referred to.

Author: To clarify the "performance" term, we have now revised the following text in the manuscript (Abstract):

"This study also evaluates the drought monitoring performance of G3P and JPL mascon Total Water Storage (TWS) monthly-solutions from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) satellite missions."

Line 27: The phrase "difference between the two methods" is too vague. Provide a concise explanation of what the difference entails.

Authors: To clarify the "difference between the two methods", we have now revised 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 in DRT values derived from the two methods."

Figure 5: Avoid showing panels (b) and (d) if there are no precipitation data used in the DRT estimation and they have no impact on the results.

Authors: To clarify the Figure 5, we have now revised the following text in the Section "*3.2 DRT Estimates*":

"The precipitation data are not utilized in calculating DRT estimates based on the storage deficit method (Fig. 5). However, they are used in the masking procedure for regions with weak or no linear relationship between cdPA and dTWSA. Although the unmasked regions have identical DRT values, the differences between Fig. 5a and 5b, as well as Fig. 5c and 5d, arise from whether a region is masked or unmasked. Consequently, these figures are not identical."

Figure 8: If the DRT values presented in this figure are derived as an average between DRT calculated from storage deficit and DRT calculated from required precipitation, this should be clarified explicitly in the caption.

Authors: To clarify the explanation of the figure, we have now revised the following text in the manuscript (Figure 8):

“Representation of the consistency in DRT estimates (the class of the time difference in DRT values between two methods, see Table 1), obtained from the different dTWSA (i.e., G3P, JPL mascon) and cdPA (i.e., GPCP, GPCC) datasets. (a) consistency using dTWSA from G3P and cdPA from GPCP (G3P–GPCP), and differences in consistency class relative to G3P–GPCP for (b) G3P–GPCC, (c) JPL mascon–GPCP, and (d) JPL mascon–GPCC.”