Responses to Comments from Reviewer #2

We appreciate the critical comments from Reviewer #2. In the revised manuscript, we have incorporated changes (highlighted in blue) to address all comments presented below in *italics*. Our responses to each comment are provided in **bold.**

1. Thank you for the response document. You engage with the review points carefully, and you fixed some key issues.
Still, I am unsure this is ready for publication. My main reason stays the same: the study area is ~23,000 km², far below GRACE's effective footprint, and that scale mismatch keeps leaking into your claims. You acknowledge this explicitly now, which is good.
But acknowledging it is not enough. You need to show, quantitatively, what GRACE can and cannot detect at this size. What is the minimum detectable seasonal amplitude and trend here? Without that, readers can't judge whether your ΔGWS signal is meaningful or just "difference-of-big-numbers" noise.

R1: We agree that it is important to quantify the minimum detectable seasonal amplitude and trend of GRACE-derived groundwater storage changes (ΔGWS). However, this is not possible at the current stage due to the very limited availability of piezometric observations and poorly constrained storage coefficients, which are critical to validate both the amplitude and trends in GRACE-derived ΔGWS. Nevertheless, we provide estimates of physically plausible values, which remain uncertain but useful as a provisional, preliminary assessment in a data-scarce region.

2. In your revised abstract, you add important caveats: higher-resolution grids do not provide independent signals; residual correlations after detrending/deseasonalizing suggesting leakage; only 21–60% of ΔGWS estimates are "plausible"; well agreement is generally weak; ENSO coupling is weaker for ΔGWS than for ΔTWS/ΔSMS.
This is more honest, yes. But it also makes the key question sharper: if the ΔGWS series depends heavily on filtering and only a fraction of realizations look plausible, why should the reader treat the resulting ΔGWS time series as a robust groundwater result rather than a sensitivity exercise?

R2: We apply a plausibility framework that has recently been applied at the global scale (Arifin et al., 2025) to obtain physically plausible ΔGWS estimates from GRACE. We have added this information in lines 412-415 (see below). As noted by Arifin et al. (2025),

this plausibility framework does not reduce uncertainties associated with GRACE-derived ΔTWS data or global model outputs but rather it filters out implausible ΔGWS estimates and improves the reliability of GRACE-derived ΔGWS time series, which are computed as a residual parameter. Implausible ΔGWS estimates may arise not only from limitations in GRACE signals but also from the inability of global-scale models, such as GLDAS, to adequately represent individual water-balance components required for residual-based GWS estimation. As in R1, robustness ultimately requires validation using piezometric data and storage coefficients derived from more dense monitoring networks.

Lines 412-415: Eight criteria (Table S5), developed by Arifin et al. (2025) for global tropical regions, were applied to identify physically plausible ΔGWS values within ΔTWS bounds. We find that plausible GRACE-derived ΔGWS estimates per realization range from 21 to 60% (mean = 42%). Filtering for plausible realizations improved consistency with 87% of ensemble mean ΔGWS values suitable for further analysis (Fig. 5).

3. On methods, your clarification that 0.25°–0.5° grids are used only to test coherence/leakage (not to claim added spatial detail) is right. Now please complete that logic: if they are not independent signals, why keep so much text/analysis around them in the main paper? Either streamline, or turn it into a formal detectability framework.

R3: Thank you for the comment. We employed GRACE data at different spatial scales to test internal consistency among datasets (lines 270-272, see below). We do not conduct analyses at the sub-pixel or single-grid scale. Instead, we use the ensemble mean for the entire study area, consistent with warnings against single-grid-based analyses (lines 365-367, see below).

Lines 270-272: These classifications enable a diagnostic assessment of GRACE performance across spatial scales, recognizing that the finer (0.25° to 0.5°) products are not independent of the native data. The analyses do not add spatial detail but rather test internal consistency among datasets.

Lines 365-367: Additionally, ΔTWS trend signals of each 0.5° grids within the study area (Fig. 3e) vary considerably across GRACE mascon datasets, reinforcing the recommendation to avoid single-grid-based analyses and instead employ ensemble mean values for more robust interpretations (Vishwakarma, 2020; Dahle et al., 2025).

4. I like that you repeated correlations on deseasonalized/detrended components and found weaker relationships, implying ENSO mainly modulates the seasonal cycle.
That is a strong scientific point. Bring it into the main narrative more prominently. Also, your plausibility filtering remains a big lever.

R4: Thank you for the comment. We have added findings on ENSO correlations with water storage components to the Conclusions (lines 741-743, see below).

Lines 741-743: Statistical analyses show weak-to-moderate coupling between ΔTWS, ΔSMS, and ENSO indices (r = -0.4 to -0.6), with deseasonalized and detrended components indicating that much of the ENSO-related signal is expressed through modulation of the seasonal cycle of water storage changes. ΔGWS, on the other hand, is less well correlated.

5. If "plausible" months are defined by sign logic, say so plainly, and show how conclusions change if you vary the rule. Readers will ask: could those "implausible" months reflect pumping pulses, delayed drainage, or lateral inflows rather than arithmetic artefacts?

R5: Thank you for the comment. As noted in R2, we have clarified this in lines 412-415. The criteria were developed by Arifin et al. (2025) and have recently been applied to evaluate the plausibility of GRACE-derived ΔGWS in 37 large aquifers globally. The complete criteria and corresponding physical meanings are presented in Table S5 in the Supplementary Materials.

6. The piezometer comparison is now more cautious, and you note the need for storage coefficients and the scale mismatch.
Please keep this comparison mostly about timing and event response, not volume. Your high-frequency recharge threshold result is likely your strongest piece of evidence. Do not let it get buried under uncertain GRACE residuals.

R6: Thank you. In addition to the comparison between GRACE-derived ΔGWS and estimated ΔGWS from piezometers, the comparison between GRACE-derived ΔGWS and groundwater levels in piezometers is discussed in lines 516-530. The corresponding plot is shown in Figure S26 in the Supplementary Materials.

Lines 516-530: Correlations between GRACE-based ΔGWS and ΔGWL vary substantially across sites and screened intervals. At MG (30-33 m), ΔGWS from this study correlates weakly with ΔGWL (r = 0.21), whereas ΔGWS from the GLDAS dataset shows a stronger correlation (r = 0.55) and GLWS shows nearly none (r = 0.01). At deeper intervals in MG (87-90 m; 132-135 m), ΔGWS from this study is negatively correlated with ΔGWL (r = -0.39 to -0.67), indicating a decoupling between regional storage dynamics and local aquifer responses. GLDAS-based ΔGWS also shows negative correlations (-0.38 to -0.34), whereas GLWS yields positive correlations (0.25 to 0.66), underscoring dataset-dependent discrepancies. These mismatches indicate that GRACE-derived ΔGWS may reflect regional-scale storage dynamics in the study area that contrast with piezometric observations representing localized responses influenced by aquifer heterogeneity and groundwater abstraction.

In several instances, piezometric observations show ΔGWL trends that differ from ΔGWS signals (Fig. S26). At site MG (screened interval: 87-135 m), ΔGWL exhibits a continuous decline from 2015 to 2023, with rates of -0.9 m/year ($p < 0.05$) for the 87-90 m interval and -0.6 m/year (p-value <0.05) for the 132-135 m interval. Similarly, at site MM (screened interval: 21-24 m), ΔGWL dropped significantly from 5.1 m to 0.1 m by the end of 2019, followed by a relatively stable trend through 2023. These discrepancies may be attributable to localized groundwater abstraction, which can play a dominant role in influencing local groundwater level changes from piezometry and override regional-scale ΔGWS trends observed from GRACE data.

7. My most direct advice would be to decide what paper you are writing. Is it "GRACE-derived ΔGWS for Nusantara planning," or "what GRACE can and cannot tell us in a 23,000 km² coastal basin"? At this scale, the second story is the defensible one.

R7: We agree and have changed the title to "Limits of GRACE-based groundwater storage estimates in a 23,000 km^2 coastal basin: evidence from the Lower Kutai Basin, Indonesia" We have also edited the aim of this paper on lines 87-93.

Lines 87-93: This study examines what groundwater storage changes GRACE can and cannot resolve in a small (23,000 km^2), coastal and data-scarce basin of Indonesia (the Lower Kutai Basin), where spatial scale and proximity to the ocean increase the risk of signal leakage and attenuation. Specifically, we (1) compare ΔTWS across multiple GRACE products, (2) evaluate the potential for ocean leakage into inland GRACE grids, (3) assess the plausibility of GRACE-derived ΔGWS against limited piezometric data, and (4) examine whether large-scale climate drivers, particularly ENSO, are detectable.

The aim is to evaluate the limits of GRACE in this challenging setting and to demonstrate that without a substantially expanded piezometric network, neither GRACE nor in situ observations alone can provide robust groundwater storage estimates.

References

Arifin, A., Shamsudduha, M., Ramdhan, A. M., Rateb, A., Scanlon, B. R., Setiawan, T., Iman, M. I., Husna, A., and Taylor, R. G.: Plausibility Criteria for GRACE-Derived Groundwater Storage Changes From Aquifers Globally, Geophysical Research Letters, 52, e2025GL118580, 10.1029/2025GL118580, 2025.

Dahle, C., Boergens, E., Sasgen, I., Döhne, T., Reißland, S., Dobslaw, H., Klemann, V., Murböck, M., König, R., Dill, R., Sips, M., Sylla, U., Groh, A., Horwath, M., and Flechtner, F.: GravIS: mass anomaly products from satellite gravimetry, Earth Syst. Sci. Data, 17, 611-631, 10.5194/essd-17-611-2025, 2025.

Vishwakarma, B. D.: Monitoring Droughts From GRACE, Frontiers in Environmental Science, Volume 8 - 2020, 10.3389/fenvs.2020.584690, 2020.