

Beyond GRACE: Evaluating the benefits of NGGM and MAGIC for precipitation estimation over Europe.

Author(s): Muhammad Usman Liaqat et al.

MS No.: egusphere-2025-3659

Reviewer 2

General Comments:

This ms addresses an important question, which is whether the next generation of gravity missions (NGGM/MAGIC) will deliver sufficient accuracy and temporal resolution to enable precipitation estimation via the SM2RAIN algorithm. The ms contributes to applying SM2RAIN for the first time, using TWS as input rather than soil moisture. However, several methodological choices require better justification before potential publication in HESS. The use of ERA5L as both the input proxy and the reference benchmark creates insufficient independence between the forcing and evaluation, which is not adequately acknowledged. The synthetic error framework needs stronger physical grounding, and the discussion of regional performance disparities (Alps, Baltic coast, Scandinavia) remains shallow. Furthermore, the ms could be improved by considering a deeper mechanistic analysis of failure modes, clearer separation of the validation and synthetic experiment frameworks, and more rigorous attention to equifinality in parameter calibration.

General Response:

We appreciate the reviewer's helpful comments, which improved overall quality of manuscript. In response, the evaluation framework has been made clearer, and the limitations of using ERA5-Land as both proxy input and reference dataset are now explicitly acknowledged. The synthetic experiment design has been strengthened by explicitly connecting the imposed measurement errors to the ESA NGGM Mission Requirements Document and discussing the assumptions underlying the Gaussian noise model. In addition, the analysis has been extended by stratifying results by Köppen–Geiger climate classes and precipitation intensity to better explain regional performance differences. The structure of the manuscript has also been changed to make it clear how the baseline feasibility experiment is different from the synthetic mission simulations and to make some parts of parameter calibration clearer. These changes make the methods clearer and the results easier to understand thoroughly, and we thank the reviewer again for their constructive feedback. A detailed response to each comment is provided below. For clarity, the reviewers' comments (RC1) are reported in light blue, while the author's responses (ACs) are presented in black. The text that will be added or revised in the manuscript is highlighted in red to clearly indicate the changes made to address the reviewers' comments. All suggested revisions will be incorporated in the revised manuscript submitted after the discussion phase.

Specific Comments:

RC2 Around line 20: The study uses ERA5L soil moisture layers as a TWS proxy and ERA5L precipitation as the benchmark for validation. This impacts the independence of the evaluation framework, as mentioned above, a limitation that is not adequately discussed in the framework, where the algorithm is essentially optimized and tested against the same underlying data model. The authors must explicitly acknowledge this limitation and discuss how it may inflate performance metrics. An independent validation using, e.g., GPCC or E-OBS precipitation data, could be considered, even if only for a subset of stations or a sub-period.

ACs: The objective of this study is not to perform an operational validation of precipitation estimates but rather to conduct synthetic experiments to assess the feasibility of retrieving precipitation from TWS and to evaluate the sensitivity of the SM2RAIN algorithm to the temporal sampling and accuracy expected from future gravity missions such as NGGM and MAGIC. For this purpose, the use of a single physically consistent dataset is common practice in mission simulation studies, as it ensures closure of the water balance and allows the impact of temporal resolution and

measurement uncertainty to be isolated. We agree that independent validation using observational precipitation datasets (e.g., GPCC or E-OBS) would be valuable for assessing the performance of TWS based precipitation retrievals under real world conditions. Such an analysis would require additional calibration strategies and consistent treatment of observational uncertainties and is therefore planned as part of future work using independent observations. The manuscript will be revised to explicitly acknowledge this limitation as well as explained below:

Limitations and Future Work

This study conducts a controlled sensitivity experiment utilizing ERA5L as a self-consistent synthetic reference framework; hence, the results reflect relative performance assessment rather than independent validation against in situ observations. Importantly, these experiments therefore provide a first, controlled synthetic assessment of the feasibility of retrieving precipitation from TWS. The added value of our study is not to validate the SM2RAIN algorithm (which has been done extensively using multiple observational precipitation datasets such as (e.g., GPCC or E-OBS) but rather to assess its sensitivity to the sampling and accuracy expected from future gravity missions. Further, the SM2RAIN framework does not show evapotranspiration as independent fluxes, which could affect how storage dynamics change over weeks. Neglecting evaporation can lead to biased precipitation estimates, particularly in warm, vegetated, and semiarid regions where evapotranspiration is significant. The SM2RAIN parameterization is also expected to be improved by using machine learning approaches. Machine learning can be used to learn how the SM2RAIN parameters depend on TWS dynamics, climate indicators, and land surface characteristics. This preserves the physical SM2RAIN structure while reducing calibration dependency, improving spatial coherence, and enhancing robustness in data scarce and nonstationary environments. Finally, TWS changes generally include contributions from groundwater, but the current framework does not make a clear distinction between groundwater storage and other hydrological components. Future work will address these issues by integrating flux-consistent water balance constraints (incorporating evapotranspiration and runoff), parameter estimation using machine learning an process based ground water depletion attribution with future gravity missions. The future work will also extend to larger domains using more realistic, spatially, and temporally correlated error structures derived from advanced mission simulations.

RC2 Around line 85: The third stated objective refers to "adding Gaussian error on simulated precipitation estimates," but the Gaussian errors are actually added to the TWS input, not to the precipitation estimates. Please, revise this to avoid confusion about the experimental design.

ACs: We thank the reviewer for noting this inconsistency. The Gaussian errors were added to the TWS input signal, not directly to the precipitation estimates. We have revised the objective statement in the manuscript "to analyze the impact of different synthetic mission configurations by adding Gaussian error into simulated precipitation TWS" to avoid this confusion and to clarify that the third objective is to assess the impact of adding Gaussian error to the simulated TWS observations used as input to the SM2RAIN algorithm.

RC2 Around line 115: The TWS estimate (Eq. 1) is computed exclusively from the four soil moisture layers of ERA5L and does not account for snow water equivalent, groundwater, surface water, or canopy interception, all of which are components of the full GRACE-measured TWS signal. The authors should either redefine their variable explicitly as "soil water storage" (SWS) or justify why this simplification is appropriate for the European domain and the 2003–2012 period. This distinction has direct implications for interpreting how faithfully ERA5L-TWS proxies GRACE-based TWS.

ACs: In this study, terrestrial water storage (TWS) was approximated using the four soil moisture layers available in ERA5L. We acknowledge that this representation does not explicitly include other storage components such as snow water equivalent, groundwater, surface water, or canopy interception, which are part of the total storage measured by GRACE and GRACE-FO. However, for the selected European domain in this study, the contribution of some of these

components is relatively limited for the purposes of the present sensitivity experiments. In particular, the contributions of canopy interception and surface water storage are generally small at the spatial scale considered (100 km). Snow water equivalent may locally influence storage variability in northern and mountainous regions, but its contribution is comparatively limited across most of the study's domain during the analyzed period. Furthermore, groundwater storage is not explicitly represented in ERA5L and therefore cannot be included in the proxy dataset used for the synthetic experiments.

Given the objective of this study to assess the sensitivity of precipitation retrieval to temporal sampling and measurement errors expected from future gravity missions, we considered the integrated soil water storage from ERA5L as a suitable proxy for TWS variability. Nevertheless, we acknowledge that this simplification does not represent the full GRACE observed storage signal. The manuscript will be revised to clarify this limitation, and future work will extend the analysis to larger domains and incorporate additional storage components when suitable datasets become available.

RC2 Around lines 135: The SM2RAIN parameters (a, b, Z*) are calibrated point-by-point against ERA5L precipitation. No discussion is provided about parameter sensitivity, equifinality, or the spatial coherence of calibrated parameters across Europe. A supplementary figure showing the spatial distribution of calibrated parameters would greatly strengthen the methodology and allow readers to assess whether unrealistic parameter values emerge in certain regions (e.g., the Alps, Nordic regions).

ACs: We thank the reviewer for this valuable suggestion. In the revised manuscript, we expanded the discussion on the calibration of SM2RAIN parameters. The parameters (Z*, a, b) are estimated independently for each grid cell through point-by-point calibration against ERA5-Land precipitation. To assess the robustness and spatial consistency of the calibrated parameters, we computed summary statistics of their distributions across the European domain. The results show physically reasonable parameter ranges without unrealistic values in regions with complex hydrology such as the Alps or northern Europe. The spatial distributions of the calibrated SM2RAIN parameters are summarized in Table 1. The storage capacity parameter “Z*” ranges from 38.85 to 1008.70, with a median value of 506.80 and an interquartile range (IQR) of 432.67–584.10. These values are substantially larger than those typically reported for surface soil moisture based SM2RAIN applications, reflecting the deeper and more integrated nature of terrestrial water storage signals derived from gravimetry.

Table 1. Summary statistics of calibrated SM2RAIN parameters across all grid cells. Q1 and Q3 denote the first and third quartiles of the dataset, respectively. IQR represents the interquartile range (Q3 – Q1) while Min and Max indicate the minimum and maximum values

Parameter	Min	Q1	Median	Q3	IQR	Max
Z*	38.85	432.67	506.8	584.1	151.43	1008.7
a	0	0.73	1.95	3.29	2.56	42.29
b	1	1	1	6.84	5.84	50
Tpot	0.12	0.14	0.15	0.17	0.03	0.75
Tbase	0.05	0.09	0.12	0.18	0.09	3

The drainage coefficient “a” exhibits a median of 1.95 (IQR: 0.73–3.29), indicating relatively smooth storage discharge dynamics at a basin scale. The exponent parameter “b” shows a median of 1.00 (IQR: 1.00–6.84), with some grid cells reaching the imposed parameter bounds. The prevalence of lower “b” values suggests near linear drainage behavior for many regions, consistent with large-scale aggregated storage processes rather than highly nonlinear near-surface responses. The exponential filter parameters (Tpot and Tbase) display narrow interquartile ranges, indicating

stable noise reduction behavior across grid cells. Overall, the calibrated parameter distributions demonstrate physically consistent behavior and support the applicability of the SM2RAIN framework to high-frequency TWS observations. The obtained parameter ranges are physically plausible and exhibit clear spatial variability across Europe, reflecting differences in climate regime and storage characteristics. Compared to typical SM2RAIN applications based on surface soil moisture, the calibrated values are generally larger, which is expected since terrestrial water storage integrates water variations over the full soil column and shallow subsurface rather than being limited to the near-surface layer. Similarly, the drainage-related parameters “b” tends to reflect smoother and slower storage discharge dynamics, consistent with the deeper and more integrated nature of TWS compared to surface soil moisture. These differences do not indicate a change in the conceptual structure of SM2RAIN but rather reflect the different hydrological control exerted by TWS relative to surface soil moisture. We will add a short discussion of these parameter statistics and their interpretation to the revised manuscript as well as in the supplementary material. This analysis helps to assess potential equifinality and provides an overview of parameter variability across the study domain. We agree that a more detailed sensitivity analysis would be valuable, but this falls outside the scope of the present study and will be considered in future work.

RC2. Around line 155: The synthetic errors (1.9, 4.2, 19, 42 mm) are introduced as additive Gaussian noise drawn from the NGGM mission error table (Daras et al., 2023). However, real mission errors are not necessarily Gaussian in space and time since they may be correlated, anisotropic, and orographically structured. The authors should discuss the limitations of the Gaussian error assumption and whether correlated or spatially structured noise could materially change the conclusions.

ACs: We thank the reviewer for raising this point. In this study, measurement uncertainties were represented by additive Gaussian noise with standard deviations derived from the NGGM mission error specifications reported by Daras et al. (2023). We acknowledge that real satellite gravimetry errors may exhibit spatial and temporal correlations and may also depend on geographical factors such as latitude, orbit geometry, and topography. Therefore, the Gaussian noise assumption used here represents a simplified error model intended to provide a first-order sensitivity assessment of how measurement uncertainty may influence precipitation retrieval. The objective of this work is to investigate the relative impact of temporal sampling and measurement uncertainty on precipitation estimation from TWS rather than to reproduce the full complexity of mission error characteristics. A more realistic representation of correlated or spatially structured noise would require mission specific error covariance models and is therefore beyond the scope of the present study. We will add a discussion of this limitation in the manuscript and clarify that the Gaussian noise assumption should be interpreted as a simplified representation of measurement uncertainty. We will also highlight that future work will extend this analysis to larger domains using more realistic, spatially and temporally correlated error structures derived from advanced mission simulations.

RC2. Around line 120: The bilinear interpolation from ERA5L’s native 0.1° resolution to 100 km is described briefly but without justification. How does this resampling affect the TWS signal and the precipitation estimates, particularly in areas with complex topography such as the Alps or the Scandinavian mountains? The authors should verify that no significant smoothing bias is introduced by this step. Furthermore, a realistic GRACE-TWS signal would require a spherical harmonic representation up to a degree and order corresponding to NGGM.

ACs: We thank the reviewer for this comment. The 100 km grid was adopted as a common analysis scale to apply the SM2RAIN framework consistently across the European domain and to compare different synthetic mission configurations within a unified spatial framework. This resolution does not represent the native resolving capability of GRACE, GRACE-FO, or future NGGM/MAGIC observations. We acknowledge that bilinear interpolation smooths small scale spatial variability and may reduce local gradients in both precipitation and terrestrial water storage, particularly in regions with complex topography such as the Alps. This spatial smoothing may partially contribute to the reduced model performance observed in these regions. The objective of this study is to investigate the relative sensitivity of precipitation retrieval to temporal sampling and measurement error, rather than to reproduce the full

spatial resolution of gravity observations. A fully realistic simulation of gravity mission observations would require spherical harmonic representation and mission-specific filtering consistent with NNGM/MAGIC retrieval chains. Such an end-to-end mission simulation is beyond the scope of this preliminary sensitivity study but represents an important direction for future research.

RC2. Around line 175–185: The ms identifies systematic underperformance in the Alpine region of northern Italy, in Estonia and Latvia, and along the Norwegian coast, but does not provide a mechanistic explanation. Are these failures related to orographic precipitation, snow processes, coastal sea-level effects on TWS, or inherent limitations of SM2RAIN? Please consider exploring this issue, and a dedicated paragraph examining the physical drivers of regional failure would improve the scientific depth of the results section.

ACs: In the revised manuscript, we will extend the discussion to investigate the weaker performance observed in the Alpine region of northern Italy, the Baltic countries, and coastal Norway. It can be explained by several hydrological processes that are not fully captured by the simplified SM2RAIN inversion framework. First, mountainous regions such as the Alps are characterized by strong orographic precipitation gradients and complex topography, which introduce significant sub-grid variability at the 100 km resolution used in this study. This spatial heterogeneity weakens the relationship between precipitation and terrestrial water storage variations. Second, in northern regions including Estonia, Latvia, and coastal Norway, snow accumulation and snowmelt processes play a dominant role in seasonal water storage dynamics. These processes introduce temporal delays between precipitation input and storage changes, violating the assumption of an immediate storage response inherent in the SM2RAIN formulation. To better interpret these spatial patterns, we additionally performed a stratified analysis based on Köppen Geiger climate classes and precipitation intensity (Tables S2–S3 in the Supplementary Material). These analyses confirm that SM2RAIN performs best in rainfall-dominated Mediterranean and oceanic climates and shows reduced performance in colder continental and snow affected regions.

RC2. Around line 200: In Section 3.1, ERA5L precipitation serves as the reference; in Section 3.2, the SM2RAIN-derived output from Section 3.1 is used as the reference for the synthetic experiments. This might imply that any systematic errors in the Section 3.1 output are inherited and compounded in Section 3.2. The authors should discuss how errors in the first step propagate into the second and explicitly state this design choice.

ACs: In Section 3.1, ERA5L precipitation is used as the reference to examine the capability of SM2RAIN to estimate precipitation from TWS. In Section 3.2, the precipitation simulated in Section 3.1 is used as the reference for the synthetic experiments to isolate the effects of temporal sampling and measurement errors introduced in the TWS signal. We acknowledge that this experimental design may propagate systematic errors from the first step into the second. However, the objective of the synthetic experiments is not to provide an independent validation of the precipitation retrieval, but rather to assess the relative sensitivity of the precipitation retrieval to different NNGM/MAGIC sampling configurations. Using the SM2RAIN derived precipitation as a reference ensures internal consistency of the experimental framework and allows the analysis to focus specifically on the impact of TWS sampling and measurement uncertainty. It will be explained in detail at the beginning of Section 3.2.

RC2. Section 3.2 appears twice, first for "Performance of NNGM synthetic TWS for precipitation estimation" (p. 7) and again for "Performance of NNGM\MAGIC mission with respect to GRACE-FO" (p. 10). This should be corrected throughout.

ACs: We thank the reviewer for pointing out these numbering inconsistencies. Corrections have been carried out.

RC2. Around line 275: The authors define $R > 0.7$ as the threshold for "satisfactory" performance, but this threshold is neither justified in the text nor referenced to any established community standard. Given that the application context is precipitation estimation for hydrological use, the justification should be tied to downstream impacts (e.g., flood

forecasting, drought detection), or a sensitivity analysis should show how conclusions change under alternative thresholds such as 0.6 or 0.8.

ACs: The threshold $R > 0.7$ was used as an indicative benchmark of satisfactory performance based on previous SM2RAIN applications and satellite precipitation validation studies (Brocca et al., 2015, 2019). As an example, global precipitation products derived using the SM2RAIN approach typically report correlations in the range 0.6–0.8 (e.g. with respect to reference precipitation datasets), which is deemed acceptable given the indirect nature of the inversion and uncertainties in the satellite observations. Brocca et al. (2019) reported similar performances in the global SM2RAIN precipitation dataset, where correlations above 0.7 were found to be representative of good precipitation retrievals. Considering the inherent uncertainties in precipitation estimation from terrestrial water storage and the large spatial scales analyzed in this study, correlations above 0.7 generally indicate that the main temporal variability of precipitation is well captured. The following discussion will be incorporated into section 3.1.

RC2. Around line 280: The statement that the GRACE-FO error is 25 mm and its temporal resolution is 30 days is used as the baseline for comparison. Please add a citation or derivation that supports these specific values in the main text. The source and spatial scale at which this 25 mm figure applies must be explicitly cited and contextualized, as the GRACE-FO error itself might be resolution-dependent.

ACs: In this study, we used terrestrial water storage (TWS) from ERA5-Land as a proxy dataset. We did not use real GRACE or GRACE-FO observations in our analysis. Rather, we used a proxy data set and degraded its temporal sampling and measurement errors to mimic different mission configurations, in which the GRACE-FO-like configuration serves as a synthetic benchmark that approximates the nominal sampling of current satellite gravimetry missions. The measurement error levels used in the synthetic experiments are derived from the ESA Next-Generation Gravity Mission (NGGM) Mission Requirements Document (MRD), which specifies expected uncertainty levels for different spatial and temporal scales of future gravity observations (ESA, 2023). Based on these specifications, we defined a reference configuration representing present-day gravimetry capability with approximately 30-day temporal sampling and ~25 mm equivalent water height uncertainty, which we refer to as a GRACE-FO-like synthetic benchmark. This configuration serves only as a reference scenario to evaluate the potential improvements achievable with future missions such as NGGM and MAGIC. The manuscript will be revised to clarify that this configuration is synthetic, to explicitly cite the ESA MRD, and to report the temporal cadence and error magnitude together wherever the GRACE-FO-like configuration is discussed.

RC2. Section 4, Discussion and Conclusion: This section reads as a summary of the results without engaging with the literature on gravity-based hydrological applications or competing approaches. The authors should discuss how their findings offer any complementary advantage over SM2RAIN-SM, which is briefly mentioned but deserves more elaboration.

ACs: We thank the reviewer for this suggestion. We will also add this explanation in the revised manuscript discussing the potential complementarity between precipitation retrieval from terrestrial water storage (TWS) and the more widely used SM2RAIN approach based on soil moisture (SM). In particular, the use of TWS in respect to standard SM2RAIN applications based on satellite soil moisture observations (e.g., Brocca et al., 2014, Brocca et al., 2019) provides a more integrated representation of the terrestrial water cycle by combining contributions from soil moisture, groundwater, and surface water storage. While soil moisture measurements typically represent only the near-surface layer, gravity-based TWS observations capture deeper storage components and large-scale water mass redistribution. Moreover, precipitation estimation from TWS may be complementary to soil-moisture-based approaches for large spatial scales and over regions where surface soil moisture observations are sparse and/or strongly impacted by local processes (e.g., evaporation, vegetation dynamics, densely vegetated areas, and frozen soils).

References

Brocca, L., Massari, C., Ciabatta, L., Moramarco, T., Penna, D., Zuecco, G., Pianezzola, L., Borga, M., Matgen, P., and Martínez-Fernández, J.: Rainfall estimation from in situ soil moisture observations at several sites in Europe: an evaluation of the SM2RAIN algorithm, *Journal of Hydrology and Hydromechanics*, 201–209, <https://doi.org/10.1515/johh-2015-0016>, 2015.

Brocca, L., Filippucci, P., Hahn, S., Ciabatta, L., Massari, C., Camici, S., Schüller, L., Bojkov, B. and Wagner, W., 2019. SM2RAIN–ASCAT (2007–2018): Global daily satellite rainfall data from ASCAT soil moisture observations. *Earth System Science Data*, 11(4), pp.1583-1601.

Daras, I., March, G., Pail, R., Hughes, C., Braitenberg, C., Güntner, A., Eicker, A., Wouters, B., Heller-Kaikov, B., and Pivetta, T.: Level-2a simulated gravity field solutions of ESA's science support study to Mass change And Geosciences International Constellation (MAGIC) Phase AV 1.0., <https://doi.org/10.5880/icgem.2023.005>, 2023.

ESA (2023). Next-Generation Gravity Mission (NGGM) Mission Requirements Document (MRD), Version 1.0. European Space Agency, Future Earth Observation Programme.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., and Levizzani, V.: Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, *Journal of Geophysical Research: Atmospheres*, 119, 5128–5141, <https://doi.org/10.1002/2014JD021489>, 2014.