Supplement of Hydrol. Earth Syst. Sci., 30, 4969–4983, 2026
https://doi.org/10.5194/hess-30-4969-2026-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

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

Muhammad Usman Liaqat et al.

Correspondence to: Muhammad Usman Liaqat (muhammadusman.liaqat@cnr.it)

The results in Fig. S1 show spatial distribution of correlation coefficients (R) between ERA5-Land-derived TWS and GRACE TWS over Europe. The left panel shows correlations computed over the extended ERA5-Land period (2002–2022), while the right panel is restricted to the common GRACE observation period (2002–2015). Median correlation values increase from 0.66 to 0.79 when restricting the analysis to the overlapping period, indicating strong consistency between ERA5-Land and GRACE at large spatial scales. Higher correlations are observed across central and southern Europe, confirming that ERA5-Land captures gravity-relevant TWS variability when viewed at appropriate scales.

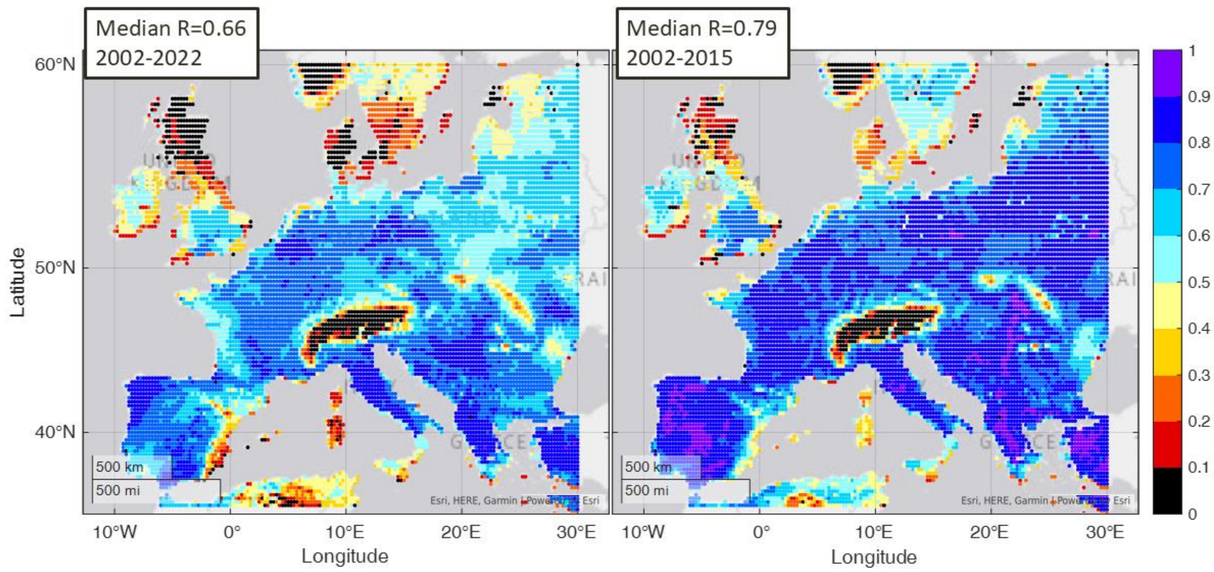

Figure S1. Spatial correlation between ERA5-Land TWS and GRACE TWS. Left panel shows correlations computed over the extended ERA5-Land period (2002–2022), while the right panel is restricted to the common GRACE observation period (2002–2015).

Additionally, Fig. 2 shows the comparison of TWS variability derived from ERA5-Land, ESM, and GRACE over Europe, illustrating the consistency of large scale signals across datasets with different native spatial resolutions. Panels include (i) spatial correlations between ERA5-Land TWS and GRACE TWS over different analysis periods and (ii) correlation maps between model-derived TWS (ESM and ERA5-Land) and GRACE solutions from two independent processing centers (CSR and JPL). Despite differences in native resolution and processing approaches, ERA5-Land and ESM exhibit coherent large-scale TWS patterns that closely match GRACE observations, with spatially consistent correlations approximately (R = 0.8) across large parts of Europe and across both GRACE solutions. Fine-scale variability present in model-based products is largely smoothed in GRACE, confirming that gravity-observable TWS signals are dominated by large-scale spatial variability.

These results demonstrate that resampling a common analysis grid preserves gravity relevant information and does not introduce artificial small-scale signals, supporting the methodological choice adopted in the synthetic experiments. Hence, these analyses demonstrate that the dominant terrestrial water storage (TWS) signals relevant for gravity missions are governed by large scale spatial variability, rather than by small scale features introduced through interpolation. We will clarify this rationale in the methods section of the revised manuscript and will add the corresponding analyses to the supplementary material. The main conclusions of the study are robust with respect to the chosen spatial resolution and should be interpreted as relative sensitivity assessments of future gravity mission configurations, rather than as claims about the native resolving power of NGGM or MAGIC.

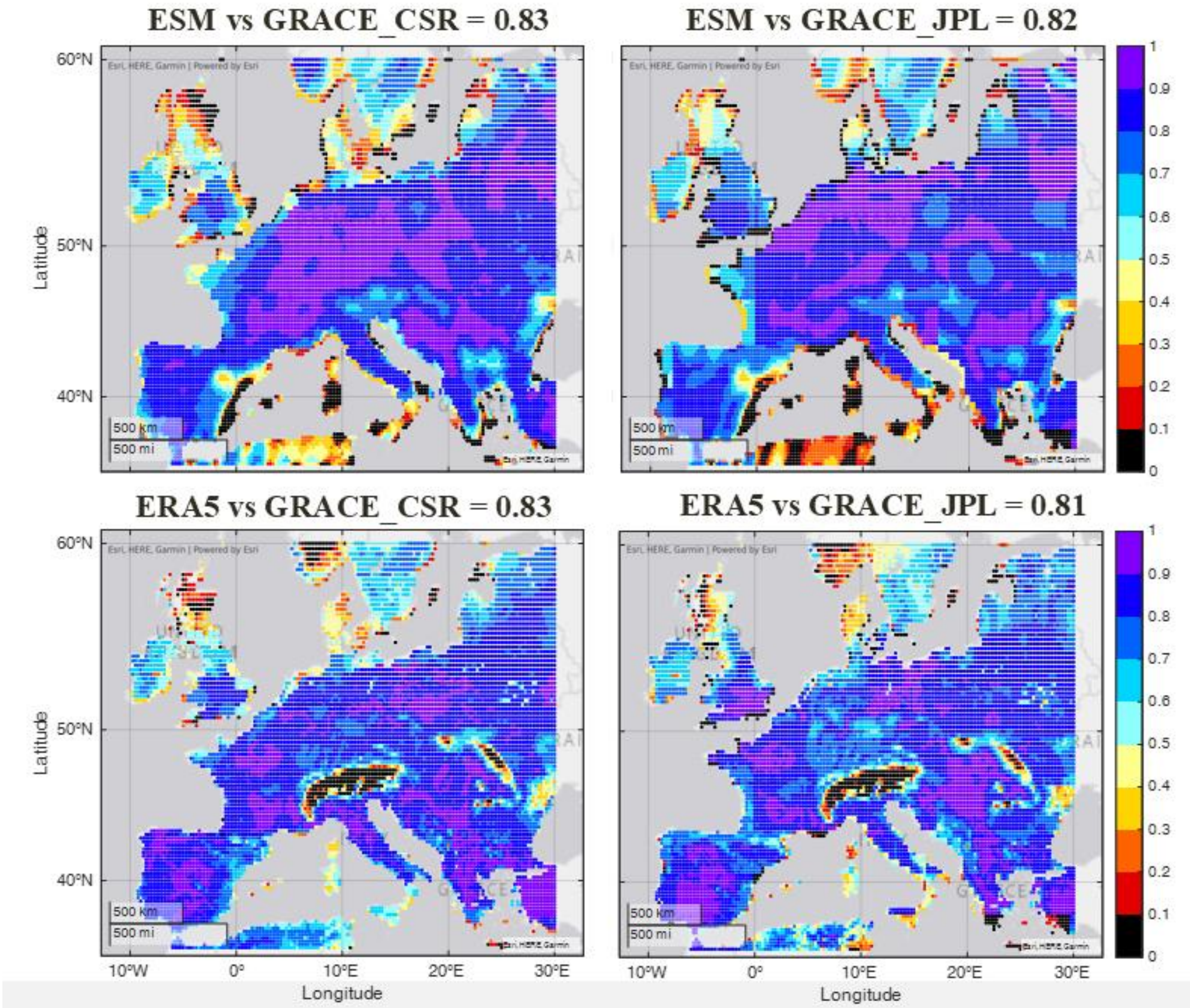

Figure S2. Scale consistency between model-derived and gravimetric terrestrial water storage over Europe.

The spatial distributions of the calibrated SM2RAIN parameters are summarized in Table S1. 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 S1. 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 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 behaviour across grid cells. Although the calibrated parameter distributions are generally physically plausible, some parameters reach the prescribed calibration bounds, which may also indicate equifinality or weak parameter identifiability in certain regions where multiple parameter combinations produce similarly good model performance. Overall, the calibrated parameter distributions support the applicability of the SM2RAIN framework to high-frequency TWS observations while highlighting the inherent uncertainty associated with large-scale parameter estimation.

Table S2. Performance of SM2RAIN derived precipitation estimates stratified by Köppen Geiger climate classes across the study domain. The table reports the number of grid cells (N), the mean correlation coefficient (R) between SM2RAIN-derived and ERA5L reference precipitation, and the corresponding root mean square error (RMSE) for 15 (15) and 30 (30D) day aggregated precipitation. Grid cells were assigned to climate classes using the Köppen–Geiger classification map. This stratification highlights the influence of regional hydroclimatic conditions on the relationship between terrestrial water storage variations and precipitation.

Class	N	R_mean_15D	R_mean_30D	RMSE_mean_15D	RMSE_mean_30D
B	137	0.786	0.792	18.505	29.227
Csa	241	0.972	0.976	6.544	10.162
Csb	172	0.958	0.959	9.039	14.459
Cfb	50	0.950	0.940	9.757	16.246
Dfb	200	0.785	0.773	15.693	24.854
Dfc	144	0.847	0.846	14.856	22.756
ET	142	0.868	0.857	13.626	21.835

To further investigate the spatial variability of the SM2RAIN performance, the results were stratified according to Köppen–Geiger climate classes (Table S2). The analysis shows that the highest skill is obtained in Mediterranean climates (Csa and Csb), with mean correlation coefficients of 0.97 and 0.96, respectively. These regions are dominated by rainfall-driven hydrological processes, which allows terrestrial water storage variations to respond more directly to precipitation inputs. Similarly, high performance is observed in oceanic climates (Cfb), where frequent precipitation events and relatively smooth hydrological responses lead to strong correspondence between precipitation and storage variations. In contrast, lower performance is observed in colder continental climates (Dfb) and arid regions (B), where the mean correlation drops to approximately 0.78. In cold regions, the presence of seasonal snow accumulation and freeze–thaw processes can delay the response of terrestrial water storage to precipitation, weakening the direct relationship assumed by the SM2RAIN inversion. In mountainous areas such as the Alps, strong orographic precipitation gradients and subgrid variability further complicate the link between precipitation and storage changes. Similar limitations are expected in northern regions such as the Baltics and coastal Norway, where snow processes and maritime precipitation variability influence the storage dynamics.

Table. S3 Performance of SM2RAIN-derived precipitation estimates stratified by precipitation intensity using ERA5L reference precipitation. The table reports the number of samples (N), correlation coefficient (R), root mean square error (RMSE), and the mean reference precipitation within each intensity bin. Precipitation events were grouped into four intensity categories (0–2, 2–5, 5–10, and >10 mm day^{-1}). This analysis evaluates the sensitivity of precipitation estimation from terrestrial water storage signals to rainfall magnitude.

Intensity Bin	**R**	**RMSE**	**MeanPobs**	**Interpretation**
0-2 mm/day	0.133	1.281	0.336	Very weak signal
2-5 mm/day	0.311	2.746	3.283	Slight improvement
5-10 mm/day	0.457	3.233	7.069	Moderate skill
>10 mm/day	0.774	6.099	16.871	Strong skill

To further investigate the sensitivity of the method to rainfall magnitude, model performance was stratified by precipitation intensity using ERA5L reference precipitation (Table S3). The results indicate a strong dependence of performance on rainfall intensity. For very light precipitation events (0–2 mm day^{-1}), the correlation between SM2RAIN-derived precipitation and the reference dataset is relatively low ($R = 0.13$), reflecting the limited detectability of small storage variations in terrestrial water storage signals. Performance improves progressively with increasing precipitation intensity, reaching $R = 0.46$ for events between 5 and 10 mm day^{-1} and $R = 0.77$ for heavy precipitation events exceeding 10 mm day^{-1}. This behavior is consistent with the physical nature of gravimetric observations, which capture integrated water storage changes and therefore respond more clearly to large rainfall inputs than to weak precipitation events. The increase in RMSE with precipitation intensity reflects the larger absolute magnitude of precipitation amounts rather than a deterioration of relative performance. The set of 100 grid cells was selected by randomly sampling from all valid land pixels within the study domain that contained complete TWS and precipitation records (i.e., no missing values). To ensure reproducibility of the sampling procedure, the random selection was performed using a fixed random seed as shown in Fig. S4.

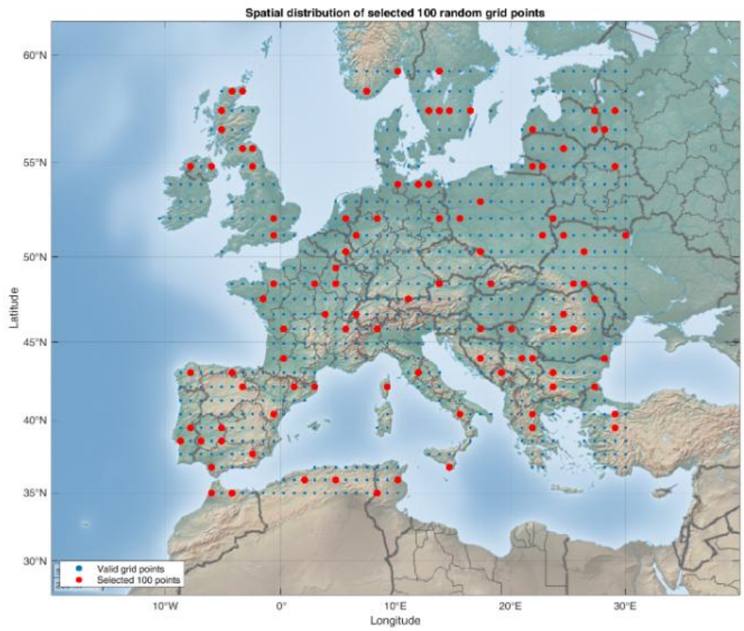

Figure S3**.** Spatial distribution of the 100 randomly selected grid points (red) used in the sensitivity analysis, overlaid on all valid grid cells (blue) across Europe. Sampling was performed using a fixed random seed to ensure reproducibility.

The purpose of this experiment was to provide a representative sensitivity analysis to evaluate the impact of temporal aggregation and measurement error on precipitation estimation using SM2RAIN. Therefore, no explicit climatic or spatial stratification was imposed in this step. The eight countries shown in the regional analysis were selected intentionally to provide broad geographic and hydro climatic coverage across Europe, including southern, central, northern, and eastern regions, and to represent contrasting environments such as Mediterranean, temperate, continental, and colder climates. They were therefore used as illustrative regional examples rather than as a statistically stratified sample of all European conditions.

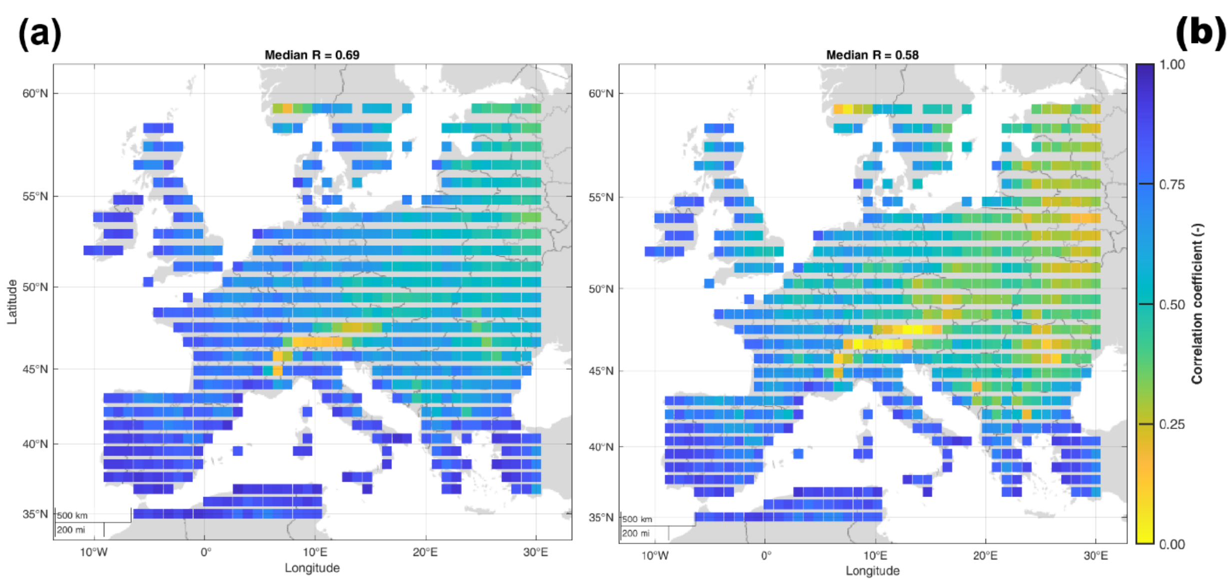

Figure S4. Spatial distribution of grid-cell-wise Pearson correlation coefficients between ERA5L terrestrial water storage (TWS) changes and ERA5L accumulated precipitation for (a) 15-day and (b) 30-day aggregation periods. For each grid cell, temporal correlations were calculated between TWS changes over each aggregation interval and the corresponding accumulated precipitation.