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:

I appreciate the authors' detailed responses and the revisions made to the manuscript. The revised framing as a controlled synthetic sensitivity experiment is clearer and the manuscript has improved substantially. I only have a few remaining points that I think should be clarified before publication.

General Response:

We appreciate the reviewer's helpful comments, which improved overall quality of manuscript. These additional changes make the revised manuscript more clearer and readable. A detailed response to each comment is provided below. For clarity, the reviewers' comments (RC) are reported in light blue, while the author's responses (**ACs)** are presented in black.

Specific Comments:

RC: It is now acknowledged that ERA5-Land is used both to derive the storage input and as a precipitation reference; while this can be acceptable within the revised synthetic sensitivity-study framing, the manuscript would benefit from a simple additional diagnostic showing the direct relationship between ERA5-Land-derived storage and ERA5-Land precipitation. For example, the addition of a spatial correlation analysis between ERA5-Land-derived TWS/SWS changes and ERA5-Land precipitation accumulations at the relevant 15-day and 30-day windows is recommended to help readers better understand the degree of coupling within the ERA5-Land framework.

ACs: We thank the reviewer for this valuable suggestion. Following this recommendation, we performed an additional diagnostic analysis to quantify the direct relationship between ERA5L derived terrestrial water storage (TWS) changes and ERA5L precipitation accumulations. Specifically, for each grid cell, temporal correlation coefficients were calculated between 15-day (30-day) TWS changes and the corresponding 15-day (30-day) accumulated precipitation during 2003-2012. The resulting spatial correlation maps have been added to the supplementary material (Fig. S4) as well as below. The analysis confirms a substantial but spatially variable coupling between precipitation and storage changes within the ERA5L framework, with spatial median correlation coefficients of 0.69 for the 15-day aggregation and 0.58 for the 30-day aggregation. The reduction in correlation at the longer aggregation period reflects the increasing influence of additional hydrological processes, including evapotranspiration, runoff, groundwater exchange, snow processes, and antecedent storage conditions, on storage dynamics. This diagnostic therefore provides useful context for interpreting the synthetic sensitivity experiments.

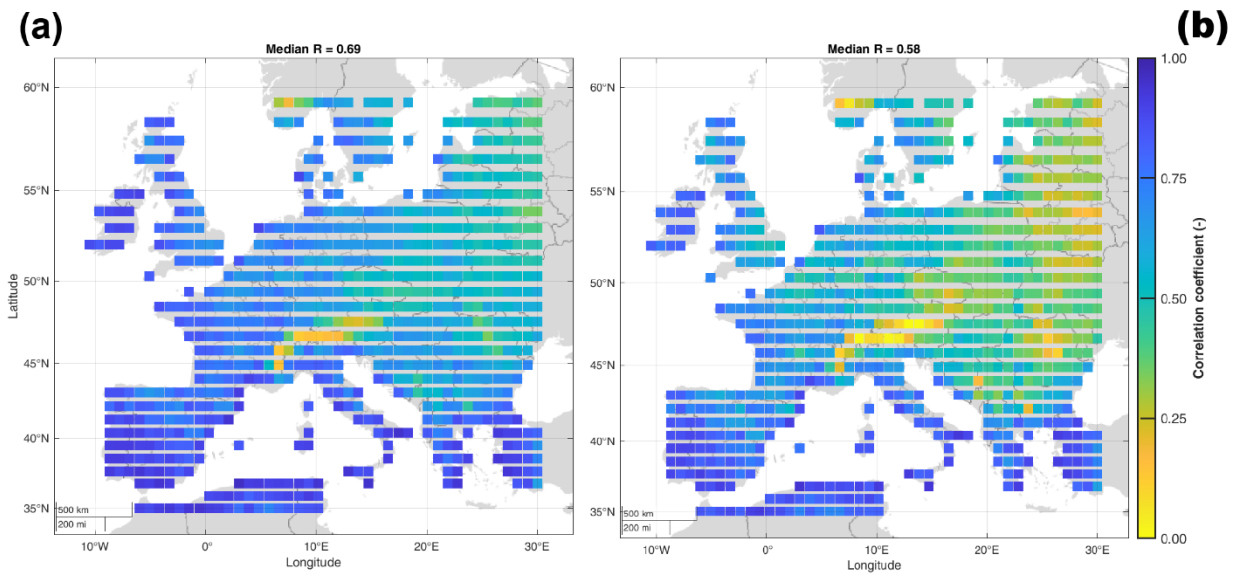

Figure S4. Spatial distribution of grid-cell-wise Pearson correlation coefficients between ERA5L terrestrial water storage (TWS) changes and ERA5L accumulated precipitation for (a) 15-day and (b) 30-day aggregation periods during 2003-2012. For each grid cell, temporal correlations were calculated between TWS changes over each aggregation interval and the corresponding accumulated precipitation.

RC: It is acknowledged that no train/validation split or spatial cross-validation was applied; although this is acceptable for a controlled synthetic sensitivity study, it is suggested to state more explicitly in the limitations that the reported R/RMSE values represent in-sample performance against ERA5-Land and should not be interpreted as independent predictive skill.

ACs: We thank the reviewer for this helpful suggestion. We agree that this limitation should be stated more explicitly. Accordingly, we have revised the Limitations and Future Work section to clarify that the reported performance metrics (e.g., R and RMSE) represent in sample agreement between SM2RAIN derived precipitation and the ERA5L reference dataset within a controlled synthetic framework, as the full available ERA5L time series was used for both calibration and evaluation. We now explicitly state that these metrics should not be interpreted as independent predictive skill or evidence of model generalization. Instead, they quantify the relative sensitivity of the proposed framework to different temporal sampling and measurement-error scenarios under internally consistent conditions. Evaluating the framework against observational datasets using independent temporal and spatial cross validation remains an important direction for future research.

RC The added parameter statistics are useful, but some parameters appear to reach the imposed bounds; therefore, a brief discussion of this as a possible indication of equifinality or weak parameter identifiability in some regions rather than only as physically consistent behaviour is suggested.

ACs: We appreciate the reviewer for this suggestion. We agree that parameters reaching the prescribed calibration bounds may indicate not only physically plausible behavior but also equifinality or weak parameter identifiability in some regions,

where multiple parameter combinations can provide similarly good model performance. Accordingly, we have revised the discussion of the calibrated parameter distributions in the Supplementary Material to explicitly acknowledge this limitation and clarify that the calibrated parameters reflect both physically plausible solutions and the inherent uncertainty associated with large-scale parameter estimation.

RC Since the "TWS" proxy is computed only from the four ERA5-Land soil moisture layers and excludes snow, groundwater, surface water and canopy storage, the presentation of this variable as full TWS without qualification should be avoided; a term such as "ERA5-Land TWS proxy" would more clearly distinguish it from satellite-gravity TWS.

AC: We thank the reviewer for this observation and corrections have been incorporated into relevant positions.

RC The Gaussian-noise experiment should be consistently described as a first-order sensitivity test rather than as a mission-realistic error simulation, since real gravity-mission errors may be spatially and temporally correlated.

AC: We thank the reviewer for this helpful suggestion. We agree that the Gaussian noise experiments should be consistently presented as a first-order sensitivity analysis rather than as a mission realistic error simulation. Accordingly, we have revised the manuscript to consistently describe the introduced Gaussian noise as a simplified first-order sensitivity test designed to evaluate the relative impact of measurement uncertainty on precipitation retrieval.