Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR042324

GRACE TWSA Data Assimilation in Land Surface Models-The Role of State Vector and Model Uncertainty Representation

Yorck Ewerdwalbesloh[1], Carsten Montzka[2], Jürgen Kusche[1], and Anne Springer[1]

[1]Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany, [2]Institute of Bio- and Geosciences: Agrosphere (IBG-3), Research Centre Jülich, Jülich, Germany

Key Points:

- Gravity recovery and climate experiment (GRACE) terrestrial water storage anomaly assimilation is strongly affected by state vector design and model uncertainty representation
- Assimilation increments vary across setups, especially in deep soil layers, showing sensitivity to methodological choices
- Surface soil moisture is well represented in the model; GRACE data assimilation mainly affects water contents in deep layers

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Ewerdwalbesloh,
ewerdwalbesloh@geod.uni-bonn.de

Citation:

Ewerdwalbesloh, Y., Montzka, C., Kusche, J., & Springer, A. (2026). GRACE TWSA data assimilation in land surface models-The role of state vector and model uncertainty representation. *Water Resources Research*, *62*, e2025WR042324. https://doi.org/10.1029/2025WR042324

Received 19 SEP 2025
Accepted 1 MAY 2026

Author Contributions:

Conceptualization: Yorck Ewerdwalbesloh
Data curation: Yorck Ewerdwalbesloh
Formal analysis: Yorck Ewerdwalbesloh
Funding acquisition: Jürgen Kusche
Investigation: Yorck Ewerdwalbesloh
Methodology: Yorck Ewerdwalbesloh
Project administration: Jürgen Kusche
Software: Yorck Ewerdwalbesloh, Anne Springer
Supervision: Carsten Montzka, Jürgen Kusche, Anne Springer
Validation: Yorck Ewerdwalbesloh
Visualization: Yorck Ewerdwalbesloh

Abstract Data assimilation (DA) of terrestrial water storage anomalies (TWSA) from the GRACE and GRACE Follow–On (GRACE/–FO) satellite missions has become a valuable tool for improving large-scale hydrological and land surface model (LSM) simulations. However, existing GRACE/–FO DA frameworks differ substantially in model structure, assimilation algorithms, GRACE/–FO data products and processing methods, and treatment of uncertainty–making it difficult to isolate the effects of individual design choices on DA increments and simulated water storages. Here, we systematically evaluate, for the first time, how state vector design and the representation of model uncertainty–both of which determine how TWSA observations translate into model updates–affect GRACE/–FO DA in the encore Community Land Model. Specifically, we test three alternative state vector configurations, ranging from explicitly representing individual storage compartments to a single integrated storage variable, and four ensemble perturbation methods, including perturbations to atmospheric forcings and to soil hydraulic parameters. We find that, while terrestrial water storage (TWS) increments are broadly similar across configurations, their vertical distribution differs markedly: in deep soil layers of a single representative grid cell, monthly mean volumetric soil water content differs by up to 20 mm, leading to differences in the partitioning of TWS. We argue that these overlooked structural choices in DA design must be carefully considered in GRACE/–FO DA systems, in particular in multi-layered LSMs as well as in multivariate DA frameworks.

1. Introduction

Land surface models (LSMs) simulate dynamic processes such as soil moisture, evapotranspiration, and energy exchanges. Their ability to capture complex hydrological processes, along with their ability to provide continuous, spatially distributed estimates of water storages and fluxes, makes them critical for understanding storage dynamics and climate system feedbacks. Nonetheless, these models are inherently limited by structural assumptions, simplifications in parameterization, uncertainties in external forcing data, and the exclusion of relevant processes, all of which can introduce systematic biases and reduce predictive accuracy (Fisher & Koven, 2020; Ukkola et al., 2016).

Constraining LSMs toward reality by incorporating independent observations via data assimilation (DA) techniques helps to reduce uncertainty in model states, improve their representation, and enhance forecast skill (De Lannoy et al., 2007; de Rosnay et al., 2013). The assimilation of in situ and satellite derived measurements such as soil moisture or snow is already widely applied in LSM DA (De Lannoy et al., 2019; Girotto et al., 2020; Houser et al., 1998; Naz et al., 2019; Reichle et al., 2008; Rodell et al., 2004; Slater & Clark, 2006; C. Sun et al., 2004; Zhao et al., 2025).

LSM DA frameworks increasingly also incorporate terrestrial water storage anomalies (TWSA) derived from the Gravity Recovery And Climate Experiment (GRACE, Tapley et al., 2019) and its successor GRACE Follow–On (GRACE–FO, hereafter both GRACE/–FO). Since 2002, these missions have provided global measurements of changes in terrestrial water storage (TWS) with a spatial resolution of a few hundred kilometers. TWSA represent vertically integrated water storage changes, including groundwater, soil moisture, surface water, snow, and vegetation. Assimilating GRACE/–FO observations has been shown to substantially improve modeled groundwater storage in a wide range of hydro-climatic settings (Girotto et al., 2019; B. Li et al., 2019; Tangdamrongsub et al., 2020; Zaitchik et al., 2008) and to support applications such as drought monitoring (Gerdener et al., 2023; Houborg et al., 2012).

Writing – original draft:
Yorck Ewerdwalbesloh, Anne Springer
Writing – review & editing:
Yorck Ewerdwalbesloh, Carsten Montzka, Jürgen Kusche, Anne Springer

At the same time, the vertically integrated and spatially coarse nature of GRACE/–FO observations poses fundamental challenges for DA. Because GRACE/–FO provides only a vertically integrated constraint that is affected by spatially correlated errors, its assimilation into vertically resolved LSMs does not constrain the redistribution of water among individual storage components. As a result, GRACE/–FO assimilation can leave specific variables weakly constrained and may even degrade model performance for near-surface quantities such as soil moisture. For example, Tian et al. (2017) reported no improvement in surface soil moisture when assimilating GRACE alone, and showed that this limitation could be overcome by jointly assimilating soil moisture observations. Similarly, Tangdamrongsub et al. (2020) found a negative impact on shallow soil moisture that was mitigated when GRACE/–FO was combined with dedicated soil moisture observations. Evidence from multi-sensor and multivariate DA studies further suggests that incorporating complementary observations can reduce unrealistic redistribution of increments among storage components and better constrain individual variables (Girotto et al., 2019; Schulze et al., 2024; Soltani et al., 2024; Wongchuig et al., 2024).

A key element in GRACE/–FO DA is the representation of model uncertainty. Most GRACE/–FO DA frameworks employ variants of the Ensemble Kalman Filter (EnKF, Evensen, 1994), in which model uncertainty is represented through the spread of an ensemble of model realizations, reflecting variability arising from uncertain inputs, parameters, or initial conditions (Springer et al., 2026). This ensemble spread determines both the magnitude of the analysis increments and the cross-covariances between state variables that govern how observational information from TWSA is propagated within the model. Because GRACE/–FO provides only coarse, vertically integrated constraints, updates to individual storage components rely on these ensemble-estimated relationships, making the partitioning of increments strongly dependent on the ensemble representation and the chosen state vector configuration.

Motivated by the above challenges, it is important to examine how GRACE/–FO DA setup choices affect not only agreement with TWSA but also the vertical distribution and physical plausibility of the resulting increments. This issue is particularly relevant for multi-layer LSMs such as the Community Land Model (CLM), where soil moisture and related storages are represented across multiple vertical layers, yet GRACE/–FO provides only a coarse, vertically integrated constraint. Previous GRACE/–FO DA studies have investigated the sensitivity of assimilation results to a range of methodological setup choices, including the characterization of observation uncertainty and its spatial error correlations (Khaki, Schumacher, et al., 2017; Schumacher, Kusche, & Döll, 2016), comparisons of alternative filtering and assimilation algorithms (Khaki, Hoteit, et al., 2017), and strategies for applying analysis increments (Girotto et al., 2016). In this study, we examine, for the case of a high-resolution implementation of CLM over Europe, the sensitivity of GRACE/–FO assimilation results to (a) the specification of the model state vector and (b) the strategy used to generate the ensemble that represents model uncertainty. Understanding these sensitivities is crucial because they determine how specific setup choices govern the redistribution of TWS increments among individual storage compartments and thus also influence physical interpretability of the updated model states. We address the two aspects by testing different configurations of the model's state vector and ensemble representation using both, the global EnKF and the more robust Local Ensemble Subspace Transform Kalman Filter (LESTKF), which applies localization to limit the influence of observations to nearby state variables and thereby reduce the impact of sampling-induced long-range correlations when translating vertically integrated TWSA into updates of individual storage components.

In previous GRACE/–FO DA studies, ensemble generation has been implemented using a range of strategies that can broadly be grouped into three categories: perturbation of forcing data, model parameters, and state variables (Springer et al., 2026). These strategies can be applied individually or in combination and strongly influence both the magnitude of ensemble spread and the statistical relationships between storage components that govern how observational information is propagated within the model.

Perturbing meteorological forcing data–such as precipitation, temperature, and radiation–is the most commonly applied approach and has, to our knowledge, been used in all existing GRACE/–FO DA frameworks. However, forcing perturbations alone mainly inject variability through near-surface processes and may therefore generate limited spread in deeper storages that dominate TWS. Contrary, perturbing model parameters introduces variability for example in flow dynamics and storage capacities (e.g., Eicker et al., 2014; Springer et al., 2019). In previous GRACE/–FO DA studies employing CLM, parameter uncertainty has often been represented by perturbing soil texture to affect derived hydraulic properties (Springer et al., 2019). For global hydrological models such as WaterGAP Global Hydrology Model (WGHM), studies commonly varied calibrated parameters

(e.g., lake depth or groundwater-related coefficients) to represent uncertainty in physio-geographic characteristics and conceptual parameters (Eicker et al., 2014; Gerdener et al., 2023; Schumacher, Eicker, et al., 2016). However, if parameter estimates are biased or based on limited data, such perturbations may underestimate or misrepresent true uncertainty.

A third approach directly perturbs state variables that contribute to or are related to TWS. For example, studies using the Catchment Land Surface Model (CLSM) have perturbed catchment deficit and root zone moisture (e.g., Forman et al., 2012; Houborg et al., 2012; B. Li et al., 2012; Zaitchik et al., 2008) and, in some cases, additional storages such as snow water equivalent (Girotto et al., 2016). However, sole perturbation of state variables does not explicitly represent the physical processes through which uncertainties in atmospheric forcing or model structure propagate into the system, and may therefore provide a less physically consistent representation of uncertainty than approaches that perturb forcings or parameters, particularly if perturbations are applied independently and do not preserve physically consistent relationships among contributing variables (Springer et al., 2026).

In addition to ensemble generation, the configuration of the model state vector plays a central role in determining how GRACE/–FO information is distributed across individual water storage components. By defining which variables are explicitly represented in the assimilation, the state vector determines which relationships between state variables can be exploited during the update and the degree of vertical disaggregation that is possible. Previous GRACE/–FO DA studies have commonly adopted two main strategies to configure the state vector.

The first strategy introduces individual water storage compartments as separate state variables, enabling vertically disaggregated updates but leading to very large state dimensions and more complex observation operators. Previous studies using CLSM often introduced catchment deficit and root zone excess moisture (e.g., Forman et al., 2012; Zaitchik et al., 2008) or included additional variables related to surface excess, canopy interception and snow water equivalent (e.g., Girotto et al., 2016; B. Li et al., 2019). In CLM, where soil moisture is resolved in multiple layers, GRACE/–FO DA studies have often introduced soil moisture for each layer, together with other storages such as snow (Springer et al., 2019).

The second strategy introduces a single aggregated variable per grid cell–such as TWS or a closely related integrated variable–into the state vector. This approach substantially reduces state dimension and simplifies the observation operator but sacrifices the ability to obtain vertically disaggregated updates. For instance, Chi et al. (2024) introduced TWS into the state vector for CLM version 5 (CLM5), whereby storage compartments and related hydrological variables were updated as a function of the assimilated TWS increment, while Soltani et al. (2024) introduced the grid cell saturation degree into the state vector for CLM version 3.5 (CLM3.5) coupled to the sub-surface model ParFlow. A related approach was explored by Retegui-Schiettekatte et al. (2025), who investigated the assimilation of GRACE/–FO-derived TWSA using an aggregated state representation and its impact on simulated water storages for the World Wide Water Resources Assessment model (W3RA).

Together, these considerations motivate a targeted assessment of how DA configuration choices–particularly ensemble generation strategies and state vector design–control the vertical distribution and physical plausibility of GRACE/–FO-induced updates. We hypothesize that both the perturbation strategy used to generate the ensemble and the configuration of the state vector significantly influence the representation of model uncertainty. Through their impact on the ensemble cross-correlation structure, these choices determine how information from GRACE/–FO observations is transferred to individual storage components.

To test this hypothesis, we evaluate three different state vector configurations and four ensemble generation methods. This study addresses three main questions:

1. To what extent do differences in state vector configurations influence GRACE DA results?
2. To what extent do differences in perturbation methods influence GRACE DA results?
3. How do global and local ensemble filters differ in their response to these methodological variations?

To answer these questions, we use the encore Community Land Model (eCLM), a streamlined fork of CLM5 with a simplified build infrastructure, over Europe with a spatial resolution of 12.5 km. Monthly TWSA observations are assimilated using the Parallel Data Assimilation Framework (PDAF, Nerger & Hiller, 2013). All experiments are conducted for the simulation period from 2003 to 2009. We analyze the assimilation increments to identify the impact on the model's mass balance, that is where water is introduced or removed within the model. The

differences between the individual approaches are evaluated using an open–loop (OL) run for reference. In particular, we investigate how different setups of the state vector and perturbation methods influence the distribution of water within the soil. Finally, we evaluate our simulations against in situ soil moisture observations contributing to the International Soil Moisture Network (ISMN, Dorigo et al., 2011; Dorigo et al., 2013, 2021).

This paper is organized as follows: Section 2 provides an overview of the DA framework, including a short description of model set up, observations, and DA algorithm. Section 3 introduces the design of our experiments including a detailed description about the applied state vector configuration and perturbation methods. In Section 4, the results are presented and discussed, and Section 5 offers conclusions and outlines directions for future research.

2. Set-Up of the Data Assimilation Framework

We implement the assimilation of GRACE-derived TWSA into eCLM within the Terrestrial Systems Modeling Platform version 2 (TSMP2) coupled to the PDAF framework (Kurtz et al., 2016; Nerger & Hiller, 2013; Poll et al., 2025; Strebel et al., 2022). TSMP2 provides a scale-consistent, fully integrated soil-vegetation-atmosphere modeling system with ICON as the atmosphere component, eCLM as the land component, and ParFlow for the subsurface hydrology. Here, we use eCLM as a stand-alone model. PDAF supplies a highly parallelized environment that supports a range of ensemble-based assimilation algorithms.

eCLM is a streamlined fork of CLM5 that preserves the scientific core, physical parameterizations, and prognostic state variables of CLM5, while modifying the build infrastructure and model interfaces to facilitate coupling and ensemble DA within TSMP2 and PDAF. To our knowledge, this study represents the first implementation of GRACE/–FO DA within an eCLM–PDAF framework. We extend the CLM3.5–PDAF setup of Springer et al. (2019) to eCLM by incorporating new storage variables, updated process representations and changes in the model setup.

2.1. The Community Land Model Version 5

CLM5, the land component of the Community Earth System Model and the basis of eCLM, simulates key surface processes such as infiltration, runoff, canopy interception, and snow accumulation (Lawrence et al., 2019). It integrates a suite of modules that represent biogeochemical and biogeophysical processes, as well as dynamic vegetation.

Spatial heterogeneity of the land surface is realized through a nested subgrid hierarchy. Each grid cell is subdivided into multiple land units, columns and patches, which account for different surface types such as vegetation, lakes, or cropland. These subgrid types serve as the basis for process-level computations where simulations of water storage changes are performed exclusively on hydrologically active sub-components. The soil column is divided into up to 20 layers, reaching a maximum depth of 8.03 m. Layer thickness increases with depth, ranging from a few centimeters at the surface to over a meter in the deepest layers. Unlike earlier versions of CLM, a grid cell dependent spatially variable soil depth with an underlying bedrock was introduced in CLM5; simulations of water storage changes are conducted only until the respective bedrock depth.

CLM5 simulates various water storage components, that is canopy water, surface water, snow, soil liquid water as well as soil ice. TWS represents a diagnostic variable, computed by aggregating all individual storage components at the end of each model time step. While TWS is represented at the grid cell level, most water storage components are simulated at the column level, with canopy water being an exception and simulated at the patch level. Snow is represented using up to 12 layers with variable thickness and density; for each layer, thickness, water mass, and ice mass are used in hydrological computations, while snow water itself is treated as a diagnostic variable. Soil water is simulated for each layer until bedrock and changes are predicted based on mass conservation for one-dimensional vertical water flow in soils (Lawrence et al., 2019).

Improvements in hydrological schemes, processes, and subgrid representations have led to better agreement with large-scale observations. CLM5 reproduces seasonal TWSA more accurately than CLM4.5 when evaluated against GRACE/–FO, both over the contiguous United States and across global river basins, and performs on par or better than other LSMs relative to GRACE/–FO data (Y. Cheng et al., 2021; Scanlon et al., 2019).

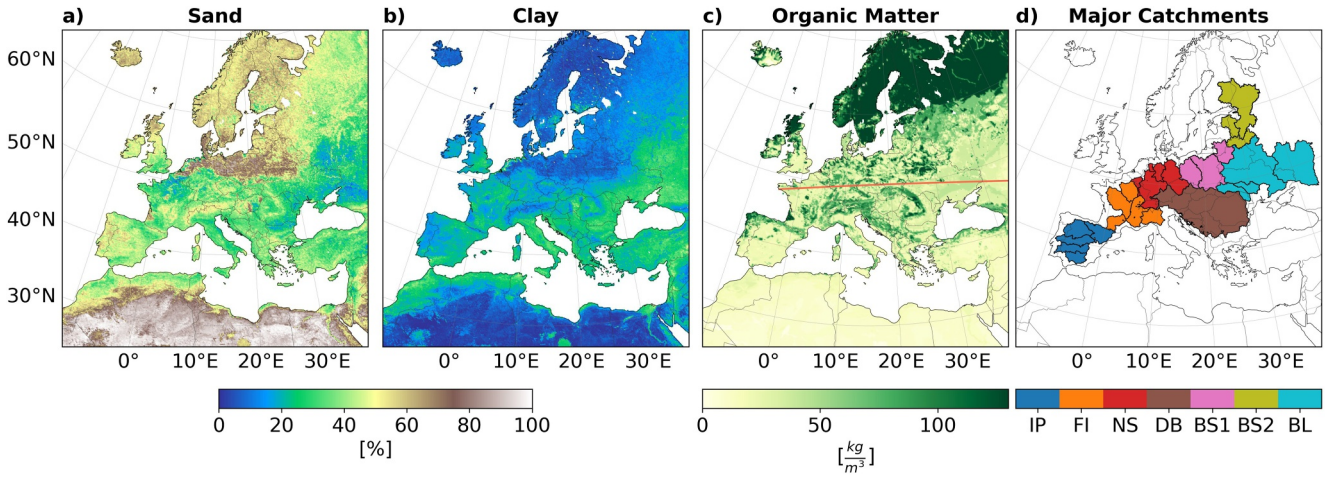


Figure 1. Texture inputs in top layer for (a) percentage sand, (b) percentage clay and (c) organic matter. Red line in (c) displays a transect used for a more in-depth analysis. Major river catchments for basin averaged terrestrial water storage evaluation are displayed in (d) including: Iberian Peninsular IP (Ebro, Duero, Tagus, Guadiana, Guadalquivir), France-Italy FI (Garonne, Loire, Seine, Rhone, Po), North Sea NS (Meuse, Rhine, Ems, Weser, Elbe), Danube DB, Baltic Sea 1 BS1 (Oder, Vistula, Neman), Baltic Sea 2 BS2 (Daugava, Narva, Neva), Black Sea BL (Dniester, Southern Bug, Dnepr, Don).

2.2. Model Setup and Input Data

eCLM is set up over Europe for the Euro–CORDEX domain (Gutowski et al., 2016), using the EUR–11 grid (Jacob et al., 2020) with a spatial resolution of 0.11° (12.5 km) and run at hourly time steps. The simulations incorporate static input variables for land surface and soil properties, alongside time-varying atmospheric forcings. Land cover information is based on the Global Land Cover 2000 data set (Bartholomé & Belward, 2005), which integrates 18 regional land cover products into a harmonized global data set (Neumann et al., 2007).

Soil texture information, including sand and clay fractions as well as organic matter content, is obtained from SoilGrids (Poggio et al., 2021) for each grid cell and input depth layer. Figure 1 gives an overview of the model domain along with the corresponding texture inputs in the top layer. Additionally, Figure 1d shows major European river catchments that are used to evaluate basin averaged TWSA. Sandy soils dominate the regions surrounding the Baltic Sea and Northern Africa, while clay fractions are generally higher in Southern Europe compared to Northern Europe. Organic matter content is significantly higher in Northern Europe and generally only present in upper soil layers.

Atmospheric forcings are obtained from the European Centre for Medium–Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) hourly data on single levels (Hersbach et al., 2020; Copernicus Climate Change Service (C3S), 2023) and include temperature, precipitation, surface pressure, specific humidity, wind speed as well as incoming short- and longwave radiation. ERA5 provides hourly global estimates of atmospheric variables with a horizontal resolution of 31 km. The atmospheric data is regridded to the EUR–11 grid and resampled to 3-hourly intervals.

2.3. GRACE TWSA

GRACE gravity field solutions, provided either as spherical harmonic (SH) coefficients or mass concentration blocks (mascons), primarily reflect gravity changes caused by hydrological mass variations, with mass variations due to oceans and atmosphere already reduced. Here, TWSA over Europe are derived using Level-2 SH coefficients of the ITSG 2018 solution (Kvas et al., 2019; Mayer-Gürr et al., 2018) up to degree/order 96. Effects from glacial isostatic adjustment are reduced using the model from Caron et al. (2018). Due to large errors associated with low-degree coefficients (M. Cheng et al., 2011; Swenson et al., 2008), degree 1 and c20 coefficients are substituted following the recommendations from M. Cheng et al. (2011) and Y. Sun et al. (2016). A temporal mean over 2003–2016 is subtracted from the coefficients, and DDK3 filtering (Kusche, 2007) is applied to reduce striping effects in the TWSA fields. The coefficients are transformed to TWSA maps on a 2° grid based

on Wahr et al. (1998), leading to monthly TWSA fields from January 2003 to December 2009, with the exception of June 2003 due to missing data.

Previous assimilation experiments using CLM3.5 over the same model domain with similar input fields showed that a 2° observation grid yielded the best performance when spatial correlations in the observation errors were considered (Springer, 2019). Although such correlations are not included in this study, the chosen resolution supports a straightforward extension to include them in future work.

Observations over Northern Africa–though part of the model domain (Figure 1)–are excluded. GRACE signals in many parts of this region are dominated by long-term groundwater signals in large fossil aquifer systems (e.g., Mohamed, 2019). These storage variations occur on multi-decadal timescales and are largely disconnected from short-term climate forcing. Moreover, eCLM does not include a deep groundwater storage, with simulations of water storages and fluxes terminating at bedrock depth, and thus cannot simulate such aquifer dynamics.

2.4. Data Assimilation Algorithm

In this study, GRACE-derived monthly gridded TWSA are assimilated into eCLM using ensemble-based sequential DA algorithms that account for uncertainties in both model and observations. Here, we assume a uniform TWSA observation error of 2 cm and neglect spatial error correlations. This value is chosen as a conservative estimate to avoid overconfident assimilation updates, particularly when spatial error correlations are not considered, and is similar to the magnitude of observation uncertainties adopted in previous GRACE/–FO DA studies (e.g., Girotto et al., 2016; Tangdamrongsub et al., 2015; Zaitchik et al., 2008). Model uncertainty is represented by an ensemble of model runs, as described in Section 3.2. Two ensemble-based filtering algorithms are employed: the EnKF (Evensen, 1994) and the LESTKF (Nerger et al., 2012), both of which have been successfully applied in previous GRACE/–FO DA studies (e.g., Eicker et al., 2014; Springer, 2019; Tangdamrongsub et al., 2015). The LESTKF is a deterministic square-root filter that computes the analysis in the ensemble (error) subspace and applies localization in observation space. Localization restricts the update of each model grid cell to GRACE/–FO observations within a specified neighborhood (here, a radius of 7°, following Springer et al. (2019)), thereby damping the impact of long-range correlations that can arise in high-dimensional systems from the finite ensemble.

A key challenge in GRACE/–FO DA is the spatial and temporal resolution mismatch between model and observations. GRACE provides monthly averages of TWSA at approximately 300 km resolution, whereas eCLM simulates up to 43 water storage compartments per grid cell at 12.5 km resolution with hourly time steps. In line with Springer et al. (2019), compartments included in the state vector (Section 3.1) are mapped to GRACE/–FO TWSA observations via an observation operator that performs temporal, vertical, and horizontal aggregation, as described in Section 3.1.1.

3. Experimental Design

DA experiments are performed from 2003 to 2009 with 50 ensemble members to balance uncertainty representation and computational resources. Initial conditions are based on a 50–year spin-up with unperturbed conditions, followed by a 20–year spin-up for each perturbation method (Section 3.2).

3.1. State Vector Configurations

To assess how the choice of state vector structure influences the assimilation, we conduct DA experiments using three distinct state vector configurations (Table 1). For each configuration, assimilation increments are computed for all elements introduced into the respective state vector, regardless of whether these entries represent individual storage components or aggregated variables. For aggregated variables, the resulting increments can subsequently be redistributed among the contributing storage compartments. Although the predicted observations are identical (or nearly identical) for all configurations, differences in the background error covariance structure lead to different increments and, consequently, different assimilation outcomes. In addition, the choice of state vector directly determines the degree of vertical disaggregation that can be achieved when translating TWS increments

Table 1
Overview Over State Vector Configurations

Method	State vector configuration	#per grid cell	#Domain
V1	All compartments are introduced separately: sum of liquid and ice water for each layer, snow, canopy water, surface water	up to 23	1.912.216
V2	TWS is directly introduced	1	103.152
V3	Soil layers are aggregated into three main layers: surface soil moisture (top 10 cm), root zone soil moisture (10–208 cm), deep soil moisture (208 cm–bedrock); snow	up to 4	373.752

Note. Displayed are abbreviations, variables introduced into the state vector for each grid cell, number of variables per grid cell and number of state variables for the whole domain. For V1 and V3, the number of state variables for one grid cell depends on the bedrock depth and is therefore variable.

into updates of individual storage components, thereby affecting how GRACE information is distributed across the soil profile.

In configuration V1 (subdivided configuration), all water storage compartments contributing to TWS are explicitly represented in the state vector as monthly averages. Compared to the setup of Springer et al. (2019) for CLM3.5, we additionally include surface water–representing wetlands and small, sub-grid scale water bodies–which was not included in the earlier model version, and exclude water in the unconfined aquifer as the corresponding parameterization was removed in CLM5 (Niu et al., 2007). The state vector therefore includes soil liquid water and soil ice across all layers down to bedrock, as well as snow water, surface water, and canopy water. Since these compartments are simulated at different sub-grid levels, they are aggregated over hydrologically active grid cells before being added to the state vector. To prevent inconsistencies due to temperature-induced differences among ensemble members–leading to variations in the partitioning of soil water into liquid and ice–soil liquid water and ice are combined into a single variable.

For configuration V2 (TWS configuration), the state vector includes a single TWS variable per grid cell, representing the sum of all storage compartments. This results in a much smaller state vector and simplifies the observation operator, but sacrifices the ability to obtain vertically disaggregated updates. Instead, increments are proportionally distributed among the individual compartments based on their relative contributions to the total storage (Section 3.1.2).

Finally, we introduce a third strategy that reduces the number of state variables while still preserving some vertical structure. In configuration V3 (aggregated configuration), the soil profile down to bedrock is grouped into three vertical components where soil liquid water and ice are aggregated: surface soil moisture (0–10 cm), root zone soil moisture (10 cm–2.08 m), and deep soil moisture (>2.08 m). Snow water is also included in the state vector, while surface water and canopy water are excluded due to their relatively small contribution to TWS.

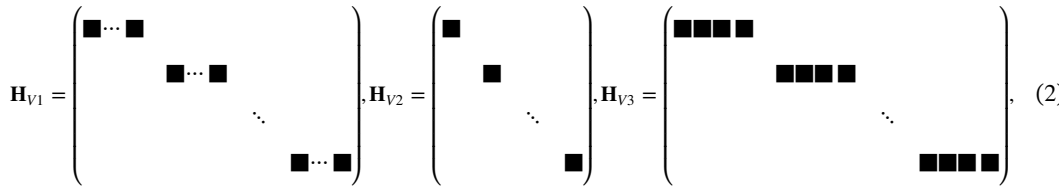
3.1.1. Observation Operator

TWSA observations are predicted through the observation operator according to

$$\mathbf{y} = \mathbf{H}_{avg}\left(\mathbf{H}_{Vi} \cdot \mathbf{x}_{Vi} - \overline{\mathbf{S_{TWS}}}\right), \tag{1}$$

where $\mathbf{y}$ represents the model prediction of the observed variable, and $\mathbf{x}_{Vi}$ denotes the model state vector for configurations $i = 1,2,3$. The matrices $\mathbf{H}_{Vi}$ and $\mathbf{H}_{avg}$ serve as measurement operators: $\mathbf{H}_{Vi}$ aggregates water compartments into TWS, and $\mathbf{H}_{avg}$ performs spatial averaging. $\overline{\mathbf{S_{TWS}}}$ denotes the long-term mean TWS of the model, computed from 2003 to 2016, consistent with GRACE processing, which is subtracted from the simulated TWS to compute TWSA. A separate temporal averaging operator is not required, as the state vector already contains monthly means.

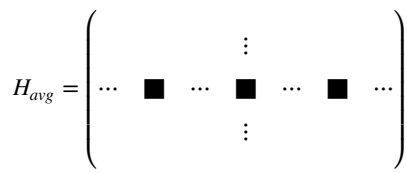
The state vector $\mathbf{x}_{Vi}$ is mapped to TWS by summing over the compartments included in each configuration (V1, V2, and V3), using the corresponding measurement matrix $\mathbf{H}_{Vi}$, with



(2)

where each non-zero element ■ is 1. The structure of $\mathbf{H}_{Vi}$ is block-diagonal, with each block associated with a single grid cell and containing a row of ones corresponding to the compartments present in that cell. The number of compartments per cell varies depending on bedrock depth, with a maximum of 23 in configuration V1 and up to four in V3. In the special case of V2, where TWS is included directly in the state vector, $\mathbf{H}_{V2}$ reduces to the identity matrix.

To account for differences in spatial resolution between model and observations, $\mathbf{H}_{avg}$ is applied to spatially average TWSA. Observations are defined on a 2° grid (Section 2.3); therefore, simulated grid cells are averaged within a 1° radius around each observation to ensure that no grid cell contributes to more than one observation.



(3)

In $\mathbf{H}_{avg}$, each column represents one grid cell and each row corresponds to a single observation. Non-zero elements in row i indicate grid cells contributing to the i-th observation and are set to $1/n_i$, where n_i is the number of grid cells contributing to that observation.

Notably, the simulated TWSA produced by the full observation operator is identical for all three configurations, with the exception of small differences in V3 resulting from the exclusion of surface and canopy water. However, these differences remain well below the observation uncertainty level.

3.1.2. Application of Assimilation Increments

The assimilation algorithm updates the monthly mean values introduced into the state vector. To apply these updates to the model, the difference between the updated and original monthly means is computed. The resulting increments are then applied to the instantaneous model variables at the end of the month. In configurations V2 and V3, where the state vector includes aggregated variables that represent the sum over multiple storage compartments, the corresponding increments are redistributed to the individual compartments according to their relative contribution to the aggregated state variable. In line with Springer (2019), the same is applied to liquid and ice water in V1 where the sum of both is included in the state vector.

For example, the increment applied to snow storage, ΔS_{snow}, in configuration V2–where TWS is directly included in the state vector–is computed as

$$\Delta S_{snow} = \Delta S_{TWS} \cdot \frac{\bar{S}_{snow}}{\bar{S}_{TWS}} \tag{4}$$

where ΔS_{TWS} is the TWS increment, $\bar{S}_{snow}$ is the monthly mean snow storage and $\bar{S}_{TWS}$ is the monthly mean TWS. To maintain physical realism, each compartment-specific increment is limited to no more than half of the corresponding storage. Further constraints are imposed to ensure that the updated state remains consistent with the model's physical boundaries, including non-negativity of snow and ice storage, a minimum threshold for liquid water (0.01 kg/m^2), and upper bounds on soil water content (SWC) given by the saturated water content.

Table 2
Overview Over Ensemble Generation Methods

Method	Atmospheric forcings	Input soil texture	Hydraulic properties
E1	✓	X	X
E2	✓	Sand, clay percentages, organic matter	X
E3	✓	Organic matter	Mineral hydraulic properties
E4	✓	X	Combined hydraulic properties

Note. Displayed are abbreviations and perturbed variables in different perturbation classes: perturbation of atmospheric forcings, input soil texture and computed hydraulic properties. An overview of the computation of mineral, organic and combined hydraulic properties based on soil texture inputs (Figure 1) can be found in Appendix A.

3.2. Perturbation Methods

In this study, we evaluate the impact of different ensemble generation strategies on GRACE DA results by considering perturbations to atmospheric forcings and to model parameters related to hydraulic properties. Table 2 summarizes all four ensemble generation strategies used. All methods incorporate perturbations to atmospheric forcings, while methods E2 to E4 additionally introduce variability in soil hydraulic properties by perturbing either input soil texture (i.e., sand and clay percentage as well as organic matter content), or derived parameter fields (i.e., mineral and/or organic hydraulic properties).

To represent uncertainties in atmospheric forcings, we perturb surface precipitation, 2–m temperature, downward longwave radiation, and downward shortwave radiation. The applied perturbation ranges, spatial/temporal correlation scales, and cross-correlations follow previous studies (Han et al., 2014; Reichle et al., 2007; Springer, 2019) and are summarized in Table 3.

Soil hydraulic properties control soil water retention and movement, directly impacting plant water availability, crop productivity, and environmental quality (Indoria et al., 2020). These properties are largely influenced by soil texture and organic matter content where soil texture refers to the relative proportion of sand, silt, and clay; it determines pore size distribution and thereby influences infiltration capacity, drainage, and the amount of water that can be stored in the soil. Organic matter content affects soil structure by increasing porosity and water-holding capacity, which enhances the ability of soils to retain moisture. In eCLM, soil hydraulic properties are represented using pedotransfer functions by Cosby et al. (1984), which relate mineral soil texture inputs (sand and clay percentages) to parameters such as porosity and hydraulic conductivity. Starting with CLM4, the effect of organic matter on these properties was explicitly included (Lawrence & Slater, 2008) to better capture the high soil water contents observed in organic-rich soils (Strebel et al., 2022). A detailed overview of the derivation of soil hydraulic properties based on soil texture and organic matter is given in Appendix A.

In perturbation method E2 (texture perturbation), hydraulic properties are varied among ensemble members by perturbing the input soil texture information, specifically sand and clay percentages, as well as organic matter content, following the methodologies and perturbation ranges from previous studies (Springer et al., 2019; Strebel et al., 2022). Sand and clay percentages are varied using a uniform distribution with zero mean and a range of −10% and 10%. Organic matter is perturbed by sampling values from a uniform distribution between −10 kg/m^3

Table 3
Overview of the Atmospheric Forcing Variables Perturbed for Ensemble Generation, Including Type of Noise Applied (Additive or Multiplicative), Standard Deviation, Minimum and Maximum Bounds for the Perturbations, the Spatial Correlation Length Scale, and the Cross-Correlation Coefficients Between Variables

Variable	Noise	Standard deviation	Min	Max	Correlation distance	Cross correlation
Precipitation	mult.	0.3	0.3	1.7	80 km	1.0 0.0 −0.8 0.5
Temperature	add.	2	−5	5	250 km	0.0 1.0 0.4 0.4
Short wave radiation	mult.	0.3	0.3	1.7	250 km	−0.8 0.4 1.0 −0.5
Long wave radiation	add.	30	−70	70	250 km	0.5 0.4 −0.5 1.0

Note. For additive noise, temperature perturbations are given in K and longwave radiation perturbations in W/m^2.

to 10 kg/m^3. For each ensemble member, the perturbed texture inputs are then used to recompute the soil hydraulic properties using the same pedotransfer relationships as in the unperturbed model setup (Appendix A). To maintain physical consistency within each ensemble member, a single random perturbation value is sampled per texture variable and ensemble member, which is then applied uniformly across all grid cells and soil layers. Sand and clay percentages are constrained to sum to a maximum of 100%, while organic matter values are limited to a range of 0–130 kg/m^3.

Perturbation method E3 (Cosby perturbation) introduces variability directly at the level of mineral hydraulic properties, rather than through perturbations of the underlying texture inputs (Eloundou et al., 2025). Hydraulic properties–soil matric potential (Ψ), saturated porosity (Θ), the shape parameter (b), and saturated hydraulic conductivity (K_{sat})–are first computed from the unperturbed sand and clay percentages using the Brooks-Corey framework (Brooks & Corey, 1964) and the pedotransfer functions of Cosby et al. (1984) (Appendix A). Variability is then introduced by perturbing these mineral hydraulic properties using parameter-specific standard deviations derived from Cosby et al. (1984). These pedotransfer functions are retained because they are used in eCLM to compute hydraulic properties and provide physically consistent, parameter-specific uncertainty estimates. This distinction is important because perturbing texture inputs (E2) propagates uncertainty through the full nonlinear pedotransfer relationships and preserves internal parameter dependencies, whereas directly perturbing mineral hydraulic properties (E3) leads to a different structure and vertical distribution of ensemble spread that is independent of the underlying pedotransfer relationships.

The variable-dependent standard deviations are applied separately to each mineral property: soil matric potential $\sigma_{\Psi_{\min,i}}$, porosity $\sigma_{\Theta_{\min,i}}$, shape parameter $\sigma_{b_{\min,i}}$ and saturated hydraulic conductivity $\sigma_{K_{\min,i}}$, following Cosby et al. (1984):

$$\log(\sigma_{\Psi_{\min,i}}) = 0.72 - 0.0026 * (100 - (sand + clay)) + 0.0012 * clay \tag{5}$$

$$\sigma_{\Theta_{\min,i}} = (7.73 - 0.073 * clay)/100 \tag{6}$$

$$\sigma_{b_{\min,i}} = 0.0500 * clay + 1.34 \tag{7}$$

$$\log(\sigma_{K_{\min,i}}) = 0.459 + 0.00321 * (100 - (sand + clay)) \tag{8}$$

To avoid unrealistically large variability in modeled TWS, the standard deviations for soil matric potential, porosity, and the shape parameter are scaled by a factor of 0.1. The scaling factor was chosen as the original perturbation bounds produced a spread in TWS that substantially exceeded the chosen observation uncertainty and such that the resulting ensemble spread in TWS is of similar magnitude to that obtained with the texture-based perturbations in E2. The scaled standard deviations are then applied to perturb the corresponding hydraulic parameters for each ensemble member. While each ensemble member is perturbed using one unique set of random numbers per property, Equations 5–8 are evaluated for each grid cell and soil layer, leading to spatially and vertically varying standard deviations that depend on the original input sand and clay percentages. Since Cosby et al. (1984) does not account for organic matter–leading to no perturbations in organic-rich soils–we also perturb the organic matter content in E3 using the same ranges as in E2 to ensure variability in those areas.

In perturbation method E4 (combined perturbation), both mineral and organic properties are varied simultaneously by perturbing their combined hydraulic property values. First, mineral properties are calculated from unperturbed sand and clay percentages using the pedotransfer functions from Cosby et al. (1984). Then, organic properties derived from unperturbed input organic matter content are incorporated following Lawrence and Slater (2008). These represent combined hydraulic properties that would be used in an unperturbed model run (Appendix A). To introduce variability, each combined property is multiplied by a random factor drawn from a uniform distribution, with perturbation ranges of 0.9–1.1 for the shape parameter and saturated hydraulic conductivity, 0.8–1.2 for soil matric potential, and 0.95–1.05 for porosity, yielding four random scaling factors per ensemble member. These perturbation ranges are chosen because they ensure variability in both TWS and hydraulic properties comparable to methods E2 and E3. A key advantage of this approach is that it preserves a consistent vertical structure of variability across all soil layers, regardless of whether mineral or organic components dominate.

It is important to note that perturbing static soil texture and hydraulic parameters introduces model errors that are persistent in time, thereby violating the assumption of temporally white forecast errors underlying the classical Kalman filter derivation (Evensen, 2009). This limitation is well recognized in ensemble DA, particularly in LSM applications where dominant sources of uncertainty–such as soil properties, structural simplifications, and missing processes–are inherently time correlated. In practice, such behavior is commonly introduced in GRACE/–FO DA studies, for instance by perturbing calibrated model parameters and keeping them fixed for the duration of the simulation (e.g., Eicker et al., 2014; Schumacher, Eicker, et al., 2016), or by perturbing soil texture or hydraulic properties that remain constant in time (e.g., Springer et al., 2019), with filter performance depending on maintaining realistic spread and avoiding divergence.

3.3. Experiments

To evaluate the sensitivity of our DA framework to different state vector configurations and perturbation strategies, we design three types of simulations:

1. OL: Ensemble OL simulations, using perturbation method E4; differences in OL ensemble means across perturbation methods remain small.
2. State vector tests: Assimilation runs for each state vector configuration (V1 to V3, Section 3.1), all using perturbation method E4.
3. Ensemble perturbation tests: Assimilation runs for each ensemble generation method (E1 to E4, Section 3.2), all using state vector configuration V3.

Each experiment is performed with both the EnKF, which performs a global update, and the LESTKF, which applies localization. We examine assimilation increments and compare results against the OL run and GRACE observations to quantify the impact of each configuration and perturbation approach. Finally, we validate all simulations against in situ soil moisture observations from measurement networks delivering data to the ISMN that adhere to the ISMN quality standards (Dorigo et al., 2011, 2013, 2021) in terms of Normalized Centered Root Mean Square Error (NCRMSE). Following Taylor (2001), the centered root mean square difference E' is defined as

$$E'^2 = \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y r, \tag{9}$$

where σ_x and σ_y denote the standard deviations of the simulated and observed time series, respectively, and r is their Pearson correlation coefficient. The NCRMSE is obtained by normalizing E' with the standard deviation of the observations,

$$\mathrm{NCRMSE} = \sqrt{1 + \left(\frac{\sigma_x}{\sigma_y}\right)^2 - 2\frac{\sigma_x}{\sigma_y}r}. \tag{10}$$

4. Results and Discussion

4.1. Impact of State Vector Configuration on Assimilation Results

In the following, we assess how the state vector design affects GRACE DA outcomes. We compare three state vector configurations that differ in the degree of vertical aggregation (V1–V3; Section 3.1). For each configuration, we (a) first evaluate the impact of assimilation in terms of Root Mean Square Deviation (RMSD) reduction relative to GRACE TWSA, (b) analyze the resulting increments across storage components and soil layers, and (c) assess how these updates translate into simulated water contents.

Figure 2 shows the RMSD between the OL simulation and GRACE TWSA for different regions in Europe (Figure 1d) as well as the RMSD reduction arising from the assimilation using different state vector configurations and filter algorithms. It is important to note that these RMSD reductions reflect the direct impact of GRACE DA on simulated TWSA and are not meant to serve as independent validation.

DA generally reduces RMSD relative to GRACE compared to the OL simulation across all regions except DB for the EnKF, with a greater reduction for the LESTKF than for the EnKF (Figure 2). The stronger reduction obtained

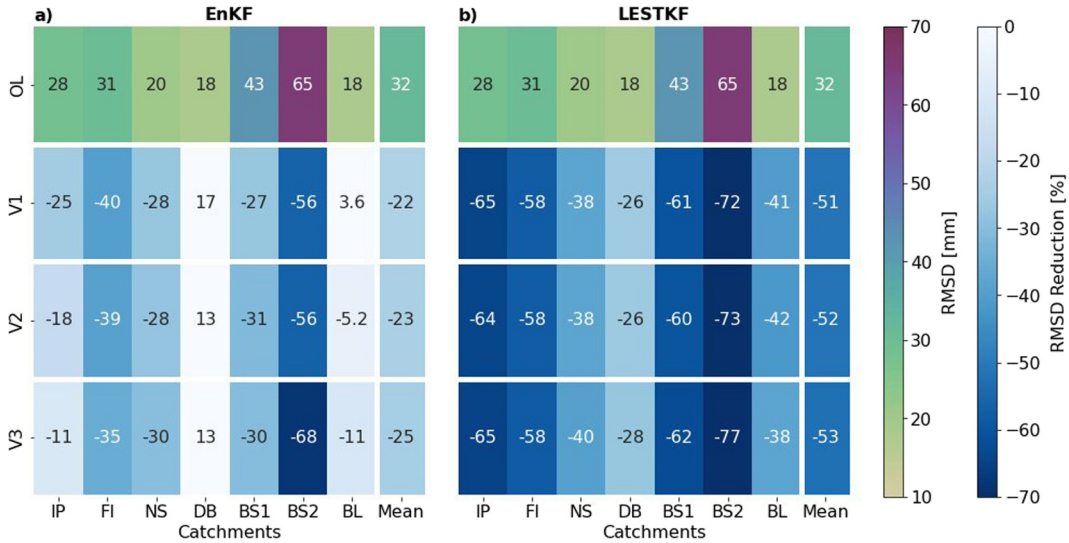


Figure 2. Root mean square deviations (RMSD) between basin averaged time series of terrestrial water storage anomalies of the open–loop simulation and gravity recovery and climate experiment (GRACE) as well as RMSD reductions using GRACE–DA with different state vector configurations V1 (subdivided), V2 (terrestrial water storage), and V3 (aggregated) for major European river catchments as defined in Figure 1d). Results are shown for (a) the Ensemble Kalman Filter and (b) the Local Ensemble Subspace Transform Kalman Filter.

with the LESTKF is consistent with its localized and spatially confined update, which limits the impact of long-range correlations and thus yields more physically plausible increments than the non-localized EnKF. V3 leads to the highest mean RMSD reductions for both filters. Still, RMSD reductions are very similar for the LESTKF, with differences only reaching 5% for region BS2 between V1 and V3. This indicates that the assimilation consistently improves model TWSA across configurations. Additional performance metrics, such as the Nash-Sutcliffe Efficiency (NSE), yield similar results (Figure S1 in Supporting Information S1).

It is particularly insightful to examine in which soil layers water is added to or removed from the model through DA, as this reveals how DA interacts with the model's vertical structure. We therefore first analyze the vertical distribution of assimilation increments before examining how these updates translate to basin-averaged time series and spatial patterns.

Figure 3 illustrates the vertical distribution of assimilation increments across soil layers for August 2003, a month characterized by severe drought conditions across large parts of Europe (Fink et al., 2004), which we use here as a representative example. While increments in other months follow similar patterns, this case highlights the characteristic differences between state vector configurations under extreme hydrological conditions. Although all configurations receive similar TWS updates, especially for the LESTKF (Figure 3h), the distribution of these updates across soil layers varies between configurations. In surface layers, increments are generally similar in sign across all configurations. However, in configuration V1–where each soil layer is explicitly represented in the state vector–some deep soil layers receive increments of opposite sign (Figures 3a and 3b). These are offset by compensating increments in other layers, such that the net update to TWS remains consistent with that of the other configurations (Figure 3h).

To assess how these vertical update characteristics translate into temporal behavior, Figure 4 shows basin-averaged SWC time series for representative layers within the surface (0.04 m), root zone (2.08 m), and deep soil moisture (5.95 m) ranges for the Dnepr basin. Differences between surface SWC (Figures 4a and 4b) are most pronounced during wet periods, when perturbed precipitation inputs exert the strongest influence. In configuration V2, a single TWS increment is computed and then distributed proportionally to the water contents of each contributing storage. Since surface layers have lower water contents then deeper layers–due to their shallower depth–this scaling yields relatively small updates. Consequently, V2 tracks the OL run more closely than the other configurations, with a maximum deviation of 0.35 mm compared to 1.07 and 0.92 mm for V1 and V3, respectively.

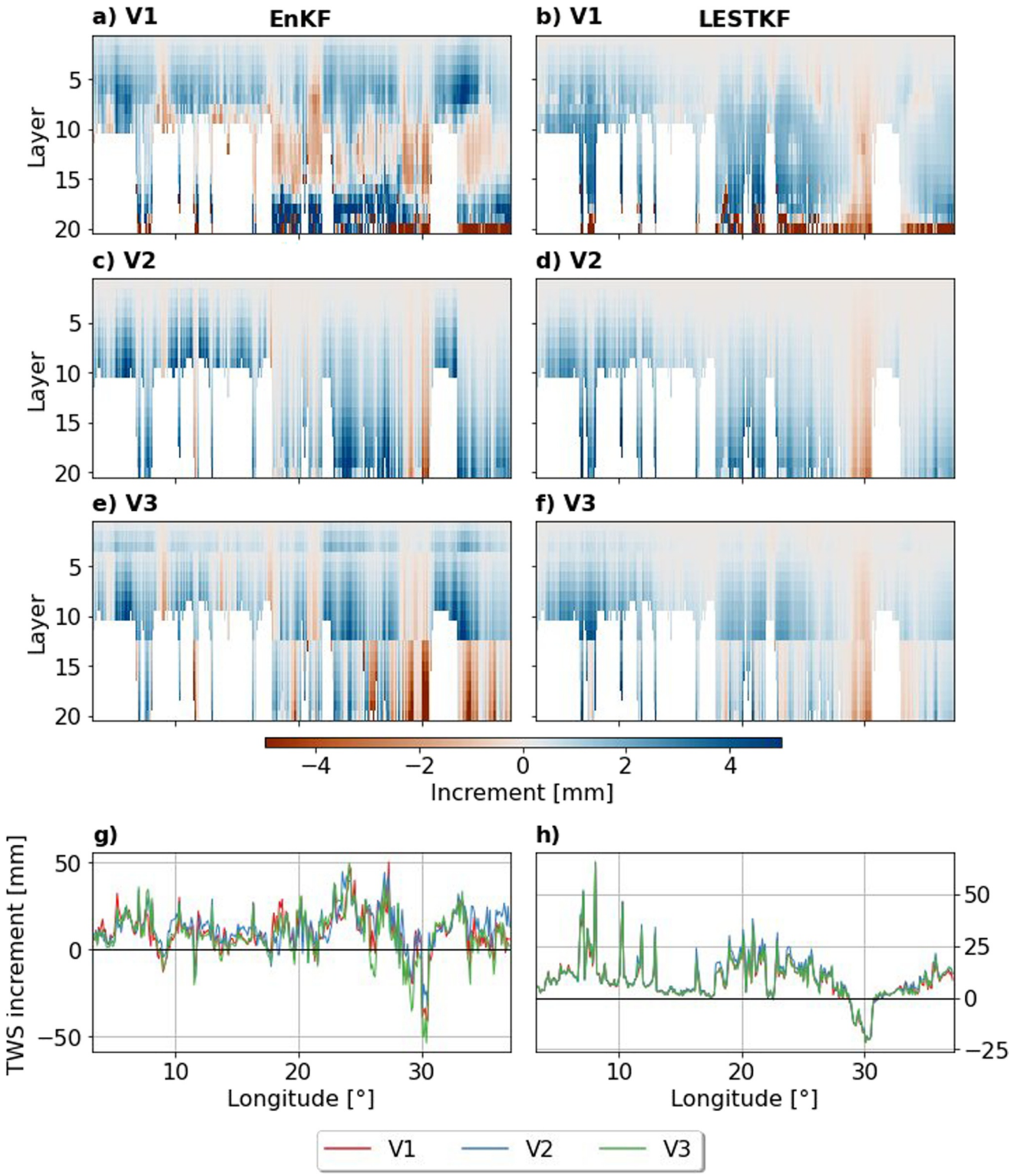


Figure 3. Increments in soil water content for different depth layers (maximum 20 layers, total depth 8.03 m) along a transect (red line in Figure 1c) in 08/2003 for different state configurations. (a and b): V1 (subdivided), (c and d): V2 (terrestrial water storage (TWS)), (e and f): V3 (aggregated). The last row (g and h) displays the increment in TWS for all state vector configuration methods. All simulations are shown for Ensemble Kalman Filter (left column) and Local Ensemble Subspace Transform Kalman Filter (right column).

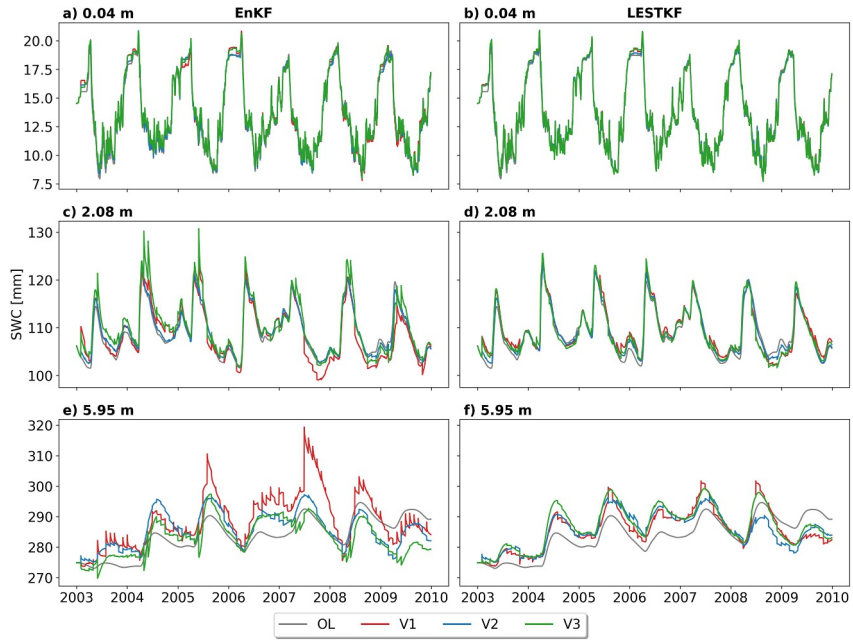


Figure 4. Basin averaged time series of soil water content in different depth layers (0.04 m (first row), 2.08 m (second row), 5.95 m (last row)) of simulations using different state vector configurations. Simulations are conducted using the Ensemble Kalman Filter (left column) and Local Ensemble Subspace Transform Kalman Filter (right column). Displayed are results for the Dnepr basin.

A similar pattern emerges in root zone layers (Figures 4c and 4d). While deviations between V1 and V3 can mostly be mitigated using the LESTKF, differences to V2 remain as they stem from the scaling procedure: here, the maximum difference from the OL is 1.8 mm for V2, compared to 4.1 mm for V1 and 5.0 mm for V3.

Larger differences in SWC between state vector configurations are observed in deep soil layers (Figures 4e and 4f), reflecting the larger fraction contributing to TWSA, that is, the greater water volume which can be stored in these thicker layers, and the behavior observed for increments in deep layers (Figure 3). Configuration V1 exhibits particularly pronounced deviations in deep-layer SWC estimates when using the EnKF, with updates that are not only larger in magnitude but also frequently of opposite sign compared to the other configurations–resulting in maximum differences of up to 28.15 mm in July 2007. This behavior reflects the state-specific background error covariances captured by the EnKF, whereby individual storage components can exhibit differing–and in some cases opposing–relationships with TWSA. Such state-dependent updates are suppressed in configuration V2, where increments are distributed solely based on fractional water volume. Applying the LESTKF substantially reduces these discrepancies, although V1 still exhibits greater variability in its deep-layer SWC even with localization and differences reaching up to 14 mm between V1 and V3.

To place these vertical differences into a spatial context, Figure 5 shows grid cell-level correlations between monthly SWC time series from the assimilation runs and the OL simulation. Examining the geographical distribution of these correlations may help to identify regions where DA induces particularly strong or weak adjustments and to link these patterns to land surface properties. Only results for the LESTKF are shown here; EnKF results are provided in the Supporting Material (Figure S2 in Supporting Information S1). In this context, high correlations indicate small DA increments, while lower correlations reflect larger updates. White areas over land in Figures 5d and 5g–and corresponding areas in panels for the other configurations (Figures 5e, 5f, 5h, and 5i)–represent regions where the depicted soil layers extend below the grid cell-specific bedrock depth.

For surface SWC (Figures 5a–5c), correlations remain largely unchanged across all configurations, indicating that the assimilation has little effect in the uppermost layers. In the root zone, V2 (Figure 5e) shows mostly positive correlation changes, particularly around the Baltic Sea–with values up to 0.2 –, reflecting the small increments characteristic of this configuration in upper layers. V3 (Figure 5f) also shows positive correlation changes in that region, but these are generally lower than for V2.

10.1029/2025WR042324

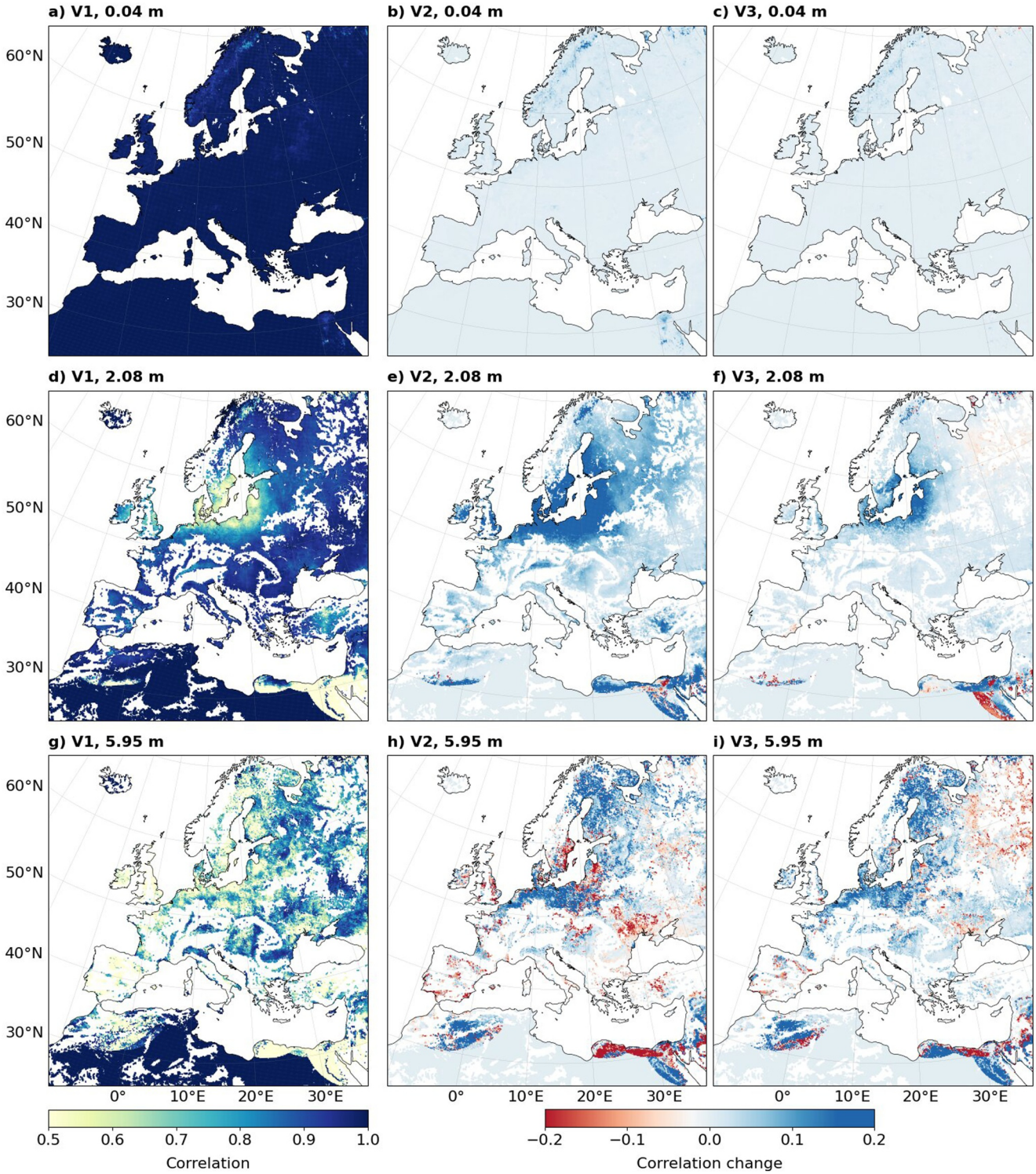


Figure 5. First column: correlations between soil water content time series in different depth layers for surface (0.04 m), root zone (2.08 m) and deep (5.95 m) soil moisture between the open–loop simulation and GRACE–DA analysis using state vector configuration V1 (subdivided) for each grid cell between 2003 and 2009. Second column: Correlation change compared to the first column when using V2 (terrestrial water storage). Last column: Correlation change compared to the first column when using V3 (aggregated). Simulations are conducted using the Local Ensemble Subspace Transform Kalman Filter.

Consistent with the basin-averaged results, correlation changes are substantially larger in deep soil layers than in upper layers (Figures 5g–5i). V2 (Figure 5h) and V3 (Figure 5i) exhibit similar spatial patterns across several regions of Europe, including the Iberian Peninsula, Northern Germany, and Scandinavia. These spatial patterns resemble the distribution of soil texture, particularly regions with a high sand fraction (Figure 1a). Because sandy soils are characterized by high saturated hydraulic conductivity and low porosity (Appendix A), we hypothesize that the resulting fast drainage and weak vertical coupling in these regions cause TWSA observations to impose a similar constraint when the state is represented by aggregated variables (V2, V3), resulting in comparable correlation-change patterns. By contrast, regions such as Eastern and Northeastern Europe often show correlation changes of differing magnitudes and signs, suggesting that other factors such as stronger vertical coupling or finer soil textures modify the response there.

Finally, we attribute the occurrence of opposite-signed increments in deep soil moisture under V1 to specific vertical correlation structures between soil layers. Deep layers are more likely to be saturated than upper layers, meaning that their water content is predominantly governed by porosity rather than hydraulic conductivity. Since these parameters vary between ensemble members (Section 3.2), the resulting inter-layer correlations–particularly between saturated and unsaturated zones–can differ substantially. In some cases, this leads to negative vertical correlations and, consequently, to increments of opposing sign across layers. This effect is most pronounced in configuration V1, where each soil layer is represented individually in the state vector, allowing such inter-layer relationships to directly influence the update. In configuration V2, where only TWS is included in the state vector, this layer-wise effect is absent, while in V3–where layers spanning both saturated and unsaturated zones are aggregated–the effect is reduced but not entirely removed.

In summary, the choice of state vector configuration can significantly influence water content estimates. Configuration V2 offers smooth updates with depth but overlooks the spread structure of individual storage compartments and does not allow for a true vertical disaggregation of GRACE observations. In contrast, V1 captures compartment-specific uncertainties but can produce opposite-signed increments between shallow and deep soil layers, leading to abrupt vertical redistributions that are difficult to reconcile with expected soil-water coupling under a vertically integrated TWS constraint. This behavior likely arises from treating state variables in saturated and unsaturated zones independently, allowing negative vertical correlations to directly translate into compensating updates across layers. Additional sensitivity tests with larger ensemble sizes (100 and 200 members) show that opposite-signed deep-layer increments in V1 remain evident (Figure S3 in Supporting Information S1). Within the tested range, this indicates that increasing ensemble size alone does not remove the behavior; it may therefore reflect not only sampling effects but also the limited vertical information content of a vertically integrated observation when combined with a highly resolved vertical state representation. Configuration V3 addresses both limitations: it provides depth-consistent updates similar to V2, while still allowing for partial downscaling of GRACE TWSA, as reflected by its similarity to V1 in surface and root zone soil moisture, and to V2 in deeper layers (Figures 4 and 5).

The differing behavior of the EnKF and LESTKF reflects how each filter handles the large-scale and vertically integrated nature of GRACE observations. In the global EnKF, increments at a given grid cell can be influenced by a broad set of TWSA observations whose footprints extend far beyond the local region. Because GRACE DA distributes information through ensemble-estimated covariances across many spatial and vertical state variables, these long-range influences can lead to less consistent vertical redistribution, particularly for fully disaggregated state vectors (V1). In contrast, the LESTKF localizes the analysis in observation space, so that each grid cell is updated using only nearby observations. This results in more spatially confined and internally consistent updates and reduces the apparent sensitivity to state vector composition. Moreover, because localization is applied in observation space, the LESTKF also implicitly limits the influence of spatially correlated GRACE observation errors on the update at a given grid cell, thereby acting as an additional regularization mechanism even when a full observation error covariance matrix is not explicitly represented.

Since eCLM subdivides soil moisture into multiple layers, configuration V3 resembles state vector approaches used in LMSs with coarser vertical structure, such as CLSM (Girotto et al., 2016; B. Li et al., 2019; Zaitchik et al., 2008) and WGHM (Eicker et al., 2014; Gerdener et al., 2023), which represent storage compartments without fine vertical discretization. Based on our findings, we suggest that future studies using vertically detailed models consider simplifying the state vector instead of including all compartments explicitly. However, we do not recommend using only TWS–as done in some studies (e.g., Chi et al., 2024)–since this prevents meaningful

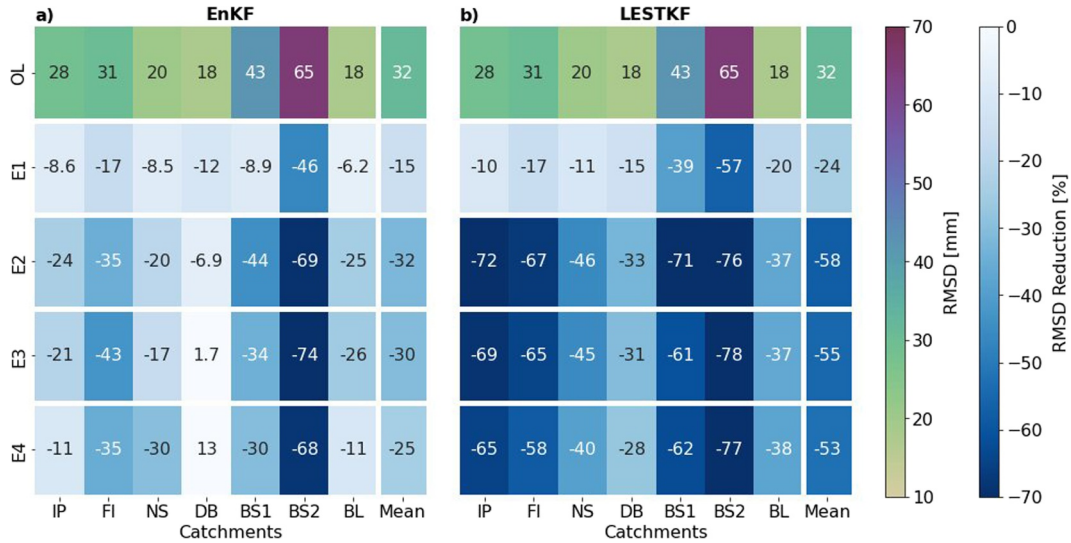


Figure 6. Root mean square deviations (RMSD) between time series of terrestrial water storage anomalies of the open–loop simulation and gravity recovery and climate experiment (GRACE) as well as RMSD reductions using GRACE–DA with different ensemble generation methods E1 (forcing), E2 (texture), E3 (Cosby), and E4 (combined) for major European river catchments as defined in Figure 1d. Results are shown for (a) the Ensemble Kalman Filter and (b) the Local Ensemble Subspace Transform Kalman Filter.

vertical disaggregation. Furthermore, we advocate the use of localized filters, as they help to account for variations in relationships between state variables for different state vector configurations and improve the representation of local hydrological conditions.

4.2. Impact of Ensemble Generation on Assimilation Results

We next assess the sensitivity of GRACE DA results to ensemble generation. We focus on whether different perturbation strategies lead to significant differences in water storage updates across soil layers. Assimilation runs using perturbation methods E1 to E4 (Section 3.2) and a consistent state vector configuration (V3) are compared to the OL for TWSA and SWC. We consider perturbations to atmospheric forcings and to model parameters related to hydraulic properties. Given the complexity of the numerous contributing compartments to TWS, we refrain from perturbing model state entries directly.

All four ensemble generation methods reduce the RMSD compared to GRACE observations over all regions–except DB (Danube) for E3 and E4 for the EnKF (Figure 6). E1 results in the smallest RMSD reduction–averaging 15% for the EnKF and 24% for the LESTKF. As only atmospheric inputs are perturbed, the resulting variability is concentrated in upper soil layers and snow (Zheng et al., 2023). Since these compartments hold significantly less water than deeper layers, the overall spread in TWS is lower compared to the other methods, leading to a lower model uncertainty and consequently to smaller increments, especially in deep soil layers.

E2 achieves the largest RMSD reduction in most regions for both filters, averaging 32% with the EnKF and 58% with the LESTKF. Although E2–E4 were designed to generate a comparable spread in TWS by construction, small differences remain: over BL, for instance, E2 produces a spread of 26 mm, compared to 23 and 20 mm for E3 and E4, respectively. These differences likely explain the slightly larger increments and RMSD reductions achieved by E2. The use of the LESTKF further reduces deviations in regions where E2, E3, and E4 diverge (e.g., BS1, BL), suggesting differing relationships between state variables across grid cells between the perturbation methods. Again, additional performance metrics, such as the NSE, show a similar behavior (Figure S4 in Supporting Information S1).

Again, we examine where water is added or removed through DA by analyzing modeled SWC in different depth layers. Figure 7 illustrates the impact of the different ensemble generation methods on the vertical distribution of assimilation increments. While TWS increments remain largely consistent across all perturbation methods–except for E1 (Figure 7j)–their distribution across soil layers differs substantially. Perturbation method

E3 frequently produces increments of opposite sign between surface, root zone, and deep soil layers when using the LESTKF (Figure 7f). In these cases, positive updates in deep soil moisture are often offset by negative updates in the root zone, such that the net TWS update remains unchanged. A similar, though less pronounced, behavior is observed for E2 (Figure 7d). In contrast, E4 generally yields increments of the same sign across all soil layers within a grid cell (Figure 7h), indicating a more vertically coherent redistribution of water storage.

These vertical increment characteristics are reflected in the basin-averaged SWC time series shown in Figure 8 for the Dnepr basin. Differences in SWC among the ensemble generation methods are larger than those arising from different state vector configurations (Figure 4). While all ensemble methods start from similar initial conditions–indicated by only minor shifts at the beginning of the simulation–DA-driven differences emerge over time.

As expected, E1 yields simulations that remain closest to the OL run. Marked deviations first appear in root zone layers, with perturbation method E3 producing noticeably higher SWC–particularly for the EnKF (Figure 8c), where differences reach up to 37 mm relative to other methods. An opposite pattern emerges in deeper layers, where E3 frequently results in lower SWC values (Figures 8e and 8f). The LESTKF mitigates some of the large EnKF updates differing in magnitudes and signs in the root zone–such as the strong E3 updates during summer months–yet substantial differences remain in deep SWC.

In deep layers, E2 often produces considerably higher SWC than the other methods–averaging 10 mm more than E4–whereas E3 is only slightly higher than E4 (1.6 mm). By contrast, in the root zone both E2 and E3 regularly provide slightly lower SWC than E4, with mean differences of −1.3 mm and −0.8 mm to E4, respectively. Thus, E2 and E3 commonly generate diverging updates between root zone and deep layers under both filters. When using the LESTKF, these divergences are largely confined to the deeper layers, while differences in the root zone are smaller.

We attribute the differences in water content between root zone and deep SWC for E2 and E3 to the independent perturbation of mineral and organic properties (Section 3.2). Treating mineral and organic components separately likely influences the relationship between mineral- and organic-dominated zones. Since organic matter is confined to upper soil layers, this separation can lead to a partial decoupling of root zone and deep soil moisture, which is reflected in the opposing increments observed for these methods.

In summary, the choice of perturbation method substantially affects SWC estimates across the soil profile. Perturbing only atmospheric inputs (E1) results in limited spread in SWC, particularly in deep soil layers where most water can be stored. As a result, the assimilation framework assigns low uncertainty to these states, leading to small assimilation increments. While E2–E4 yield comparable reductions in RMSD relative to GRACE observations, they differ markedly in how updates are distributed across soil layers. E2 and E3 often produce opposite-signed increments in deep layers compared to root zone layers, likely due to the independent perturbation of mineral and organic hydraulic properties. In contrast, E4 produces depth-consistent updates by perturbing mineral and organic properties simultaneously, generating a more coherent vertical spread structure–regardless of whether hydraulic properties are dominated by mineral texture or organic matter.

This analysis highlights that ensemble generation strategies can strongly influence DA outcomes by shaping the relationships between soil layers and thus the covariance structure that governs the vertical redistribution of TWS updates. Although ensemble generation is inherently model-specific and the approaches discussed here apply to eCLM, our results indicate that it is useful to evaluate perturbation strategies not only in terms of the resulting TWS spread, but also in terms of the implied cross-covariances and depth-dependent update behavior. While perturbing only atmospheric inputs appears insufficient to represent uncertainty in deeper storages, it is important to recognize that perturbing other input data, computed or calibrated parameters, or even state variables directly, can influence DA increments by altering the relationships between state variables. Notably, even using different approaches on perturbing the same parameters has a substantial effect.

4.3. Evaluation With In Situ Soil Moisture Station Data

Before evaluating the impact of the DA experiments against in situ observations from measurement networks contributing to the ISMN, we first assess the baseline performance of the OL model in representing soil moisture dynamics. Here, we consider only stations that provide data at least in half of the simulation period (42 months), leading to 26 stations in the top 10 cm, 28 stations between 10 and 50 cm and 29 stations below 50 cm. The

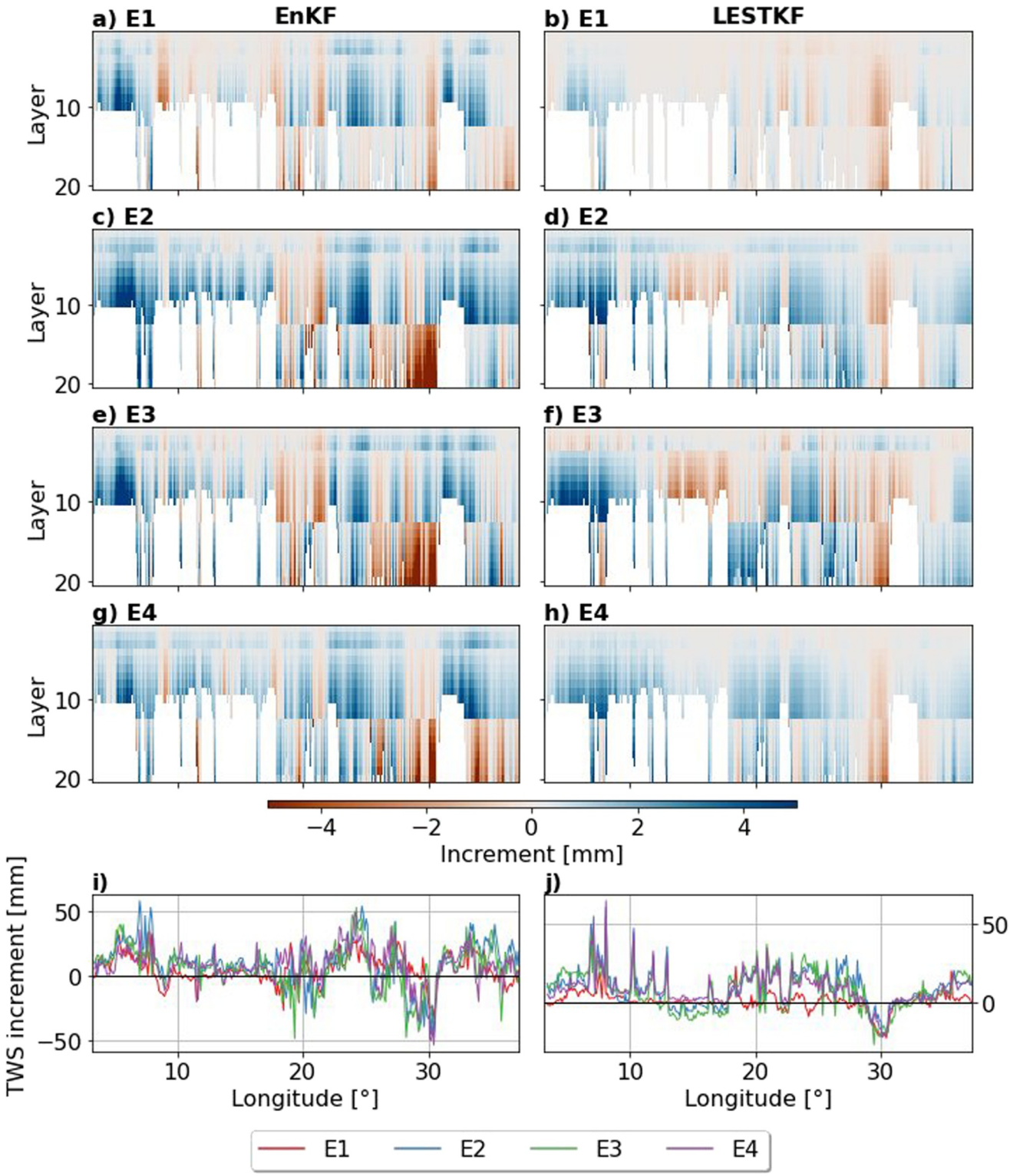


Figure 7. Increments in soil water content for different depth layers (maximum 20 layers, total depth 8.03 m) along a transect (red line in Figure 1c) in 08/2003 for different ensemble generation methods: (a and b): E1 (forcing), (c and d): E2 (texture), (e and f): E3 (Cosby), (g and h): E4 (combined). The last row (i and j) displays the increment in terrestrial water storage for all ensemble generation methods. All simulations are shown for Ensemble Kalman Filter (left column) and Local Ensemble Subspace Transform Kalman Filter (right column).

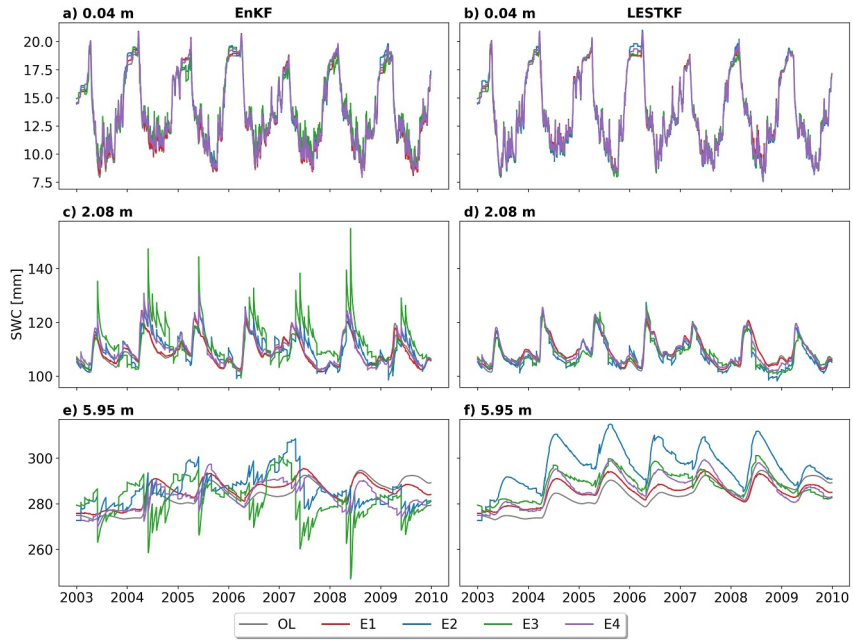


Figure 8. Basin averaged time series of soil water content in different depth layers (0.04 m (first row), 2.08 m (second row), 5.95 m (last row)) of simulations using different ensemble generation methods. Simulations are conducted using the Ensemble Kalman Filter (left column) and Local Ensemble Subspace Transform Kalman Filter (right column). Displayed are results for the Dnepr basin.

stations are not evenly distributed across Europe but are largely concentrated in Spain, France, Germany, Italy, and Finland (Figure S5 in Supporting Information S1). Moreover, many stations are co-located at the same sites but sample different depths, so the effective number of independent locations is smaller. For each in situ soil moisture station, the nearest model grid cell is selected and modeled SWC is vertically interpolated to the measurement depth. Figure 9 shows Taylor diagrams comparing modeled (OL) and observed SWC across different depth ranges. Stations sampling the top 10 cm exhibit the best agreement in phase, with most correlations clustering around 0.8 and only one station below 0.6, and a mean NCRMSE of 1.314. Stations sampling 10–50 cm show greater spread in skill but slightly smaller mean NCRMSE values (1.25) whereas deeper station (>50 cm depth) show a slightly larger mean NCRMSE (1.392).

Using DA with different state vector configurations and perturbation methods does not consistently improve SWC estimates compared to the OL, consistent with the mixed findings of previous studies (e.g., Girotto et al., 2016; Houborg et al., 2012; Tian et al., 2017). We focus here on results from the LESTKF, as it shows substantially higher consistency across the DA setups. Overall, changes in mean NCRMSE are small (Table 4) and do not indicate a clear advantage for any single configuration. The largest NCRMSE reductions are visible for deep stations, with the largest mean improvement observed for E3 at depths >50 cm.

Figure 10 illustrates an example comparison of observed and modeled SWC at the Falkenberg station in Eastern Germany, where measurements are available at multiple depths. Both observed and modeled SWC are scaled by layer thickness to reflect each layer's contribution to TWS. Time series are shown for different state vector configurations (Figures 10a–10c) and perturbation methods (Figures 10d–10f), for both 0.08 and 0.6 m depth; no observations are available at 5.95 m.

Consistent with previous results (Figures 4, 8, and 9, Table 4), the variability of modeled SWC relative to station data is well captured and differences across DA configurations remain small at 0.08 m depth. At 0.6 m depth, some divergences appear: V1 and V3 agree closely, while V2 deviates by up to 5 mm due to weaker updates in root zone soil moisture. Among the perturbation methods, slightly larger differences are observed, with E2 and E3 often yielding higher SWC values than the OL and the other experiments.

The same pattern holds at 5.95 m depth, where the largest discrepancies occur (Figures 10c and 10f). Here, all simulations show exceptionally high SWC at the start of the simulation period, reflecting the unusually wet year

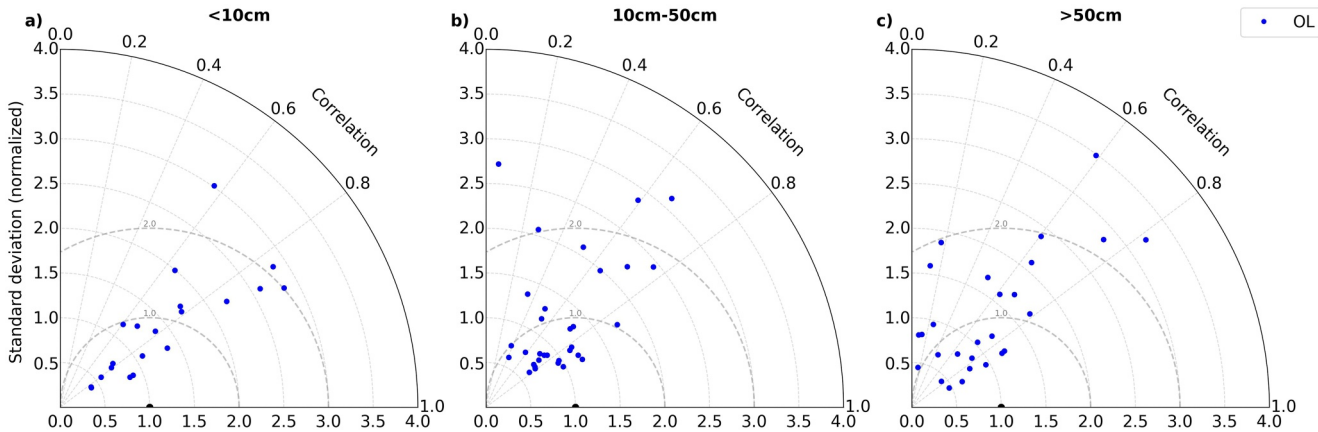


Figure 9. Taylor diagram evaluating modeled monthly soil water content from the open–loop against station data from the ISMN. Panels show depths of <10 cm (a), 10–50 cm (b), and >50 cm (c). The black dot marks the reference point corresponding to the observations (correlation of 1 and normalized standard deviation of 1). Radial distance from the origin indicates the normalized standard deviation, and angular position the correlation coefficient. The distance from the reference point corresponds to the NCRMSE; the two concentric circles centered on the reference point denote Normalized Centered Root Mean Square Error values of 1.0 and 2.0.

2002 in this region (Ulbrich et al., 2003). DA simulations exhibit a similar initial drop of about 100 mm, reaching values that the OL attains only after 1 year. In July 2008, differences between state vector configurations are most pronounced: V1 exceeds V2 by 20 mm and V3 by 8 mm. Across perturbation methods, E2 generally produces the highest deep-layer SWC, frequently 15 mm above E3 and up to 20 mm above E1 and E4. These differences are comparable to the assumed TWSA observation error (20 mm), highlighting the sensitivity of deep soil moisture to state vector design and ensemble perturbation strategy. This sensitivity is also reflected in TWSA differences: for July 2008, V1 exceeds V3 by 7 mm and V2 by 18 mm, while E2 surpasses E1 by 4 mm, E3 by 7 mm, and E4 by 6 mm. Thus, even when total anomalies are similar, the partitioning of moisture across soil layers can differ.

Overall, OL correlation values indicate generally good phase agreement, although amplitudes are often over- or underestimated across depths and stations. DA does not lead to systematic improvements in SWC compared to in situ data, and the tested methods show no major differences in overall skill. However, the experiments demonstrate that DA choices–both state vector configuration and perturbation strategy–can strongly influence the vertical partitioning of TWS, especially in deeper layers. Because eCLM does not explicitly simulate groundwater and in situ SWC measurements are sparse at depth, these differences cannot be fully evaluated here. Using a model that better represents 3-dimensional groundwater-related processes could allow different state vector configurations and perturbation methods to be assessed against well measurements, providing a more complete evaluation of state vector configuration and perturbation method impacts on assimilation results.

5. Conclusions and Outlook

This study is the first to systematically assess how state vector configuration and ensemble generation strategies influence GRACE–DA simulations within a high-resolution LSM.

While TWS estimates were broadly similar across the different state vector configurations, notable differences emerged in how assimilation increments were distributed across the soil profile, with implications for the depth-dependent evolution of SWC. In case V1, where individual storage compartments were explicitly introduced into

Table 4
Mean NCRMSE Between the OL and In Situ Stations From the ISMN as Well as GRACE DA Induced Changes in Mean NCRMSE Compared to the OL for Each Experiment

	OL	E1	E2	E3	E4	V1	V2	V3
<10 cm	1.314	−0.006	−0.046	−0.030	−0.015	−0.016	−0.005	−0.015
10–50 cm	1.250	−0.016	−0.030	−0.033	−0.022	−0.045	−0.027	−0.022
>50 cm	1.392	−0.050	−0.078	−0.104	−0.045	−0.048	−0.052	−0.045

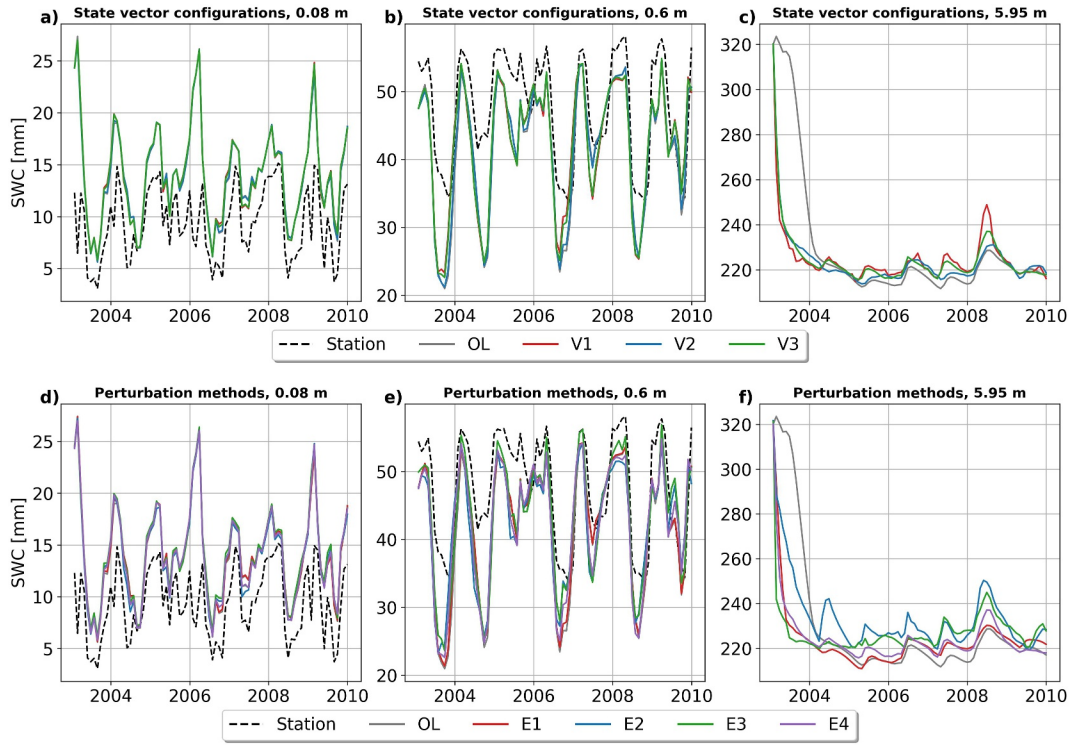


Figure 10. Observed and modeled soil water content at Falkenberg station for (a–c) state vector configurations; and (d–f) ensemble perturbation methods. Panels show depths of 0.08 m (a, d), 0.6 m (b, e), and 5.95 m (c, f). All time series are scaled by model layer thickness. Simulations are shown for the Local Ensemble Subspace Transform Kalman Filter.

the state vector, the assimilation produced strongly depth-varying and partially compensating increments across soil layers, including opposite-signed updates in deeper layers, possibly due to vertical correlations between saturated and unsaturated zones. Configuration V2, in which each state variable represented aggregated TWS, lacked vertical disaggregation, while configuration V3, where soil layers were aggregated into surface, root-zone, and deep soil moisture, offered a trade-off by producing vertically consistent increment patterns without fully collapsing the vertical structure, consistent with expected hydrological behavior, although these patterns cannot be independently validated in the absence of observations across the full soil profile.

Similar sensitivities to state representation have been reported by Retegui-Schiettekatte et al. (2025), who compared fully disaggregated and aggregated state formulations (analogous to V1 and V2) and showed that aggregation can reduce or avoid degradations associated with fully disaggregated configurations, particularly in deeper storage components such as groundwater. Their findings support our interpretation that partial or full aggregation can act as a stabilizing constraint on the vertical redistribution of GRACE/–FO increments. Rather than implying that aggregation is theoretically optimal, we interpret V3 as a pragmatic stabilizing choice under a vertically integrated observational constraint. By reducing the effective degrees of freedom of the vertical update, this configuration helps to limit internally compensating layer-wise adjustments that are weakly constrained by the observation and difficult to interpret physically.

Ensemble perturbation strategies had an even greater effect on the magnitude and vertical distribution of assimilation increments. Perturbing only atmospheric forcings produced generally small increments, mostly limited to near-surface layers due to limited ensemble spread at depth. Including perturbations of soil hydraulic properties increased ensemble spread and led to larger deep-layer increments; however, consistent treatment of both mineral texture and soil organic matter was essential to avoid inconsistencies between the root zone and deeper compartments.

Importantly, these differences arose even when perturbation methods were designed to generate comparable spread in TWS. This demonstrates that, for GRACE/–FO DA, matching TWS variability alone is not sufficient: the perturbation strategy must also produce a covariance structure that yields a coherent redistribution of a

vertically integrated TWS constraint across depths and storage components. In this sense, our results highlight why it is particularly important in GRACE/–FO DA to evaluate how ensemble generation shapes cross-covariances, because these cross-covariances largely determine how TWSA information is partitioned within the model. This is especially relevant because many GRACE/–FO DA frameworks adopt or extend perturbation strategies originally developed for near-surface DA (e.g., soil moisture), often without a systematic assessment of how these choices affect vertical repartitioning under a vertically integrated TWSA constraint. Although the specific perturbation implementations tested here are tailored to eCLM and not intended as a prescriptive ensemble design for other models, this mechanism can be transferred to other GRACE/–FO DA frameworks as well: whenever vertically integrated TWSA is assimilated into models with multiple storages and vertical structure (LSMs and global hydrological models), ensemble design choices can materially alter the inferred vertical and compartmental update patterns, even if the fit to TWSA and its spread are similar.

The magnitude of soil moisture differences–on the order of 20 mm in deep layers–is substantial when compared to the typical impact of GRACE/–FO DA on individual storage components. For example, Tangdamrongsub et al. (2017) reported an average RMSD of 20.6 mm between an OL simulation and in situ groundwater observations in a semi-arid region, which was reduced to 11.9 mm through assimilation. Similarly, Getirana et al. (2020) found that GRACE–DA altered modeled groundwater storage by up to 100 mm when averaged across large regions of the United States. These values indicate that GRACE/–FO DA commonly induces groundwater changes of comparable magnitude. Against this background, the differences we observe across state vector configurations and perturbation strategies in individual deep soil layers are not negligible relative to the observational constraint itself. Although compensating increments in other layers may result in similar TWSA, these findings underscore that internal DA design choices in multi-layered LSMs can strongly influence the vertical distribution of updates, particularly in deeper soil compartments.

Differences between the EnKF and LESTKF further emphasize the role of filter design in moderating these sensitivities. In high-dimensional systems such as eCLM, fully disaggregated state vectors substantially increase the number of state variables whose cross-covariances with TWSA must be estimated from a finite ensemble. In a global EnKF, this can amplify sampling-induced long-range covariances and noisy cross-layer relationships, resulting in counter-intuitive vertical update patterns. In contrast, the LESTKF localizes the analysis in observation space, restricting the influence of each observation to nearby state variables. This reduces the impact of sampling noise and yields more spatially confined and internally consistent updates, making the filter less sensitive to the internal structure of the state vector.

Nevertheless, although the use of localization in the LESTKF reduced differences between DA experiments compared to the global EnKF, substantial discrepancies persisted in deeper soil layers. Localization therefore improves the robustness and spatial coherence of updates but does not fully resolve uncertainties related to vertical coupling and state representation. While previous GRACE/–FO DA studies have shown that localization improves agreement with TWSA observations at regional scales (Khaki, Schumacher, et al., 2017), our results demonstrate that this improved fit does not necessarily translate into reduced sensitivity of vertical redistribution to state representation.

There is a general trend toward more detailed, multi-layered LSMs, also within the framework of fully coupled Earth system models. Based on our experiments with eCLM, we highlight the following key insights to inform future GRACE/–FO DA studies:

1. Aggregating state vectors across multiple soil layers into a few robust variables is recommended, as it reduces state dimensionality, mitigates depth-dependent artifacts, and still enables vertical downscaling of TWSA observations.
2. Ensemble design strongly controls the cross-covariances that partition GRACE/–FO updates across storages and depth; therefore, perturbation strategies should be evaluated for both TWS spread and the resulting vertical redistribution behavior.
3. Applying a localized filter is advisable, as it can partially mitigate the influence of differing state vector configurations and perturbation methods, while also improving the representation of local hydrological conditions.

Recent advances in multivariate DA frameworks increasingly aim to integrate multiple observation types–next to GRACE/–FO TWSA also in situ measurements and further remote sensing products such as surface SWC or land

surface temperature (X. Li et al., 2024; Montzka et al., 2012; Soltani et al., 2024; Tangdamrongsub et al., 2020). In such multi-sensor settings, the design of the state vector and ensemble generation becomes even more critical. Different observation types require distinct observation operators, are sensitive to different hydrological compartments, and may necessitate observation specific perturbation strategies. Our findings highlight the need for carefully harmonized DA framework designs to maintain internal consistency across components. For instance, state vectors tailored to surface or shallow soil moisture observations may require finer vertical disaggregation than those aimed at deep compartments primarily constrained by GRACE/–FO. Future research should explore how to balance observation specific needs with generalizability, and how to incorporate observation type specific uncertainties without introducing inconsistencies. These efforts will be essential for building multi-sensor DA frameworks that consistently represent the water cycle across compartments and spatial scales.

Appendix A: Computation of Soil Properties

Mineral properties are computed according to the relationship to texture (Cosby et al., 1984). Soil matric potential $\Psi_{\min,i}$, porosity $\Theta_{\min,i}$, the shape parameter $b_{\min,i}$ and saturated hydraulic conductivity $k_{sat,min,i}$ of the mineral part are computed according to

$$\Psi_{\min,i} = 10 * \left(10^{1.88-0.0131 * sand}\right) \tag{A1}$$

$$\Theta_{\min,i} = 0.489 - 0.00126 * sand \tag{A2}$$

$$b_{\min,i} = 2.91 + 0.159 * clay \tag{A3}$$

$$k_{sat,min,i} = 0.0070556 * 10^{-0.884+0.0153 * sand}, \tag{A4}$$

where i corresponds to the vertical level. The organic part of the hydraulic properties are computed according to Lawrence and Slater (2008), with

$$\Psi_{om,i} = \min(10.1, 10.3 - 0.2D_i) \tag{A5}$$

$$\Theta_{om,i} = \max(0.83, 0.93 - 0.1D_i) \tag{A6}$$

$$b_{om,i} = \max(12, 2.7 + 9.3D_i) \tag{A7}$$

$$k_{sat,om,i} = \max(K_{\min,i}, 0.28 - 0.2799D_i) \tag{A8}$$

and

$$D_i = \frac{z_i}{zsapric}, \tag{A9}$$

where z_i is the depth of the vertical level and $zsapric$ is the depth where organic matter takes on the characteristics of sapric peat which is fixed to 0.5 m. The mineral properties are adjusted using a weighted averaging depending on the organic fraction $f_{om,i}$ which is computed via

$$f_{om,i} = \frac{om}{om_{\max}}, \tag{A10}$$

with om being the organic matter content and $om_{\max} = 130\ \mathrm{kg/m^3}$ being the maximum organic matter content. The hydraulic properties are adjusted via

$$\Psi_i = \left(1 - f_{om,i}\right)\Psi_{\min,i} + f_{om,i}\Psi_{om,i} \tag{A11}$$

$$\Theta_i = \left(1 - f_{om,i}\right)\Theta_{\min,i} + f_{om,i}\Theta_{om,i} \tag{A12}$$

$$b_i = \left(1 - f_{om,i}\right) b_{\min,i} + f_{om,i} b_{om,i} \tag{A13}$$

Saturated hydraulic conductivity is assumed to correspond to flow pathways that pass through the mineral and organic components, with

$$k_{sat,uncon} = f_{uncon}\left(\frac{\left(1 - f_{om}\right)}{k_{sat,min}} + \frac{\left(f_{om} - f_{perc}\right)}{k_{sat,om}}\right)^{-1} \tag{A14}$$

Here, f_{perc} is the connected portion of the grid and computed via

$$f_{perc} = \begin{cases} N_{perc}\left(f_{om} - f_{threshold}\right)^{\beta_{perc}} f_{om}, & f_{om} \geq f_{threshold} \\ f_{perc} = 0, & f_{om} < f_{threshold} \end{cases}, \tag{A15}$$

with $\beta_{perc} = 0.139$, $f_{threshold} = 0.5$ and $N_{perc} = \left(1 - f_{threshold}\right)^{-\beta_{perc}}$. f_{uncon} is the unconnected portion of the grid cell according to

$$f_{uncon} = \left(1 - f_{perc}\right) \tag{A16}$$

Finally, the weighted saturated hydraulic conductivity can be computed via

$$k_{sat,i} = f_{uncon,i} k_{sat,uncon,i} + \left(1 - f_{uncom,i}\right) k_{sat,om,i} \tag{A17}$$

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The simulations presented in this study have been carried out using eCLM within the TSMP–PDAF framework, available at https://github.com/HPSCTerrSys/. Atmospheric forcings from ERA5 (Copernicus Climate Change Service (C3S), 2023; Hersbach et al., 2020) can be downloaded from the Copernicus Climate Data Store: https://cds.climate.copernicus.eu/. The GRACE–TWSA data are available at https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018, see also Mayer-Gürr et al. (2018). In situ soil moisture data from the ISMN can be downloaded from https://ismn.earth/en/. The model output data supporting the findings of this study are available via Zenodo (Ewerdwalbesloh, 2026). Due to the large volume of data produced by the full simulations, a representative subset consisting the first full year of output has been archived. The complete simulation output is available from the corresponding author upon reasonable request.

Acknowledgments
Y.E., C.M., J.K. and A.S. acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the CRC DETECT (DFG–SFB 1502/1–2022–Projektnummer: 450058266). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). We sincerely thank the two anonymous reviewers for their valuable feedback, which greatly improved this paper. Open Access funding enabled and organized by Projekt DEAL.

References

Bartholomé, E., & Belward, A. S. (2005). GLC2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, *26*(9), 1959–1977. https://doi.org/10.1080/01431160412331291297

Brooks, R. H., & Corey, A. T. (1964). *Hydraulic properties of porous media (hydrology paper no. 3)*. Colorado State University.

Caron, L., Ivins, E. R., Larour, E., Adhikari, S., Nilsson, J., & Blewitt, G. (2018). GIA model statistics for GRACE hydrology, cryosphere, and ocean science. *Geophysical Research Letters*, *45*(5), 2203–2212. https://doi.org/10.1002/2017GL076644

Cheng, M., Ries, J. C., & Tapley, B. D. (2011). Variations of the Earth's figure axis from satellite laser ranging and GRACE. *Journal of Geophysical Research*, *116*(B1), B01409. https://doi.org/10.1029/2010JB000850

Cheng, Y., Huang, M., Zhu, B., Bisht, G., Zhou, T., Liu, Y., et al. (2021). Validation of the community land model version 5 over the contiguous United States (CONUS) using in situ and remote sensing data sets. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033539. https://doi.org/10.1029/2020JD033539

Chi, H., Seo, H., & Kim, Y. (2024). Hydrological trends captured by assimilating GRACE total water storage data into the CLM5-BGC model. *Journal of Hydrology*, *629*, 130527. https://doi.org/10.1016/j.jhydrol.2023.130527

Copernicus Climate Change Service (C3S). (2023). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, *20*(6), 682–690. https://doi.org/10.1029/WR020i006p00682

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N., & Verhoest, N. E. C. (2007). Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resources Research*, *43*(9). https://doi.org/10.1029/2006WR005449

De Lannoy, G. J. M., Rosnay, P., & Reichle, R. H. (2019). Soil moisture data assimilation. https://doi.org/10.1007/978-3-642-39925-1_32

de Rosnay, P., Drusch, M., Vasiljevic, D., Balsamo, G., Albergel, C., & Isaksen, L. (2013). A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1199–1213. https://doi.org/10.1002/qj.2023

Dorigo, W. A., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., et al. (2021). The international soil moisture network: Serving Earth system science for over a decade. *Hydrology and Earth System Sciences*, *25*(11), 5749–5804. https://doi.org/10.5194/hess-25-5749-2021

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011). The international soil moisture network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5), 1675–1698. https://doi.org/10.5194/hess-15-1675-2011

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A., et al. (2013). Global automated quality control of in situ soil moisture data from the international soil moisture network. *Vadose Zone Journal*, *12*(3). https://doi.org/10.2136/vzj2012.0097

Eicker, A., Schumacher, M., Kusche, J., Döll, P., & Schmied, H. M. (2014). Calibration/data assimilation approach for integrating GRACE data into the WaterGAP global hydrology model (WGHM) using an ensemble kalman filter: First results. *Surveys in Geophysics*, *35*(6), 1285–1309. https://doi.org/10.1007/s10712-014-9309-8

Eloundou, F. B., Strebel, L., Naz, B. S., Teran, C. P., Vereecken, H., & Franssen, H.-J. H. (2025). Disentangling sources of uncertainty in clm5 model predictions: Water, energy, and carbon fluxes at European observation sites. *ESSOAr Preprint*. https://doi.org/10.22541/essoar.174526373.38206874/v1

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143–10162. https://doi.org/10.1029/94JC00572

Evensen, G. (2009). Data assimilation: The ensemble kalman filter. https://doi.org/10.1007/978-3-642-03711-5

Ewerdwalbesloh, Y. (2026). eCLM simulation outputs for GRACE data assimilation experiments on state vector and ensemble design [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.18590052

Fink, A. H., Brücher, T., Krüger, A., Leckebusch, G. C., Pinto, J. G., & Ulbrich, U. (2004). The 2003 European summer heatwaves and drought –Synoptic diagnosis and impacts. *Weather*, *59*(8), 209–216. https://doi.org/10.1256/wea.73.04

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

Forman, B. A., Reichle, R. H., & Rodell, M. (2012). Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resources Research*, *48*(1). https://doi.org/10.1029/2011WR011239

Gerdener, H., Kusche, J., Schulze, K., Döll, P., & Klos, A. (2023). The global land water storage data set release 2 (GLWS2.0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model. *Journal of Geodesy*, *97*(7), 73. https://doi.org/10.1007/s00190-023-01763-9

Getirana, A., Rodell, M., Kumar, S., Beaudoing, H. K., Arsenault, K., Zaitchik, B., et al. (2020). GRACE improves seasonal groundwater forecast initialization over the United States. *Journal of Hydrometeorology*, *21*(1), 59–71. https://doi.org/10.1175/JHM-D-19-0096.1

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, *52*(5), 4164–4183. https://doi.org/10.1002/2015WR018417

Girotto, M., Musselman, K. N., & Essery, R. L. H. (2020). Data assimilation improves estimates of climate-sensitive seasonal snow. *Current Climate Change Reports*, *6*(3), 81–94. https://doi.org/10.1007/s40641-020-00159-7

Girotto, M., Reichle, R. H., Rodell, M., Liu, Q., Mahanama, S., & De Lannoy, G. J. M. (2019). Multi-sensor assimilation of SMOS brightness temperature and GRACE terrestrial water storage observations for soil moisture and shallow groundwater estimation. *Remote Sensing of Environment*, *227*, 12–27. https://doi.org/10.1016/j.rse.2019.04.001

Gutowski, W. J., Jr., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., et al. (2016). WCRP COordinated regional downscaling EXperiment (CORDEX): A diagnostic MIP for CMIP6. *Geoscientific Model Development*, *9*(11), 4087–4095. https://doi.org/10.5194/gmd-9-4087-2016

Han, X., Franssen, H.-J. H., Montzka, C., & Vereecken, H. (2014). Soil moisture and soil properties estimation in the community land model with synthetic brightness temperature observations. *Water Resources Research*, *50*(7), 6081–6105. https://doi.org/10.1002/2013WR014586

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated gravity recovery and climate experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7). https://doi.org/10.1029/2011WR011291

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, *34*(12), 3405–3420. https://doi.org/10.1029/1998WR900001

Indoria, A. K., Sharma, K. L., & Reddy, K. S. (2020). Chapter 18 - Hydraulic properties of soil under warming climate. *Climate Change and Soil Interactions*, 473–508. https://doi.org/10.1016/B978-0-12-818032-7.00018-7

Jacob, D., Teichmann, C., Sobolowski, S., Katragkou, E., Anders, I., Belda, M., et al. (2020). Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Regional Environmental Change*, *20*(2), 51. https://doi.org/10.1007/s10113-020-01606-9

Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A. I. J. M., et al. (2017). Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources*, *107*, 301–316. https://doi.org/10.1016/j.advwatres.2017.07.001

Khaki, M., Schumacher, M., Forootan, E., Kuhn, M., Awange, J. L., & van Dijk, A. I. J. M. (2017). Accounting for spatial correlation errors in the assimilation of GRACE into hydrological models through localization. *Advances in Water Resources*, *108*, 99–112. https://doi.org/10.1016/j.advwatres.2017.07.024

Kurtz, W., He, G., Kollet, S. J., Maxwell, R. M., Vereecken, H., & Hendricks Franssen, H.-J. (2016). TerrSysMP–PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface–subsurface model. *Geoscientific Model Development*, *9*(4), 1341–1360. https://doi.org/10.5194/gmd-9-1341-2016

Kusche, J. (2007). Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. *Journal of Geodesy*, *81*(11), 733–749. https://doi.org/10.1007/s00190-007-0143-3

Kvas, A., Behzadpour, S., Ellmer, M., Klinger, B., Strasser, S., Zehentner, N., & Mayer-Gürr, T. (2019). ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series. *Journal of Geophysical Research: Solid Earth*, *124*(8), 9332–9344. https://doi.org/10.1029/2019JB017415

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

Lawrence, D. M., & Slater, A. G. (2008). Incorporating organic soil into a global climate model. *Climate Dynamics*, *30*(2), 145–160. https://doi.org/10.1007/s00382-007-0278-1

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., et al. (2019). Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges. *Water Resources Research*, *55*(9), 7564–7586. https://doi.org/10.1029/2018WR024618

Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D., & van Dam, T. M. (2012). Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *Journal of Hydrology*, *446–447*, 103–115. https://doi.org/10.1016/j.jhydrol.2012.04.035

Li, X., Liu, F., Ma, C., Hou, J., Zheng, D., Ma, H., et al. (2024). Land data assimilation: Harmonizing theory and data in land surface process studies. *Reviews of Geophysics*, *62*(1), e2022RG000801. https://doi.org/10.1029/2022RG000801

Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., & Zehentner, N. (2018). ITSG-Grace2018–Monthly, daily and static gravity field solutions from GRACE [Dataset]. *GFZ Data Services*. https://doi.org/10.5880/ICGEM.2018.003

Mohamed, A. (2019). Hydro-geophysical study of the groundwater storage variations over the Libyan area and its connection to the Dakhla basin in Egypt. *Journal of African Earth Sciences*, *157*, 103508. https://doi.org/10.1016/j.jafrearsci.2019.05.016

Montzka, C., Pauwels, V. R. N., Franssen, H.-J. H., Han, X., & Vereecken, H. (2012). Multivariate and multiscale data assimilation in terrestrial systems: A review. *Sensors*, *12*, 16291–16333. https://doi.org/10.3390/s121216291

Naz, B. S., Kurtz, W., Montzka, C., Sharples, W., Goergen, K., Keune, J., et al. (2019). Improving soil moisture and runoff simulations at 3 km over Europe using land surface data assimilation. *Hydrology and Earth System Sciences*, *23*(1), 277–301. https://doi.org/10.5194/hess-23-277-2019

Nerger, L., & Hiller, W. (2013). Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, *55*, 110–118. https://doi.org/10.1016/j.cageo.2012.03.026

Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A unification of ensemble square root kalman filters. *Monthly Weather Review*, *140*(7), 2335–2345. https://doi.org/10.1175/MWR-D-11-00102.1

Neumann, K., Herold, M., Hartley, A., & Schmullius, C. (2007). Comparative assessment of CORINE2000 and GLC2000: Spatial analysis of land cover data for Europe. *International Journal of Applied Earth Observation and Geoinformation*, *9*(4), 425–437. https://doi.org/10.1016/j.jag.2007.02.004

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. *Journal of Geophysical Research*, *112*(D7). https://doi.org/10.1029/2006JD007522

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. https://doi.org/10.5194/soil-7-217-2021

Poll, S., Rigor, P., Brdar, S., Ho-Hagemann, H. T. M., Hartick, C., van Hulten, M., et al. (2025). The new tsmp2 coupled Earth system model. *EGUsphere*, *2025*, 1–28. https://doi.org/10.5194/egusphere-2025-5468

Retegui-Schiettekatte, L., Schumacher, M., Yang, F., Madsen, H., & Forootan, E. (2025). An ensemble Kalman filter with rescaling disaggregation for assimilating terrestrial water storage into hydrological models. *Scientific Reports*, *15*(1), 28675. https://doi.org/10.1038/s41598-025-13602-2

Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3). https://doi.org/10.1029/2007WR006357

Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P. P., Njoku, E. G., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the advanced microwave scanning radiometer for the Earth observing system (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). *Journal of Geophysical Research*, *112*(D9). https://doi.org/10.1029/2006JD008033

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Scanlon, B. R., Zhang, Z., Rateb, A., Sun, A., Wiese, D., Save, H., et al. (2019). Tracking seasonal fluctuations in land water storage using global models and GRACE satellites. *Geophysical Research Letters*, *46*(10), 5254–5264. https://doi.org/10.1029/2018GL081836

Schulze, K., Kusche, J., Gerdener, H., Döll, P., & Müller Schmied, H. (2024). Benefits and pitfalls of GRACE and streamflow assimilation for improving the streamflow simulations of the WaterGAP global hydrology model. *Journal of Advances in Modeling Earth Systems*, *16*(10), e2023MS004092. https://doi.org/10.1029/2023MS004092

Schumacher, M., Eicker, A., Kusche, J., Schmied, H. M., & Döll, P. (2016). Covariance analysis and sensitivity studies for GRACE assimilation into WGHM. https://doi.org/10.1007/1345_2015_119

Schumacher, M., Kusche, J., & Döll, P. (2016). A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. *Journal of Geodesy*, *90*(6), 537–559. https://doi.org/10.1007/s00190-016-0892-y

Slater, A. G., & Clark, M. P. (2006). Snow data assimilation via an ensemble kalman filter. *Journal of Hydrometeorology*, *7*(3), 478–493. https://doi.org/10.1175/JHM505.1

Soltani, S. S., Ataie-Ashtiani, B., Al Bitar, A., Simmons, C. T., Younes, A., & Fahs, M. (2024). Assimilating multivariate remote sensing data into a fully coupled subsurface-land surface hydrological model. *Journal of Hydrology*, *641*, 131812. https://doi.org/10.1016/j.jhydrol.2024.131812

Springer, A. (2019). A water storage reanalysis over the European continent: Assimilation of grace data into a high-resolution hydrological model and validation. *Doctoral Dissertation, Rheinische Friedrich-Wilhelms-Universität Bonn*. Retrieved from https://hdl.handle.net/20.500.11811/7987

Springer, A., De Lannoy, G., Rodell, M., Ewerdwalbesloh, Y., Gerdener, H., Khaki, M., et al. (2026). A review of current best practices and future directions in assimilating grace/-fo terrestrial water storage data into numerical models. *Hydrology and Earth System Sciences*, *30*(4), 985–1022. https://doi.org/10.5194/hess-30-985-2026

Springer, A., Karegar, M. A., Kusche, J., Keune, J., Kurtz, W., & Kollet, S. (2019). Evidence of daily hydrological loading in GPS time series over Europe. *Journal of Geodesy*, *93*(10), 2145–2153. https://doi.org/10.1007/s00190-019-01295-1

Strebel, L., Bogena, H. R., Vereecken, H., & Hendricks Franssen, H.-J. (2022). Coupling the community land model version 5.0 to the parallel data assimilation framework PDAF: Description and applications. *Geoscientific Model Development*, *15*(2), 395–411. https://doi.org/10.5194/gmd-15-395-2022

Sun, C., Walker, J. P., & Houser, P. R. (2004). A methodology for snow data assimilation in a land surface model. *Journal of Geophysical Research*, *109*(D8). https://doi.org/10.1029/2003JD003765

Sun, Y., Riva, R., & Ditmar, P. (2016). Optimizing estimates of annual variations and trends in geocenter motion and J2 from a combination of GRACE data and geophysical models. *Journal of Geophysical Research: Solid Earth*, *121*(11), 8352–8370. https://doi.org/10.1002/2016JB013073

Swenson, S., Chambers, D., & Wahr, J. (2008). Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research*, *113*(B8). https://doi.org/10.1029/2007JB005338

Tangdamrongsub, N., Han, S.-C., Yeo, I.-Y., Dong, J., Steele-Dunne, S. C., Willgoose, G., & Walker, J. P. (2020). Multivariate data assimilation of GRACE, SMOS, SMAP measurements for improved regional soil moisture and groundwater storage estimates. *Advances in Water Resources*, *135*, 103477. https://doi.org/10.1016/j.advwatres.2019.103477

Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., Sutanudjaja, E. H., Sun, Y., et al. (2017). Improving estimates of water resources in a semi-arid region by assimilating GRACE data into the PCR-GLOBWB hydrological model. *Hydrology and Earth System Sciences*, *21*(4), 2053–2074. https://doi.org/10.5194/hess-21-2053-2017

Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., & Weerts, A. H. (2015). Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin. *Hydrology and Earth System Sciences*, *19*(4), 2079–2100. https://doi.org/10.5194/hess-19-2079-2015

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., et al. (2019). Contributions of GRACE to understanding climate change. *Nature Climate Change*, *9*(5), 358–369. https://doi.org/10.1038/s41558-019-0456-2

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I. J. M., Walker, J. P., Pauwels, V. R. N., & Allgeyer, S. (2017). Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research*, *53*(3), 1820–1840. https://doi.org/10.1002/2016WR019641

Ukkola, A. M., De Kauwe, M. G., Pitman, A. J., Best, M. J., Abramowitz, G., Haverd, V., et al. (2016). Land surface models systematically overestimate the intensity, duration and magnitude of seasonal-scale evaporative droughts. *Environmental Research Letters*, *11*(10), 104012. https://doi.org/10.1088/1748-9326/11/10/104012

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., & Pinto, J. G. (2003). The central European floods of August 2002: Part 1 – Rainfall periods and flood development. *Weather*, *58*(10), 371–377. https://doi.org/10.1256/wea.61.03A

Wahr, J., Molenaar, M., & Bryan, F. (1998). Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research*, *103*(B12), 30205–30229. https://doi.org/10.1029/98JB02844

Wongchuig, S., Paiva, R., Siqueira, V., Papa, F., Fleischmann, A., Biancamaria, S., et al. (2024). Multi-satellite data assimilation for large-scale hydrological-hydrodynamic prediction: Proof of concept in the Amazon Basin. *Water Resources Research*, *60*(8), e2024WR037155. https://doi.org/10.1029/2024WR037155

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the mississippi River Basin. *Journal of Hydrometeorology*, *9*(3), 535–548. https://doi.org/10.1175/2007JHM951.1

Zhao, H., Montzka, C., Keller, J., Li, F., Vereecken, H., & Hendricks Franssen, H.-J. (2025). How does assimilating SMAP soil moisture improve characterization of the terrestrial water cycle in an integrated land surface-subsurface model? *Water Resources Research*, *61*(6), e2024WR038647. https://doi.org/10.1029/2024WR038647

Zheng, H., Fei, W., Yang, Z.-L., Wei, J., Zhao, L., Li, L., & Wang, S. (2023). An ensemble of 48 physically perturbed model estimates of the 1/8° terrestrial water budget over the conterminous United States, 1980–2015. *Earth System Science Data*, *15*(7), 2755–2780. https://doi.org/10.5194/essd-15-2755-2023