PyGLDA: a fine-scale Python-based Global Land Data Assimilation system for integrating satellite gravity data into hydrological models

Fan Yang[1,2], Maike Schumacher[1], Leire Retegui-Schiettekatte[1], Albert I.J.M. van Dijk[3], and Ehsan Forootan[1]

[1]Geodesy Group, Department of Sustainability and Planning, Aalborg University, Aalborg 9000, Denmark
[2]School of Physics, Huazhong University of Science and Technology, Wuhan 430074, China
[3]Fenner School of Environment & Society, College of Science, Australian National University, Canberra 2600, Australia

Correspondence: Fan Yang (fany@plan.aau.dk)

Abstract.

Data assimilation (DA) of time-variable satellite gravity observations, such as those from the Gravity Recovery and Climate Experiment (GRACE), GRACE Follow-On (GRACE-FO), and future gravity missions, can be used to constrain simulations of the vertical sum of water storage in Global Hydrological Models (GHMs). However, current DA implementations of these Terrestrial Water Storage (TWS) changes are often performed at regional scales or, if applied globally, at low spatial resolutions. This limitation is primarily due to the high computational demands of DA and numerical challenges, such as instabilities in covariance matrix inversion. To fully exploit the potential of satellite gravity observations and the high spatial resolution of GHMs, we developed PyGLDA, an open-source Python-based system that enables fine-scale and computationally efficient global DA. The key innovations of PyGLDA include (1) a global patch-wise DA approach using domain localization and neighboring-weighted global aggregation and (2) seamless compatibility between basin-scale and grid-scale DA implementations. PyGLDA represents a significant functional improvement over previous DA systems, offering wide-ranging and flexible options for user-specific applications. The modular structure of the system allows users to customize water storage compartments, modify observation representations, and potentially select different GHMs. This paper provides a comprehensive description of PyGLDA and its application in a case study of the Danube River Basin, along with a demonstration of global DA, where experiments involve integrating monthly GRACE TWS fields (2002-2010) with the daily W3RA water balance model at 0.1° spatial resolution.

1 Introduction

Accurate monitoring of the global water cycle is essential to understand natural climate change, as well as human

water resource management (Rodell et al., 2018). However, measuring global changes in Terrestrial Water Storage (TWS) —

25 which represents the total vertical sum of water stored on land, including ice, snow, wetlands, lakes, rivers, soil moisture and groundwater — has long been a challenge (Rodell and Reager, 2023). This difficulty arises from the vast spatial coverage required (approximately 150 million km^2 of the global land area) and the lack of comprehensive observational data for different hydrological components. In situ measurements of key variables, such as groundwater, are often spatially and temporally limited, subject to political constraints, and prone to data gaps and systematic errors. These factors reduce their suitability for

30 large-scale hydrological studies [Reformulated as suggested] (Li et al., 2019).

~~Therefore~~To address these challenges, Global Hydrological Models (GHMs) ~~of various structures~~ have been developed to simulate the spatially distributed hydrological response to weather and climate variations, as well as human water use. Most GHMs are conceptually or physically based, and one can see [the text here and the remainder of this paragraph is reformulated as suggested by reviewer] Sutanudjaja et al. (2018) for an extensive list of existing GHMs. These models operate at various spatial resolutions, ranging

35 from coarser scales (e.g., $0.5° \times 0.5°$ for WaterGap, Müller Schmied et al., 2021) to finer resolutions (e.g., $0.01° \times 0.01°$ for the hyper-resolution PCR-GLOBWB, Hoch et al., 2023). Despite their advancements, these models often struggle to realistically simulate certain water storage and flux processes due to oversimplified representations of natural processes or a lack of constraints on the temporal variability of some components [reformulated as suggested], such as groundwater and human interventions. This issue persists in most GHMs (Scanlon et al., 2018; Mehrnegar et al., 2020), although some models may perform better

40 than others in specific river basins. Recent efforts have aimed to improve the representation of certain components, such as groundwater processes, to more accurately simulate groundwater levels and surface–groundwater interactions. However, significant modeling challenges persist [reformulated] (see Niu et al., 2007; de Graaf et al., 2017; Decharme et al., 2019). Furthermore, calibrating model parameters to better capture local water variability remains a major challenge, which requires innovative observational and calibration techniques (Schumacher et al., 2018a).

45 ~~Direct remote sensing observations~~Since 2002, satellite observations of time-variable gravity changes have been available through the US-German Gravity Recovery and Climate Experiment (GRACE) mission (Tapley et al., 2004). By analyzing these ~~fields~~data, ~~one~~researchers can estimate global TWS changes with reasonable precision (Landerer and Swenson, 2012). ~~Over last~~During the past two decades, GRACE and its Follow-On mission (GRACE-FO; Landerer et al., 2020) have ~~helped hydrologists and meteorologists~~contributed significantly to hydrological and meteorological studies. These satellite-based TWS observations

50 have been used to close the global water budget (e.g., Sheffield et al., 2009; Castle et al., 2016), estimate groundwater depletion (e.g., Scanlon et al., 2012; Voss et al., 2013; Castellazzi et al., 2016), analyze droughts and flood potential (e.g., Gouweleeuw et al., 2018; Forootan et al., 2019), validate and constrain hydrological models [revised as suggested] (e.g., Ngo-Duc et al., 2007; Eicker et al., 2016; Mehrnegar et al., 2020, 2023). However, despite its advantages, the GRACE-derived TWS anomalies (abbreviated as TWS hereinafter) [revised as suggested] are limited by coarse spatial and temporal resolution ($\sim$~~500~~330 km or $3° \times 3°$

55 on a monthly basis) due to the presence of ~~its~~ considerable correlated noise (Kusche et al., 2009; Chen et al., 2022; Yang et al., 2024b).

To introduce realistic variability from observations into model simulations, ~~the Data Assimilation (DA) technique has found a great interest~~Data Assimilation (DA) techniques have gained significant attention. ~~Especially~~In particular, DA of ~~satellite based~~GRACE-based

TWS observations can enhance the temporal, horizontal, and vertical disaggregation of the estimated TWS signal (Zaitchik
60 et al., 2008; Schumacher, 2016). ~~TWS~~GRACE data provide unique constraints on model-based water storage estimates that cannot be obtained through other observational techniques. ~~A majority of DA studies in previous have confirmed its unique advantage of improving TWS, as well as its compartments on regional/basin scale~~Numerous studies have demonstrated the effectiveness of DA with GRACE in improving TWS simulations and its individual components, particularly at regional and basin scales (Eicker et al., 2014; van Dijk et al., 2014; Schumacher et al., 2016; Girotto et al., 2016; Khaki et al., 2017; Tian et al., 2017). ~~In recent years, there~~
65 ~~is an emerging trend to implement the DA on global scale to better reveal~~More recently, there has been a growing interest in implementing GRACE-TWS DA on a global scale to better capture the global water cycle (Li et al., 2019; Felsberg et al., 2021; Gerdener et al., 2023; Forootan et al., 2024).

[a new paragraph is started as suggested]However, few studies have achieved high-resolution global DA of GRACE-TWS and hydrological models. Li et al. (2019) reported a 0.25° global DA resolution, which, to the best of our knowledge represents the
70 highest resolution achieved so far. Although the value of high-resolution hydrological modeling (for example the resolution of 0.1°) has been widely acknowledged (Benedict et al., 2017; Sutanudjaja et al., 2018; Springer et al., 2019), the implementation of high-resolution global DA remains highly challenging due to numerical instabilities and computational constraints. The latter, in particular, poses a critical challenge, often rendering traditional DA frameworks impractical for large-scale applications. Additionally, none of the existing global DA implementations for GRACE-TWS have made their software publicly available,
75 limiting transparency and reproducibility within the scientific community.

In this ~~study~~article, we introduce PyGLDA, the first open-source and Python-based Global Land Data Assimilation (GLDA) system designed to address the challenges of high-resolution global DA. PyGLDA enables the integration of satellite-based TWS observations from GRACE, GRACE-FO, and likely future gravity missions (Daras et al., 2023) with GHMs. In this version, W3RA model (van Dijk, 2010) is selected as the GHM. The software is a key outcome of the ~~DANSk-LSM~~DANSK-LSM
80 project (https://aaugeodesy.com/dansk-lsm/) and will be expanded in the future for multi-sensor hydrological DA applications. Unlike previous studies, PyGLDA introduces a fully restructured DA framework to efficiently and stably implement high-resolution global DA at 0.1-degree spatial resolution and a daily time step. The system includes several key innovations: [The following text has been reformulated as suggested, to avoid overselling.]

– A unified framework for both basin-scale and grid-scale DA, where the ~~are to be computed~~area to be covered is defined as a
85 'basin' and each grid cell acts as a 'sub-basin,' allowing for flexible spatial configurations via standard shapefiles;

– A novel approach for transitioning from regional DA to global DA, achieved through domain localization and neighboring-weighted aggregation algorithms;

– The ability to compute spatial covariance of GRACE-based TWS observations, enabling full covariance matrices for sub-basins and grids, which improves DA reliability by considering spatial correlations;

90 – ~~Flexible options can be selected for the grid resolution of GHMs, where here a choice of 0.1° and 0.5° is available for the employed global W3RA model~~

- Flexible ~~choices~~choice of the grid/sub-basin ~~TWS~~GRACE-TWS observations to be assimilated into GHMs, e.g., from $1°$ to $5°$ with an increment of $0.5°$;

- A range of perturbation strategies for generating ensemble members, allowing users to select which forcing data and which model parameters are to be perturbed following which noise distribution (Gaussian or Triangle distribution, as well as additive or multiplicative noise);

- The Ensemble Kalman filter (EnKF similar to Schumacher, 2016), and an Ensemble Kalman Smoother (EnKS) that optimally desegregates monthly increments to daily ones (similar to Tian et al., 2017), are implemented as mergers.

The architecture of the code includes the following features:

- A flexible modular structure to decouple PyGLDA into three individual modules: (1) GHM, (2) ~~gravity data post-processing~~GRACE processing, and (3) ~~mathematical~~ DA integration. This choice is made to make it possible to ~~easily~~ develop/modify/replace any individual module;

- PyGLDA is implemented in object-oriented Python, making it accessible and easy to use, while maintaining computational efficiency comparable to C++, Fortran, and MATLAB. The Python translation of the W3RA mode [reformulated]l (van Dijk, 2010) is distributed as well;

- Cross-platform compatibility, allowing installation on Windows, Linux, and High Performance Computing (HPC) clusters; Configurable settings via JSON files, enabling users to adjust parameters and options without modifying the source code; Efficient data management, using the modern HDF5 format for handling large spatial-temporal datasets [reformulated].

The paper is structured as follows. In Section 2, we describe the data ~~used in~~required for PyGLDA. In Section 3, a detailed description of employed GHM and the development of PyGLDA is presented. Section 4 introduces some application examples to demonstrate the effectiveness and flexibility of PyGLDA. The current limitations of PyGLDA are summarized in [a new section to address the limitations is added as suggested] Section 5. Finally, conclusions and outlooks are demonstrated in Section 6.

2 Data

2.1 Masks

~~Various data masks have been defined by PyGLDA to help manage the DA processes. A flexible combination of these masks could allow for various user-specific studies as long as the masks are specified, whether the desired DA is basin-scale or grid-scale and whether the DA is run globally or locally.~~PyGLDA defines various data masks to facilitate the management of DA processes. A flexible combination of these masks allows users to conduct a wide range of studies, whether at the basin scale or grid scale, and whether DA is performed globally or locally. The system supports five primary mask types [their definitions are slightly revised for better clarity], which must be properly configured: (1) land mask that identifies grid points at land; (2) data mask that indicates areas where forcing fields and model parameter

fields are available; (3) region mask that specifies where one wishes to run GHM; (4) basin mask that defines where one wishes to perform DA; (5) observation mask that represents sub-basins or grid cells where observational data are assimilated within the DA process. PyGLDA includes default options for the land mask and data mask, but users must define the region mask, basin mask, and observation mask using a standard shapefile format (.shp). This approach ensures flexibility and adaptability, allowing users to tailor PyGLDA to their specific research needs.

2.2 Meteorological forcing field

So far, only ERA5-land reanalysis (Hersbach et al., 2020) data has been integrated into PyGLDA. However, this does not mean that the implementation is restricted to this forcing field. Instead, one can easily modify the interface module to load other meteorological forcing data. In the current PyGLDA, the ERA5-land at a native (also its highest) resolution of $0.1°$ is required to run W3RA (as an example), which includes the daily precipitation, the maximal/minimal temperature and the surface solar radiation. In particular, PyGLDA provides a function to download these forcing data from the ECMWF (European Centre for Medium Range Weather Forecast) website automatically, which one can use to keep the forcing field up-to-date. In addition, an up(or down)-scaling of forcing fields can be done with the official 'metview' package [added as suggested by reviewer] (from ECMWF, https://metview.readthedocs.io/en/latest/) integrated in PyGLDA, if one wishes to run PyGLDA at a different spatial resolution.

2.3 Parameters and Climatologies

The parameters and climate input data are originally distributed along with the employed GHM, i.e., W3RA, which has an original resolution of $0.5°$. PyGLDA provides a finer resolution of $0.1°$ as a default option. For a detailed review of the definition of these parameters and climatological resources, one can refer to the W3RA technical document (van Dijk, 2010).

2.4 GRACE Level-2 product and its variance-covariance information

In PyGLDA, the monthly GRACE-based TWS is computed from standard level-2 gravity field products, which are expressed in terms of spherical harmonic coefficients up to degree/order 60 or 96 (Yang et al., 2017). These L2 products are provided by three official data centers, i.e., CSR (Center for Space Research from the University of Texas at Austin, Texas, USA), GFZ (GeoForschungsZentrum, Potsdam, Germany) and JPL (Jet Propulsion Laboratory, USA). Additionally, to generate realistic perturbations for GRACE-based TWS, PyGLDA requires the full error covariance matrix of the L2 products. However, this information is not included in the official L2 products. To overcome this limitation, PyGLDA uses ITSG-2018 temporal gravity field solutions (Kvas et al., 2019; Mayer-Gürr et al., 2018), which offer key advantages: (i) access to the full error variance-covariance matrix; and (ii) better quality than official products due to the additional background model uncertainty considered [more details on the reason why ITSG is selected are given] (Kvas and Mayer-Gürr, 2019; Meyer et al., 2019).

3   Model description

3.1   Structure

PyGLDA consists of three core scientific modules: (1) GRACE processing, (2) W3RA model, and (3) DA, each of which is described in detail below. Additionally, PyGLDA includes three supporting modules [slightly reformulated]: (4) a postprocessing
155   module for data analysis and visualization, (5) a flow control module that integrates the necessary sub-modules into a streamlined processing chain for specific studies, and (6) an auxiliary module for processes outside DA, e.g., downloading forcing fields and formatting output. Figure 1 provides an overview of the structure of PyGLDA, illustrating the interaction between these modules.

GRACE processing. This module is responsible for *conversion*, *correction* and *propagation* of GRACE level-2 data and
160   its associated uncertainty. By default, GRACE level-2 products are distributed as geopotential spherical harmonic coefficients (SHCs), which must be transformed into physically meaningful variables, such as gridded TWS anomalies, for use in DA. This transformation, known as *conversion*, is performed using the harmonic synthesis approach (Wahr et al., 1998). Furthermore, GRACE data contain considerable noise due to imperfect background models [more clarifications are added as suggested] (Yang et al., 2021; Zhou et al., 2023) and instrument noise (Yang et al., 2018; Patel, 2020; Yang et al., 2022a). Therefore, a
165   series of *corrections* is applied to ensure the extraction of a desired signal, including low-degree coefficient replacement (e.g., Loomis et al., 2020), de-correlation and filtering (e.g., Swenson and Wahr, 2006; Kusche, 2007; Klees et al., 2008; Horvath et al., 2018), spatial leakage correction (e.g., Chen et al., 2015; Vishwakarma et al., 2018), Glacial Isostatic Adjustment (GIA) removal (e.g., Richard Peltier et al., 2018). PyGLDA selects and implements the most widely used *correction* strategies, see Table 1. However, be aware that Table 1 is not exhaustive; many more corrections such as the ellipsoidal correction (e.g.,
170   Yang et al., 2022b) are also implemented. Furthermore, to ensure an optimal balance between observational and model-derived water storage estimates, the GRACE-TWS uncertainty must be properly quantified. PyGLDA achieves this through Monte Carlo *propagation* of the full error covariance matrix from the SHC domain to the gridded (or basin-averaged) TWS domain (Wahr et al., 2006; Boergens et al., 2022; Yang et al., 2024a). Be aware that, due to the fact that filtering introduces signal covariance, the ultimate covariance of GRACE reflects both noise and signal's behaviors [clarification is added as suggested]. Given
175   its independent functionality, all GRACE processing related tasks have been integrated into a standalone toolkit called SaGEA, which is also Python-based and open source [added to indicate where to see a detailed description and explanation of GRACE-related terminology as suggested by reviewer] (Liu et al., 2025). SaGEA can be seamlessly coupled to PyGLDA.

W3RA GHM. W3RA is an open-source and global realization of the AWRA-L hydrological model (Australian Water Resources Assessment System, version 0.5) (van Dijk, 2010). It is a grid-based one-dimensional water balance model that rep-
180   resents water storage and flows from soil, groundwater, and surface water. The model employs well-established equations to describe radiation, energy, and water balance processes [reformulated as suggested], requiring minimal meteorological input, including daily gridded precipitation, incoming shortwave radiation, and daytime temperature (typically estimated from minimum and maximum temperature). Although its publicly available version is implemented at $0.5°$ resolution, we have adapted W3RA to $0.1°$ resolution to align with the available high resolution meteorological data sets (Mehrnegar et al., 2021). However, a limi-
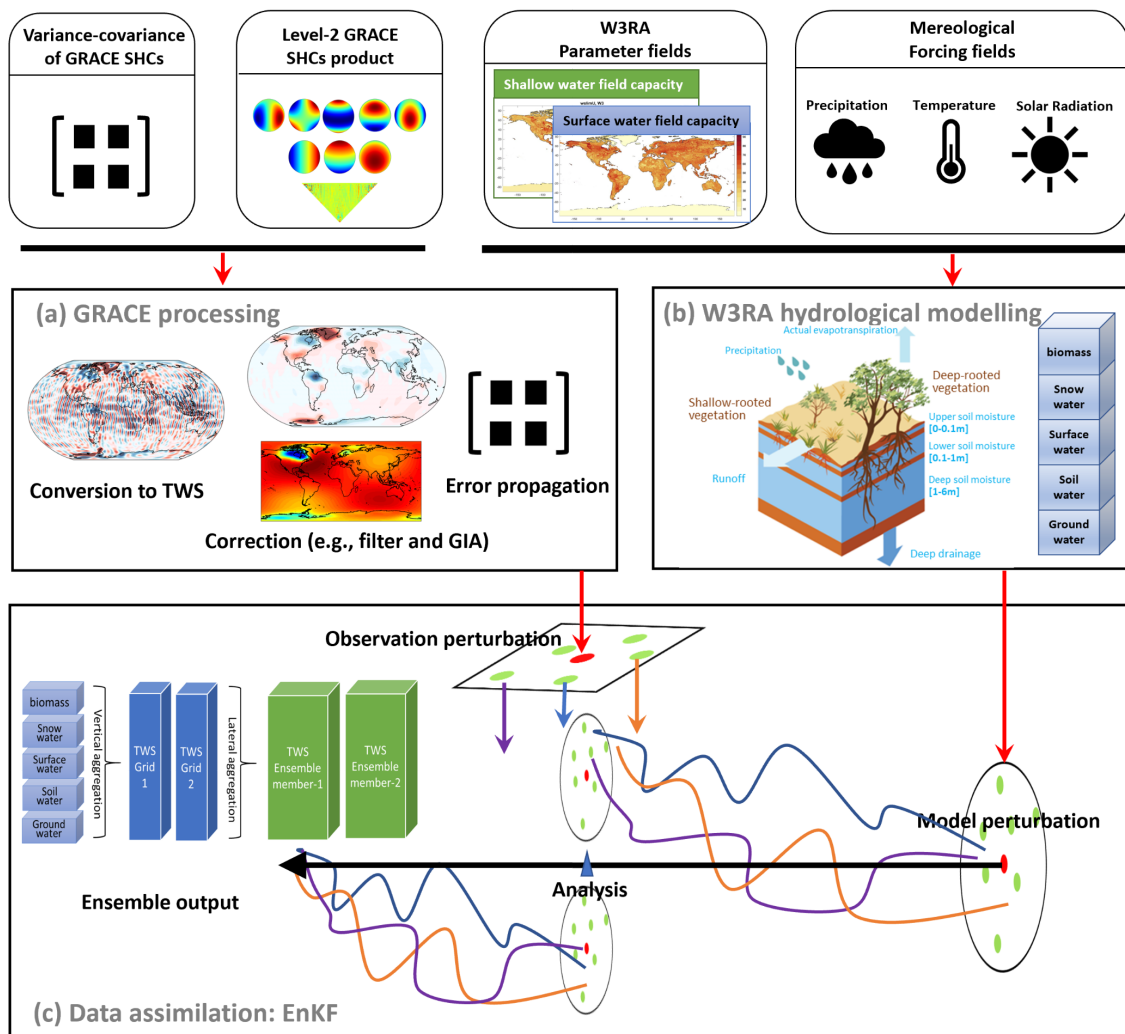
Figure 1. [The figure has been updated to improve the visualization as suggested by reviewer]An overview of PyGLDA structure that mainly consists of three modules: (a) GRACE processing, (b) W3RA GHM (van Dijk, 2010) and (c) Data Assimilation (DA).

185 tation of the current model version is its simplified assumption of lateral water redistribution [revised to avoid confusion as suggested], which reduces reliability in areas influenced by irrigation, inundation, or groundwater inflows (van Dijk, 2010). This limitation highlights the importance of incorporating GRACE-TWS observations via data assimilation. The original W3RA model was released as an open-source MATLAB script, but it has been fully translated into Python and integrated into PyGLDA. ~~Original W3RA was released in an open-source Matlab script, but it has been translated to Python and integrated into PyGLDA, where intensive code validations across~~

190 ~~various combinations of areas/resolution/forcing have been made to verify the same precision and comparable numerical efficiency between before/after trans-~~ ~~lation (based on 0.5 degree solution). The TWS simulations of W3RA is compared against GRACE TWS and with other commonly used GHMs, which were~~ ~~forced by the same climate data within the eartH2Observe project. Their comparison indicated that the skills of W3RA in resembling the global water storage~~

Table 1. An overview of the available *correction* options in PyGLDA's GRACE processing toolkit, see Liu et al. (2025) for more details [add reference].

Correction	Options	Remark	Reference
Low degrees replacement	Degree-1	$[C_{1,0}, C_{1,1}, S_{1,1}]$ by model	Swenson et al. (2008)
	$C_{2,0}$	by TN-11[1] file as of December 2019.	Loomis et al. (2020)
	$C_{3,0}$	by TN-14	Loomis et al. (2020)
De-correlation	Sliding window fitting	Empirical selection of window length	Swenson and Wahr (2006)
	PnMm	A polynomial fitting for degree n and order m	Chen et al. (2006)
Averaging filter	DDK[2]	DDK1 - DDK8	Kusche (2007)
	Gaussian	with a radius r, $200 < r < 1000$ (km)	Jekeli (1981)
	Non-isotropic Gaussian	r_1-radius of degree, r_2-radius of order, m_0-truncated order	Han et al. (2005)
	Fan	r_1-radius of degree, r_2-radius of order	Zhang et al. (2009)
Leakage correction	Forward modeling	Iteration times n, acceleration factor s	Chen et al. (2015)
	Classical[3]	$\hat{f}_c = \overline{f}_c - l_c$	Wahr et al. (1998)
	Additive[4]	$\hat{f}_c = \overline{f}_c - l_c^m + b_c^m$	Klees et al. (2007)
	Multiplicative[5]	$\hat{f}_c = k(\overline{f}_c - l_c^m)$	Longuevergne et al. (2010)
	Scaling	$\hat{f}_c = k^m \overline{f}_c$, basinal or gridded	Landerer and Swenson (2012)
	Data-driven[6]	$\hat{f}_c = \overline{f}_c - \overline{\delta F}_c - l_c$	Vishwakarma et al. (2018)
GIA correction	ICE6G-D (ICE series)	Removal in spectral domain or in spatial domain	Richard Peltier et al. (2018)
	Geruo2013		A et al. (2013)
	Caron2018		Caron et al. (2018)

[1] The Technical Note (TN-XX) files are provided along with the official GRACE products.

[2] DDK is actually a de-correlation and averaging filter (Kusche et al., 2009).

[3] where $\hat{f}_c$, $\overline{f}_c$ denote the signal after and before correction; and l_c indicates the spatial 'leakage'.

[4] where b_c indicates the 'bias'; and the superscript m indicates that a hydrological model is required.

[5] where k is the derived scaling factor.

[6] where $\overline{\delta F}_c$ indicates the deviation integral.

~~variability are similar to the existing models.~~Extensive validation tests across different study areas and forcing datasets confirm that the Python implementation maintains numerical precision and computational efficiency comparable to the original MATLAB

version. Furthermore, previous studies have demonstrated the capability of W3RA to capture global water storage variability, making it a suitable choice for integration into PyGLDA (Schellekens et al., 2017; Mehrnegar et al., 2020).

Data Assimilation (DA). This module consists of four main steps, which are *perturbation*, *initialization*, *forecast*, and *analysis*. Since Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKS) are selected as our merger in PyGLDA, proper *perturbation* of both model and observations ~~has to be performed~~is required. In PyGLDA, model perturbations — specifically for parameters and forcing fields — can be generated using additive or multiplicative noise, following a Gaussian or triangular distribution (Schumacher, 2016). In contrast, perturbations for GRACE-TWS observations are generated from its full error [added as suggested] covariance matrix of the GRACE data. Be aware that the *perturbation* of the initial field (state vector) is currently not considered in PyGLDA due to its limited impact on the DA of land water (Reichle and Koster, 2003). ~~n the next step, an *initialization*, i.e., a spin-up of the error-free W3RA GHM (given, for example, two years at least) to acquire a reliable initial field. Subsequently, in the *forecast* step, the state of the model is continuously evolving on a daily basis until being informed to have an observation available.~~The *initialization* step involves a spin-up phase, where the W3RA model is run for at least two years without perturbations, ensuring a stable and physically meaningful initial state. In the forecast step, the model evolves forward on a daily timescale until the next observation becomes available. Finally, in the analysis step, observational data and model forecasts are combined to update the state variables for the past month, following the standard EnKF or EnKS framework; see Appendix A for details.

3.2   DA workflow and parallel computation

[this paragraph has been reformulated with the emphasis on the conceptual distinction of ensemble and ensemble members]State-of-the-art sequential DA methods are widely used in hydrological applications, with Kalman filter-based approaches gaining significant attention due to their relatively ~~straight forward~~straightforward matrix implementation (Schumacher, 2016). The classical Kalman filter assumes that all probability distributions involved in the system are Gaussian and provides algebraic formulations for updating the mean and covariance matrices under the assumption of a linear system. However, maintaining a full covariance matrix is computationally impractical for high-dimensional systems. To address this problem, the EnKF was developed, representing the probability distribution of the system using an ensemble of state vectors rather than an explicitly stored covariance matrix. This approach enables a Monte Carlo approximation of the Bayesian update, where the accuracy of the covariance estimate depends on the number of ensemble members. Since there is no universally optimal ensemble size, PyGLDA allows users to customize the number of ensemble members through a JSON-based configuration file; see Appendix B for details. We recommend an ensemble of 30 members to balance computational efficiency and accuracy [added as suggested](Schumacher et al., 2016, 2018a).

In PyGLDA, the processing steps (including perturbation, model forecast, state analysis, and data management) for each ensemble member are completely independent of the others [reformulated to avoid confusion, as suggested]. This independence makes PyGLDA highly suitable for parallel computing, allowing full utilization of CPU resources. In PyGLDA, each ensemble member is assigned to a separate computational thread, significantly improving efficiency. Parallelization is implemented using Message Passing Interface (MPI), a widely adopted standard for parallel computing. Specifically, PyGLDA leverages the MPI4PY package (https://mpi4py.readthedocs.io/en/stable/), which provides MPI bindings for Python, enabling multi-core and multi-node execution on High-Performance Computing (HPC) clusters.
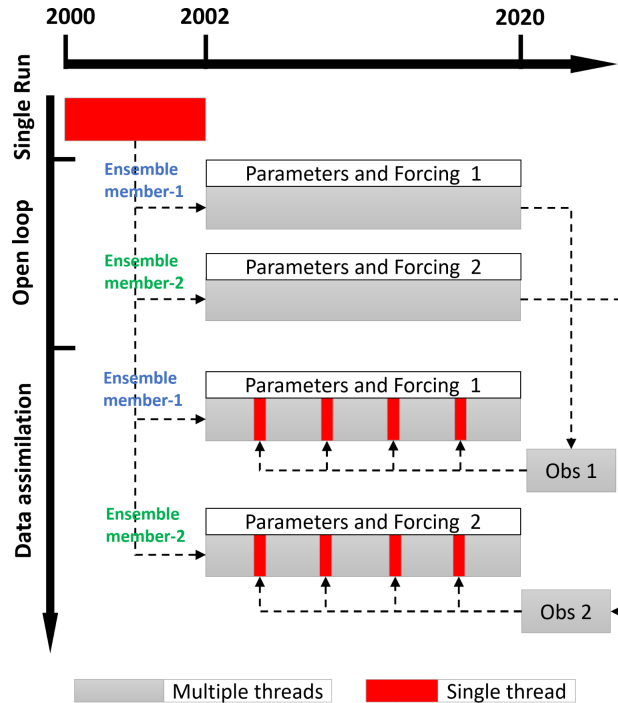
Figure 2. [This figure is updated by changing the 'ensemble' to 'ensemble member' in response to reviewer's comment]A schematic diagram to manage the threading pool of the DA workflow in PyGLDA, where an ensemble of two members for 2002-2020 is shown as an example.

Figure 2 provides a schematic overview of how parallel DA ~~works~~is managed in PyGLDA~~, where an ensemble of two members has~~
230 ~~been shown as an example. In practice, one use more ensemble members, for example 30~~. The general DA workflow, as defined in PyGLDA, is divided into three primary stages [the description of each stage is slightly changed]: (1) a *single-run*, i.e., model spin-up, (2) an *open loop* run, that is, ensemble-based forecasts with perturbed input, and (3) the *data assimilation* run, that is, merging ensemble forecasts with GRACE observations. In the *single-run* stage, the hydrological model is initialized and spun up for at least two years to ensure the equilibrium of the system, with only one single thread required. In the *open loop* stage, all members of
235 the ensemble are run simultaneously, each initialized from the spin-up state and driven by perturbed forcing fields and model parameters. ~~Subsequently, for each ensemble member, the long-term mean of the state vector is calculated and subtracted from the GRACE observa-~~ ~~tion.~~The aim of this stage is to obtain a long-term mean of the states and subtract it from the GRACE observations. Finally, in the *assimilation* stage, multiple treads are launched again, whereas only one of them is meant to deal with the Kalman update when new observation becomes available. In summary, a complete DA job will go through these three stages to arrive at the
240 final assimilated results, where PyGLDA automatically manages the threading pool to distribute the appropriate ~~threads~~resources for each stage.

3.3 GLDA configuration

3.3.1 Patch-wise domain localization

[slightly reformulated]Running a high-resolution GHM with an ensemble-based DA system presents significant computational
challenges, particularly when the ensemble size exceeds a few dozen members. Computational constraints and numerical
stability issues become critical concerns, especially in large-scale Global Land Data Assimilation (GLDA) applications (Re-
ichle and Koster, 2003). Given a limited ensemble size, spurious long-range correlations are easily introduced (Forman and
Reichle, 2013; Khaki et al., 2017). These artificial correlations often degrade the DA performance by causing numerical in-
stabilities during the Kalman update step, potentially leading to filter divergence. Consequently, compared to regional DA,
traditional GLDA implementations face additional challenges related to computational feasibility and numerical robustness,
making GLDA significantly more complex and less frequently attempted in previous studies.
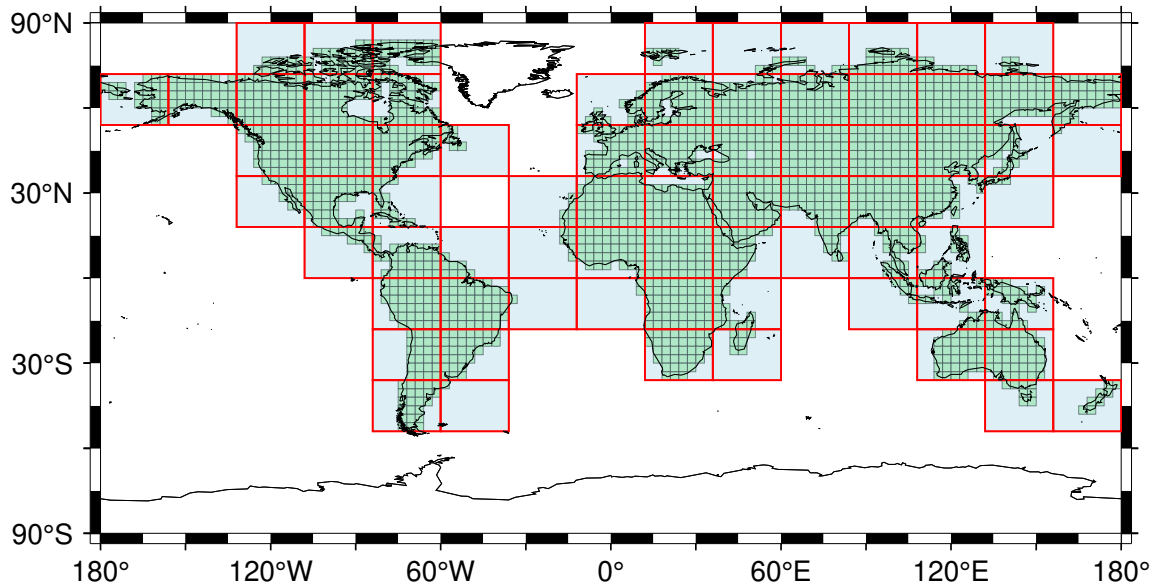
Figure 3. Configuration of the patch-wise domain localization, where each tile (denoted as the rectangle filled by the light blue) represents
the 'domain' to perform DA, individually. The unit box in black, contained in each tile, represents the smallest area to average GRACE
observations to be assimilated.

[slightly reformulated]To address these challenges, it is beneficial to restrict the spatial extent to which covariance matrices are
computed. For global DA, it would be effective to divide the global domain into smaller localized patches. This technique,
known as domain localization, helps stabilize the covariance matrix by assuming that significant correlations do not extend
beyond a certain region. By enforcing this assumption, PyGLDA reduces the impact of spurious long-range correlations,
improving numerical stability. Additionally, if both the hydrological model and GRACE-TWS observations exhibit only weak
cross-domain correlations, DA computations within each patch can be treated as independent processes, making the GLDA

problem computationally tractable. Therefore, in PyGLDA, the global domain is partitioned into patches (tiles), each operating as an independent DA unit. The patch size is flexible, but a recommended configuration for GRACE-based global DA is $18° \times 24°$, as illustrated in Figure 3. Each tile contains a set of grid cells, $3° \times 3°$, which represent the smallest units used to aggregate GRACE observations. For future gravity missions with improved spatial resolution, such as MAGIC and NGGM (Daras et al., 2023), smaller grid sizes can be used accordingly.

Additionally, in PyGLDA, the domain size and configuration can be fully customized, allowing users to define different patch sizes depending on their specific application. The default setup excludes regions with no land coverage (such as oceans and ice sheets), leaving a total of 74 valid computational domains for DA [reformulated as suggested], see Figure 3. This flexible approach ensures that PyGLDA can be efficiently adapted to various spatial resolutions and study objectives. For example, users can modify the unit size (for example, using grid cells of $4°$ or $5°$) or redefine domain boundaries based on a specific land mask. By employing this patch-wise DA strategy, PyGLDA effectively mitigates numerical instability issues and reduces computational costs, enabling high-resolution GLDA to be executed efficiently, even on standard computational platforms. Furthermore, this approach facilitates parallel computing, allowing multiple domain patches to be processed simultaneously, significantly accelerating GLDA execution [reformulated].

3.3.2 Weighted global aggregation

In practice, ~~the hypothesis of independence between adjacent domain/tiles is a strong assumption~~the assumption of complete independence between adjacent domains (patches) is a simplification. Although it is reasonable to assume that correlations between the center of a patch and locations outside the patch are negligible—provided that the patch is sufficiently large—the same assumption does not hold for areas near the boundary of the patch (Reichle and Koster, 2003). Since PyGLDA performs DA independently within each patch, ignoring cross-boundary correlations can result in artificial discontinuities at the edges of neighboring patches. This issue, commonly referred to as the edge effect, manifests itself as visible artifacts in the global DA results, where abrupt transitions appear between adjacent patches. Therefore, to ensure smooth transitions across patch boundaries, PyGLDA introduces a weighted global aggregation strategy that effectively blends the DA results of neighboring patches [reformulated].

[Intensive revisions have been made to this paragraph to address reviewers' comments: (1) language (2) mathematical formulation, and (3) more justification]To mitigate the edge effect, PyGLDA extends the computational domain of each patch by introducing overlapping regions with adjacent patches. These overlapping transition zones allow the DA results from neighboring patches to be smoothly blended, reducing artificial discontinuities. Specifically, each patch is extended by a predefined margin (for example $6°$ beyond its original boundaries, as shown in Figure 4a), ensuring that DA calculations consider interactions between patch borders. After DA is performed within each extended patch, the overlapping regions are merged using a weighted averaging approach, see Figure 4b-c. In this approach, each overlapping grid cell receives contributions from multiple patches, with weights assigned based on the distance from its outermost border. For example, $f^a(x) = \sum_{i=1}^{2} w^i(|x - b^i|) f^i(x)$, where f^a denotes the averaged results of two patches (f^1 and f^2), x denotes a certain point within the overlap area, b^i denotes the boundary of the patch i^{th}, and w indicates the weighted function related to the distance $|x - b^i|$. The inverse distance weighting method used is based on the principle that correlations generally decay with increasing spatial separation (Reichle and Koster, 2003). The
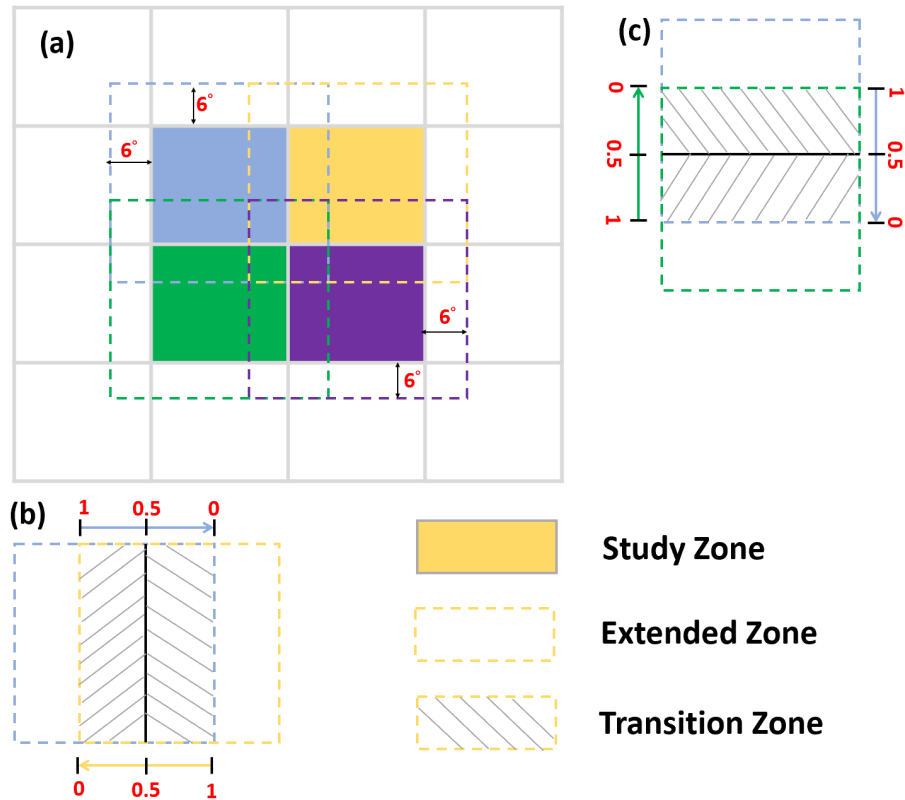
Figure 4. [The figure has been updated to change the layout as suggested]An illustration of the proposed global aggregation strategy. In Figure 4(a), the box in grey replicates the tile/patch of Figure 3. Four adjacent tiles (in blue/green/yellow/purple) are selected as examples to show the extension of the zone (in dashed line, e.g., $6°$). A zoom-in is demonstrated in Figure 4(b-c), where the shaded area indicates the overlap of two extended tiles, and the red number indicates the weight assigned to each tile for the averaging over the overlapping areas.

extension of $6°$ is considered based on our previous study on the length of GRACE's correlation [justification of our choice in response to reviewer's comment] (Yang et al., 2024a). This ensures a gradual transition between neighboring DA solutions, preserving spatial continuity.

4 Application

4.1 Basic-scale DA for Danube

~~The Danube River Basin (DRB) is Europe's second largest river basin, with a total area of 801,463 km². It is the world's most international river basin, flowing through the territory of 19 countries.~~The Danube River Basin (DRB) is the second largest river basin in Europe, covering an area of approximately 801,463 km² and spanning 19 countries. Given its extensive transboundary nature, accurately monitoring terrestrial water storage changes within the DRB is crucial for water resource management, climate studies, and hydrological

forecasting. To demonstrate the capability of PyGLDA in a regional DA context, we selected DRB as a case study. In this experiment, we performed basin-scale DA, a widely adopted approach in previous studies. However, in the next section, we extend this experiment to a grid-scale DA setup over the DRB to evaluate the added value of higher spatial resolution.
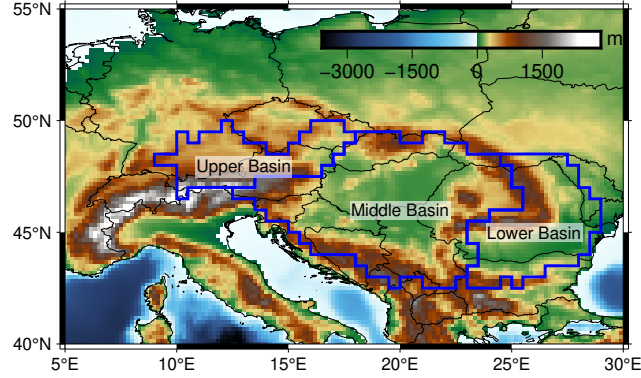
Figure 5. The Danube River Basin (DRB) and its three major subbasins. The basin's topography is shown as background picture.

[reformulated]To implement the basin-scale DA, the boundaries of the Danube River Basin (DRB) and its sub-basins must
305 be properly defined. Conventionally, the DRB is divided into three main subregions based on elevation and hydrological characteristics (Stagl and Hattermann, 2015): (1) the Upper Basin (UB), extending from the source of the Danube in Germany to Bratislava, Slovakia; (2) the Middle Basin (MB), the largest of the three subregions, stretching from Bratislava to the Iron Gate Gorge, a dammed section on the Serbia-Romania border; and (3) the Lower Basin (LB), covering the lowlands, plateaus and mountains of Romania and Bulgaria. Figure 5 illustrates the DRB and its sub-basin divisions, where the sub-basins serve
310 as the smallest spatial units to average GRACE observations before assimilating into the model.

Table 2. GRACE processing choice for the case study of DRB, see Table 1 for more details.

Step	Option
Low-degree replacement	degree-1 terms and $C_{2,0}, C_{3,0}$
Filtering	DDK3
GIA correction	ICE6G-D
Mean removal	temporal mean of 2002-2010

[slightly modified]The GRACE processing module is first executed to generate the observations required for assimilation. The specific GRACE processing configurations used in this study are summarized in Table 2. Then, the full propagation of variance-covariance from the harmonic domain to the gridded domain (that is, the basin-average TWS) is performed following the method of Yang et al. (2024a). Unlike many previous studies, which assume zero correlation between sub-basins, our analysis

315 indicates that significant correlations exist, which cannot be ignored. For example, the correlation coefficient between UB and LB reaches 0.50 in January 2009, highlighting the importance of accounting for spatial correlations in DA. Furthermore, the temporal variability of the variance-covariance structure is considered, since changes in the orbital height of the GRACE mission affect the error characteristics [reformulated to avoid potential confusion] (Boergens et al., 2022).
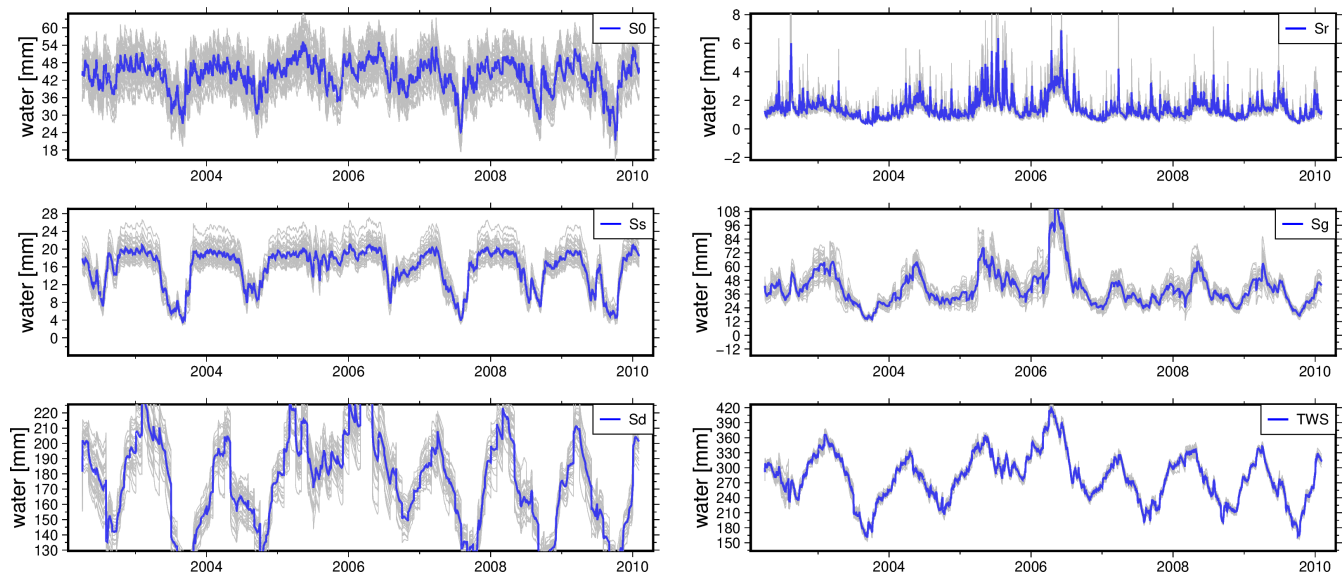
Figure 6. An overview of the mean (in blue) of the ensemble (shaded) estimates after Basin-scale DA experiment over DRB is illustrated. Plots include the TWS estimate and its major vertical compartments: surface water (Sr), topsoil water ($S0$), shallow soil water (Ss), deep soil water (Sd), and groundwater (Sg) [added as suggested].

Previous studies suggest that an ensemble size of 30 provides a reasonable balance between computational efficiency and
320 required data storage capacity on the one hand and representative error statistics on the other hand [reformulated as suggested] (Schumacher et al., 2018b). It is also adopted for this case study. It is important to note that the number of MPI threads initialized must match the defined ensemble size; otherwise, an error will occur due to insufficient computational threads to handle all ensemble members. In this study, the ensemble members are generated by perturbing both the meteorological forcing fields and the model parameters. Specifically, random Monte Carlo sampling is applied using a triangular probability
325 density function, introducing a multiplicative perturbation 30% to these fields. The current version of PyGLDA assumes that perturbations are spatially correlated (homogeneous) but temporally uncorrelated (independently), though future versions will allow users to define arbitrary spatial-temporal correlation structures for generating perturbations.

~~Following the strategy described before, we perform the DA over DRB from years 2002 to 2010 and collect the ensemble. Therefore, by convention, the average of ensembles can be assumed to be the assimilated result, and moreover, the spread of the ensemble should represent the uncertainty, while it is~~
330 ~~suspicious how reliable and to what extent the realistic uncertainty can be represented~~Following the outlined strategy, we perform DA in the DRB for the period 2002–2010. Conventionally, the ensemble mean is considered the assimilated TWS estimate, while the

ensemble spread serves as an approximation of uncertainty. However, the extent to which this spread of the ensemble accurately represents the true uncertainty remains an open question and requires further validation (Gerdener et al., 2023). Figure 6 illustrates the temporal variability of TWS and its major vertical components, providing insights into how DA redistributes the

335   GRACE-based TWS increments across different water storage compartments. Among the five modeled components—surface water (*Sr*), topsoil water (*S0*), shallow soil water (*Ss*), deep soil water (*Sd*), and groundwater (*Sg*)—the largest DA-induced changes occur in groundwater and deep soil water, while surface and topsoil water capture mainly high frequency variations from GRACE. This finding is consistent with previous studies that indicate that GRACE-based TWS anomalies primarily constrain slow-varying subsurface water storage components rather than short-term surface fluxes [reformulated as suggested to avoid

340   confusion].
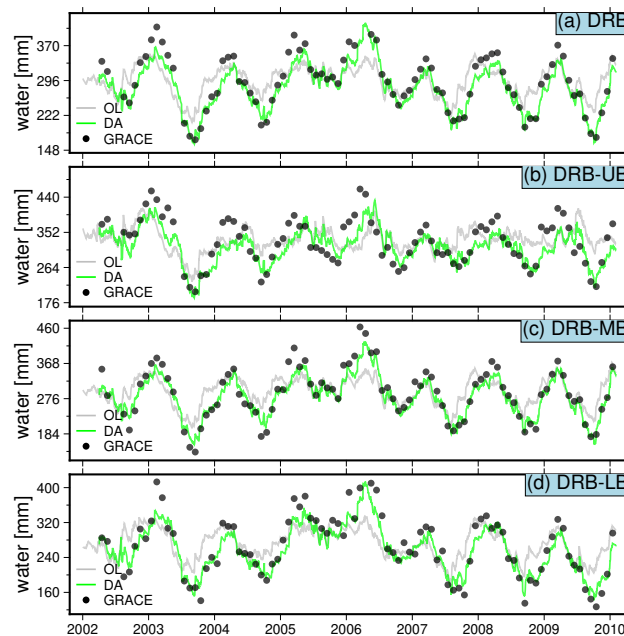
Figure 7. An overview of the basin averaged TWS of DRB and its three sub-basins (LB, MB, and UB) for OL, DA and GRACE observations.

Furthermore, Figure 7 presents the basin-averaged TWS time series for the entire DRB and its three sub-basins (UB, MB, and LB), comparing the results from the open-loop (OL) simulation, the DA run, and GRACE observations. A comparison between GRACE (black dots) and OL (shaded line) reveals that while the model preserves more short-term dynamics, GRACE exhibits larger magnitude variations, highlighting the need for DA to optimally merge both sources of information. As expected, the DA

345   result (green line) adjusts the magnitude of the TWS closer to GRACE while retaining the temporal variability of the model. In particular, in Figure 7b, an evident phase shift exists between OL and GRACE, probably due to a delay in the streamflow response. For example, across the entire DRB, the annual signal phase for OL, GRACE, and DA is found to be -41°, 29°, and 11°. Therefore, the DA process successfully reduces this phase mismatch [statistic added as suggested], although further validation is required using in situ observations to confirm the accuracy of this correction.
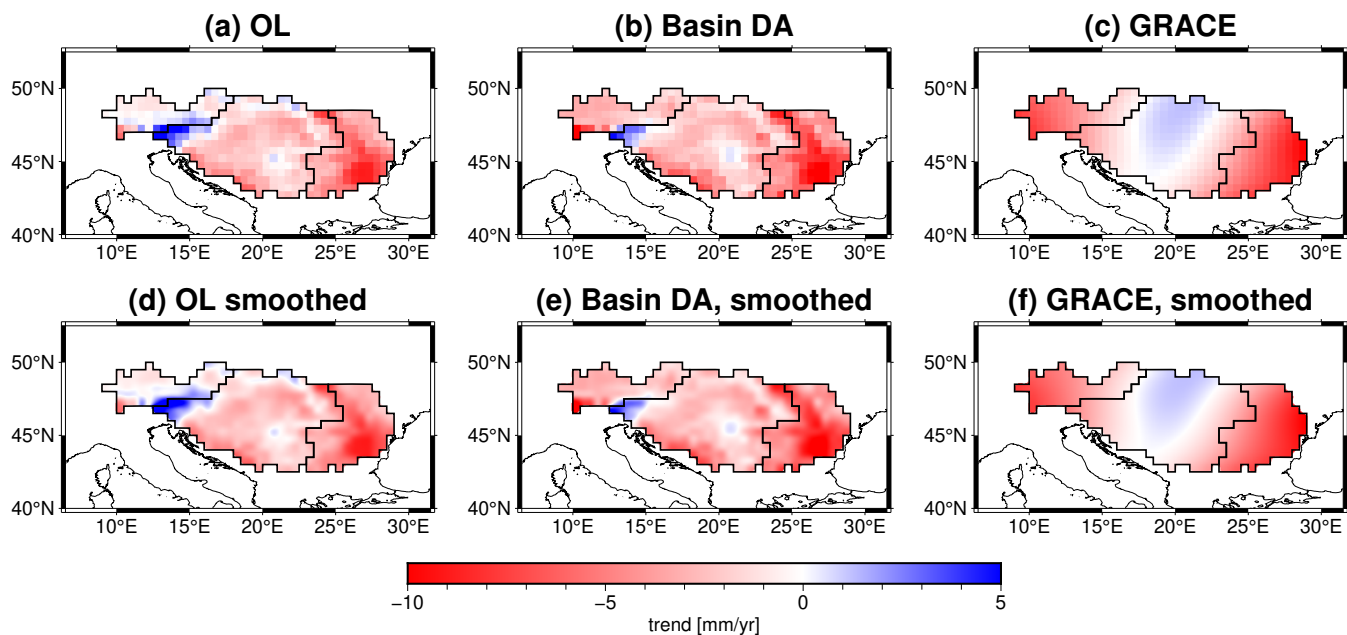
Figure 8. [the figure is updated by reversing the colorbar as suggested]The secular trend of TWS over 2002-2010 for DRB, where the OL, DA and GRACE results are listed at (a-c), and a smoothed version is respectively shown at (d-f) for a better visualization.

350   To further analyze the spatial characteristics of the assimilation results, we extract the long-term trend and seasonal amplitude [modified as suggested] of the TWS estimates based on OL, DA and GRACE and present them as 2D spatial maps in Figures 8 and 9. The original spatial resolution of the OL and DA outputs is $0.1°$, but for comparison with GRACE, an up-scaling procedure is applied, aggregating the data to the resolution $0.5°$. Furthermore, a 2D spline smoothing technique is applied to improve visualization clarity, as shown in Figures 8d–f and 9d–f. The trend analysis in Figure 8 reveals that while DA largely

355   preserves the spatial pattern of the OL model, it brings the upper Danube Basin (UB) trends closer to the GRACE observations. Quantitative comparisons indicate a spatial correlation of 0.74, 0.98, and 0.89 between DA and OL in the UB, MB, and LB sub-basins, respectively. However, the correlation between DA and GRACE is lower (0.79, 0.35, and 0.75), particularly in the MB region, suggesting that DA is more dependent on the model when the discrepancies between GRACE and the model are large. Likewise, by looking at the annual amplitude in Figure 9, one can see that DA perfectly replicates the spatial pattern

360   of OL, with a spatial correlation coefficient up to 0.92; but meanwhile, its spatial magnitude has been tuned much (almost twice larger than OL) toward GRACE observations. In this sense, PyGLDA has fulfilled its goal of balancing the model and observations to reach the statistically optimal combination.

4.2   Grid-scale DA within DRB

In recent years, there has been a growing interest in DA on a grid scale using GRACE gridded data products, such as the

365   $3° \times 3°$ mascon solutions from JPL. Thus, we have added this possibility to PyGLDA to accept Mascon solutions directly, for
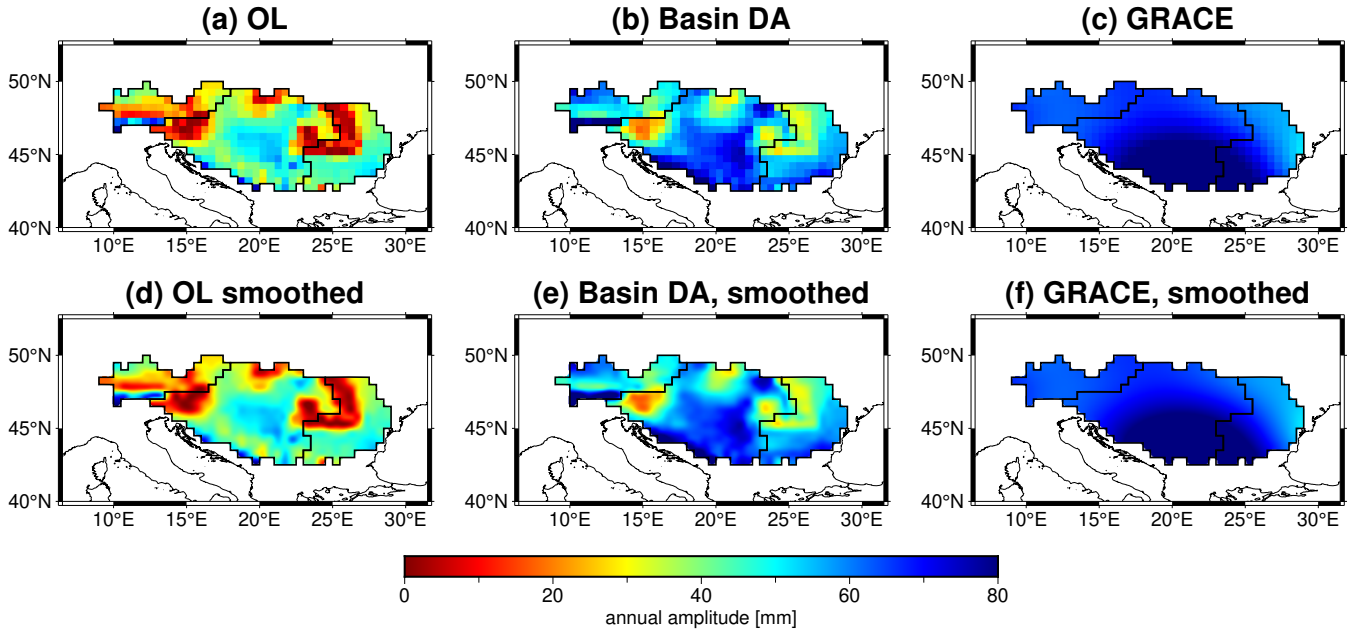
Figure 9. An overview of the annual amplitude of TWS within DRB over 2002-2010, where the OL, DA and GRACE results are shown in (a-c), and a smoothed version is respectively shown at (d-f) for a better visualization.

example, the CSR mascon solutions (Save et al., 2016) are integrated in this study. Compared to basin-scale DA, grid-scale DA allows for a finer spatial representation of GRACE-TWS signals, potentially improving the assimilation of satellite gravity data into high-resolution hydrological models (Girotto et al., 2016; Khaki et al., 2017; Li et al., 2019; Gerdener et al., 2023). In this section, we extend our previous basin-scale DA experiment over the DRB by implementing a grid-scale DA approach.

370 It is important to note that PyGLDA does not require structural modifications or particular operations to perform grid-scale DA; users simply need to define the target region using a shapefile, regardless of whether the intended DA is at the basin or grid scale. For a grid-scale DA, the shapefile defines each grid cell as the sub-basin. This feature highlights the flexibility of PyGLDA, which allows seamless transitions between different spatial scales. [reformulated for a better readability]
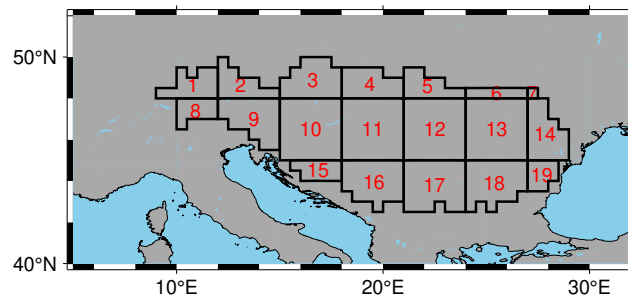
Figure 10. An overview of the $3° \times 3°$ cell definition for DRB, where the identification number (ID) of each cell is highlighted in red.

[slightly modified]For this experiment, we adopt a widely used grid resolution of $3° \times 3°$ for GRACE observations, ensuring consistency with previous grid-scale DA studies. The grid cells are defined to fully cover the DRB, as illustrated in Figure 10, where each numbered cell represents an individual assimilation unit. However, be aware that a uniform grid structure cannot be ensured, and one can see from Figure 10 that some cells are irregular in shape (for example, cell 7), with their boundaries intersecting the edges of the basin. In total, the DRB is divided into 19 computational cells, each serving as an independent observation region for GRACE-TWS averaging and assimilation. Although the irregularity of some cells does not significantly affect the assimilation process, optimizing the grid generation approach for basins with complex geometries remains a potential area for future improvement.
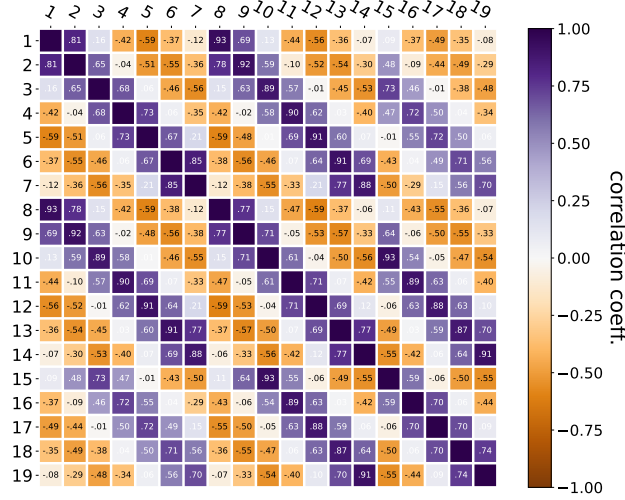
Figure 11. Spatial correlation coefficients of GRACE observations for cells within DRB. The label of X(Y) axis denotes the cell ID as indicated by Figure 10.

[slightly modified]Following the grid definition in Figure 10, we calculate GRACE-TWS anomalies for each $3° \times 3°$ cell by averaging the original $0.5°$ gridded GRACE data within the limits of each cell. In addition, we investigated the spatial correlation structure of GRACE observations in these grid cells, as shown in Figure 11. The results reveal that strong spatial correlations exist between many grid cells, with some correlation coefficients exceeding 0.9. Additionally, evident negative correlations appear, likely due to residual correlated noise from imperfect GRACE decorrelation filtering. These findings suggest that assuming uncorrelated observation errors, as done in some traditional DA implementations, may not be entirely realistic. Instead, PyGLDA explicitly incorporates the full variance-covariance matrix of GRACE observations, ensuring that spatial correlations are properly accounted for during the assimilation process.

[slightly modified]Keeping the GRACE processing and perturbation (including model) strategies unchanged from the basin-scale DA experiment, we now perform grid-scale DA over the DRB using PyGLDA. The ensemble-averaged results are presented in Figure 12, where, for brevity, only the basin-averaged TWS (rather than the individual compartment) is shown. A comparison between grid-scale DA (Figure 12) and basin-scale DA (Figure 7a) indicates that at the basin-wide scale, the two
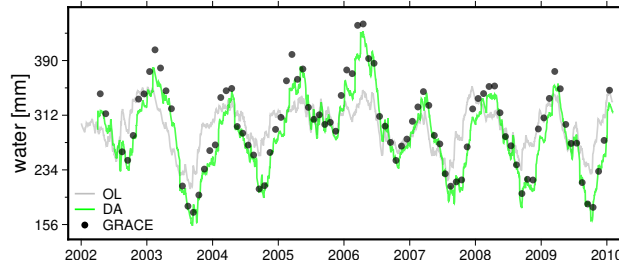
Figure 12. An overview of the basin averaged TWS of DRB for OL, grid-scale DA and from GRACE observations

assimilation strategies produce generally similar results. The temporal correlation between the basin-averaged TWS estimates
395 from the two DA approaches is found to be 0.99, suggesting that, at large spatial scales, grid-scale DA does not significantly
alter basin-wide statistics. However, this does not diminish the importance of grid-scale DA, as its primary advantage likely
lies in capturing finer spatial details, which are further explored in the subsequent 2D spatial analysis. Furthermore, the con-
sistency between basin-scale and grid-scale DA results confirms the correctness and robustness of PyGLDA's grid-scale DA
implementation.
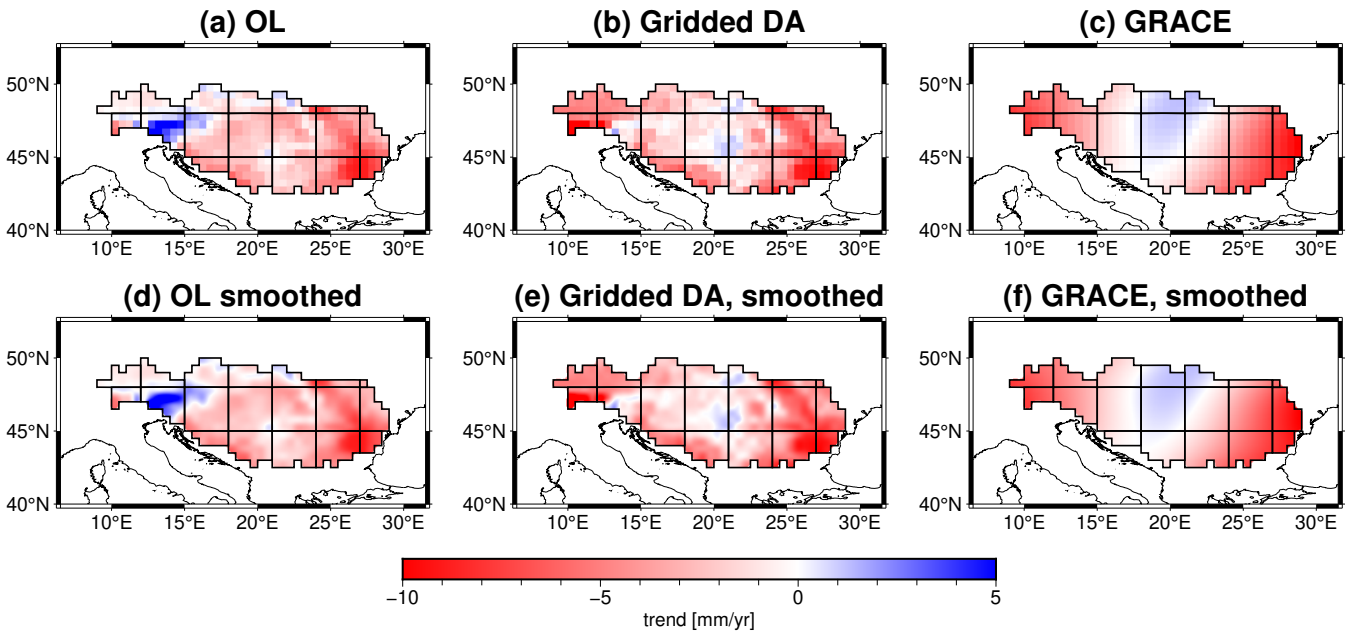
Figure 13. As that of Figure 8 but this is for the Gridded DA. [the figure is updated by reversing the colorbar as suggested]

400   [slightly modified]To further examine the spatial characteristics of grid-scale DA, we compare the long-term trends and seasonal
amplitudes of the OL, DA, and GRACE-TWS estimates at a resolution of $0.5°$, as shown in Figures 13 and 14. Comparing
the DA trend maps in Figure 13e (grid-scale DA) and Figure 8e (basin-scale DA), we observe that grid-scale DA more closely
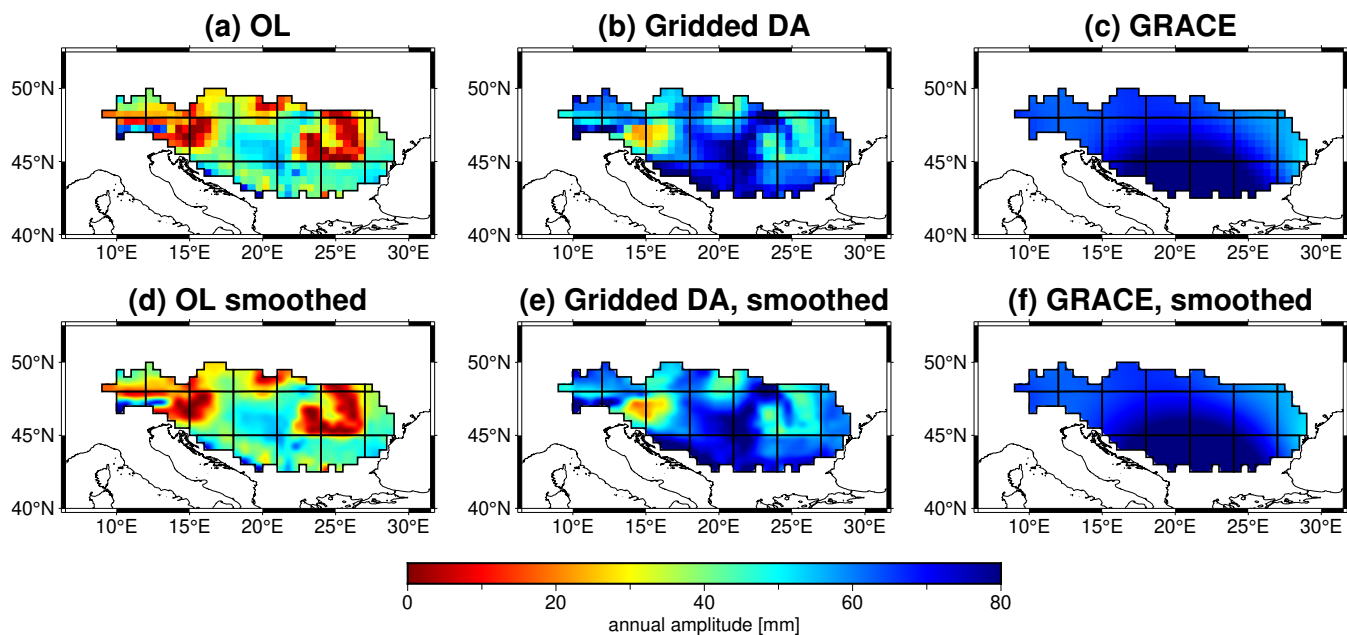
Figure 14. As that of Figure 9 but this is for the Gridded DA. [the figure is updated by reversing the colorbar as suggested]

follows GRACE-TWS trends, particularly in the Middle Basin (MB). For example, the positive trend signal (indicating a water gain) found in the heart of MB ($\sim 20°$E, $\sim 45°$N) appears more evident in Figure 13e than in Figure 8e, and such a

405 positive signal is quite consistent with that of GRACE-TWS. Similarly, in Figure 14, we see that DA on the grid scale better captures the magnitude of the annual amplitude variations observed in GRACE, reducing the Root Mean Squared of Differences (RMSD) between GRACE and DA from 22.1 mm (basin scale) to 16.5 mm (grid scale). Furthermore, the spatial correlation between DA and GRACE improves from 0.45 (basin scale) to 0.51 (grid scale), further demonstrating the added value of grid-scale DA in preserving finer-scale spatial information. These findings suggest that incorporating higher-resolution GRACE-

410 TWS observations into the DA process allows a better spatial representation of water storage changes, particularly in regions where the hydrological model and observations exhibit significant differences. By integrating finer-scale satellite-derived TWS anomalies, grid-scale DA provides a more detailed and accurate spatial distribution of water storage variability compared to basin-scale DA. This is particularly relevant for future satellite gravity missions, such as MAGIC and NGGM (Daras et al., 2023), which are expected to offer higher spatial and temporal resolution with improved error characteristics (Daras et al.,

415 2023). However, while these results highlight the advantages of grid-scale DA, it is important to note that neither approach can be universally considered superior. The optimal choice between basin-scale and grid-scale DA depends on factors such as study objectives, model resolution, and observation uncertainties. Future research should include independent validation using in situ measurements, such as groundwater well data and river discharge records, to further assess the accuracy of these DA strategies.
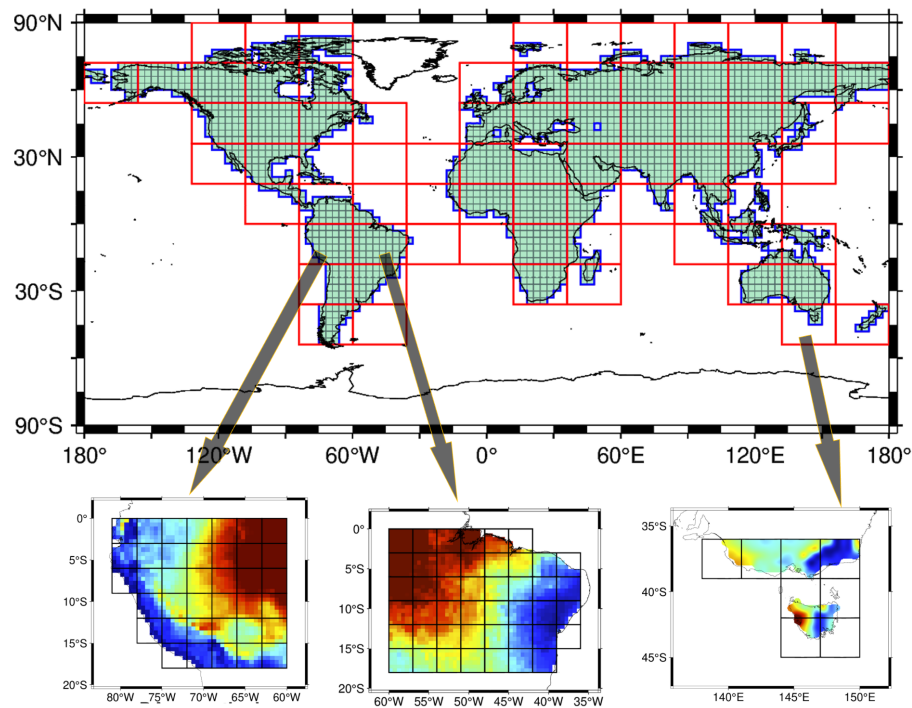
Figure 15. A schematic diagram of the GLDA implementation, where extension of the zone (6° as indicated by Figure 4) has been applied but not shown here for simplifying the visualization.

[slightly modified]In this case study, we performed the GLDA experiment covering 2002-2010 by merging monthly GRACE TWS into daily W3RA globally at high spatial resolution, that is, 0.1° or equivalently 10-12 km, which is likely the highest resolution among similar previous efforts. Here, we follow the patch-wise domain localization and weighted global aggregation as described in Section 3.3 to ensure numerical stability and avoid artificial discontinuities in the DA results. The overall

425 implementation of GLDA is conceptually illustrated in Figure 15, where one can see the global distribution of the patches (red rectangles): the shape of the 'basin' inside each patch (in blue) and the 'sub-basins' inside each 'basin' (black boxes). The global domain is divided into 74 computational patches, each spanning $18° \times 24°$, with overlapping regions (extended by 6°) incorporated to ensure smooth transitions across patch boundaries, see Figure 15. Greenland and Antarctica are excluded because GHMs often do not cover these regions. All the shapefiles related to the aforementioned definitions are shared in the

430 PyGLDA package. The DA is performed independently for each patch, followed by a weighted global aggregation step to merge neighboring DA solutions.

[slightly modified]Furthermore, the detailed ~~roadmap~~execution of global DA (GLDA) in our implementation is outlined in Table 3, summarizing how each sub-module of PyGLDA is utilized and how efficient it is for specific tasks. The workflow is

Table 3. An overview of the numerical efficiency tests for PyGLDA, where the computation is designed from years 2002-2010 with an ensemble size of 30.

Step	Sub-step	Target area	Resources needed[1]	Time (hour per core)
Data pre-process or preparation	GRACE signal process[2]	globe	1-core, up to 10 GB	~ 0.8
	GRACE covariance estimation	globe	1-core, up to 15 GB	~ 3.0
	W3RA data pre-process[3]	1-patch[4]	1-core, up to 6 GB	~ 0.5
Data process	Perturbation	1-patch	1-core, up to 4 GB	~ 0.2
	Spin-up (2-years)	1-patch	1-core, up to 3 GB	~ 0.04
	Open loop	1-patch	30 cores, up to 60 GB	~ 0.3
	DA	1-patch	30 cores, up to 70 GB	~ 0.5
Data post-process	Statistical analysis	1-patch	30 cores, up to 10 GB	~ 0.1
	Global aggregation	globe	1-core, up to 15 GB	~ 1.0
	Visualization	1-patch	1-core, up to 2 GB	~ 0.01

[1] CPU cores and running memory for serial/parallel computations.

[2] This includes conversion from GRACE level-2 to TWS and necessary corrections as indicated by Table 2.

[3] This includes the crop of ERA5-land forcing field (as well as the model parameters and climatology) at intended area, and its upscaling to a desired resolution.

[4] This indicates the performance is evaluated on one patch as defined in Figure 15.

divided into three main stages: (1) data pre-processing, (2) data processing, and (3) data post-processing, each comprising multiple sub-steps. This structured approach facilitates performance evaluation and optimization of computational efficiency. We conducted the GLDA tests on an in-house computing cluster, using a single node equipped with 32 cores (Intel Xeon Gold 613 @1.59 GHz) and 192 GB of RAM. While computational demands vary based on the user's specific setup, our test results provide a baseline reference for PyGLDA's efficiency. From Table 3, our tests indicate that the data preparation is most time-consuming, followed by the data processing and the post-processing. Fortunately, the data preparation only requires to be run once and the results can be saved for repeat use in later stages. In addition, currently the running time is tested for data preparation based on a single thread, which means that a considerable reduction of computation time could be expected through multi-thread parallelization in future update. In contrast, the data processing (responsible for DA) as the core of the GLDA, is super fast and should be affordable even on a personal desktop. For example, the total time of running data processing for a regional DA in DRB (as indicated by Section 4.2) is less than 15 minutes in our platform. For the GLDA in this section, the total time for running the global DA experiment is approximately 5 days at the single node of the cluster. Then, one can actually expect a higher efficiency by making use of multiple nodes. Despite the satisfactory efficiency, PyGLDA has a high demand of the data storage for GLDA. In this case study, the amount of temporary files and output overall has already been up to 15

TB, because of the high temporal (daily), horizontal (0.1°, i.e., 2142930 grid points), and vertical (seven layers) resolution at global scale [clarification added as suggested].
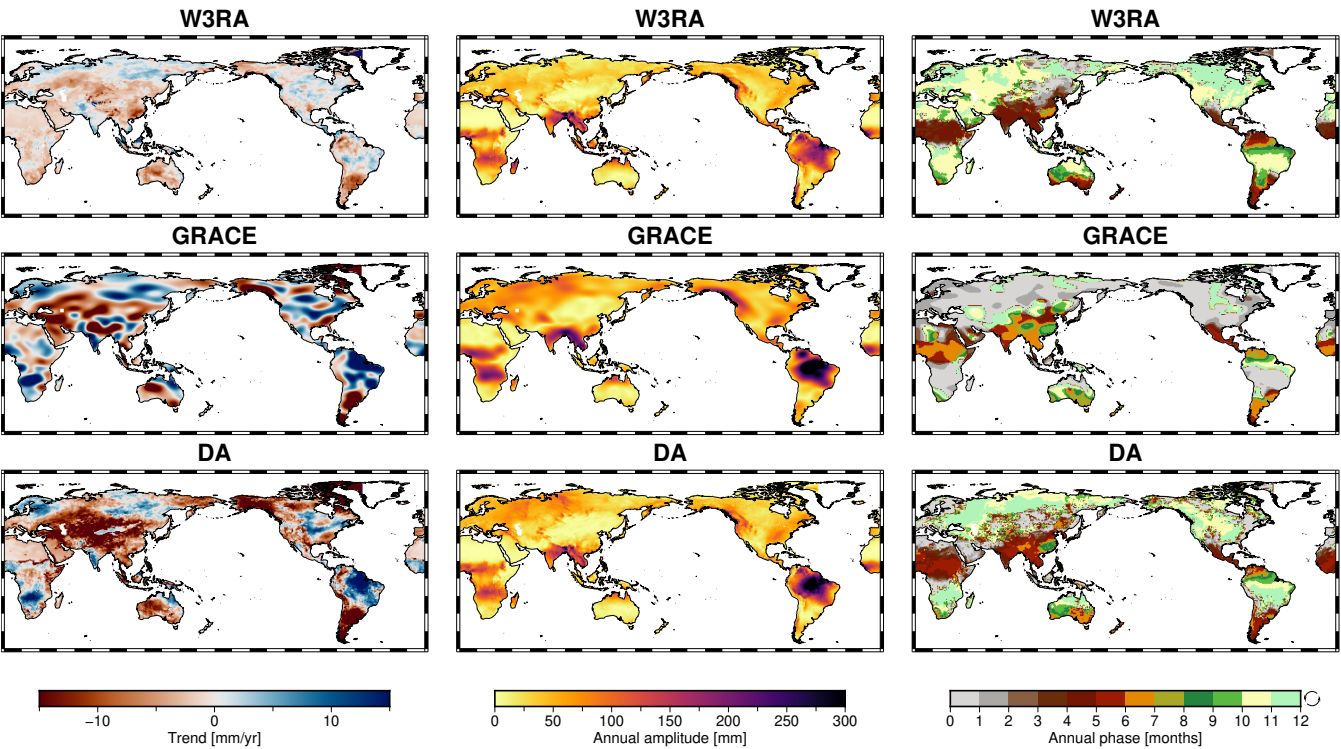
Figure 16. TWS results (2002-2010) in terms of secular trend (the 1st column), annual amplitude (the 2nd column) and annual phase (the 3rd column). Plots from the top to the bottom correspond to W3RA (open-loop ensemble-mean model output), GRACE (perturbation-free observations), and the assimilated results. For appropriate visual comparison, the results of W3RA and DA are upscaled from 0.1° to 0.5° to be consistent with GRACE-TWS. The annual phase is defined as the month of year when TWS reaches its maximum.

450   [modified for a better clarity]Then, the TWS results (from the model, GRACE observation, and GLDA) are collected in Figure 16. By comparing the trend map between W3RA and GRACE, one can see a very evident underestimation of the trend in TWS on a global scale in the model, which was also confirmed by previous studies that current models cannot well reproduce long-term change, particularly in groundwater (Scanlon et al., 2018; Mehrnegar et al., 2020; Forootan et al., 2024), and other evidence that shows groundwater is declining worldwide (Jasechko et al., 2024). GRACE TWS shows larger trends, but its spatial resolution

455   is too coarse for hydrological interpretation. In short, the model is poor at revealing the magnitude of trends, but can benefit from a high spatial resolution. Fortunately, GLDA can use these two features in a complementary way, where, for example, in Figure 16, one can see that DA results adopt the resolution of W3RA and the trend magnitudes are successfully tuned to GRACE. In statistics, as for the trend map, the spatial correlation between W3RA and DA is found to be 0.12, which is because the original model does not represent trends well. The correlation between GRACE and DA is increased to 0.35. The global

460   mean (latitude weighted) RMS (Root Mean Squared) of the differences between W3RA and DA is found to be 50.0 mm / year,

while that of GRACE and DA is 45.1 mm / year. Regarding the annual amplitude, the spatial correlation between W3RA and DA is found to be 0.86, while that between GRACE and DA is 0.92; the global mean RMS of differences between W3RA and DA is 26.0 mm, whereas that between GRACE and DA is 29.0 mm. Based on these results, it seems that the DA is more dependent on GRACE than on the W3RA model, in terms of trend and seasonality. Nevertheless, the current experiment might not be optimal, since the weight of the model also depends on how the ensemble is generated, and proper tuning of the model covariance should be subject to more investigations. In addition, validations of the proposed GLDA against independent data sets are desirable, which will be addressed in our future release of the GLDA product.

5 Limitations

[This Section is newly added, to address a few major concerns that reviewers have.]

PyGLDA consists of three main modules: (1) GHM, (2) GRACE observations, and (3) the DA system. Consequently, each of these modules introduces potential limitations that can affect the performance of PyGLDA, and they should be individually addressed. However, the limitations listed below are not exhaustive but represent the most critical challenges that should be prioritized in future updates.

GHM. (1) The W3RA model does not account for anthropogenic water use, such as irrigation, which significantly impacts the global water cycle. For example, in regions such as the north China plain, excessive groundwater extraction for irrigation is a major contributor to water depletion, but it is not well represented in W3RA. Consequently, model estimates in these regions may be less reliable. (2) Although W3RA includes a snow module, it does not adequately capture water storage changes in glaciated regions. For example, GRACE detects a strong ice-melting signal over Alaska, which is poorly represented in W3RA, leading to less accurate vertical disaggregation of water storage in glacier-covered areas.

GRACE. (1) Spatial leakage remains a significant limitation. Although PyGLDA excludes lake storage, the inherent spatial leakage of satellite gravity data can introduce artificial signals into adjacent areas, affecting the DA results in regions with significant changes in lake water, such as areas near the Caspian Sea. (2) Uncertainty in the true spatial resolution of GRACE. Larger grid cells improve the signal-to-noise ratio, but result in a loss of spatial detail. Although PyGLDA allows users to flexibly define the grid size for GRACE observations, determining the optimal balance between spatial resolution and noise remains an open question.

DA. (1) The current perturbation module in PyGLDA is sub-optimal for both forcing fields and model parameters. The current implementation assumes spatial homogeneity in perturbations to prevent spatially independent perturbations from canceling out at the basin scale. However, this simplification may lead to unrealistic model uncertainty representations, ultimately degrading DA performance. Future improvements should explore spatially correlated perturbation strategies to enhance covariance realism. (2) Covariance localization is currently absent. Although no clear signs of filter divergence have been observed in PyGLDA due to its patch-wise and coarse grid implementation (e.g., 3° grid cells), users may face potential risks of filter divergence when using finer grid configurations, such as 1° resolution. Incorporating traditional covariance localization techniques,

such as the Schur product method (Webber and Morzfeld, 2023), could improve the stability and robustness of PyGLDA in future updates.

6   Conclusions

[This section has been reformulated by (i) removing some of limitation statement, and (2) adding more possible extensions of current study.]

This study introduces PyGLDA, a community-level and open-source high-resolution global land data assimilation (DA) system designed to integrate GRACE/GRACE-FO-derived TWS observations with global hydrological models (GHMs). Our major contribution is that PyGLDA can be used globally and regionally at various resolutions over arbitrary regions of the land. In addition, PyGLDA can allow users to flexibly switch between grid-scale and basin-scale DA. Unlike previous global DA implementations, PyGLDA features a patch-wise domain localization strategy and a weighted global aggregation scheme, which enhance computational efficiency and ensure smooth spatial transitions across assimilation regions. Our results demonstrate that PyGLDA effectively reduces discrepancies between model-based TWS estimates and GRACE observations, providing an improved representation of long-term trends and seasonal water storage variations at global and regional scales.

Beyond its current capabilities, PyGLDA provides a flexible and scalable DA framework that can be extended for multi-sensor hydrological data assimilation. Future developments will focus on incorporating additional observational datasets, such as satellite-based soil moisture retrievals, to further improve the reliability of water storage estimates. Additionally, advances in satellite gravimetry, including upcoming missions like MAGIC and NGGM, may offer higher spatial and temporal resolution with improved error characteristics, enabling even more precise DA applications. The modular nature of PyGLDA ensures that it can seamlessly integrate with these future datasets, making it a valuable tool for hydrological modeling, climate studies, and water resource management.

PyGLDA also faces challenges as addressed in Section 5. Although PyGLDA is designed for multi-GHM compatibility, this study focuses primarily on W3RA [clarified as suggested]. In future updates, a more advanced GHM such as WaterGAP 2.2e (Müller Schmied et al., 2024), could be considered an effective alternative to account for the use of anthropogenic water, reservoirs, lakes, and glaciers. In particular, this model is released in Python and can be compatible with our PyGLDA framework in principle. Other numerical issues should also be considered, for example, (i) using covariance localization to stabilize the filter and exploit the spatial details, (ii) accounting for reasonable (e.g., local correlation) spatio-temporal perturbation to better simulate the model covariance, and (iii) simulating lake signal and reducing spatial leakage from observations (Deggim et al., 2021). All these extensions shall be the major focus of PyGLDA's next updates.

Code and data availability. PyGLDA is programed in Python as an open-source project licensed under MIT, and its latest stable version is publicly available at Zenodo repository (Yang, 2024), which comes with the code, a guide to installation, and a few demos to showcase various DA experiments. In that repository, we also offer the necessary sample data to ensure an instant running of all those demonstrations. In addition, PyGLDA is being actively developed in a GitHub repository; see https://github.com/AAUGeodesyGroup/PyGLDA, where one can obtain the newest version. The original Matlab implementation of the W3RA water balance model can be found in https://www.dropbox.

The shape files (.shp) of DRB and global patches used in this study, as well as other auxiliary data to run PyGLDA demos are available at Yang (2024). GRACE level-2 monthly gravity fields are available at ICGEM (International Center for Global Gravity Field Models) via https:

530   //icgem.gfz-potsdam.de/home. The ITSG-2018 data (the normal equations in Sinex) used to estimate GRACE's variance-covariance matrices are available at http://ftp.tugraz.at/outgoing/ITSG/GRACE/ITSG-Grace2018/monthly/normals_SINEX/monthly_n96/. The propagation of covariance from the harmonic domain to the gridded domain (TWS) follows the methodology of Yang et al. (2024a). The meteorological forcing fields, i.e., ERA5-land, are all downloaded from https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview. The $0.1°$ ERA5-land data have been used to force the W3RA model.

535   **Appendix A: EnKF and EnKS**

The Ensemble Kalman Filter (EnKF) is a merger that applies observations to update the state and parameters of the model based on their relative errors (Evensen et al., 2009). Model errors are generated by a Monte Carlo simulation that considers ~~i (i.e., $i = 1,...,n$)~~ n ensemble members of the model parameters (Θ_i^f) expressed as:

$$\boldsymbol{\Theta}_i = \boldsymbol{\Theta}_0 + \xi_i,\ i = 1,...n, \tag{A1}$$

540   where $\boldsymbol{\Theta}_{0\,m_1 \times 1}$ (m_1 being the number of parameters) is a vector representing the initial values of the model parameters. Random errors (ξ_i) are introduced to perturb the initial values of the parameters. The distribution of these errors can be Gaussian and triangle as (Schumacher, 2016). In the example of PyGLDA as described in Section 4.1, nine model parameters are selected to be perturbed, thus $m_1 = 9$. Perturbations of the forcing fields follow

$$\boldsymbol{M}_i = \boldsymbol{M}_0 + r_i,\ i = 1,...n, \tag{A2}$$

545   where $\boldsymbol{M}_{0\,p \times q}$ (p being the number of grid points, q being the number of forcing field) is a vector representing the initial values of the forcing inputs. Random errors (r_i) are considered to perturb the initial values of these errors. ~~we consider multiplicative errors for precipitation, thus $q = 1$~~In PyGLDA, for example, we consider multiplicative errors for precipitation, temperature, and shortwave solar radiation, thus $q = 3$. The model's forecasting (denoted by ^{f}, which can represent the simulation of W3RA or any other GHMs) ensembles are then generated as follows:

550    $$\mathbf{X}_i^f = F(\Theta_i, \boldsymbol{M}_i),\ i = 1,...n. \tag{A3}$$

For example, we use $n = 30$ ensemble members to perform our experiments of Section 4.

In each step of GLDA (TWS data are introduced), each ensemble of model states (e.g., in forecasting mode) is represented by $\mathbf{X}_{p \times n}^f$ as:

$$\mathbf{X}^f = (\mathbf{x}_1^f, \cdots, \mathbf{x}_n^f), \tag{A4}$$

where $\mathbf{x}^f_{i\ p\times 1}(i=1,\cdots,n)$ is the i'th ensemble. The ensemble mean is defined as:

$$\bar{\mathbf{x}}^f = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^f, \tag{A5}$$

and the error covariance matrix of the model forecast ($\mathbf{C}_{m\times m}^f$) is computed as follows:

$$\mathbf{C}^f = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)(\mathbf{x}_i^f - \bar{\mathbf{x}}^f)^T. \tag{A6}$$

The analysis step (shown by the upper-index a) corrects the states and their uncertainties using the GRACE-TWS observations as:

$$\mathbf{X}^a = (\mathbf{x}_1^a,\cdots,\mathbf{x}_n^a) = \mathbf{X}^f + \mathbf{K}(\mathbf{Y} - \mathbf{H}\mathbf{X}^{\mathrm{f}}), \tag{A7}$$

$$\mathbf{C}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{C}^f, \tag{A8}$$

where $Y_{p\times n}$ is an ensemble of actual observations, which are perturbed by the estimated noise from the covariance matrix of observations (Yang et al., 2024a). This means that, for the example of Section 4.2, the dimension is 19×30, and for the global GLDA (Section 4.3) with grid resolution $3°$, the dimension is roughly 1700×30. The analyzed unknowns ($\mathbf{X}_{p\times n}^a$) and their covariance matrix ($\mathbf{C}_{p\times p}^a$) depend on the estimation of the Kalman gain matrix ($\mathbf{K}$) and the design matrix $\mathbf{H}$, which are defined as

$$\mathbf{K} = \mathbf{C}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{C}^{\mathrm{f}}\mathbf{H}^{\mathrm{T}} + \mathbf{C}^{\mathrm{r}}\right)^{-1}, \tag{A9}$$

where C^{r} is the covariance matrix of the satellite-derived TWS estimates. However, the formulation of $\mathbf{H}$ is highly dependent on the definition of the state vector $\mathbf{X}$. For the W3RA model, in terms of each ensemble member, $\mathbf{X}$ includes seven vertical layers following

$$\mathbf{X} = (\mathbf{s}_1,\cdots,\mathbf{s}_v), \tag{A10}$$

where $\mathbf{s}$ represents one of the surface water, the topsoil water, the shallow soil water, the deep soil water and the groundwater, the biomass water, and the open free water. However, since some of these vertical compartments might have more than one column, the dimension (v) along the vertical profile is 10 in our study. Laterally, the state vector $\mathbf{X}$ adheres to the native spatial resolution of the W3RA model, for example, $0.1°$, and therefore for GLDA in Section 4.3 the dimension is more than 2 million on a global scale. In this case, which is exactly the case for EnKF, $\mathbf{H}$ represents the vertical aggregation (into TWS) and the lateral aggregation (from $0.1°$ to $3°$).

Our EnKS implementation is similar to that of Tian et al. (2017), where $\mathbf{X}$ no longer represents the monthly mean like GRACE-TWS. Instead, $\mathbf{X}$ consists of daily state vectors of one month from the model, and in this case, the design matrix $\mathbf{H}$ of EnKS shall be responsible for the temporal aggregation other than the vertical and lateral aggregation of EnKF. The benefit of EnKS is potentially better tuning of daily output by accounting for the temporal covariance, but the price is a heavier computational burden as the dimension along time increases to 30 (or 28 or 31) from 1 of EnKF. ~~More details on the derivation of the design matrix and EnKS can refer to~~ More details on the derivation of the design matrix and EnKS can be found in Tian et al. (2017).

Table B1. A summary of input-output of PyGLDA

Input	Class	Sub-class
	GRACE	L1b-L2 normal equations
		L2 monthly gravity fields
		GIA model
		DDK filtering matrix
		Low-degree coefficients
	GHM	Forcing field from ERA5-land
		Climatologies
		Parameter fields
	Mask (.shp)	GRACE global land mask
		GHM global land mask
		Forcing field global land mask
		GHM running mask
		Basin and subbasin mask
	Setting (.json)	Data preprocess
		W3RA-GHM setting
		Perturbation
		Data assimilation
Output	**Class**	**Sub-class**
	GRACE (.h5df)	Grid/Basin-scale TWS
		Perturbed TWS ensemble
		Full variance-covariance
	GHM (.h5df)	Initial field (spin-up)
		Cropped input fields
		Perturbed input field ensemble
		Daily state vector ensemble
		Else output, e.g., evaporation
	DA (.h5df)	Ensemble mean of state vector
		Ensemble spread of state vector
		Monthly mean output
		Statistics, e.g., basin TWS

585 **Appendix B: A summary of input-output of PyGLDA**

Please refer to Table B1 for a record of the input and output of the software. Details on the inputs (GRACE, GHM, and Mask) can also be found in Section 2. Information on the setting files (.json) can be found on our Github or in the data repository (Yang, 2024). The output can also be obtained by running the demo offered in that repository.

Author contributions. Fan Yang led the writing of the manuscript and the design of the software. Maike Schumacher led the design of DA algorithms, Leire Retegui-Schiettekatte helped with the DA implementation, Albert van Dijk offered the original W3RA code and helped with technical discussions, and Ehsan Forootan led the project and improved the writing. All authors commented on the manuscript draft.

Competing interests. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgements. This work is supported by the Danmarks Frie Forskningsfond [10.46540/2035-00247B] through the DANSk-LSM project. Fan Yang acknowledges financial support from the National Natural Science Foundation of China (Grant Nos. 42274112 and 41804016). Maike Schumacher acknowledges financial supports through VILLUM FONDEN under the research grant number VIL60779.

590

595

References

A, G., Wahr, J., and Zhong, S.: Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, Geophysical Journal International, 192, 557–572, https://doi.org/10.1093/gji/ggs030, 2013.

600 Benedict, I., van Heerwaarden, C. C., Weerts, A. H., and Hazeleger, W.: An evaluation of the importance of spatial resolution in a global climate and hydrological model based on the Rhine and Mississippi basin, Hydrology and Earth System Sciences Discussions, 2017, 1–28, https://doi.org/10.5194/hess-2017-473, 2017.

Boergens, E., Kvas, A., Eicker, A., Dobslaw, H., Schawohl, L., Dahle, C., Murböck, M., and Flechtner, F.: Uncertainties of GRACE-Based Terrestrial Water Storage Anomalies for Arbitrary Averaging Regions, Journal of Geophysical Research: Solid Earth, 127, 605 https://doi.org/10.1029/2021JB022081, 2022.

Caron, L., Ivins, E. R., Larour, E., Adhikari, S., Nilsson, J., and Blewitt, G.: GIA Model Statistics for GRACE Hydrology, Cryosphere, and Ocean Science, Geophysical Research Letters, 45, 2203–2212, https://doi.org/10.1002/2017gl076644, 2018.

Castellazzi, P., Martel, R., Rivera, A., Huang, J., Pavlic, G., Calderhead, A. I., Chaussard, E., Garfias, J., and Salas, J.: Groundwater depletion in Central Mexico: Use of GRACE and InSAR to support water resources management, Water resources research, 52, 5985–6003, 610 https://doi.org/10.1002/2015WR018211, 2016.

Castle, S. L., Reager, J. T., Thomas, B. F., Purdy, A. J., Lo, M.-H., Famiglietti, J. S., and Tang, Q.: Remote detection of water management impacts on evapotranspiration in the Colorado River Basin, Geophysical Research Letters, 43, 5089–5097, https://doi.org/10.1002/2016GL068675, 2016.

Chen, J., Wilson, C. R., and Seo, K.-W.: Optimized smoothing of Gravity Recovery and Climate Experiment (GRACE) time-variable gravity 615 observations, Journal of Geophysical Research: Solid Earth, 111, https://doi.org/10.1029/2005JB004064, 2006.

Chen, J., Wilson, C. R., Li, J., and Zhang, Z.: Reducing leakage error in GRACE-observed long-term ice mass change: a case study in West Antarctica, Journal of Geodesy, 89, 1432–1394, https://doi.org/10.1007/s00190-015-0824-2, 2015.

Chen, J., Cazenave, A., Dahle, C., Llovel, W., Panet, I., Pfeffer, J., and Moreira, L.: Applications and Challenges of GRACE and GRACE Follow-On Satellite Gravimetry, Surveys in Geophysics, 43, 305–345, https://doi.org/10.1007/s10712-021-09685-x, 2022.

620 Daras, I., March, G., Pail, R., Hughes, C. W., Braitenberg, C., Güntner, A., Eicker, A., Wouters, B., Heller-Kaikov, B., Pivetta, T., and Pastorutti, A.: Mass-change And Geosciences International Constellation (MAGIC) expected impact on science and applications, Geophysical Journal International, 236, 1288–1308, https://doi.org/10.1093/gji/ggad472, 2023.

de Graaf, I. E., van Beek, R. L., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in Water Resources, 102, 53–67, 625 https://doi.org/10.1016/j.advwatres.2017.01.011, 2017.

Decharme, B., Delire, C., Minvielle, M., Colin, J., Vergnes, J.-P., Alias, A., Saint-Martin, D., Séférian, R., Sénési, S., and Voldoire, A.: Recent Changes in the ISBA-CTRIP Land Surface System for Use in the CNRM-CM6 Climate Model and in Global Off-Line Hydrological Applications, Journal of Advances in Modeling Earth Systems, 11, 1207–1252, https://doi.org/10.1029/2018MS001545, 2019.

Deggim, S., Eicker, A., Schawohl, L., Gerdener, H., Schulze, K., Engels, O., Kusche, J., Saraswati, A. T., van Dam, T., Ellenbeck, L., 630 Dettmering, D., Schwatke, C., Mayr, S., Klein, I., and Longuevergne, L.: RECOG RL01: correcting GRACE total water storage estimates for global lakes/reservoirs and earthquakes, Earth System Science Data, 13, 2227–2244, https://doi.org/10.5194/essd-13-2227-2021, 2021.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Schmied, H. M.: 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, 1285–1309, https://doi.org/10.1007/s10712-014-9309-8, 2014.

635 Eicker, A., Forootan, E., Springer, A., Longuevergne, L., and Kusche, J.: Does GRACE see the terrestrial water cycle "intensifying"?, Journal of Geophysical Research: Atmospheres, 121, 733–745, https://doi.org/10.1002/2015JD023808, 2016.

Evensen, G. et al.: Data assimilation: the ensemble Kalman filter, vol. 2, Springer, 2009.

Felsberg, A., Lannoy, G. J. M. D., Girotto, M., Poesen, J., Reichle, R. H., and Stanley, T.: Global Soil Water Estimates as Landslide Predictor: The Effectiveness of SMOS, SMAP, and GRACE Observations, Land Surface Simulations, and Data Assimilation, Journal of

640 Hydrometeorology, 22, 1065 – 1084, https://doi.org/10.1175/JHM-D-20-0228.1, 2021.

Forman, B. A. and Reichle, R. H.: The spatial scale of model errors and assimilated retrievals in a terrestrial water storage assimilation system, Water Resources Research, 49, 7457–7468, https://doi.org/10.1002/2012WR012885, 2013.

Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., van Dijk, A., Brocca, L., Farzaneh, S., Akinluyi, F., Ramillien, G., Shum, C., Awange, J., and Mostafaie, A.: Understanding the global hydrological droughts of 2003-2016 and their relationships with

645 teleconnections, Science of The Total Environment, 650, 2587–2604, https://doi.org/10.1016/j.scitotenv.2018.09.231, 2019.

Forootan, E., Mehrnegar, N., Schumacher, M., Schiettekatte, L. A. R., Jagdhuber, T., Farzaneh, S., van Dijk, A. I., Shamsudduha, M., and Shum, C.: Global groundwater droughts are more severe than they appear in hydrological models: An investigation through a Bayesian merging of GRACE and GRACE-FO data with a water balance model, Science of The Total Environment, 912, 169 476, https://doi.org/10.1016/j.scitotenv.2023.169476, 2024.

650 Gerdener, H., Kusche, J., Schulze, K., Döll, P., and Klos, A.: 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, 73, https://doi.org/10.1007/s00190-023-01763-9, 2023.

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

655 Gouweleeuw, B. T., Kvas, A., Gruber, C., Gain, A. K., Mayer-Gürr, T., Flechtner, F., and Güntner, A.: Daily GRACE gravity field solutions track major flood events in the Ganges-Brahmaputra Delta, Hydrology and Earth System Sciences, 22, 2867–2880, https://doi.org/10.5194/hess-22-2867-2018, 2018.

Han, S.-C., Shum, C. K., Jekeli, C., Kuo, C.-Y., Wilson, C., and Seo, K.-W.: Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement, Geophysical Journal International, 163, 18–25, https://doi.org/10.1111/j.1365-246X.2005.02756.x, 2005.

660 Hersbach, H., Bell, B., Berrisford, P., and et al: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., van Beek, R. L. P. H., and Bierkens, M. F. P.: Hyper-resolution PCR-GLOBWB: opportunities and challenges from refining model spatial resolution to 1 km over the European continent, Hydrology and Earth System Sciences, 27, 1383–1401, https://doi.org/10.5194/hess-27-1383-2023, 2023.

665 Horvath, A., Murböck, M., Pail, R., and Horwath, M.: Decorrelation of GRACE time variable gravity field solutions using full covariance information, Geosciences, 8, 323, https://doi.org/10.3390/geosciences8090323, 2018.

Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., Fallatah, O., and Kirchner, J. W.: Rapid groundwater decline and some cases of recovery in aquifers globally, Nature, 625, 715–721, https://doi.org/10.1038/s41586-023-06879-8, 2024.

Jekeli, C.: Alternative methods to smooth the Earth's gravity field, Report No. 327, Reports of the Department of Geodetic Science and Surveying, Ohio State Univ. Columbus, OH, United States, 1981.

Khaki, M., Schumacher, M., Forootan, E., Kuhn, M., Awange, J. L., and van Dijk, A. I.: 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, 2017.

Klees, R., Zapreeva, E. A., Winsemius, H. C., and Savenije, H. H. G.: The bias in GRACE estimates of continental water storage variations, Hydrology and Earth System Sciences, 11, 1227–1241, https://doi.org/10.5194/hess-11-1227-2007, 2007.

Klees, R., Revtova, E. A., Gunter, B. C., Ditmar, P., Oudman, E., Winsemius, H. C., and Savenije, H. H. G.: The design of an optimal filter for monthly GRACE gravity models, Geophysical Journal International, 175, 417–432, https://doi.org/10.1111/j.1365-246X.2008.03922.x, 2008.

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

Kusche, J., Schmidt, R., Petrovic, S., and Rietbroek, R.: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, Journal of Geodesy, 81, 903–913, https://doi.org/10.1007/s00190-009-0308-3, 2009.

Kvas, A. and Mayer-Gürr, T.: GRACE gravity field recovery with background model uncertainties, Journal of geodesy, 93, 2543–2552, https://doi.org/10.1007/s00190-019-01314-1, 2019.

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

Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48, https://doi.org/10.1029/2011WR011453, 2012.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., Bettadpur, S. V., Byun, S. H., Dahle, C., Dobslaw, H., Fahnestock, E., Harvey, N., Kang, Z., Kruizinga, G. L. H., Loomis, B. D., McCullough, C., Murböck, M., Nagel, P., Paik, M., Pie, N., Poole, S., Strekalov, D., Tamisiea, M. E., Wang, F., Watkins, M. M., Wen, H.-Y., Wiese, D. N., and Yuan, D.-N.: Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance, Geophysical Research Letters, 47, https://doi.org/10.1029/2020GL088306, 2020.

Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., et al.: Global GRACE data assimilation for groundwater and drought monitoring: Advances and challenges, Water Resources Research, 55, 7564–7586, https://doi.org/10.1029/2018WR024618, 2019.

Liu, S., Yang, F., and Forootan, E.: SAGEA: A toolbox for comprehensive error assessment of GRACE and GRACE-FO based mass changes, Computers & Geosciences, 196, 105 825, https://doi.org/10.1016/j.cageo.2024.105825, 2025.

Longuevergne, L., Scanlon, B. R., and Wilson, C. R.: GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, Water Resources Research, 46, https://doi.org/10.1029/2009wr008564, 2010.

Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W., and Luthcke, S. B.: Replacing GRACE/GRACE-FO With Satellite Laser Ranging: Impacts on Antarctic Ice Sheet Mass Change, Geophysical Research Letters, 47, https://doi.org/10.1029/2019gl085488, 2020.

Mayer-Gürr, T., Behzadpour, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., and Zehentner, N.: ITSG-Grace2018-monthly, daily and static gravity field solutions from GRACE, 2018.

Mehrnegar, N., Jones, O., Singer, M. B., Schumacher, M., Bates, P., and Forootan, E.: Comparing global hydrological models and combining them with GRACE by dynamic model data averaging (DMDA), Advances in Water Resources, 138, 103 528, https://doi.org/10.1016/j.advwatres.2020.103528, 2020.

710 Mehrnegar, N., Jones, O., Singer, M. B., Schumacher, M., Jagdhuber, T., Scanlon, B. R., Rateb, A., and Forootan, E.: Exploring groundwater and soil water storage changes across the CONUS at 12.5 km resolution by a Bayesian integration of GRACE data into W3RA, Science of The Total Environment, 758, 143 579, https://doi.org/10.1016/j.scitotenv.2020.143579, 2021.

Mehrnegar, N., Schumacher, M., Jagdhuber, T., and Forootan, E.: Making the Best Use of GRACE, GRACE-FO and SMAP Data Through a Constrained Bayesian Data-Model Integration, Water Resources Research, 59, https://doi.org/10.1029/2023WR034544, 2023.

Meyer, U., Jean, Y., Kvas, A., Dahle, C., Lemoine, J.-M., and Jäggi, A.: Combination of GRACE monthly gravity fields on the normal

715 equation level, Journal of geodesy, 93, 1645–1658, https://doi.org/10.1007/s00190-019-01274-6, 2019.

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geoscientific Model Development, 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.

Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., Kynast, E., Peiris, T. A., Schiebener, L., Schu-

720 macher, M., and Döll, P.: The global water resources and use model WaterGAP v2.2e: description and evaluation of modifications and new features, Geoscientific Model Development, 17, 8817–8852, https://doi.org/10.5194/gmd-17-8817-2024, 2024.

Ngo-Duc, T., Laval, K., Ramillien, G., Polcher, J., and Cazenave, A.: Validation of the land water storage simulated by Organising Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) with Gravity Recovery and Climate Experiment (GRACE) data, Water Resources Research, 43, https://doi.org/10.1029/2006WR004941, 2007.

725 Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., and Su, H.: Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007522, 2007.

Patel, C. R.: Analyzing and Monitoring GRACE-FO Star Camera Performance in a Changing Environment , Master's thesis, The University of Texas at Austin, 2020.

730 Reichle, R. H. and Koster, R. D.: Assessing the Impact of Horizontal Error Correlations in Background Fields on Soil Moisture Estimation, Journal of Hydrometeorology, 4, 1229 – 1242, https://doi.org/10.1175/1525-7541(2003)004<1229:ATIOHE>2.0.CO;2, 2003.

Richard Peltier, W., Argus, D. F., and Drummond, R.: Comment on "An Assessment of the ICE-6G_C (VM5a) Glacial Isostatic Adjustment Model" by Purcell et al., Journal of Geophysical Research: Solid Earth, 123, 2019–2028, https://doi.org/10.1002/2016JB013844, 2018.

Rodell, M. and Reager, J. T.: Water cycle science enabled by the GRACE and GRACE-FO satellite missions, Nature Water, 1, 47–59,

735 https://doi.org/10.1038/s44221-022-00005-0, 2023.

Rodell, M., Famiglietti, J., Wiese, D., and et al: Emerging trends in global freshwater availability, Nature, 557, 651–659, https://doi.org/10.1038/s41586-018-0123-1, 2018.

Save, H., Bettadpur, S., and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons, Journal of Geophysical Research: Solid Earth, 121, 7547–7569, https://doi.org/10.1002/2016JB013007, 2016.

740 Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., and McMahon, P. B.: Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, Proceedings of the National Academy of Sciences, 109, 9320–9325, https://doi.org/10.1073/pnas.1200311109, 2012.

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., Van Beek, L. P., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., et al.: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data, Proceedings of the National Academy of Sciences, 115, E1080–E1089, https://doi.org/10.1073/pnas.170466511, 2018.

Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., Minvielle, M., Calvet, J.-C., Decharme, B., Eisner, S., Fink, G., Flörke, M., Peßenteiner, S., van Beek, R., Polcher, J., Beck, H., Orth, R., Calton, B., Burke, S., Dorigo, W., and Weedon, G. P.: A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, Earth System Science Data, 9, 389–413, https://doi.org/10.5194/essd-9-389-2017, 2017.

Schumacher, M.: Methods for assimilating remotely-sensed water storage changes into hydrological models, Ph.D. thesis, Universitäts-und Landesbibliothek Bonn, 2016.

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

Schumacher, M., Forootan, E., van Dijk, A. I., Schmied, H. M., Crosbie, R. S., Kusche, J., and Döll, P.: Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model, Remote Sensing of Environment, 204, 212–228, https://doi.org/10.1016/j.rse.2017.10.029, 2018a.

Schumacher, M., Forootan, E., van Dijk AIJM, Müller Schmied, H., Crosbie, R., Kusche, J., and Döll, P.: Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model, Remote Sensing of Environment, 204, 212–228, doi:10.1016/j.rse.2017.10.029, 2018b.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., and McCabe, M. F.: Closing the terrestrial water budget from satellite remote sensing, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL037338, 2009.

Springer, A., Karegar, M. A., Kusche, J., Keune, J., Kurtz, W., and Kollet, S.: Evidence of daily hydrological loading in GPS time series over Europe, Journal of geodesy, 93, 2145–2153, 2019.

Stagl, J. C. and Hattermann, F. F.: Impacts of Climate Change on the Hydrological Regime of the Danube River and Its Tributaries Using an Ensemble of Climate Scenarios, Water, 7, 6139–6172, https://doi.org/10.3390/w7116139, 2015.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geoscientific Model Development, 11, 2429–2453, https://gmd.copernicus.org/articles/11/2429/2018/, 2018.

Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL025285, 2006.

Swenson, S., Chambers, D., and Wahr, J.: Estimating geocenter variations from a combination of GRACE and ocean model output, Journal of Geophysical Research: Solid Earth, 113, https://doi.org/10.1029/2007jb005338, 2008.

Tapley, B. D., Bettadpur, S., Watkins, M., and Reigber, C.: The gravity recovery and climate experiment: Mission overview and early results, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL019920, 2004.

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

van Dijk, A.: AWRA Technical Report 3. Landscape Model (version 0.5) Technical Description, 2010.

780 van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., and Tregoning, P.: A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, Hydrology and Earth System Sciences, 18, 2955–2973, https://doi.org/10.5194/hess-18-2955-2014, 2014.

Vishwakarma, B. D., Devaraju, B., and Sneeuw, N.: What Is the Spatial Resolution of GRACE Satellite Products for Hydrology?, Remote Sensing, 10, https://doi.org/10.3390/rs10060852, 2018.

785 Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., and Swenson, S. C.: Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, Water resources research, 49, 904–914, https://doi.org/10.1002/wrcr.20078, 2013.

Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, Journal of Geophysical Research: Solid Earth, 103, 30 205–30 229, https://doi.org/10.1029/98jb02844, 1998.

790 Wahr, J., Swenson, S., and Velicogna, I.: Accuracy of GRACE mass estimates, Geophysical Research Letters, 33, https://doi.org/10.1029/2005GL025305, 2006.

Webber, R. J. and Morzfeld, M.: Localized covariance estimation: A Bayesian perspective, arXiv preprint arXiv:2301.04828, 2023.

Yang, F.: PyGLDA: Python-based Global Land Data Assimilation, https://doi.org/10.5281/zenodo.12206756, 2024.

Yang, F., Kusche, J., Forootan, E., and Rietbroek, R.: Passive-ocean radial basis function approach to improve temporal gravity recovery
795 from GRACE observations, Journal of Geophysical Research: Solid Earth, 122, https://doi.org/10.1002/2016JB013633, 2017.

Yang, F., Forootan, E., Schumacher, M., Shum, C., and Zhong, M.: Evaluating non-tidal atmospheric products by measuring GRACE K-band range rate residuals, Geophysical Journal International, 215, 1132–1147, https://doi.org/10.1093/gji/ggy340, 2018.

Yang, F., Forootan, E., Wang, C., Kusche, J., and Luo, Z.: A New 1-Hourly ERA5-Based Atmosphere De-Aliasing Product for GRACE, GRACE-FO, and Future Gravity Missions, Journal of Geophysical Research: Solid Earth, 126, e2021JB021 926,
800 https://doi.org/10.1029/2021JB021926, 2021.

Yang, F., Liang, L., Wang, C., and Luo, Z.: Attitude Determination for GRACE-FO: Reprocessing the Level-1A SC and IMU Data, Remote Sensing, 14, https://doi.org/10.3390/rs14010126, 2022a.

Yang, F., Luo, Z., Zhou, H., and Kusche, J.: On study of the Earth topography correction for the GRACE surface mass estimation, Journal of Geodesy, 96, https://doi.org/10.1007/s00190-022-01683-0, 2022b.

805 Yang, F., Forootan, E., Liu, S., and Schumacher, M.: A Monte Carlo Propagation of the Full Variance-Covariance of GRACE-Like Level-2 Data With Applications in Hydrological Data Assimilation and Sea-Level Budget Studies, Water Resources Research, 60, e2023WR036 764, https://doi.org/10.1029/2023WR036764, 2024a.

Yang, F., Liu, S., and Forootan, E.: A spatial-varying non-isotropic Gaussian-based convolution filter for smoothing GRACE-like temporal gravity fields, Journal of Geodesy, 98, 66, https://doi.org/10.1007/s00190-024-01875-w, 2024b.

810 Zaitchik, B. F., Rodell, M., and Reichle, R. H.: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin, Journal of Hydrometeorology, 9, 535–548, https://doi.org/10.1175/2007JHM951.1, 2008.

Zhang, Z.-Z., Chao, B. F., Lu, Y., and Hsu, H.-T.: An effective filtering for GRACE time-variable gravity: Fan filter, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL039459, 2009.

Zhou, H., Zheng, L., Pail, R., Liu, S., Qing, T., Yang, F., Guo, X., and Luo, Z.: The impacts of reducing atmospheric and
815 oceanic de-aliasing model error on temporal gravity field model determination, Geophysical Journal International, 234, 210–227, https://doi.org/10.1093/gji/ggad064, 2023.