



# Multidecadal reconstruction of terrestrial water storage changes by combining pre-GRACE satellite observations and climate data

Charlotte Hacker<sup>1</sup>, Benjamin D. Gutknecht<sup>1</sup>, Anno Löcher<sup>1</sup>, and Jürgen Kusche<sup>1</sup>

<sup>1</sup>Institute for Geodesy and Geoinformation, University of Bonn, Bonn, Germany

**Correspondence:** Charlotte Hacker (chacker@uni-bonn.de)

**Abstract.** The Gravity Recovery And Climate Experiment (GRACE) and its follow-on mission, GRACE-FO, have observed global mass changes and transports, expressed as terrestrial water storage anomalies (TWSA), for over two decades. However, for climate model evaluation, climate change attribution and other applications, multi-decadal TWSA time series are required. This need has triggered several studies on reconstructing TWSA via regression approaches or machine learning techniques, with the help of predictor variables such as rainfall, land or sea surface temperature. Here, we combine such an approach, for the first time, with large-scale time-variable gravity information from geodetic satellite laser ranging (SLR) and Doppler Orbitography by Radiopositioning Integrated on Satellite (DORIS) tracking. The new reconstruction TWSTORE (Terrestrial Water STorage REconstruction) is formulated in a GRACE-derived empirical orthogonal functions (EOFs) basis and complemented with the Löcher et al. (2025) approach, in which global gravity fields are solved from SLR ranges and DORIS observations in EOF space for the pre-GRACE time frame. Our approach is highly modular, allowing to use different data sets at several steps in the workflow.

We reconstruct GRACE-like TWSA for the global land, excluding Greenland and Antarctica, from 1984 onward. We find that the new combined reconstruction inherits information from the geodetic method, mainly at longer timescales. In contrast, at the seasonal scale, the climate-driven reconstruction and the geodetic product are already surprisingly consistent. In comparison to other reconstructions, we find thus major differences mainly at the multi-decadal timescale. All in all, our study confirms the presence of significant changes in storage trends, showing that GRACE-derived results should not be extrapolated to the past. The reconstructed fields and corresponding uncertainty information are available at <https://doi.org/10.5281/zenodo.15827789> (Hacker, 2025). We also derive evaporation based on the water balance equation and the presented reconstruction for 11 river basins. The corresponding time series are available at <https://doi.org/10.5281/zenodo.16643628> Gutknecht (2025).

## 1 Introduction

The GRACE and GRACE-FO missions (GRACE/-FO, Tapley et al., 2019) have provided an unprecedented data record for understanding our Earth system, beginning in 2002 and interrupted only by an eleven-month gap between the missions and a few missing months due to instrument problems. Monthly GRACE/-FO gravity models, after conversion to terrestrial water storage anomaly (TWSA) grids, have an effective spatial resolution of a few hundred kilometers. They have become a mainstay in global hydrological modeling (Rodell and Reager, 2023), sea level and ocean mass studies (Chambers et al., 2010), and the



assessment of large earthquakes (Han et al., 2024). When assimilated into hydrological or land surface models, at basin scale (Zaitchik et al., 2008) or grid level (Eicker et al., 2014), they improve the spatio-temporal resolution and realism of hydrological modeling, e.g., (Li et al., 2019; Gerdener et al., 2023). However, applications that rely on estimates of land water storage trends, such as assessing continental wetting and drying in response to climate forcing (Jensen et al., 2019) or separating anthropogenic and natural variability (Hosseini-Moghari et al., 2023), require long, multi-decadal time series. The same holds if the aim is to evaluate the increase in the frequency of extreme events, e.g., across land water storage, precipitation, and surface temperatures. Several papers have demonstrated that the current GRACE/-FO record is too short to enable robust assessments of increasing extremes in storage deficit or surplus (Kusche et al., 2016) and, e.g., climate model evaluation (Jensen et al., 2024).

Several "TWSA reconstruction" approaches and data products have emerged, typically based on climate data time series and the training of a regression or machine learning model on GRACE/FO-derived maps of TWSA. Global reconstructions of gridded terrestrial water storage (Humphrey and Gudmundsson, 2019b; Li et al., 2021; Chandanpurkar et al., 2022b; Palazzoli et al., 2025; Mandal et al., 2025) provide long and gapless time series with high spatial resolution since their predictor climate data records are typically much longer than the GRACE/-FO record, albeit they can differ substantially (Hacker and Kusche, 2024). Reconstruction methods enable spatial downscaling (Gou and Soja, 2024) and causal inference (Nowack et al., 2020). On the downside, uncertainty information may be missing or ambiguous since it is challenging and complex to represent the various types of uncertainty in machine learning predictions and to assess the effect of unknown climate data biases. Another problem is that the GRACE/FO training data contains human water use, meaning that learned relations between climate and water storage can almost certainly not be transferred straightforwardly to the past.

On the other hand, satellite-geodetic tracking data enable the retrieval of the gravity field and, thus, water storage changes in the pre-GRACE era. These data are entirely independent of climate data and do not require any assumptions inherent to regression and machine learning; however, their spatial resolution is limited. Laser ranges measured to passive spherical satellites (satellite laser ranging, SLR) have been reanalyzed in several studies (cf. Galdyn et al. (2024)). The resulting spherical harmonic solutions have a maximum spherical harmonic degree of 5–10, corresponding to the resolution of several thousands of kilometers. Other approaches, such as GNSS tracking to active satellites in low Earth orbit (Weigelt et al., 2024) or inverting the elastic deformation experienced by global networks of ground stations (Nowak et al., 2025), provide little information prior to the GRACE era. Recently, some studies (Löcher and Kusche, 2021; Cheng and Ries, 2023; Löcher et al., 2025b) have turned to directly evaluating the SLR ranges in the EOF space spanned by the leading modes of the GRACE/-FO TWSA record, instead of the conventional analysis in spherical harmonics, to alleviate spatial aliasing and retain more signals in their retrievals. Löcher et al. (2025b) showed that this approach, albeit not free of constraints inherited from the GRACE/-FO period, allowed the integration of ranging data with the DORIS system and stabilized the SLR TWSA retrievals in the mid-80s.

In this study, we present a first attempt to combine low-resolution geodetic TWSA records obtained from SLR DORIS with climate data regression, resulting in a four-decade reconstruction of GRACE-like TWSA data over the global land. Our aim is thus to close the resolution gap of the geodetic data. We anticipate that our new synthesis data set TWSTORE will be more long-term consistent than the 'pure' reconstructions. We base our combination approach on the variance component estimation



method (Koch and Kusche, 2002; Förstner, 1979) to overcome the problem of inconsistent uncertainty representation. Instead of performing the combination in the spherical harmonic basis or grid space, we first project both data sets into a GRACE/FO EOF space, which has been utilized in both the original SLR/DORIS analysis (Löcher and Kusche, 2021) and in our reconstruction based on climate data. Our ensemble reconstruction method is similar to Li et al. (2021), but it uses a Random Forest (RF) instead of an artificial neural network, in addition to the autoregressive (ARX) and multilinear regression (MLR) approaches. Furthermore, in addition to correlation, we employ Granger causality to rank the predictors according to their impact on GRACE.

We provide a reconstruction on a  $0.5^\circ$  grid with a monthly temporal resolution from 1984 onwards. Similar to the original GRACE data, the grid resolution of the data set is higher than the actual resolution. The effective resolution of the reconstruction aligns with the GRACE data, which is about 330 km. For the time frame from 2002 onward, we compare the reconstruction to GRACE and GRACE-FO (referred to as GRACE/FO hereafter) data. We also compare it to existing reconstructions from Humphrey and Gudmundsson (2019b) (HG19), Li et al. (2021) (Li21), and Chandanpurkar et al. (2022b) (CHR22) to access the reliability of the long term TWSA signal. We find a high degree of agreement regarding the annual, interannual, and subseasonal signals. However, our reconstructions exhibit strong accelerations that are only partly reflected in the other reconstructions. Further analysis suggests that the strong acceleration signal is inherited from SLR.

To investigate the occurrence probability of anomalous monthly water storage deficits or surpluses, we then fit a generalized extreme value (GEV) distribution to the storage maps after removing the climatology. The fitted GEV distribution enables us to estimate return levels for anomalously high and low TWSA events, similar to Kusche et al. (2016). Furthermore, we derive multi-decadal area-averaged evapotranspiration from reconstructed TWSA changes for a selection of river catchments using terrestrial water budgets. The derived time series for evapotranspiration are available at 10.5281/zenodo.16643628 (Gutknecht, 2025).

## 2 Methods

In this study, we develop a modular approach to create a multi-decadal reconstruction of TWSA, which combines (1) geodetic satellite tracking data prior to the launch of GRACE and (2) climate data time series. While the link to terrestrial water storage maps in (1) is obtained from physical/geodetic modeling, and data uncertainties are propagated from tracking misfits, in (2), the link is established by training regression or random forest models, and uncertainties are derived from an ensemble approach. The two methods are summarized below. Thereon, we describe how we combine the outcome of both in a statistically optimal way, taking into account that the two uncertainty representations are not entirely consistent with each other.

### 2.1 Deriving large-scale maps of terrestrial water storage from geodetic SLR and DORIS data analysis

In satellite laser ranging (SLR), passive spherical satellites have been tracked from a network of ground stations since the mid-1970s. This technique is well-established for measuring Earth rotation variations and determining a global reference frame, including its origin. From satellite orbit perturbations, time-variable gravity field models were derived and large-scale mass



changes in the Earth system could be quantified, e.g., Cox and Chao (2002). However, due to the slow evolution of the ground station network and the technology on the stations, uncertainties were considerable in the early years. An excellent review is provided in Pearlman et al. (2019).

In an attempt to stabilize early gravity field solutions, Löcher et al. (2025b) suggest adding more satellites to their analysis, which are tracked through the DORIS system. After applying correction steps consistent with the GRACE/-FO data reduction, e.g., removing non-tidal atmospheric and ocean mass variations using the same background models, monthly GRACE-like maps of terrestrial water storage variation,  $\sigma(\lambda, \phi)$  can be derived as

$$\sigma(\lambda, \phi, t) = L(\lambda, \phi) \frac{3a}{\rho_e} \sum_{n=1}^{\bar{n}} \sum_m \frac{1+k'_n}{2n+1} (c_{nm} \cos m\lambda + s_{nm} \sin m\lambda) P_{nm}(\sin \phi) \quad (1)$$

where  $L = 1 - O$  is the land function (i.e. 1 over land, 0 over ocean),  $a$  the Earth's radius,  $\rho_e = \frac{3M_e}{4\pi a^3}$  Earth's mean density,  $k'_n$  the potential load Love number of harmonic degree  $n$ ,  $P_{nm}$  the fully normalized associated Legendre functions,  $m$  the harmonic order, and  $c_{nm}(t)$  and  $s_{nm}(t)$  are the fully normalized Stokes coefficients retrieved from the orbit reconstruction of geodetic (SLR and/or DORIS) satellites, with the temporal mean removed. In the above, the truncation degree  $\bar{n}$  for SLR is typically low, e.g., 5-10. This limited resolution is due to the high satellite altitudes (compared to GRACE), insufficient tracking network coverage, and measurement errors, which all translate to a spatial resolution of several thousand kilometers. From the SLR and DORIS analysis, one obtains a variance-covariance matrix of the  $c_{nm}$ , which indeed represents the uncertainty introduced by the geodetic tracking systems, i.e., the effect of the errors of the tracking instruments, SLR and DORIS range-rate errors, given the particular network configuration and coverage.

The spherical harmonics basis is not optimal for combining with a climate-variables-driven reconstruction of TWSA, which by definition exists only over land. Therefore, we begin with the gridded maps from Eq. (1) and facilitate a transformation into a spatial EOF basis. After this step, we find in the GRACE EOF basis (cf. Löcher and Kusche (2021), Fig. 2)

$$\sigma(\lambda, \phi, t) = \sum_{i=1}^m c_i(t) \sigma_i(\lambda, \phi) \quad (2)$$

In the above equation,  $\sigma_i(\lambda, \phi)$  refers to the normalized eigenvectors of the signal covariance. The signal covariance often resembles spatial patterns affected by independent large-scale climate modes. However, due to the presence of common drivers and real teleconnections in the climate system, these patterns are often non-contiguous. We subdivide the land mass into hydrological river basins,  $L = L_1 \cup L_2 \cup \dots \cup L_{\bar{p}}$ , and then derive a finite number of EOF modes for each basin. Subdividing the global land mass leads to a more localized yet physically motivated representation, as the EOFs capture the dominant spatial patterns of the  $p$ -th catchment, which can be written as follows

$$\sigma(\lambda, \phi, t) = \sum_{p=1}^{\bar{p}} L_p(\lambda, \phi) \sigma^p(\lambda, \phi, t) = \sum_{p=1}^{\bar{p}} \sum_{i=1}^m c_i^p(t) \sigma_i^p(\lambda, \phi) . \quad (3)$$

We obtained the covariance matrix for the SLR data by propagating the given variances of the spherical harmonics coefficients. Similar to the spherical harmonic coefficients, their variance-covariance matrix can be projected onto the basin EOF space. The



projected uncertainties for the SLR/DORIS-derived time series per basin are then consistent with the representation in Eq. (3). The limited spatial resolution of SLR and DORIS data leads to large-scale spatial error correlation. As a consequence of our  
 125 basin-wise approach, the large-scale correlations between SLR and DORIS are mapped to spatial correlations within a basin but not across basins. In other words,

$$\mathbf{C}(c_i^p, c_j^p) = \mathbf{\Sigma} \mathbf{C}(\lambda_i, \phi_i, \lambda_j, \phi_j) \mathbf{\Sigma}^T$$

$$= \begin{bmatrix} V(c_1^p) & C(c_1^p, c_2^p) & \cdots & C(c_1^p, c_m^p) \\ C(c_2^p, c_1^p) & V(c_2^p) & \cdots & C(c_2^p, c_m^p) \\ \vdots & \ddots & \ddots & \vdots \\ C(c_m^p, c_1^p) & \cdots & C(c_m^p, c_{m-1}^p) & V(c_m^p) \end{bmatrix}. \quad (4)$$

where  $\mathbf{\Sigma}$  is a matrix with entries  $\sigma_i^p(\lambda, \phi)$  for the  $p$ -th basin.

## 130 2.2 Deriving maps of terrestrial water storage from climate data

Recent studies have employed statistical or Machine Learning (ML) methods to predict TWSA maps from a time series of climate inputs or predictors. Predicted TWSA maps are generally derived from a data-based model fitted to real GRACE TWSA data during a training step (Forootan et al., 2014; Humphrey and Gudmundsson, 2019b; Hacker and Kusche, 2024). While not aiming at physical consistency, these approaches have been demonstrated to outperform hydrological and land  
 135 surface modeling when the aim was to predict TWSA during unobserved periods.

Reconstructing TWSA from a given set of predictors requires learning a relationship between the target variable and the predictors. The relationship is described by linear or nonlinear statistical models and/or machine learning (ML) architecture. A set of weights inherent to the respective method is estimated during a selected time frame when both the target variable and predictors are available; i.e. the training period. While early approaches derived TWSA maps on a grid-scale, e.g., individual  
 140 grid points or neighborhoods (in CNN architectures), we follow Li et al. (2019), who suggested describing the reconstruction in an EOF space and reconstructing the dominant temporal modes. It is then generally possible to express the reconstructed fields for  $m$  dominant modes in basin  $p$  as

$$\sigma^p(\lambda, \theta, t) = \sum_{j=1}^m \underbrace{f_j^A(\mathbf{p}(t), \mathbf{p}(t-1), \dots, \Theta_j)}_{\hat{c}_j(t)} \sigma_j^p(\lambda, \theta) \quad (5)$$

where  $f_j^A$  relates to the chosen algorithm/architecture (see below), the vector  $\mathbf{p}(t), \dots, \mathbf{p}(t-n)$  refers to the specific predictor  
 145 time series and  $\Theta_j$  denotes the fitted weights or parameters. For example, if precipitation anomalies are chosen as a predictor for TWSA, the corresponding entry of  $\mathbf{p}$  contains the time series of these anomalies over the river basin in question and after projection onto the GRACE-derived EOF modes. In this study, we use three different algorithms to identify the relationship between GRACE TWSA and the predictor time-series, i.e.



1. Autoregressive process with exogenous variables (ARX). An ARX model (Ljung, 1987) is a linear, recursive filter that links past observations of the predictand to external, independent input variables. In our case, it links predicted TWSA time series per EOF mode to multiple inputs (predictors and past predicted TWSA, for the same EOF mode) (Forootan et al., 2014)

$$y(t) + \sum_{i=1}^{n_a} a_i y(t - t_i) = \sum_{q=1}^m \sum_{l=1}^{n_b} b_{q,l} p_q(t - k_q - (l - 1)) \quad (6)$$

where  $n_a$  is the order of the ARX model concerning the predictand,  $n_b$  is the order for the predictors, and  $m$  is the total amount of exogenous variables used. Due to the stationarity assumption of the ARX model, the derived model coefficients are constant over the reconstruction period. The coefficients are estimated using a least squares adjustment during the training period. The coefficient  $k_q$  allows considering a time lag between the observations and the predictors. Once the coefficients  $a_i$  and  $b_{q,l}$  are estimated the reconstructed TWSA are given as

$$f_j^A = f^{ARX} : \tilde{c}(t) = - \sum_{i=1}^{n_a} \tilde{a}_i \tilde{c}(t - t_i) + \sum_{q=1}^m \sum_{l=1}^{n_b} \tilde{b}_{q,l} p_q(t - k_q - (l - 1)) \quad (7)$$

2. Multiple linear regression (MLR). Multiple linear regression is a statistical technique used to model the linear relationship between a single dependent variable and two or more independent variables. The target variable is the given as:

$$f_j^A = f^{MLR} : c(t) = p_1(t)\alpha_1 + p_2(t)\alpha_2 + \dots + p_n(t)\alpha_n \quad (8)$$

with  $p_1(t) \dots p_n(t)$  denoting the predictors and  $\Theta_j = [\alpha_1(t) \dots \alpha_n(t)]$  the corresponding scaling factors.

3. Random forest. Random Forest is an ensemble learning method. It builds multiple decision trees (the forest) and combines their output to model complex, nonlinear relationships. Each tree is trained on a random subset of the target data and predictors, also called features, which reduces overfitting and improves generalization (Breiman, 2001). The reconstructed PCs are given as

$$f_j^A = f^{RF} : \frac{1}{B} \sum_{b=1}^B f_b(p(t)) \quad (9)$$

with  $B$  indicating the tree built by  $f_b(p(t))$ .

In our approach, this step is conducted independently of the SLR/DORIS retrieval mentioned above. Thus, any other published TWSA could act as a replacement. To obtain a preliminary reconstruction, we first reduce the input data, GRACE, and the predictors to a given partitioning of the global land excluding Greenland and Antarctica (cf. A1). The TWSA EOF analysis is performed for every polygon, and the first  $m$  modes are selected. The number of modes  $m$  is determined such that the dominating modes explain  $> 95\%$  of the signal. The predictors are then projected onto the GRACE EOF basis (section 2.3 and A1), reducing them to temporal modes. We use correlation and nonlinear Granger causality (Granger (1969); Papagiannopoulou



et al. (2017) and section A3) to identify the most important predictors for each polygon. We establish a relationship between the GRACE principal components and the projected principal components of the predictor data using either multiple linear regression (MLR), autoregressive with exogenous inputs (ARX), or random forest (RF). The found relationship is then used to generate reconstructed temporal modes. We recover the full, reconstructed signal by remapping the reconstructed principal components (PCs) using the GRACE empirical orthogonal functions (EOFs). We used different data sets for the predictors and GRACE solutions (section 3) to assess robustness by generating an ensemble and to represent both aleatoric (statistical) and epistemic uncertainty. Finally, we derive the preliminary reconstruction as the mean over all ensemble members. From the ensemble spread, we estimate an error covariance matrix of the preliminary reconstruction. However, when we derive the full error covariance of the preliminary reconstruction from the reconstruction ensemble, we encounter a rank deficit due to the limited number of realizations. Therefore, we focus on the variances and disregard correlations between grid points. Similarly to Section 2.1, we project the errors of the preliminary reconstruction onto the GRACE-derived EOF basis. The outcome is

$$\mathbf{C}(c_i^p, c_j^p) = \Sigma \mathbf{C}(\lambda_i, \phi_i, \lambda_j, \phi_j) \Sigma^T$$

$$= \begin{bmatrix} V(c_1^p) & C(c_1^p, c_2^p) & \cdots & C(c_1^p, c_m^p) \\ C(c_2^p, c_1^p) & V(c_2^p) & \cdots & C(c_2^p, c_m^p) \\ \vdots & \ddots & \ddots & \vdots \\ C(c_m^p, c_1^p) & \cdots & C(c_m^p, c_{m-1}^p) & V(c_m^p) \end{bmatrix}. \quad (10)$$

As a result of the projection, the resulting covariance matrix is fully populated; however, it is important to note that the covariances only account for correlations arising from the projection operation.

### 2.3 Optimal data combination in EOF space

As mentioned in section 2.2, our analyses include an EOF decomposition on the GRACE TWSA data. Both climate predictors and the SLR-DORIS data are projected onto the  $m$  first GRACE-derived EOF basis, which provides the  $c_i^p$  time series in Eq. (3) and Eq. (5). Therefore, the projection allows the description of the reconstruction and data combination within a single orthogonal basis per basin. The common GRACE-derived EOF basis provides a shared description of spatial variability, making it easier to relate variations in the predictors to the patterns that drive mass variability observed by GRACE. Additionally, noise and correlations within and between the predictors and the target variables are reduced. From the projection, we obtain the PCs, the temporal evolution of the GRACE data, predictors, and SLR-DORIS data. Within the reconstruction, the PCs of the reconstruction are linked to the PCs of GRACE. However, depending on the climate regime, the importance of predictors differs, making it difficult to select "the best" predictors for a specific area. To overcome this issue, we use correlation and Granger causality (Granger (1969); Papagiannopoulou et al. (2017) and section A3) to determine which predictor has the highest impact on TWSA in a selected region.

The SLR/DORIS data and the preliminary reconstruction are finally combined via variance component estimation (VCE) ((Koch, 2018; Förstner, 1979) and section A2). The estimated variance components scale the observation covariance matrix (Eq. (4) and (10)) and can be interpreted as weights among the observation groups contributing to the overall solution. Similar



| GRACE product          | Time period                 | Spatial resolution                                          | GIA correction        | Access                                                                                                                                                                            | Citation                                      |
|------------------------|-----------------------------|-------------------------------------------------------------|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| ITSG-Grace2018         | April 2002 to December 2020 | order and degree of 96 synthesized on a $0.5^\circ$ grid    | Peltier et al. (2015) | <a href="https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018">https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018</a> | Kvas et al. (2019); Mayer-Gürr et al. (2018b) |
| COST-GRL01             | April 2002 to December 2020 | order and degree of 90 synthesized on a $0.5^\circ$ grid    | Peltier et al. (2015) | <a href="https://doi.org/10.5880/ICGEM.COST-G.001">https://doi.org/10.5880/ICGEM.COST-G.001</a>                                                                                   | (Jäggi et al., 2022)                          |
| GSFC mascons RL06 v2.0 | April 2002 to December 2020 | $1^\circ$ equal-area mascons, sampled on a $0.5^\circ$ grid | Peltier et al. (2015) | <a href="https://earth.gsfc.nasa.gov/geo/data/grace-mascons">https://earth.gsfc.nasa.gov/geo/data/grace-mascons</a>                                                               | Loomis et al. (2019)                          |

**Table 1.** Overview of used GRACE products

to the reconstruction, we apply the VCE to the principal components derived from SLR and the preliminary reconstruction by mapping the data and covariance matrices onto a GRACE-derived EOF basis. We set the a-priori variance factor for all data sets to one. Due to the lower resolution of the SLR data compared to GRACE, we aggregate the polygons used for the reconstruction to "super" polygons (figure B1). The data combination is performed for every super polygon's first  $m$  dominant modes. The number of dominant modes per polygon is illustrated in figure C1. The derived variance factors are adjusted by a factor of up to about 10 for SLR and up to about 1000, depending on the mode and basin, for the reconstruction, leading to a higher influence of SLR on the final solution. After deriving and applying the variance components, we remap the adjusted PCs using the GRACE EOF basis.

We derive variances for TWSTORE by propagating the derived uncertainties of SLR/DORIS and the climate data reconstruction through our combination approach. The averaged standard deviations exhibit uncertainties around 2-6 cm per grid cell. Higher uncertainties are found in regions with greater variability in the TWSA signal (Figure D1).

### 3 Data

Three data types are important for this study: GRACE data, the climate data used as predictors, and the SLR/DORIS data. The GRACE level-2 data is converted to TWSA maps. They serve to generate the EOF bases, in which the outcome of the preliminary reconstruction and the SLR/DORIS data are represented and combined, and for training the preliminary reconstruction in Section 2.2. After describing the GRACE postprocessing, we summarize the climate data and the SLR/DORIS data set used in this study.



### 3.1 GRACE products

225 In this study, we used three different GRACE solutions: monthly ITSG-Grace2018 gravity fields computed at Graz University of Technology (Kvas et al., 2019; Mayer-Gürr et al., 2018b) of order and degree 96 for the years April 2002 to December 2020, the International Combination Service for Time-variable Gravity Fields (COST-G) RL01 solution from the University of Bern (Jäggi et al., 2022) of order and degree 90 for the years April 2002 to December 2020 and the NASA GSFC global mascon product on a 0.5 x 0.5-degree grid (table 1). For the two L2 products (ITSG-Grace2018 and COST-G), we reduced the  
 230 temporal mean from 2003 to 2020 from the spherical harmonics (geopotential) coefficients. We replaced the degree 1 and  $c_{20}$  coefficients with those from NASA/JPL (Cheng and Ries, 2023; Swenson et al., 2008), as GRACE measures these coefficients poorly. Following Loomis et al. (2020), we also replaced the  $c_{30}$  coefficient for the late GRACE months, compensating for the degrading estimates from GRACE due to accelerator problems. The coefficients were smoothed with the anisotropic DDK3 filter (Kusche, 2007; Kusche et al., 2009) to reduce the noise level of the GRACE-derived coefficients. The post-glacial rebound  
 235 (glacial isostatic adjustment (GIA)) of the Earth causes a redistribution of interior Earth mass, leading to a long-term gravity trend unrelated to water storage variations. The effect of GIA was removed from the coefficients. We synthesized the gravity fields (Wahr et al., 1998) from the coefficients on a 0.5 x 0.5-degree grid and interpolated missing months using Akima Spline interpolation. It should be noted that the actual resolution of GRACE is much coarser (around 330 km at the equator) and is not reflected by the grid (around 55 km for a 0.5-degree grid) (Humphrey et al., 2023).

### 240 3.2 Climate data

We used five different predictors, all of which were provided in gridded fields. Two of these predictors were related to water cycle variables (precipitation,  $P$ , and soil moisture,  $SM$ ), two to the energy cycle (sea surface temperature,  $SST$ , and air temperature,  $T$ ), and one to the carbon cycle (leaf area index,  $LAI$ ).

$SST$  and  $SM$  influence atmospheric moisture and heat transfer. The variables are closely connected to evapotranspiration rates  
 245 and rainfall. High  $SST$  values lead to high air moisture content and heavy rainfalls, which can cause floods.  $SM$  determines plant water availability, which in turn influences plant growth and  $CO_2$  uptake. Low  $SM$  values can indicate droughts, whereas high soil moisture contents can cause higher runoff rates and possibly floods.  $T$  is the driver for the movement of water masses, influencing the phase changes of water (evaporation, condensation, freezing) and energy fluxes.  $P$  patterns determine regional climates and ecosystems, replenishing freshwater supplies and maintaining  $SM$ . Like  $SM$ ,  $P$  is linked to droughts and flood  
 250 events.  $LAI$  is an indicator of water uptake by plants, evapotranspiration, and photosynthesis, which connects the water cycle to the carbon cycle.

We refrain from including climate indices here. Although they have been helpful in studies aimed at optimizing prediction skills, they are compounded by the variables mentioned, and we expect them to obscure our causality framework. Moreover, some are derived directly from  $SST$ .

255 The data sets used in this study are listed in table 2. We use two datasets for  $SST$ ,  $P$ , and  $SM$  and one dataset for  $T$  and  $LAI$ . For  $SST$ , we utilized the Met Office Hadley Centre's sea ice and sea surface temperature (SST) dataset, HadISST1, and the



| Data set                      | Time period  | Spatial resolution       | Description                                                                                                                                                                                         | Access                                                                                                                                          | Citation                    |
|-------------------------------|--------------|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| HadISST1                      | 1870 to 2024 | 1° grid                  | Reconstructed monthly, global sea ice and SST fields using a two-stage reduced-space optimal interpolation procedure. Observations allow to restore local detail                                    | <a href="https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html">https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html</a>     | Rayner et al. (2003)        |
| COBE-SST 2 and Sea Ice        | 1850 to 2024 | 1° grid                  | Constructed daily SST fields using in situ SST and sea ice observations, corrected for biases, incorporating satellite data for sparse regions, and providing uncertainty estimates for reliability | <a href="https://psl.noaa.gov/data/gridded/data.cobe2.html">https://psl.noaa.gov/data/gridded/data.cobe2.html</a>                               | Hirahara et al. (2014)      |
| GPCC Version 2022             | 1891 to 2020 | 0.5° grid                | Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data                                                                                                          | <a href="https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2022_050">https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2022_050</a>                               | Schneider et al. (2022)     |
| CPC global precipitation V1.0 | 1971 to 2024 | 0.5° grid                | Global Unified Gauge-Based Analysis of Daily Precipitation                                                                                                                                          | <a href="https://www.psl.noaa.gov/data/gridded/data.cpc.globalprecip.html">https://www.psl.noaa.gov/data/gridded/data.cpc.globalprecip.html</a> | Chen et al. (2008)          |
| CPC Soil Moisture v2          | 1948 to 2024 | 0.5° grid                | Modeled monthly averaged soil moisture water height equivalents                                                                                                                                     | <a href="https://www.psl.noaa.gov/data/gridded/data.cpcsoil.html">https://www.psl.noaa.gov/data/gridded/data.cpcsoil.html</a>                   | Fan and van den Dool (2004) |
| GLEAM Soil Moisture v4.1a     | 1980 to 2023 | 0.1° grid                | Modeled root-zone soil moisture enhanced by assimilation of satellite observations of surface soil moisture                                                                                         | <a href="https://www.gleam.eu/">https://www.gleam.eu/</a>                                                                                       | Miralles et al. (2025)      |
| GHCN CAMS Monthly Temperature | 1948 to 2024 | 0.5° grid                | Gridded 2 m temperature over land                                                                                                                                                                   | <a href="https://www.psl.noaa.gov/data/gridded/data.ghcncams.html">https://www.psl.noaa.gov/data/gridded/data.ghcncams.html</a>                 | Fan and van den Dool (2008) |
| Globmap LAI v3                | 1981 to 2020 | 8km resolution (≈ 0.05°) | Fusion of Moderate Resolution Imaging Spectroradiometer (MODIS) and historical Advanced Very High Resolution Radiometer (AVHRR) data                                                                | <a href="https://doi.org/10.5281/zenodo.4700264">https://doi.org/10.5281/zenodo.4700264</a>                                                     | Liu et al. (2012)           |

**Table 2.** Overview of the climate data sets used in the preliminary prediction



COBE-SST 2 and Sea Ice dataset provided by the National Oceanic and Atmospheric Administration (NOAA) Physical Sciences Laboratory (PSL). HadISST is global monthly averaged SST anomalies derived from in situ ship-based measurements, drifting and moored buoys, and satellite observations. HadISST1 employs interpolation techniques and statistical methods to fill data gaps. COBE-SST 2, in contrast, does not include satellite observations. The data set focuses on in situ data. SST values are interpolated by minimizing the differences between observed and reconstructed SST values. For precipitation, we utilize the Global Precipitation Climatology Centre (GPCC) dataset, operated by the German Meteorological Service (Deutscher Wetterdienst, DWD), and the Climate Prediction Center's (CPC) Global Precipitation V1.0 Data Set, provided by NOAA. The GPCC data set is based on in situ observations from over 85,000 rain gauges worldwide, provided by national meteorological and hydrological services (NMHSs), global data collections (e.g., SYNOP), and research networks. The CPC data set, provided by NOAA, combines in situ gauge-based precipitation observations with satellite-derived precipitation estimates. The CPC soil moisture data set is not directly measured but rather modeled using a water balance model that incorporates precipitation data from NOAA's CPC Unified Precipitation dataset and temperature inputs from NCEP-NCAR reanalysis data. GHCN CAMS Monthly Temperature data is derived from in situ observations from surface weather stations that report to the Global Historical Climatology Network (GHCN). Globmap LAI v3 is based on satellite observations using data from sensors such as the AVHRR (Advanced Very High-Resolution Radiometer) (pre-2000) and MODIS (Moderate Resolution Imaging Spectroradiometer) (after 2000).

All data sets were trimmed to the period 1984-2020, averaged or summed to monthly data, and, except for the SST data, resampled to a  $0.5^\circ$  grid. Missing data points (due to different land masks between GRACE and the predictors) are interpolated.

### 3.3 Large-scale gravity fields from SLR and DORIS

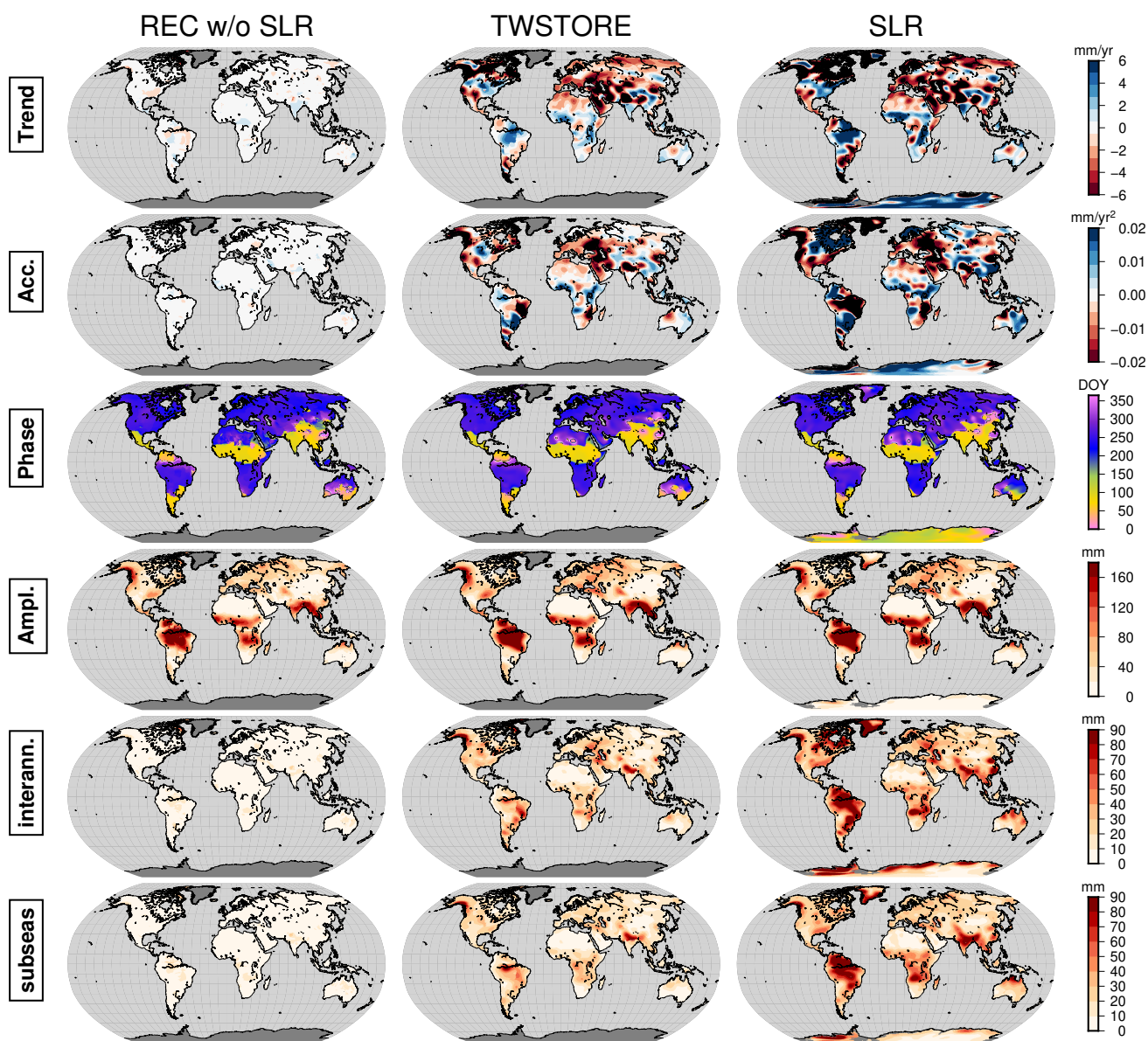
There are only a few reconstructions of pre-GRACE terrestrial water storage maps. In this study, we utilize the solution proposed by Löcher et al. (2025b). The dataset is a combination of satellite laser ranges from six satellites and ten satellites tracked by Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS). For pre-1992 full-rate laser ranging data, normal points were formed as described in Löcher et al. (2025b). The range rates derived from the Doppler observations were integrated into biased ranges. The observations were combined at the level of the normal equations and solved on the basis of the EOF derived from GRACE. Eventually, the results were subsequently converted to SHC complete to degree and order 60.

## 4 Evaluation of the combined reconstruction

Following, we will assess the impact of the combination procedure by comparing it to the SLR/DORIS and preliminary climate data reconstructions. Afterward, we compare terrestrial water storage maps from GRACE-FO, which was not used in the training process, and finally, we compare our reconstruction, TWSTORE, to other published reconstructions. TWSTORE and associated uncertainties are available as NetCDF at <https://doi.org/10.5281/zenodo.15827789> (Hacker, 2025).



#### 4.1 Impact of the data combination



**Figure 1.** Trend, acceleration, annual amplitude and phase, interannual and subseasonal signal of the preliminary reconstruction (without SLR), TWSTORE and the SLR fields for a spherical harmonic degree of  $n = 60$

290 Our terrestrial water storage reconstruction is, to our knowledge, the first attempt to integrate satellite-geodetic records with data-based methods. In this section, we therefore assess the impact of the data combination (section 2.3). We analyze the trend,

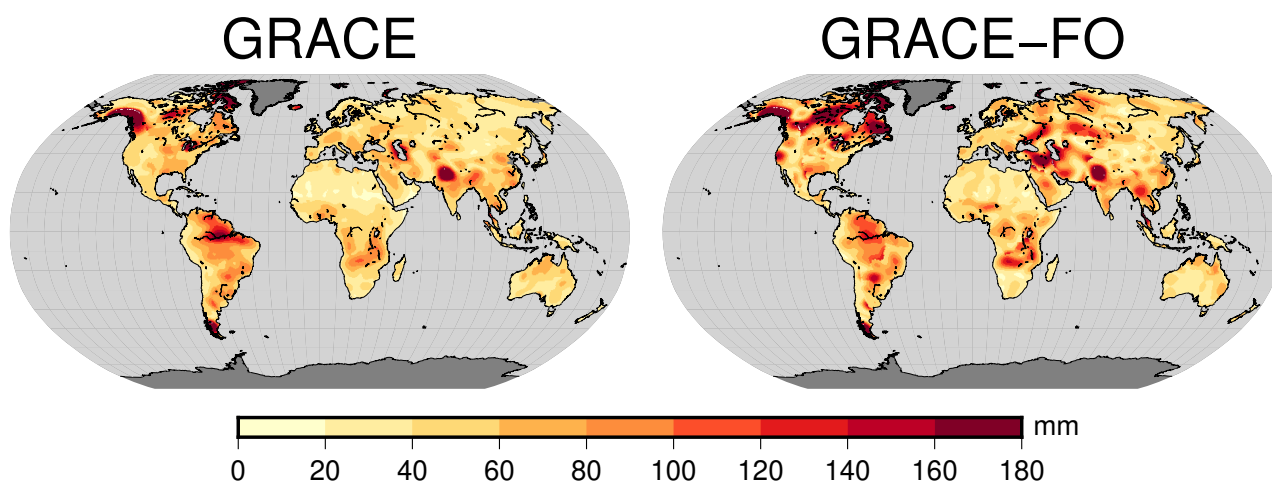


acceleration, annual amplitude and phase, inter-annual, and sub-seasonal signal for 1984-2020 for the preliminary reconstruction from climate data without integrating SLR-DORIS (REC w/o SLR), the SLR-DORIS solution (SLR), and the combined reconstruction (TWSTORE) (cf. Fig. 1). We expect that the influence of the satellite-geodetic data set will vary by region, as  
295 its weight in the combination is controlled by both the uncertainty description of the SLR/DORIS covariance matrix and the ensemble of the preliminary reconstruction.

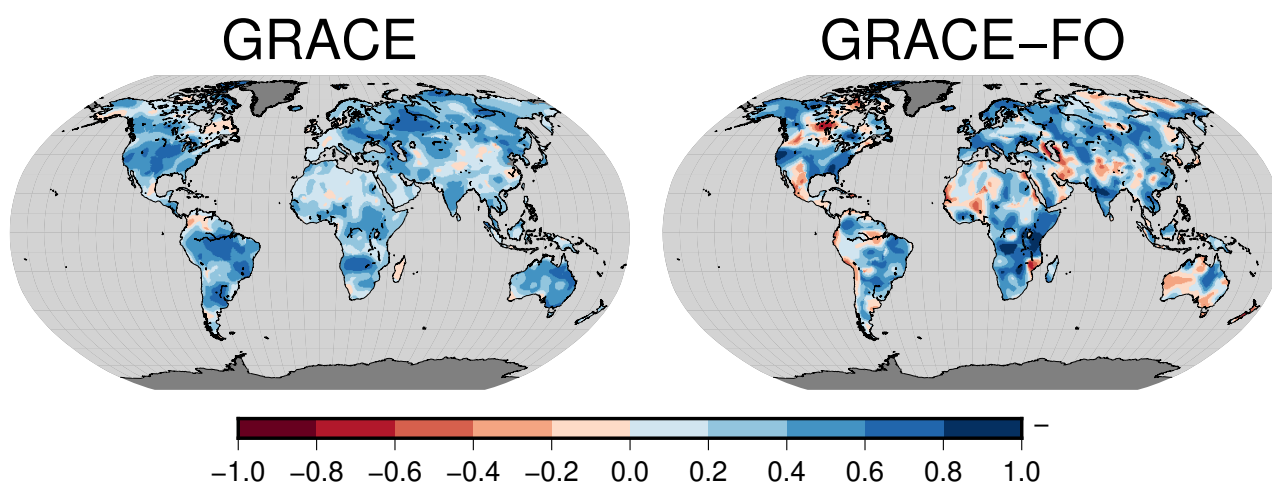
The SLR/DORIS TWSA fields exhibit strong signal magnitudes across all metrics. On the other side, the preliminary reconstruction displays smaller magnitudes in all metrics except for the annual amplitude and phase, which are consistent with the satellite-geodetic solution. The annual cycle in the preliminary reconstruction is well captured and thus only slightly adjusted  
300 by the data combination. Compared to the climate data reconstruction, we find that the data combination for Europe and Africa moderately magnifies the magnitude of the water storage change signal. In contrast, the trend rate and acceleration in our reconstruction (TWSTORE) are controlled by the SLR/DORIS solution. This effect is desired since the tracking data should be superior to climate data reconstruction at these timescales. However, the water storage trends displayed by TWSTORE show a regional dependency: For instance, the trend for the Arabian Peninsula appears consistent between SLR and TWSTORE, while  
305 in contrast, in Africa, it seems somewhat lower for TWSTORE compared to SLR. We find a similar pattern for the acceleration. It is essential to note that the weighting between climate-driven and geodetic reconstruction is determined by the ensemble uncertainty of the preliminary reconstruction in comparison to the SLR/DORIS uncertainty, and this uncertainty varies by basin. The influence of SLR/DORIS is less pronounced for inter-annual and sub-seasonal timescales than for trend and acceleration. Nevertheless, a noticeable increase in signal magnitude can be observed in the inter-annual and sub-seasonal components when  
310 comparing the preliminary reconstruction to TWSTORE. It is also worth noting that the SLR/DORIS contribution led to higher uncertainty from 1984 to 1992, as these observations were noisier during this period (Löcher et al., 2025b).



## 4.2 Comparison against GRACE/FO (2002-2020)



**Figure 2.** Root Mean square deviation between GRACE/FO and the reconstruction



**Figure 3.** Correlation (of de-seasonalized, de-trended anomalies) between GRACE/FO and the reconstruction

We first compare the reconstruction with terrestrial water storage anomalies from the GRACE mission. However, this cannot be considered an independent evaluation, as our preliminary reconstruction was trained using the GRACE data. Nevertheless, comparing the performance of the reconstruction with both GRACE and GRACE-FO provides an impression of the reconstruction's prediction skill.

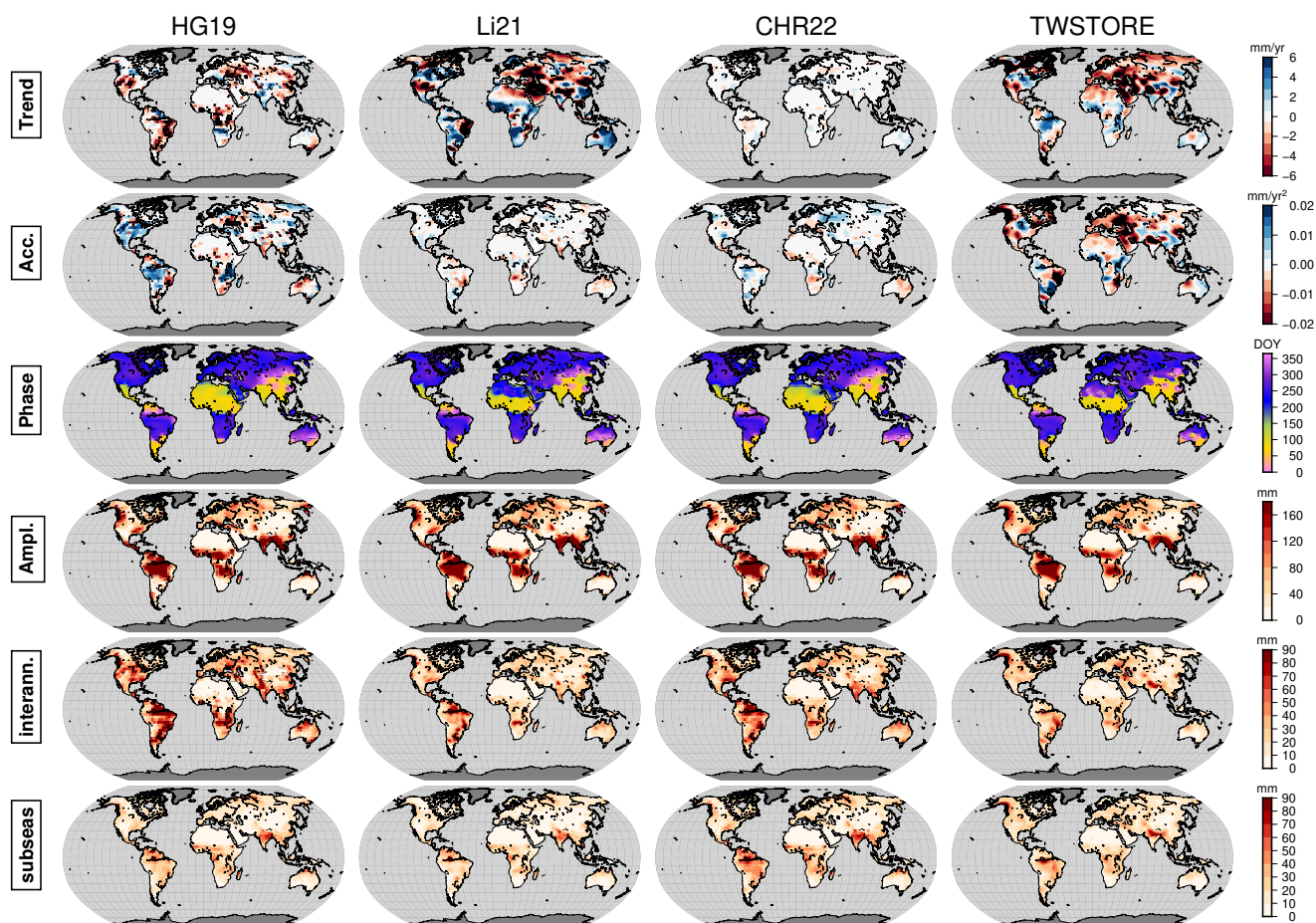


Figure 3 shows the Pearson correlation coefficient between the detrended and deseasonalized data sets, specifically TWSTORE vs. GRACE (January 2003 - December 2016) and TWSTORE vs. GRACE-FO (July 2018 - December 2020). We generally find lower correlations for arid and highland climate regions. The low correlation may be attributed to the small TWSA signal and the inability of the predictors, primarily precipitation, soil moisture, and leaf area index, to accurately represent the changes in TWSA in these regions. On the other hand, for regions with a humid climate, we find a high correlation. As may be expected, when advancing the reconstruction to the more recent GRACE-FO data, we find a loss in correlation compared to GRACE, most pronounced in Australia, the Amazon basin, the region around the Caspian Sea, and the Sahara. Although not optimal, a slight phase shift can be expected when predicting a time series.

Figure 2 illustrates the root mean square deviation (RMSD) between the reconstruction and GRACE/FO terrestrial water storage. We find lower skills and higher RMSD values in Alpine and highland regions, which we attribute to the reconstruction's limited ability to describe TWSA in the presence of glacier and snowpack changes. Compared to GRACE-FO, high RMSD values are observed in the regions around Hudson Bay and the Gulf of Alaska. The high values could indicate a mismatch in the GIA correction, but this requires further investigation. However, the RMSD values for Fennoscandia and Northern Europe are quite small. Notable are also the lower prediction skills for Turkey, the region around the Caspian Sea, the Volga-Don River, and the area around the Aral Sea. We suspect that this could be due to anthropogenic effects, such as damming activities or irrigation with river water, which are not or are only partially represented in the reconstruction.



### 4.3 Comparison to other reconstructions



**Figure 4.** Trend, acceleration, annual amplitude and phase, sub-seasonal and inter-annual signal for HG19, Li21, CHR22, the reconstruction and the SLR-DORIS data set for 1984–2019 on a global scale.

A comparison with GRACE data offers limited insight into the performance of the reconstruction, as it cannot be considered an independent dataset. Point-wise validation data is also unusable due to the resolution of the reconstruction, and the SLR data from Löcher and Kusche (2021) is not independent either. Therefore, we provide a comparison with other GRACE-like TWSA reconstructions. The comparison with other reconstructions can be seen as an internal evaluation, determining consistency across different data sets. The employed reconstruction differs in terms of used data sets and reconstruction principle (Humphrey and Gudmundsson, 2019b; Li et al., 2021; Chandanpurkar et al., 2022b). We derived a trend, acceleration, annual amplitude, and phase, sub-seasonal, and interannual signals from the reconstruction and the SLR-DORIS data set used in the reconstruction for 1984 - 2020. An in-depth evaluation of the used reconstructions can be found in Hacker and Kusche (2024).



The reconstructions agree on negative trends in regions affected by (historical) droughts, like the US and the Arabian Peninsula, regions affected by ice mass loss, e.g., Siberia and the Andes, and regions affected by water loss due to anthropogenic effects, such as the region around the Caspian Sea and Aral Sea. Notable differences arise for the Amazon basin, Africa, and China.

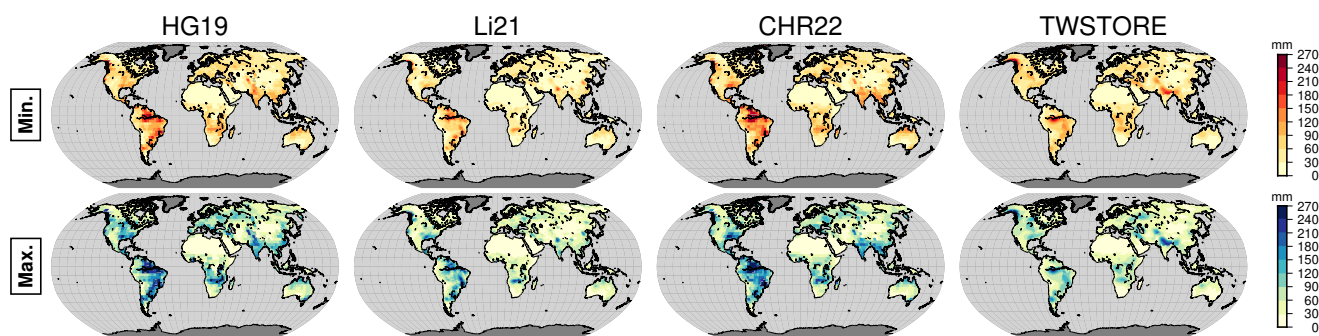
345 We hypothesize that these differences are due to dynamic processes, partly influenced by humans, that are not well captured in the predictors of the reconstruction. Our reconstruction exhibits the strongest accelerations in terms of signal magnitude, which originates from SLR. Acceleration and trend exhibit similar signs for most regions, indicating respective drying and wetting conditions. This behavior is only partly reflected by the other reconstructions. The acceleration is part of the inter-annual signal. Except for HG19, all reconstructions were trained on a small training time frame within the GRACE period. The ability

350 of algorithms to accurately extrapolate interannual changes from a limited training dataset is questionable and may contribute to the variability in the signal magnitude and sign observed here. The data sets exhibit nearly identical annual phases. For the yearly amplitude, all data sets agree on high signal amplitude along the equatorial band and higher signal magnitudes for the more humid, northern climate regions. Differences arise in the strength of the signal between Africa and North America, which is likely related to the GRACE products and predictor data sets used. The high agreement on the annual scale is due to

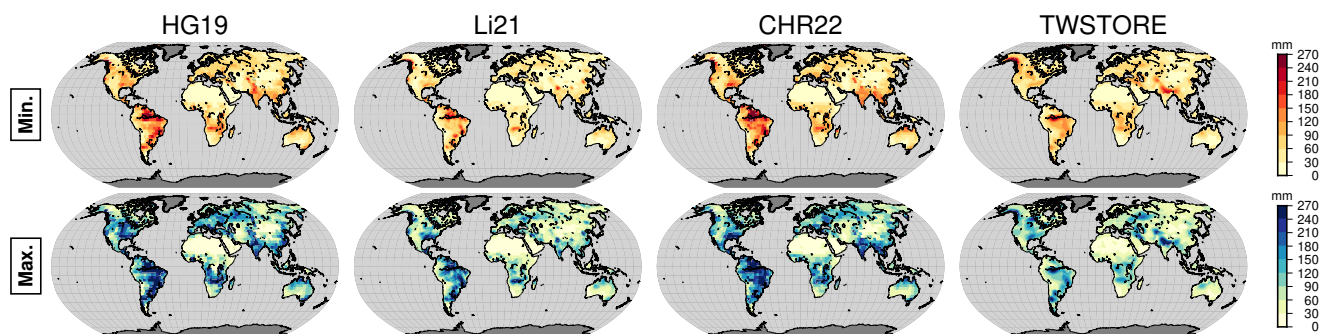
355 the dominance of the yearly cycle in climate data. As for the acceleration, the interannual signal exhibits the most variation in terms of signal magnitude and pattern throughout the datasets. The data sets agree on high inter-annual signal variations in the Amazon Basin and along the Atlantic coast of South America, the Zambezi Basin, the region around the Don Basin, and India. On the sub-seasonal scale, all data sets show high signal magnitudes for the equatorial band, reflecting precipitation patterns.

## 5 Results

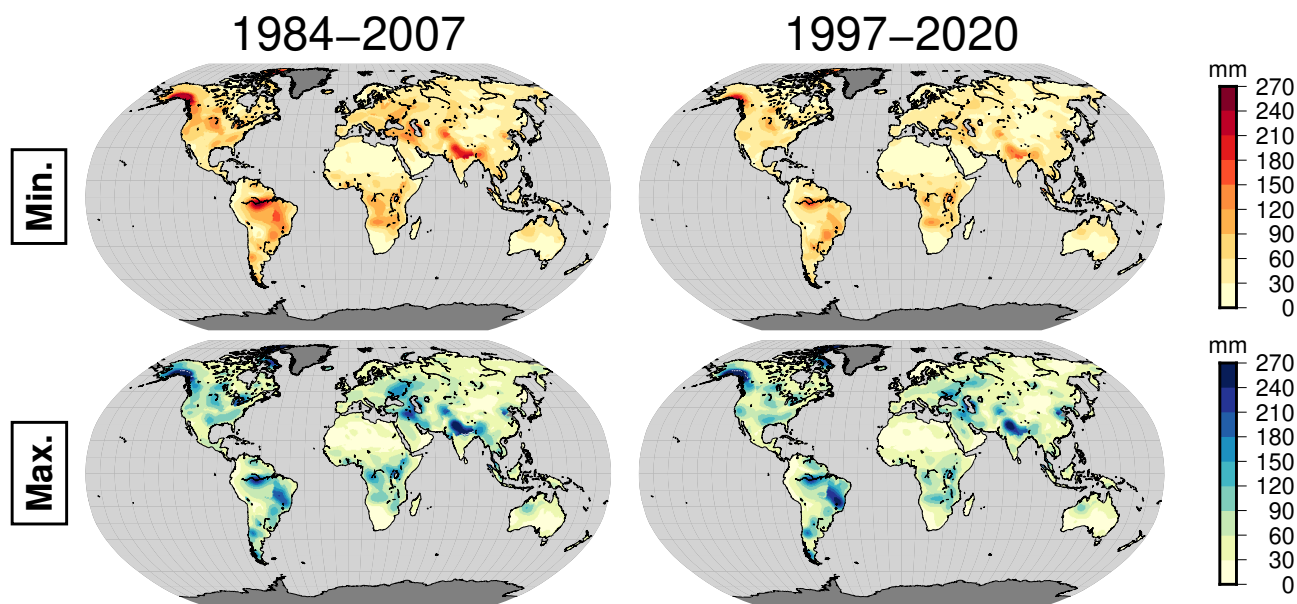
### 360 5.1 How often did extreme terrestrial water storage deficit or surplus occur during the past decades?



**Figure 5.** One-in-five-year return levels of anomalously low(top) and high (bottom) terrestrial water storage (TWS) with respect to climatology for different reconstructions for 1984-2019



**Figure 6.** One-in-ten-year return levels of anomalously low(top) and high (bottom) terrestrial water storage (TWS) with respect to climatology for different reconstructions for 1984-2019



**Figure 7.** One-in-five-year return levels of anomalously low(top) and high (bottom) terrestrial water storage (TWS) with respect to climatology for two 24-year periods 1984-2007 and 1997-2020

Due to the still short GRACE/-FO data record, only a few studies (Moore and Williams, 2014; Humphrey et al., 2016; Kusche et al., 2016) have assessed the statistics of extreme TWSA events, i.e., beyond analyzing either episodic droughts and floods, or the second moments of the underlying probability distribution. It is an open question whether the expected intensification of the water cycle due to global warming (Huntington, 2006; Durack et al., 2012) has already been observed in space gravimetry and if yes, whether it can be attributed to climate drivers. For example, Yoon et al. (2015) project an increase in the variance of



California top-1 m soil moisture by at least 50% towards the end of the 21st century.

Based on our multidecadal reconstruction of terrestrial water storage, we first determine the return levels of one in five and one in ten years for extreme anomalous storage conditions (see section A4 for the computational details). These return levels represent the frequency of annual extreme storage surpluses and deficits that exceed average climatological patterns. We then  
 370 examine whether temporal changes in return levels emerge by partitioning the time frame into overlapping 24-year periods. We focus on annual maximum return surplus and deficit levels for a given return time, as these are sensitive indicators of magnitude increases in the underlying distribution's tails (Allen and Ingram, 2002). It is essential to note however that although our reconstruction is based on GRACE and GRACE-FO data, it is not entirely independent of climate datasets.

Following (Kusche et al., 2016), we derive annual maximum and minimum anomalies per grid cell and fit a Generalized  
 375 Extreme Value (GEV) function to these time series by using the moment method (Martins and Stedinger, 2000). The fitted distribution enables the drawing of samples, i.e., return values, resulting in maps of one-in-five (Fig. 5) and one-in-ten (Fig. 6) year return levels of anomalously high and low TWSA events from the reconstructions for 1984-2020. Again, we note that "anomalously" in these maps refers to the TWSA climatology, which varies for each grid point. The figures demonstrate that, due to droughts and rainfall extremes over the last few decades, a one-in-five-year storage deficit (surplus) has reached 20-30  
 380 cm below (above) the climatological conditions for some of the major river basins. In contrast, at the one-in-ten-year scale, this could be 60 cm (in the Amazon basin).

We find return levels of extreme storage anomalies generally moderately higher for HG19 and CHR22 than for Li21 and our new reconstruction. We expect that extreme storage surplus or deficit primarily reflects patterns in rainfall, and we speculate that a possible explanation for the differences may be related to the specific choice of predictors, precipitation and temperature,  
 385 employed by HG19 and CHR22. As expected, due to its strong TWSA signal driven by high amounts of precipitation and influenced by ENSO, leading to droughts and floods, the Amazon basin is identified as the global hotspot here with the highest (up to 0.4 - 0.5 m) and lowest (0.3 m) one-in-five-year return levels in all data sets. Additionally, all reconstructions indicate peak annual anomalous surpluses and low water storage levels for the Zambezi Basin, the Congo Basin, and the region surrounding Lake Victoria. High return levels are also identified for the Orinoco, Essequibo, São Francisco, and La Plata/Paraná  
 390 river basins. HG19 and our reconstruction reveal moderate return storage levels for the US, primarily in the Mississippi Basin, the region surrounding the Caspian Sea, and the Volga-Don Basin; the latter is also identified in the CHR22 and Li21 reconstructions.

When compared to Fig. 1 in Kusche et al. (2016), we observe close similarities in the spatial patterns, although the return levels in the present study are generally lower. However, this is an expected outcome, as we analyze anomalies concerning the  
 395 climatology of multi-decadal reconstructions trained on more recent GRACE data, whereas in Kusche et al. (2016), only 12 years of data were available.

As expected, one-in-ten-year return levels are larger as compared to the one-in-five-year return levels. The spatial patterns are similar, of course, since both one-in-five and one-in-ten year return level maps are based on the same estimates for the GEV distribution parameters.

400 Lastly, to identify temporal changes, we divided our water storage reconstruction into two overlapping 24-year periods. For



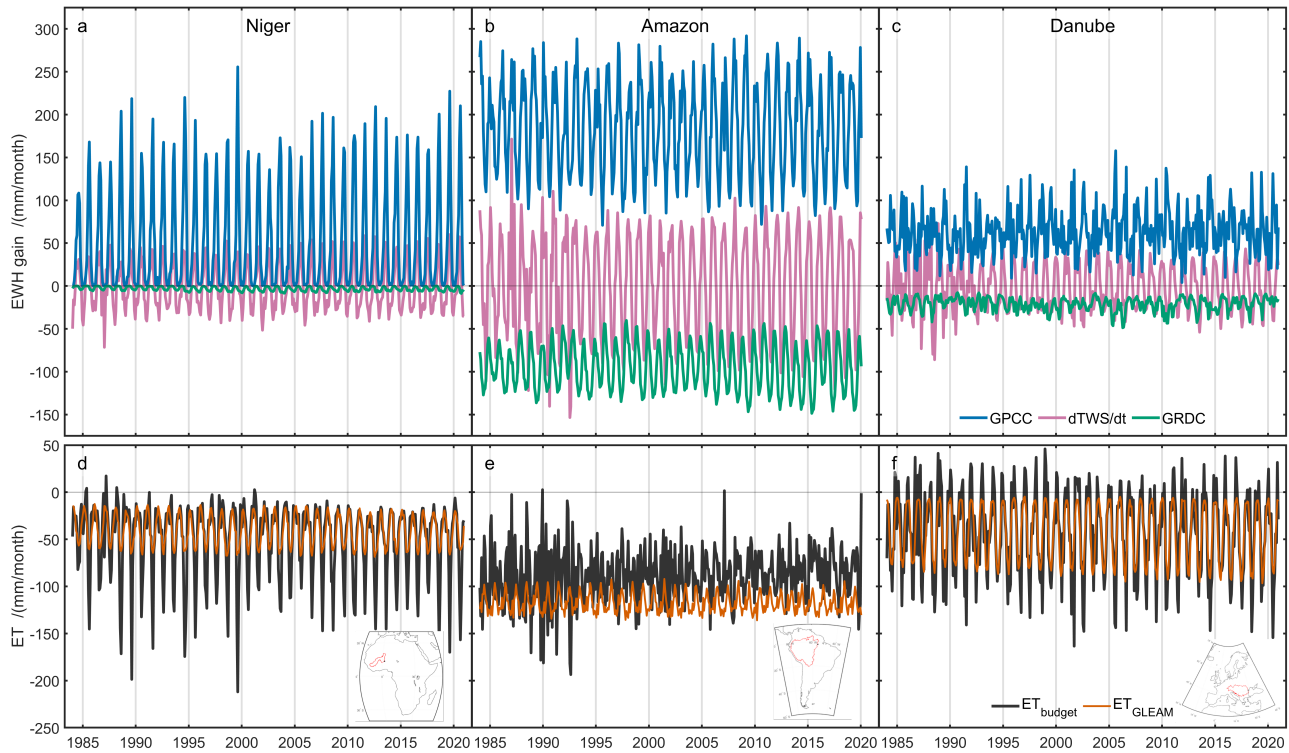
consistency, we derive return levels for the annual water storage surplus and deficit concerning the same long-term climatology. As the sliding window overlaps by about 50%, the results cannot be entirely independent. However, shortening the analysis window further would result in a loss of robustness for the moment estimates, as shown by simulations in (Kusche et al., 2016). We find for most regions somewhat higher return levels for storage deficits for the 1984-2007 timeframe compared to the more recent period, which at first glance seems counterintuitive in light of the narrative of more severe and more frequently occurring drought conditions. However, return levels express the magnitude of a hypothetical anomalous event that occurs under a given probability relative to the mean storage. In other words, our analysis tells that the variability of annual minimum storage conditions with respect to the long-term mean was higher from 1984 to 2007 than in recent years. Fluctuations in yearly water storage lows in recent years were less pronounced. However, this should not be confused with the magnitude or duration of the recent annual water storage deficit, which contributed to an observed multi-decadal decrease in water storage. These storage trends have been found to bias the computation of drought indicators, as these typically rely on the assumption of nonstationarity (Gerdener et al., 2020).

In contrast, we find only a few differences between the two time frames when comparing the return levels for extreme water storage surplus. A visual inspection suggests that return levels have been larger in the Amazon, Zambezi, and East African Rift Valley from 1984 to 2007 compared to 1997-2020. In contrast, in the La Plata/Paraná and, for example, the North China Plain, return levels were higher in the more recent period. However, we caution that these changes in one-in-five-year return levels - mainly of the order of a few cm within a 13-year window shift - may be easily caused by one or two extreme events, e.g., during ENSO years.

## 5.2 Terrestrial water budget and evapotranspiration

Given that multidecadal time series for measured precipitation and river discharge exist for several river basins, it is tempting to evaluate the terrestrial water budget with the reconstructed terrestrial water storage records now and solve for basin-averaged evapotranspiration, as done, e.g., earlier in several studies for the GRACE/-FO period (e.g., Xiong et al. (2023)). The term "evapotranspiration" (Miralles et al., 2020) refers to the total freshwater flux between the atmosphere and the surface through all types of evaporation and condensation, including those from transpiration and interception by vegetation, sublimation, and deposition, where downward fluxes are defined positive. Evapotranspiration plays a crucial role in the water and energy cycles, as it is sensitive to climate change and anthropogenic land cover modifications, providing insight into the history of the land surface's response to various forcings. Several studies suggest that global evapotranspiration might have changed significantly at decadal timescales (Douville et al., 2013), both due to changes in evaporative demand (temperatures) and moisture supply (vegetation, rainfall), but direct (flux-tower) measurements are scarce. Independent data sets from the budget approach can also aid in the evaluation of meteorological reanalyses (e.g., Springer et al. (2014)).

We, therefore, decided to evaluate basin-averaged monthly evapotranspiration  $\overline{ET}$  for three of the largest river basins of the world situated in different climate zones, the Amazon, the Niger and the Danube basin. Results for eight additional catchments



**Figure 8.** Terrestrial water budget monthly fluxes over the Niger, Amazon and Danube river catchments. The first row (a,b,c) shows in- and output expressed in equivalent water height from (1) precipitation from GPCC (blue), (2) river discharge from GRDC (green), and (3) water storage change from reconstructed TWS (purple). The second row (d,e,f) shows monthly  $\overline{ET}$  (black) as derived using the budget equation with the upper row components, and 'Actual Evaporation' from GLEAM (orange) for comparison, respectively. Positive fluxes represent net monthly water mass gain for the catchment and negative numbers represent net losses.

are provided in supplementary table A1. We use the formulation of the terrestrial water budget (where  $\sigma$  equals TWSA, as above)

$$\frac{d\sigma}{dt} = P - ET - R \quad (11)$$

with precipitation  $P$  and river discharge  $R$ , which is rearranged for evapotranspiration  $ET$  and numerically evaluated as in Springer et al. (2014). For terrestrial water storage changes, we evaluate TWSTORE We utilize the GPCC 2022 dataset for precipitation, as shown in Table 2. The river discharge is obtained from the Global Runoff Data Center GRDC for the Obidos (Amazon river), Ceatal (Danube river), and Niamey (Niger river) stations (table B1 and figure E1). All data sets span the period from 1984 to 2020. In addition, we compare the derived  $\overline{ET}$  with the Actual Evaporation product from GLEAM after temporal binning and averaging over the basin area.



Table 3 summarizes the mean basin-averaged water mass fluxes over the entire reconstruction period and the two shorter  
 445 periods as above, i.e. 1984 to 2007 and 1997 to 2020, derived from forming the budget. During all three assessed time frames,  
 evapotranspiration rates for the Danube are found to be approximately 42 mm/mon, which balances two thirds of the total  
 precipitation input. TWS net losses increased by 0.2 mm/a<sup>2</sup> and annual precipitation intensified by 2.5 mm each year on average  
 over 1984–2020. As monthly river discharge increased only by 0.7 mm between the two sub-periods,  $\overline{ET}$  rates changed by  
 2.4 mm/a<sup>2</sup> to account for increased P input and TWS losses. The mean of the budget-derived  $\overline{ET}$  is in perfect agreement with  
 450 GLEAM throughout the three assessed periods.

However, for the rather arid Niger upstream from Niamey, the net loss to the atmosphere from the budget is 8.1 mm/mon  
 stronger than GLEAM suggests for the period 1984–2007 and 8.4 mm/mon for the period 1997–2020. The differences in  
 evapotranspiration are mirrored in the precipitation differences recorded by GPCC and the ERA5 reanalysis (Hersbach et al.,  
 2020), with GPCC depicting higher precipitation rates compared to ERA5, which can be attributed to more pronounced peaks  
 455 in GPCC precipitation during summer months (figure 8a) and supplementary figures F1 and G1. We observe the opposite  
 behavior over the mostly tropical Amazon region, where the budget-derived  $\overline{ET}$  appears weaker by around 30 mm/mon than  
 seen in GLEAM. Furthermore, ERA5 mean monthly accumulated precipitation is 24 mm/mon stronger than GPCC. However,  
 it is obvious that observational and reanalysis products in the Niger and Amazon domains are hampered due to low observation  
 density. Our results suggest that in the Amazon basin, evapotranspiration rates have weakened by 4.4 mm between the two  
 460 (overlapping) analysis periods. The decrease in evapotranspiration is only partially explained by a decrease in  $P$  (2.1 mm) and  
 an increase in  $R$  (2.9 mm), while the water storage change remained comparably constant (decreased by less than 1 mm). We  
 speculate that the lower evapotranspiration rate and the higher river discharge in the Amazon catchment could be related to the  
 ongoing deforestation of the region (Coe et al., 2009). For the Niger basin, we observe an increase in  $P$  (+3.0 mm/a<sup>2</sup>) rates.  
 The higher amount of  $P$  is compensated by 73% by an increase in  $\overline{ET}$ , by 19% through more discharge, and by 8% due to  
 465 storage recharge.

It should be noted that opposite-sign  $\overline{ET}$  fluxes (cf. figure 8f), i.e. from the atmosphere to the ground by means of condensa-  
 tion and deposition, can represent existing physical processes that are frequently observed during winter months in temperate  
 and boreal climate zones. However, the amplitude of the respective events in the budget-derived  $\overline{ET}$  is not in agreement with  
 those of GLEAM or ERA5 reanalysis and could hint at unidentified offsets in some of the budget components.  
 470 The derived ET time series are available at <https://doi.org/10.5281/zenodo.16643628>, Gutknecht (2025).

## 6 Conclusions

We present a GRACE-like reconstruction of terrestrial water storage anomalies for the period 1984–2020, covering the entire  
 global land area excluding Greenland and Antarctica. In contrast to earlier data sets, our reconstruction is derived as an optimal  
 combination of a geodetic TWSA time series (derived from time-variable gravity data from the SLR and DORIS techniques)  
 475 and a data-driven TWSA reconstruction based on climate data sets. The data-driven TWSA reconstruction is trained on the  
 GRACE data from 2003 to 2010. Both TWSA data sets are transformed to a truncated EOF basis to remove and minimize



| Catchment | Component          | 1984–2007 |           | 1997–2020 |           | 1984–2020 |           |
|-----------|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|           |                    | $\mu$     | $\beta_2$ | $\mu$     | $\beta_2$ | $\mu$     | $\beta_2$ |
| Amazon    | $d\sigma_{rec}/dt$ | +0.1      | -0.4      | +0.0      | -0.1      | +0.3      | -0.0      |
|           | $P_{GPCC}$         | +183.9    | -3.3      | +181.8    | -0.9      | +183.8    | -1.4      |
|           | $R_{GRDC}$         | -93.3     | +0.6      | -96.2     | -5.8      | -95.3     | -2.6      |
|           | $ET_{budget}$      | -90.5     | +2.4      | -86.1     | +4.7      | -88.2     | +4.0      |
|           | $ET_{GLEAM}$       | -117.3    | -0.7      | -118.0    | -0.5      | -117.6    | -0.6      |
| Danube    | $d\sigma_{rec}/dt$ | -0.2      | -0.3      | -0.4      | -0.2      | -0.3      | -0.2      |
|           | $P_{GPCC}$         | +62.5     | +4.2      | +65.3     | +0.7      | +63.6     | +2.5      |
|           | $R_{GRDC}$         | -21.4     | -2.4      | -22.1     | +2.1      | -21.3     | -0.4      |
|           | $ET_{budget}$      | -41.3     | -2.1      | -43.7     | -3.1      | -42.5     | -2.4      |
|           | $ET_{GLEAM}$       | -41.3     | -2.4      | -43.7     | -2.1      | -42.5     | -2.2      |
| Niger     | $d\sigma_{rec}/dt$ | -0.0      | +0.2      | +0.3      | -0.0      | +0.1      | +0.3      |
|           | $P_{GPCC}$         | +51.0     | +2.0      | +54.0     | +3.7      | +52.9     | +3.0      |
|           | $R_{GRDC}$         | -3.1      | -1.3      | -3.7      | +0.1      | -3.3      | -0.6      |
|           | $ET_{budget}$      | -48.0     | -0.4      | -50.1     | -3.9      | -49.6     | -2.2      |
|           | $ET_{GLEAM}$       | -39.8     | -1.6      | -41.7     | -1.0      | -40.8     | -1.6      |

**Table 3.** Basin mean fluxes from evaluating the terrestrial water budget over the three periods.  $\mu$ : temporal mean over monthly data;  $\beta_2$ : linear trend derived from a 6-parameter model. Mean fluxes are in units of mm/mon, trends in mm/a<sup>2</sup>. Positive fluxes are defined as input into the integration domain, negative fluxes represent losses. Note that by this definition negative trends can also mean increased losses.

noise in the input data, harmonized, and finally combined. Since it is challenging to prescribe the uncertainty levels of the two TWSA data sets, we apply an iterative variance component estimation procedure in their combination.

A comparison between the final reconstruction (TWSTORE), the preliminary reconstruction, i.e., before the data combination, and the SLR-DORIS-only dataset revealed that the SLR-DORIS data influence the combination primarily on longer time scales. We show that the reconstruction is consistent with other independently derived reconstructions. The data sets exhibit a good agreement on shorter time scales. On longer time scales, the differences between the data sets increase, raising doubts about the reconstructions' ability to capture long-term signals based on a relatively short training period. We find significant TWSA rate changes and accelerations over the entire four-decade time frame at the scale of large river basins. Our findings suggest that GRACE trends should not be projected unquestioningly to the past. The results also indicate that trend error models should consider long-term fluctuations. The inter-annual and inter-decadal signals are much more prominent in the satellite-geodetic data compared to the reconstructions. We speculate they may be missing in pure data-driven reconstructions.

A formal analysis of extreme values reveals that, over the last few decades, one-in-five-year storage deficits (surpluses) have reached 20–30 cm below (above) climatological conditions for some major river basins and up to 60 cm for one-in-ten-year deficits and surpluses. Through a sliding-window analysis, we find only minor temporal changes in these return levels, indicating no intensification of extremes in water storage. In our opinion, it is unclear whether extreme value theory provides a



suitable framework to investigate hypotheses on water cycle intensification and more frequent extremes. It is important to note that at the spatial and temporal scale of our reconstruction, extreme weather events, such as convective precipitation and the subsequent flooding, are not evident. The hydrological response to such events is probably more affected by river runoff and evapotranspiration than by changes in storage soil moisture, surface water levels and groundwater recharge.

As a case study application, we derive evapotranspiration rates based on terrestrial water budgets using the presented reconstruction for three major river basins, utilizing observed river streamflow data. We found a reduced evapotranspiration flux for the Amazon basin that could only partly be attributed to reduced precipitation rates and increased runoff, suggesting a change in land cover. For the Niger basin, we find increased evapotranspiration rates that compensate for the increase in precipitation. Our analysis demonstrates the potential of such long-term reconstructed data records to study changes in energy fluxes, water balance, and climate variables. We conclude that our extended observational TWSTORE record, although not directly observed, is beneficial for analyzing terrestrial water budgets despite utilizing spatial constraints and containing higher noise in the earlier 1984-1992 period. The reconstruction also enables the evaluation of modeled water storage before the GRACE time frame at a spatial scale that does not conform to GRACE but is still valuable, for example, for validating CMIP runs.

We recognize that evaluating our own and others' reconstructions is a challenging task, and we invite readers to propose new ideas for this purpose.

## 7 Data availability

The presented datasets are publicly available from: TWSTORE (<https://doi.org/10.5281/zenodo.15827789>, Hacker (2025))  
 Budget-derived ET time series (<https://doi.org/10.5281/zenodo.16643628>, Gutknecht (2025))  
 All datasets used in this article are available at the following locations: CSR mascon (<https://www2.csr.utexas.edu/grace>, last access: 6 August 2025); GRACE L2 data: COST-G ([https://doi.org/10.5880/COST-G.IGEM\\_02\\_L2](https://doi.org/10.5880/COST-G.IGEM_02_L2), Meyer et al. (2025)), ITSG2018 (<https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018>, Mayer-Gürr et al. (2018a)); HadISST1 (<https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>, last access: 6 August 2025), COBE-SST 2 (<https://psl.noaa.gov/data/gridded/data.cobe2.html>, last access: 6 August 2025), GPCC Version 2022 ([https://doi.org/10.5676/DWD\\_GPCC/CLIM\\_M\\_V2022\\_050](https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2022_050), Elke et al. (2022)), CPC global precipitation V1.0 (<https://www.psl.noaa.gov/data/gridded/data.cpc.globalprecip.html>, last access: 6 August 2025), CPC Soil Moisture v2 (<https://www.psl.noaa.gov/data/gridded/data.cpcsoil.html>, last access: 6 August 2025), GLEAM Soil Moisture v4.1a (<https://www.gleam.eu/>, last access: 6 August 2025), GHCN CAMS Monthly Temperature (<https://www.psl.noaa.gov/data/gridded/data.ghcncams.html>, last access: 6 August 2025), Globmap LAI v3 (<https://doi.org/10.5281/zenodo.4700264>, Liu et al. (2021)), Reconstruction by Li (2021): <https://doi.org/10.5061/dryad.z612jm6bt>, Reconstruction by Humphrey and Gudmundsson (2019a): <https://doi.org/10.6084/m9.figshare.7670849>, Reconstruction by Chandanpurkar et al. (2022a): <https://doi.org/10.5281/zenodo.6659543>; DORIS-SLR hybrid solution by Löcher et al. (2025a): <https://doi.org/10.5880/ICGEM.2025.001>; "ERA5 monthly data on single levels from 1940 to present" Total precipitation: <https://doi.org/10.24381/cds.f17050d7>, accessed 15 April 2025; GLEAM Actual evaporation v4.2a: accessed 10



525 February 2025 from sftp://hydras.ugent.be through https://www.gleam.eu; GRDC streamflow: https://portal.grdc.bafg.de, accessed 11 April 2025.

## Appendix A: Methods

### A1 Signal decomposition: Identifying dominant spatial and temporal patterns

530 Principal component analysis (PCA) is commonly used to reduce the dimensionality of a data set while preserving as much information contained in the data as possible (Jolliffe, 2002). GRACE L3 data is mostly given on a  $1^\circ$  ( $\approx 110$  km) or  $0.5^\circ$  grid ( $\approx 55$  km). Reconstructing the GRACE time series for every grid point would be computationally expensive. Therefore, we reduce the dimensionality of the GRACE data using PCA, expressing the GRACE data into temporal (principal components (PCs) and spatial (empirical orthogonal functions (EOFs)) dominant modes. Assuming the spatial modes are constant over the  
 535 reconstruction period, only the temporal modes are reconstructed.

Within the PCA, a new set of (transformed) uncorrelated variables, the principal components  $P_{n \times t}$  (PCs),

$$P_{n \times t} = X_{n \times t} E_{n \times n} \quad (A1)$$

are gained by projecting the data,  $X_{n \times t}$ , with  $n$  being the spatial grid cells and  $t$  the epochs, onto a new orthogonal space,  $E_{n \times n}$ , which represents the empirical orthogonal functions. The first  $m$  PCs contain most information, mostly 95 percent of  
 540 the total signal variance, whereas noise is mapped to the  $m + 1 \dots n$  PCs (Wold et al., 1987). The orthogonal space is gained from the eigenvectors of the empirical covariance matrix  $C = \frac{1}{t}(X - \bar{x})^T(X - \bar{x})$  of the data. Using the  $m$  first modes, the data matrix can be restored by

$$\bar{X}_{n \times t} = \bar{E}_{n \times m} \bar{P}_{m \times t} \quad (A2)$$

We apply the PCA to the GRACE data splitting it into PCs and EOFs. The predictors and the SLR-DORIS data is then projected  
 545 onto the GRACE derived EOF basis as

$$PC_{n \times t} = D_{n \times t} E_{n \times n} \quad (A3)$$

with  $PC$  denoting the projected data and  $D$  the specific data set.

### A2 Variance component estimation

Variance component estimation (VCE) is a weighting technique for observations in the context of least-squares adjustment. A  
 550 well-known example of the application of VCE is the combination of gravity data in the context of gravity field modeling or geopotential determination (Koch and Kusche, 2002; Brockmann and Schuh, 2010; Fecher et al., 2015).

Given a linear normal equation system in terms of a least squares adjustment (linearized Gauss- Markoff Model (Koch, 2004))

$$l + v = Ax \quad \Sigma\{\mathcal{L}\} = \sigma_0^2 Q_{ll} \quad (A4)$$



$\mathbf{x}$  describes the parameters of the adjustment. The matrix  $\mathbf{A}$ , also called design matrix, maps the observation to the parameter space,  $\mathbf{v}$  are the residuals, and  $\Sigma\{\mathcal{L}\}$  the covariance matrix of the observations. The variance component  $\sigma_0^2$  scales the co-factor matrix  $\mathbf{Q}_{ll}$  and is a mostly unknown random variable. The least squares solution for the unknown parameters  $\mathbf{x}$  reads

$$\hat{\mathbf{x}} = \mathbf{N}^{-1} \mathbf{n} \quad (\text{A5})$$

with  $\mathbf{N} = \mathbf{A}^T (\sigma_k^2 \mathbf{Q}_{l_k l_k})^{-1} \mathbf{A}$  being the normal equation matrix and  $\mathbf{n} = \mathbf{A}^T (\sigma_k^2 \mathbf{Q}_{l_k l_k})^{-1} \mathbf{l}$  the normal equation vector. If  $\mathbf{A}$  has full rank, solving the above system of equations leads to the optimal solution in the least squares sense.

Assuming  $\mathbf{l}$  is composed of  $n$  independent groups, we can sequentially sum up the normal equations for each observation group, leading to

$$\sum_{k=1}^n \sigma_k^2 \mathbf{N}_k \mathbf{x} = \sum_{k=1}^n \sigma_k^2 \mathbf{n}_k \quad (\text{A6})$$

The unknown variance components  $\sigma_k^2$  are estimated iteratively from the residuals, catching as many uncertainties from the data as possible (Förstner, 1979). Based on the maximum-likelihood method or the best invariant quadratic unbiased estimation (Koch, 2018; Förstner, 1979), the estimated variance components are given as

$$\sigma_k^{2(p)} = \frac{\mathbf{v}_k^T (\sigma_k^{2(p-1)} \mathbf{Q}_{l_k l_k})^{-1} \mathbf{v}_k}{r_k} \quad \text{with} \quad \mathbf{v}_k = \mathbf{A}_k \hat{\mathbf{x}}_k - \mathbf{l}_k \quad \text{and} \quad r_k = u_k - \text{tr}(\mathbf{N}_k \mathbf{N}^{-1}) \quad (\text{A7})$$

with  $p$  referring to the iteration step,  $r_k$  is the redundancy and  $u_k$  the number of observations of the individual observation group  $k$ .

We use VCE for the data combination not on the normal equation level but on the observational level, stating that the parameters  $\mathbf{x}$  are directly observed, leading to  $\mathbf{A}_k$  being the identity matrix.

### A3 Granger causality

Granger causality (Granger, 1969), is a statistical method to determine temporal, predictive causality between two or more time series. Given two time series  $\mathbf{l} = [l_1, l_2, \dots, l_n]$  and  $\mathbf{y} = [y_1, y_2, \dots, y_n]$  with  $n$  being the length of the time series,  $\mathbf{y}$  Granger causes  $\mathbf{l}$  if incorporating  $\mathbf{y}$  in the forecast of  $\mathbf{l}$  improves the forecast of  $\mathbf{l}$ .

Under the assumption of stationarity (marginal distribution does not change over time/space), causal sufficiency (all influencing variables are observed) and linear dependencies between the time series Granger (1969) employed an autoregressive (AR) process of order  $p$  to describe the relationship between the time series, leading to the two predictive models.

$$\hat{l}_t = \beta_{01} + \sum_{p=1}^p (\beta_{11p} l(t-p) + \beta_{12p} y(t-p)) + \epsilon \quad \text{full model} \quad (\text{A8})$$

$$\hat{l}_t = \beta_{01} + \sum_{p=1}^p (\beta_{11p} l(t-p)) + \epsilon \quad \text{baseline model} \quad (\text{A9})$$



$y$  Granger causes  $l$  if at least one of the parameters  $\beta_{12p}$  is non zero, leading to the null hypothesis (e.g Attanasio (2012))

$$H_0 : \beta_{121} = \beta_{122} = \dots = \beta_{12p} = 0 \quad \text{non causality} \quad (\text{A10})$$

A F-test is typically used to determine, whether incorporating  $y$  leads to an improvement in the prediction of  $l$ . The F-test assumes that the residuals are Gaussian distribution. However, earth-related observations are not normal distributed. Papa-  
 giannopoulou et al. (2017).

Instead of a F-test we follow Papagiannopoulou et al. (2017) and employ the coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=p+1}^N (l_i - \hat{l}_i)^2}{\sum_{i=p+1}^N (l_i - \bar{l})^2} \quad (\text{A11})$$

with  $\hat{l}_i$  being the forecast of  $l_i$  and  $\bar{l}$  the mean of  $l_i$ . In the literature,  $R^2$  is also known as the Nash–Sutcliffe model efficiency coefficient, a common metric in hydrology and climate science.  $R^2$  can also be interpreted as the forecasting model's fraction  
 of the explained variance. The optimal value for  $R^2$  is 1, which corresponds to an error-free forecast ( $\hat{l}_i = l_i$ ). Values  $R^2 < 0$   
 indicate that the forecast error is larger than the observed error and  $R^2 = 0$  indicates that the model has the same predictive skill  
 as the mean of the observed time series. If  $R^2(l, \hat{l})$  increases when  $y(t-1), y(t-2), \dots, y(t-n)$  is included in the prediction  
 of  $l$ , then  $y$  Granger causes  $l$ .

#### A4 Generalized extreme value (GEV) distribution

Extreme value theory provides the statistical framework for estimating the probability of singular or extreme events. General-  
 ized extreme value (GEV) distribution functions are a family of continuous probability distributions within the extreme value  
 theory. Three parameters describe the GEV: shape  $\xi$  (tail behavior of the distribution), location  $\mu$  ("location or shift of the  
 distribution), and scale  $\sigma$  (spread of the distribution). The GEV is built upon the three subfamilies: The Gumbel ( $\xi=0$ ), the  
 Fréchet ( $\xi>0$ ), and the Weibull ( $\xi<0$ ) distributions, leading to the cumulative distribution function

$$F(z) = \begin{cases} \exp \left( - \left( 1 + \xi \frac{z - \mu}{\sigma} \right)^{-1/\xi} \right) & \xi \neq 0 \\ \exp \left( \exp \left( - \frac{z - \mu}{\sigma} \right) \right) & \xi = 0 \end{cases} \quad (\text{A12})$$

The quantiles  $z_p$  at probability  $p$  can be derived as

$$z_p = \begin{cases} \mu - \sigma / \xi (1 - (-\log(1 - p)))^{-\xi} & \xi \neq 0 \\ \mu - \sigma \log(-\log(1 - p)) & \xi = 0 \end{cases} \quad (\text{A13})$$

We interpret  $z_p$  as the return level interval for probability  $p$ , which means that an expected extreme occurs after  $p$  years.

We follow the approach of Kusche et al. (2016), which is based on Martins and Stedinger (2000) to derive the parameters of  
 the GEV.

To estimate the GEV distribution, we first reduced a six-parameter model and the climatology from the grid time series. We



then select the annual maximum and minimum anomalous storage values. Using this shortened time series, we estimate the GEV parameters through Moment Estimation (ME) (Martins and Stedinger, 2000). The GEV parameters are calculated based on the first moment (mean), the second moment (standard deviation), and the third moment (skewness) of the time series,

610 leading to

$$\hat{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad \hat{s} = \left( \frac{1}{n} \sum_{j=1}^n (x_j - \hat{x})^2 \right)^{1/2}, \quad \hat{g} = \frac{1}{n} \sum_{j=1}^n \frac{(x_j - \hat{x})^3}{\hat{s}^3} \quad (\text{A14})$$

With  $x$  denoting the time series and  $n$  the number of observations. From the skewness, the shape parameter is iteratively solved via

$$\text{sign}(\hat{\xi}) \frac{-\Gamma_3(\hat{\xi}) + 3\Gamma_1(\hat{\xi})\Gamma_2(\hat{\xi}) - 2\Gamma_1^3(\hat{\xi})}{(\Gamma_2(\hat{\xi}) - \Gamma_1^2(\hat{\xi}))^{3/2}} = \hat{g} \quad (\text{A15})$$

615 where  $\Gamma_n(\hat{\xi}) = \Gamma(1 + n\hat{\xi})$  and  $\Gamma()$  is the Gamma function. The location and shape parameter is then derived as

$$\hat{\sigma} = \frac{\hat{s}|\hat{\xi}|}{(\Gamma_2(\hat{\xi}) - \Gamma_1^2(\hat{\xi}))^{1/2}}, \quad \hat{\mu} = \hat{x} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - \Gamma_1(\hat{\xi})) \quad (\text{A16})$$



## Appendix B: Terrestrial water budget and evapotranspiration

In addition to the three river catchments presented in the main article, we derived  $\overline{ET}$  using the terrestrial water budget equation for a wider selection of basins. Table A1 summarizes the fluxes of the budget components, respectively. Table B1 gives an overview of the exact stations and corresponding HydroSHED-based shapes used for the budget analysis. For the location of the analyzed catchments and the corresponding gauge stations, see Figure E1.

| Catchment   | Component          | 1984–2007 |           | 1997–2020 |           | 1984–2020 |           |
|-------------|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|             |                    | $\mu$     | $\beta_2$ | $\mu$     | $\beta_2$ | $\mu$     | $\beta_2$ |
| Orange      | $d\sigma_{rec}/dt$ | +0.0      | +0.0      | +0.1      | +0.0      | +0.0      | +0.0      |
|             | $P_{GPCC}$         | +30.6     | +2.1      | +31.0     | -2.6      | +30.4     | +0.0      |
|             | $R_{GRDC}$         | -0.6      | +0.1      | -0.5      | +0.4      | -0.5      | +0.2      |
|             | $ET_{budget}$      | -30.1     | -2.1      | -30.4     | +2.3      | -29.9     | -0.1      |
|             | $ET_{GLEAM}$       | -32.0     | -2.1      | -31.2     | +5.3      | -31.0     | +1.1      |
| Zambesi     | $d\sigma_{rec}/dt$ | -0.1      | +0.5      | +0.1      | +0.1      | -0.0      | +0.2      |
|             | $P_{GPCC}$         | +87.0     | +5.0      | +89.1     | -5.0      | +87.3     | +0.7      |
|             | $R_{GRDC}$         | -7.7      | -1.4      | -10.1     | -1.2      | -8.9      | -1.7      |
|             | $ET_{budget}$      | -79.3     | -3.1      | -79.0     | +6.4      | -78.5     | +1.3      |
|             | $ET_{GLEAM}$       | -67.1     | -0.2      | -67.3     | +0.5      | -67.1     | +0.0      |
| Yenisey     | $d\sigma_{rec}/dt$ | -0.1      | -0.1      | -0.1      | +0.0      | -0.1      | -0.0      |
|             | $P_{GPCC}$         | +34.8     | +1.3      | +35.9     | +0.6      | +35.1     | +0.8      |
|             | $R_{GRDC}$         | -21.2     | -1.0      | -20.9     | +2.1      | -20.8     | +0.3      |
|             | $ET_{budget}$      | -13.7     | -0.4      | -14.9     | -1.9      | -14.3     | -1.1      |
|             | $ET_{GLEAM}$       | -27.7     | -0.7      | -28.4     | -0.9      | -27.9     | -0.7      |
| Mississippi | $d\sigma_{rec}/dt$ | -0.3      | +1.1      | +0.0      | +0.6      | -0.1      | +0.5      |
|             | $P_{GPCC}$         | +50.0     | -0.8      | +50.2     | +0.9      | +50.4     | +0.2      |
|             | $R_{GRDC}$         | -8.8      | +1.4      | -10.0     | -2.5      | -9.8      | -0.9      |
|             | $ET_{budget}$      | -41.4     | +0.5      | -40.2     | +2.2      | -40.7     | +1.2      |
|             | $ET_{GLEAM}$       | -39.8     | -0.3      | -40.4     | -0.8      | -40.2     | -0.5      |
| Murray      | $d\sigma_{rec}/dt$ | +0.0      | +0.1      | +0.1      | +0.1      | +0.1      | +0.1      |
|             | $P_{GPCC}$         | +42.6     | -4.1      | +41.6     | -3.1      | +42.4     | -2.0      |
|             | $R_{GRDC}$         | -0.5      | +0.4      | -0.3      | -0.1      | -0.5      | +0.1      |
|             | $ET_{budget}$      | -42.1     | +3.9      | -41.3     | +3.3      | -42.0     | +1.9      |
|             | $ET_{GLEAM}$       | -40.1     | +4.8      | -37.5     | +4.5      | -39.1     | +3.3      |
| Loire       | $d\sigma_{rec}/dt$ | -0.1      | -0.3      | -0.3      | -0.2      | -0.2      | -0.2      |
|             | $P_{GPCC}$         | +68.8     | +2.5      | +68.7     | -4.7      | +68.2     | -0.5      |



|         |                    |       |      |       |      |       |      |
|---------|--------------------|-------|------|-------|------|-------|------|
|         | $R_{GRDC}$         | -19.8 | +0.5 | -18.4 | +2.2 | -19.0 | +1.3 |
|         | $ET_{budget}$      | -49.2 | -3.3 | -50.5 | +2.3 | -49.4 | -1.0 |
|         | $ET_{GLEAM}$       | -42.0 | -2.1 | -44.1 | -1.8 | -43.0 | -2.0 |
| Vistula | $d\sigma_{rec}/dt$ | -0.1  | +0.6 | -0.1  | +0.1 | -0.1  | +0.1 |
|         | $P_{GPCC}$         | +52.1 | +3.0 | +55.9 | +0.7 | +53.9 | +2.8 |
|         | $R_{GRDC}$         | -13.3 | -1.0 | -13.8 | +2.1 | -13.4 | -0.1 |
|         | $ET_{budget}$      | -38.8 | -1.4 | -42.2 | -2.7 | -40.6 | -2.6 |
|         | $ET_{GLEAM}$       | -40.8 | -2.8 | -43.5 | -2.7 | -41.7 | -2.6 |
| TISA    | $d\sigma_{rec}/dt$ | -0.1  | -0.2 | -0.4  | -0.2 | -0.3  | -0.3 |
|         | $P_{GPCC}$         | +57.1 | +5.5 | +60.5 | -0.4 | +58.2 | +2.8 |
|         | $R_{GRDC}$         | -14.8 | -3.8 | -15.0 | +4.5 | -14.1 | +0.0 |
|         | $ET_{budget}$      | -42.5 | -1.9 | -45.9 | -4.4 | -44.4 | -3.1 |
|         | $ET_{GLEAM}$       | -41.5 | -3.1 | -44.9 | -3.3 | -43.3 | -3.1 |

Table A1: Basin-averaged fluxes from evaluating the terrestrial water budget over the three periods.  $\mu$ : temporal mean over monthly data;  $\beta_2$ : linear trend derived from a 6-parameter model. Mean fluxes are in units of mm/mon, trends in mm/a<sup>2</sup>. Positive fluxes are defined as input into the integration domain, negative fluxes represent losses. Note that by this definition negative trends can also mean increased losses.



| GRDC ID | River Name              | Station Name       | Area /km <sup>2</sup> | N months         |
|---------|-------------------------|--------------------|-----------------------|------------------|
| 1159100 | *ORANGE RIVER           | VIOOLSDRIF         | 784763                | <sup>a</sup> 443 |
| 1234150 | RIVER NIGER             | NIAMEY             | 665238                | 444              |
| 1291100 | *ZAMBEZI RIVER          | KATIMA MULILO      | 334883                | <sup>b</sup> 443 |
| 2909150 | *YENISEY                | IGARKA             | 2431665               | <sup>c</sup> 408 |
| 3629001 | *AMAZON RIVER           | OBIDOS - LINIGRAFO | 4671462               | <sup>d</sup> 433 |
| 4127503 | MISSISSIPPI RIVER       | ST. LOUIS, MO      | 1776973               | 444              |
| 5204268 | *MURRAY RIVER           | LOCK 9 UPSTREAM    | 754963                | <sup>e</sup> 442 |
| 6123100 | LOIRE                   | MONTJEAN           | 110119                | 444              |
| 6458010 | VISTULA                 | TCZEW              | 193089                | 444              |
| 6544100 | TISA (Danube tributary) | SENTA              | 141085                | 444              |
| 6742900 | DANUBE RIVER            | CEATAL IZMAIL      | 779812                | 444              |

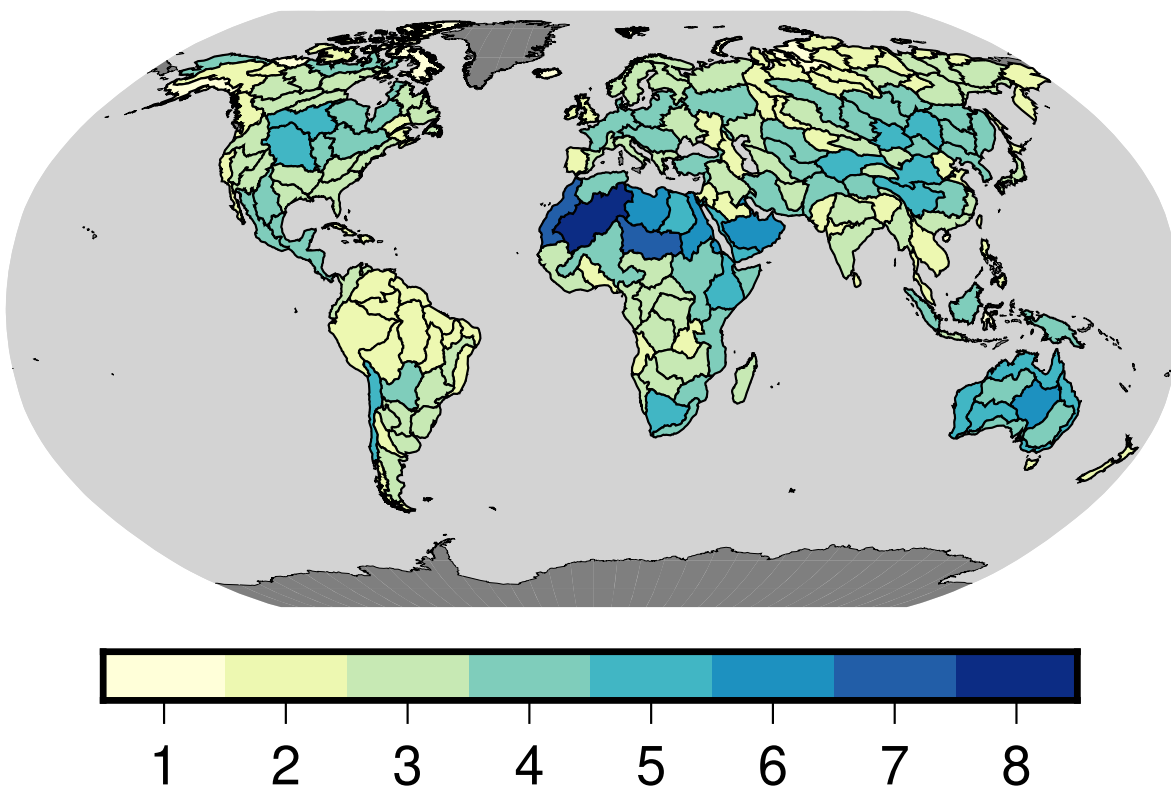
**Table B1.** Detailed list of GRDC stations and corresponding shape area used in the budget analysis. Last column is the number of months with discharge observations available over the targeted 1984–2020 time frame (i.e. N=444). a) Missing month in 1994-06; b) missing month in 2019-10; c) no data after 2017-12; d) no data after 2020-01; e) missing months in 2017-04 and 2017-05.



## Appendix C: Additional plots

The reconstruction is performed on a polygon scale. For every polygon, the reconstruction procedure is applied. The number of dominant modes selected in the PCA differs depending on the polygon's size and location. Figure A1 illustrates the number of dominant modes per polygon such that 95 % of the signal properties are covered. For the data combination, the polygons used for the reconstruction are aggregated to "bigger" polygons to account for the lower spatial resolution of the SLR data (figure B1). The number of modes selected for the aggregated polygons is shown in Figure C1. We also provide uncertainty information for our reconstructions. The information is given as variances only. Figure D1 displays the temporally averaged standard deviation for TWSTORE. The selected catchments for the budget analysis are depicted in Figure E1. For a comparison of the impact on budget-derived  $\overline{ET}$  of using ERA5 precipitation instead of GPCC, see figures F1 and G1. In the Niger catchment, the stronger peak precipitation in GPCC during summer causes higher  $\overline{ET}$ . And in the Amazon catchment, the relatively 'wetter' ERA5 precipitation leads to stronger  $\overline{ET}$  fluxes compared to GPCC.

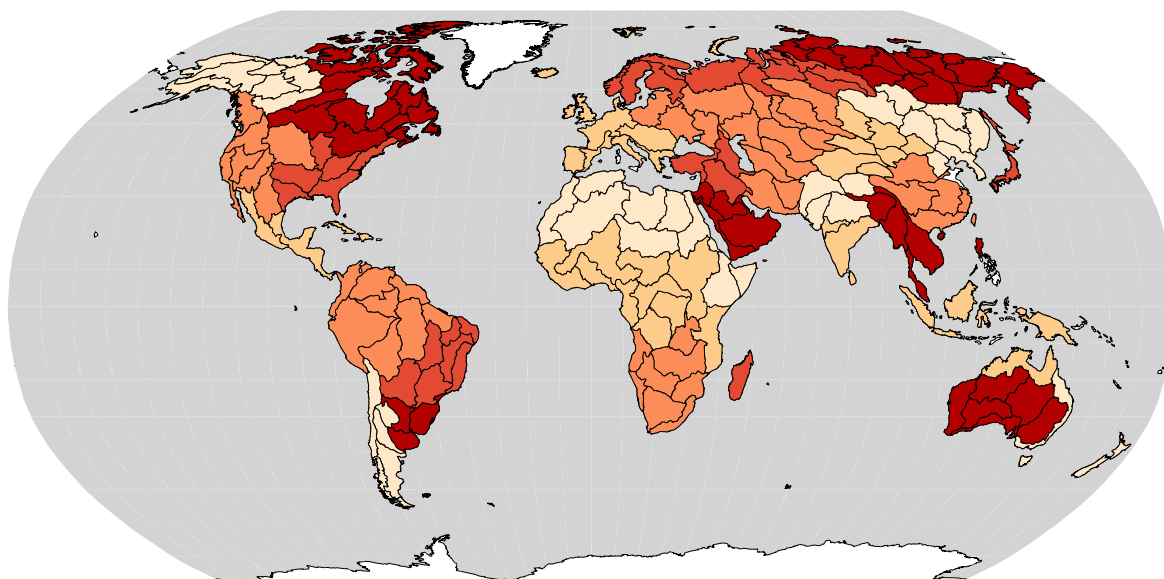
### A0.1 Number of modes per polygon



**Figure A1.** Number of modes per polygon used for the reconstruction. The number of polygons is 222.



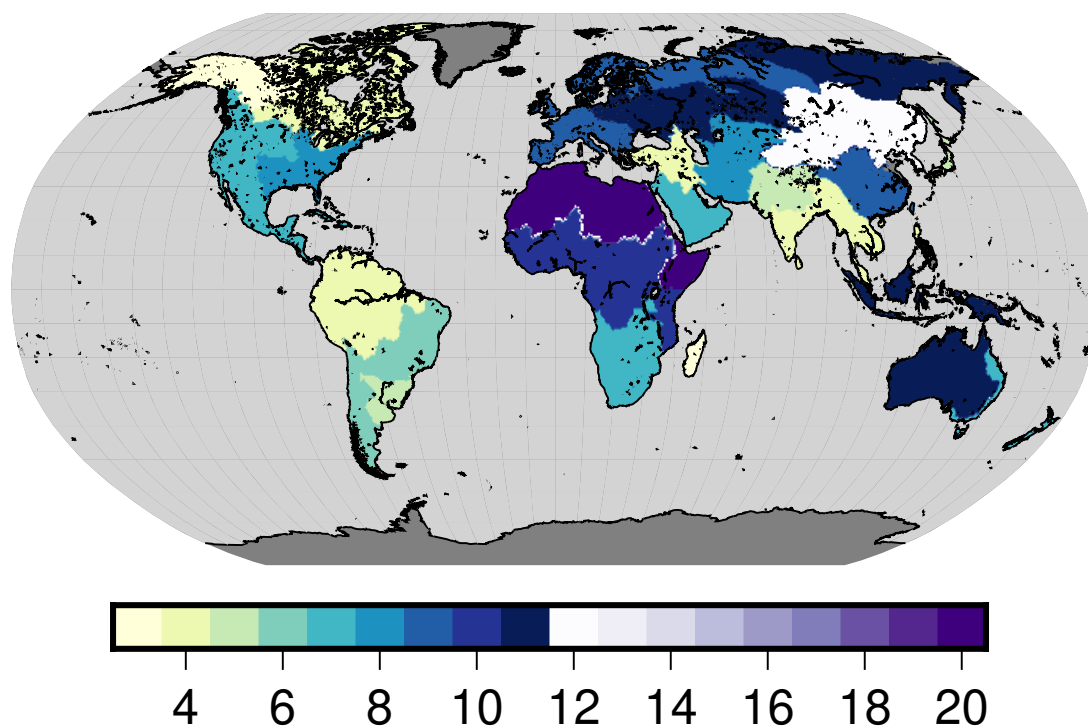
## A0.2 Polygons used for the data combination



**Figure B1.** Overview of the polygons used for the data combination. Polygons of the same color and spatial orientation are grouped together.



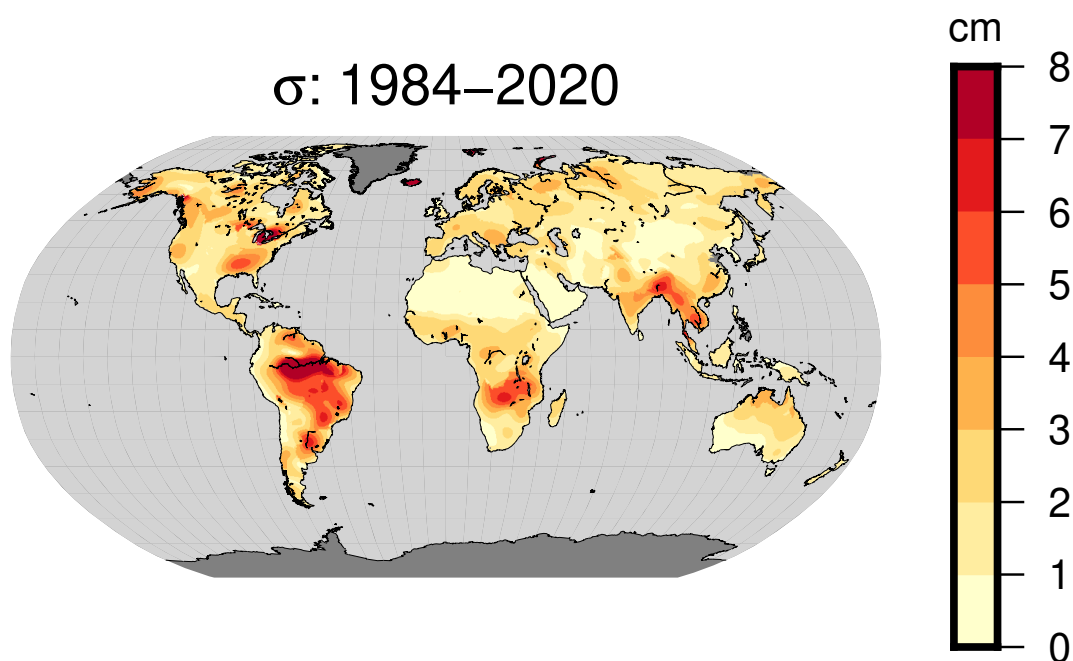
635 **B0.1 Number of dominant modes per polygon for the data combination**



**Figure C1.** Number of modes per polygon used for the data combination.



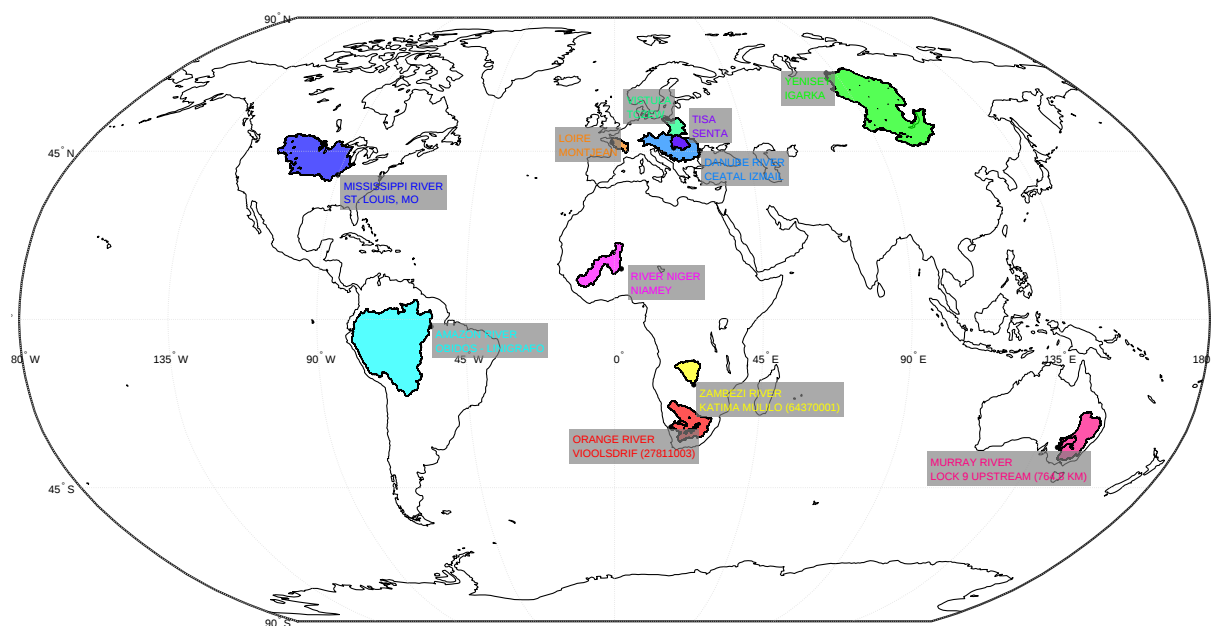
## C1 Variances of TWSTORE



**Figure D1.** Averaged standard deviation of TWSTORE over 1984 to 2020. High standard deviations correspond to regions with high TWSA signal variations



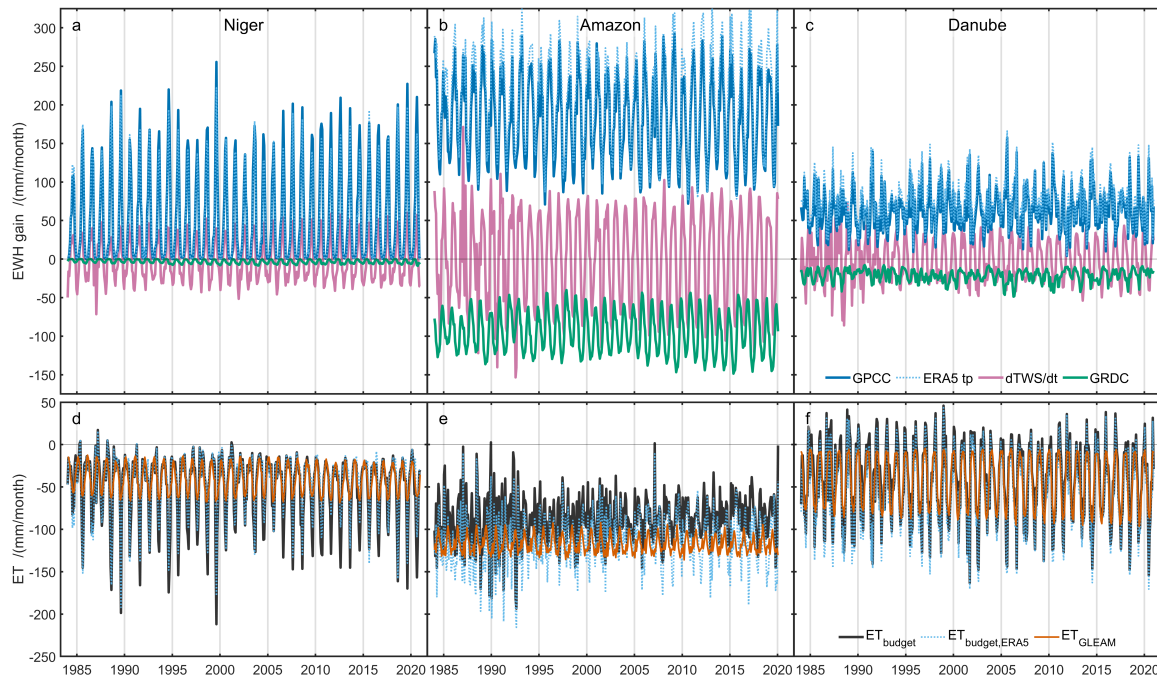
## E0.1 Location of river catchments used in the budget analysis



**Figure E1.** World map with locations of catchments analyzed for water budgets.



## F0.1 $ET_{budget}$ timeseries in comparison with ERA5



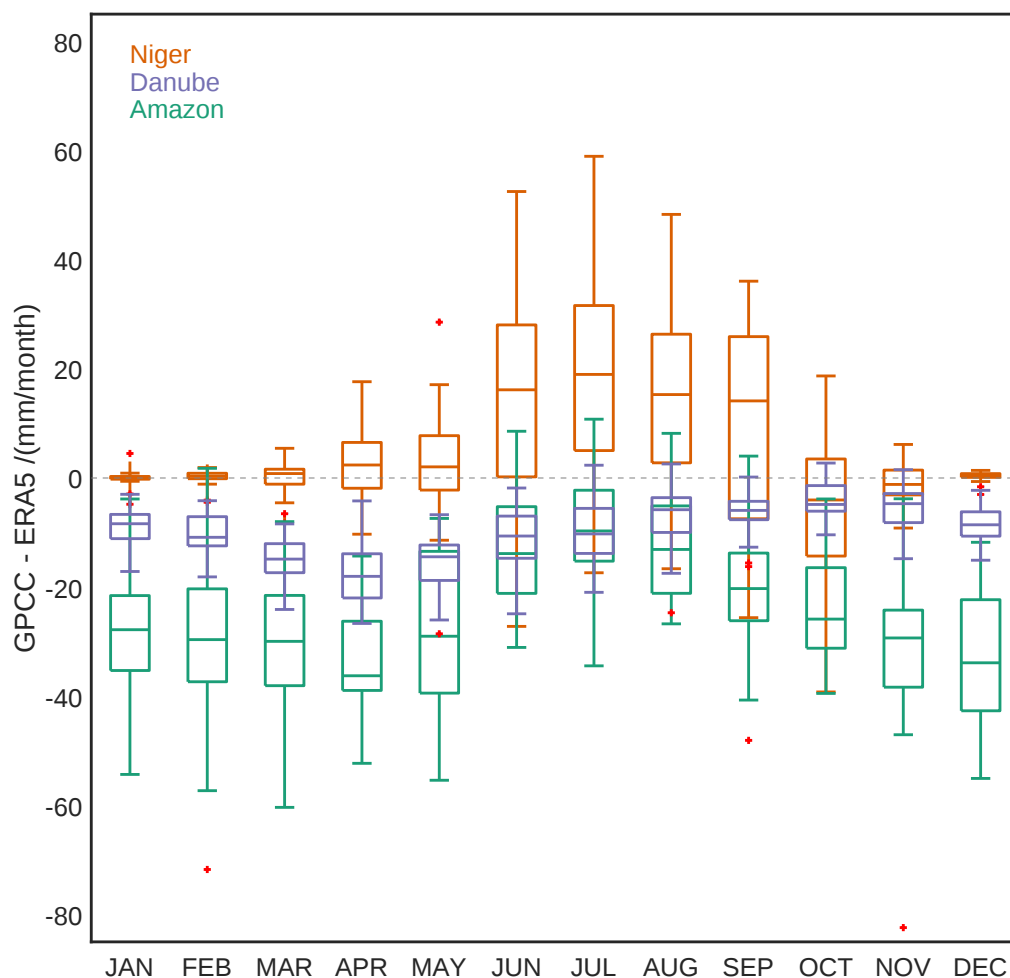
**Figure F1.** Terrestrial water budget monthly fluxes over the Niger, Amazon and Danube river catchments. The first row (a,b,c) shows in- and output expressed in equivalent water height from (1) precipitation from GPCC (solid blue), (2) river discharge from GRDC (green), and (3) water storage change from reconstructed TWS (purple). The second row (d,e,f) shows monthly  $\overline{ET}$  (black) as derived using the budget equation with the upper row components, and 'Actual Evaporation' from GLEAM (orange) for comparison, respectively. Positive fluxes represent net monthly water mass gain for the catchment and negative numbers represent net losses. The dotted blue curves in a–c shows precipitation from ERA5 reanalysis for comparison. And the dotted blue curves in d–f depict budget-derived  $\overline{ET}$  when ERA5 total precipitation instead of GPCC is used.

<https://doi.org/10.5194/essd-2025-461>  
Preprint. Discussion started: 30 October 2025  
© Author(s) 2025. CC BY 4.0 License.





## G0.1 GPCC and ERA5 precipitation comparison



**Figure G1.** Monthly climatological comparison of GPCC and ERA5 precipitation over the Niger-, Amazon- and Danube-catchments, derived from the monthly aggregated area-averaged precipitation timeseries for GPCC minus those for ERA5. Note how the (relative) ERA5 wet bias in the Amazon region can explain the offset in the budget-derived  $\overline{ET}$ , whereas the ERA5 dry-bias during peak-precipitation months in the summer in the Niger catchment leads to less pronounced budget-derived  $\overline{ET}$ . On each box, the central mark indicates the median, and bottom and top edges indicate the 25th to 75th percentile, respectively. Thin vertical lines represent the spread range, excluding outliers (red crosses).



640 *Author contributions.* CH: Conceptualization, Fomal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing (original draft preparation), Writing (review and editing); BDG: Investigation, Software, Validation, Visualization, Writing (original draft preparation), Writing (review and editing); AL: Data curation, Resources, Writing (review and editing) ; JK: Project administration, Conceptualization, Funding acquisition, Supervision, Writing (original draft preparation), Writing (review and editing)

*Competing interests.* The authors declare that they have no known competing financial interests or personal relationships that could have  
645 appeared to influence the work reported in this paper.

*Acknowledgements.* The authors acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 1502/1-2022—Project No. 450058266. We acknowledge the Global Runoff Data Centre, 56068 Koblenz, Germany for providing observational streamflow data and watershed boundaries.



## References

- 650 Allen, M. R. and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, *Nature*, 419, 224–232, <https://doi.org/10.1038/nature01092>, 2002.
- Attanasio, A.: Testing for linear Granger causality from natural/anthropogenic forcings to global temperature anomalies, *Theoretical and Applied Climatology*, 110, <https://doi.org/10.1007/s00704-012-0634-x>, 2012.
- Breiman, L.: Random Forests, *Machine Learning*, 45, <https://doi.org/10.1023/A:1010933404324>, 2001.
- 655 Brockmann, J. and Schuh, W. D.: Fast Variance Component Estimation in GOCE Data Processing, in: *Gravity, Geoid and Earth Observation*, edited by Mertikas, S. P., pp. 185–193, Springer, Berlin, Heidelberg, ISBN 978-3-642-10634-7, [https://doi.org/10.1007/978-3-642-10634-7\\_25](https://doi.org/10.1007/978-3-642-10634-7_25), 2010.
- Chambers, D. P., Wahr, J., Tamisiea, M. E., and Nerem, R. S.: Ocean mass from GRACE and glacial isostatic adjustment, *Journal of Geophysical Research: Solid Earth*, 115, <https://doi.org/10.1029/2010JB007530>, 2010.
- 660 Chandanpurkar, H. A., Hamlington, B. D., and Reager, J. T.: Global terrestrial water storage reconstruction using cyclostationary empirical orthogonal functions (1979 - present), <https://doi.org/10.5281/zenodo.6659543>, data set, 2022a.
- Chandanpurkar, H. A., Hamlington, B. D., and Reager, J. T.: Global Terrestrial Water Storage Reconstruction Using Cyclostationary Empirical Orthogonal Functions (1979–2020), *Remote Sensing*, 14, <https://doi.org/10.3390/rs14225677>, 2022b.
- Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Wayne Higgins, R., and Janowiak, J. E.: Assessing objective techniques for gauge-  
 665 based analyses of global daily precipitation, *Journal of Geophysical Research: Atmospheres*, 113, <https://doi.org/10.1029/2007JD009132>, 2008.
- Cheng, M. and Ries, J.: C20 and C30 Variations From SLR for GRACE/GRACE-FO Science Applications, *Journal of Geophysical Research: Solid Earth*, 128, <https://doi.org/10.1029/2022JB025459>, 2023.
- Coe, M. T., Costa, M. H., and Soares-Filho, B. S.: The influence of historical and potential future deforestation on the  
 670 stream flow of the Amazon River – Land surface processes and atmospheric feedbacks, *Journal of Hydrology*, 369, 165–174, <https://doi.org/10.1016/j.jhydrol.2009.02.043>, 2009.
- Cox, C. M. and Chao, B. F.: Detection of a Large-Scale Mass Redistribution in the Terrestrial System Since 1998, *Science*, <https://doi.org/10.1126/science.1072188>, publisher: American Association for the Advancement of Science, 2002.
- Douville, H., Ribes, A., Decharme, B., Alkama, R., and Sheffield, J.: Anthropogenic influence on multidecadal changes in reconstructed  
 675 global evapotranspiration, 3, 59–62, <https://doi.org/10.1038/nclimate1632>, number: 1 Publisher: Nature Publishing Group, 2013.
- Durack, P. J., Wijffels, S. E., and Matear, R. J.: Ocean Salinities Reveal Strong Global Water Cycle Intensification During 1950 to 2000, *Science*, 336, 455–458, <https://doi.org/10.1126/science.1212222>, publisher: American Association for the Advancement of Science, 2012.
- 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,  
 680 1285–1309, <https://doi.org/10.1007/s10712-014-9309-8>, 2014.
- Elke, R., Hänsel, S., Finger, P., Schneider, U., and Ziese, M.: GPCC Climatology Version 2022 at 0.25°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historical Data., [https://doi.org/10.5676/DWD\\_GPCC/CLIM\\_M\\_V2022\\_050](https://doi.org/10.5676/DWD_GPCC/CLIM_M_V2022_050), data set, 2022.
- Fan, Y. and van den Dool, H.: Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present, *Journal of Geophysical Research: Atmospheres*, 109, <https://doi.org/10.1029/2003JD004345>, 2004.



- Fan, Y. and van den Dool, H.: A global monthly land surface air temperature analysis for 1948–present, *Journal of Geophysical Research: Atmospheres*, 113, <https://doi.org/10.1029/2007JD008470>, 2008.
- Fecher, T., Pail, R., and Gruber, T.: Global gravity field modeling based on GOCE and complementary gravity data, *International Journal of Applied Earth Observation and Geoinformation*, 35, <https://doi.org/10.1016/j.jag.2013.10.005>, 2015.
- 690 Forootan, E., Kusche, J., Loth, I., Schuh, W.-D., Eicker, A., Awange, J., Longuevergne, L., Diekkrüger, B., Schmidt, M., and Shum, C. K.: Multivariate Prediction of Total Water Storage Changes Over West Africa from Multi-Satellite Data, *Surveys in Geophysics*, 35, 913–940, <https://doi.org/10.1007/s10712-014-9292-0>, 2014.
- Förstner, W.: Ein verfahren zur schätzung von varianz-und kovarianzkomponenten, *Allgemeine Vermessungsnachrichten*, 1979.
- Galdyn, F., Sosnica, K., Zajdel, R., Meyer, U., and Jäggi, A.: Long-term ice mass changes in Greenland and Antarctica derived from satellite  
 695 laser ranging, *Remote Sensing of Environment*, 302, 113 994, <https://doi.org/10.1016/j.rse.2024.113994>, publisher: Elsevier, 2024.
- Gerdener, H., Engels, O., and Kusche, J.: A framework for deriving drought indicators from the Gravity Recovery and Climate Experiment (GRACE), 24, 227–248, <https://doi.org/10.5194/hess-24-227-2020>, publisher: Copernicus GmbH, 2020.
- 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, [https://doi.org/10.1007/s00190-023-](https://doi.org/10.1007/s00190-023-01763-9)  
 700 01763-9, 2023.
- Gou, J. and Soja, B.: Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms, *Nature Water*, 2, 139–150, <https://doi.org/10.1038/s44221-024-00194-w>, number: 2 Publisher: Nature Publishing Group, 2024.
- Granger, C. W. J.: Investigating Causal Relations by Econometric Models and Cross-spectral Methods, *Econometrica*, 37, <https://doi.org/10.2307/1912791>, publisher: [Wiley, Econometric Society], 1969.
- 705 Gutknecht, B.: Catchment-Averaged Monthly Evaporation Timeseries 1984–2020 Derived from GRACE-like TWS Change via Terrestrial Water Budgets (1.0.0), <https://doi.org/10.5281/zenodo.16643628>, data set, 2025.
- Hacker, C.: Multidecadal statistical reconstruction of GRACE (Gravity Recovery And Climate Experiment) like terrestrial water storage anomalies (TWSA) incorporating geodetic tracking data (1.0.0), <https://doi.org/10.5281/zenodo.15827789>, data set, 2025.
- Hacker, C. and Kusche, J.: How realistic are multi-decadal reconstructions of GRACE-like total water storage anomalies?, *Journal of Hy-*  
 710 *drology*, 645, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2024.132180>, 2024.
- Han, S.-C., Sauber, J., Broerse, T., Pollitz, F., Okal, E., Jeon, T., Seo, K.-W., and Stanaway, R.: GRACE and GRACE Follow-On Gravity Observations of Intermediate-Depth Earthquakes Contrasted With Those of Shallow Events, *Journal of Geophysical Research: Solid Earth*, 129, <https://doi.org/10.1029/2023JB028362>, 2024.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Sim-  
 715 mons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Vil-  
 laume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, <https://doi.org/https://doi.org/10.1002/qj.3803>, 2020.
- 720 Hirahara, S., Ishii, M., and Fukuda, Y.: Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty, *Journal of Climate*, 27, <https://doi.org/10.1175/JCLI-D-12-00837.1>, publisher: American Meteorological Society Section: Journal of Climate, 2014.



- Hosseini-Moghari, S.-M., Araghinejad, S., Ebrahimi, K., Tang, Q., and AghaKouchak, A.: Using GRACE satellite observations for separating meteorological variability from anthropogenic impacts on water availability, 10, 1–12, <https://doi.org/10.1038/s41598-020-71837-7>, number: 1 Publisher: Nature Publishing Group, 2023.
- 725 Humphrey, V. and Gudmundsson, L.: GRACE-REC: A reconstruction of climate-driven water storage changes over the last century, <https://doi.org/10.6084/m9.figshare.7670849.v3>, data set, 2019a.
- Humphrey, V. and Gudmundsson, L.: GRACE-REC: a reconstruction of climate-driven water storage changes over the last century, *Earth System Science Data*, 11, 1153–1170, <https://doi.org/10.5194/essd-11-1153-2019>, 2019b.
- Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes, *Surveys in Geophysics*, 37, 357–395, <https://doi.org/10.1007/s10712-016-9367-1>, 2016.
- 730 Humphrey, V., Rodell, M., and Eicker, A.: Using Satellite-Based Terrestrial Water Storage Data: A Review, *Surveys in Geophysics*, <https://doi.org/10.1007/s10712-022-09754-9>, 2023.
- Huntington, T. G.: Evidence for intensification of the global water cycle: Review and synthesis, *Journal of Hydrology*, 319, 83–95, <https://doi.org/10.1016/j.jhydrol.2005.07.003>, 2006.
- 735 Jensen, L., Eicker, A., Dobslaw, H., Stacke, T., and Humphrey, V.: Long-Term Wetting and Drying Trends in Land Water Storage Derived From GRACE and CMIP5 Models, *Journal of Geophysical Research: Atmospheres*, 124, 9808–9823, <https://doi.org/10.1029/2018JD029989>, 2019.
- Jensen, L., Gerdener, H., Eicker, A., Kusche, J., and Fiedler, S.: Observations indicate regionally misleading wetting and drying trends in CMIP6, *npj Climate and Atmospheric Science*, 7, <https://doi.org/10.1038/s41612-024-00788-x>, 2024.
- 740 Jolliffe, J.: *Principal Component Analysis*, Springer Series in Statistics, Springer-Verlag, New York, ISBN 978-0-387-95442-4, <https://doi.org/10.1007/b98835>, 2002.
- Jäggi, A., Meyer, U., Lasser, M., Jenny, B., Lopez, T., Flechtner, F., Dahle, C., Förste, C., Mayer-Gürr, T., Kvas, A., Lemoine, J.-M., Bourgogne, S., Weigelt, M., and Groh, A.: International Combination Service for Time-Variable Gravity Fields (COST-G), pp. 57–65, [https://doi.org/10.1007/1345\\_2020\\_109](https://doi.org/10.1007/1345_2020_109), 2022.
- 745 Koch, K. R.: *Parameterschätzung und Hypothesentests in linearen Modellen*, Ehemals Ferd. Dümmlers Verlag, Bonn, 4 überarb. Aufl. edn., 2004.
- Koch, K. R.: Bayesian statistics and Monte Carlo methods, *Journal of Geodetic Science*, 8, <https://doi.org/10.1515/jogs-2018-0003>, publisher: De Gruyter Open Access, 2018.
- Koch, K.-R. and Kusche, J.: Regularization of geopotential determination from satellite data by variance components, *Journal of Geodesy*, 76, <https://doi.org/10.1007/s00190-002-0245-x>, 2002.
- 750 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*, pp. 83:903–913, 2009.
- 755 Kusche, J., Eicker, A., Forootan, E., Springer, A., and Longuevergne, L.: Mapping probabilities of extreme continental water storage changes from space gravimetry, *Geophysical Research Letters*, 43, 8026–8034, <https://doi.org/10.1002/2016GL069538>, 2016.
- 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.



- 760 Li, B., Rodell, M., Kumar, S., Beaudoin, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I. B., Daira, D., Bila, M., de Lannoy, G., Mocko, D., Steele-Dunne, S. C., Save, H., and Bettadpur, S.: 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.
- Li, F.: Data from: Long-term (1979–present) total water storage anomalies over the global land derived by reconstructing GRACE data, <https://doi.org/10.5061/dryad.z612jm6bt>, data set, 2021.
- 765 Li, F., Kusche, J., Chao, N., Wang, Z., and Löcher, A.: Long-Term (1979–Present) Total Water Storage Anomalies Over the Global Land Derived by Reconstructing GRACE Data, *Geophysical Research Letters*, 48, <https://doi.org/10.1029/2021GL093492>, 2021.
- Liu, R., Liu, Y., and Chen, J.: GLOBMAP global Leaf Area Index since 1981 (Version 3.0), <https://doi.org/10.5281/zenodo.4700264>, data set, 2021.
- 770 Liu, Y., Liu, R., and Chen, J. M.: Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data, *Journal of Geophysical Research: Biogeosciences*, 117, <https://doi.org/10.1029/2012JG002084>, 2012.
- Ljung, L.: System identification: theory for the user, Beijing: Tsinghua University Press, 1987.
- Loomis, B. D., Luthcke, S. B., and Sabaka, T. J.: Regularization and error characterization of GRACE mascons, *Journal of Geodesy*, 93, <https://doi.org/10.1007/s00190-019-01252-y>, 2019.
- 775 Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W., and Luthcke, S. B.: Replacing GRACE/GRACE-FO C30 With Satellite Laser Ranging: Impacts on Antarctic Ice Sheet Mass Change, *Geophysical Research Letters*, 47, <https://doi.org/10.1029/2019GL085488>, 2020.
- Löcher, A. and Kusche, J.: A hybrid approach for recovering high-resolution temporal gravity fields from satellite laser ranging, *Journal of Geodesy*, 95, 6, <https://doi.org/10.1007/s00190-020-01460-x>, 2021.
- Löcher, A., Kusche, J., and Nie, Y.: IGG-SLR-DORIS: Monthly gravity field solutions from SLR and DORIS, <https://doi.org/10.5880/ICGEM.2025.001>, data set, 2025a.
- 780 Löcher, A., Kusche, J., and Nie, Y.: A 40-year record of the Earth’s time-variable gravity field from SLR and DORIS, *Advances in Space Research*, pp. 1281–1291, <https://doi.org/10.1016/j.asr.2025.05.089>, 2025b.
- Mandal, N., Das, P., and Chanda, K.: Machine-learning-based reconstruction of long-term global terrestrial water storage anomalies from observed, satellite and land-surface model data, *Earth System Science Data*, 17, <https://doi.org/10.5194/essd-17-2575-2025>, publisher: Copernicus GmbH, 2025.
- 785 Martins, E. S. and Stedinger, J. R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, *Water Resources Research*, 36, <https://doi.org/10.1029/1999WR900330>, 2000.
- Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., and Zehentner, N.: ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE, <https://doi.org/10.5880/ICGEM.2018.003>, data set, 2018a.
- 790 Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., and Zehentner, N.: ITSG-Grace2018 - Monthly, Daily and Static Gravity Field Solutions from GRACE, <https://doi.org/10.5880/ICGEM.2018.003>, 2018b.
- Meyer, U., Jäggi, A., Dahle, C., Flechtner, F., Kvas, A., Behzadpour, S., Öhlinger, F., Mayer-Gürr, T., Lemoine, J.-M., Bourgogne, S., Lasser, M., Koch, I., Flury, J., Chen, Q., Wang, C., Yan, Z., Zhou, H., and Feng, W.: International Combination Service for Time-variable Gravity Fields (COST-G) Monthly GRACE/GRACE-FO RL02 Series, [https://doi.org/10.5880/COST-G.ICGEM\\_02\\_L2](https://doi.org/10.5880/COST-G.ICGEM_02_L2), data set, 2025.
- 795 Miralles, D. G., Brutsaert, W., Dolman, A. J., and Gash, J. H.: On the Use of the Term “Evapotranspiration”, *Water Resources Research*, 56, <https://doi.org/10.1029/2020wr028055>, 2020.



- Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W., Verhoest, N. E. C., and Haghdoust, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, *Scientific Data*, 12, <https://doi.org/10.1038/s41597-025-04610-y>, publisher: Nature Publishing Group, 2025.
- 800 Moore, P. and Williams, S. D. P.: Integration of altimetric lake levels and GRACE gravimetry over Africa: Inferences for terrestrial water storage change 2003–2011, *Water Resources Research*, 50, 9696–9720, <https://doi.org/10.1002/2014WR015506>, 2014.
- Nowack, P., Runge, J., Eyring, V., and Haigh, J. D.: Causal networks for climate model evaluation and constrained projections, *Nature Communications*, 11, 1–11, <https://doi.org/10.1038/s41467-020-15195-y>, number: 1 Publisher: Nature Publishing Group, 2020.
- Nowak, A., Zajdel, R., Gałdyn, F., and Sośnica, K.: Low-degree gravity field coefficients based on inverse GNSS method: insights into hydrological and ice mass change studies, *GPS Solutions*, 29, 1–16, <https://doi.org/10.1007/s10291-024-01760-1>, company: Springer  
 805 Distributor: Springer Institution: Springer Label: Springer Number: 1 Publisher: Springer Berlin Heidelberg, 2025.
- Palazzoli, I., Ceola, S., and Gentile, P.: GRAiCE: reconstructing terrestrial water storage anomalies with recurrent neural networks, *Scientific Data*, 12, <https://doi.org/10.1038/s41597-025-04403-3>, publisher: Nature Publishing Group, 2025.
- Papagiannopoulou, C., Miralles, D. G., Decubber, S., Demuzere, M., Verhoest, N. E. C., Dorigo, W. A., and Waegeman, W.:  
 810 A non-linear Granger-causality framework to investigate climate–vegetation dynamics, *Geoscientific Model Development*, 10, <https://doi.org/10.5194/gmd-10-1945-2017>, publisher: Copernicus GmbH, 2017.
- Pearlman, M., Arnold, D., Davis, M., Barlier, F., Biancale, R., Vasiliev, V., Ciufolini, I., Paolozzi, A., Pavlis, E. C., Sośnica, K., and Bloßfeld, M.: Laser geodetic satellites: a high-accuracy scientific tool, *Journal of Geodesy*, 93, 2181–2194, <https://doi.org/10.1007/s00190-019-01228-y>, company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 11 Publisher: Springer Berlin Heidelberg,  
 815 2019.
- Peltier, W. R., Argus, D. F., and Drummond, R.: Space geodesy constrains ice age terminal deglaciation: The global ICE-6G\_C (VM5a) model, *Journal of Geophysical Research: Solid Earth*, 120, <https://doi.org/10.1002/2014JB011176>, 2015.
- Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/10.1029/2002JD002670>, 2003.  
 820
- Rodell, M. and Reager, J.: Water cycle science enabled by the GRACE and GRACE-FO satellite missions, *Nature Water*, 1, 47–59, 2023.
- Schneider, U., Hänsel, S., Finger, P., Rustemeier, E., and Ziese, M.: GPCC Full Data Monthly Product Version 2022 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historical Data, [https://doi.org/10.5676/DWD\\_GPCC/FD\\_M\\_V2022\\_050](https://doi.org/10.5676/DWD_GPCC/FD_M_V2022_050), 2022.
- 825 Springer, A., Kusche, J., Hartung, K., Ohlwein, C., and Longuevergne, L.: New Estimates of Variations in Water Flux and Storage over Europe Based on Regional (Re)Analyses and Multisensor Observations, *Journal of Hydrometeorology*, 15, 2397–2417, <https://doi.org/10.1175/JHM-D-14-0050.1>, 2014.
- 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.
- 830 Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Sasgen, I., Famiglietti, J. S., Landerer, F. W., Chambers, D. P., Reager, J. T., Gardner, A. S., Save, H., Ivins, E. R., Swenson, S. C., Boening, C., Dahle, C., Wiese, D. N., Dobslaw, H., Tamisiea, M. E., and Velicogna, I.: Contributions of GRACE to understanding climate change, *Nature Climate Change*, 9, 358–369, <https://doi.org/10.1038/s41558-019-0456-2>, 2019.



- Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible  
835 detection using GRACE, *Journal of Geophysical Research: Solid Earth*, 103, <https://doi.org/10.1029/98JB02844>, 1998.
- Weigelt, M., Jäggi, A., Meyer, U., Arnold, D., Mayer-Gürr, T., Öhlinger, F., Sośnica, K., Ebadi, S., Schön, S., and Steffen, H.: Bridging  
the gap between GRACE and GRACE Follow-On by combining high–low satellite-to-satellite tracking data and satellite laser ranging,  
*Journal of Geodesy*, 98, 1–20, <https://doi.org/10.1007/s00190-024-01888-5>, company: Springer Distributor: Springer Institution: Springer  
Label: Springer Number: 9 Publisher: Springer Berlin Heidelberg, 2024.
- 840 Wold, S., Esbensen, K., and Geladi, P.: Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52,  
[https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9), 1987.
- Xiong, J., Abhishek, Xu, L., Chandanpurkar, H. A., Famiglietti, J. S., Zhang, C., Ghiggi, G., Guo, S., Pan, Y., and Vishwakarma,  
B. D.: ET-WB: water-balance-based estimations of terrestrial evaporation over global land and major global basins, 15, 4571–4597,  
<https://doi.org/10.5194/essd-15-4571-2023>, 2023.
- 845 Yoon, J.-H., Wang, S.-Y. S., Gillies, R. R., Kravitz, B., Hipps, L., and Rasch, P. J.: Increasing water cycle extremes in California and in  
relation to ENSO cycle under global warming, 6, 1–6, <https://doi.org/10.1038/ncomms9657>, number: 1 Publisher: Nature Publishing  
Group, 2015.
- 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.