Author Comments to Referee Comment 1

Manuscript number: EGUSPHERE-2025-5379

Title: High-resolution terrestrial water storage dynamics in Central Asia: Evaluating hydrological forcing datasets for GRACE downscaling

Dear Reviewer,

We are very grateful for your constructive comments and suggestions to our manuscript entitled “High-resolution terrestrial water storage dynamics in Central Asia: Evaluating hydrological forcing datasets for GRACE downscaling” [#EGUSPHERE-2025-5379]. These comments are very valuable and helpful for improving our manuscript. In the following, the comments are in black and numbered. Our responses are in blue, and quoted texts from the manuscript are in orange and italic.

General comments

This study focuses on the downscaling of terrestrial water storage change (TWSC) derived from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) missions to a daily temporal resolution and 1 km spatial resolution over the Naryn–Kara Darya basins and the Fergana Valley in Central Asia. The authors implement and refine a three-step downscaling framework that integrates GRACE observations with high-resolution hydrological forcing datasets of different natures (reanalysis- or observation-based), namely precipitation (P), evapotranspiration (E), and runoff (R). The objectives of the study are threefold: (i) to evaluate the sensitivity of the downscaling accuracy to different hydrological forcing datasets, (ii) to develop validation strategies in data-scarce regions, and (iii) to demonstrate the added value of high-resolution TWSC products for improving hydrological process understanding.

The manuscript addresses relevant scientific questions within the scope of *Hydrology and Earth System Sciences* (HESS) and evaluates recent methodological advances in GRACE downscaling. The paper present important bottleneck: the use of input in the physical downscaling while addressing indirect validation. The methods and assumptions are valid and overall presentation is well structured and clear. While the study shows potential, major revisions are required to meet the standards of the journal. In particular, the manuscript suffers from issues related to notation consistency, dataset description, methodological clarity, and the depth of the evaluation. The comments below are intended to help the authors improve the manuscript accordingly.

Response: We sincerely thank you for the thoughtful summary and constructive suggestions, which have helped us improve the manuscript substantially. In response, we have made the following main revisions:

(1) The notation has been standardized by consistently using TWSA/TWSC throughout the manuscript.

(2) The descriptions of ITSG-Grace2018 product and hydrological forcing datasets have been revised for improved clarity and detail.

(3) The assumptions and methodological procedures of downscaling have been more comprehensively described.

(4) An indirect evaluation using soil moisture data has been conducted and presented.

(5) Comparative analyses have been added to the precipitation event-based evaluation. Evaluation results have been organized for improved structural clarity.

Major comments

Notation and Conceptual Clarity

1. There are several instances of ambiguity in the notation used throughout the manuscript, particularly regarding: Total Water Storage (TWS), Total Water Storage Anomalies (TWSA), Total Water Storage Change (TWSC), and alternative notations such as S, dS, dTWS/dt, "TWS change" etc. This lack of consistency makes it difficult to follow the manuscript. I recommend adopting a single, consistent notation throughout the manuscript, including the main text, equations (Eqs. 1–4), and figure labels. Using either: TWSA / TWSC, or S / dS. A careful revision of the entire manuscript is required to resolve these ambiguities.

Response: Thanks for your suggestion. We have standardized the notation by consistently using TWSA/TWSC throughout the manuscript to avoid confusion. In the main text and in Figs. 4 – 11, the previous notation (S, dS, and "TWS change") have been replaced with TWSA and TWSC. In Figs. 4 – 10, the label dTWS/dt has been revised to TWSC.

For the equations, Equation (1) has been revised as "*GRACE products provide monthly terrestrial water storage anomalies relative to the long-term mean, defined as* $TWSA = TWS - \overline{TWS}$*. To analyze temporal variations in storage, we compute the time derivative of TWSA (i.e. TWSC) in units of mm per month using a centred difference scheme applied to the JPL mascon data (Kalu et al., 2024):*

$$TWSC|_{month} \approx \frac{TWSA_{m_{i+1}} - TWSA_{m_{i-1}}}{2} \quad (1)$$".

Equation (3) has been revised as "*Daily changes in terrestrial water storage (mm* d^{-1}*) are then derived from the water balance (WB) equation:*

$$TWSC|_{day} = TWSC_t = P_t - ET_t - (I - Q) \cdot R_t \quad (3)$$ ".

Equation (4) has been revised as "*Since daily fluxes are available, GRACE-based monthly TWSCs can be compared with WB-derived estimates. Following Rodell et al. (2004a) and Humphrey et al. (2023), monthly TWSCs based on WB-derived data are computed by Eq. (1) and*

$$TWSA_{m_i} = \frac{1}{d_e^{m_i} - d_s^{m_i} + 1} \sum_{d=d_s^{m_i}}^{d_e^{m_i}} \sum_{t=1}^{d} TWSC_t + TWSA_{d=0} \quad (4)$$".

Dataset Description

2. The ITSG-2018 dataset used for evaluation is not clearly introduced and should be described in Section 3.1, alongside the GRACE Mascon dataset. Furthermore, ITSG-2018 is available for periods beyond 2017 and can be downloaded from: https://ftp.tugraz.at/outgoing/ITSG/GRACE/ITSG-Grace_operational/daily_kalman/netcdf/.

Response: Thank you for your suggestion. We have added a description of the ITSG-Grace2018 dataset at the end of Sect. 3.1, reading "*In addition, we use the ITSG-Grace2018 daily gravity field solutions for evaluation purposes. The ITSG-Grace2018 product, developed by Graz University of Technology, provides a time series of constrained daily GRACE solutions (Kvas et al., 2019). The solutions are generated using a Kalman smoother approach based on GRACE Level-1B observations (Kurtenbach et al., 2012). An autoregressive (AR) model of order three is applied to*

express the spatio-temporal correlations between epochs. The dataset is provided on a 1° × 1° global grid and covers the period from 2002 to 2024.".

We have also downloaded the ITSG-Grace2018 dataset for the period 2017 – 2023, so that the comparison between our downscaled results and the ITSG daily solution can be performed over the entire study period (except for the data gap of GRACE/-FO). Accordingly, Figs. 5 and 6 have been changed as follows. The conclusions derived from the figure analysis remain unchanged despite the revisions.

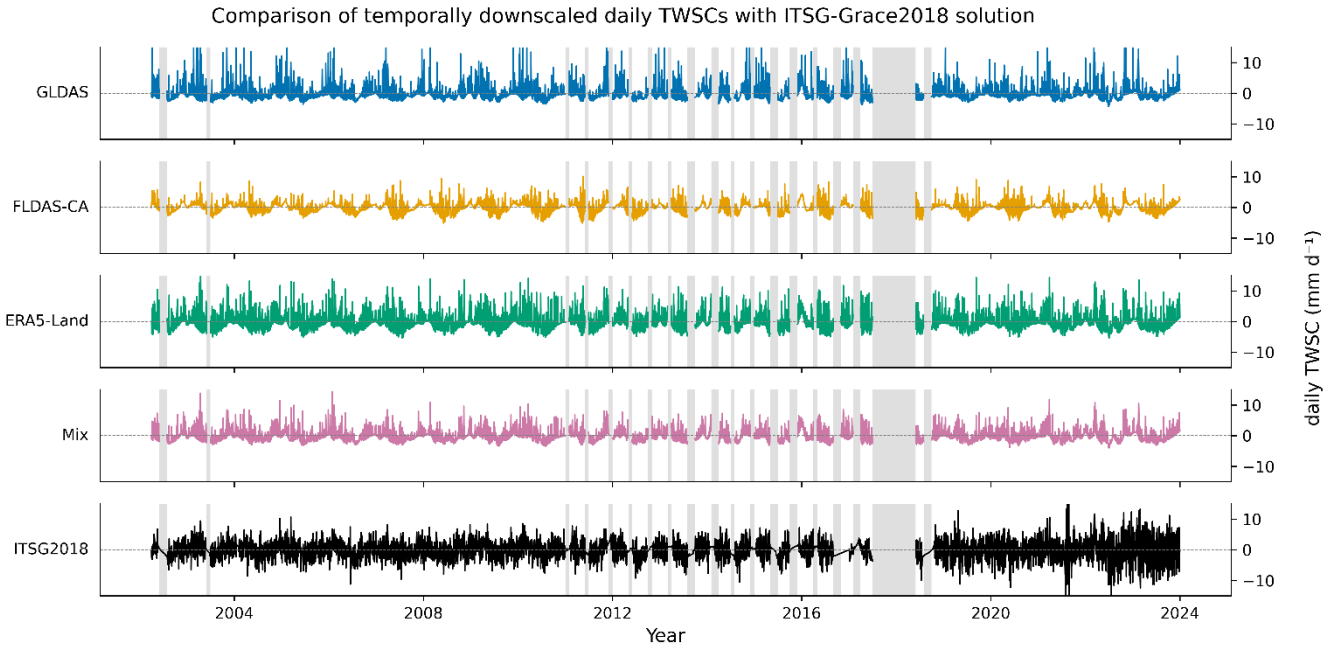

Figure 5: Results of temporal downscaling (Step 1) compared with the daily ITSG-Grace2018 solution. Shown are basin-averaged daily TWSCs at 0.5° resolution, computed from the four hydrological forcing scenarios (GLDAS, FLDAS-CA, ERA5-Land, and Mix) and from ITSG-Grace2018. Grey-shaded areas indicate periods when GRACE data are unavailable.

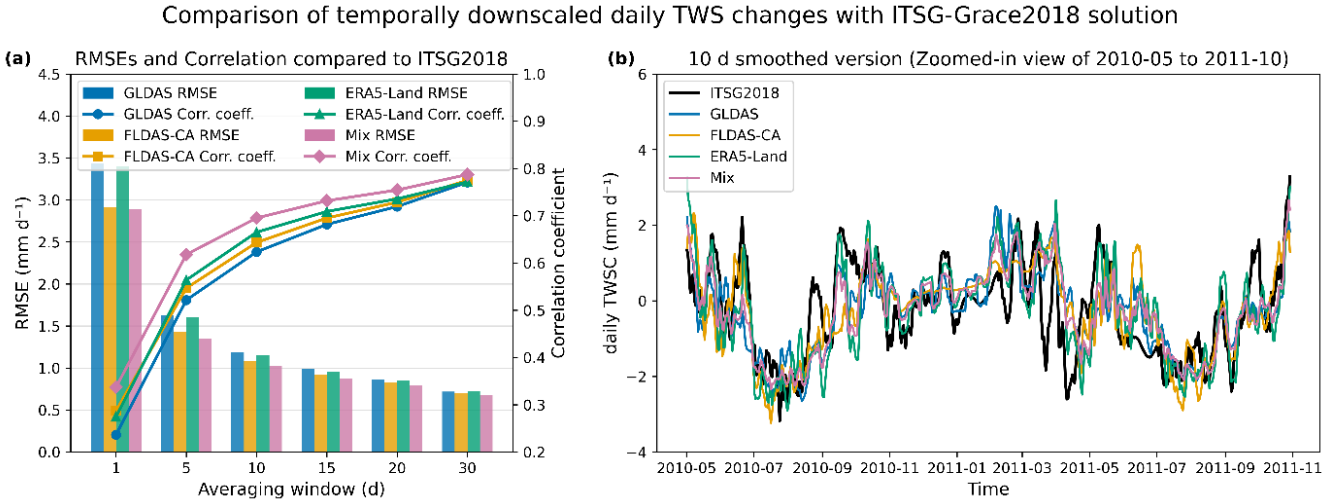

Figure 6: (a) RMSEs (left axis) and correlation coefficients (right axis) between daily TWSCs derived from temporal downscaling and from the ITSG-Grace2018 solution. Both metrics are computed for the entire study period. (b) Example time series of daily TWSCs between May 2010 and October 2011 from the four hydrological forcing scenarios and the ITSG-Grace2018 solution, smoothed using a 10 day moving window.

3. The hydrological datasets (Section 3.2) are insufficiently described. In particular, the nature of each dataset (observed, reanalysis-based, or model-derived) should be clearly stated, especially for precipitation and evapotranspiration. For evapotranspiration, the underlying model(s) should be specified. This information is crucial

when discussing the limitations and uncertainties associated with different forcing combinations and directly relates to the core objectives of the study. Also it is not clear how GLOFAS runoff differ from ERA5 runoff.

Response: Thank you for pointing this out. We have added descriptions of the nature of each dataset in Sect. 3.2 and summarized them in Table R1. We have also specified the underlying model of evapotranspiration and clarified how GloFAS runoff differs from ERA5 runoff in the section.

Table R1: The nature of each hydrological forcing dataset.

	Precipitation	Evapotranspiration	Runoff
GLDAS	observation-based	model-derived	model-derived
FLDAS-CA	observation-based	model-derived	model-derived
ERA5-Land	reanalysis-based	model-derived	model-derived
Mix (MSWEP, GLEAM, GloFAS)	observation-based	observation-based	model-derived

Section 3.2 has been revised as "

3.2.1 GLDAS

The Global Land Data Assimilation System (GLDAS) integrates satellite and ground-based observations with land surface modelling to provide globally consistent hydrological fields (Rodell et al., 2004b). We use P, ET, and R from the GLDAS Noah V2.1 model at 0.25° and 3-hourly resolution. The Noah land surface model (LSM) simulates surface energy and water fluxes driven by observation-based meteorological forcing including precipitation. Evapotranspiration and runoff are internally derived within the Noah LSM. Daily totals are obtained by summing 3-hourly values, and then resampled to 1 km × 1 km using cubic interpolation (Arshad et al., 2022).

3.2.2 FLDAS Central Asia

The Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) is a NASA-developed modelling framework designed for food and water security monitoring in data-scarce regions (McNally et al., 2022). The Central Asia configuration (FLDAS-CA) is based on the Noah land surface model forced by observation-based precipitation and other meteorological inputs. Evapotranspiration and runoff are model-derived outputs computed internally within the Noah LSM. The FLDAS-CA product provides daily P, ET, and R at 0.01° spatial resolution over the Central Asia domain (30 – 100°E, 21 – 56°N).

3.2.3 ERA5-Land

The land component of the Fifth Generation European ReAnalysis (ERA5-Land; Muñoz-Sabater et al., 2021) is a global land-focused reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It is generated using the H-TESSEL (Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land) land surface model forced by ERA5 atmospheric reanalysis, and provides hourly land surface variables at 0.1° spatial resolution. Precipitation is obtained from ERA5 reanalysis, while ET and R are simulated by H-TESSEL. In this study, daily P, ET, and R are derived from the daily accumulated outputs and subsequently resampled to 1 km × 1 km resolution.

3.2.4 MSWEP

The Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset combines gauge observations, satellite products, and reanalysis data to provide global precipitation from 1979 (Beck et al., 2019). We use daily data at 0.1°, interpolated to 1 km × 1 km.

3.2.5 GLEAM

The Global Land Evaporation Amsterdam Model (GLEAM) estimates terrestrial evapotranspiration and its components from remote sensing inputs combined with a simplified land surface model (Miralles et al., 2011). We use GLEAM V4.2a (Miralles et al., 2025) daily ET at 0.1°, resampled to 1 km × 1 km. The dataset is based on satellite and reanalysis data.

3.2.6 GloFAS

The Global Flood Awareness System (GloFAS), developed by the European Commission and ECMWF, is a global hydrological forecasting and monitoring system that provides gridded daily hydrological simulations. It is generated by forcing the open-source LISFLOOD hydrological model with ERA5 meteorological reanalysis data (Alfieri et al., 2013). We use daily runoff fields at 0.05° from version 4.0 and resample them to 1 km × 1 km. Unlike ERA5-Land runoff, which represents grid-scale surface and subsurface runoff generation within the H-TESSEL LSM, GloFAS runoff is produced by the LISFLOOD rainfall – runoff – routing model, in which runoff generated at each grid cell is routed through the river network using a kinematic wave approach.".

4. The authors should better emphasize that the MIX combination relies primarily on satellite observations, whereas the other combinations do not, which could explain why the downscaling performs better in this case. Indeed, MSWEP aggregates multiple datasets, several of which are based on Earth observations; GLEAM is derived from satellite and reanalysis data.

Response: Thank you for your suggestion. We have added the following text to Sect. 5.2:

"*The differences in downscaling performance can be attributed to the characteristics of the input hydrological datasets. The Mix combination relies more directly on observation-based datasets than the other combinations. MSWEP integrates gauge and satellite precipitation products, while GLEAM derives evapotranspiration primarily from satellite observations. In contrast, the other combinations represent internally consistent land surface model simulations driven by meteorological forcing. The stronger observational constraint in the Mix combination may partially explain its better performance in the downscaling.*"

Methodological Issues

5. In line 205, it is unclear whether the connectivity matrix Q accounts only for direct upstream ("parent") pixels or for all upstream contributors (i.e., the entire upstream catchment). This distinction is critical, as it determines whether runoff exchanges are evaluated locally or integrated over the full drainage area. In addition, the methodology implicitly treats runoff as a proxy for discharge. However, water mass conservation in hydrology is governed by discharge, not runoff. Considering daily runoff exchanges between adjacent pixels implies assumptions about water travel time (e.g., 1 km per day), whereas aggregating runoff over the full upstream area implicitly treats runoff as a net flux at the pixel scale. These assumptions should be explicitly stated and discussed in the manuscript.

Response: Thank you for this comment. In our implementation, the connectivity matrix Q encodes only direct downstream connectivity between neighboring pixels (D8 "parent – child" relationship), rather than the full upstream contributing area. Specifically, $Q(p, q) = 1$ only when pixel q drains directly into its direct downstream neighbor p. Therefore, runoff exchanges are evaluated locally between adjacent pixels along the drainage network.

Here, we use gridded runoff as a proxy for lateral outflow at the pixel scale. Under this approximation, $(I - Q) \cdot R_t$ represents the discrete divergence of runoff, i.e. the difference between the runoff leaving a pixel and the runoff entering it from its direct upstream neighbors. This formulation implicitly assumes that, at the chosen spatial (~1 km) and temporal (daily) resolution, the effective travel time for lateral transfer between a pixel and its direct downstream neighbor is treated as occurring within the daily time step.

We have added "*In this formulation, Q encodes only direct neighbor-to-neighbor connectivity (i.e., immediate upstream contributors), and does not aggregate contributions over the entire upstream catchment. R_t denotes grid-cell runoff rather than routed river discharge. The lateral exchange term $(I - Q) \cdot R_t$ therefore represents the discrete divergence of runoff under the adopted D8 connectivity, i.e. the imbalance between runoff leaving a pixel and that entering it from its direct upstream neighbours. This implicitly assumes that lateral transfer between adjacent pixels is not resolved explicitly and is treated as occurring within the daily time step at the ~1 km resolution.*" to the METHODS section.

6. In the temporal downscaling step (Section 4.2.1), it is unclear how months with missing GRACE TWSC data prior to 2017 are handled. Figure 5 does not indicate these gaps, although they exist. Moreover, the time series suggests that linear interpolation at the monthly scale may have been applied prior to temporal downscaling (e.g., in 2017). I recommend masking all months with missing GRACE data rather than interpolating them, and clearly documenting this choice in both the methodology and the figures.

Response: Thank you for the comment and suggestion. We have revised the temporal downscaling procedure such that months without GRACE observations are masked rather than interpolated. Figure 5 has been updated accordingly (see Response to Comment 2) to explicitly indicate these missing periods. We have also clarified this treatment in METHODS section as "*Months without GRACE observations are masked.*".

7. The description of the spatial downscaling procedure (Section 4.2.2) should be expanded as the manuscript should provide a self-contained explanation of the main principles and assumptions underlying the spatial disaggregation the methodology.

Response: Thanks for the remark. We agree that this section lacked a sufficiently coherent and self-contained explanation of the underlying methodology, as key steps were previously provided only in the Supplement. We have revised Sect. 4.2.2 accordingly as

"*Step 2 applies a Partial Least Squares (PLS) regression to relate GRACE-derived TWSA at coarse resolution to WB-based estimates at fine resolution. PLS is chosen for its ability to exploit spatial gradients and covariance structures across entire images. The regression establishes a statistical relationship between the temporally downscaled 0.5° TWSA and the 1 km WB-based predictors. The method assumes that the relationship between coarse-scale GRACE storage and fine-scale WB patterns is linear, scale-consistent, and temporally stationary. The detailed steps (Vishwakarma et al., 2021; Pellet et al., 2024) are as follows.*

Let $L \in R^{d \times m}$ denote the temporally downscaled daily TWSCs at coarse resolution, and $F \in R^{d \times n}$ the corresponding daily WB-based TWSCs at 1 km resolution, where d is the number of days in the study period, and m and n are the numbers of grid cells at coarse and fine resolutions, respectively. The cross-covariance matrix between coarse- and fine-scale fields is computed as

$$C = L^T F \quad (5)$$

Singular Value Decomposition (SVD) is then applied to extract the dominant coupled modes:

$$C = U_C \Sigma V_C^T \quad (6)$$

where U_C and V_C represent the canonical modes, and Σ is a diagonal matrix containing covariance between L and F.

The prediction matrix $\widehat{H}$ is subsequently constructed as follows:

$$U_L = LU_C, \quad (7)$$

$$\widehat{K} = (U_L^T U_L)^{-1} U_L^T F, \quad (8)$$

$$\widehat{H} = U_C \widehat{K}. \quad (9)$$

Finally, the spatially downscaled daily TWSC matrix M are obtained as

$$M = L\widehat{H}. \quad (10)$$".

Limitations of the Analysis and Evaluation

8. The evaluation presented in Figure 9 could be moved to the appendix as this analysis does not really focus on the high-resolution product and mainly serves as a sanity check, since the monthly GRACE Mascon dynamics are imposed by construction on the downscaled product. Therefore trends and dynamics at monthly and longer time scales are linked to the original GRACE data.

Response: Thank you for this suggestion. We agree that the analysis of long-term trend and annual amplitude primarily serves as a sanity check. To improve the focus of the manuscript on the added value of the high-resolution product, Figure 9 and the related analysis have been moved to the Appendix.

9. A disentanglement of the FLDAS-CA bias into its individual components (E, P, and R) could strengthen the analysis by highlighting the sensitivity to each hydrological variable.

Response: Thank you for this valuable suggestion. We quantified the bias of FLDAS-CA relative to the Mix forcing by decomposing the basin-averaged TWSC difference into contributions from P, ET, and R based on the water balance equation (TWSC = P – ET - R). Basin-mean P, ET, R, and TWSC over the study period were first computed for both FLDAS-CA and the Mix forcing, yielding a TWSC difference of +303 mm yr^{-1}. The corresponding differences in P, ET, and R are −117, +285, and +135 mm yr^{-1}, respectively.

These results show that the weaker water storage decline simulated by FLDAS-CA relative to the Mix forcing (+303 mm yr^{-1}) is mainly explained by the evapotranspiration component (+285 mm yr^{-1}), with runoff playing a secondary role (+135 mm yr^{-1}). In contrast, precipitation exerts a small compensating effect (−117 mm yr^{-1}). In other words, the

underestimation of long-term water storage loss in FLDAS-CA primarily arises from reduced evaporative and runoff losses rather than from enhanced precipitation input.

10. While the evaluation against the ITSG-2018 dataset is highly valuable (particularly Figure 6, which shows that downscaling relying on satellite observations (MIX) can recover sub-weekly processes (correlation > 0.6)), I encourage adding an indirect evaluation using soil moisture data for three reasons:

1. Blank et al. (2023, https://doi.org/10.5194/hess-27-2413-2023) demonstrated that HR GRACE products provide insights into subsurface water storage dynamics. Exploring similar relationships with downscaled TWSC would be highly informative.

2. Recent soil moisture products are now available at daily, 1 km resolution (e.g., Brocca et al., 2025; https://doi.org/10.1016/j.scitotenv.2024.174087), enabling co-evaluation at comparable spatio-temporal scales.

3. Soil moisture evaluation could help disentangle precipitation-driven and irrigation-driven storage changes, particularly relevant for the Fergana Valley.

Response: Thanks for this valuable suggestion. We have conducted an indirect evaluation using soil moisture data. Specifically, we use the ESACCI-Zheng product (Zheng et al., 2023; described in Brocca et al., 2024), which provides global gap-free surface soil moisture at daily 1 km resolution for the period 2000 – 2020. It enables evaluation at comparable spatio-temporal scales. The dataset can be downloaded at https://data.tpdc.ac.cn/zh-hans/data/30131436-88d1-4be3-8e3d-14905a29d6d6. Daily soil moisture data is stored in MOIDIS Sinusoidal Tile Grid. For our study region, four tiles h23v04, h23v05, h24v04, h24v05 for each year are downloaded.

Due to the temporal coverage of soil moisture (SM) data, the analysis is restricted to April 2002 – December 2020. TWSC is converted to TWS (here simply referred to as TWS although it represents storage relative to TWS at epoch 0) because TWS is comparable with SM. Prior to the comparison, the linear trend and long-term mean are removed from the time series of SM and TWS. The resulting detrended series are hereafter referred to as SM and TWS anomalies. It should be noted that a direct comparison of absolute values between SM and TWS is not feasible due to differences in integration depths and physical units. Therefore, the Pearson's correlation coefficient is used to analyze their relationship. And time lags between TWS and SM anomalies are computed using a cross-correlation analysis.

Figure R1 presents the basin-averaged time series of soil moisture and TWS anomalies from the downscaled products in the four hydrological forcing scenarios (GLDAS, FLDAS-CA, ERA5-Land, and Mix). TWS exhibits comparatively smooth behavior at short timescales and a dominant seasonal signal, whereas the SM time series shows substantially higher variability at short timescales.

A general correspondence is observed between TWS and SM anomalies in terms of their seasonal dynamics, with a strong rise during spring followed by a quick decline, indicating strong climatic control on terrestrial water variability. In addition to the primary seasonal peak, a minor secondary maximum is observed in the SM time series during October and November. Correspondingly, a more rapid increase in TWS can partially be observed during this period. This can be attributed to reduced evapotranspiration after the summer period, which allows water to accumulate. The correlation coefficients between TWS and SM are 0.50, 0.45, 0.46 and 0.48 for the four downscaled products (GLDAS, FLDAS-CA, ERA5-land and Mix), respectively. Their relative performance is consistent with that already analyzed in the main manuscript. The GLDAS performs well in the basin-averaged values. The Mix downscaled

product's overall performance is stable according to evaluation shown in the main manuscript and this additional evaluation with SM data.

The time shift is identified by computing the cross-correlation between two time series, resulting in a lag of approximately one month (31 – 34 days for the four downscaled results). The seasonal maxima and minima of the TWS time series occur about one month later than those of SM. This is reasonable and can be attributed to the slower and delayed response of deeper water storage components observed by GRACE, whereby transitions between dry and wet conditions evolve more slowly than in near surface soil layers.

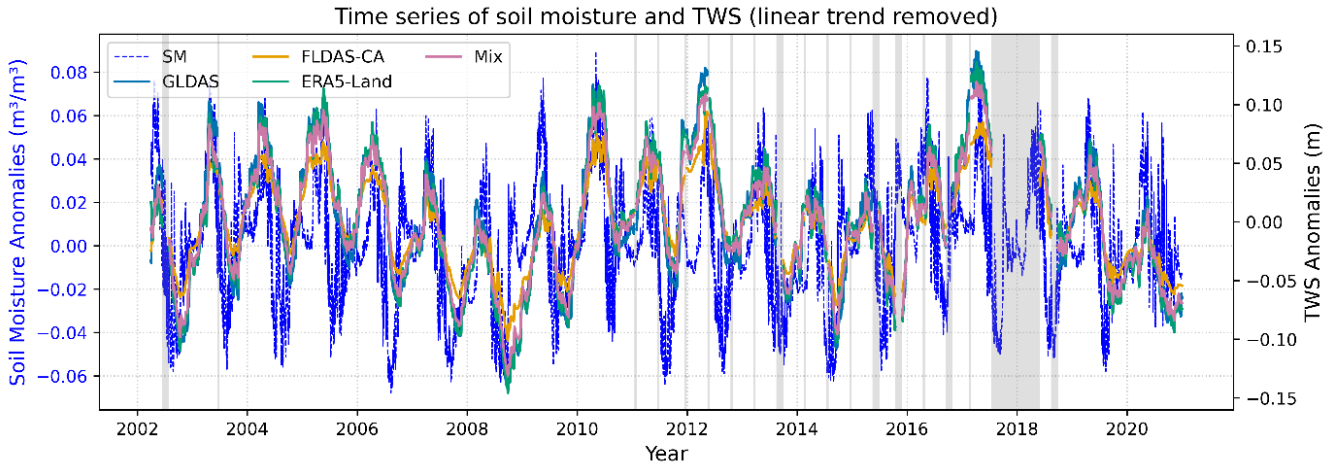

Figure R1: Time series of soil moisture (left axis) and TWS anomalies derived from the downscaled products in the four hydrological forcing scenarios (GLDAS, FLDAS-CA, ERA5-Land, and Mix) (right axis), averaged over the study region for the period 2002 – 2020. Grey-shaded areas indicate periods when GRACE data are unavailable.

Figure R2 displays the correlation coefficients between daily time series of TWS from downscaled products and SM data for each grid cell along with the corresponding time shifts derived from cross-correlation analysis. All downscaled products exhibit predominantly positive correlations with SM across the study region. Negative correlations are mainly found in areas containing reservoirs (Toktogul Reservoir), lakes (Song Kul Lake), and southern glacier regions, which are known to be challenging for reliable SM observations.

The spatial patterns of time shifts indicate that TWS variations generally lag behind SM, as reflected by the dominance of positive time shifts (green shading). This lag is particularly pronounced in the central and southern regions, where delays of approximately 30 – 90 days are common. Such behavior reflects the slower and integrated response of terrestrial water storage relative to surface soil moisture, likely mediated by subsurface processes and soil water redistribution. It should be noted that time shifts reaching ±120 days do not necessarily indicate an exact lag of 120 days, but rather reflect cases where the maximum allowable lag is exceeded. This suggests weak correlations in these regions, which generally correspond to areas with low or negative correlation coefficients in the correlation map (left column).

It can be observed that the correlation between two variables in highly irrigated area is relatively lower than that in the surrounding area, and the lag of TWS relative to SM is also longer. This suggests that additional water redistribution processes may be present in highly irrigated area, e.g., potential return flows from surface water irrigation.

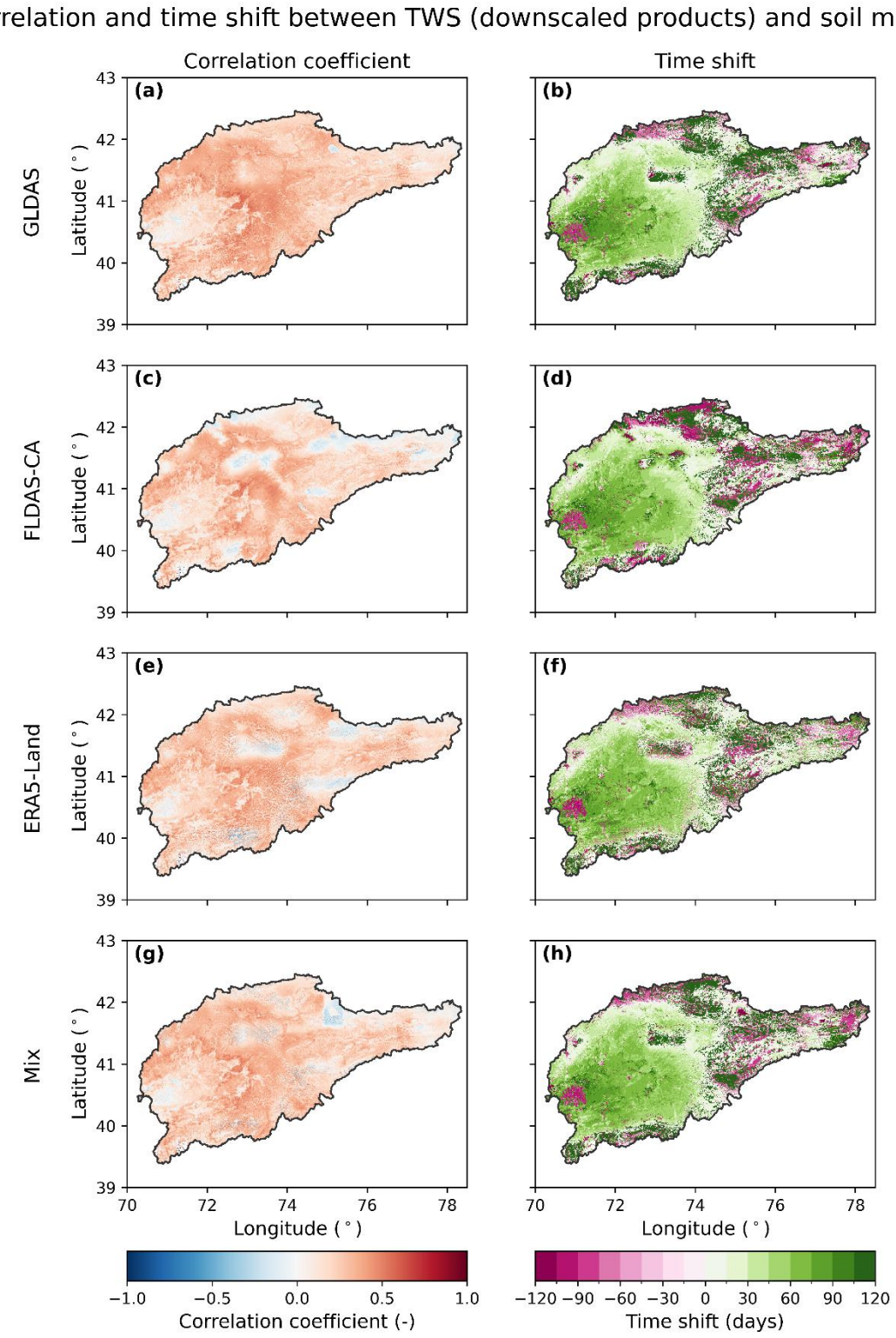

Figure R2: Spatial distribution of correlation coefficients and time shifts between TWS anomalies derived from the downscaled products in the four hydrological forcing scenarios (GLDAS, FLDAS-CA, ERA5-Land, and Mix) and SM anomalies at each grid point. Positive time shifts mean that TWS is delayed in comparison to SM.

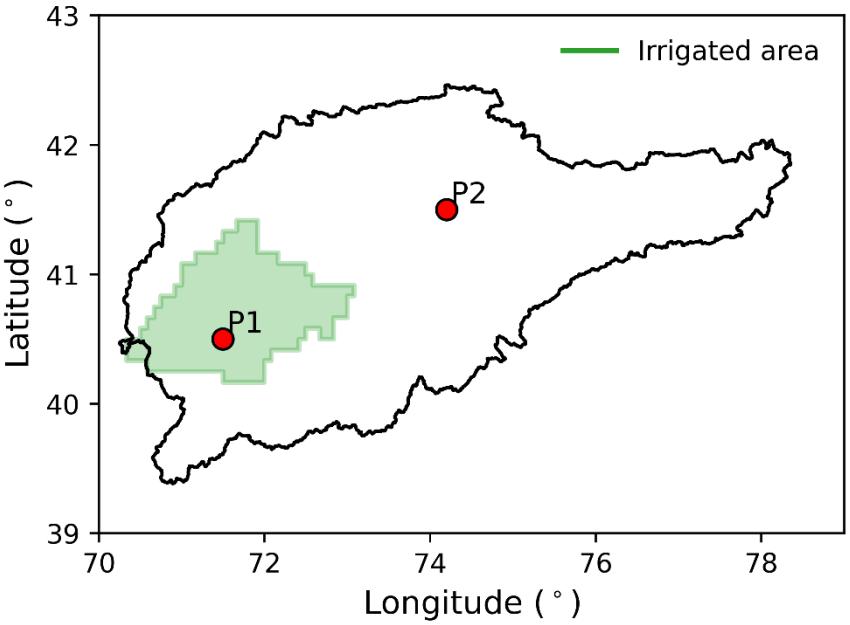

Figure R3: The study region and highly irrigated area. P1 and P2 are chosen to show SM and TWS time series.

We choose two grid cells, P1 and P2 as shown in Fig. R3, located in highly irrigated and non-irrigated area, respectively. P2 is located at an elevation of approximately 2000 m, on moderately sloping terrain that is typical in non-irrigated area. It is situated far from glacier region and is therefore not directly influenced by glacier processes. Figures R4a and R4c show the daily TWS anomalies in these grid cells in comparison with SM. Seasonal patterns are clearly evident at both locations. The seasonal cycle at P2 is comparatively more moderate than that observed at P1. For both sites, SM exhibits substantially higher variability than TWS products, with pronounced short-term fluctuations superimposed on a clear seasonal cycle. In contrast, all TWS datasets (GLDAS, ERA5-Land, FLDAS-CA, and Mix) display smoother temporal variations, reflecting their integrated nature. Despite these differences in variability, the temporal evolution of TWS anomalies is broadly consistent with that of SM. However, a temporal lag between TWS and SM is observed at both sites. TWS variations generally lag behind SM. This lag is more pronounced at P1 (70 – 80 days for the four downscaled products) than at P2 (20 – 30 days for the four downscaled products), indicating potential spatial differences in hydrological processes.

Figures R4b and R4d illustrate the corresponding daily climatology. For both sites, TWS reaches its maximum during late spring to early summer (approximately April – June), followed by a gradual decline towards autumn and a recovery during winter. This seasonal variation is consistent across all downscaled products, although ERA5-Land generally exhibit slightly higher amplitudes compared to others. In contrast, SM shows a different seasonal phase relative to TWS. At P1, SM peaks earlier in the year (late winter to early spring) and declines during summer, reaching minimum values around August – September. At P2, SM reaches its maximum around April, followed by a decline through summer, with minimum values also occurring in August – September. Subsequently, SM exhibits a secondary peak around November, after which it decreases slightly and remains relatively stable from December to March. We recognize during winter, microwave-based soil moisture retrievals are known to be unreliable due to frozen soil and snow cover, often resulting in missing or underestimated values. As temperatures rise in early spring, the rapid thawing of the soil and snowpack leads to a sudden increase in liquid water availability, which is captured by the satellite signal as a sharp rise in soil moisture. However, SM at P1 remains consistently high throughout winter

(December to March). This high level indicates a prolonged wet state during the cold season, suggesting stronger water accumulation prior to winter, potentially associated with moisture accumulation following the summer irrigation.

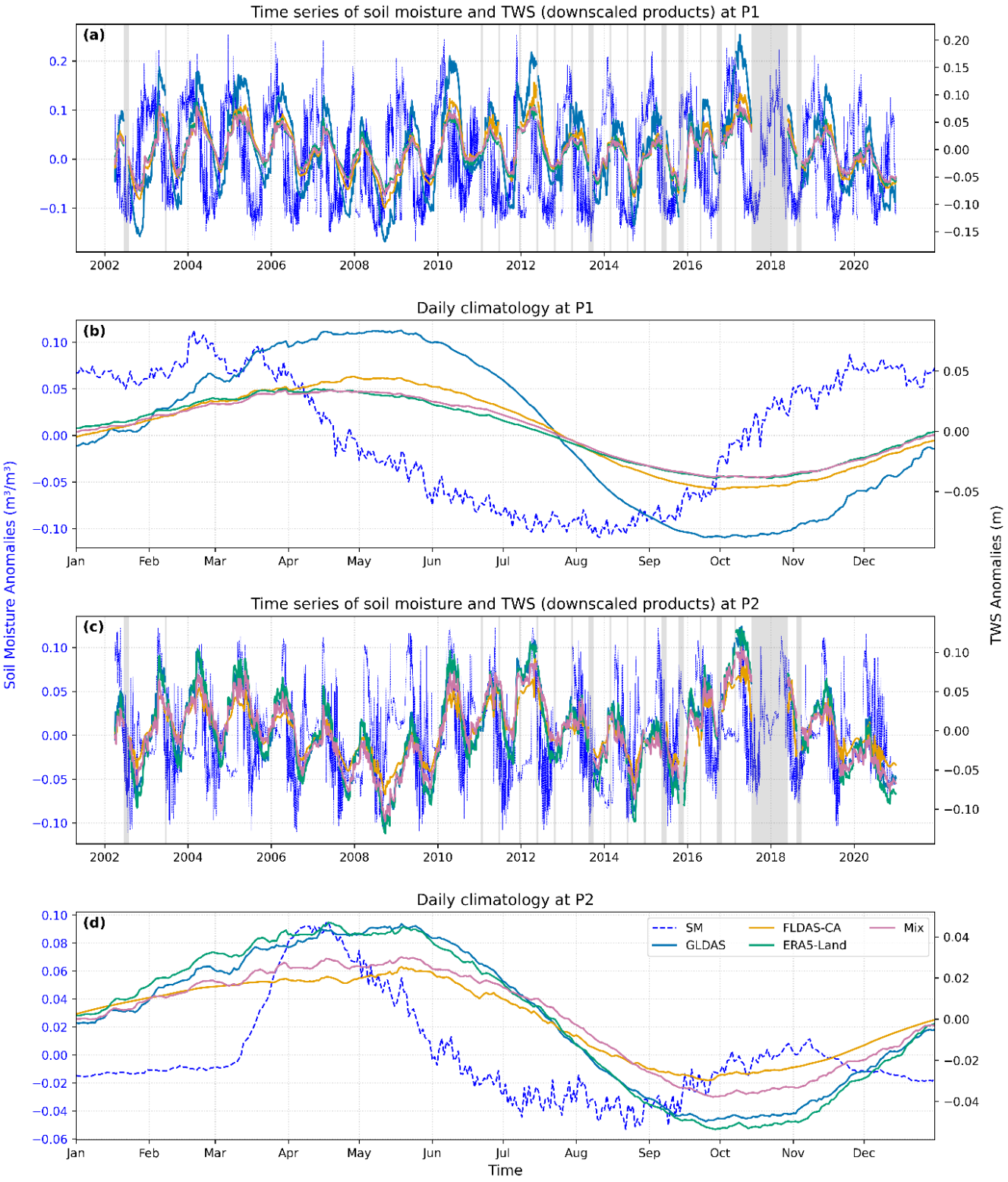

Figure R4: Time series of SM and TWS anomalies from four downscaled products (GLDAS, FLDAS-CA, ERA5-Land, and Mix) at P1 and P2. The linear trend and long-term mean have been removed from all time series. (a) SM (left axis) and TWS anomalies (right axis) from downscaled products at P1. (b) Daily climatology of SM (left axis) and TWS anomalies (right axis) at P1. (c) SM (left axis) and TWS anomalies (right axis) from downscaled products at P2. (d) Daily climatology of SM (left axis) and TWS anomalies (right axis) at P2.

To isolate sub-monthly variations, following Blank et al. (2023), a third-order Butterworth high-pass filter with a cutoff frequency of 30 d is applied in the forward and backward directions. This filter preserves the phase while removing signals with periods longer than 30 days, which dominate the original time series. Figure R5 illustrates the high-pass filtered (sub-monthly) variations of SM and TWS from the four downscaled products (GLDAS, FLDAS-CA, ERA5-Land, and Mix) at P1 and P2 between May 2010 and October 2011, highlighting short-term hydrological fluctuations after removing low-frequency components.

Compared with the total signals shown in Figs. R4a and R4c, the high-frequency component of SM accounts for a substantially larger proportion of the total signal than that of TWS at both P1 and P2. SM exhibits larger and more frequent sub-monthly fluctuations than TWS. This reflects its rapid response to precipitation and evapotranspiration at the land surface, whereas TWS integrates water storage changes across soil, groundwater, and surface water components, resulting in attenuation of high-frequency signals. Moreover, the intrinsic characteristics of GRACE observations, including their coarse spatial resolution and inherent temporal smoothing, further reduce the sensitivity of TWS to sub-monthly variability. The correlations between the high-pass filtered TWS and SM time series are considerably lower than those of the unfiltered time series. This is expected, as the seasonal storage variations that largely drive high correlations for the unfiltered time series are not present anymore in the filtered ones. Nevertheless, the correlation between SM and TWS at sub-monthly timescales remains moderate (approximately $r \approx 0.3$), with the Mix forcing achieving the highest values around 0.35 at both sites. This suggests that downscaled TWS is still able to capture part of the day-to-day variability in SM.

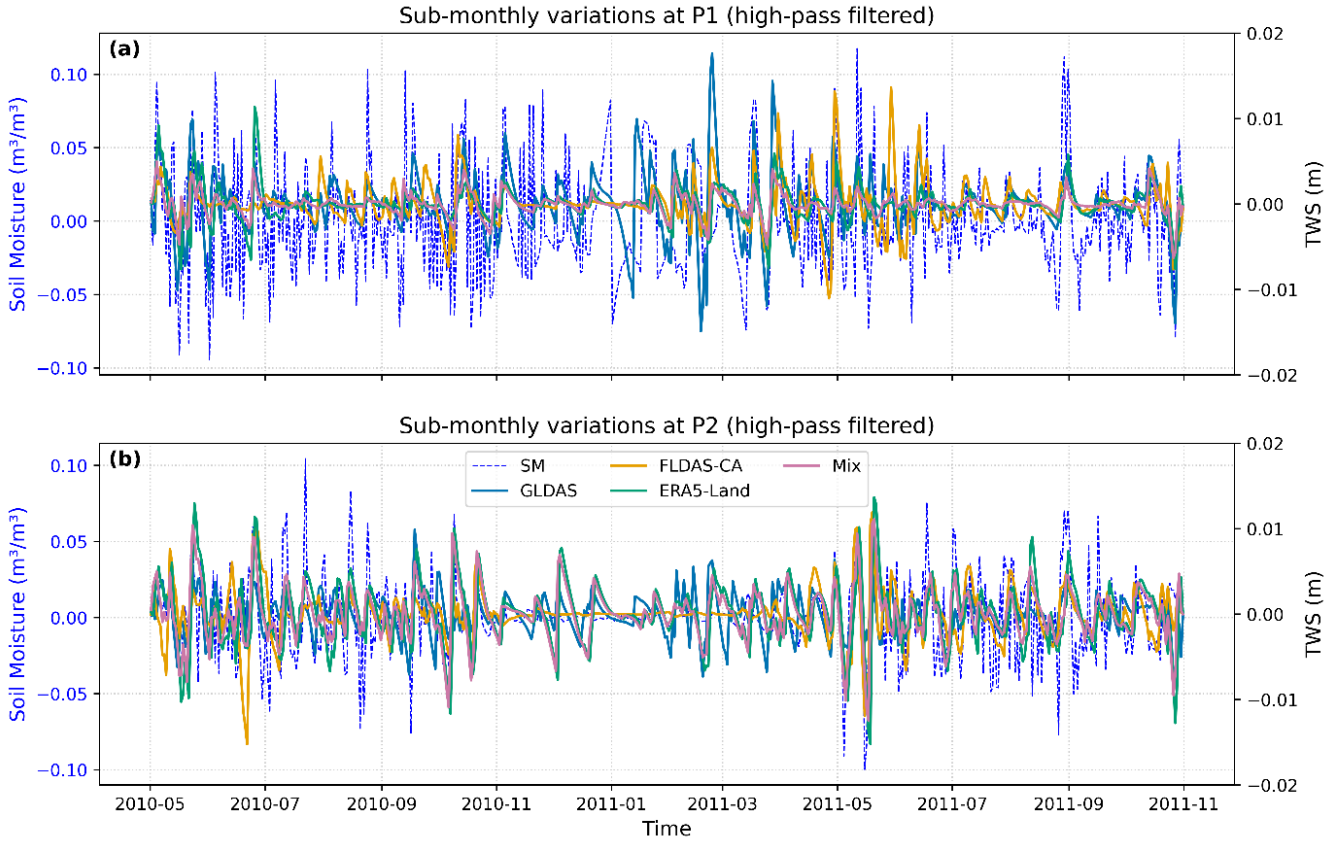

Figure R5: High-pass filtered time series of SM and TWS anomalies from four downscaled products (GLDAS, FLDAS-CA, ERA5-Land, and Mix) at P1 and P2.

To address the third point raised in this comment by the reviewer, we computed the correlation coefficients among daily soil moisture changes (ΔSM), daily TWSC from the Mix downscaled product, and daily P in highly irrigated and non-irrigated regions. This is also a supplementary analysis to Fig. 10 in the main manuscript. ΔSM is used instead of SM to ensure comparability with P, which is defined as a daily flux. As shown in Fig. R6, in both regions, TWSC exhibits a strong positive correlation with P (r = 0.80/0.81), indicating that precipitation is one of the dominant drivers of terrestrial water storage variability in the downscaling. This strong relationship is consistent across irrigated and non-irrigated areas, suggesting that large-scale water storage dynamics are primarily controlled by hydrological forcing inputs.

In contrast, the correlations between ΔSM and both TWSC and P are notably weaker. In highly irrigated areas (Fig. R6a), ΔSM shows low correlations with TWSC (r = 0.16) and P (r = 0.15). This weak relationship likely reflects the influence of irrigation, which introduces additional water inputs independent of precipitation. In non-irrigated areas (Fig. R6b), the correlations between ΔSM and both TWSC and P are slightly higher ($r \approx 0.22$). The absence of irrigation allows soil moisture to respond more directly to precipitation and to co-vary more consistently with terrestrial water storage.

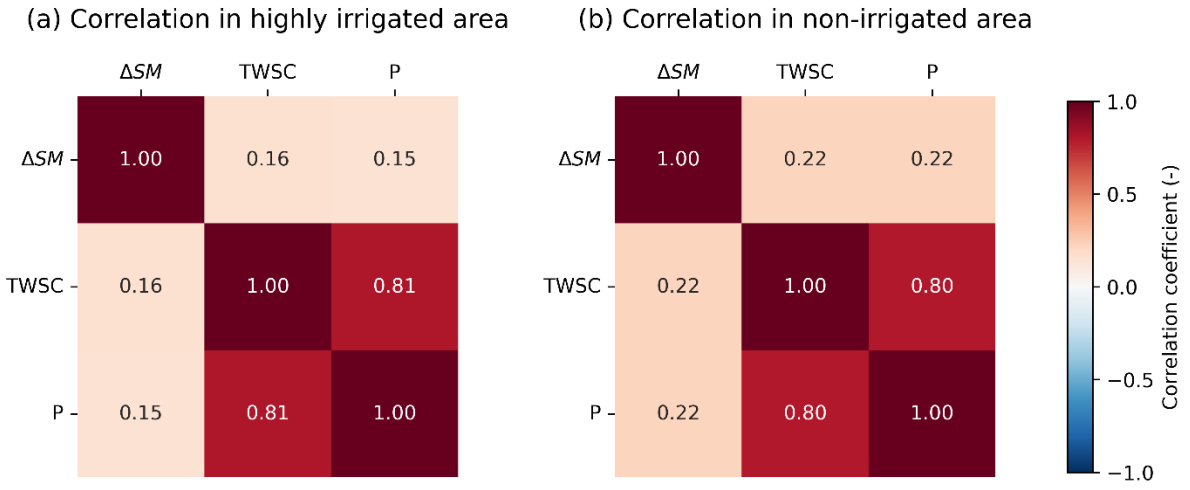

Figure R6: Correlation coefficients among region-averaged ΔSM, TWSC, and P from Mix downscaled product in highly irrigated and non-irrigated areas.

In addition to the evaluation using soil moisture data, we include two existing GRACE downscaling datasets (Gou and Soja, 2024; Li and Kusche, 2026) for comparison. To the best of our knowledge, these are currently the only two open-access global downscaled total water storage datasets available. The correlation coefficients and root mean square error (RMSE) between the downscaled products and the JPL mascon are computed at each mascon grid point, as shown in Fig. R7. This is also a supplementary analysis to Fig. 8 in the main manuscript.

All downscaled results in our study (Figs. R7a, c, e, g) exhibit consistently high correlations with the JPL mascon, with values approaching or exceeding 0.8 across most of the study region. The mean correlation coefficients are 0.88, 0.79, 0.82, and 0.85 for the GLDAS, FLDAS-CA, ERA5-Land, and Mix downscaled results, respectively. In comparison, the result of Gou and Soja (2024) in Fig. R7i displays lower and more spatially heterogeneous correlations, particularly in the central and northern regions, with a mean value of 0.76. The results of Li and Kusche

(2026) in Fig. R7k shows uniforms high correlations across the domain, with a mean value of 0.98, exceeding those of other downscaled products.

The RMSE patterns (Figs. R7b, d, f, h) further differentiate the performance of these datasets. The downscaled results in our study generally exhibit low to moderate RMSE values, with relatively higher errors concentrated in the southern region. The basin-averaged RMSE values are 1.90, 2.44, 2.04, and 1.94 cm for the GLDAS, FLDAS-CA, ERA5-Land, and Mix downscaled results, respectively. In contrast, the result of Gou and Soja (2024) in Fig. R7j shows substantially higher RMSE values across most of the region, with a mean value of 8.42 cm. The result of Li and Kusche (2026) in Fig. R7l exhibits pronounced spatial variability in RMSE (mean value of 4.52 cm), with particularly high errors in the central areas, despite its high correlation values. This suggests that while this product captures temporal variability well, it may suffer from amplitude mismatches.

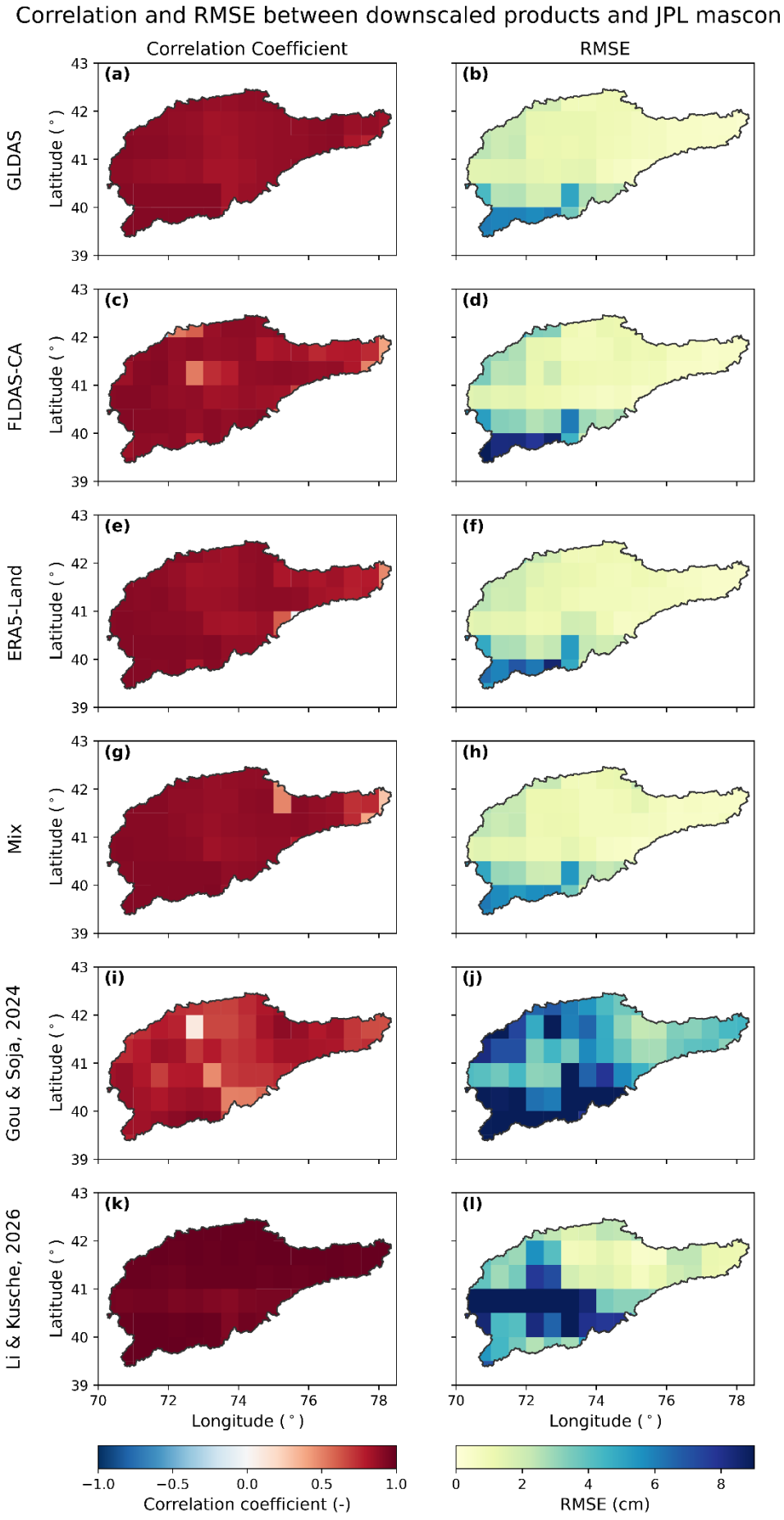

Figure R7: Correlation coefficients and RMSEs between monthly TWSCs from the downscaled products obtained using different hydrological forcing datasets (GLDAS, FLDAS-CA, ERA5-Land, and Mix) and the JPL mascon product. Results from Gou and Soja (2024, https://doi.org/10.1038/s44221-024-00194-w) and Li and Kusche (2026, https://doi.org/10.1029/2025GL119881), two global downscaled total water storage anomaly products are also included.

11. The evaluation based on precipitation events should be presented in a dedicated subsection, as it is currently placed within the irrigated-area section without a clear link. To better address the main objective of the paper, I recommend comparing downscaling results obtained with different precipitation forcings and explicitly showing how precipitation errors propagate into HR TWSC estimates. Including precipitation time series in Figure 12 would greatly strengthen this discussion. Finally, if river discharge observations are available for this period and for these rivers, they would further strengthen the analysis by showing that in situ surface water observations capture this additional amount of water.

Response: Thank you for pointing this out. We have added a dedicated Sect. 5.4 to present the evaluation based on precipitation events. The comparison of downscaled results obtained with different precipitation forcing was previously included in the Supplement (Figs. S1 – S3 and Fig. 11). To improve clarity and completeness, we are considering including these figures in Sect. 5.4. We have also revised Fig. 12 to include precipitation time series from different forcing datasets, enabling a more direct comparison with meteorological station observations. For the in situ river discharge observations, we have checked the Global Runoff Data Centre (GRDC, https://grdc.bafg.de/data/data_portal/) which provides river discharge data on global scale. Unfortunately, no station records are available in Central Asia for the interested period here (2000 to present). We have also checked the dataset of Marti et al. (2023), which compiles 295 gauge locations in mountainous Central Asia and provides norm discharge as well as time series of river discharge for 135 of these locations collected from hydrological yearbooks in Central Asia. However, for the stations located within our interested area, only monthly discharge data are available, which are not suitable for observing short-term water storage changes.

The revised Sect. 5.4 now reads as follows (figure numbering is temporarily referring to the previous version of the manuscript).

"*There is a distinct positive peak observed in both regions at the beginning of 2006, as shown in Fig. 10c and Fig. 10d, corresponding to the day of maximum TWS increase and precipitation. According to ReliefWeb (2006), a strong snowfall event occurred between 26 and 28 January 2006. This event is further corroborated by meteorological records from ground stations (Menne et al., 2012) (Fig. 12), which show daily precipitation at the FERGANA (40.4 °N, 71.8 °E) and ANDIZAN (40.8 °N, 72.3 °E) stations between 1 January and 25 February 2006. Both stations record peaks exceeding 20 mm on 27 January.*

To evaluate the performance of the downscaled results during precipitation events, Figure 12 compares daily precipitation time series from station observations and different hydrological datasets (GLDAS, FLDAS-CA, ERA5-Land, and MSWEP) at two stations. MSWEP shows the best agreement with station observations in capturing both the timing and magnitude of precipitation events. FLDAS-CA captures the peak on 27 January but substantially underestimates its magnitude. ERA5-Land shows a comparable peak magnitude but with a one-day delay. GLDAS fails to capture the event accurately. Figures S1 – S3 and 11 further illustrate the spatial distribution of daily TWSCs and precipitation between 25 and 30 January 2006 for the four forcing scenarios. This event is clearly reflected in Fig. 11 in both the precipitation fields (Part II, panels h – j) and the downscaled daily TWSCs from the Mix product (Part I, panels b – d). On 26 January, the onset of snowfall produced moderate precipitation across the basin, accompanied by positive TWSC values of 5 – 10 mm d-1 in the central and southern regions. The event peaked on 27 January, when heavy snowfall exceeded 25 mm d-1 over the southern basin (panel i), and the TWSC response surpassed 20 mm d-1 (panel c), indicating rapid accumulation of water mass within the snowpack. Residual positive

anomalies persisted on 28 January (panels d and j), reflecting continued storage of snow water equivalent even after the main precipitation had subsided. By 29 – 30 January, both precipitation and storage changes returned to near-background levels. In contrast, GLDAS fails to reproduce the event (Fig. S1), while FLDAS-CA (Fig. S2) captures a TWSC peak on 27 January consistent with the Mix product (Fig. 11), but with a much lower magnitude. ERA5-Land, by contrast, shows a one-day delay, with the TWSC peak occurring on 28 January and being spatially displaced (Fig. S3).

Across all products, precipitation signals are mirrored by TWSCs. The discrepancies in precipitation propagates into the downscaled HR TWSC estimates through the water balance equation, leading to differences in both amplitudes and short-term variability among downscaled products. It highlights the central role of hydrological forcing data quality in determining downscaling performance. This case study demonstrates the ability of the downscaled GRACE product to capture short-term hydrological responses to extreme events. The strong spatial and temporal correspondence between the recorded snowstorm and the positive TWS changes demonstrates that the Mix downscaled product effectively detects daily water mass variations associated with snow accumulation.

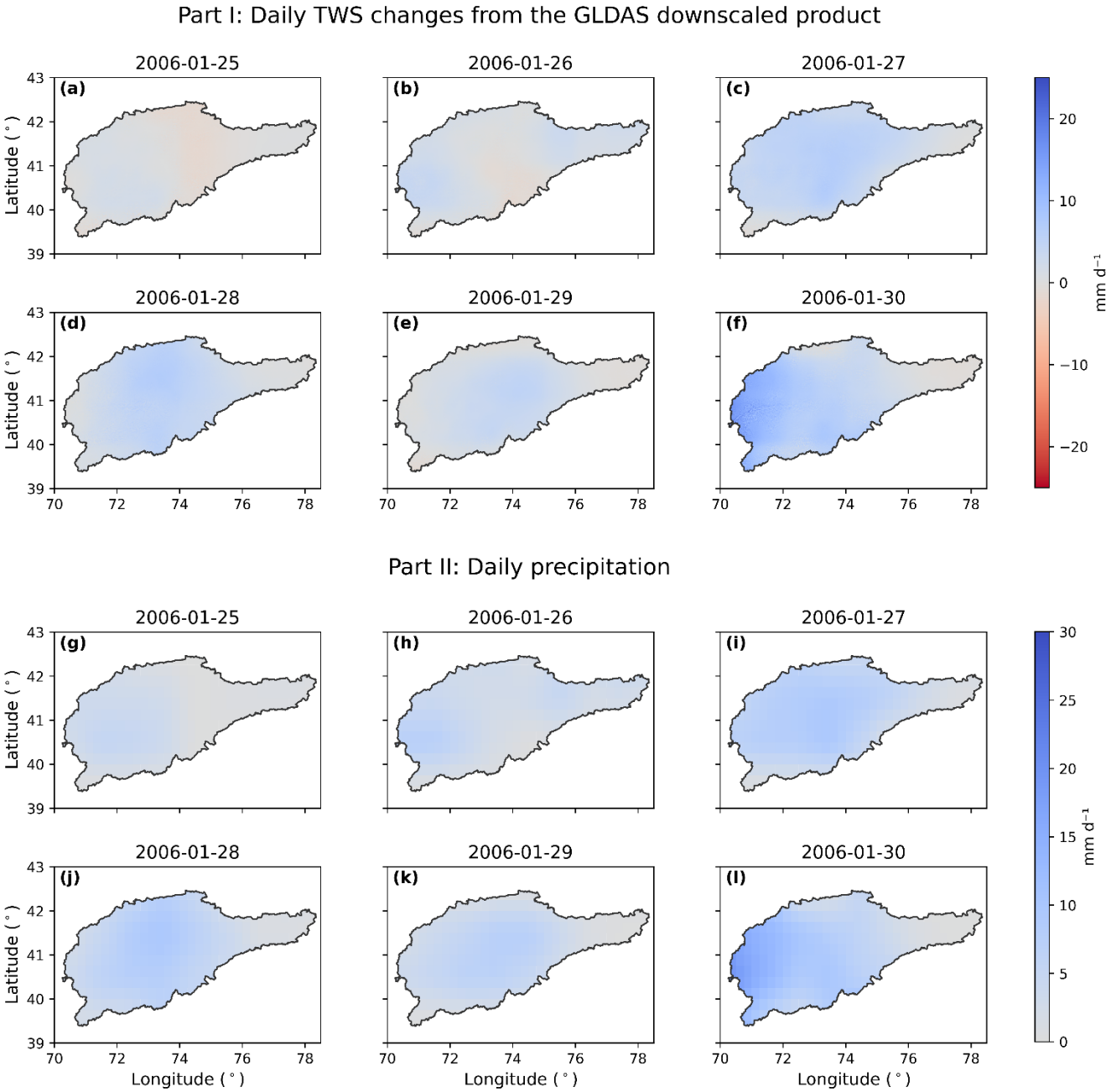

Figure S1: Daily TWS changes (Part I) from the GLDAS downscaled product and corresponding precipitation patterns (Part II) for the period 25 – 30 January 2006.

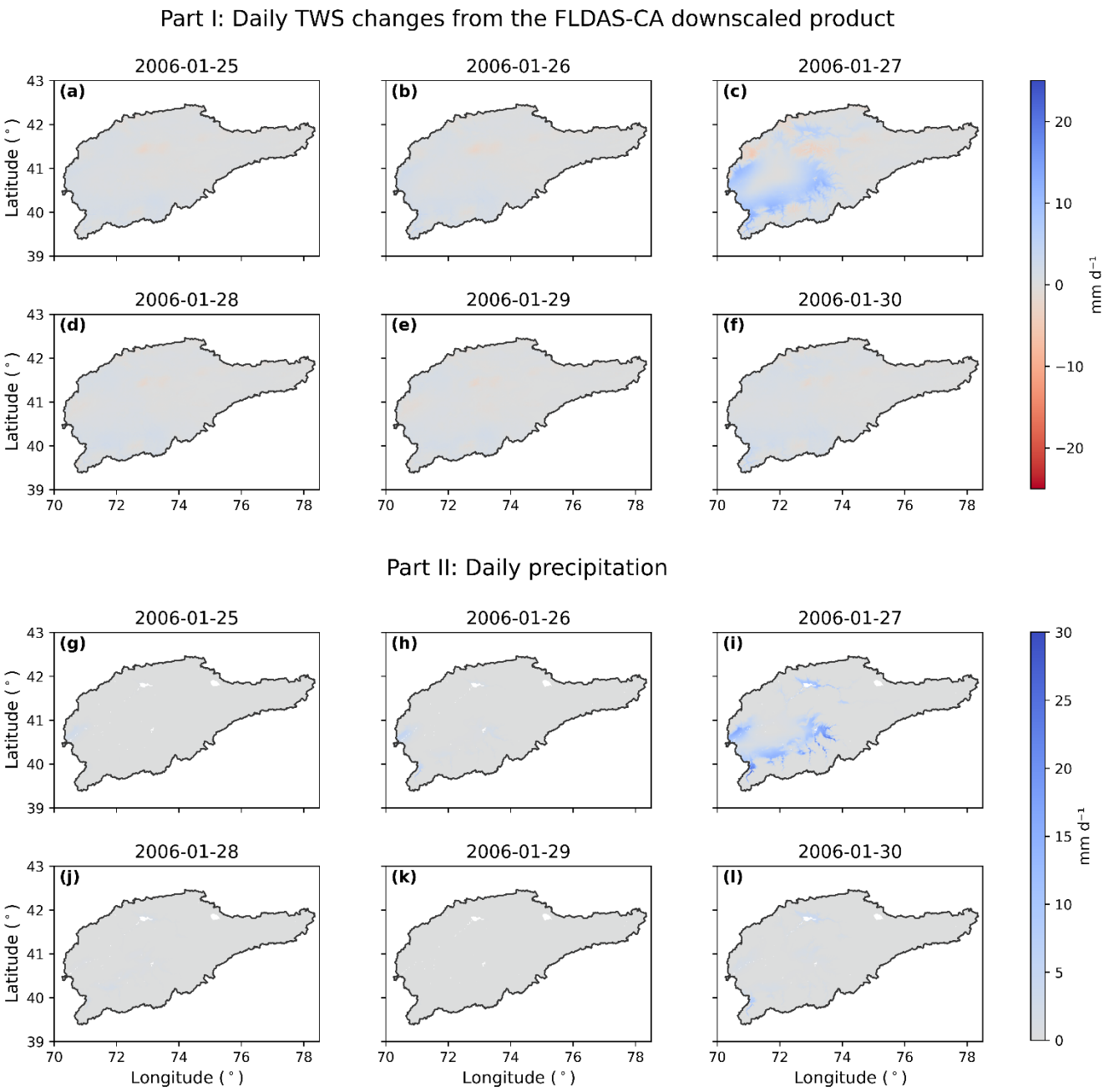

Figure S2: Daily TWS changes (Part I) from the FLDAS-CA downscaled product and corresponding precipitation patterns (Part II) for the period 25 – 30 January 2006.

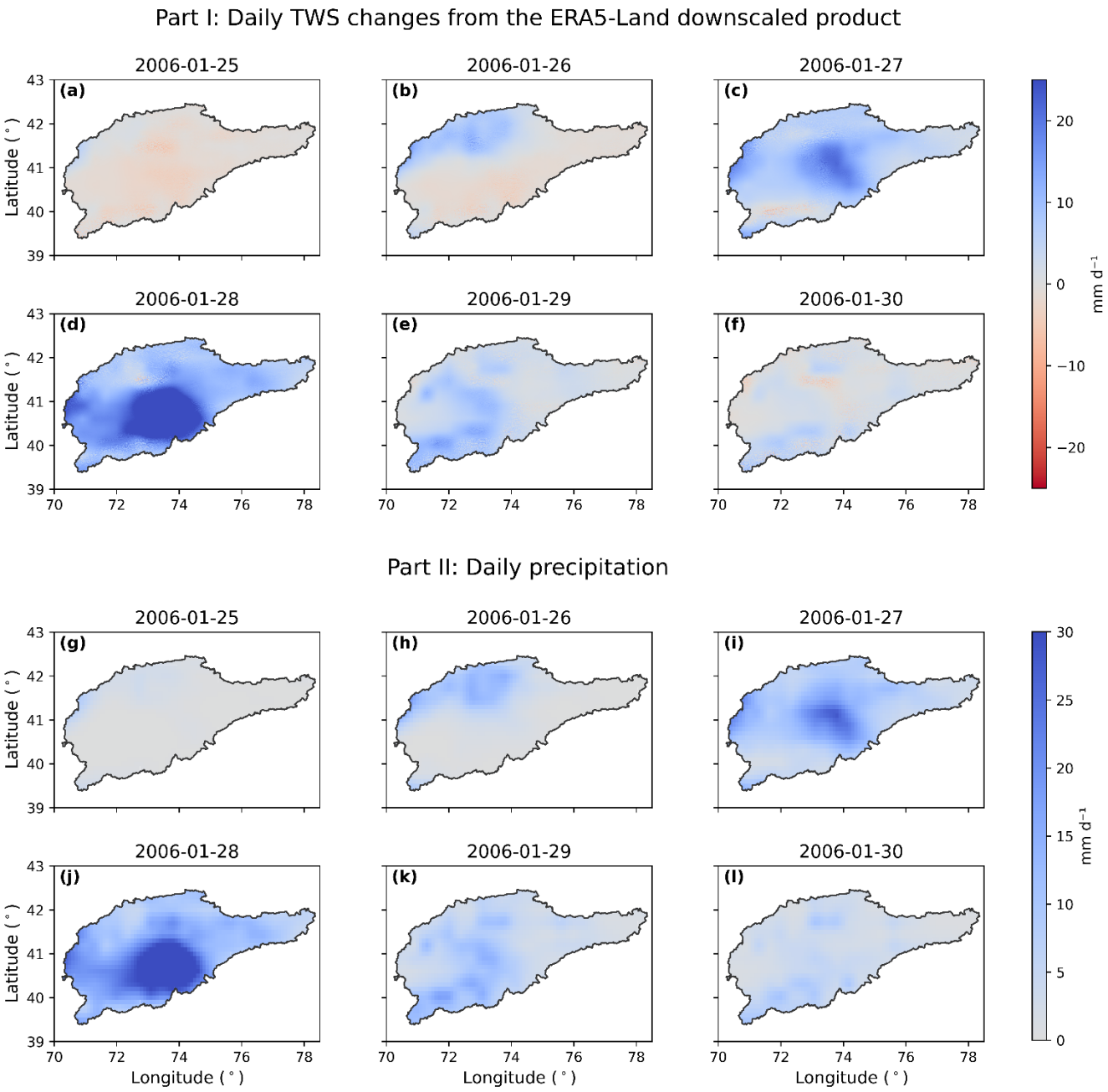

Figure S3: Daily TWS changes (Part I) from the ERA5-Land downscaled product and corresponding precipitation patterns (Part II) for the period 25 – 30 January 2006.

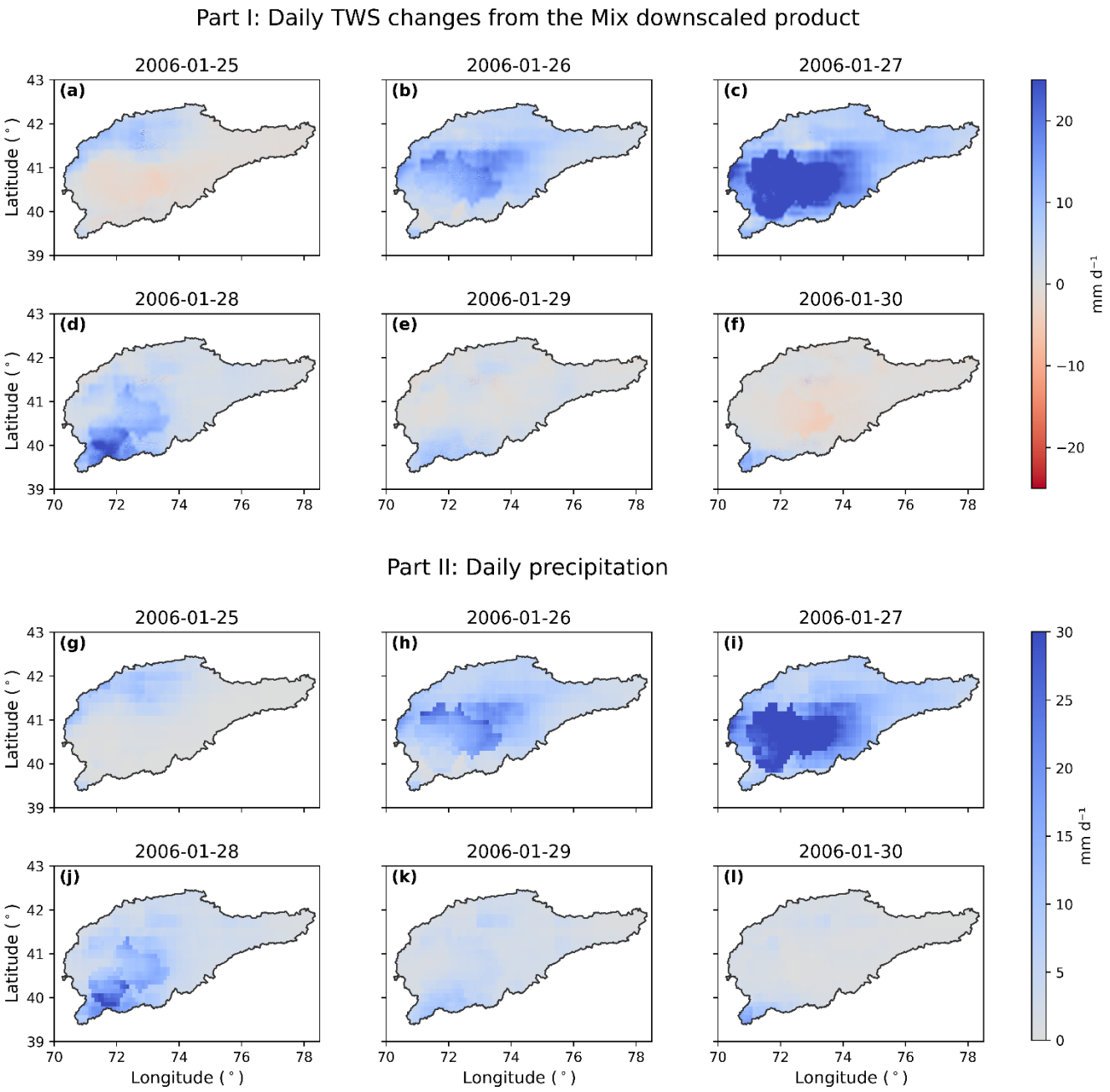

Figure 11: Daily TWS changes (Part I) from the Mix downscaled product and corresponding precipitation patterns (Part II) for the period 25 – 30 January 2006.

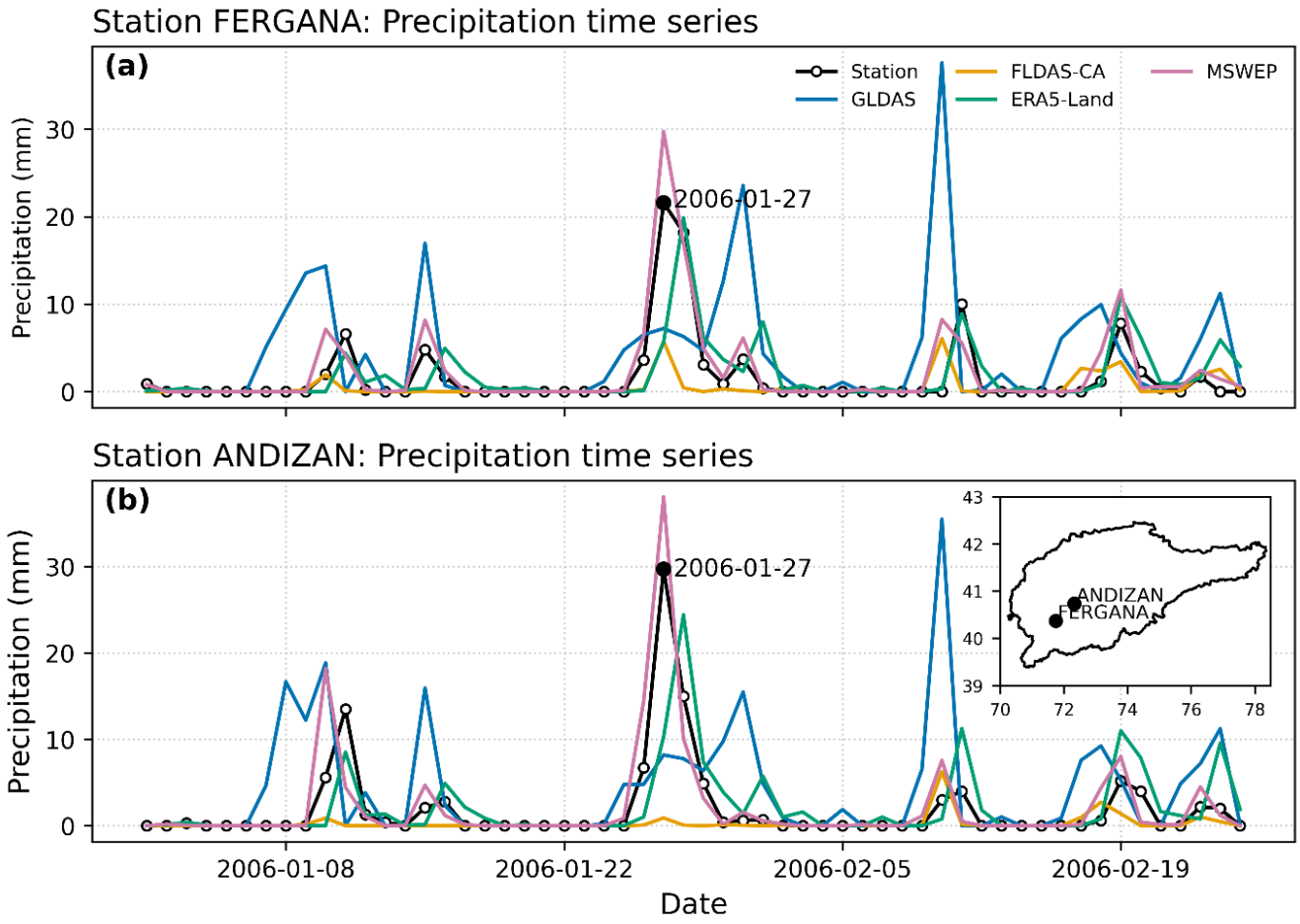

Figure 12: Precipitation records from the meteorological stations and different hydrological forcing (GLDAS, FLDAS-CA, ERA5-land, and Mix) between 1 January and 25 February 2006. The inset in panel (b) indicates their locations within the study region."

Specific comments

1. L41: The native GRACE spatial resolution is approximately 300 km.

Response: We have changed it to "*However, GRACE's coarse spatial (~300 km) and temporal (~monthly) resolutions limit their direct application in regional hydrological studies*".

2. L57: Kalu et al., 2024 (not 2014).

Response: We have corrected this typo to "*Kalu et al. (2024)*".

3. L139: "00:00 record" ?

Response: ERA5-Land hourly data are provided as accumulated values within a day, where the maximum accumulation is over 24 hours, i.e., from day = D, time = 0 to day = D + 1, time = 0 (step = 24) (https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation#ERA5Land:datadocumentation-accumulationsAccumulations). Therefore, daily totals are derived by downloading data at 00:00 for each day.

4. L190: In Eq. (1), reverting to TWS in the central difference is confusing. Since absolute TWS is not observable from GRACE, the last equality can be removed, and TWSA should remain the variable of interest.

Response: Equation (1) has been revised as "

$TWSC|_{month} \approx \frac{TWSA_{m_{i+1}} - TWSA_{m_{i-1}}}{2}$ *(1)*".

5. L210: Eq. (4) appears to compute TWS rather than TWSC. Similar ambiguities between S and dS occur elsewhere and should be systematically corrected.

Response: Equation (4) is used to compute TWS. By combining Eq. (4) and Eq. (1), monthly TWSC based on water balance-derived data can be computed. We have corrected the notation throughout the manuscript. Please refer to Comment 1 under Major comments for details.

6. Figure 3:

The top-right box ("GRACE- vs WB-based") should be placed after the downscaling steps for consistency with the text.

The box should refer to WB-based S (prior to central differentiation).

Use NSE instead of correlation and RMSE, to be consistent with Figure 8.

Use consistent notation (S/dS or TWSA/TWSC) throughout the figure.

Response: The top-right box ("GRACE- vs. WB-based") is placed before the downscaling steps, as it compares monthly TWSC derived from JPL mascon and WB-based data (prior to downscaling) to assess their consistency (Fig. 4).

The box now refers to WB-based TWSA prior to central differentiation.

In the bottom-right box, Nash – Sutcliffe Efficiency (NSE) is used, which is consistent with Fig. 8. The correlation coefficient and RMSE are presented in the box "Comparison with ITSG-Grace2018", which is consistent with Fig. 6. The notation has been updated throughout the figure. The revised Fig. 3 is shown below.

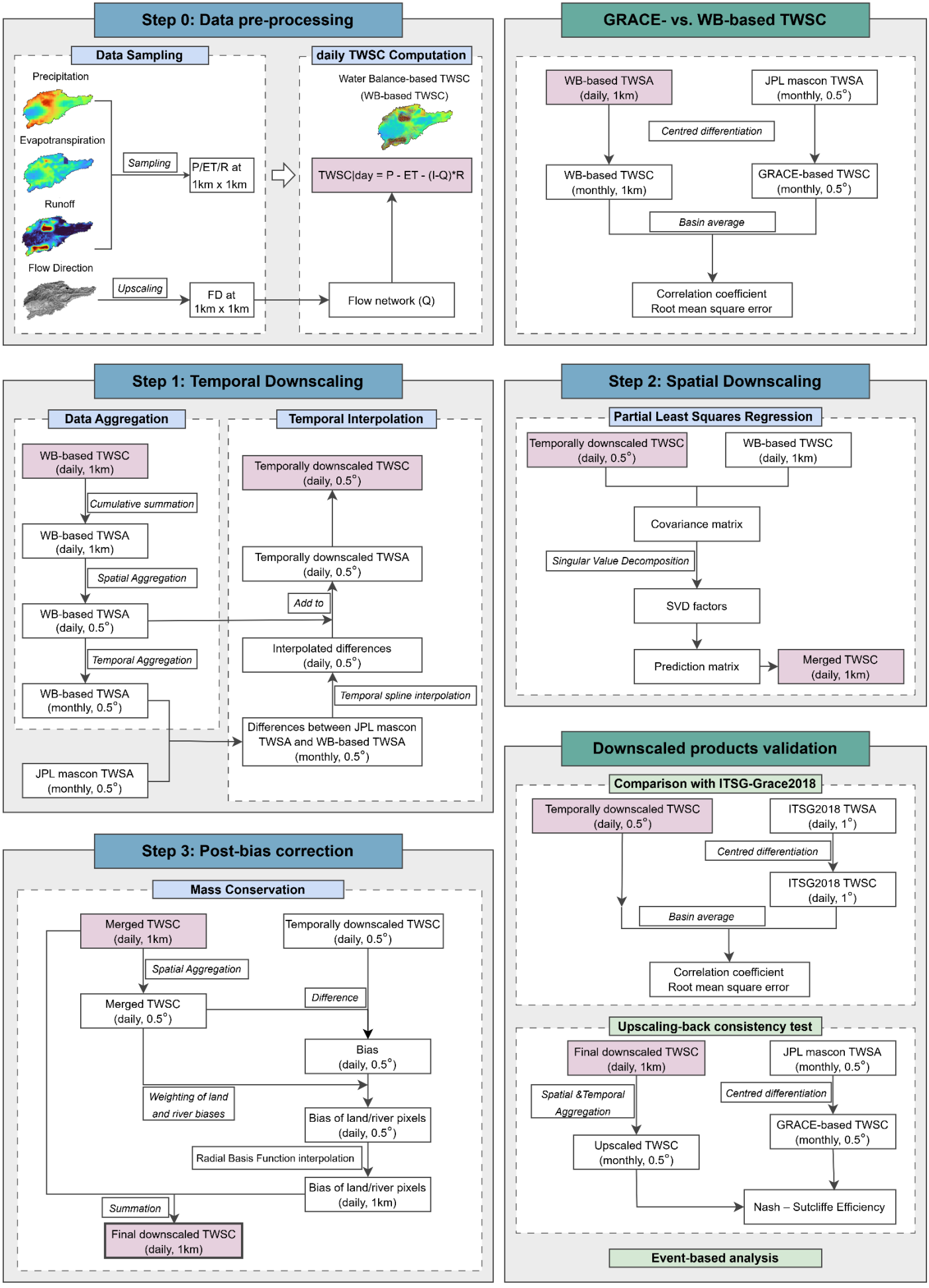

Figure 3: Workflow of GRACE data downscaling and evaluation of results.

7. Figure 4:

Y-axis label should be TWSC or dS, not dTWS/dt (same comment applies to other figures).

Avoid visual linear interpolation in periods with missing data (e.g., 2017).

Response: The y-axis labels in Figs. 4, 5, 6b, 10c, and 10d in the manuscript have been changed from "*dTWS/dt*" to "*TWSC*". The colorbar labels in Figs. 7 and 8 in the manuscript have also been changed accordingly. As noted in the response to Comment 6 under Major comments, months with missing GRACE data are now masked rather than interpolated. We have updated Fig. 4 shown below.

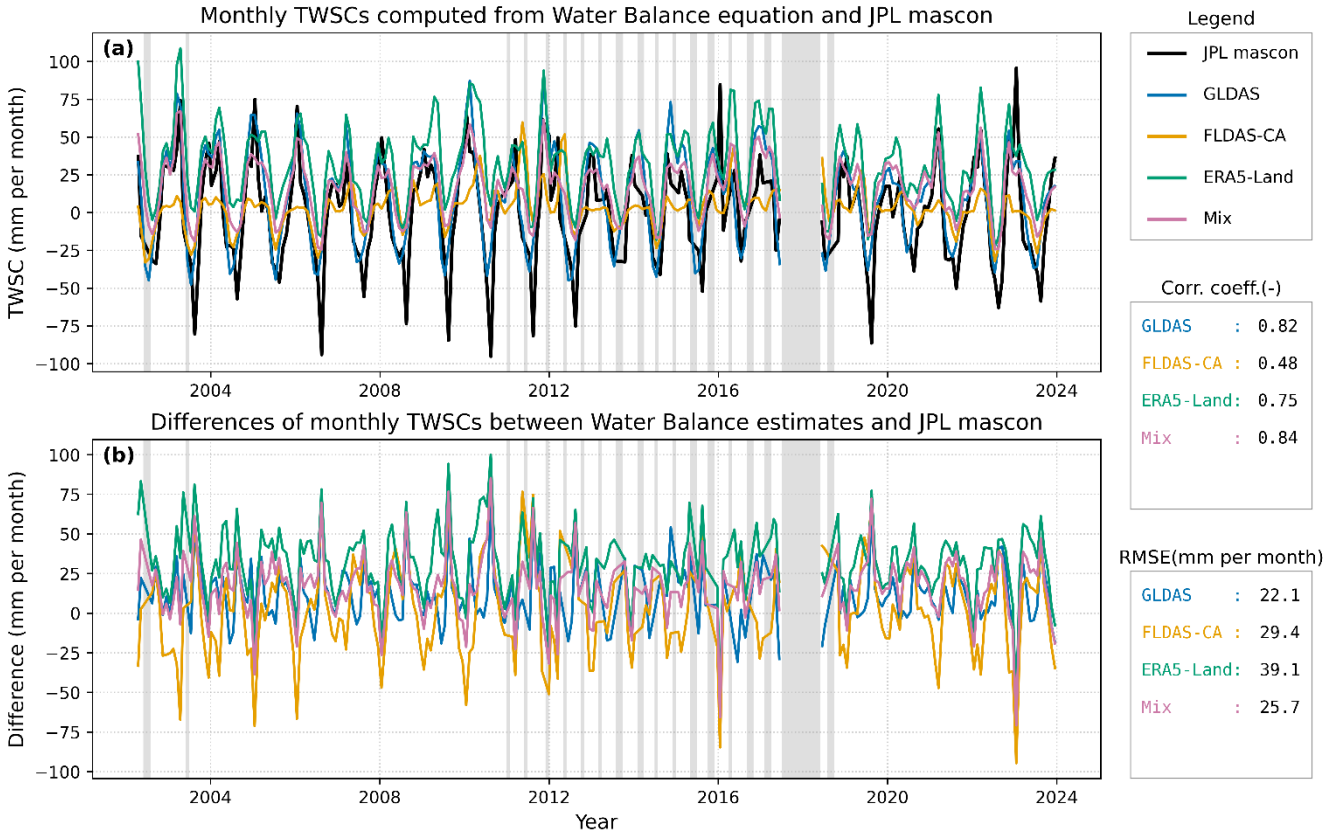

Figure 4: Comparison of monthly TWSCs between April 2002 and December 2023 derived from four hydrological forcing scenarios (GLDAS, FLDAS-CA, ERA5-Land, and Mix) and the JPL mascon solution, averaged over the study region. Panel (a) shows monthly TWSCs computed from hydrological forcing datasets alongside the JPL mascon solution, while panel (b) presents their differences relative to JPL mascon. Grey-shaded areas indicate data gaps in the GRACE record. Correlation coefficients and root mean square errors with respect to the JPL mascon solution are shown on the right.

8. Figure 10b: The title should refer to TWSC, not TWS.

Response: The title has been changed to TWSA, as Fig. 10b shows variations of TWSA computed by accumulating daily TWSCs.

We sincerely thank you again for your valuable and insightful feedback that helps to improve the quality of our manuscript.

References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger, F.: GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013, 2013.

Arshad, A., Mirchi, A., Samimi, M., and Ahmad, B.: Combining downscaled-GRACE data with SWAT to improve the estimation of groundwater storage and depletion variations in the Irrigated Indus Basin (IIB), Sci. Total Environ., 838, 156044, https://doi.org/10.1016/j.scitotenv.2022.156044, 2022.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, Bull. Amer. Meteor. Soc., 100(3), 473-500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Blank, D., Eicker, A., Jensen, L., and Güntner, A.: A global analysis of water storage variations from remotely sensed soil moisture and daily satellite gravimetry, Hydrol. Earth Syst. Sci., 27, 2413–2435, https://doi.org/10.5194/hess-27-2413-2023, 2023.

Brocca, L., Gaona, J., Bavera, D., Fioravanti, G., Puca, S., Ciabatta, L., Filippucci, P., Mosaffa, H., Esposito, G., Roberto, N., Dari, J., Vreugdenhil, M., and Wagner, W.: Exploring the actual spatial resolution of 1 km satellite soil moisture products, Sci. Total Environ., 945, 174087, https://doi.org/10.1016/j.scitotenv.2024.174087, 2024.

Gou, J., and Soja, B.: Global high-resolution total water storage anomalies from self-supervised data assimilation using deep learning algorithms, Nature Water, 2(2), 139-150, https://doi.org/10.1038/s44221-024-00194-w, 2024.

Humphrey, V., Rodell, M., and Eicker, A.: Using satellite-based terrestrial water storage data: a review, Surv Geophys, 44(5), 1489-1517, https://doi.org/10.1007/s10712-022-09754-9, 2023.

Kalu, I., Ndehedehe, C. E., Ferreira, V. G., Janardhanan, S., Currell, M., and Kennard, M. J.: Statistical downscaling of GRACE terrestrial water storage changes based on the Australian Water Outlook model, Sci. Rep., 14(1), 10113, https://doi.org/10.1038/s41598-024-60366-2, 2024.

Kurtenbach, E., Eicker, A., Mayer-Gürr, T., Holschneider, M., Hayn, M., Fuhrmann, M., and Kusche, J.: Improved daily GRACE gravity field solutions using a Kalman smoother, J. Geodyn., 59, 39-48, https://doi.org/10.1016/j.jog.2012.02.006, 2012.

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, J. Geophys. Res. Solid Earth, 124, https://doi.org/10.1029/2019JB017415, 2019.

Li, F., and Kusche, J.: Reproducing GRACE total water storage change at finer spatial scales, Geophys. Res. Lett., 53(1), e2025GL119881, https://doi.org/10.1029/2025GL119881, 2026.

Marti, B., Yakovlev, A., Karger, D.N., Ragettli, S., Zhumabaev, A., Wakil, A.W., and Siegfried, T.: CA-discharge: Geo-Located Discharge Time Series for Mountainous Rivers in Central Asia, Sci Data 10, 579, https://doi.org/10.1038/s41597-023-02474-8, 2023.

McNally, A., Jacob, J., Arsenault, K., Slinski, K., Sarmiento, D. P., Hoell, A., Pervez, S., Rowland, J., Budde, M., Kumar, S., Peters-Lidard, C., and Verdin, J. P.: A Central Asia hydrologic monitoring dataset for food and water security applications in Afghanistan, Earth Syst. Sci. Data, 14, 3115–3135, https://doi.org/10.5194/essd-14-3115-2022, 2022.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Oceanic Technol., 29, 897-910, https://doi.org/10.1175/JTECH-D-11-00103.1, 2012.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Miralles, D. G., Bonte, O., Koppa, A., Baez-Villanueva, O. M., Tronquo, E., Zhong, F., Beck, H. E., Hulsman, P., Dorigo, W. A., Verhoest, N. E. C., Haghdoost, S.: GLEAM4: global land evaporation and soil moisture dataset at 0.1° resolution from 1980 to near present, Sci Data, 12, 416, https://doi.org/10.1038/s41597-025-04610-y, 2025.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and Thépaut, J.-N.: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications, Earth Syst. Sci. Data, 13, 4349–4383, https://doi.org/10.5194/essd-13-4349-2021, 2021.

Pellet, V., Aires, F., Alfieri, L., and Bruno, G.: A physical/statistical data-fusion for the dynamical downscaling of GRACE data at daily and 1 km resolution, J. Hydrol., 628, 130565, https://doi.org/10.1016/j.jhydrol.2023.130565, 2024.

ReliefWeb: Tajikistan: Avalanche – Jan 2006, https://reliefweb.int/disaster/av-2006-000015-tjk, last access: 29 October 2025, 2006.

Rodell, M., Famiglietti, J. S., Chen, J., Seneviratne, S. I., Viterbo, P., Holl, S., and Wilson, C. R.: Basin scale estimates of evapotranspiration using GRACE and other observations, Geophys. Res. Lett., 31(20), https://doi.org/10.1029/2004GL020873, 2004a.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, Bull. Amer. Meteor. Soc., 85(3), 381-394, https://doi.org/10.1175/BAMS-85-3-381, 2004b.

Vishwakarma, B. D., Zhang, J., and Sneeuw, N.: Downscaling GRACE total water storage change using partial least squares regression, Sci data, 8, 95, https://doi.org/10.1038/s41597-021-00862-6, 2021.

Zheng, C., Jia, L., and Zhao, T.: A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution, Scientific Data, 10(1), 139, https://doi.org/10.1038/s41597-023-01991-w, 2023.