



An improved empirical framework for long-term reconstruction of terrestrial water storage and runoff in Pan-Arctic basins

Bingshi Liu^{1,2,3}, Xiancai Zou^{1,2,3,*}, Jiancheng Li^{1,2,3,4}, and Taoyong Jin^{1,2,3}

¹ School of Geodesy and Geomatics, Wuhan University, Wuhan, China

² Hubei Luojia Laboratory, Wuhan University, Wuhan, China

³ Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan University, Wuhan, China

⁴ School of Geosciences and Info-Physics, Central South University, Changsha, China

Corresponding author: Xiancai Zou (xczou@whu.edu.cn).

Abstract: Long-term and continuous observations of terrestrial water storage (TWS) and runoff remain limited in the Pan-Arctic region, constraining our understanding of freshwater variability before the satellite era. Here, we developed two empirical models within a common framework for long-term TWS and runoff reconstruction. Monthly TWS is reconstructed from precipitation, air temperature, and snow information using an improved state-update empirical model, and runoff is subsequently estimated from TWS using a refined runoff–TWS relationship. The TWS model is constrained by observations from the Gravity Recovery and Climate Experiment (GRACE) mission, whereas the runoff model is calibrated against gauge observations. Using this framework, we reconstructed monthly TWS and runoff for 12 major Pan-Arctic basins from 1952 to 2025. The improved TWS model closely reproduces the mean seasonal cycle observed by GRACE and GRACE Follow-On (GRACE-FO), with a median Nash–Sutcliffe efficiency of 0.97. It also captures the major non-seasonal variations in most basins, with a median correlation coefficient of 0.74. Evaluation using GRACE (–FO) TWS shows that the refined runoff–TWS relationship improves monthly runoff reconstruction relative to the previous formulation, increasing the median Kling–Gupta efficiency (KGE) from 0.82 to 0.90 and reducing systematic seasonal biases. When applied to reconstructed TWS, the relationship maintains similar performance beyond the parameter-estimation periods, with median monthly KGE values of 0.92 during the 8-year parameter-estimation periods and 0.93 over the remaining available observations. Reconstructed annual runoff also captures the major interannual variations in observed runoff, with a median annual KGE of 0.65 across the 12 stations. This agreement with long-term gauge observations provides a complementary indirect assessment of the historical TWS reconstruction. Comparisons with existing GRACE-like TWS products and a gridded runoff reconstruction further support the long-term consistency of the framework, while sensitivity experiments show limited dependence on the parameter-estimation period. The proposed approach provides a simple means of extending Pan-Arctic TWS records into the pre-GRACE era and producing continuous runoff reconstructions over the full study period, thereby supporting investigations of long-term Pan-Arctic freshwater variability.

1 Introduction

The Arctic is warming at nearly three times the global mean rate (Zhou et al., 2024) and is experiencing sustained increases in precipitation (Bintanja and Andry, 2017; Box et al., 2019). These changes are reshaping the Arctic freshwater cycle, as reflected by increasing river discharge (Feng et al., 2021; Gudmundsson et al., 2025; Makarieva et al., 2019; Peterson et al., 2002), declining snow cover duration (Callaghan et al., 2011; Derksen and Brown, 2012; Qing et al., 2025), permafrost degradation (Biskaborn et al., 2019; Hamm et al., 2023; Nitzbon et al., 2024; Smith et al., 2022), and widespread loss of lowland lakes (Chen et al., 2023;



35 Jones et al., 2022; Webb et al., 2022; Webb and Liljedahl, 2023). These hydrological changes can further affect Arctic ecosystems and biogeochemical processes (Bring et al., 2016; Chin et al., 2016; Miner et al., 2022).

Terrestrial water storage (TWS) and runoff are key indicators of freshwater availability and hydrological change. However, long-term and continuous observations remain limited in the Pan-Arctic region. In situ hydrological observations are sparse and unevenly distributed, and many records are short or discontinuous because of environmental, economic, and political constraints
40 (Feng et al., 2021). Hydrological models can provide spatially and temporally continuous simulations, but their performance depends strongly on climate forcing, model structure, and parameterization (Lin et al., 2019). Models still differ substantially (Mudryk et al., 2025; Yi et al., 2023), and some may overestimate historical snow storage and increases in runoff (Chai et al., 2025; Zhang et al., 2026).

Satellite observations have substantially expanded hydrological monitoring capabilities (Andreadis et al., 2025; Landerer et al.,
45 2010; Leopardi et al., 2026; Webb et al., 2022). In particular, the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) missions have provided large-scale observations of TWS variability since 2002 (Rodell et al., 2018; Rodell and Reager, 2023; Tapley et al., 2019). Nevertheless, satellite records remain too short to fully characterize multidecadal variability and long-term hydrological change.

Recent studies have therefore used empirical models and machine learning methods to extend historical TWS records (Gentner et al., 2026; Hacker et al., 2026; Humphrey et al., 2017; Humphrey and Gudmundsson, 2019; Li et al., 2020, 2021; Liu et al., 2021, 2022; Xie and Yi, 2026; Yin et al., 2023) and reconstruct runoff (Ghiggi et al., 2019; Gudmundsson and Seneviratne, 2016). However, important limitations remain. Machine learning methods can capture complex relationships between hydrological variables and climate forcing, but their physical interpretability is generally limited and their performance may vary considerably among regions (Sun et al., 2020). Empirical models generally have simpler structures, but existing formulations are often developed
55 for specific climatic or hydrological conditions. Some models perform poorly in high-latitude basins (Liu et al., 2021) where snow accumulation and melt strongly regulate seasonal water storage, whereas others do not adequately reproduce the seasonal variability of TWS (Humphrey and Gudmundsson, 2019; Xie and Yi, 2026). More generally, a unified basin-scale framework for long-term reconstruction of both TWS and runoff that can accommodate high-latitude snow processes remains limited.

In this study, we developed two empirical models within a common basin-scale framework for reconstructing monthly TWS and
60 runoff in 12 major Pan-Arctic basins from 1952 to 2025. Monthly TWS is reconstructed from precipitation, air temperature, and snow information using an improved state-update empirical model, whereas runoff is estimated from TWS using a refined runoff–TWS relationship. The TWS model is constrained by GRACE TWS observations, and the runoff relationship is calibrated against gauge observations. We first evaluated the TWS reconstruction against GRACE (-FO) observations and assessed the refined runoff–TWS relationship using GRACE (-FO) TWS as input. The refined relationship was then applied to reconstructed TWS to
65 evaluate long-term runoff and provide a complementary indirect assessment of the historical TWS reconstruction. We further examined model sensitivity to the parameter-estimation period the effects of model simplification, comparisons with existing reconstruction products, and the influence of human regulation on runoff evaluation. The proposed framework provides an alternative means of extending Pan-Arctic TWS into the pre-GRACE period and producing continuous long-term runoff reconstructions, thereby supporting investigations of long-term freshwater variability.



70 2 Data

2.1 Reference Data

Monthly gauge observations were used as reference data for the runoff reconstruction. To obtain long and reliable in situ records, we compiled observations from three archives, including the Arctic Great Rivers Observatory (ArcticGRO), the Global Runoff Data Centre (GRDC), and R-ArcticNet. We selected 12 gauging stations located near the outlets of major Pan-Arctic rivers with relatively complete observational records. TWS reconstruction and evaluation were conducted over the corresponding catchments (Figure 1).

Three GRACE (-FO) mascon products from the Center for Space Research (CSR; Save et al., 2016), the Goddard Space Flight Center (GSFC; Loomis et al., 2019; Luthcke et al., 2013), and the Jet Propulsion Laboratory (JPL; Watkins et al., 2015; Wiese et al., 2016) were used as reference data for TWS reconstruction. The CSR product is provided on a 0.25° grid, whereas the GSFC and JPL products are provided on 0.5° grids. For consistency, the CSR product was resampled to 0.5° before calculating catchment-scale TWS anomaly time series.

Glacier mass changes are not explicitly represented in our TWS reconstruction model. To reduce their influence, 0.5° grid cells containing glaciers were excluded from the calculation of basin-scale TWS anomalies based on glacier distributions from the Randolph Glacier Inventory 7.0 (RGI Consortium, 2023). However, because of the effective spatial resolution of GRACE (-FO) and associated signal leakage, spatial masking cannot fully remove contributions from nearby glaciers. This issue is particularly relevant to the JPL product (Figure S1), which is based on mascon elements with a characteristic scale of approximately 3° , although the product is distributed on a 0.5° grid. We therefore further removed the linear trend from the GRACE (-FO) TWS time series before TWS-model parameter estimation to reduce the influence of residual glacier mass changes on the reference TWS. This treatment may also reduce the influence of long-term TWS changes associated with permafrost degradation (Yi et al., 2023), which is also not explicitly represented in our model.

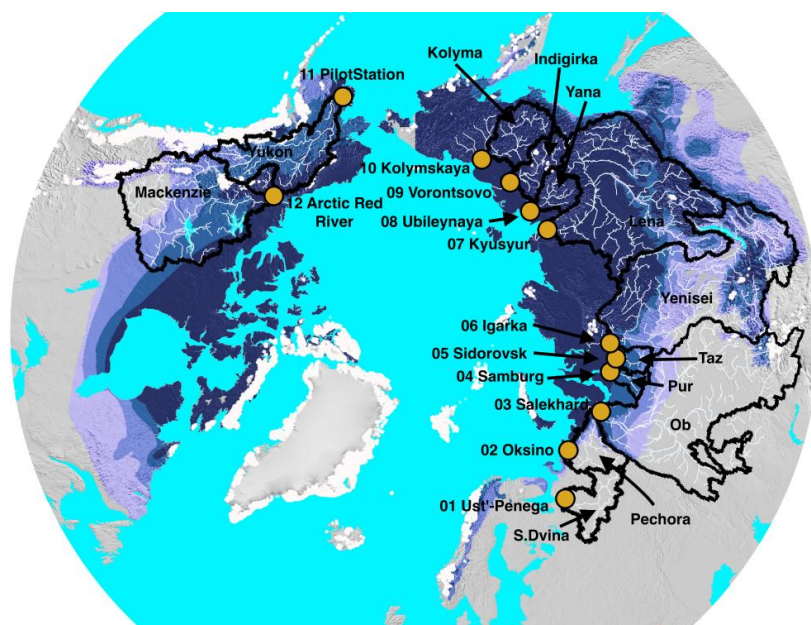


Figure 1. Spatial distribution of the 12 river gauging stations and their corresponding catchments used in this study. Yellow circles indicate gauging stations, black outlines delineate the corresponding catchments, and blue shading indicates permafrost distribution. White symbols indicate glacier locations identified from the Randolph Glacier Inventory 7.0 (RGI Consortium, 2023). They are used only



95 to show the spatial distribution of glaciers and are not proportional to the actual glacier area. The 0.5° grid cells containing glaciers were excluded from the catchment-scale TWS calculation.

2.2 Forcing and auxiliary data

Monthly runoff reconstruction uses TWS as the primary input, together with information on snow storage and snow cover fraction. Snow storage is estimated from snow water equivalent (SWE) products. Although TWS can be obtained from GRACE (-FO) observations, their temporal coverage is insufficient for long-term historical runoff reconstruction, particularly for extending the record into the twentieth century. Reliable historical TWS reconstruction is therefore required before reconstructing long-term runoff.

In this study, TWS reconstruction is primarily driven by precipitation and air temperature, with snow information used as auxiliary data. We selected five combinations of precipitation and air temperature datasets (Table 1), five SWE products (Table 2), and two snow cover fraction products (Table 3) for the TWS and runoff reconstructions. The TWS model also requires separating total precipitation into liquid and solid components. JRA-3Q and ERA5 provide explicit rainfall and snowfall fields. For the remaining datasets, precipitation was partitioned using a gridded 50% rain–snow air temperature threshold for the Northern Hemisphere following Jennings et al. (2018). Further details on the forcing and auxiliary datasets are provided in Supplementary Information Text S1.

110 **Table 1. Gridded precipitation and temperature combinations used in this study. P represents precipitation, and T represents air temperature.**

Combination	Production	Version	Type	Spatial resolution	Resample	Time resolution	Time coverage	Reference
CRU	CRU	ts 4.09	P & T	0.50°	/	Monthly	1901–2024	Harris et al. (2020)
GPCC	GPCC	v2025	P	0.50°	/	Monthly	1891–2024	Schirmeister et al. (2025)
	BEST	/	T	1.00°	0.50°	Monthly	1850–2024	Rohde & Hausfather, (2020)
U. Delaware	U. Delaware	5.01	P & T	0.50°	/	Monthly	1900–2017	Willmott & Matsuura, (1995)
JRA -3Q	JRA-3Q		P & T	0.25°	0.50°	Monthly	1947–2025	Kosaka et al. (2024)
ERA5 single	ERA5 single	/	P & T	0.25°	0.50°	Monthly	1940–2025	Hersbach et al. (2020)

Table 2. Gridded Snow Water Equivalent (SWE) datasets used in this study.

Name	Version	Spatial resolution	Resample	Time resolution	Time coverage	Reference
ERA5-land		0.10°	0.50°	Monthly	1950–2025	Muñoz-Sabater et al. (2021)
Crocus-ERA5		0.25°	0.50°	Monthly	1950–2023	Ramos Buarque et al. (2025)
WaterGAP 20crv3-w5e5	2.2e	0.50°	/	Monthly	1901–2019	Müller Schmied et al. (2024)
GLDAS NOAH	020	0.25°	0.50°	Monthly	1948–2014	Rodell et al. (2004)
JRA -3Q		0.25°	0.50°	Monthly	1947–2025	Kosaka et al. (2024)

Table 3. Gridded snow cover datasets used in this study.

Name	Spatial resolution	Resample	Time resolution	Time coverage	Reference
ERA5-land	0.10°	0.50°	Monthly	1950–2025	Muñoz-Sabater et al. (2021)



Crocus-ERA5	0.25°	0.50°	Monthly	1950–2023	Ramos Buarque et al. (2025)
-------------	-------	-------	---------	-----------	-----------------------------

3 Method

3.1 Improved state-update empirical model for TWS reconstruction

3.1.1 Previous model and its limitations

Liu et al. (2021) proposed a state-update empirical model that updates monthly TWS using antecedent TWS, precipitation, and air temperature. The model is expressed as:

$$TWS_{t_{i+1}} + d_w = (TWS_{t_i} + d_w) \cdot e^{\tau_{t_{i+1}}} + P_{t_{i+1}} \quad (1)$$

$$\tau_{t_i} = a + b \cdot T_{t_i}^{\text{melt}'} + c \cdot P_{t_i}' \quad (2)$$

where TWS_{t_i} represents the TWS anomaly at time t_i . T^{melt} and P represent air temperature above 0 °C and precipitation, respectively; $'$ represents the normalized values of these variabilities. The residual function τ , which depends on precipitation and air temperature and is constrained to remain below zero, controls the TWS decay factor e^{τ} . d_w is a storage-offset parameter because GRACE (-FO) provide TWS anomalies rather than absolute water storage.

The model is based on three main assumptions. First, precipitation represents the primary water input to the basin. Second, antecedent TWS controls the capacity of the basin to store additional water. Third, precipitation and air temperature jointly regulate the release of stored water. This formulation describes the relationship between climate forcing and TWS variability using relatively few parameters.

However, two limitations are particularly important for Pan-Arctic applications. First, the model does not distinguish rainfall from snowfall and therefore does not explicitly represent snowfall accumulation in solid TWS. Second, it does not explicitly represent meltwater generation. Although temperature is included, the temperature-dependent decay term acts on total TWS, whereas meltwater generation depends on both temperature and the available solid TWS.

3.1.2 Improved model formulation

To address these limitations, we modified the previous model in three main aspects.

First, we restructured the model to explicitly separate solid (TWS^{solid}) and liquid TWS (TWS^{liquid}). Because the previous framework is better suited to representing liquid TWS dynamics, solid TWS is subtracted from total TWS, and the state variable is reformulated as liquid TWS. Accordingly, liquid precipitation (P^{liquid}) and meltwater (MW) are retained as the main water inputs to liquid TWS. We further introduce an empirical scaling parameter α_w , with $0 < \alpha_w < 1$, to represent the effective contribution of solid precipitation (P^{solid}) to changes in liquid TWS. This parameter may also partly account for uncertainty associated with the partitioning of precipitation into liquid and solid phases. The resulting formulation is given in Eq. (3). Solid TWS may include snow, ground ice, and glacier ice. In this study, after excluding grid cells containing glaciers, we approximate solid TWS by snow storage. Snow storage is more readily available than other solid-water components at large spatial scales and over long continuous periods. Moreover, seasonal snow represents the dominant active solid-water component in non-glaciated northern high-latitude basins and contributes strongly to seasonal TWS variability (Trautmann et al., 2018). It should be noted that meltwater MW in Eq. (5) is an approximation because a month-to-month decrease in solid TWS may reflect sublimation losses in addition to melting. Therefore, MW should be interpreted as a meltwater proxy rather than a direct estimate of actual meltwater production.



$$TWS_{t_{i+1}}^{\text{liquid}} + d_w = (TWS_{t_i}^{\text{liquid}} + d_w) \cdot e^{\tau_{t_{i+1}}^*} + P_{t_{i+1}}^{\text{liquid}} + \alpha_w P_{t_{i+1}}^{\text{solid}} + MW_{t_{i+1}} \quad (3)$$

$$TWS_{t_i}^{\text{liquid}} = TWS_{t_i} - TWS_{t_i}^{\text{solid}} \quad (4)$$

$$MW_{t_{i+1}} \approx \max(TWS_{t_i}^{\text{solid}} - TWS_{t_{i+1}}^{\text{solid}}, 0) \quad (5)$$

150

Second, we reformulated the residual function τ^* . To reduce potential multicollinearity between precipitation and air temperature, the precipitation term was removed from the residual function, while temperature was retained. Temperature is further separated into above-zero (T^{melt}) and below-zero (T^{freeze}) components to represent liquid TWS losses under different thermal conditions. The revised function also includes a lag coefficient ω , with $0 < \omega < 1$, to account for potential delayed hydrological responses, which may arise from processes such as slow drainage in low-lying areas and gradual active-layer thaw. In addition, normalized meltwater (MW'), referred to here as meltwater intensity, is included to represent the influence of meltwater input on the residual function.

155

$$\tau_{t_{i+1}}^* = a + b_1 \cdot (T_{t_{i+1}}^{\text{freeze}} - \min(T_{t_{i+1}}^{\text{freeze}}))' + b_2 \cdot \omega \cdot T_{t_i}^{\text{melt}'} + b_2 \cdot (1 - \omega) \cdot T_{t_{i+1}}^{\text{melt}'} + c \cdot MW'_{t_{i+1}} \quad (6)$$

$$T^{\text{melt}} = \max(T, 0), \quad T^{\text{freeze}} = \min(T, 0) \quad (7)$$

160

Finally, we revised the normalization scheme by adopting the Tanh estimator to reduce the influence of temperature outliers on parameter estimation. Here, $\text{mean}(T)$ and $\text{std}(T)$ are calculated over the full forcing period and are used only for normalization of the temperature forcing.

$$T'_{t_i} = \frac{1}{2} \cdot \left(\tanh\left(\frac{T_{t_i} - \text{mean}(T)}{\text{std}(T)}\right) + 1 \right) \quad (8)$$

165

The improved model contains seven parameters, ($a, b_1, b_2, c, d_w, \alpha_w$, and ω), whose estimation is described in Section 3.3. When solid precipitation and solid TWS are negligible, the additional solid-water terms in Eq. (3) vanish, and the model retains the same basic state-update structure as the previous formulation. The new model therefore preserves the core structure of the previous model while introducing additional terms to represent snow-related processes in high-latitude basins.

3.2 Refined runoff-TWS relationship

3.2.1 Previous runoff-TWS relationship

170

Monthly runoff (R) can be estimated from liquid TWS using an empirical relationship previously validated for Pan-Arctic basins (Yi et al., 2023).

$$R_{t_{i+1}} = \alpha_1 \cdot (1 - \beta \cdot SC_{t_i}) \cdot (TWS_{t_i}^{\text{liquid}} + d_r) + \alpha_2 \cdot (1 - \beta \cdot SC_{t_{i+1}}) \cdot (TWS_{t_{i+1}}^{\text{liquid}} + d_r) \quad (9)$$

175

Where α_1 and α_2 conversion coefficients linking previous- and current-month liquid TWS, respectively, to runoff. SC denotes snow cover fraction, and β is a dimensionless scaling coefficient ranging from 0 to 1. The term $(1 - \beta \cdot SC)$ modulates the contribution of liquid TWS to runoff under different snow-cover conditions. d_r is a storage-offset parameter analogous to that d_w in Eqs. (1) and (3), accounting for the fact that liquid TWS is expressed as an anomaly rather than absolute storage.

The relationship contains four parameters to be estimated, namely α_1 , α_2 , β , and d_r . Their estimation is described in Section 3.3.



3.2.2 Refined relationship

In Pan-Arctic basins, the runoff response to liquid TWS may differ between snowmelt and non-snowmelt periods. However, Yi et al. (2023) used two time-invariant conversion coefficients, α_1 and α_2 , so that the estimated coefficients represent an average response over different hydrological conditions. Such averaging may smooth seasonal differences in the liquid TWS–runoff relationship.

To account for this difference, we refined the relationship by introducing two meltwater-dependent perturbation coefficients, μ_1 and μ_2 , into α_1 and α_2 , respectively. This allows the runoff response to liquid TWS to vary with meltwater conditions.

$$R_{t_{i+1}} = (\alpha_1 + \mu_1 \cdot MW'_{t_i}) \cdot (1 - \beta \cdot SC_{t_i}) \cdot (TWS_{t_i}^{\text{liquid}} + d_r) + (\alpha_2 + \mu_2 \cdot MW'_{t_{i+1}}) \cdot (1 - \beta \cdot SC_{t_{i+1}}) \cdot (TWS_{t_{i+1}}^{\text{liquid}} + d_r) \quad (10)$$

Although the snow-cover modulation term $(1 - \beta \cdot SC)$ and the meltwater-dependent adjustment $\mu \cdot MW'$ may be active during partly overlapping periods, they represent different hydrological effects. The former represents changes in runoff release associated with snow-cover conditions, whereas the latter adjusts the sensitivity of runoff to liquid TWS according to meltwater intensity.

Yi et al. (2023) detrended GRACE TWS before runoff estimation to reduce long-term non-hydrological or unmodeled signals, including residual glacial isostatic adjustment (GIA) effects and leakage from glacier mass loss. These non-target signals are expected to have a smaller influence on the reconstructed TWS used in this study because the reconstruction is driven by climate forcing and snow information. We therefore did not apply additional detrending before runoff reconstruction, thereby preserving the long-term variability generated by the climate and snow forcing.

3.3 Ensemble reconstruction, parameter estimation, and uncertainty

Both TWS and runoff were reconstructed using an ensemble method. For TWS reconstruction, three GRACE mascon products were used as reference data in combination with five precipitation–air temperature dataset combinations and five SWE products. Each data combination was calibrated independently, yielding $3 \times 5 \times 5 = 75$ reconstruction members for each catchment. Their ensemble mean was used as the final TWS estimate. For runoff reconstruction, the ensemble-mean reconstructed TWS was combined with five SWE products and two snow cover fraction products. Each combination was independently calibrated, yielding $5 \times 2 = 10$ runoff reconstruction members for each gauging station, whose ensemble mean was used as the final runoff estimate.

Model parameters were estimated using the Markov chain Monte Carlo (MCMC) algorithm (Haario et al., 2006), with parameter bounds defined based on model characteristics and physical constraints. Initial parameter values were obtained using a constrained least-squares optimization within the prescribed bounds by minimizing the unweighted sum of squared residuals between monthly GRACE TWS observations and reconstructed TWS over the parameter-estimation period. Starting from these initial values, the MCMC analysis was performed until 20,000 accepted parameter samples were obtained. For each reconstruction member, the final parameter estimates were taken as the mean of the 20,000 accepted samples.

The TWS reconstruction model contains seven parameters, $a, b_1, b_2, c, d_w, \alpha_w$, and ω . The ranges of a, b_1, b_2 , and c were set to $[-5, 0]$, while the range of d_w was set to $[|\min(TWS_{\text{GRACE}}^{\text{liquid}})|, |\min(TWS_{\text{GRACE}}^{\text{liquid}})| + 10,000] \text{ km}^3$. Based on their physical definitions, both α_w and ω were constrained to $[0, 1]$. GRACE TWS observations from 2003 to 2010 were used as the reference period for parameter estimation. The distributions of parameter estimates obtained from the 75 ensemble members are shown in Figure 2.

Estimated parameters vary across catchments and input-data combinations. Parameter b_1 is close to zero in most catchments, whereas b_2 and c generally show larger absolute values. This suggests that below-zero temperature variations generally exert a weaker influence on monthly liquid TWS changes than above-zero temperature variations and meltwater processes. The distribution of α_w is relatively concentrated in some catchments but spans much of its allowable range in others, suggesting



sensitivity to the choice of input datasets or relatively weak parameter identifiability. In contrast, ω is mainly distributed between approximately 0.25 and 0.55 in most catchments, suggesting a generally delayed response of liquid TWS to temperature variations during melt periods. The necessity of these parameters and their effects on model performance are further discussed in Section 5.1.

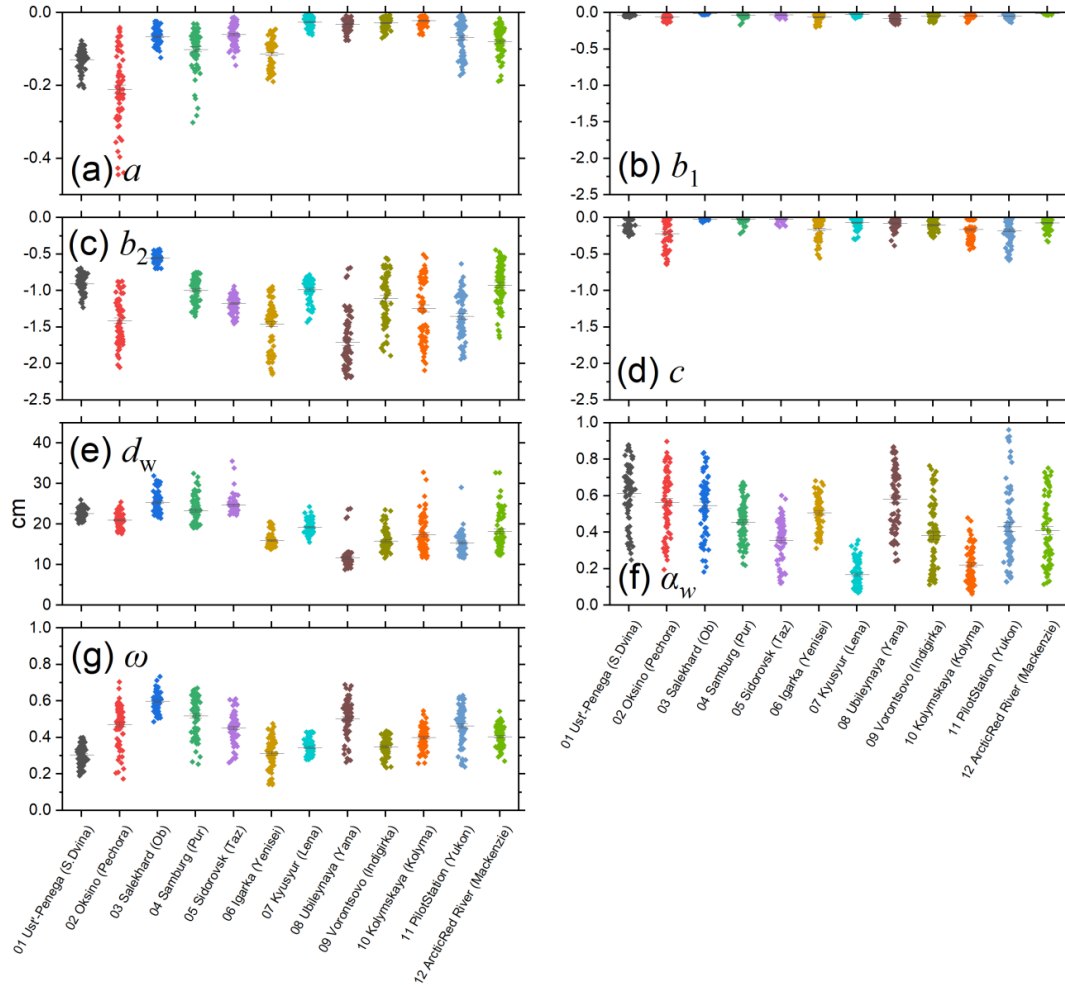


Figure 2. Distributions of parameter estimates for the TWS reconstruction model obtained from the 75 ensemble members in each catchment. GRACE TWS observations from 2003 to 2010 were used for parameter estimation. For each ensemble member, the parameter estimate shown is the mean of the 20,000 accepted MCMC samples. Parameter d_w is expressed as equivalent water height (unit: cm) to facilitate comparison among catchments.

The refined runoff–TWS relationship contains six parameters, $\alpha_1, \beta, d_r, \alpha_2, \mu_1$ and μ_2 . The ranges of $\alpha_1, \alpha_2, \mu_1$, and μ_2 were set to $[0, 4]$, while d_r was constrained to $[\min(TWS^{\text{liquid}}), |\min(TWS^{\text{liquid}})| + 10,000]$ km³. Based on its physical definition, β was constrained to $[0, 1]$. When the TWS reconstructed in this study was used for runoff reconstruction, an 8-year period of gauge-derived runoff was used for parameter estimation at each gauging station, with the specific periods shown in Figure 7a. The corresponding parameter distributions obtained from the 10 runoff ensemble members are shown in Figure 3.

Estimated β values are relatively high at most stations, generally ranging from 0.7 to 1.0, suggesting a strong modulation of catchment drainage connectivity by snow cover. The current-month conversion coefficient α_2 is generally larger than the previous-month coefficient α_1 , indicating a stronger runoff response to current-month liquid TWS. A similar difference is found for the



meltwater-related perturbation coefficients, μ_1 and μ_2 . At several stations, the coefficient μ_1 approaches zero, suggesting that the enhanced conversion of liquid TWS to runoff associated with meltwater occurs mainly within the current month.

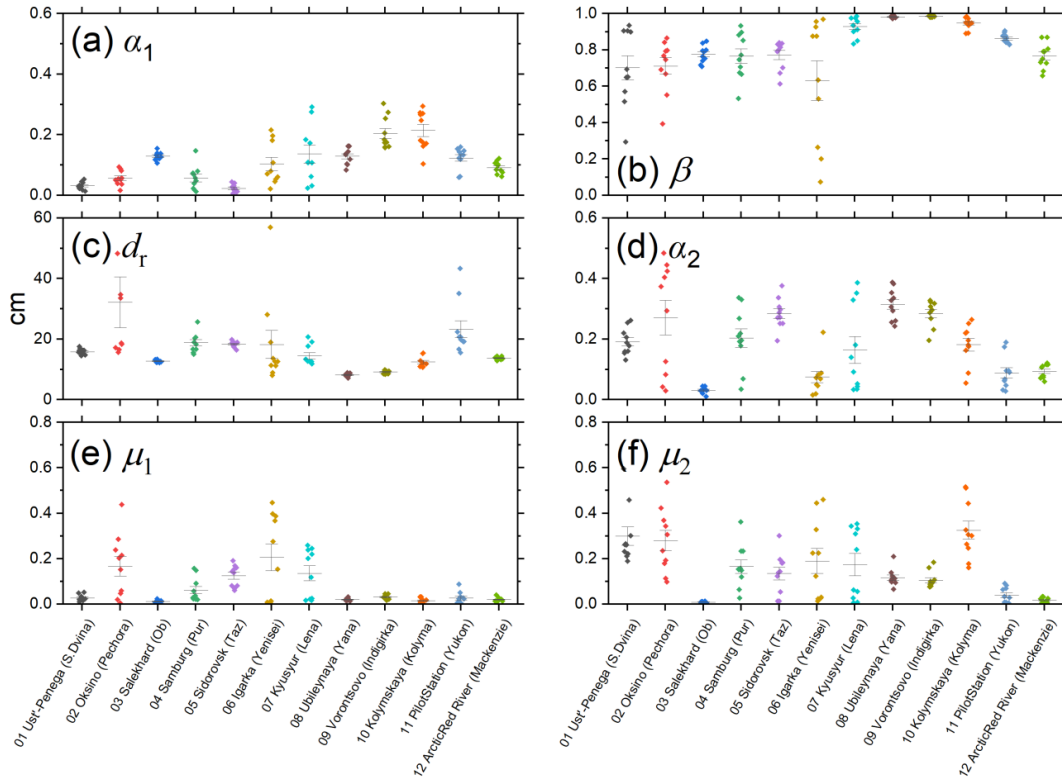


Figure 3. Distributions of parameter estimates for the refined runoff–TWS relationship obtained from the 10 runoff ensemble members at each gauging station. The runoff reconstructions used the TWS reconstructed in this study as input. An 8-year period was used for parameter estimation at each station, with the specific periods shown in Figure 7a. For each ensemble member, the parameter estimate shown is the mean of the 20,000 accepted MCMC samples. Parameter d_r is expressed as equivalent water height (unit: cm) to facilitate comparison among stations.

The uncertainty of the ensemble-mean reconstruction was quantified from parameter uncertainty and data-source uncertainty. For each ensemble member, 20,000 accepted parameter samples from the MCMC analysis were propagated through the reconstruction model. The parameter uncertainty of ensemble member s , $\sigma_{p,s}$, was defined as the standard deviation of the resulting reconstructions at each time step. Assuming negligible covariance in parameter uncertainty among ensemble members, their contribution to the uncertainty of the ensemble mean was calculated as $\sum_{s=1}^N \sigma_{p,s}^2 / N^2$. Data-source uncertainty, σ_d , was represented by the ensemble spread, calculated as the standard deviation among ensemble reconstructions at each time step. This term reflects differences arising from the forcing, auxiliary, and reference datasets. The total reconstruction uncertainty was then calculated as

$$\sigma_{\text{Recon}} = \sqrt{\left(\sum_{s=1}^N \sigma_{p,s}^2 / N^2 \right) + \sigma_d^2} \quad (11)$$

Here, N denotes the number of ensemble members, with $N=75$ for TWS reconstruction and $N=10$ for runoff reconstruction.



3.4 Evaluation metrics

Reconstruction performance was evaluated using the correlation coefficient (r), Nash–Sutcliffe efficiency (NSE), Kling–Gupta efficiency (KGE; Gupta et al., 2009), and root mean square error (RMSE). The correlation coefficient measures the consistency of temporal variations between reconstructed and reference time series, whereas RMSE measures the absolute magnitude of reconstruction errors. Both were used as complementary metrics for TWS and runoff evaluation.

NSE evaluates reconstruction errors relative to the variance of the reference data and was used as the primary metric for TWS anomalies. KGE combines correlation, relative variability, and mean bias. Because its bias component depends on the ratio of reconstructed to reference means, KGE can become unstable when the reference mean approaches zero, as is typical for TWS anomalies defined relative to a reference baseline. KGE was therefore not used as the primary metric for TWS evaluation. In contrast, runoff has a well-defined positive mean, allowing the KGE components to be interpreted more reliably. KGE was therefore used as the primary metric for runoff evaluation.

4 Results

4.1 Evaluation of the improved TWS reconstruction model

To evaluate the reliability of the improved TWS reconstruction model, we first conducted training and testing experiments. The training experiment assessed the model's ability to reproduce TWS during the parameter-estimation period, whereas the testing experiment assessed its temporal extrapolation beyond this period. The training period corresponded to the parameter-estimation period from January 2003 to December 2010. The testing period covered the remaining available GRACE (-FO) observations from January 2011 to December 2025, excluding the data gap between the two missions. The mean TWS time series from the CSR, JPL, and GSFC mascon products was used as the reference for calculating the evaluation metrics.

Figure 4a shows the monthly NSE values during the training and testing periods for the 12 study catchments. For the 10 Eurasian catchments (IDs: 01–10), the reconstructed TWS showed consistently good agreement with GRACE (-FO) observations during both periods, with most catchments lying close to the $y=x$ line. For the aggregated Eurasian region (sum of 01–10), NSE reached 0.94 during training and 0.92 during testing (Figure 4b). Lower testing performance was mainly found in the northern lowland catchments of Yana (ID: 08), Indigirka (ID: 09), and Kolyma (ID: 10), where the disagreement was primarily associated with weaker reconstructed long-term trends (Figure S2). This mismatch may partly reflect long-term TWS changes associated with permafrost degradation, which are not explicitly represented in the reconstruction model.

The Yukon (ID: 11) and Mackenzie (ID: 12) catchments in North America also show clear differences in long-term trends between reconstructed and GRACE (-FO) TWS (Figure 4c and S2), resulting in substantially lower testing NSE values. The approximately parallel distribution of points relative to the $y=x$ line in Figure 4e further suggests that much of the disagreement is associated with low-frequency differences rather than monthly variability. For Yukon, the coarse effective spatial resolution of the JPL mascon product may allow leakage from glacier mass loss in the Gulf of Alaska into the study catchment, resulting in a stronger negative TWS trend. In contrast, the reconstructed TWS shows closer agreement with the CSR and GSFC products (Figure S1). For Mackenzie, uncertainty in GIA corrections may also contribute to the trend discrepancy, as GRACE (-FO) TWS estimates vary substantially among different GIA corrections (Figure S3). In addition, the limited representation of storage changes in large lakes may contribute to the weaker recent TWS decline in the reconstruction, as discussed further in Section 5.6. After removing the linear trends, NSE increased for both Yukon and Mackenzie (Figure S4), indicating that differences in long-term trends account for much of their lower testing performance.

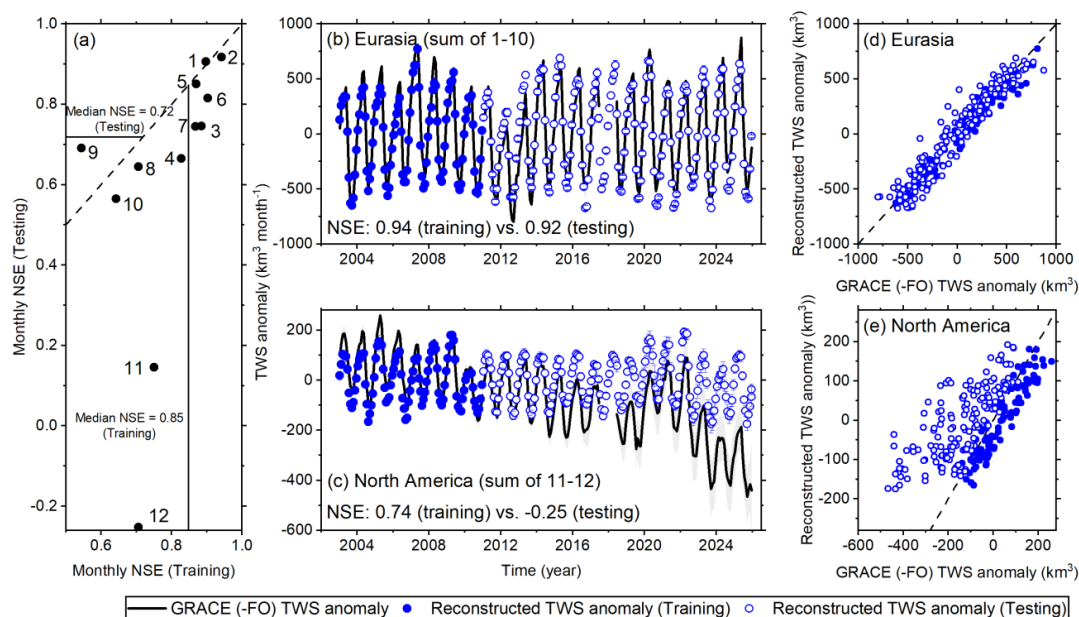


Figure 4. Training and testing evaluation of the improved TWS reconstruction model. The mean TWS time series from the CSR, JPL, and GSFC GRACE (-FO) mascon products was used as the reference for calculating the evaluation metrics. (a) Monthly NSE during the training and testing periods for the 12 study catchments. The dashed line indicates the $y=x$ relationship. (b, c) Comparison of reconstructed and GRACE (-FO) TWS anomalies for the aggregated Eurasian catchments (IDs 01–10) and North American catchments (IDs 11–12), respectively. Filled and open blue circles indicate reconstructed TWS during the training and testing periods, respectively. (d, e) Corresponding scatterplots between reconstructed and GRACE (-FO) TWS anomalies for Eurasia and North America.

We further decomposed the monthly TWS time series into the mean seasonal cycle and the non-seasonal component. The reconstructed mean seasonal cycles agree closely with GRACE (-FO) observations in all 12 catchments (Figure 5a and b), with a median NSE of 0.97, indicating that the improved model effectively reproduces seasonal TWS variability in the Pan-Arctic region. For the non-seasonal component, the reconstructed TWS also captures the major interannual variations observed by GRACE (-FO) in most regions (Figure 5a and c), with a median correlation coefficient of 0.74. Lower correlations remain concentrated in the northern Eurasian lowland and the two North American catchments and are largely associated with differences in linear trends (Figure S5). Correlations between reconstructed and observed non-seasonal components further increase after linear detrending (Figure S6).

These results show that the improved model effectively captures both the seasonal cycle and major non-seasonal TWS variability during the GRACE (-FO) period. However, this evaluation is limited to the GRACE (-FO) era and cannot directly assess the reliability of the reconstructed TWS for 1952–2002. In the following section, we evaluate the historical TWS reconstruction indirectly by using reconstructed TWS as input to runoff reconstruction and comparing the resulting runoff with long-term gauge observations.

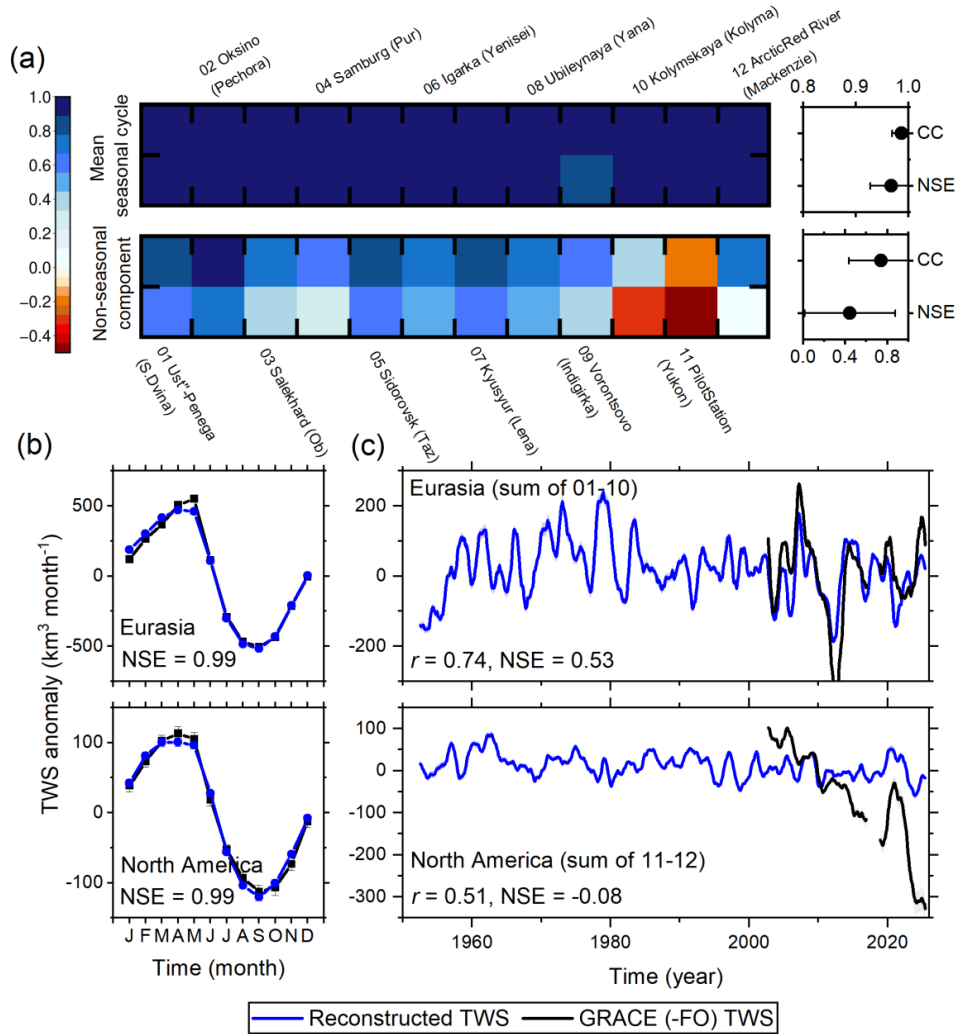


Figure 5. Evaluation of the mean seasonal cycle and non-seasonal component of reconstructed TWS. (a) Correlation coefficient (r) and NSE for the mean seasonal cycle and the non-seasonal component in the 12 catchments, with a 1-year sliding window applied to the non-seasonal component. The black dots and error bars indicate the median and standard deviation of the evaluation metrics, respectively. (b) Comparison of reconstructed and GRACE and GRACE-FO mean seasonal cycles for the aggregated Eurasian catchments (IDs: 01–10) and North American catchments (IDs: 11–12). (c) Corresponding comparison of the non-seasonal components.

4.2 Evaluation of the refined runoff–TWS relationship

4.2.1 Evaluation using GRACE (-FO) TWS

To evaluate the refined runoff–TWS relationship independently of errors in reconstructed TWS, we used GRACE (-FO) TWS directly as input for runoff reconstruction. This allows the refined relationship to be compared more directly with the previous relationship of Yi et al. (2023). For a consistent comparison, parameters for both relationships were estimated using data from January 2003 to December 2010.

Figure 6a shows that the refined runoff–TWS relationship achieved higher monthly KGE values at nearly all gauging stations, with the median increasing from 0.82 to 0.90, indicating a widespread improvement rather than one limited to a few specific catchments.



The mean seasonal runoff residuals across the 12 gauging stations further illustrate the differences between the two relationships (Figure 6b). Runoff residuals were calculated as observed runoff minus reconstructed runoff. The previous relationship uses time-invariant conversion coefficients α_1 and α_2 , which represent an average runoff response to liquid TWS over different hydrological conditions. Consequently, reconstructed runoff tends to be underestimated during snowmelt periods, particularly in June, and overestimated in late summer, particularly in August and September. In contrast, the refined relationship introduces the meltwater-intensity-dependent perturbation coefficients μ_1 and μ_2 , allowing runoff sensitivity to liquid TWS to vary with meltwater conditions. The resulting mean seasonal residuals are generally closer to zero, indicating improved agreement with observed runoff. The time series and residual distributions further illustrate these differences (Figure 6 c and d). For the Eurasian aggregate, the refined relationship more closely follows the observed runoff peaks and subsequent rapid recession, increasing KGE from 0.85 to 0.96 and reducing the residual standard deviation from 50.77 to 32.91 km³ month⁻¹. For the North American aggregate, the two relationships show similar overall performance, with a KGE of 0.94 for both, although the refined relationship slightly reduces the residual standard deviation from 6.50 to 5.57 km³ month⁻¹.

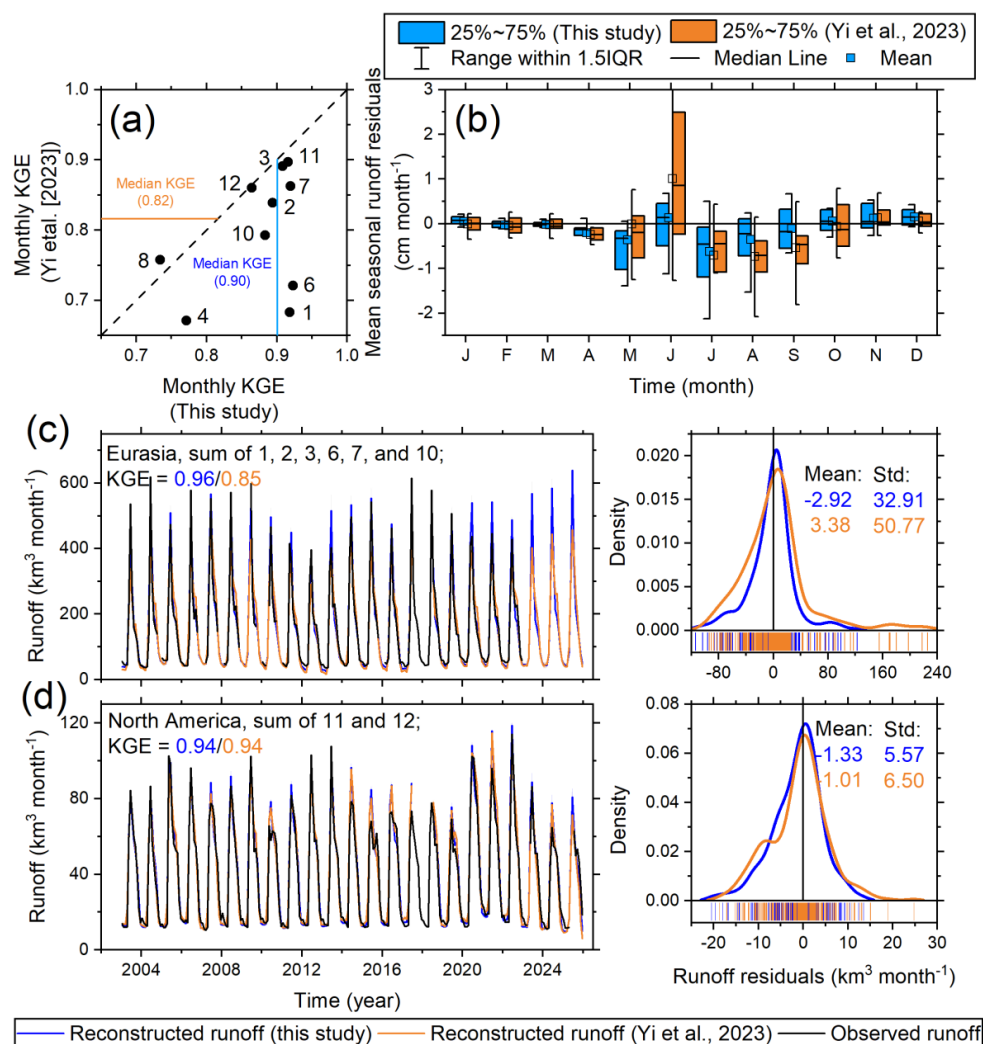


Figure 6. Evaluation of monthly runoff reconstructed from GRACE (-FO) TWS using the refined runoff-TWS relationship and the previous relationship of Yi et al. (2023). (a) Comparison of monthly KGE values at the 12 gauging stations. (b) Monthly distributions of



mean seasonal runoff residuals across the 12 stations. Residuals were calculated as observed runoff minus reconstructed runoff, with positive and negative values indicating underestimation and overestimation, respectively. (c, d) Aggregated monthly runoff time series and corresponding residual distributions for selected Eurasian catchments (IDs: 01, 02, 03, 06, 07, and 10) and the two North American catchments (IDs: 11–12), respectively.

4.2.2 Evaluation using reconstructed TWS

The previous section demonstrated the performance of the refined runoff–TWS relationship when GRACE (-FO) TWS is used as input. We then applied the relationship to the TWS reconstructed in this study to estimate long-term runoff. Comparison with gauge observations also provides a complementary indirect assessment of the historical TWS reconstruction.

Because the 12 gauging stations have different periods of available observations (Figure 7a), a common parameter-estimation period could not be used. Instead, an 8-year period was selected separately for each station and placed as early as possible within the available record to retain a longer testing period. The 8-year parameter-estimation period was defined as the training period, while the remaining available observations were used for testing.

The reconstructed monthly runoff shows stable performance between the training and testing periods (Figure 7b). The median monthly KGE values are 0.92 and 0.93 for the training and testing periods, respectively. Most stations lie close to the $y=x$ line, indicating that the model maintains good temporal extrapolation ability over several decades.

For the mean seasonal cycle, reconstructed runoff agrees closely with observations at the 12 stations, with a median KGE of 0.94 and relatively small variability among stations (Figure 7c). The examples in the left column of Figure 7d further show that the reconstruction captures the rapid rise, peak, and subsequent recession of seasonal runoff in different Pan-Arctic catchments.

To evaluate interannual runoff variability, we used annual runoff rather than the monthly non-seasonal component, because the latter has a mean close to zero after removal of the seasonal cycle and is therefore less suitable for KGE-based evaluation. Reconstructed runoff captures the major interannual variations in observed runoff, with a median annual KGE of 0.65 across the 12 stations (Figure 7c). Agreement is weaker at several stations, particularly Igarka (ID: 06) in the Yenisei basin, Pilot Station (ID: 11) in the Yukon basin, and Arctic Red River (ID: 12) in the Mackenzie basin (Figure 7d). The discrepancies may partly reflect processes not represented in the reconstruction framework, including reservoir regulation in the Yenisei basin (Yang et al., 2004), glacier meltwater (The GLaMBIE Team et al., 2025) contributions in the Yukon basin, and large-lake storage dynamics in the Mackenzie basin. These effects are discussed further in Section 5.5 and 5.6.

Overall, runoff derived from the reconstructed TWS agrees well with observations at most stations. These results support the long-term applicability of the combined TWS–runoff reconstruction framework and provide an indirect assessment of the historical TWS reconstruction. However, this assessment should be regarded as complementary rather than independent, because calibration of the runoff–TWS relationship may partly compensate for errors in the reconstructed TWS.

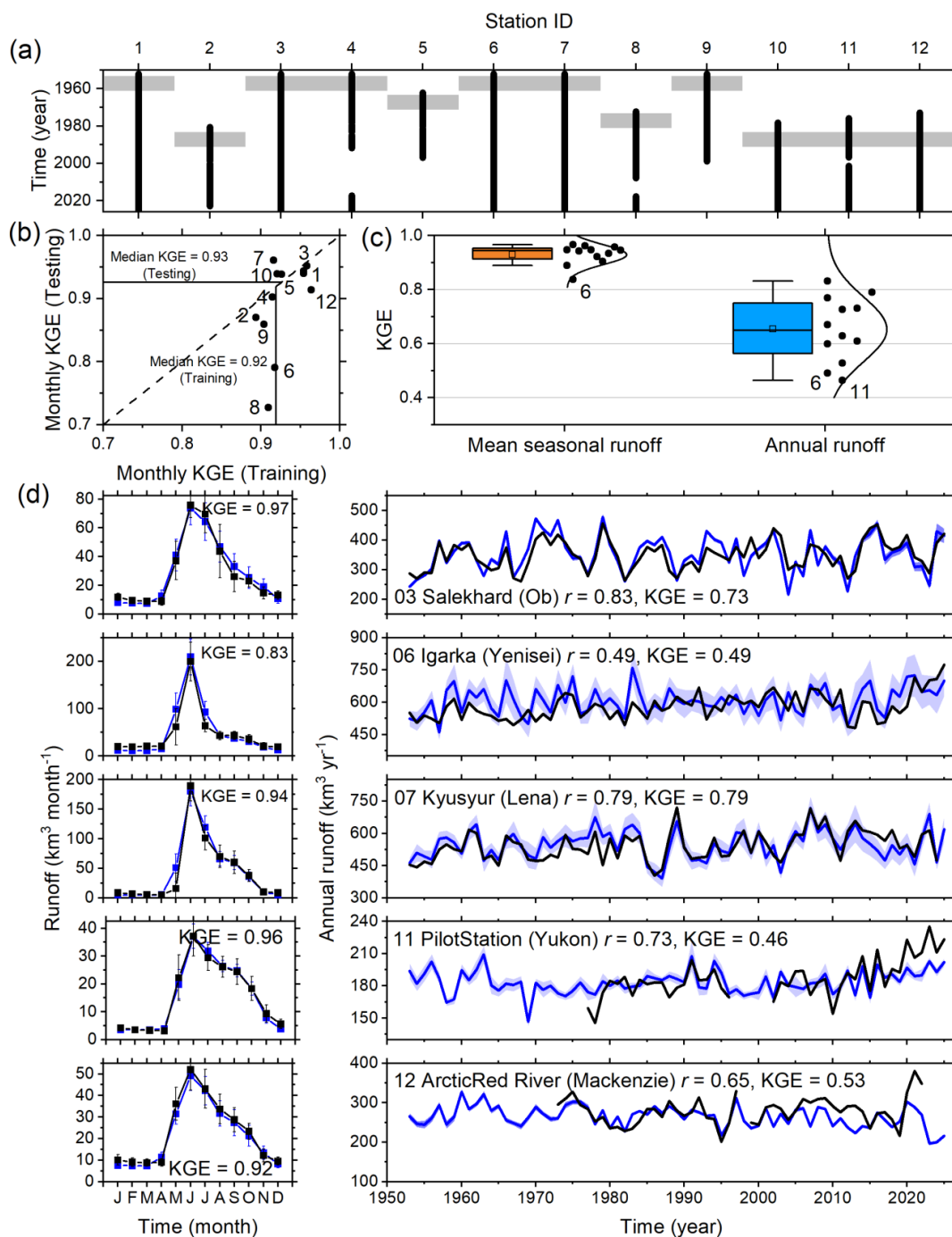


Figure 7. Evaluation of runoff reconstructed using the TWS developed in this study and the refined runoff–TWS relationship. (a) Periods of available observed runoff at the 12 gauging stations (black lines) and the 8-year periods used for parameter estimation (gray shading). (b) Comparison of monthly KGE values between the training and testing periods. The training period corresponds to the parameter-estimation period, whereas the testing period covers the remaining available observations. (c) Distributions of KGE values for the mean seasonal cycle and annual runoff at the 12 stations. Annual runoff was calculated for the hydrological year from September to August. (d) Comparison of reconstructed and observed mean seasonal cycles (left) and annual runoff time series (right) at selected stations. Blue error bars and shading indicate reconstruction uncertainty. Complete station-level time series comparisons are shown in Figure S7.



375 5 Discussion

5.1 Sensitivity to the parameter-estimation period

The improved TWS reconstruction model and the refined runoff–TWS relationship contain seven and six parameters to be estimated, respectively. This raises the question of whether their performance depends strongly on the period used for parameter estimation. We therefore conducted additional sensitivity experiments using different parameter-estimation periods.

380 For the improved TWS reconstruction model, parameter-estimation windows of 5, 10, and 15 years were considered and shifted through the available GRACE (-FO) period at 1-year intervals. Each parameter set was used to reconstruct monthly TWS from 1952 to 2025, and the resulting time series were first evaluated against GRACE (-FO) TWS. Because the satellite record covers only a limited part of the full reconstruction period, we also conducted a complementary indirect assessment using runoff. TWS reconstructed from each parameter set was used as input to the refined runoff–TWS relationship, and the resulting runoff was
 385 compared with observations.

For the refined runoff–TWS relationship, parameter-estimation windows of 5, 15, 25, 35, 45, 55, and 65 years were used and shifted at 5-year intervals. If monthly runoff observations had been continuously available from 1952 to 2025, these settings would have yielded approximately 56 possible parameter-estimation periods. In practice, the number differs among stations because of differences in observational coverage (Figure 7a).

390 The sensitivity experiments show that dependence on the parameter-estimation period is generally limited for the mean seasonal cycle (Figure 8a, c and 9a). For most catchments and stations, TWS and runoff reconstructed using different parameter-estimation periods retain high NSE or KGE values and show relatively small variations in RMSE. This indicates that the timing and amplitude of the reconstructed seasonal cycle are not strongly controlled by the choice of parameter-estimation period.

Sensitivity is somewhat greater for interannual variations (Figure 8b, d and 9b), especially for annual runoff (Figure 9b), with
 395 larger variations in the evaluation metrics at several basins and stations. This difference is expected because the seasonal cycle is repeatedly sampled within each parameter-estimation window and can therefore be constrained relatively consistently. In contrast, interannual and lower-frequency variability depends more strongly on whether the selected period includes representative wet, dry, and extreme hydrological conditions, particularly when short windows (Figures S8 and S9) such as 5 years are used. The resulting sensitivity is therefore concentrated mainly in non-seasonal and interannual variability and differs among basins, rather than
 400 indicating systematic instability of the models. Overall, both models show relatively low sensitivity to reasonable changes in the parameter-estimation period at the seasonal scale, whereas greater caution is warranted when interpreting interannual and lower-frequency variability.

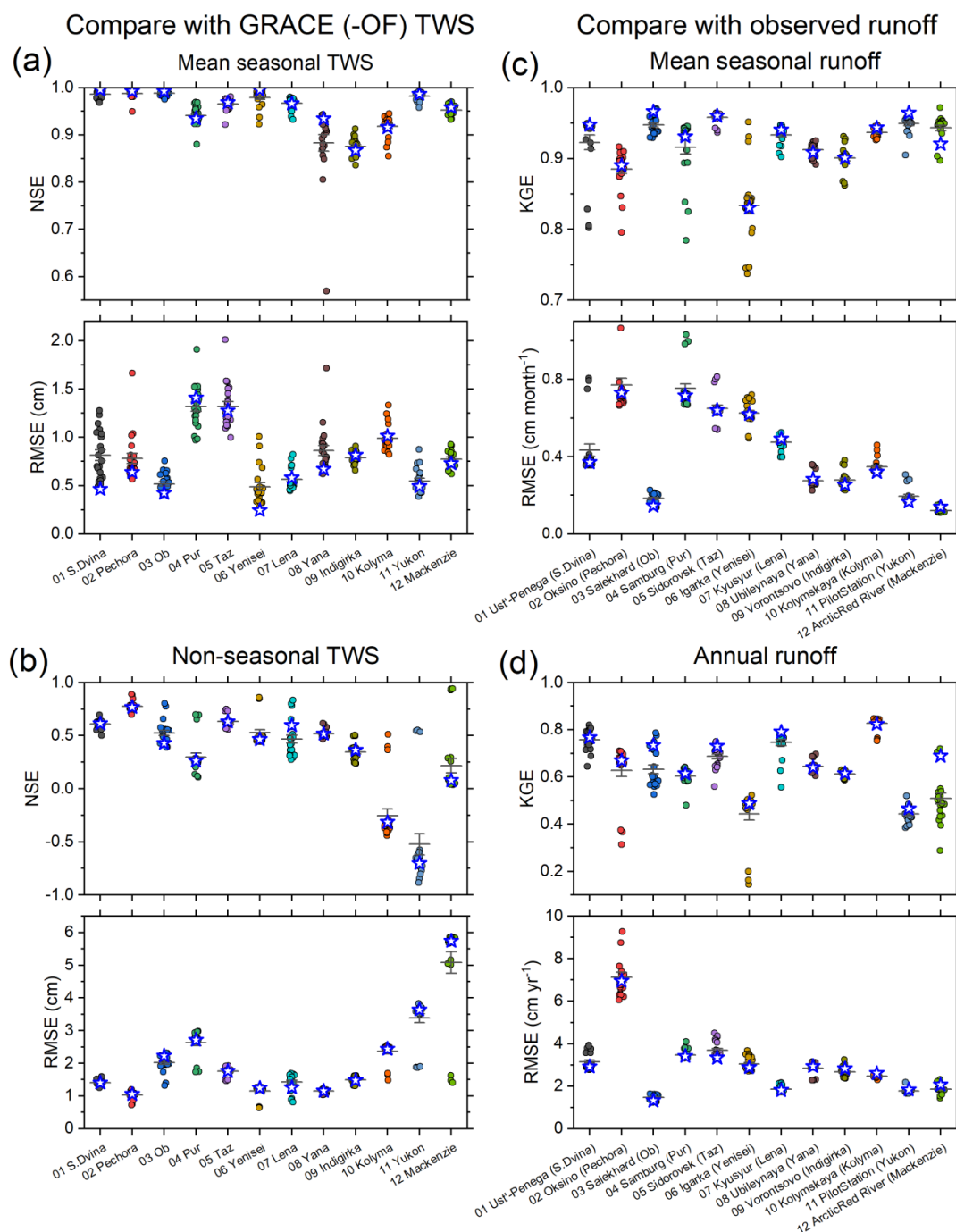


Figure 8. Sensitivity of the TWS reconstruction to the parameter-estimation period. Parameter-estimation windows of 5, 10, and 15 years were shifted through the available GRACE (-FO) record at 1-year intervals, and each parameter set was used to reconstruct monthly TWS from 1952 to 2025. (a, b) NSE and RMSE of the reconstructed mean seasonal cycle and non-seasonal component evaluated against GRACE (-FO) TWS. (c, d) KGE and RMSE of mean seasonal and annual runoff reconstructed from the corresponding TWS and evaluated against observed runoff. Each circle represents one parameter-estimation period. Dots indicate different parameter-estimation periods, and stars indicate the performance obtained using the 8-year parameter-estimation periods adopted in this study.

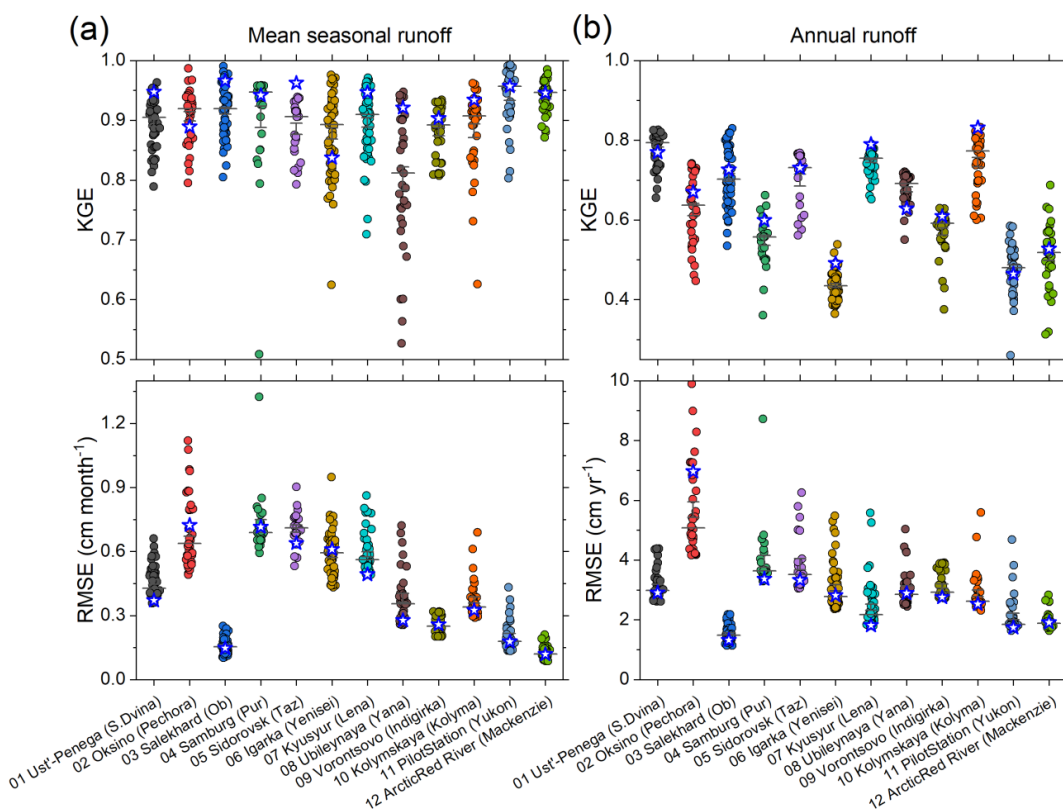


Figure 9. Sensitivity of the refined runoff–TWS relationship to the parameter-estimation period. Parameter-estimation windows of 5, 15, 25, 35, 45, 55, and 65 years were shifted through the available runoff records at 5-year intervals. (a) KGE and RMSE for the mean seasonal runoff cycle. (b) KGE and RMSE for annual runoff. Each circle represents one available parameter-estimation period, and stars indicate the periods adopted in this study (Figure 7a). Dots indicate different parameter-estimation periods, and stars indicate the performance obtained using the 8-year parameter-estimation periods adopted in this study.

5.2 Role of model parameters and model simplification

The improved TWS reconstruction model contains seven parameters, raising the question of whether all parameters make a meaningful contribution to reconstruction performance. We therefore tested a series of simplified model configurations by progressively removing selected parameters. The parameter settings for each configuration are listed in Table 4, with Scenario 1 representing the full model used in this study.

We examined model performance using both direct and complementary indirect approaches. In the direct assessment, TWS reconstructed using each scenario was compared with GRACE (-FO) TWS. In the complementary indirect assessment, each reconstructed TWS series was used for runoff reconstruction, and the resulting runoff was compared with observations. This extends the evaluation beyond the GRACE (-FO) period.

The mean seasonal cycle of TWS is more sensitive to model simplification than the non-seasonal component (Figure 10a, b). Removing b_1 in Scenario 2 has little effect on reconstruction performance, consistent with its near-zero estimates in most catchments (Figure 2b). This suggests that b_1 provides only a limited additional contribution to the current model formulation in most catchments. This is also consistent with a relatively weak influence of below-zero temperature variations on monthly liquid TWS changes. In contrast, removing the lag parameter ω in Scenario 4 clearly reduces performance for the mean seasonal cycle. This indicates that including a lagged temperature response improves the representation of the mean seasonal TWS cycle.



Parameters c and α_w show a relatively strong interaction. Removing c alone in Scenario 5 substantially reduces performance for the mean seasonal cycle, whereas removing both c and α_w in Scenario 6 restores much of the lost performance. The two parameters may therefore partly compensate for each other. The broad distribution of estimated α_w values in several catchments (Figure 2f) further suggests sensitivity to the choice of input datasets and possible weak identifiability in some catchments.

435 Notably, Scenario 8, which retains only a, b_2, d_w , and ω , achieves performance close to that of the full model in the direct TWS evaluation. This indicates that a reduced four-parameter formulation can reproduce much of the full model performance during the satellite period. However, when the reconstructed TWS is subsequently used for runoff reconstruction, Scenario 8 performs worse than Scenario 1 at several stations (Figure 10c, d). Thus, a simplified formulation that performs well during the GRACE (-FO) period does not necessarily preserve performance over longer periods or in downstream runoff reconstruction.

440 These experiments do not indicate that all seven parameters are equally important or individually necessary in every catchment. Rather, they show that no single simplified configuration consistently reproduces the performance of the full model across catchments and evaluation scales. We therefore retain Scenario 1 as a common formulation for Pan-Arctic TWS reconstruction. Together with the sensitivity analysis in Section 5.1, these results support the use of the full formulation for Pan-Arctic applications, whereas basin-specific studies may consider further simplification after evaluating parameter sensitivity and identifiability.

445 **Table 4. Parameter configurations used to evaluate simplifications of the TWS reconstruction model.**

Scenario	Estimated parameters	Number of estimated parameters
Scenario 1	$a, b_1, b_2, c, d_w, \alpha$, and ω	7
Scenario 2	a, b_2, c, d_w, α , and ω	6
Scenario 3	a, b_1, b_2, c, d_w , and ω	6
Scenario 4	a, b_1, b_2, c, d_w , and α	6
Scenario 5	a, b_1, b_2, d_w, α , and ω	6
Scenario 6	a, b_1, b_2, d_w , and ω	5
Scenario 7	a, b_1, b_2 , and d_w	4
Scenario 8	a, b_2, d_w , and ω	4
Scenario 9	a, b_2 , and d_w	3

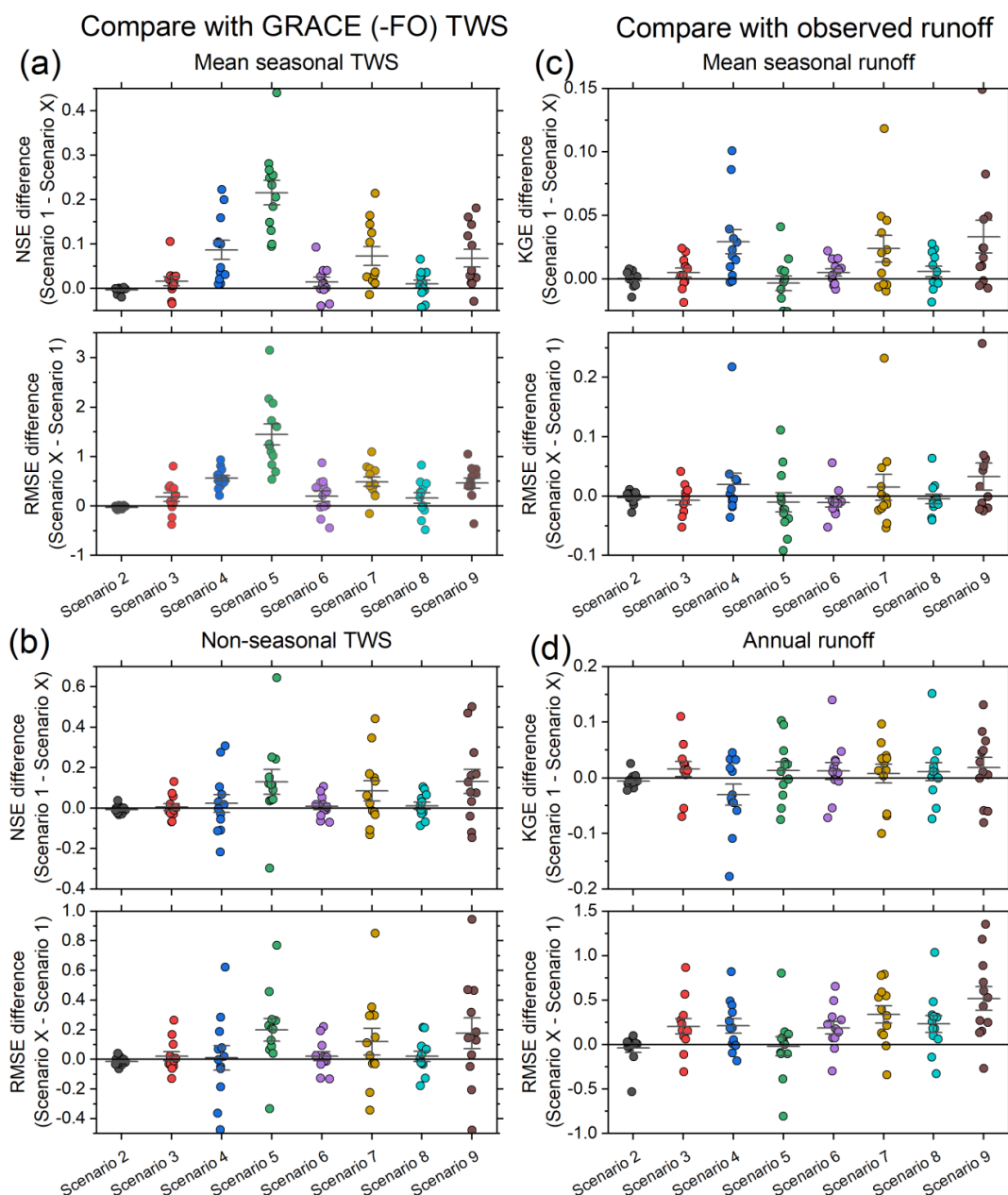


Figure 10. Effects of model simplification on TWS reconstruction performance relative to the full model (Scenario 1). Parameter configurations for each scenario are given in Table 4. (a, b) Direct evaluation against GRACE (-FO) TWS for the mean seasonal cycle and non-seasonal component, respectively. (c, d) Complementary indirect evaluation using runoff reconstructed from the corresponding TWS for the mean seasonal cycle and annual runoff, respectively. NSE and KGE differences are calculated as Scenario 1 minus Scenario X, whereas RMSE differences are calculated as Scenario X minus Scenario 1. Positive values therefore indicate better performance of Scenario 1. Each point represents one catchment or gauging station. Gray horizontal lines and error bars indicate the median and standard deviation, respectively.



5.3 Comparison with existing GRACE-like TWS reconstructions

455 Several recent studies have reconstructed historical GRACE-like TWS from meteorological and hydrological data using statistical or machine learning methods. We therefore compared our reconstruction with five publicly available GRACE-like TWS products. Four were developed using machine learning methods (Gentner et al., 2026; Hacker et al., 2026; Li et al., 2021; Yin et al., 2023), and one was based on an empirical model (Humphrey and Gudmundsson, 2019). Although these products generally reproduce the GRACE (-FO) mean seasonal cycle well, their non-seasonal components diverge substantially in the pre-GRACE period (Figure 460 11a). The reconstruction of Hacker et al. (2026), in particular, shows a clear long-term trend. This may partly reflect the additional low-resolution TWS constraint derived from satellite laser ranging used in that reconstruction, which may be more susceptible to spatial signal leakage, including contributions from glacier mass loss around the Gulf of Alaska.

To extend the comparison beyond the GRACE and GRACE-FO period, we used each TWS reconstruction as input to the refined runoff–TWS relationship to reconstruct runoff (Figure 11b and c). For each TWS product, the parameters of the runoff–TWS 465 relationship were re-estimated independently using the same station-specific parameter-estimation period. Most products reproduce the mean seasonal runoff reasonably well, consistent with their similar seasonal TWS cycles. In contrast, their performance differs much more at the annual scale (Figure 11b). Among the compared products, our reconstruction yields the highest median annual KGE and one of the lowest RMSE values. Its annual runoff residuals are centered close to zero, with a mean of -0.72 cm yr^{-1} and a standard deviation of 2.58 cm yr^{-1} (Figure 11c). The mean seasonal runoff residuals are also small, with a 470 mean of $-0.07 \text{ cm month}^{-1}$ and a standard deviation of $0.44 \text{ cm month}^{-1}$, although some products, such as Gentner et al. (2026), show comparable performance at this timescale.

These comparisons indicate that good agreement with GRACE (-FO) during the satellite era alone is insufficient to establish the reliability of historical TWS reconstructions. Long-term runoff observations provide a complementary indirect assessment of historical TWS through the refined runoff–TWS relationship.

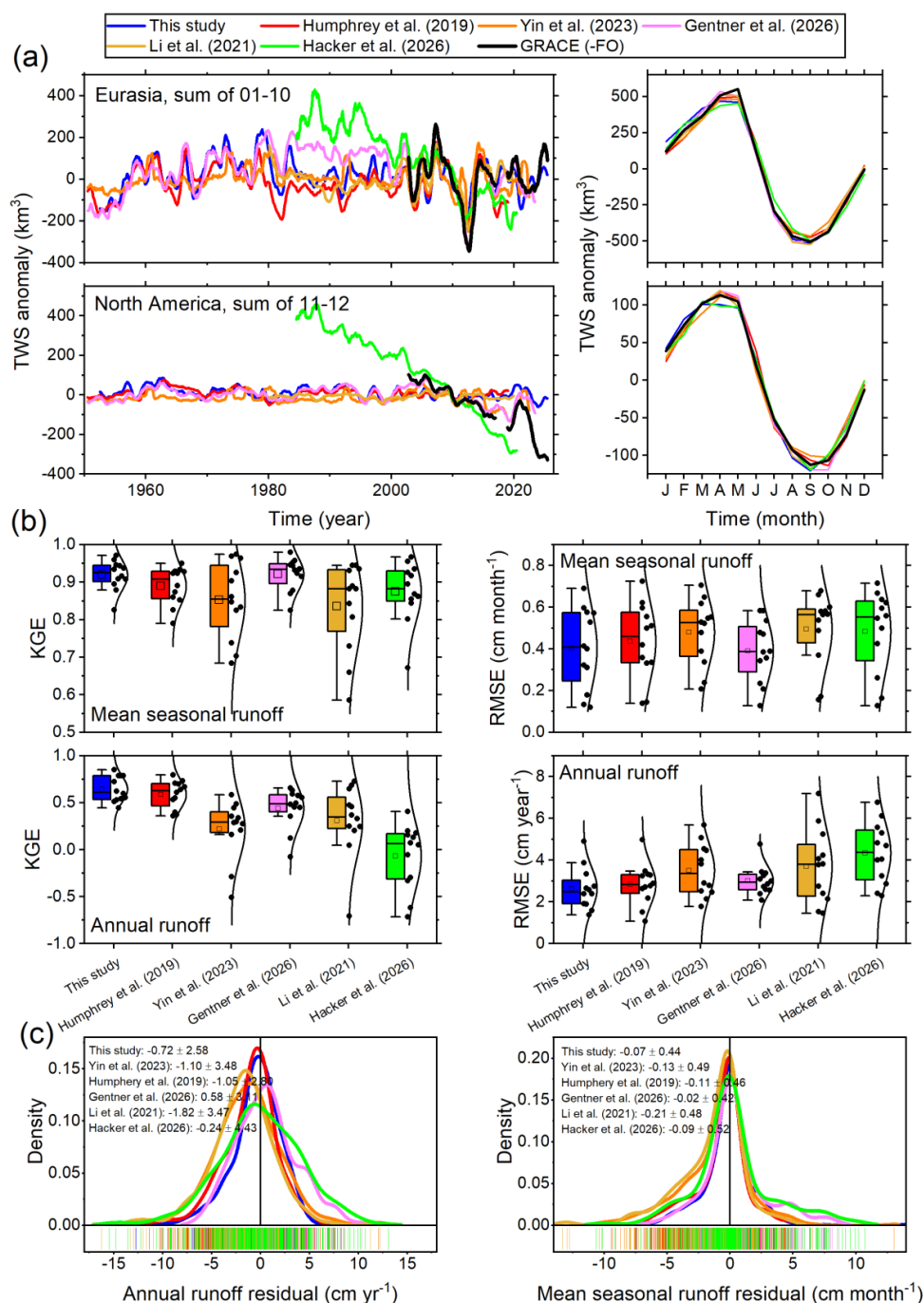


Figure 11. Comparison of the TWS reconstruction developed in this study with five existing GRACE-like TWS products. (a) Non-seasonal TWS time series and mean seasonal cycles for the aggregated Eurasian basins (IDs: 01–10) and North American basins (IDs: 11–12). Mean seasonal cycles were calculated over April 2002 to December 2025. (b) KGE and RMSE of mean seasonal and annual runoff reconstructed from the different TWS products using the refined runoff–TWS relationship. The evaluation was conducted over the common period 1985–2019. (c) Distributions of annual and mean seasonal runoff residuals. Residuals were calculated as observed runoff minus reconstructed runoff.



5.4 Comparison with an existing runoff reconstruction

We further compared our reconstructed runoff with GRUN, a publicly available 0.5° gridded runoff reconstruction based on machine learning (Ghiggi et al., 2019). GRUN provides a useful external reference because its reconstruction is also primarily climate driven and does not explicitly represent human regulation or glacier melt. However, GRUN is a gridded runoff product, and catchment-aggregated runoff is not strictly equivalent to gauge-derived runoff at the basin outlet (Ghiggi et al., 2019) because routing and storage processes are not represented in the same way. To reduce the influence of these scale differences, we focused on runoff anomalies relative to a common reference period rather than absolute runoff. And, residuals were calculated as observed runoff anomalies minus reconstructed runoff anomalies. The purpose of this comparison is therefore not to rank the two reconstructions, but to examine whether they recover similar runoff variability.

The mean seasonal runoff residuals from our reconstruction are more tightly centered around zero than those from GRUN (Figure 12a), and seasonal RMSE values are generally lower at the representative stations shown in Figure 12c. This indicates closer agreement with the observed timing and amplitude of the seasonal runoff cycle. At the annual scale, the residual distributions are more similar, and the relative performance varies among stations (Figure 12b and c). Neither reconstruction shows a consistent advantage in representing interannual runoff variability.

The similarity in annual-scale performance is nevertheless noteworthy because the two reconstructions follow different modeling pathways. GRUN reconstructs runoff directly from climate-related information, whereas our runoff is inferred indirectly from reconstructed TWS through the refined runoff–TWS relationship. Despite these differences, both capture broadly consistent interannual variations in observed runoff. This provides complementary support that the reconstructed historical TWS retains hydrologically meaningful interannual variability.

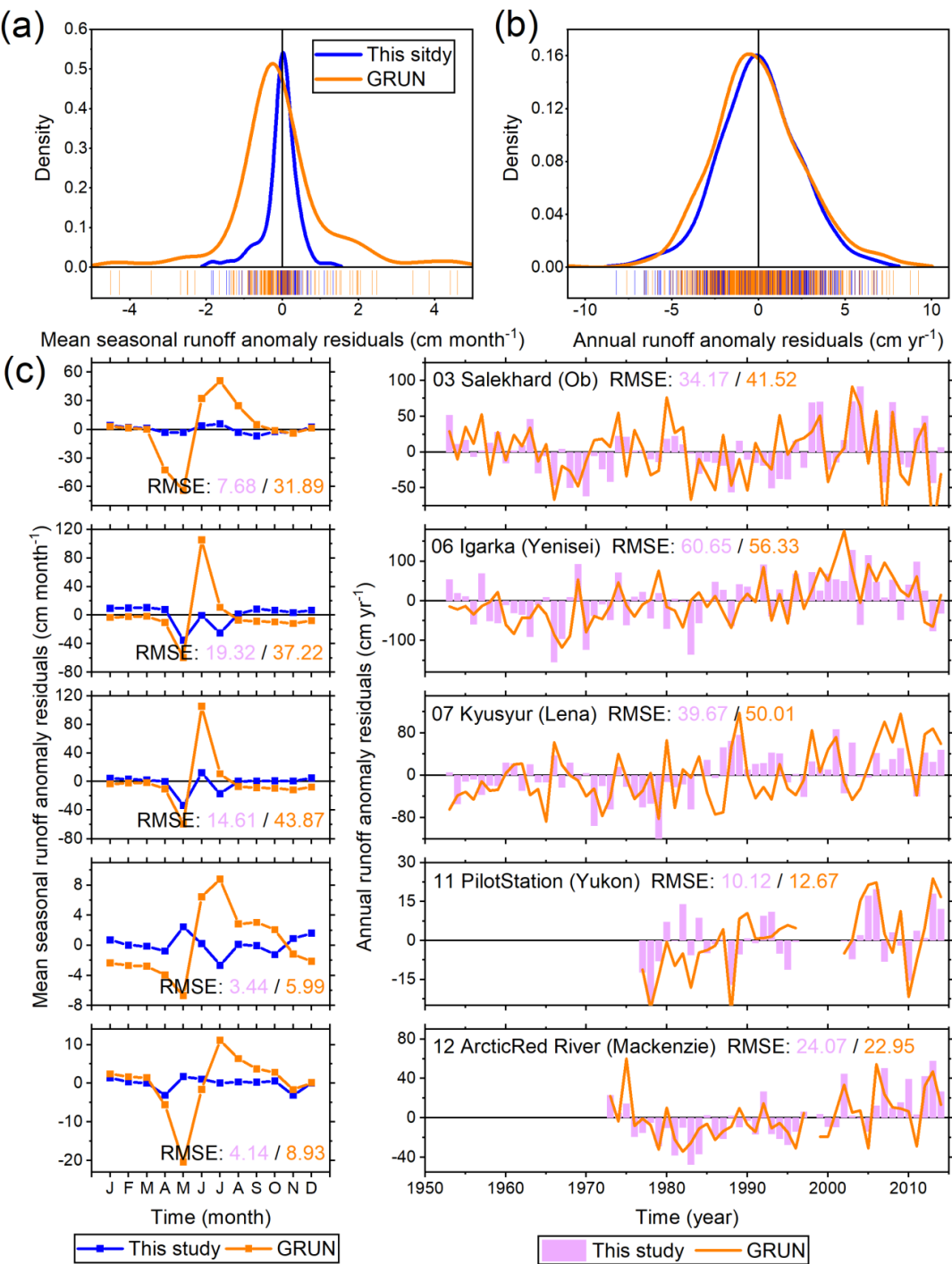


Figure 12. Comparison of runoff anomaly residuals from the reconstruction developed in this study and GRUN at the 12 Pan-Arctic gauging stations. Residuals were calculated as observed runoff anomalies minus reconstructed runoff anomalies. For each comparison, runoff anomalies were calculated relative to the common reference period shared by the observations and both reconstructions. (a) Distributions of mean seasonal runoff residuals. (b) Distributions of annual runoff residuals. (c) Mean seasonal runoff residuals (left) and annual runoff residuals (right) at selected gauging stations.

505



5.5 Effects of human regulation on runoff evaluation

Since the late 1950s, several large dams have been constructed for hydropower generation in the upper Yenisei basin (Davidson, 2016), substantially altering regional runoff regimes (Yang et al., 2004). To examine whether the discrepancy between reconstructed and observed runoff at Igarka (ID: 06) is associated with human regulation, we extended the analysis to 10 additional gauging stations in the Yenisei basin. The 11 stations were classified as regulated or unregulated according to the presence of major upstream dams (Figure 13a). To avoid incorporating regulation effects into the parameter estimates, only pre-regulation observations were used at regulated stations, with all parameter-estimation periods ending before 1960.

At the unregulated stations, reconstructed monthly runoff agrees well with observations during both the training and testing periods (Figure 13b). In contrast, the regulated stations generally show high KGE values during the pre-regulation training period but substantially lower values during the testing period, which includes reservoir operation. This contrast is also evident in the mean seasonal cycle (Figure 13c and d). Observed runoff at regulated stations shows clear seasonal redistribution, consistent with reservoir operation that stores water during high-flow summer months and increases releases during low-flow winter months for hydropower generation (Yang et al., 2004). Agreement at the annual scale is also weaker at some regulated stations (Figure 13c and d), suggesting that reservoir operation and multiyear changes in reservoir storage may also affect interannual runoff variability. By comparison, reconstructed and observed runoff remain in close agreement at the unregulated stations (Figure 13c and e), supporting the ability of the model to recover seasonal and interannual runoff variations primarily driven by climate conditions.

These results also highlight the importance of the parameter-estimation period at regulated stations. Using post-regulation runoff for parameter estimation may cause regulation-induced changes to be incorrectly attributed to TWS, snow cover, or snow storage, because these predictors do not explicitly encode reservoir operation in the current framework. We therefore restricted parameter estimation at regulated stations to the period before major reservoir operation. The resulting reconstructions can therefore be interpreted as an approximation of the runoff expected under climate-driven hydrological conditions, whereas subsequent differences from observed runoff may reflect reservoir regulation and other processes not represented in the model. This comparison should therefore be interpreted as evidence consistent with the influence of reservoir regulation rather than as a quantitative attribution of anthropogenic effects.

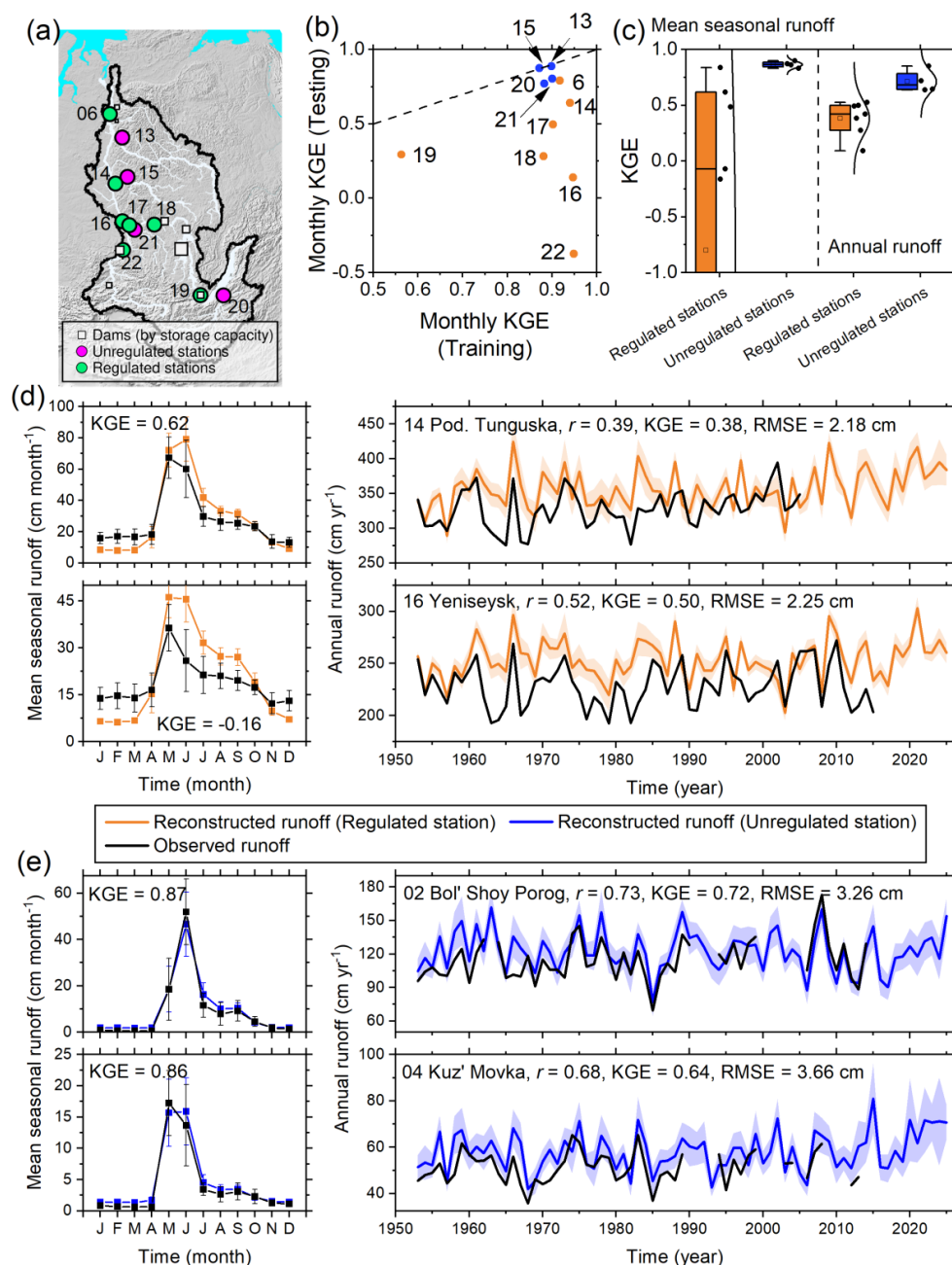


Figure 13. Effects of human regulation on runoff evaluation in the Yenisei basin. (a) Locations of the 11 gauging stations and major dams (Lehner et al., 2024) used in the analysis. Stations were classified into four unregulated and seven regulated stations according to the presence of major upstream dams. Dam symbol size represents storage capacity. **(b)** Comparison of monthly KGE values between the training and testing periods. For regulated stations, parameter estimation was restricted to the pre-regulation period before 1960. **(c)** Distributions of KGE values for the mean seasonal cycle and annual runoff at regulated and unregulated stations. **(d, e)** Comparisons of reconstructed and observed mean seasonal cycles and annual runoff at selected regulated and unregulated stations, respectively.

535



5.6 Limitations of the reconstruction framework

Although the evaluations above show generally good performance of the improved TWS reconstruction model and the refined runoff–TWS relationship, several limitations remain. First, similar to the previous model of Liu et al. (2021), the current TWS reconstruction framework does not explicitly represent the storage and release dynamics of large lakes. In the Mackenzie basin, satellite altimetry and in situ observations show clear fluctuations in large-lake storage during the later part of the study period (Figure 14b). However, the reconstructed TWS for catchments containing these lakes (Figure 14d) does not fully capture the corresponding storage variations (Figure 14a), and the runoff derived from reconstructed TWS also fails to reproduce some of the observed fluctuations during these periods (Figure 14c). Large-lake storage dynamics may therefore contribute to reconstruction errors in some basins.

Second, the improved TWS reconstruction model does not explicitly represent ground-ice changes or the hydrological responses associated with permafrost degradation. Permafrost degradation can alter groundwater flow pathways and hydrological connectivity, thereby changing basin-scale water-storage components (Bring et al., 2016). However, these effects remain difficult to quantify because long-term observations are sparse and discontinuous. Recent studies have reported widespread decreases in thermokarst lake area in northern lowland permafrost regions during the past two decades (Chen et al., 2023; Jones et al., 2022; Webb et al., 2022; Webb and Liljedahl, 2023), which have been linked to permafrost degradation (Webb and Liljedahl, 2023). These changes are broadly consistent with the persistent TWS losses observed by GRACE (-FO) in the Yana, Indigirka, and Kolyma basins, whereas our reconstruction does not fully capture these long-term declines (Figure S5). The absence of explicit permafrost-related processes may therefore limit the ability of the model to reproduce long-term TWS changes in some northern lowland basins.

Finally, because the TWS reconstruction model is driven by precipitation, air temperature, and snow-related inputs, direct human influences are not included as independent drivers. The reconstructed TWS should therefore be interpreted primarily as TWS variability driven by the climate-related inputs represented in the model. GRACE (-FO) observations, however, contain signals from both climate variability and human activities. Removing the linear trend before parameter estimation reduces the influence of persistent anthropogenic signals, such as long-term groundwater depletion (Rodell et al., 2018), but does not eliminate shorter-term or interannual effects associated with human regulation. For example, reservoir impoundment and operation may introduce interannual storage variations (Xie et al., 2022) that are not fully removed by detrending and may contribute to uncertainty in parameter estimation. Such effects are particularly relevant in strongly regulated basins, such as the Yenisei. Region-specific applications should therefore evaluate potential anthropogenic signals in the reference TWS and, where possible, account for them before parameter estimation. A detailed attribution and correction of anthropogenic effects is beyond the scope of this paper and warrants further study.

These limitations are most relevant in basins strongly affected by large-lake storage, permafrost change, glacier-related signals, or human regulation and should be considered when interpreting long-term TWS and runoff variability. Future improvements could incorporate additional information on lake storage, permafrost dynamics, glacier contributions, and human water management to better represent these processes.

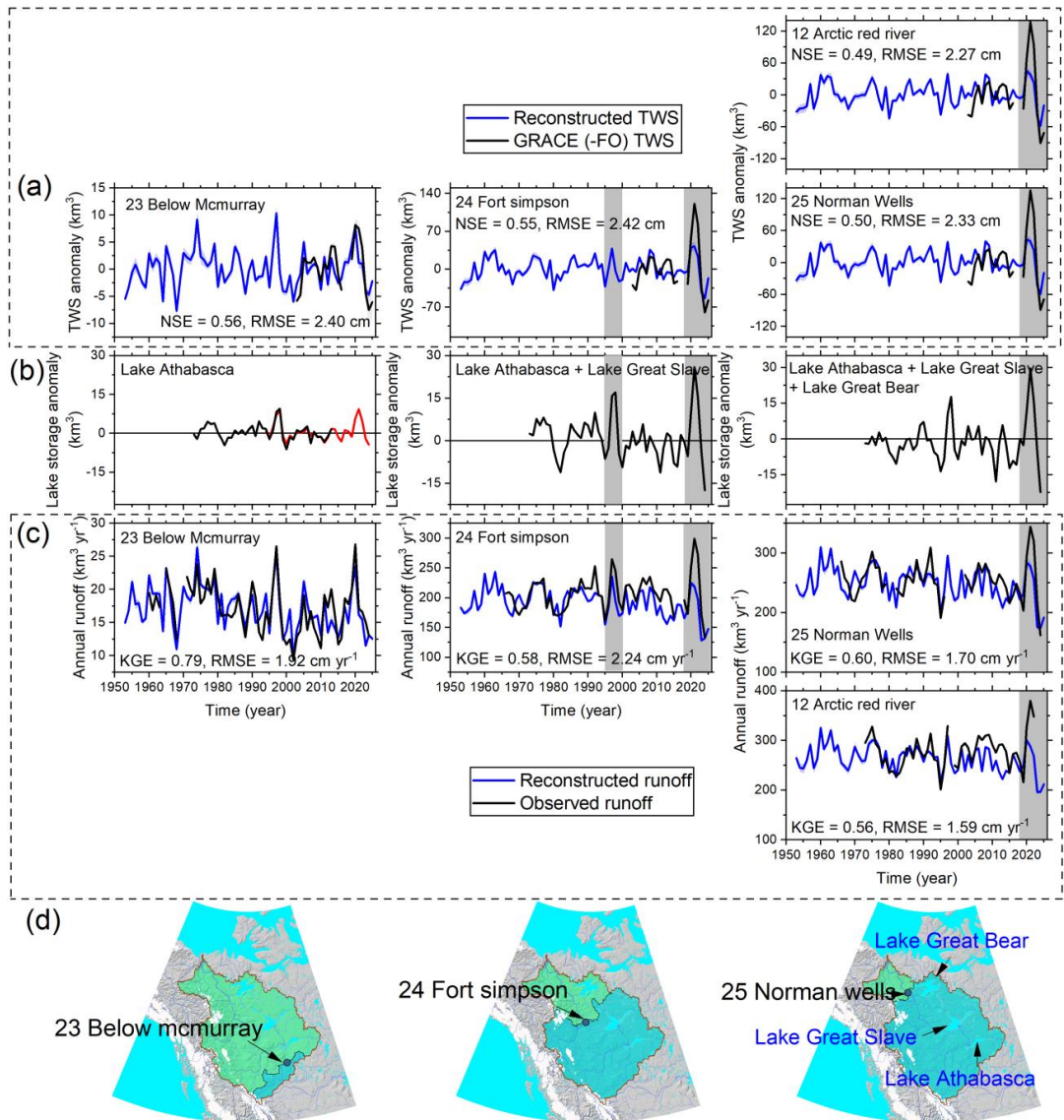


Figure 14. Comparison of large-lake storage variations with TWS and runoff reconstruction in the Mackenzie basin. (a) Comparison of annual reconstructed and GRACE (-FO) TWS anomalies for selected gauging-station basins. Linear trends were removed from both time series to highlight differences during periods of lake-storage variability. (b) Annual storage anomalies of the three major lakes in the Mackenzie basin derived from satellite altimetry and in situ observations. Combined series represent the summed storage anomalies of the indicated lakes. (c) Comparison of annual runoff derived from reconstructed TWS with observed runoff at the corresponding gauging stations. (d) Locations of the gauging stations and their upstream drainage basins. Gray shading indicates periods of large-lake storage variability, during which the reconstructed TWS and runoff show larger differences from observations.

6 Conclusion

In this study, we developed two empirical models within a common framework to reconstruct long-term TWS and runoff in the Pan-Arctic region. Monthly TWS was reconstructed from precipitation, air temperature, and snow information using an improved state-update empirical model, with GRACE TWS observations used for parameter estimation. Runoff was subsequently estimated



from TWS using a refined runoff–TWS relationship calibrated against gauge observations. Using this framework, we reconstructed monthly TWS and runoff for 12 major Pan-Arctic basins from 1952 to 2025.

Multiple evaluations support the reliability of the proposed framework. The improved TWS model captures both the mean seasonal cycle observed by GRACE (-FO) and the major non-seasonal variations in most basins, with median NSE and correlation coefficient values of 0.97 and 0.74, respectively. Evaluation using GRACE (-FO) TWS shows that the refined runoff–TWS relationship improves monthly runoff reconstruction relative to the previous formulation, increasing the median KGE from 0.82 to 0.90 and reducing systematic seasonal biases. When applied to reconstructed TWS, the relationship maintains similar performance beyond the parameter-estimation periods, with median monthly KGE values of 0.92 during the 8-year parameter-estimation periods and 0.93 over the remaining available observations. Reconstructed annual runoff also captures the major interannual variations in observed runoff, with a median annual KGE of 0.65. This agreement with long-term gauge observations provides a complementary indirect assessment of the historical TWS reconstruction.

Comparisons with existing GRACE-like TWS products and a gridded runoff reconstruction provide further support for the long-term consistency of the reconstructed TWS and runoff. Sensitivity tests show that the reconstructions are not strongly dependent on the choice of parameter-estimation period, particularly for seasonal variability. Model simplification experiments further show that although individual parameters have different importance among basins, no single simplified configuration consistently reproduces the performance of the full model across catchments and evaluation scales.

Important limitations remain because large-lake storage dynamics, permafrost-related hydrological changes, glacier processes, and human regulation are not explicitly represented and may affect reconstruction performance in some basins. Overall, the proposed framework provides a simple empirical approach for extending Pan-Arctic TWS records into the pre-GRACE era and producing continuous runoff reconstructions over the full study period, thereby supporting investigations of long-term Pan-Arctic freshwater variability.

Data availability

GRACE (-FO) mascon products (https://www2.csr.utexas.edu/grace/RL06_mascons.html; https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/; <https://earth.gsfc.nasa.gov/geo/data/grace-mascons>); In-situ runoff data In-situ runoff data (<https://grdc.bafg.de/>; <https://arcticgreatrivers.org/>; <https://www.r-arcticnet.sr.unh.edu/v4.0/index.html>); Dam and reservoir information (Lehner et al., 2024); GRACE-like TWS reconstruction products (Gentner et al., 2026; Hacker et al., 2026; Humphrey and Gudmundsson, 2019; Li et al., 2021; Yin et al., 2023). The sources of the forcing data sets used in this research are listed in Text S1 of the Supplementary Information.

Author contributions

Bingshi Liu: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing (original draft preparation); Writing (review and editing). **Xiancai Zou:** Conceptualization; Funding acquisition; Project administration; Resources; Supervision; Writing (review and editing). **Jiancheng Li:** Funding acquisition; Project administration; Resources; Supervision; Writing (review and editing). **Taoyong Jin:** Writing (review and editing).



Conflict of Interest

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

620 Acknowledgements

The authors gratefully appreciate these publicly accessible data sets used in this research.

Financial support

This work was supported by the National Natural Science Foundation of China [grant numbers 42304002, 42388102, 42192532]; the China Postdoctoral Science Foundation (CPSF) [grant numbers 2024T170678, 2023M732685]; the Postdoctoral Fellowship
625 Program of CPSF [grant number GZB20230536].

References

- Andreadis, K. M., Coss, S. P., Durand, M., Gleason, C. J., Simmons, T. T., Tebaldi, N., Bjerklie, D. M., Brinkerhoff, C., Dudley, R. W., Gejadze, I., Larnier, K., Malaterre, P., Oubanas, H., Allen, G. H., Bates, P. D., David, C. H., Domeneghetti, A., Elmi, O., Marc, L. F., Prata De Moraes Frasson, R., Friedmann, E., Garambois, P., Gehring, J., Getirana, A., Hughes, M., Lee, J., Matte, P., Minear, J. T., Monnier, J., Muhebwa, A., Tourian, M. J., Pavelsky, T. M., Riggs, R. M., Rodríguez, E., Sikder, M. S., Smith, L. C., Stuurman, C., Taneja, J., Tarpanelli, A., Wang, J., Williams, B. A., and Yadav, B.: A First Look at River Discharge Estimation From SWOT Satellite Observations, *Geophys. Res. Lett.*, 52, e2024GL114185, <https://doi.org/10.1029/2024GL114185>, 2025.
- 630
- Bintanja, R. and Andry, O.: Towards a rain-dominated Arctic, *Nat. Clim. Change*, 7, 263–267, <https://doi.org/10.1038/nclimate3240>, 2017.
- 635
- Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, *Nat. Commun.*, 10, 264, <https://doi.org/10.1038/s41467-018-08240-4>, 2019.
- 640
- Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F.-J. W., Brown, R., Bhatt, U. S., Euskirchen, E. S., Romanovsky, V. E., Walsh, J. E., Overland, J. E., Wang, M., Corell, R. W., Meier, W. N., Wouters, B., Mernild, S., Mård, J., Pawlak, J., and Olsen, M. S.: Key indicators of Arctic climate change: 1971–2017, *Environ. Res. Lett.*, 14, 045010, <https://doi.org/10.1088/1748-9326/aafc1b>, 2019.
- 645
- Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mård, J., Mernild, S. H., Prowse, T., Semenova, O., Stuefer, S. L., and Woo, M. -K.: Arctic terrestrial hydrology: A synthesis of processes, regional effects, and research challenges, *J. Geophys. Res. Biogeosciences*, 121, 621–649, <https://doi.org/10.1002/2015JG003131>, 2016.
- 650
- Callaghan, T. V., Johansson, M., Brown, R. D., Groisman, P. Ya., Labba, N., Radionov, V., Barry, R. G., Bulygina, O. N., Essery, R. L. H., Frolov, D. M., Golubev, V. N., Grenfell, T. C., Petrushina, M. N., Razuvaev, V. N., Robinson, D. A., Romanov, P., Shindell, D., Shmakin, A. B., Sokratov, S. A., Warren, S., and Yang, D.: The Changing Face of Arctic Snow Cover: A Synthesis of Observed and Projected Changes, *AMBIO*, 40, 17–31, <https://doi.org/10.1007/s13280-011-0212-y>, 2011.
- 655
- Chai, Y., Miao, C., Gentine, P., Mudryk, L., Thackeray, C. W., Berghuijs, W. R., Wu, Y., Fan, X., Slater, L., Sun, Q., and Zwiers, F.: Constrained Earth system models show a stronger reduction in future Northern Hemisphere snowmelt water, *Nat. Clim. Change*, 15, 514–520, <https://doi.org/10.1038/s41558-025-02308-y>, 2025.



- Chen, Y., Cheng, X., Liu, A., Chen, Q., and Wang, C.: Tracking lake drainage events and drained lake basin vegetation dynamics across the Arctic, *Nat. Commun.*, 14, 7359, <https://doi.org/10.1038/s41467-023-43207-0>, 2023.
- 660 Chin, K. S., Lento, J., Culp, J. M., Lacelle, D., and Kokelj, S. V.: Permafrost thaw and intense thermokarst activity decreases abundance of stream benthic macroinvertebrates, *Glob. Change Biol.*, 22, 2715–2728, <https://doi.org/10.1111/gcb.13225>, 2016.
- Davidson, N. C.: Yenisei River Basin and Lake Baikal (Russia), in: *The Wetland Book*, edited by: Finlayson, C. M., Milton, G. R., Prentice, R. C., and Davidson, N. C., Springer Netherlands, Dordrecht, 1–8, https://doi.org/10.1007/978-94-007-6173-5_276-1, 2016.
- 665 Derksen, C. and Brown, R.: Spring snow cover extent reductions in the 2008–2012 period exceeding climate model projections, *Geophys. Res. Lett.*, 39, 2012GL053387, <https://doi.org/10.1029/2012GL053387>, 2012.
- Feng, D., Gleason, C. J., Lin, P., Yang, X., Pan, M., and Ishitsuka, Y.: Recent changes to Arctic river discharge, *Nat. Commun.*, 12, 6917, <https://doi.org/10.1038/s41467-021-27228-1>, 2021.
- 670 Gentner, L. Q., Gou, J., Tourian, M. J., Börger, L., Sneeuw, N., and Soja, B.: DeepRec: Global Terrestrial Water Storage Reconstruction Since 1941 Using Spatiotemporal-Aware Deep Learning Model, *J. Geophys. Res. Mach. Learn. Comput.*, 3, e2025JH000889, <https://doi.org/10.1029/2025JH000889>, 2026.
- Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, *Earth Syst. Sci. Data*, 11, 1655–1674, <https://doi.org/10.5194/essd-11-1655-2019>, 2019.
- Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUNversion 1.1), *Earth Syst. Sci. Data*, 8, 279–295, <https://doi.org/10.5194/essd-8-279-2016>, 2016.
- 675 Gudmundsson, L., Brunner, M. I., Döll, P., Fluet-Chouinard, E., Frolova, N., Gosling, S. N., Hirabayashi, Y., Kireeva, M. B., Liu, X., Müller Schmied, H., Magritskiy, D., Slater, L. J., Stein, L., Trambly, Y., Wang, K., Wasko, C., Yamazaki, D., and Zhou, X.: Past and future change in global river flows, *Nat. Rev. Earth Environ.*, 7, 7–23, <https://doi.org/10.1038/s43017-025-00745-z>, 2025.
- 680 Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K., and Merz, B.: A global analysis of temporal and spatial variations in continental water storage, *Water Resour. Res.*, 43, 2006WR005247, <https://doi.org/10.1029/2006WR005247>, 2007.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, 377, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>, 2009.
- 685 Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: Efficient adaptive MCMC, *Stat. Comput.*, 16, 339–354, <https://doi.org/10.1007/s11222-006-9438-0>, 2006.
- Hacker, C., Gutknecht, B. D., Löffler, A., and Kusche, J.: Multidecadal reconstruction of terrestrial water storage changes by combining pre-GRACE satellite observations and climate data, *Earth Syst. Sci. Data*, 18, 1747–1781, <https://doi.org/10.5194/essd-18-1747-2026>, 2026.
- 690 Hamm, A., Magnússon, R. Í., Khattak, A. J., and Frampton, A.: Continentality determines warming or cooling impact of heavy rainfall events on permafrost, *Nat. Commun.*, 14, 3578, <https://doi.org/10.1038/s41467-023-39325-4>, 2023.
- Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, *Sci. Data*, 7, 109, <https://doi.org/10.1038/s41597-020-0453-3>, 2020.
- 695 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, 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., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, *Q. J. R. Meteorol. Soc.*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.



- 700 Humphrey, V. and Gudmundsson, L.: GRACE-REC: a reconstruction of climate-driven water storage changes over the last century, *Earth Syst. Sci. Data*, 11, 1153–1170, <https://doi.org/10.5194/essd-11-1153-2019>, 2019.
- Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: A global reconstruction of climate-driven subdecadal water storage variability, *Geophys. Res. Lett.*, 44, 2300–2309, <https://doi.org/10.1002/2017GL072564>, 2017.
- Jennings, K. S., Winchell, T. S., Livneh, B., and Molotch, N. P.: Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere, *Nat. Commun.*, 9, 1148, <https://doi.org/10.1038/s41467-018-03629-7>, 2018.
- 705 Jones, B. M., Grosse, G., Farquharson, L. M., Roy-Léveillé, P., Veremeeva, A., Kanevskiy, M. Z., Gaglioti, B. V., Breen, A. L., Parsekian, A. D., Ulrich, M., and Hinkel, K. M.: Lake and drained lake basin systems in lowland permafrost regions, *Nat. Rev. Earth Environ.*, 3, 85–98, <https://doi.org/10.1038/s43017-021-00238-9>, 2022.
- Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M., Goto, N., Chiba, J., Miyaoka, K., Sekiguchi, R., Deushi, M., Kamahori, H., Nakaegawa, T., Tanaka, T. Y., Tokuhiro, T., Sato, Y., Matsushita, Y., and Onogi, K.: The JRA-3Q Reanalysis, *J. Meteorol. Soc. Jpn. Ser II*, 102, 49–109, <https://doi.org/10.2151/jmsj.2024-004>, 2024.
- 710 Landerer, F. W., Dickey, J. O., and Güntner, A.: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009, *J. Geophys. Res. Atmos.*, 115, 2010JD014584, <https://doi.org/10.1029/2010JD014584>, 2010.
- Lehner, B., Beames, P., Mulligan, M., Zarfl, C., De Felice, L., Van Soesbergen, A., Thieme, M., Garcia De Leaniz, C., Anand, M., Belletti, B., Brauman, K. A., Januchowski-Hartley, S. R., Lyon, K., Mandle, L., Mazany-Wright, N., Messenger, M. L., 715 Pavelsky, T., Pekel, J.-F., Wang, J., Wen, Q., Wishart, M., Xing, T., Yang, X., and Higgins, J.: The Global Dam Watch database of river barrier and reservoir information for large-scale applications, *Sci. Data*, 11, 1069, <https://doi.org/10.1038/s41597-024-03752-9>, 2024.
- Leopardi, F., Dari, J., Saltalippi, C., Brocca, L., Volden, E., Fernández-Prieto, D., Saemian, P., Sneeuw, N., Tourian, M. J., and Camici, S.: A satellite-based approach for estimating runoff and river discharge in the Pan-Arctic region from 2003 to 2022, 720 *Remote Sens. Environ.*, 338, 115353, <https://doi.org/10.1016/j.rse.2026.115353>, 2026.
- Li, F., Kusche, J., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., and Lück, C.: Comparison of Data-Driven Techniques to Reconstruct (1992–2002) and Predict (2017–2018) GRACE-Like Gridded Total Water Storage Changes Using Climate Inputs, *Water Resour. Res.*, 56, e2019WR026551, <https://doi.org/10.1029/2019WR026551>, 2020.
- 725 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, *Geophys. Res. Lett.*, 48, e2021GL093492, <https://doi.org/10.1029/2021GL093492>, 2021.
- Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., and Wood, E. F.: Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches, *Water Resour. Res.*, 55, 6499–6516, <https://doi.org/10.1029/2019WR025287>, 2019.
- 730 Liu, B., Zou, X., Yi, S., Sneeuw, N., Cai, J., and Li, J.: Identifying and separating climate- and human-driven water storage anomalies using GRACE satellite data, *Remote Sens. Environ.*, 263, 112559, <https://doi.org/10.1016/j.rse.2021.112559>, 2021.
- Liu, B., Zou, X., Yi, S., Sneeuw, N., Li, J., and Cai, J.: Reconstructing GRACE-like time series of high mountain glacier mass anomalies, *Remote Sens. Environ.*, 280, 113177, <https://doi.org/10.1016/j.rse.2022.113177>, 2022.
- Loomis, B. D., Luthcke, S. B., and Sabaka, T. J.: Regularization and error characterization of GRACE mascons, *J. Geod.*, 93, 1381–1398, <https://doi.org/10.1007/s00190-019-01252-y>, 2019.
- 735 Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J., and Camp, J.: Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution, *J. Glaciol.*, 59, 613–631, <https://doi.org/10.3189/2013JoG12J147>, 2013.
- 740 Makarieva, O., Nesterova, N., Post, D. A., Sherstyukov, A., and Lebedeva, L.: Warming temperatures are impacting the hydrometeorological regime of Russian rivers in the zone of continuous permafrost, *The Cryosphere*, 13, 1635–1659, <https://doi.org/10.5194/tc-13-1635-2019>, 2019.



- Miner, K. R., Turetsky, M. R., Malina, E., Bartsch, A., Tamminen, J., McGuire, A. D., Fix, A., Sweeney, C., Elder, C. D., and Miller, C. E.: Permafrost carbon emissions in a changing Arctic, *Nat. Rev. Earth Environ.*, 3, 55–67, <https://doi.org/10.1038/s43017-021-00230-3>, 2022.
- 745 Mudryk, L., Mortimer, C., Derksen, C., Elias Chereque, A., and Kushner, P.: Benchmarking of snow water equivalent (SWE) products based on outcomes of the SnowPEX+ Intercomparison Project, *The Cryosphere*, 19, 201–218, <https://doi.org/10.5194/tc-19-201-2025>, 2025.
- Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., Kynast, E., Peiris, T. A., Schiebener, L., Schumacher, M., and Döll, P.: The global water resources and use model WaterGAP v2.2e: description and evaluation of
750 modifications and new features, *Geosci. Model Dev.*, 17, 8817–8852, <https://doi.org/10.5194/gmd-17-8817-2024>, 2024.
- 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.
- 755 Nitzbon, J., Schneider Von Deimling, T., Aliyeva, M., Chadburn, S. E., Grosse, G., Laboor, S., Lee, H., Lohmann, G., Steinert, N. J., Stuenzi, S. M., Werner, M., Westermann, S., and Langer, M.: No respite from permafrost-thaw impacts in the absence of a global tipping point, *Nat. Clim. Change*, 14, 573–585, <https://doi.org/10.1038/s41558-024-02011-4>, 2024.
- Peterson, B. J., Holmes, R. M., McClelland, J. W., Vörösmarty, C. J., Lammers, R. B., Shiklomanov, A. I., Shiklomanov, I. A., and Rahmstorf, S.: Increasing River Discharge to the Arctic Ocean, *Science*, 298, 2171–2173, <https://doi.org/10.1126/science.1077445>, 2002.
- 760 Qing, Y., Wang, S., AghaKouchak, A., and Gentine, P.: Delayed formation of Arctic snow cover in response to wildland fires in a warming climate, *Nat. Clim. Change*, 15, 1091–1098, <https://doi.org/10.1038/s41558-025-02443-6>, 2025.
- Ramos Buarque, S., Decharme, B., Barbu, A. L., and Franchisteguy, L.: Insights into the North Hemisphere daily snowpack at high resolution from the new Crocus–ERA5 product, *Earth Syst. Sci. Data*, 17, 7227–7249, <https://doi.org/10.5194/essd-17-7227-2025>, 2025.
- 765 RGI Consortium: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 7, <https://doi.org/10.5067/F6JMOVY5NAVZ>, 2023.
- Rodell, M. and Reager, J. T.: Water cycle science enabled by the GRACE and GRACE-FO satellite missions, *Nat. Water*, 1, 47–59, <https://doi.org/10.1038/s44221-022-00005-0>, 2023.
- 770 Rodell, M., Houser, P. R., Jambor, U., Gottschalk, 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. Am. Meteorol. Soc.*, 85, 381–394, <https://doi.org/10.1175/BAMS-85-3-381>, 2004.
- Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., and Lo, M.-H.: Emerging trends in global freshwater availability, *Nature*, 557, 651–659, <https://doi.org/10.1038/s41586-018-0123-1>, 2018.
- 775 Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, *Earth Syst. Sci. Data*, 12, 3469–3479, <https://doi.org/10.5194/essd-12-3469-2020>, 2020.
- Save, H., Bettadpur, S., and Tapley, B. D.: High-resolution CSR GRACE RL05 mascons, *J. Geophys. Res. Solid Earth*, 121, 7547–7569, <https://doi.org/10.1002/2016JB013007>, 2016.
- 780 Schirmeister, Z., Finger, P., Rustemeier, E., and Ziese, M.: GPCC Precipitation Analysis Monthly Version 2025 (at 0.25°, 0.5°, 1.0°, 2.5°) (2025), https://doi.org/10.5676/DWD_GPCC/MONTHLY_V2025_050, 2025.
- Smith, S. L., O'Neill, H. B., Isaksen, K., Noetzli, J., and Romanovsky, V. E.: The changing thermal state of permafrost, *Nat. Rev. Earth Environ.*, 3, 10–23, <https://doi.org/10.1038/s43017-021-00240-1>, 2022.
- Sun, Z., Long, D., Yang, W., Li, X., and Pan, Y.: Reconstruction of GRACE Data on Changes in Total Water Storage Over the Global Land Surface and 60 Basins, *Water Resour. Res.*, 56, e2019WR026250, <https://doi.org/10.1029/2019WR026250>, 2020.



- 785 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, *Nat. Clim. Change*, 9, 358–369, <https://doi.org/10.1038/s41558-019-0456-2>, 2019.
- 790 The GLaMBIE Team, Zemp, M., Jakob, L., Dussailant, I., Nussbaumer, S. U., Gourmelen, N., Dubber, S., A, G., Abdullahi, S., Andreassen, L. M., Berthier, E., Bhattacharya, A., Blazquez, A., Boehm Vock, L. F., Bolch, T., Box, J., Braun, M. H., Brun, F., Cicero, E., Colgan, W., Eckert, N., Farinotti, D., Florentine, C., Floricioiu, D., Gardner, A., Harig, C., Hassan, J., Hugonnet, R., Huss, M., Jóhannesson, T., Liang, C.-C. A., Ke, C.-Q., Khan, S. A., King, O., Kneib, M., Krieger, L., Maussion, F., Mattea, E., McNabb, R., Menounos, B., Miles, E., Moholdt, G., Nilsson, J., Pálsson, F., Pfeffer, J., Piermattei, L., Plummer, S., Richter, A., Sasgen, I., Schuster, L., Seehaus, T., Shen, X., Sommer, C., Sutterley, T., Treichler, D., Velicogna, I., Wouters, B., Zekollari, H., and Zheng, W.: Community estimate of global glacier mass changes from 2000 to 2023, *Nature*, 639, 382–388, <https://doi.org/10.1038/s41586-024-08545-z>, 2025.
- 795 Trautmann, T., Koirala, S., Carvalhais, N., Eicker, A., Fink, M., Niemann, C., and Jung, M.: Understanding terrestrial water storage variations in northern latitudes across scales, *Hydrol. Earth Syst. Sci.*, 22, 4061–4082, <https://doi.org/10.5194/hess-22-4061-2018>, 2018.
- 800 Watkins, M. M., Wiese, D. N., Yuan, D., Boening, C., and Landerer, F. W.: Improved methods for observing Earth’s time variable mass distribution with GRACE using spherical cap mascons, *J. Geophys. Res. Solid Earth*, 120, 2648–2671, <https://doi.org/10.1002/2014JB011547>, 2015.
- Webb, E. E. and Liljedahl, A. K.: Diminishing lake area across the northern permafrost zone, *Nat. Geosci.*, 16, 202–209, <https://doi.org/10.1038/s41561-023-01128-z>, 2023.
- 805 Webb, E. E., Liljedahl, A. K., Cordeiro, J. A., Loranty, M. M., Witharana, C., and Lichstein, J. W.: Permafrost thaw drives surface water decline across lake-rich regions of the Arctic, *Nat. Clim. Change*, 12, 841–846, <https://doi.org/10.1038/s41558-022-01455-w>, 2022.
- Wiese, D. N., Landerer, F. W., and Watkins, M. M.: Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution, *Water Resour. Res.*, 52, 7490–7502, <https://doi.org/10.1002/2016WR019344>, 2016.
- 810 Willmott, C. J. and Matsuura, K.: Smart Interpolation of Annually Averaged Air Temperature in the United States, *J. Appl. Meteorol.*, 34, 2577–2586, [https://doi.org/10.1175/1520-0450\(1995\)034%3C2577:SIOAAA%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034%3C2577:SIOAAA%3E2.0.CO;2), 1995.
- Xie, J., Xu, Y.-P., Booij, M. J., and Guo, Y.: Influences of reservoir operation on terrestrial water storage changes detected by GRACE in the Yellow River basin, *J. Hydrol.*, 610, 127924, <https://doi.org/10.1016/j.jhydrol.2022.127924>, 2022.
- 815 Xie, P. and Yi, S.: Reconstruction of climate-driven global terrestrial water storage variations (2002–2021) using a four-parameter linear recursive model, *Hydrol. Earth Syst. Sci.*, 30, 3221–3244, <https://doi.org/10.5194/hess-30-3221-2026>, 2026.
- Yang, D., Ye, B., and Kane, D.: Streamflow changes over Siberian Yenisei River Basin, *J. Hydrol.*, 296, 59–80, <https://doi.org/10.1016/j.jhydrol.2004.03.017>, 2004.
- Yi, S., Saemian, P., Sneeuw, N., and Tourian, M. J.: Estimating runoff from pan-Arctic drainage basins for 2002–2019 using an improved runoff-storage relationship, *Remote Sens. Environ.*, 298, 113816, <https://doi.org/10.1016/j.rse.2023.113816>, 2023.
- 820 Yin, J., Slater, L. J., Khouakhi, A., Yu, L., Liu, P., Li, F., Pokhrel, Y., and Gentine, P.: GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from 1940 to present, *Earth Syst. Sci. Data*, 15, 5597–5615, <https://doi.org/10.5194/essd-15-5597-2023>, 2023.
- 825 Zhang, Y., Blöschl, G., Wei, H., Kong, D., Ma, N., Wagener, T., Tian, J., Xia, J., Li, C., Wang, L., Chiew, F. H. S., Leung, L. R., Liu, X., Zheng, H., Zhang, X., and Liu, C.: Overestimation of past and future increases in global river flow by Earth system models, *Nat. Geosci.*, <https://doi.org/10.1038/s41561-025-01897-9>, 2026.
- Zhou, W., Leung, L. R., and Lu, J.: Steady threefold Arctic amplification of externally forced warming masked by natural variability, *Nat. Geosci.*, 17, 508–515, <https://doi.org/10.1038/s41561-024-01441-1>, 2024.