Near-real-time monitoring of global terrestrial water storage anomalies and hydrological droughts

Mo, Shaoxing; Schumacher, Maike; van Dijk, Albert IJM; Shi, Xiaoqing; Wu, Jichun; Forootan, Ehsan

Published in:
Geophysical Research Letters

DOI (link to publication from Publisher):
10.1029/2024GL112677

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

AGU ADVANCING EARTH AND SPACE SCIENCES

Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL112677

Near-Real-Time Monitoring of Global Terrestrial Water Storage Anomalies and Hydrological Droughts

Shaoxing Mo[1,2], Maike Schumacher[2], Albert I. J. M. van Dijk[3], Xiaoqing Shi[1], Jichun Wu[1], and Ehsan Forootan[2]

[1]Key Laboratory of Surficial Geochemistry of Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China, [2]Department of Sustainability and Planning, Geodesy Group, Aalborg University, Aalborg, Denmark, [3]Fenner School of Environment & Society, Australian National University, Canberra, ACT, Australia

Key Points:

- Deep Learning (DL) is proposed to mitigate the latency of Gravity Recovery and Climate Experiment/Follow-On (GRACE)/FO terrestrial water storage anomaly (TWSA) products for operational use
- DL enables near-real-time global TWSA and hydrological drought characterization up to 3 months before GRACE/FO data availability
- Timely GRACE/FO data enables rapid identification and response to emerging water challenges

Supporting Information:

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

Correspondence to:

X. Shi and J. Wu,
shixq@nju.edu.cn;
jcwu@nju.edu.cn

Citation:

Mo, S., Schumacher, M., van Dijk, A. I. J. M., Shi, X., Wu, J., & Forootan, E. (2025). Near-real-time monitoring of global terrestrial water storage anomalies and hydrological droughts. *Geophysical Research Letters*, *52*, e2024GL112677. https://doi.org/10.1029/2024GL112677

Received 20 SEP 2024
Accepted 18 MAR 2025

Author Contributions:

Conceptualization: Shaoxing Mo, Maike Schumacher, Albert I. J. M. van Dijk, Ehsan Forootan
Data curation: Shaoxing Mo, Ehsan Forootan
Formal analysis: Shaoxing Mo
Investigation: Shaoxing Mo
Methodology: Shaoxing Mo
Project administration: Xiaoqing Shi, Jichun Wu
Resources: Xiaoqing Shi, Jichun Wu, Ehsan Forootan
Software: Shaoxing Mo

Abstract Global terrestrial water storage anomaly (TWSA) products from the Gravity Recovery and Climate Experiment (GRACE) and its Follow-On mission (GRACE/FO) have an approximately three-month latency, significantly limiting their operational use in water management and drought monitoring. To address this challenge, we develop a Bayesian convolutional neural network (BCNN) to predict TWSA fields with uncertainty estimates during the latency period. The results demonstrate that BCNN provides near-real-time TWSA estimates that closely match GRACE/FO observations, with median correlation coefficients of 0.92–0.95, Nash-Sutcliffe efficiencies of 0.81–0.89, and root mean squared errors of 1.79–2.26 cm for one- to three-month ahead predictions. More importantly, the model advances global hydrological drought monitoring by enabling detection up to three months before GRACE/FO data availability, with median characterization mismatches below 16.4%. This breakthrough in early warning capability addresses a fundamental constraint in satellite-based hydrological monitoring and offers water resource managers critical lead time to implement drought mitigation strategies.

Plain Language Summary Monitoring Earth's water resources from space is essential for understanding droughts and managing water supplies. The Gravity Recovery and Climate Experiment (GRACE) satellite and its Follow-On (GRACE/FO) track changes in terrestrial water storage by detecting tiny variations in Earth's gravity, providing crucial information for water management. However, there's typically a three-month delay before this satellite data becomes available due to extensive processing requirements, making it difficult to respond quickly to developing water challenges. Our study develops a deep learning approach that predicts water storage changes during this latency period. By integrating historical GRACE/FO observations with current hydrometeorological data, our method generates reliable estimates of water storage variations in near-real-time and detects hydrological droughts up to three months earlier than conventional GRACE/FO observations. This advancement helps bridge a long-standing gap in Earth observation, allowing for more timely responses to emerging water challenges. As climate change increases the severity and frequency of extreme hydrological events, our method offers a valuable tool for improving drought preparedness, water management, and disaster mitigation strategies.

1. Introduction

Terrestrial water storage anomalies (TWSAs), which integrate changes in surface water, snow and ice, canopy water, soil moisture, and groundwater, offer direct observations of total (land) water availability (Adusumilli et al., 2019; B. Li et al., 2019). This variable is now designated as an essential climate variable (Zemp et al., 2022), and its observation, globally, with known uncertainty and low latency is desirable for effective water resource management and drought mitigation (Mo, Zhong, Forootan, Shi, et al., 2022; Thomas et al., 2014). The Gravity Recovery and Climate Experiment (GRACE) mission and its successor, GRACE Follow-On (collectively GRACE/FO), provide an unprecedented way to monitor the time-variable Earth's gravity fields, from which the global-scale TWSAs can be inferred (Hu et al., 2024; Long et al., 2017; Rodell & Reager, 2023; Wahr et al., 1998). These TWSA measurements have been widely used as an indicator of hydrological droughts and flood potentials (e.g., Boergens et al., 2020; Forootan et al., 2019, 2024; Mo, Zhong, Forootan, Shi, et al., 2022; Reager et al., 2014; Rodell & Li, 2023; Thomas et al., 2014; Zhao et al., 2017). Furthermore, the assimilation of GRACE/FO TWSA data into land surface models has significantly improved their accuracy and reliability (e.g., B. Li et al., 2019; Houborg et al., 2012; Schumacher et al., 2018; van Dijk et al., 2014; Yang et al., 2024).

Validation: Shaoxing Mo, Maike Schumacher, Albert I. J. M. van Dijk, Ehsan Forootan
Visualization: Shaoxing Mo
Writing – original draft: Shaoxing Mo
Writing – review & editing: Shaoxing Mo, Maike Schumacher, Albert I. J. M. van Dijk, Xiaoqing Shi, Jichun Wu, Ehsan Forootan

A critical limitation in the operational use of GRACE/FO data is the approximately three-month latency between observation and data availability. This delay arises from the need for extensive processing, quality control, cross-product comparison, and rigorous validation to ensure the consistency and reliability of the final gravity solutions. Such latency is typical for the three major GRACE/FO Mascon products provided by the Jet Propulsion Laboratory (JPL) (Watkins et al., 2015), the Center for Space Research (Save et al., 2016), and the Goddard Space Flight Center (Loomis et al., 2019), which are widely used in hydrological applications. The extended latency period severely impedes real-time water resource monitoring and hampers timely assessment of hydrological extremes. For instance, existing GRACE/FO-based drought indices (Forootan et al., 2019; Thomas et al., 2014; Zhao et al., 2017) can only characterize historical events, limiting their utility for active drought monitoring and response.

Data-model fusion techniques have demonstrated success in utilizing GRACE/FO data, yet the latency period poses a persistent challenge for operational applications. For instance, assimilating GRACE/FO data into NASA's catchment land surface model enhances estimates of individual water storage components, such as groundwater and soil moisture at different depths (B. Li et al., 2019), enabling more precise monitoring of specific hydrological conditions. However, during the latency period, models must operate without data assimilation (B. Li et al., 2019; Houborg et al., 2012; Yang et al., 2024), leading to increased uncertainties and reduced reliability that compromise decision-making for water resource management and drought mitigation. While recent machine learning approaches have attempted to bridge this gap (Lu et al., 2024; F. Li et al., 2024), they focused primarily on basin-scale predictions and lack crucial uncertainty quantification.

To address these limitations, we present an integrated deep learning (DL) approach using a Bayesian convolutional neural network (BCNN) to predict grid-level GRACE/FO TWSA data during the latency period. Our method combines past GRACE/FO observations with hydrometeorological inputs that have substantially shorter delays (typically less than one month), enabling near-real-time TWSA monitoring. The BCNN framework integrates convolutional neural networks (Gu et al., 2018; Ronneberger et al., 2015), which are proven effective for spatial pattern extraction, (Fu et al., 2019; He et al., 2016; G. Huang et al., 2017; Mo et al., 2019; Z. Zhang et al., 2021), with series stationarization (Y. Liu et al., 2022) to mitigate the inherent nonstationarity issue in hydrological time series. Furthermore, we employ a Bayesian inference algorithm known as Stein variational gradient descent (SVGD) (Q. Liu & Wang, 2016) for model training to provide uncertainty-quantified predictions. This uncertainty quantification capability, absent in most previous prediction attempts, is crucial for both water resource monitoring and GRACE/FO data assimilation applications. Through extensive validation, we demonstrate that our method effectively bridges the GRACE/FO data latency gap and enables timely detection and characterization of global hydrological droughts.

2. Data and Methods

2.1. GRACE/FO and Hydrometeorological Data

We focus on the global land area that excludes Antarctica and Greenland (60°S–84°N and 180°W–180°E). In this study, the monthly JPL GRACE/FO Mascon product (Watkins et al., 2015) is used. JPL's processing of the GRACE/FO data includes various corrections, such as the corrections for glacial isostatic adjustment and the application of destriping and filtering techniques to minimize correlated errors and noise. These corrections improve the accuracy of the data, increasing its reliability for hydrological and climatic applications (Watkins et al., 2015). The monthly TWSA fields, representing anomalies relative to the averaging period of 2004–2009, reflect variations in the water stored on land. Although the data are provided at a spatial resolution of $0.5° \times 0.5°$, their effective resolution is approximately $3° \times 3°$. For this study, we resample the data in grids of $1° \times 1°$, since this resolution is commonly used in hydrological applications.

The precipitation (P), temperature (T), and reanalyzed/simulated TWSA (rTWSA), which are generally well correlated with variations in terrestrial water storage (Mo, Zhong, Forootan, Shi, et al., 2022; Sun et al., 2019), are used as auxiliary inputs to the BCNN model for TWSA prediction. Although evapotranspiration and runoff also influence TWSAs, they are excluded because temperature serves as a proxy for evapotranspiration (Almorox et al., 2015), and our preliminary tests revealed that runoff has a weaker correlation with TWSA compared to precipitation and temperature. Here, the ERA5-Land reanalysis dataset is considered for its good performance (Muñoz Sabater et al., 2021). To ensure spatial consistency, these data sets are also resampled to the same $1° \times 1°$ grids as those of GRACE/FO TWSA.

2.2. Bayesian Convolutional Neural Network

The BCNN model for predicting gridded TWSA fields consists of three key components: (a) a convolutional neural network (Gu et al., 2018) to effectively capture the spatial dependencies in GRACE/FO TWSA and hydrometeorological data fields; (b) a series stationarization strategy (Y. Liu et al., 2022) to mitigate the nonstationary property of time series (e.g., time-varying mean and variance), thereby enhancing the accuracy of future predictions; and finally (c) an uncertainty assessment of the TWSA predictions.

To implement the BCNN model, we reorganize each TWSA field as $\text{TWSA} \in \mathbb{R}^{H \times W}$, where $H \times W$ is the spatial resolution. This allows convolutional neural networks to effectively extract spatial dependencies from the sequence of two-dimensional fields. Specifically, the BCNN model is designed to learn the following input-output mapping:

$$(\mathbf{X}_{t-p+1:t+k}, \mathbf{TWSA}_{t-p+1:t}) \xrightarrow{f_{\boldsymbol{\theta}}(\cdot)} \mathbf{TWSA}_{t+1:t+k}, \tag{1}$$

where t is the month index, and $f_{\boldsymbol{\theta}}(\cdot)$ denotes the neural network with trainable parameters $\boldsymbol{\theta}$. The outputs are the TWSA fields during the k-month latency period (i.e., $\mathbf{TWSA}_{t+1:t+k} \in \mathbb{R}^{k \times H \times W}$), the inputs include the TWSA fields from the past $p = 12$ months (i.e., $\mathbf{TWSA}_{t-p+1:t} \in \mathbb{R}^{p \times H \times W}$) and the auxiliary hydrometeorological data during both the past and latency periods (i.e., $\mathbf{X}_{t-p+1:t+k} \in \mathbb{R}^{(p+k) \times C \times H \times W}$, where C denotes the number of auxiliary predictors). In particular, the hydrometeorological data during the latency period is also used as input because they are typically available.

In series stationarization (Y. Liu et al., 2022), the input sequences ($\mathbf{TWSA}_{t-p+1:t}$ and $\mathbf{X}_{t-p+1:t+k}$) are stationarized by subtracting the local mean and dividing by the local standard deviation computed within a sliding window. The BCNN predictions are then transformed back to their original scale by applying the inverse operation using the mean and standard deviation from the corresponding input TWSA window. This transformation normalizes the scale and ensures approximate stationarity, enabling the model to capture the underlying patterns in the data more effectively. Additionally, the BCNN model quantifies both epistemic and aleatoric uncertainties in its predictions by treating the trainable parameters $\boldsymbol{\theta}$ as random variables. To infer their posterior distribution, the SVGD algorithm (Q. Liu & Wang, 2016) is employed to approximate the distribution using a set of parameter particles: $\{\boldsymbol{\theta}_i\}_{i=1}^{n_s}$. Given an arbitrary input to BCNN, it will produce n_s TWSA predictions, from which the mean and standard deviation can be computed to quantify uncertainties. Further details on the BCNN architecture, hyperparameter settings, series stationarization strategy, and SVGD training method are provided in Texts S1–S3 and Figures S1–S3 in Supporting Information S1.

2.3. Training and Testing Data Sets

The GRACE/FO observations from April 2002 to June 2017 and June 2018 to February 2024 are respectively used to organize training and testing data for BCNN, with any one- or two-month data gaps in the GRACE/FO missions being filled using cubic interpolation. For each training/testing sample, the input variables include GRACE/FO TWSA from months $(t - 11)$ to t and auxiliary hydrometeorological inputs from months $(t - 11)$ to $(t + 3)$ (Figure S4 in Supporting Information S1). The targets are the GRACE/FO TWSAs for the next three months, that is, $(t + 1)$ to $(t + 3)$. Therefore, the GRACE/FO TWSA fields from June 2019 to February 2024 are used to evaluate the accuracy of BCNN prediction, using the commonly employed correlation coefficient (R), Nash-Sutcliffe efficiency (NSE), and root mean squared error (RMSE) metrics (Naser & Alavi, 2023).

3. Results

3.1. Global TWSA Dynamics Monitoring

Figure 1 presents the performance metrics (R, NSE, and RMSE) of BCNN in predicting the original GRACE/FO TWSA signals during the three-month latency period, along with their cumulative distributions. For the first month, BCNN obtains median R, NSE, and RMSE values of 0.95, 0.89, and 1.79 cm, respectively (Figures 1a–1c). The cumulative distributions (Figures 1j–1l) show that approximately 80% of regions exhibit $R > 0.9$, $NSE > 0.8$, and $RMSE < 3$ cm. While prediction accuracy gradually decreases over time due to decreasing

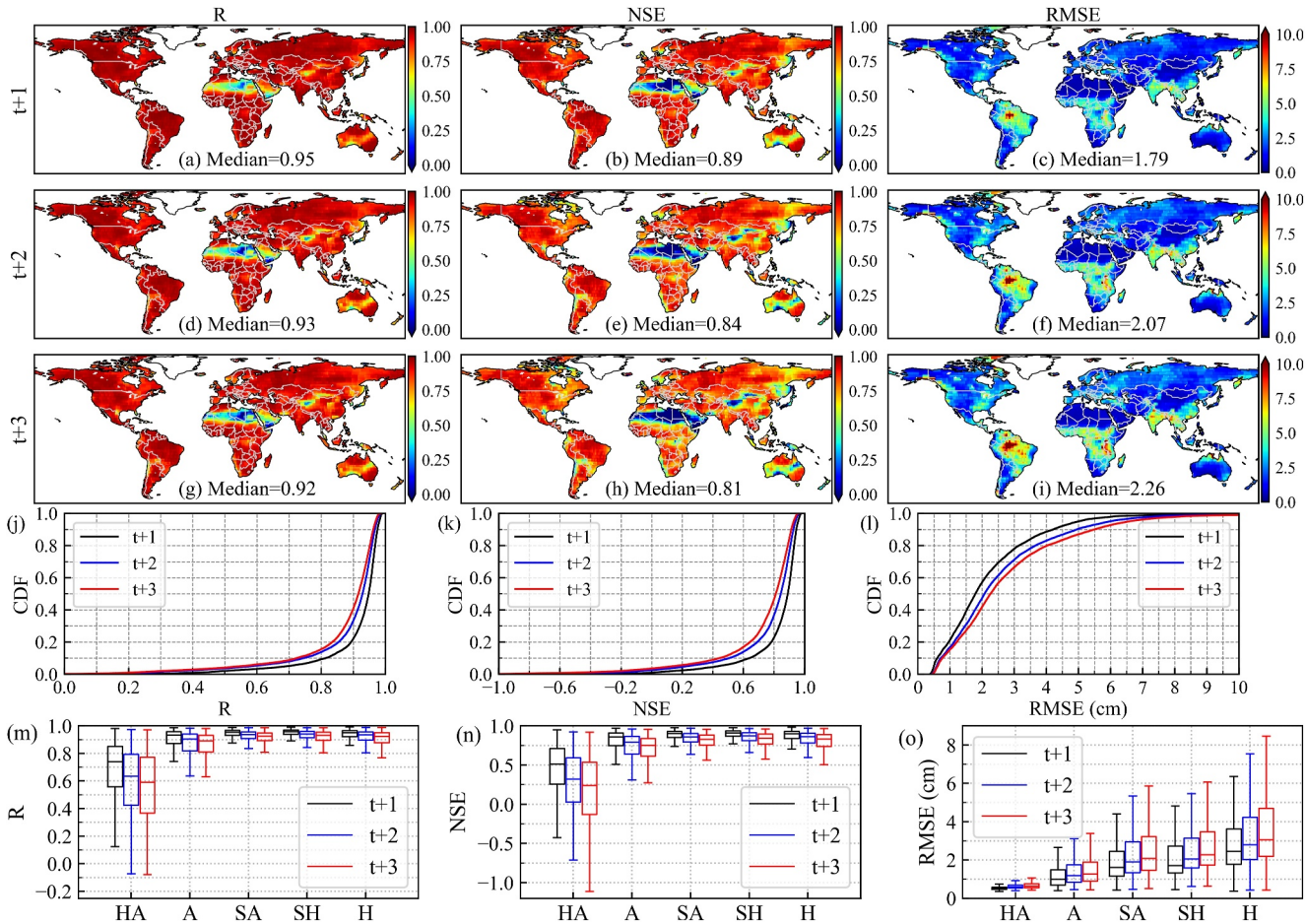

Figure 1. Performance metrics of the Bayesian convolutional neural network (BCNN) model for predicting the original terrestrial water storage anomaly (TWSA) fields derived from the Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On across the testing period (June 2019–February 2024): R (column 1), Nash-Sutcliffe efficiency (NSE) (column 2), and root mean squared error (RMSE) (column 3) values for one-month (row 1), two-month (row 2), and three-month (row 3) ahead predictions. The median of metric values in all grids is shown in each subplot. (j)–(l) Cumulative distributions of R, NSE, and RMSE values. (m)–(o) Boxplots of R, NSE, and RMSE values in hyper-arid (HA), arid (A), semi-arid (SA), semi-humid (SH), and humid (H) regions.

dependency on historical data and increasing system variability (Salinas et al., 2020), BCNN maintains relatively accurate TWSA estimates with median *R* of 0.93 and 0.92, median NSE of 0.84 and 0.81, and median RMSE of 2.07 and 2.26 cm for the second and third months, respectively. Spatial analysis indicates challenges in hyperarid regions, such as the Sahara Desert and the Arabian Peninsula (climate zone distributions shown in Figure S5 in Supporting Information S1). We summarize boxplots of the R, NSE, and RMSE values in different climate zones in Figures 1m–1o. R and NSE values in the hyperarid regions are significantly lower compared to other regions. However, the RMSE values are the lowest in these regions. This can be attributed to the low magnitude and signal-to-noise ratios (Forootan et al., 2012; Humphrey et al., 2016; Mo, Zhong, Forootan, Mehrnegar, et al., 2022), which make TWSA signals particularly difficult to predict. Furthermore, although the RMSE values are the highest in humid regions, the high R and NSE values indicate accurate predictions, as the errors remain small relative to the large magnitudes of TWSA.

To complement the global performance assessment, we evaluate BCNN's predictive capabilities across specific regions and time periods. Figure 2 compares basin-averaged TWSA time series from GRACE/FO and BCNN in six major river basins: the Amazon, Central Europe (merged basin region), Congo, Ganges-Brahmaputra, Mississippi, and Yangtze (locations are shown in Figure S6 in Supporting Information S1). These basins were selected to represent diverse geographical regions and, notably, experienced extreme drought events during the testing period (Singh et al., 2023; L. Zhang et al., 2023), providing robust test cases for model performance under challenging conditions. BCNN accurately captures the temporal dynamics of GRACE/FO TWSA signals in all

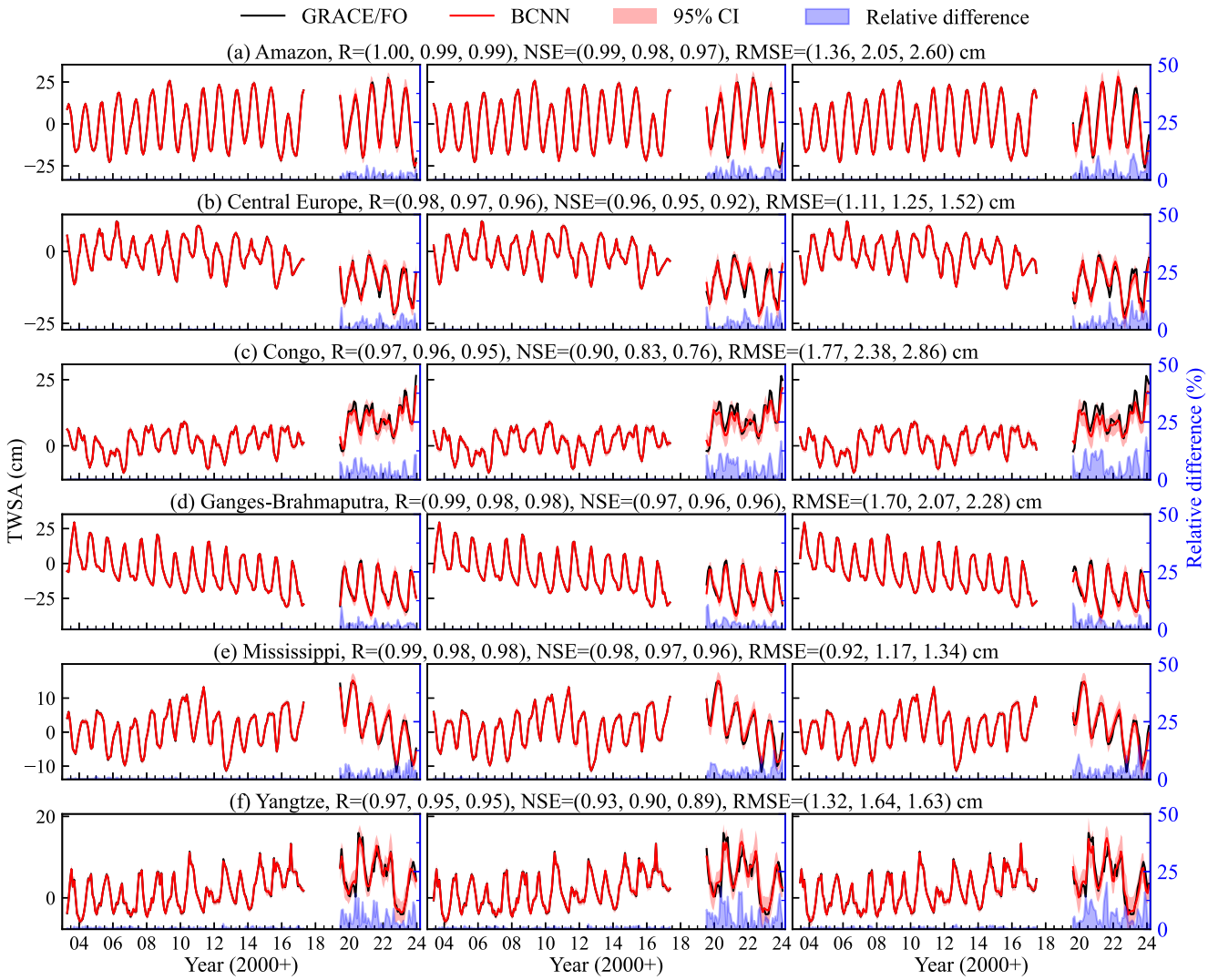

Figure 2. Basin-averaged TWSA time series derived from GRACE/FO observations and BCNN's one- (column 1) to three-month (column 3) ahead predictions for the (a) Amazon, (b) Central Europe, (c) Congo, (d) Ganges-Brahmaputra, (e) Mississippi, and (f) Yangtze River basins. The training and testing samples span the periods of April 2003–June 2017 and June 2019–February 2024, respectively. The red shaded area represents the 95% confidence interval of the predictions. The blue shared area (right *y*-axis) depicts the absolute relative difference, calculated as $\left|\frac{\mathrm{TWSA}-\widehat{\mathrm{TWSA}}}{\mathrm{TWSA}_{\max}-\mathrm{TWSA}_{\min}}\times 100\%\right|$, where $\mathrm{TWSA}_{\max}$ and $\mathrm{TWSA}_{\min}$ are the maximum and minimum GRACE/FO TWSA values during the testing period, respectively. The values in parentheses indicate predictive accuracy for one-, two-, and three-month lead times during the testing period.

basins ($R \geq 0.93$, $NSE \geq 0.76$, $RMSE < 2.28$ cm during the testing period), successfully modeling the nonstationary features including declining trends (Figures 2b and 2d), increasing trends (Figure 2c), and irregular variations (Figures 2e and 2f). As expected, prediction accuracy gradually declines with increasing lead time. Furthermore, GRACE/FO observations generally fall within the BCNN's 95% confidence interval. Among these basins, the prediction accuracy is relatively lower for the Congo and Yangtze River basins, with the relative difference exceeding 10% in some temporal periods. This decreased performance likely reflects increased TWSA variability in these regions driven by climate change and intensive anthropogenic activities (Y. Huang et al., 2015; Tourian et al., 2023). Figures S7 and S8 in Supporting Information S1 present BCNN-predicted TWSA fields for two representative test samples, alongside GRACE/FO observations, prediction errors, and uncertainties (represented by 1.96 standard deviations). BCNN successfully captures the spatial patterns of GRACE/FO fields across all three latency months, with median absolute prediction errors generally below 1.28 cm. Larger errors and uncertainties occur in humid regions such as the Amazon River Basin (>6 cm). This is consistent with their higher TWSA signal magnitudes in these regions (Humphrey et al., 2016).

The effectiveness of BCNN in effectively extracting and integrating informative features from multiple data sources for TWSA prediction is illustrated through sensitivity analysis, as detailed in Text S4 and Figure S9 in Supporting Information S1. This analysis quantifies the relative importance of auxiliary hydrometeorological predictors (P, T, and rTWSA) in improving prediction accuracy. Results underscore the critical importance of integrating physically-based model outputs with meteorological data for GRACE/FO TWSA prediction during the latency period. We further demonstrate the superiority of BCNN predictions by comparing them with the widely used ERA5-Land reanalysis (Muñoz Sabater, 2019) and Noah land surface model (Rodell et al., 2004), both of which provide open-loop TWSA estimates during the latency period. Detailed comparison results presented in Text S5 and Figure S10 in Supporting Information S1 demonstrate BCNN's superior performance.

These results highlight BCNN's potential for accurately filling the GRACE/FO latency period and thus enabling near-real-time monitoring of global TWSA dynamics. In BCNN, a computationally efficient series stationarization strategy (Y. Liu et al., 2022) is employed to alleviate the nonstationarity of TWSA time series. Figure S11 in Supporting Information S1 demonstrates that the exclusion of series stationarization leads to decreased model performance, with median NSE values dropping from (0.89, 0.84, 0.81) to (0.87, 0.81, 0.78) and median RMSE values increasing from (1.79, 2.07, 2.26) cm to (1.89, 2.23, 2.37) cm for one- to three-month ahead predictions, respectively. While our BCNN model effectively predicts TWSAs during the latency period, we recognize that the temporal out-of-sample nature of the prediction task remains a fundamental challenge. The model's reliance on historical data inherently limits its ability to fully account for unexpected changes in future TWSAs driven by extreme climate events or anthropogenic activities, leading to decreased performance as the lead time increases. To mitigate this limitation, future research could explore hybrid approaches integrating state-of-the-art Transformer-based temporal prediction models (Chen et al., 2023), which have shown promise in handling complex temporal dependencies. However, integrating these advanced techniques presents challenges due to increased model complexity and limited GRACE/FO data availability, highlighting the need for innovative integration strategies in future work.

3.2. Timely Detection and Characterization of Global Hydrological Droughts

The BCNN-predicted TWSAs during the GRACE/FO latency period enable the timely detection and characterization of hydrological droughts at a global scale. To demonstrate this capacity, we compare hydrological droughts characterized by both GRACE/FO and BCNN during the testing period. The drought condition is identified using the water storage deficit index (WSDI), defined as:

$$\mathrm{WSDI}_{i,j} = \frac{\mathrm{TWSA}_{i,j}^{\mathtt{detrend}} - \mu_i}{\sigma_i}, \tag{2}$$

where j denotes the time series index, $i = 1, \ldots, 12$ is the ith calendar month, $\mathrm{TWSA}_{i,j}^{\mathtt{detrend}}$ is the detrended TWSA time series, and μ_i and σ_i are the mean and standard deviation of the detrended TWSA for month i, respectively. The linear trend was removed from the original TWSA time series to eliminate biases from long-term systematic changes in drought assessments (Humphrey et al., 2016). The discrepancies between BCNN and GRACE/FO drought characterizations are quantified using a mismatch metric (Saemian et al., 2024), defined as the percentage of months where drought classifications differ by two or more categories (see Text S6 in Supporting Information S1). Drought categories are detailed in Table S1 in Supporting Information S1, with drought conditions defined as $\mathrm{WSDI} \leq -0.5$ persisting for three or more consecutive months.

Figure 3 quantitatively demonstrates BCNN's capability for timely detection and reliable characterization of hydrological droughts during the GRACE/FO latency period through comprehensive drought mismatch analysis. Hyper-arid regions (e.g., the Sahara Desert, Arabian Peninsula and Taklamakan Desert) are excluded due to BCNN's reduced accuracy in these areas (Section 3.1). For the first delayed month, the mismatches are below 21% across 90% of the regions (Figure 3d), with a median value of 10.9% (Figure 3a). The median mismatch increases slightly to 14.5% and 16.4% for the second and third delayed months, respectively (Figures 3b and 3c). Throughout all latency months, mismatches remain below 30% in more than 90% of regions (Figure 3d). The superior timeliness and reliability of BCNN are further confirmed through comparative analysis with drought characterization mismatches from ERA5-Land reanalysis and Noah land surface model (see Text S5 and Figure S12 in Supporting Information S1).

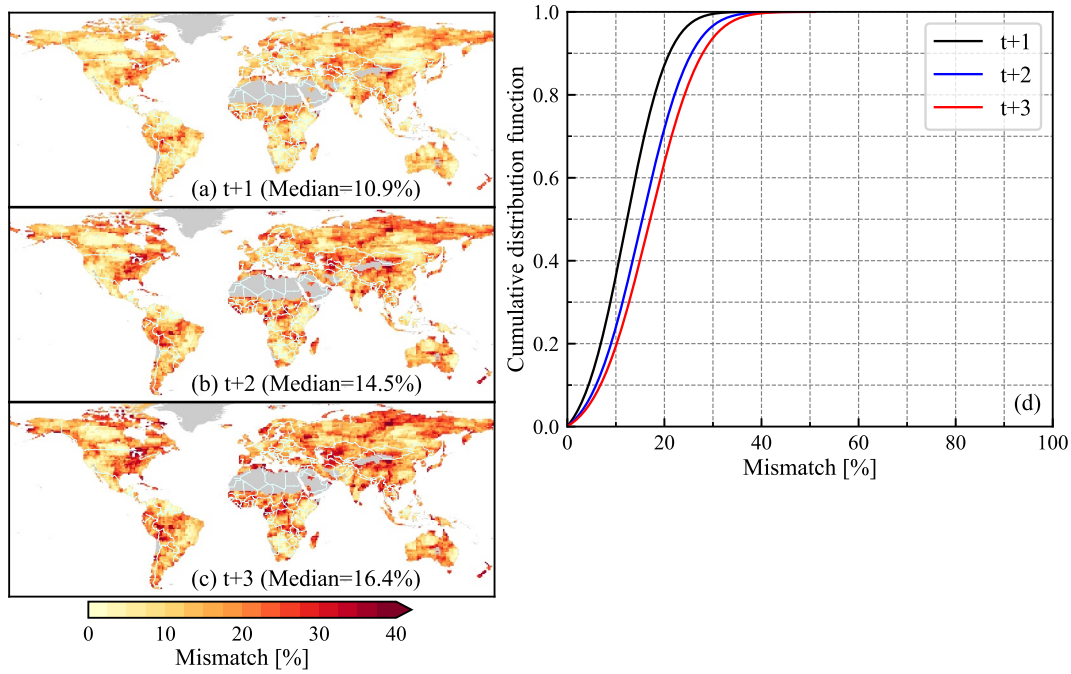

Figure 3. Distribution of disparities between BCNN-predicted and GRACE/FO-observed hydrological droughts during the testing period. The mismatch values represent the percentage of months where BCNN-identified droughts differ from GRACE/FO-identified ones by at least two categories. (a–c) Show the mismatches of BCNN's one-, two-, and three-month ahead predictions, respectively, while (d) summarizes their cumulative distributions.

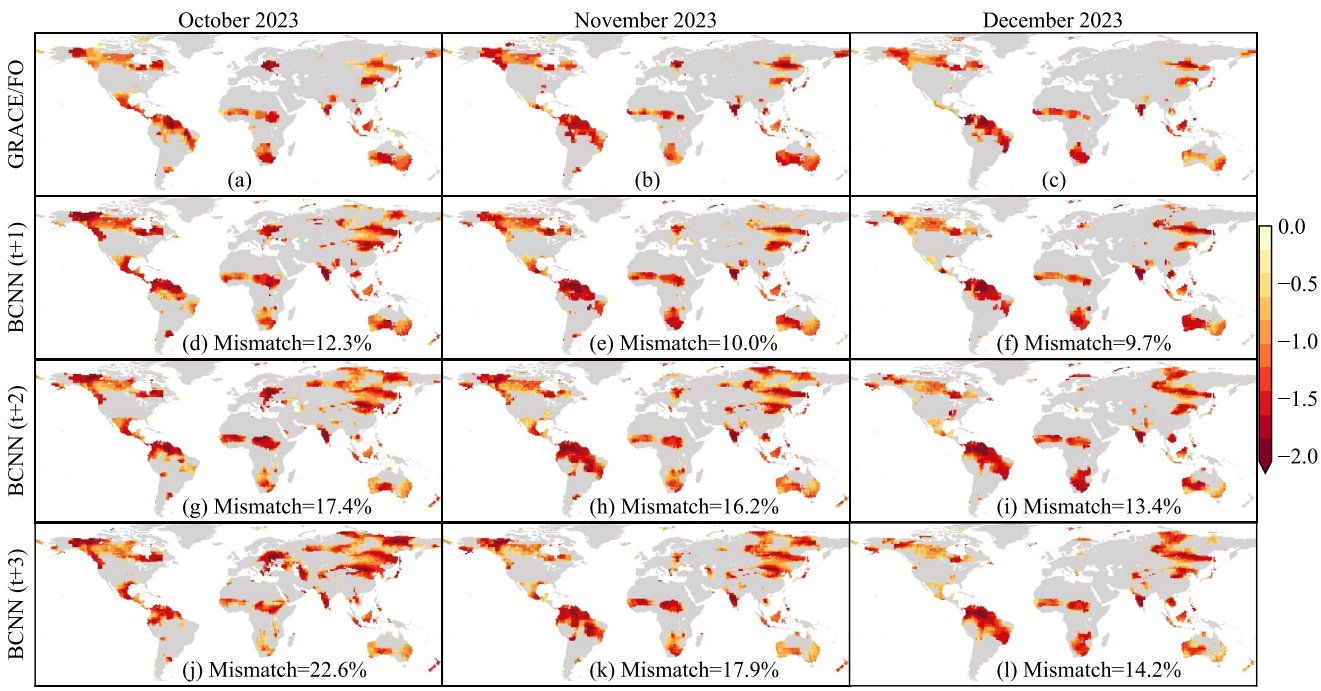

Figure 4. The hydrological drought regions identified by the water storage deficit index (WSDI) during October 2023 to December 2023. The first to fourth rows display results based on the WSDI indicators derived from GRACE/FO data and one- ($t + 1$) to three-month ($t + 3$) ahead predictions by the BCNN model, respectively. The mismatch value represents the percentage of land grids where BCNN-identified droughts differ from GRACE/FO-identified droughts by at least two categories.

Widespread hydrological droughts were documented globally during October-December 2022 and October-December 2023 (Singh et al., 2023; L. Zhang et al., 2023). Figure 4 and Figure S13 in Supporting Information S1 demonstrate the BCNN's effectiveness in predicting these established drought events, comparing GRACE/FO observations with BCNN's one- to three-month ahead predictions for both time periods. BCNN successfully captures the main spatiotemporal drought patterns identified by GRACE/FO, with spatial mismatches consistently below 23% across all months, though minor discrepancies in drought extent and severity are observed. These results validate the BCNN's capability to detect hydrological droughts up to three months before GRACE/FO data availability, effectively extending the temporal window for drought identification and mitigation.

Basin-scale validation results are presented in Figure S14 in Supporting Information S1, showing basin-averaged WSDI time series derived from GRACE/FO and BCNN predictions across six major river basins during the testing period. The accuracy scores ($R \in [0.81, 0.95]$, $NSE \in [0.65, 0.90]$) are lower than those for the original TWSA time series (Figure 2), reflecting the increased complexity in predicting WSDI signals due to the removal of trend and seasonal components (Cleveland et al., 1990; Humphrey et al., 2016; Mo, Zhong, Forootan, Shi, et al., 2022). Despite this challenge, BCNN-derived WSDI time series show good agreement with GRACE/FO observations, with the 95% prediction confidence intervals generally encompassing the GRACE/FO measurements.

The spatial characteristics of BCNN's drought predictions are further examined in Figure S15 in Supporting Information S1, which compares WSDI fields (excluding non-drought regions) between GRACE/FO observations and BCNN predictions for one to three months ahead during a month with extreme drought conditions in the six river basins. The BCNN-predicted WSDI fields closely replicate the spatial patterns observed in the GRACE/FO data, although there are some discrepancies in drought extent and severity, likely attributable to the significant fluctuations inherent in the WSDI signals (Humphrey et al., 2016; Mo, Zhong, Forootan, Shi, et al., 2022).

4. Discussion and Conclusions

Our study proposes BCNN, a novel DL framework that addresses a fundamental limitation in satellite-based hydrological monitoring: the approximately three-month data latency of GRACE/FO missions. This innovation enables near-real-time monitoring of global water availability and hydrological droughts that would otherwise remain unidentified during this crucial period. Our comprehensive evaluation demonstrates BCNN's reliable performance in predicting global TWSAs with uncertainty estimates, achieving median grid-level metrics of $R \in [0.92, 95]$, $NSE \in [0.81, 89]$, and $RMSE \in [1.79, 2.26]$ cm across the three-month latency period. The model's effectiveness extends to timely drought characterization, with consistently low median mismatches (10.9%, 14.5%, and 16.4%) over the three-month latency period, confirming its capability to reliably track hydrological drought dynamics and substantially mitigate the impact of GRACE/FO data delays.

BCNN's robust performance stems from two key innovations: the convolutional neural networks' capacity to extract informative features from multi-source image data, and the implementation of series stationarization to alleviate TWSA time series non-stationarity. Our analysis reveals that BCNN clearly outperforms traditional open-loop estimates from reanalysis data sets and land surface models during the GRACE/FO latency period. Notably, incorporating these open-loop estimates alongside meteorological predictors (P and T) as BCNN's inputs substantially enhances prediction accuracy, particularly in relatively humid regions characterized by high hydroclimatic variability.

In summary, this study presents several major advances in operational hydrology and environmental monitoring.

- Global operational hydrology advancement: BCNN bridges the critical latency gap in GRACE/FO data, enabling near-real-time monitoring of water availability and droughts up to three months ahead of conventional satellite observations. This breakthrough enhances proactive water resource management and drought mitigation strategies worldwide, particularly benefiting climate-vulnerable regions where early warning systems can significantly reduce socioeconomic impacts.
- Synergy between data-driven and physical modeling: By incorporating model-derived open-loop TWSA estimates as inputs, our BCNN framework demonstrates how DL can leverage physical modeling insights to enhance prediction accuracy. This synergy extends further, as BCNN's near-real-time predictions could be

assimilated into hydrological and land surface models, enhancing a deeper integration between DL and process-based hydrology that advances both fields.

- Scalability and flexibility: The BCNN framework's data-driven nature makes it adaptable to various hydrological and environmental applications where latency filling is crucial, extending its utility beyond TWSA prediction. Furthermore, its capacity to generate uncertainty estimates enhances its utility for both risk analysis and data assimilation processes.

While BCNN demonstrates robust performance, several avenues for enhancement remain. Integration of emerging DL techniques, such as attention mechanisms and physics-informed learning (Chen et al., 2023), could further improve prediction accuracy. Additionally, extending the framework to forecast future TWSAs and hydrological droughts beyond the GRACE/FO observation period represents a promising direction for future research. These advancements would strengthen our ability to predict and manage global water resources in an era of increasing climate variability and hydrological events.

Acronyms

BCNN	Autoregressive Bayesian convolutional neural network
GRACE/FO	Gravity Recovery and Climate Experiment and its Follow-On
NSE	Nash–Sutcliffe Efficiency
R	Correlation coefficient
RMSE	Root mean squared error
SVGD	Stein variational gradient descent
TWSA	Terrestrial water storage anomaly
WSD	Water storage deficit
WSDI	Water storage deficit index

Data Availability Statement

The JPL GRACE/FO Mascon and ERA5-Land data sets used in this study are publicly available at Wiese et al. (2023) and Muñoz Sabater (2019), respectively. The Python codes for the proposed DL method are available at Mo et al. (2025), as well as on GitHub: https://github.com/njujinchun/DL4TWSA and https://github.com/aaugeodesy/DL4TWSA.

Acknowledgments
This work was jointly funded by the National Natural Science Foundation of China (U2443201, 42330718, 42472321) and Villum Fonden (VIL60779). S. M. acknowledges financial support from the China Scholarship Council for his research stay at Aalborg University. We are grateful to the High-Performance Computing Center (HPCC) of Nanjing University for doing the numerical calculations in this paper on its blade cluster system. We also appreciate the Editors and anonymous reviewers for their helpful comments, which were used to improve the quality of this study.

References

Adusumilli, S., Borsa, A. A., Fish, M. A., McMillan, H. K., & Silverii, F. (2019). A decade of water storage changes across the contiguous United States from GPS and satellite gravity. *Geophysical Research Letters*, *46*(22), 13006–13015. https://doi.org/10.1029/2019GL085370

Almorox, J., Quej, V. H., & Martí, P. (2015). Global performance ranking of temperature-based approaches for evapotranspiration estimation considering Köppen climate classes. *Journal of Hydrology*, *528*, 514–522. https://doi.org/10.1016/j.jhydrol.2015.06.057

Boergens, E., Güntner, A., Dobslaw, H., & Dahle, C. (2020). Quantifying the central European droughts in 2018 and 2019 with GRACE follow-on. *Geophysical Research Letters*, *47*(14), e2020GL087285. https://doi.org/10.1029/2020GL087285

Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., et al. (2023). Iterative integration of deep learning in hybrid Earth surface system modelling. *Nature Reviews Earth and Environment*, *4*(8), 568–581. https://doi.org/10.1038/s43017-023-00452-7

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics*, *6*(1), 3–73.

Forootan, E., Awange, J., Kusche, J., Heck, B., & Eicker, A. (2012). Independent patterns of water mass anomalies over Australia from satellite data and models. *Remote Sensing of Environment*, *124*, 427–443. https://doi.org/10.1016/j.rse.2012.05.023

Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., van Dijk, A., et al. (2019). Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections. *Science of the Total Environment*, *650*, 2587–2604. https://doi.org/10.1016/j.scitotenv.2018.09.231

Forootan, E., Mehrnegar, N., Schumacher, M., Schiettekatte, L. A. R., Jagdhuber, T., Farzaneh, S., et al. (2024). Global groundwater droughts are more severe than they appear in hydrological models: An investigation through a Bayesian merging of GRACE and GRACE-FO data with a water balance model. *Science of the Total Environment*, *912*, 169476. https://doi.org/10.1016/j.scitotenv.2023.169476

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2019.00326

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition [Conference Proceedings]. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, *48*(7), W07525. https://doi.org/10.1029/2011WR011291

Hu, Z., Tang, S., Mo, S., Shi, X., Yin, X., Sun, Y., et al. (2024). Water storage changes (2003–2020) in the Ordos Basin, China, explained by GRACE data and interpretable deep learning. *Hydrogeology Journal*, *32*(1), 307–320. https://doi.org/10.1007/s10040-023-02730-4

Huang, G., Liu, Z., Maaten, L. V. D., & Weinberger, K. Q. (2017). Densely connected convolutional networks [Conference Proceedings]. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2261–2269). https://doi.org/10.1109/CVPR.2017.243

Huang, Y., Salama, M. S., Krol, M. S., Su, Z., Hoekstra, A. Y., Zeng, Y., & Zhou, Y. (2015). Estimation of human-induced changes in terrestrial water storage through integration of GRACE satellite detection and hydrological modeling: A case study of the Yangtze River basin. *Water Resources Research*, *51*(10), 8494–8516. https://doi.org/10.1002/2015WR016923

Humphrey, V., Gudmundsson, L., & Seneviratne, S. I. (2016). Assessing global water storage variability from GRACE: Trends, seasonal cycle, subseasonal anomalies and extremes. *Surveys in Geophysics*, *37*(2), 357–395. https://doi.org/10.1007/s10712-016-9367-1

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

Li, F., Kusche, J., Sneeuw, N., Siebert, S., Gerdener, H., Wang, Z., et al. (2024). Forecasting next year's global land water storage using GRACE data. *Geophysical Research Letters*, *51*(17), e2024GL109101. https://doi.org/10.1029/2024GL109101

Liu, Q., & Wang, D. (2016). Stein variational gradient descent: A general purpose bayesian inference algorithm. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems (NeurIPS)* (Vol. 29, pp. 2378–2386). Curran Associates, Inc.

Liu, Y., Wu, H., Wang, J., & Long, M. (2022). Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, *35*, 9881–9893. https://doi.org/10.48550/arXiv.2205.14415

Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., et al. (2017). Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models. *Remote Sensing of Environment*, *192*, 198–216. https://doi.org/10.1016/j.rse.2017.02.011

Loomis, B. D., Luthcke, S. B., & Sabaka, T. J. (2019). Regularization and error characterization of GRACE mascons. *Journal of Geodesy*, *93*(9), 1381–1398. https://doi.org/10.1007/s00190-019-01252-y

Lu, S., Li, W., Yao, G., Zhong, Y., Bao, L., Wang, Z., et al. (2024). The changes prediction on terrestrial water storage in typical regions of China based on neural networks and satellite gravity data. *Scientific Reports*, *14*(1), 16855. https://doi.org/10.1038/s41598-024-67611-8

Mo, S., Schumacher, M., van Dijk, A. I. J. M., Shi, X., Wu, J., & Forootan, E. (2025). A Bayesian convolutional neural network for predicting terrestrial water storage anomalies (TWSA) during the 3-month latency period of GRACE/FO data availability [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.14981355

Mo, S., Zhong, Y., Forootan, E., Mehrnegar, N., Yin, X., Wu, J., et al. (2022). Bayesian convolutional neural networks for predicting the terrestrial water storage anomalies during GRACE and GRACE-FO gap. *Journal of Hydrology*, *604*, 127244. https://doi.org/10.1016/j.jhydrol.2021.127244

Mo, S., Zhong, Y., Forootan, E., Shi, X., Feng, W., Yin, X., & Wu, J. (2022). Hydrological droughts of 2017–2018 explained by the Bayesian reconstruction of GRACE(-FO) fields. *Water Resources Research*, *58*(9), e2022WR031997. https://doi.org/10.1029/2022WR031997

Mo, S., Zhu, Y., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, *55*(1), 703–728. https://doi.org/10.1029/2018WR023528

Muñoz Sabater, J. (2019). ERA5-Land monthly averaged data from 1950 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.68d2bb30

Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Naser, M. Z., & Alavi, A. H. (2023). Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences. *Architectures and Structural Construction*, *3*(4), 499–517. https://doi.org/10.1007/s44150-021-00015-8

Reager, J., Thomas, B., & Famiglietti, J. (2014). River basin flood potential inferred using GRACE gravity observations at several months lead time. *Nature Geoscience*, *7*(8), 588–592. https://doi.org/10.1038/ngeo2203

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

Rodell, M., & Li, B. (2023). Changing intensity of hydroclimatic extreme events revealed by GRACE and GRACE-FO. *Nature Water*, *1*(3), 241–248. https://doi.org/10.1038/s44221-023-00040-5

Rodell, M., & Reager, J. T. (2023). Water cycle science enabled by the GRACE and GRACE-FO satellite missions. *Nature Water*, *1*, 47–59. https://doi.org/10.1038/s44221-022-00005-0

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. Wells, & A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (Vol. 9351, pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Saemian, P., Tourian, M. J., Elmi, O., Sneeuw, N., & AghaKouchak, A. (2024). A probabilistic approach to characterizing drought using satellite gravimetry. *Water Resources Research*, *60*(8), e2023WR036873. https://doi.org/10.1029/2023WR036873

Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, *36*(3), 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547–7569. https://doi.org/10.1002/2016JB013007

Schumacher, M., Forootan, E., van Dijk, A., Müller Schmied, H., Crosbie, R., Kusche, J., & Döll, P. (2018). Improving drought simulations within the Murray-Darling basin by combined calibration/assimilation of GRACE data into the WaterGAP global hydrology model. *Remote Sensing of Environment*, *204*, 212–228. https://doi.org/10.1016/j.rse.2017.10.029

Singh, M., Sah, S., & Singh, R. (2023). The 2023-24 El Niño event and its possible global consequences on food security with emphasis on India. *Food Security*, *15*(6), 1431–1436. https://doi.org/10.1007/s12571-023-01419-8

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: Can we learn from mismatch? *Water Resources Research*, *55*(2), 1179–1195. https://doi.org/10.1029/2018WR023333

Thomas, A. C., Reager, J. T., Famiglietti, J. S., & Rodell, M. (2014). A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters*, *41*(5), 1537–1545. https://doi.org/10.1002/2014GL059323

Tourian, M. J., Papa, F., Elmi, O., Sneeuw, N., Kitambo, B., Tshimanga, R. M., et al. (2023). Current availability and distribution of congo basin's freshwater resources. *Communications Earth & Environment*, *4*(1), 174. https://doi.org/10.1038/s43247-023-00836-z

van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., & Tregoning, P. (2014). A global water cycle reanalysis (2003-2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble. *Hydrology and Earth System Sciences*, *18*(8), 2955–2973. https://doi.org/10.5194/hess-18-2955-2014

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

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2648–2671. https://doi.org/10.1002/2014JB011547

Wiese, D. N., Yuan, D.-N., Boening, C., Landerer, F. W., & Watkins, M. M. (2023). JPL GRACE and GRACE-FO mascon ocean, ice, and hydrology equivalent water height CRI filtered [Dataset]. *Ver. RL06.1Mv03*. https://doi.org/10.5067/TEMSC-3JC634

Yang, F., Schumacher, M., Retegui-Schiettekatte, L., van Dijk, A. I. J. M., & Forootan, E. (2024). PyGLDA: A fine-scale python-based global land data assimilation system for integrating satellite gravity data into hydrological models. *Geoscientific Model Development Discussions*, *2024*, 1–34. https://doi.org/10.5194/gmd-2024-125

Zemp, M., Chao, Q., Dolman, A. J. H., Herold, M., Krug, T., Speich, S., et al. (2022). *The 2022 GCOS implementation plan*. World Meteorological Organization. Geneva. https://doi.org/10.5167/uzh-224271

Zhang, L., Yu, X., Zhou, T., Zhang, W., Hu, S., & Clark, R. (2023). Understanding and attribution of extreme heat and drought events in 2022: Current situation and future challenges. *Advances in Atmospheric Sciences*, *40*(11), 1941–1951. https://doi.org/10.1007/s00376-023-3000-0

Zhang, Z., Tang, Z., Wang, Y., Zhang, Z., Zhan, C., Zha, Z., & Wang, M. (2021). Dense residual network: Enhancing global dense feature flow for character recognition. *Neural Networks*, *139*, 77–85. https://doi.org/10.1016/j.neunet.2021.02.005

Zhao, M., A, G., Velicogna, I., & Kimball, J. S. (2017). A global gridded dataset of GRACE drought severity index for 2002–14: Comparison with PDSI and SPEI and a case study of the Australia Millennium drought. *Journal of Hydrometeorology*, *18*(8), 2117–2129. https://doi.org/10.1175/JHM-D-16-0182.1