Hydrological droughts of 2017-2018 explained by the Bayesian reconstruction of GRACE(-FO) fields

Mo, Shaoxing; Zhong, Yulong; Forootan, Ehsan; Shi, Xiaoqing; Feng, Wei; Yin, Xin; Wu, Jichun

Published in:
Water Resources Research

DOI (link to publication from Publisher):
10.1029/2022WR031997

Publication date:
2022

Document Version
Early version, also known as pre-print

Hydrological droughts of 2017-2018 explained by the Bayesian reconstruction of GRACE(-FO) fields

Shaoxing Mo[1], Yulong Zhong[2], Ehsan Forootan[3], Xiaoqing Shi[1], Wei Feng[4,5], Xin Yin[6], and Jichun Wu[1]

[1]Key Laboratory of Surficial Geochemistry of Ministry of Education, School of Earth Sciences and Engineering, Nanjing University, Nanjing, China
[2]School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan, China
[3]Geodesy Group, Department of Planning, Aalborg University, Aalborg, Denmark
[4]School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai, China
[5]State Key Laboratory of Geodesy and Earth's Dynamics, Institute of Geodesy and Geophysics, Innovation Academy for Precision Measurement Science and Technology, Chinese Academy of Sciences, Wuhan, China
[6]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China

Key Points:

- The missing TWSA signals between the 2017-2018 GRACE and GRACE-FO gap are reconstructed by Bayesian deep learning
- Hydrological droughts overlapping with this gap period are characterized and quantified
- The northern mid-latitudes experienced continuous water storage deficits as a result of precipitation deficiency

Corresponding author: Xiaoqing Shi, `shixq@nju.edu.cn`

Corresponding author: Wei Feng, `fengwei@sysu.edu.cn`

Abstract

Hydrological droughts are events of prolonged water scarcity and cause many devastating impacts. It is, therefore, extremely crucial to understand their spatiotemporal evolution to guide prevention and mitigation policies. The Gravity Recovery and Climate Experiment (GRACE, April 2002-June 2017) and GRACE Follow-On (GRACE-FO, June 2018-present) missions have been used to study large-scale droughts of almost two decades. But characterizing droughts during the between missions gap period of 2017-2018 has not been well addressed and will be covered here. To bridge the gap, an innovative Bayesian convolutional neural network is developed to reconstruct the missing signals from hydroclimatic inputs. The reconstruction fields and existing signals are then used to explore regions that have experienced consecutive water storage deficits during the 2017-2018 gap. We found many regions of the northern mid-latitudes exhibiting moderate to exceptional droughts in terms of water storage deficits, among which parts of Pakistan and Afghanistan, and Iberian Peninsula experienced exceptional droughts lasting for more than one year with the maximum deficits (-4.4 ± 0.8 cm and -7.2 ± 1.1 cm, respectively) being over 50% of the seasonal storage variations. Comparisons with climate indicators show that the identified droughts are predominantly caused by continuous below-normal precipitation. The recovery process correlates generally well with the accumulation rate of precipitation surpluses (the correlation coefficient (R) can be up to 0.92). Besides, the reconstructed signals, which have $R > 0.7$ with the testing GRACE(-FO) data in over 90% of the globe, reliably maintain the data continuity and therefore they are recommended for hydro-climatological studies.

1 Introduction

Drought is an event of prolonged water shortages and usually associated with long-standing (months or even years) and devastating impacts on ecosystem, agriculture, and society (Hao et al., 2018; X. He, Estes, et al., 2019; Van Loon, 2015). To better manage these events and alleviate their impacts, it is extremely essential to understand the characteristics of historical drought events well, including their duration, intensity, affected areas, severity, and the process of droughts recovery. This information is useful for developing proper water resources management policies to tackle challenges that likely occur in future (AghaKouchak et al., 2015; A. K. Mishra & Singh, 2010; Shah & Mishra, 2020; Van Loon, 2015; West et al., 2019). Droughts are generally classified into meteorological (deficit in precipitation), agricultural (deficit in soil moisture), hydrological (deficit in total water storage), and socioeconomic (water demand exceeds supply) (Wilhite & Glantz, 1985). What we are concerned in this study can be categorized as hydrological droughts, and the study addresses the spatiotemporal evolution of global water storage deficits during the major drought events of 2017-2018.

The Gravity Recovery and Climate Experiment (GRACE, 2003-2017) satellite mission and its follow-on (GRACE-FO, 2018-now) provide time-variable global gravity fields that can be converted to the terrestrial water storage anomaly (TWSA) estimates (Tapley et al., 2004). These fields are unique because they can reflect a summation of surface water (e.g., measured by altimetry and SAR missions), and shallow soil water storage (e.g., monitored by soil moisture remote sensing), as well as deep water storage in the rooted soil and groundwater compartments. The latter two compartments cannot be measured by other remote sensing techniques. Therefore, GRACE and GRACE-FO have been used widely for studying global large-scale hydrometeorological events (see, e.g., Chen et al., 2010; Eicker et al., 2016; Forootan et al., 2019; X. Liu et al., 2020; Long et al., 2013, 2014; Boergens, Güntner, et al., 2020; Humphrey et al., 2016; B. Li et al., 2019; Mehrnegar et al., 2021; Thomas et al., 2014; Yan et al., 2021).

However, there is a major research gap of the existing literature related to characterization of the hydrological drought events of 2017-2018. The key reason for this research gap is that there is an approximately one-year gap period (July 2017-May 2018) between the GRACE and GRACE-FO missions, therefore a reliable estimation of TWSA fields in this period is missing. As a result, the spatiotemporal characteristics of the hydrological drought events, whose duration is overlapped with this gap period, are still largely unknown. For instance, the missing GRACE(-FO) data records inhibit a full quantification of the 2018-2019 Central European drought (Boergens, Güntner, et al., 2020) and of the 2017-2019 Eastern Australian drought, which accounted for the severe and long-lasting 2019-2020 bushfires (Deb et al., 2020; Kumar et al., 2021). Many other regions also experienced drought conditions during the gap period, as consistently indicated in literature (see, e.g., V. Mishra et al., 2021; Theron et al., 2021) and by multiple drought monitors/indicators (e.g., US `https://droughtmonitor.unl.edu` and NASA `https://nasagrace.unl.edu/`) from the perspective of deficits in precipitation, soil moisture, or shallow groundwater. However, little is still known about the associated total water storage losses, which is a key characteristic for hydrological drought characterization. Lacking reliable characterizations for these recent historical droughts poses challenges to future drought planning and management (Haile et al., 2019; X. He, Feng, et al., 2019; A. K. Mishra & Singh, 2010).

Reconstructing gaps in GRACE(-FO) has been done in past by formulating a mathematical relationship between TWSA and its indicators (e.g., Forootan et al., 2014, 2020; Humphrey & Gudmundsson, 2019; F. Li et al., 2021; A. Y. Sun et al., 2019, 2020; Z. Sun et al., 2020; Yi & Sneeuw, 2021). However, the accuracy of reconstruction may be limited especially in (semi-)arid regions due to the high-complexity of hydrological processes in these regions (Mo et al., 2022; Z. Sun et al., 2020). Considering that the estimation of water storage deficits requires to remove the easiest-to-reconstruct seasonal component from the TWSA signals and the resulting deficit signals are subject to high-frequency variability (Humphrey et al., 2016), the limited TWSA reconstruction accuracy will lead to a poor reconstruction for the deficits. Consequently, their applications in quantifying droughts during the gap period, to the best of our knowledge, have not yet been reported.

This study fills the existing the research gap by addressing the following questions: (1) how did the hydrological drought-induced water storage deficits spatiotemporally evolve during the 2017-2018 GRACE and GRACE-FO gap? and (2) what are the major causes for these drought events? For this, an innovative Bayesian Convolutional Neural Network (BCNN) is firstly developed to reliably reconstruct the missing TWSA fields from hydroclimatic inputs (indicators) in an image-to-image (field-to-field) regression setup. The BCNN-reconstructed signals are then used together with the existing TWSA observations to bridge the mentioned research gap. The outperformance of convolutional neural network over traditional statistical and machine learning methods in reconstructing the missing TWSA fields has been illustrated in A. Y. Sun et al. (2019) and our previous work (Mo et al., 2022). Its deep architecture with multiple nonlinear processing layers enables it to extract spatially-correlated and multiscale features of the data fields for effectively learning highly-complex mappings (LeCun et al., 2015; Du et al., 2022; Mo, Zhu, et al., 2019; Mo, Zabaras, et al., 2019; Mo et al., 2020; Reichstein et al., 2019; Shen, 2018). To further improve the TWSA reconstruction accuracy, the novel components introduced in BCNN relative to the convolutional neural networks used in A. Y. Sun et al. (2019) and Mo et al. (2022) include the residual-in-residual dense block to enhance information flow through deep networks (Wang et al., 2018) and the spatial-channel attention module to enable the network to automatically focus on useful features (Fu et al., 2019). The Bayesian nature of BCNN enables it to provide uncertainty estimation for the deficit estimates. Finally, the hydrological drought quantification results will be examined by comparing with independent drought indices.

In summary, our innovative contributions are as follows. First and most importantly, we cover the aforementioned research gap that the recent hydrological drought events overlapping with the 2017-2018 GRACE and GRACE-FO gap period are characterized and quantified. The improved understanding toward spatiotemporal evolution of the associated water storage deficits, major causes for these droughts, and major factors influencing their recovery is of crucial significance for future drought management and mitigation. Second, we provide improved reconstructions of the missing 2017-2018 TWSA signals. Therefore, our reconstructions can more reliably maintain the TWSA data continuity and benefit subsequent hydroclimatic applications desiring a continuous data record. Finally, the proposed deep learning method can also be flexibly extended to other hydroclimatic problems involving learning complex mappings due to its generic nature.

2 Data and Methods

2.1 GRACE-derived TWSAs

There are multiple products of GRACE TWSA data, such as the spherical harmonic (SH) solutions provided by the Jet Propulsion Laboratory (JPL), German Research Center (GFZ), Center for Space Research (CSR) and many others, three mascons (provided by JPL, CSR, and Goddard Space Flight Center (GSFC)), and the GravIS and COST-G products by GFZ and the International Gravity Field Service (IGFS), respectively. These products differ mainly in the employed postprocessing algorithms. More details regarding GRACE data processing can be found in, for example, Save et al. (2016) and Boergens, Dobslaw, et al. (2020).

It has been shown that an ensemble of different GRACE products is beneficial for reducing the associated uncertainties caused by signal processing (Ali et al., 2022; Sakumura et al., 2014; Yan et al., 2021). Therefore, we take a weighted average of the eight products in the spatial domain as the observed TWSAs (The signal attenuation in SH products is restored by multiplying the provided scaling factors). The weights are determined using the generalized three-cornered hat method (Long et al., 2017; Xu et al., 2020; Yin et al., 2021). The time-mean baseline for computing anomalies and the spatial resolutions of these products are not completely consistent. We recompute the anomalies of all products using the same time-mean baseline (January 2004-December 2009) and resample averagely the data into $1° \times 1°$ grids.

The TWSA time series may exhibit long-term declining/rising trends caused by anthropogenic activities and/or climate change. Capturing the long-term trends is challenging as they may not be well reflected by the predictor data (Humphrey & Gudmundsson, 2019; Z. Sun et al., 2020). Fortunately, the GRACE TWSA data in the pre- (April 2002-June 2017) and post-gap (June 2018-) periods are available. We remove the linear trend ($\mathrm{trend}_{\mathrm{GRACE}}$) fitted using the available GRACE data (i.e., April 2002-June 2017 and June 2018-December 2020) from the original time series ($\mathrm{TWSA}_{\mathrm{GRACE}}$) and let BCNN learn to reconstruct the detrended TWSA signals ($\mathrm{TWSA}_{\mathrm{GRACE}}^{\mathrm{detrend}}$) instead. Finally, the reconstructions for the $\mathrm{TWSA}_{\mathrm{GRACE}}$ time series are obtained by adding the linear GRACE trend:

$$\mathrm{TWSA}_{\mathrm{BCNN}} = \mathrm{TWSA}_{\mathrm{BCNN}}^{\mathrm{detrend}} + \mathrm{trend}_{\mathrm{GRACE}}, \tag{1}$$

where $\mathrm{TWSA}_{\mathrm{BCNN}}^{\mathrm{detrend}}$ denotes BCNN's reconstruction for $\mathrm{TWSA}_{\mathrm{GRACE}}^{\mathrm{detrend}}$.

2.2 Hydroclimatic Predictors and Optimal Data Source Selection

BCNN inherits the convolutional neural network's flexible and excellent capability in efficiently handling multiple image inputs and extracting useful features from multi-source data (LeCun et al., 2015; Shen, 2018). Therefore, in addition to the commonly used predictors, namely simulated/reanalyzed TWSA (sTWSA), precipitation (P), and temperature (T) (Humphrey & Gudmundsson, 2019; Z. Sun et al., 2020), we consider additional four TWSA-related predictors: climatic water balance (CWB), cumulative water storage change (CWSC), cumulative precipitation anomaly, and cumulative CWB (CCWB). CWB is defined as the difference between P and potential evapotranspiration (PET). Thus, CCWB is written as:

$$\mathrm{CCWB}_t = \sum_{i=1}^{t} (\mathrm{P}_i - \mathrm{PET}_i), \tag{2}$$

where $t = 1, \ldots, T$, with T denoting the length of time series. CWSC is by definition the calculated TWSA calculating as the cumulative difference between the inflow (i.e., P) and outflow (i.e., evapotranspiration ET and runoff RO) of a region based on the water mass balance:

$$\mathrm{CWSC}_t = \sum_{i=1}^{t} (\mathrm{P}_i - \mathrm{ET}_i - \mathrm{RO}_i). \tag{3}$$

Human activities (e.g., reservoir operation and groundwater extraction) may also contribute to water storage changes, but they are difficult to quantify and thus not considered here.

Despite this, it is assumed that human activities are to some extent related to the climatic conditions. Therefore, the consideration of multiple hydroclimatic predictors may partially alleviate the influence of the omission of human activities. All predictor data are resampled averagely into $1° \times 1°$ grids to match the GRACE resolution.

Since multiple data products are usually available for each predictor and their quality varies with space and time, we make a grid-wise data source selection to identify the optimal data source for reconstructing GRACE TWSAs. For each grid cell, the product having the highest criterion value among all data source candidates is used as the predictor data. The selection criterion is the Nash-Sutcliffe efficiency or absolute correlation coefficient between the predictor and GRACE TWSA time series. The data source candidates considered for each predictor and the selection criterion are listed in Table C1. Similarly, the predictor time series are also detrended as GRACE TWSAs. The identified optimal data source for each predictor is shown in Figure S1.

2.3 Drought Indices

The hydrological drought index is defined on the basis of the deficit in total water storage, which is the negative deviation relative to the long-term climatology (also known as the seasonal signal) (Humphrey et al., 2016; Thomas et al., 2014), and can be calculated using the existing GRACE(-FO) TWSA observations and BCNN's reconstructions. Prior to calculating the climatology, the linear trend of the original TWSA time series is removed to eliminate the effect of non-climatic factor-induced TWSA declining/increasing trends on drought evaluation as suggested by Humphrey et al. (2016) and X. Liu et al. (2020). For instance, the consistently declining TWSA trend induced by groundwater over-exploitation, will lead to an underestimation of the drought condition in the former period and an over-estimation in the latter period (X. Liu et al., 2020). Formally, the water storage deficit (WSD) is calculated as

$$\mathrm{WSD}_{i,j} = \mathrm{TWSA}_{i,j}^{\texttt{detrend}} - \mu_i, \tag{4}$$

where j denotes the index of time series, and $i = 1, \ldots, 12$ represents the ith calendar month of a year, $\mathrm{TWSA}_{i,j}^{\texttt{detrend}}$ is the detrended TWSA, and μ_i is the long-term climatology calculated by averaging all of the ith-month observations in detrended data record. A standardized water storage deficit index (WSDI) defined as

$$\mathrm{WSDI}_{i,j} = \frac{\mathrm{WSD}_{i,j}}{\sigma_i}, \tag{5}$$

is used to facilitate the identification of droughts in different regions, where σ_i is the standard deviation of $\mathrm{TWSA}_{i,j}^{\texttt{detrend}}$ in month i.

Drought usually starts from a deficit in precipitation (i.e. meteorological drought) and further translates to agricultural/hydrological drought (Van Loon, 2015; West et al., 2019). Therefore, the SPEI-6 (6-month standardized precipitation evapotranspiration index) (Vicente-Serrano et al., 2010) meteorological drought index and the standardized soil moisture (SM-Z) agricultural drought index from the Climate Prediction Center (CPC) (van den Dool et al., 2003) are used to validate the occurrence of the identified hydrological droughts during the 2017-2018 gap. The SM-Z index are standardized similarly as WSDI in equation (5). To make the data comparable with GRACE, we apply a 300 km Gaussian filter to the gridded drought indices. Note that the meteorological/agricultural drought indices (e.g., SPEI-6 and SM-Z), which are based commonly on precipitation, evapotranspiration, or soil moisture, can only represent the dynamics of respective water storage/cycle components. One unique merit of the GRACE-based WSD indicator is the ability to quantify the deficits in total water availability (Chen et al., 2010; Z. Sun et al., 2018; Zhao et al., 2017).

2.4 BCNN Model

The BCNN deep learning model is tasked with learning the underlying relationship between the target GRACE TWSAs and seven predictors. Let $\mathbf{x} \in \mathbb{R}^{n_x \times H \times W}$ and $\mathbf{y} \in \mathbb{R}^{n_y \times H \times W}$ denote the n_x input (i.e., predictor) fields and n_y output (i.e., GRACE TWSA) fields, respectively, with $H \times W$ denoting the spatial resolution. They are treated in BCNN as images and the high-dimensional mapping learning task becomes an image-to-image regression problem between n_x and n_y images:

$$\boldsymbol{\eta}(\mathbf{x}, \mathbf{w}): \ \mathbf{x} \in \mathbb{R}^{n_x \times H \times W} \longrightarrow \mathbf{y} \in \mathbb{R}^{n_y \times H \times W}, \tag{6}$$

where $\mathbf{w}$ denotes all trainable parameters of the network $\boldsymbol{\eta}(\cdot, \cdot)$. The region spanning 60°S-84°N and 180°W-180°E is considered, excluding Greenland and Antarctica. Thus, $H \times W = 144 \times 360$.

The BCNN network architecture is depicted Figure 1. It contains multiple processing units for feature extraction and nonlinear transformations. The architecture is U-shaped that the input images are sequentially downsampled with convolutional layers and then recovered by transposed convolutional layers to extract multi-scale features (Ronneberger et al., 2015). The major improvement in network architecture design compared to A. Y. Sun

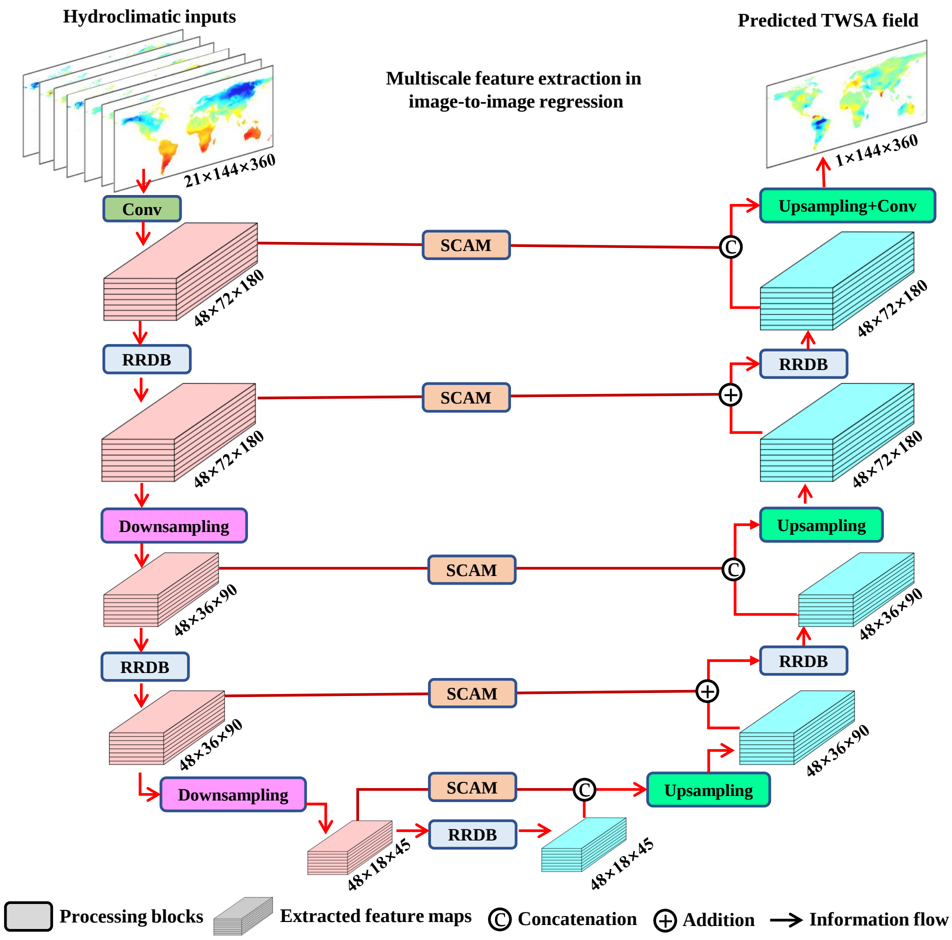

Figure 1. BCNN network architecture. It consists of alternating residual-in-residual dense blocks (RRDBs) and downsampling/upsampling layers for feature extraction and nonlinear processing in image-to-image regression. The network is U-shaped that the size of feature maps is sequentially halved by downsampling layers and then recovered by upsampling layers to extract multiscale features. ⊕ and ⓒ denote the addition (residual connection) and concatenation (dense connection) operations on feature maps, respectively. The spatial and channel attention module (SCAM) is used to enable BCNN to automatically focus on informative features. Conv: convolution.

et al. (2019) and Mo et al. (2022) is that we propose a novel basic building block for BCNN for higher performance. More details are given in Appendix A.

Our aim is to reconstruct the TWSA signals during the 2017-2018 gap between GRACE and GRACE-FO. For those few one- or two-month gaps within the GRACE and GRACE-FO missions, we fill them with the `pchip` interpolation implemented in MATLAB for simplicity. Alternatively, one can simply leave these months out or fill them with BCNN. The GRACE data of April 2002-March 2014 are used to train the BCNN network, and those of April 2014-June 2017 and June 2018-December 2020 to test the performance. Considering that there may be a lagged response of TWSA to the predictors, the predictors in months t-2 to t are used as the inputs to BCNN to reconstruct the TWSA in month t. Thus, each predictor-predictand sample contains $n_x = 21$ input images and $n_y = 1$ output image. Details on BCNN training are given in Appendix A. The benefits of considering more predictors in addition to the three commonly used ones (i.e., sTWSA, P, and T) are illustrated and discussed in Appendix B.

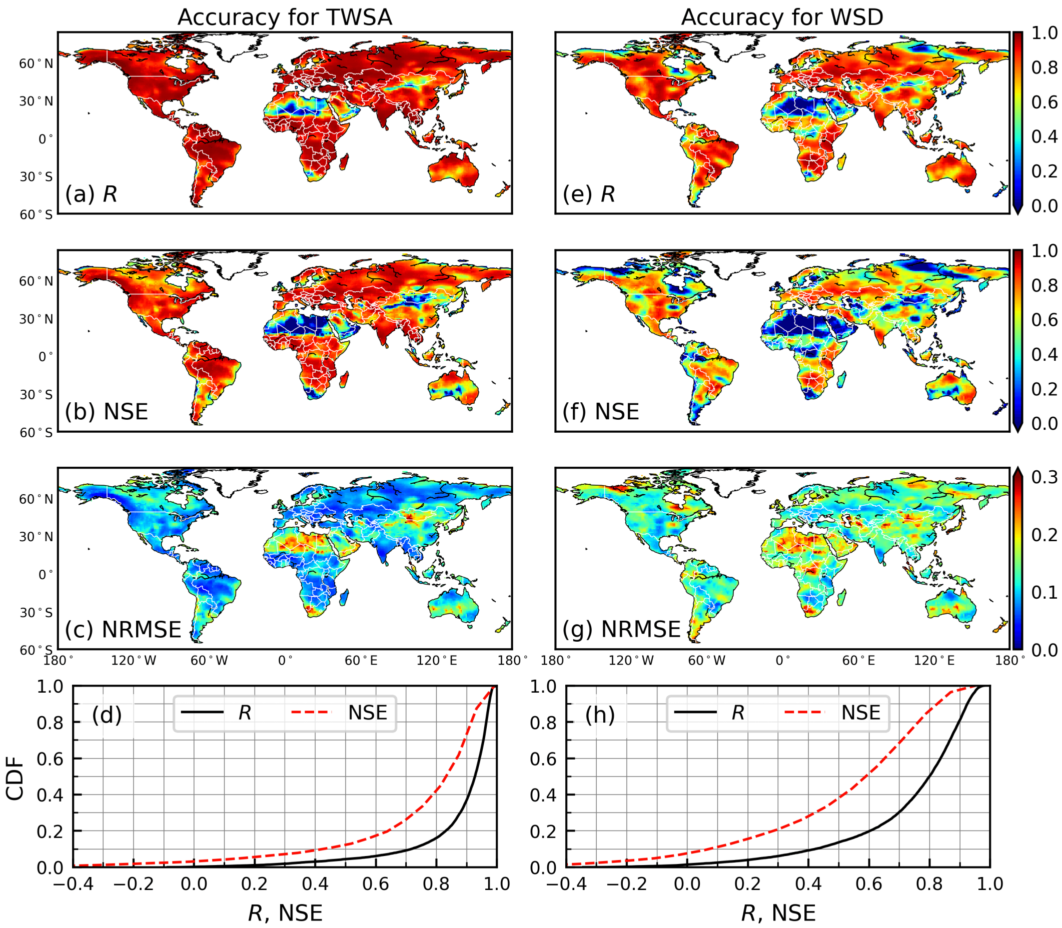

Figure 2. Reconstruction accuracy for the testing GRACE TWSA (a-d) and WSD (the detrended and deseasonalized TWSA) signals (e-h). CDF: cumulative distribution function; NRMSE: normalized root mean squared error; NSE: Nash-Sutcliffe efficiency coefficient; R: correlation coefficient.

3 Results and Discussion

3.1 TWSA Reconstruction Results

Figures 2 show BCNN's reconstruction accuracy in terms of correlation coefficient (R), Nash-Sutcliffe efficiency (NSE) coefficient, and normalized root mean squared error (NRMSE) for the original GRACE TWSA and WSD (i.e., the detrended and deseasonalized TWSA; equation (4)) signals during the testing periods. The NRMSE metric is a normalized version of the root mean squared error divided by the difference between the maximum and minimum values of the signals for each grid. For the original GRACE TWSA signals, BCNN achieves relatively high R and NSE, and low NRMSE reconstruction accuracy in most regions, with over 90% and 80% of grids having $R >$0.7 and NSE>0.6, respectively, as indicated by the cumulative distribution functions in Figure 2d. The hyper-arid regions (e.g., Sahara and Gobi) generally have low R and NSE values, which is mainly due to the low signal-to-noise ratio (Humphrey et al., 2016).For the detrened and deseasonalized WSD signals, the reconstruction accuracy show a similar spatial pattern to those of the original signals but decreases as expected since the deseasonalized time series are usually associated with high-frequency variability (Humphrey et al., 2016). However, considering the high variability of the WSD signals, BCNN still obtains a relatively high reconstruction accuracy with over 50% of grids having $R >$0.8 and NSE>0.6 (Figure 2h).

Figure 3 compares BCNN TWSAs with the testing GRACE TWSAs in December 2015 and November 2020 (The reconstructions for all testing months are shown in the GIF animation attached in the supporting information). The Bayesian nature of BCNN enables it to quantify the predictive uncertainty. Thus, the error and standard deviation of BCNN's reconstructions are also shown. The two months are shown because they experienced a very strong El Niño event (December 2015) and a moderate La Niña event (November 2020) (Figure C1). The climate extremes usually cause abnormal TWSA signals which are challenging to reconstruct. It is found that BCNN captures the spatial patterns of GRACE TWSAs relatively well and provides close reconstructions in both months. The time series of basin-averaged BCNN and GRACE TWSAs for 21 major river basins over the globe (Figure C2) are compared in Figure 4. It is observed that BCNN reliably bridges the data gap in the sense that the reconstructed TWSAs agree well with the testing GRACE TWSAs in the pre- and post-gap periods in these river basins dominated by either seasonal (e.g. Amazon, Figure 4a) or inter-annual/long-term (e.g. Murray-Darling, Figure 4k) signals

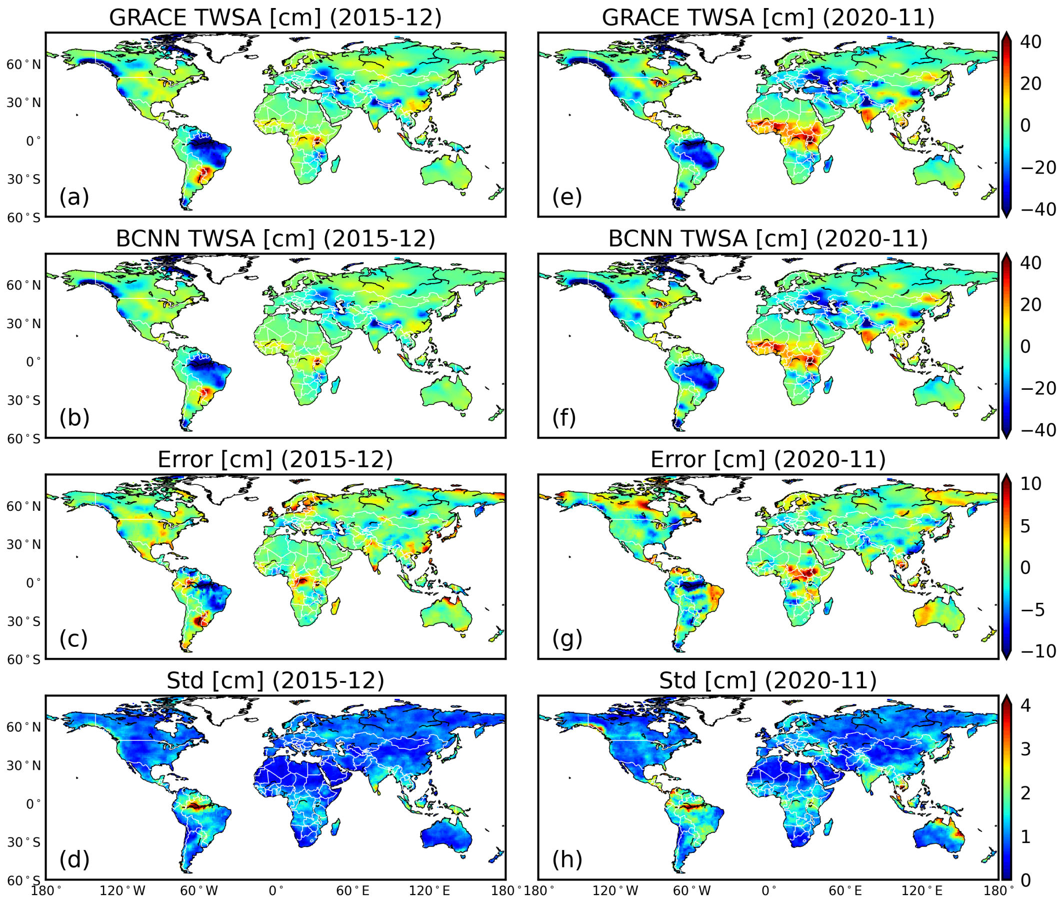

Figure 3. BCNN's reconstructions for the GRACE TWSA fields in December 2015 (a-d) and November 2020 (e-h). Error (c and g) denotes the difference between GRACE and BCNN TWSAs. (d) and (h) show the standard deviation (Std) of BCNN's reconstructions.

(see Humphrey et al. (2016) for the distribution of the relative contributions of different signal components).

The above results indicate BCNN provides close reconstructions for the unseen testing GRACE TWSAs in the pre- and post-gap periods, suggesting that BCNN reliably maintains the TWSA data continuity and thus enhance data consistency. Note that an implicit assumption here is that if BCNN is able to provide close reconstructions to the testing (unseen) GRACE observations in the pre- and post-gap periods, the gap-filling reconstructions are thought to be reliable. Figure C3 further illustrates the superior performance of BCNN in providing improved TWSA and WSD reconstructions by comparing the NSE accuracy with that obtained in Mo et al. (2022), which achieved a clearly higher reconstruction accuracy in comparison with previous studies (Humphrey & Gudmundsson, 2019; F. Li et al.,

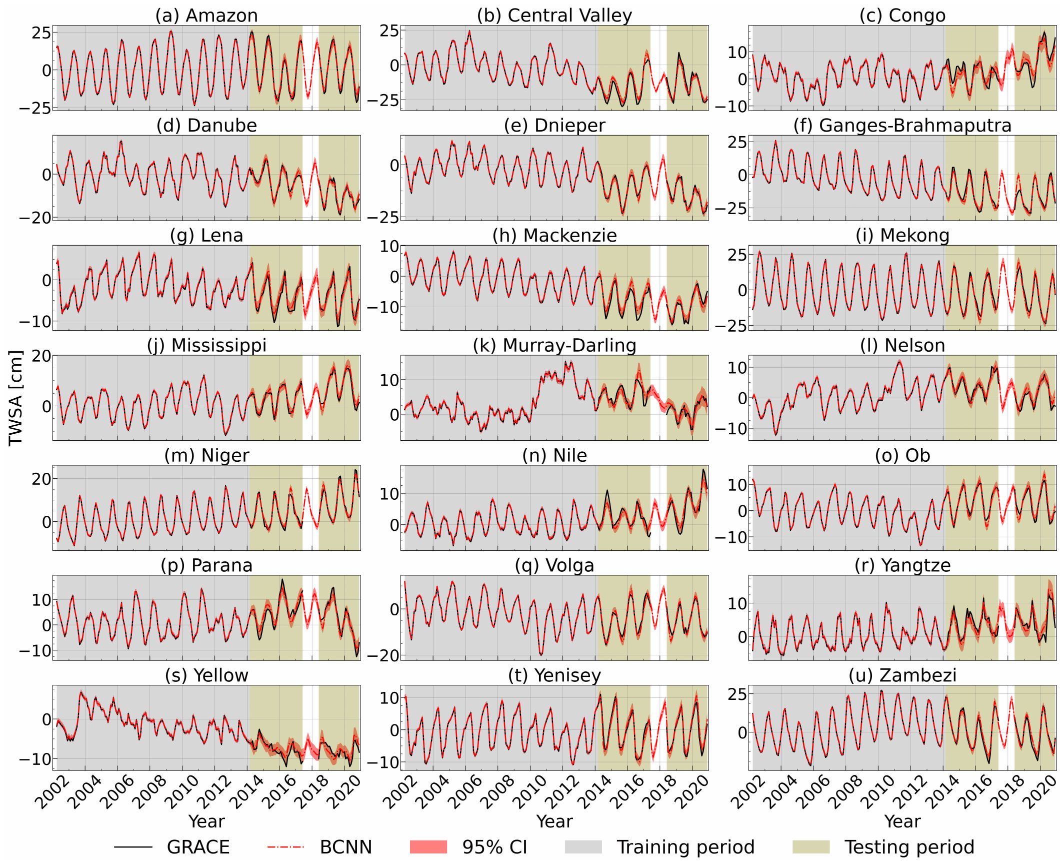

Figure 4. The basin-averaged GRACE and BCNN TWSA time series. The red shaded areas denote the 95% CI of BCNN predictions. The BCNN model's training period is April 2002-March 2014 and the testing periods are April 2014-June 2017 and June 2018-December 2020. CI: confidence interval; TWSA: terrestrial water storage anomaly.

2021; Z. Sun et al., 2020) when evaluating on the same testing GRACE data. We attribute the outperformance of our BCNN jointly to the new network architecture, the data source selection strategy, and the consideration of more predictors.

3.2 Characterization of the 2017-2018 Hydrological Droughts

The hydrological drought regions are identified using the WSDI index (equation (5)) and the associated water storage deficit is quantified using the WSD index (equation (4)). The SPEI-6 meteorological drought index and the SM-Z agricultural drought index, as introduced in section 2.3, are utilized to verify the occurrence of the WSDI-identified droughts, though they indicate droughts in different hydrological cycle components. The regions having a negative index value for at least three consecutive months with the maximum less than or equal to -0.5 are designated as experiencing droughts.

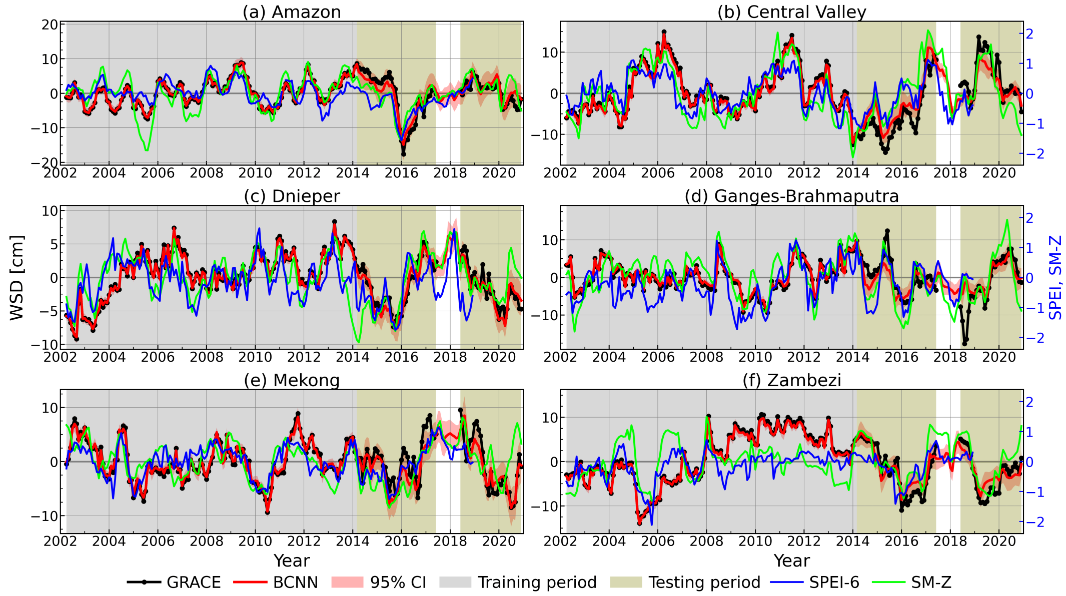

Figure 5. Basin-averaged time series of GRACE- and BCNN-derived WSDs, SPEI-6, and SM-Z in Central Valley and five river basins. These basins/regions experienced droughts during the 2015-2016 El Niño event. The red shaded areas denote the 95% CI (confidential interval) of BCNN reconstructions. SM-Z: standardized soil moisture; SPEI-6: 6-month standardized precipitation evapotranspiration index; WSD: water storage deficit.

We first demonstrate BCNN's ability in quantifying the water storage deficits induced by GRACE-recorded hydrological droughts during BCNN's testing period. The drought regions during the strongest 2015-2016 El Niño event on record (Figure C1) identified by GRACE and BCNN are compared in Figure C4. It is observed that BCNN-identified regions are generally consistent with those identified by GRACE. The basin-averaged time series of water storage deficit and two meteorological/agricultural drought indices (SPEI-6 and SM-Z) in six river basins within these regions are compared in Figure 5. BCNN successfully reproduces the GRACE-based time series of the water storage deficit during the testing period, although the reconstructions may be slightly smoothed. In addition, the two meteorological/agricultural drought indices correlate well with GRACE- and BCNN-derived water storage deficits, suggesting that they can be used later to examine the BCNN-identified droughts during the 2017-2018 GRACE and GRACE-FO gap.

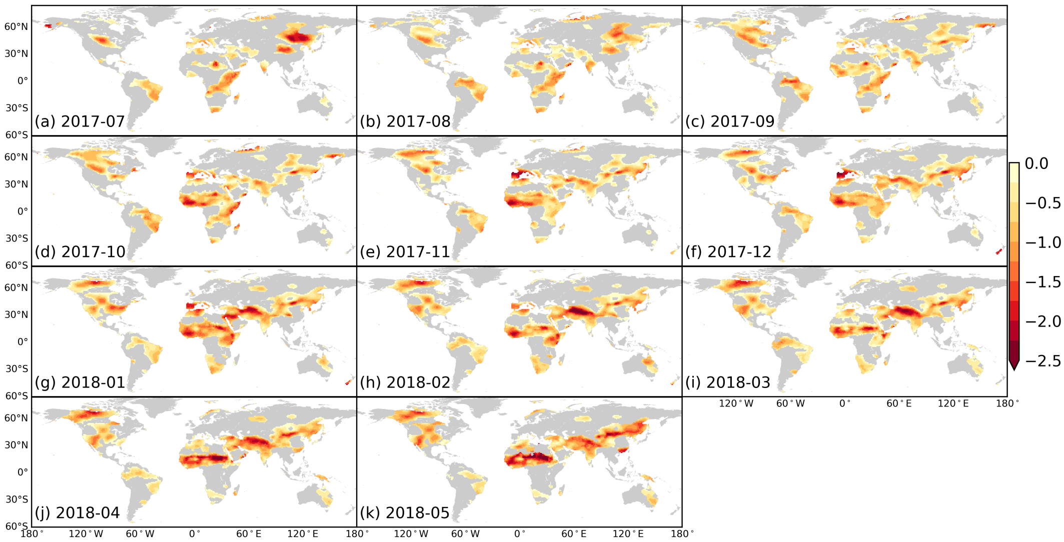

Figure 6. Drought regions identified by the BCNN WSDI index during the GRACE and GRACE-FO gap (July 2017-May 2018). WSDI: water storage deficit index.

The drought regions during the 2017-2018 gap identified by BCNN-derived WSDI, SPEI-6, and SM-Z indices are depicted in Figures 6, C5, and C6, respectively. The regions identified by WSDI show a favorable spatiotemporal agreement with those by the other two indices. They consistently indicate that the droughts occurred in the northern mid-latitudes. We now quantify the basin-scale water storage deficits during the gap for six river basins within the identified regions, including the Amur, Indus-Helmand-Amu-Darya, Tigris-Euphrates, Colorado, and Missouri River Basins and those in Iberian Peninsula. The basin-averaged time series of GRACE- and BCNN-derived WSD, SPEI-6, and SM-Z are shown in Figures 7, in which the latter three indices again consistently suggest the occurrence of 2017-2018 droughts in these basins. Basing on the BCNN-derived WSD index, the droughts are characterized and quantified in Table 1. These droughts completely covered the 11-month gap (except the one in the Corolado River Basin) and lasted for 13-25 months. The maximum deficits account for 34%-56% of the seasonal water storage oscillations (i.e., difference between the maximum and minimum of the seasonal signal). One exception is the Amur River Basin (181%) primarily due to the fact that the TWSA signals in this basin are dominated by the long-term/sub-seasonal changes and the seasonal change is relatively small (Humphrey et al., 2016). We classify the drought intensity according to the $\mathrm{WSDI}_{\max}$ values and the thresholds used by US Drought Monitor (Table C2) (Svoboda et al., 2002).

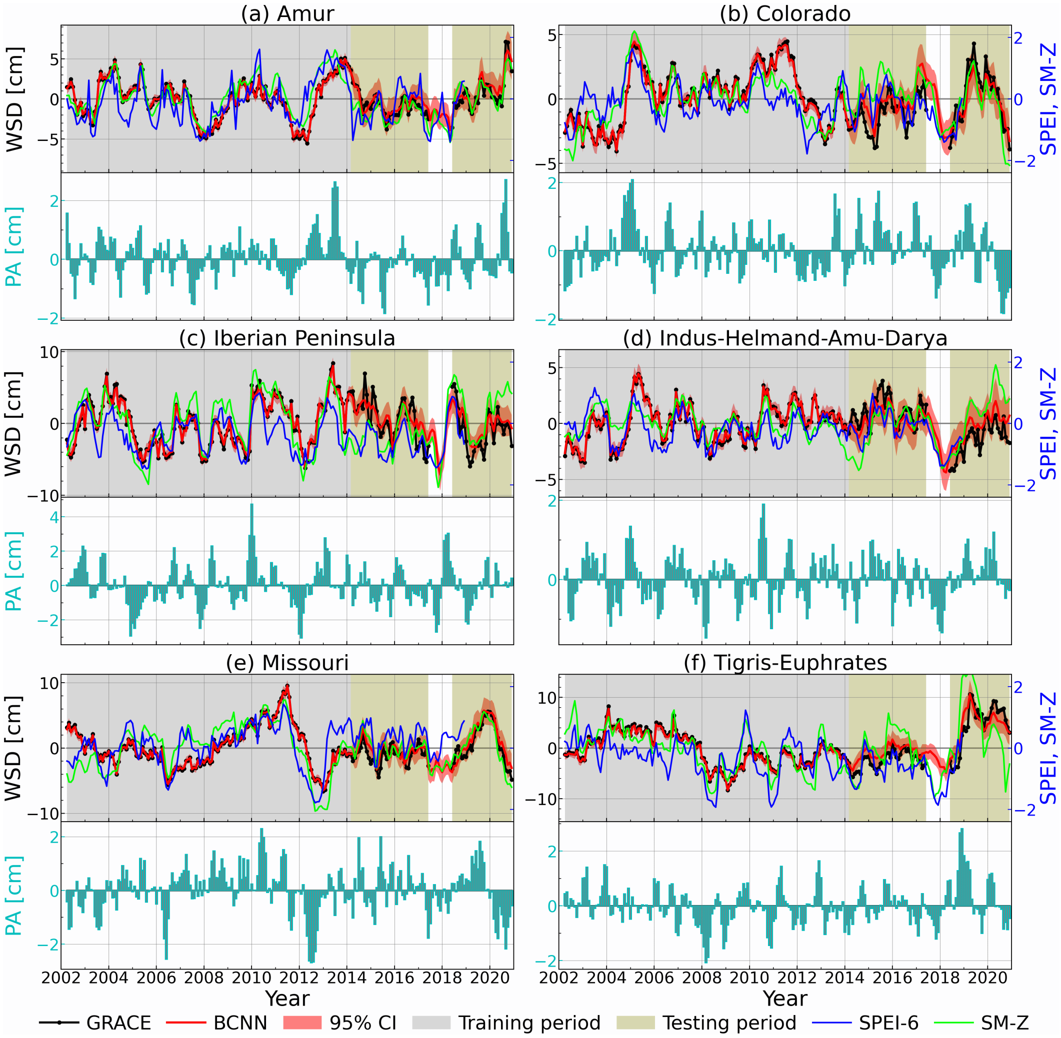

Figure 7. Basin-averaged time series of GRACE- and BCNN-derived WSD, SPEI-6, SM-Z, and PA in six regions/river basins. CI: confidential interval; PA: precipitation anomaly; SM-Z: standardized soil moisture; SPEI-6: 6-month standardized precipitation evapotranspiration index; WSD: water storage deficit.

The results indicate that the six basins experienced severe to exceptional droughts during the gap period (Table 1).

The complemented quantification of water storage deficits during the 2017-2018 GRACE and GRACE-FO gap allows for a full spatiotemporal characterization of the hydrological drought events since the GRACE era. Figure 8 illustrates the temporal evolution of percent area of different drought categories in the six basins since 2002. The missing data during the gap period are filled with BCNN's reconstructions, where the $(\mathrm{WSDI}_{\mathrm{BCNN}} - 2\sigma)$ values

Table 1. Characterization and Quantification of the Droughts in Six River Basins During the 2017-2018 GRACE and GRACE-FO Gap.

Basin (Area [km^2])	Time span	Duration [months]	$\mathrm{WSD_{max}} \pm \sigma$ [cm] (date)	$\overline{\mathrm{WSD}} \pm \sigma$ [cm]	$\frac{\mathrm{WSD_{max}}}{\mathrm{SC}}$ †	Maximum intensity ($\mathrm{WSDI_{max}}$)
Amur (2,251,276)	2017.3-2018.7	17	-3.9±0.8 (2018.4)	-1.7±0.83	-181.4%‡	D3 (-1.7)
Colorado (669,546)	2017.12-2019.1	14	-3.0±0.6 (2018.3)	-1.9±0.6	-46.5%	D2 (-1.5)
Iberian Peninsula (410,959)	2017.3-2018.3	13	-7.2±1.1 (2017.12)	-2.7±1.0	-54.5%	D4 (-2.5)
Indus-Helmand-Amu -Darya (2,258,034)	2017.9-2019.1	17	-4.4±0.8 (2018.4)	-2.2±0.8	-55.6%	D4 (-2.7)
Missouri (1,354,379)	2017.6-2019.1	20	-3.7±0.8 (2018.6)	-2.4±0.9	-38.2%	D2 (-1.3)
Tigris-Euphrates (602,127)	2016.10-2018.10	25	-4.9±0.9 (2018.4)	-1.8±0.9	-33.8%	D2 (-1.4)

† $\frac{\mathrm{WSD_{max}}}{\mathrm{SC}}$ denotes the ratio between the maximum water storage deficit (WSD) and the seasonal change (SC) of water storage.

‡ The TWSA signals in the Amur River Basin are dominated by the long-term/sub-seasonal changes, the SC is relatively small.

Note: $\mathrm{WSD_{max}}$ and $\overline{\mathrm{WSD}}$ are the maximum and mean of WSD during the drought, respectively, σ denotes one standard deviation of BCNN's reconstructions.

(σ denotes one standard deviation of BCNN's reconstruction) are utilized when classifying the drought to ensure a lower confidence level of $\sim$ 98%. It can be seen from Figure 8a that, for instance, some areas in the Amur River Basin started to occur drought since the middle of 2014. The drought areas then sequentially expanded and reached the maximum in April 2018. During the gap period, all regions of the river basins in Iberian Peninsula experienced D2 or more severe droughts (>70% regions experienced D4 drought in December 2017; Figure 8c). For the Missouri and Tigris-Euphrates River Basins, the 2017-2018 droughts may be partly related to the earlier ones since 2012 and 2008, respectively, as there have been regions experiencing droughts since then (Figures 8e and 8f).

It is worth mentioning that the maximum deficits in these basins all occurred during the gap period (July 2017-May 2018) except the Missouri River Basin (Table 1). A lack of awareness of this fact as well as that these droughts completely cover the gap period will lead to an underestimation of the drought severity and duration. The improved understanding

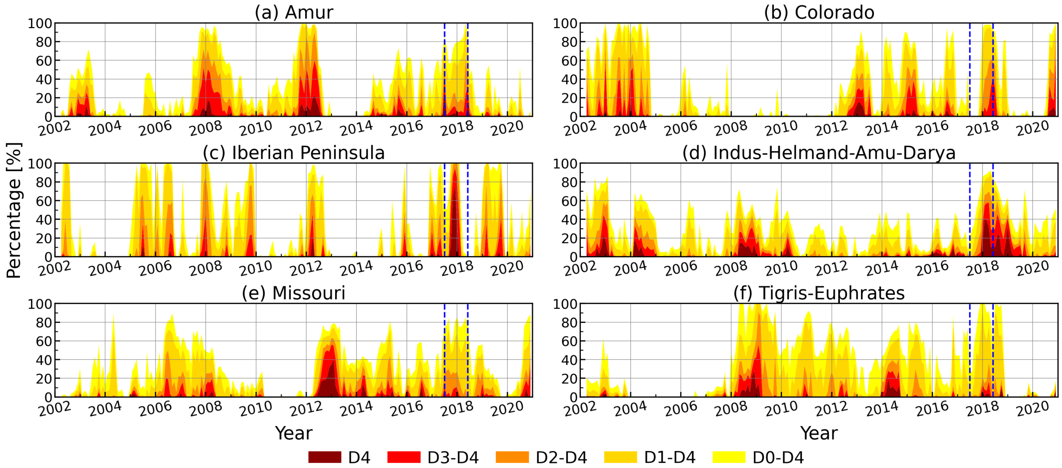

Figure 8. Temporal evolution of the percent area of different drought categories in six river basins since the GRACE era. The period between the two vertical dashed lines denotes the 2017-2018 GRACE and GRACE-FO gap.

toward the most recent 2017-2018 hydrological droughts in these regions, therefore, can better guide future drought prevention and mitigation policies.

3.3 Discussion

In terms of the maximum deficit, the droughts covering the gap period in the Iberian Peninsula and Indus-Helmand-Amu-Darya River Basin (mostly in Pakistan and Afghanistan) are the most severe on record since the launch of GRACE (Figures 7d and 7e, 8c and 8d). However, unlike the Iberian Peninsula where the water storage recovered quickly to a far above-normal condition, the drought recovery in the Indus-Helmand-Amu-Darya River Basin was slow and the water storage remained at a near- or below-normal level in post-drought years, leading to greater impacts.

To investigate the major causes for the six droughts and the major factors influencing drought recovery, the time series of precipitation anomaly and BCNN- and GRACE-derived WSDs are also compared in Figure 7 (the global precipitation anomaly fields during the 2017-2018 gap are shown in Figure S2). The precipitation is a weighted ensemble of the eight products listed in Table C1 by the generalized three-cornered hat method. We perform 3-month running mean on the precipitation time series to smooth out short-term fluctuations. The anomaly is then respectively calculated for each month of a year. The WSD and precipitation anomaly time series are visually correlated well with each other that their peaks/valleys are generally consistent, though there may be a time lag. The six basins all received below-normal precipitation during the gap period, implying that these droughts are primarily caused by continuous precipitation deficits. The Iberian Peninsula's wet 2018 summer with continuous and large precipitation surpluses contributed to a rapid drought recovery (Figure 7c). On the contrary, although the Indus-Helmand-Amu-Darya River Basin experienced a relatively wet 2018 winter, the precipitation surpluses were obviously not sufficient to recharge the large deficits during the drought, leading to a slow recovery (Figures 7d and 8d). The results are further validated in Figure 9, which compares the BCNN-derived WSD time series with the cumulative precipitation anomaly since the beginning of drought. The cumulant is compared here due to the fact that the drought propagation/recovery relies generally on the continuous precipitation shortage/surplus. A cumulant can rule out the influence of local signal fluctuations and is thus more suitable for illustrating the lasting impact. The two time series show a good correlation ($R \geq 0.77$ for all basins except the Amur River Basin ($R = 0.26$), but their fluctuations show a similar pattern in this basin) that the water storage recovery is generally consistent with the accumulation rate of precipitation surpluses. The faster recovery here indicates a higher drought resilience.

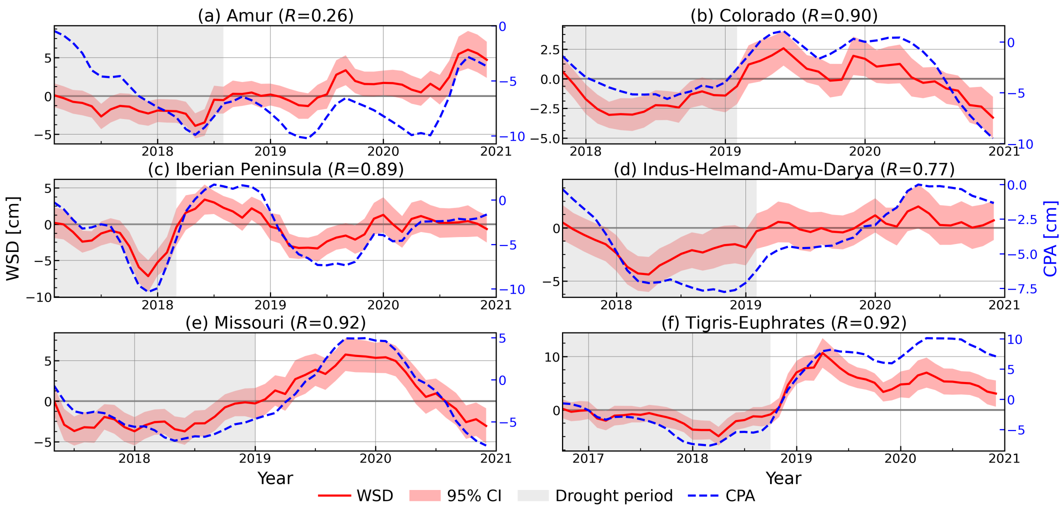

Figure 9. Basin-averaged time series of BCNN-derived WSD and cumulative precipitation anomaly (CPA) since the beginning of drought covering the 2017-2018 GRACE and GRACE-FO gap in six river basins. R denotes the correlation coefficient between the two time series. CI: confidential interval; WSD: water storage deficit.

In addition to the hydrological droughts identified and quantified above, our gap-filling results enable a full characterization of previously analyzed ones. For example, the2018-2019 Central European drought partly quantified in Boergens, Güntner, et al. (2020) and the 2017-2019 drought in Murray-Darling River Basin recorded in Deb et al. (2020) and Kumar et al. (2021), which was the most severe on record and led to the serious 2019–2020 bushfires. It can be seen from Figure 10a that prior to the 2018-2019 Central European drought, the above-normal precipitation in the latter half of 2017 resulted in an increased water storage. Then the storage declined rapidly due to continuous precipitation shortages since the beginning of 2018, thereby leading to the summer drought. The Murray-Darling Basin's 2017-2019 drought was also caused by continuous below-normal precipitation since the beginning of 2017. The drought started in February 2018 in terms of water storage deficits, which reached the maximum (-8.7 cm) in December 2019 (Figure 10b). The continuous precipitation shortages and water storage losses led to the long-lasting drought and eventually the very serious 2019–2020 bushfires (Deb et al., 2020; Kumar et al., 2021).

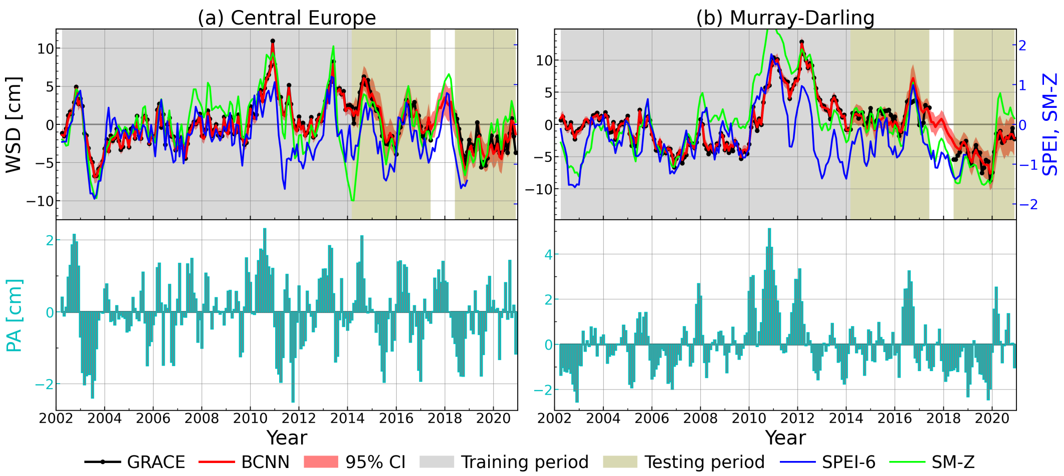

Figure 10. Basin-averaged time series of GRACE- and BCNN-derived WSD, SPEI-6, SM-Z, and precipitation anomaly (PA) in (a) Central Europe (45°N-55°N, 4°E-24°E) and (b) Murray-Darling River Basin. CI: confidential interval; SM-Z: standardized soil moisture; SPEI-6: 6-month standardized precipitation evapotranspiration index; WSD: water storage deficit.

Note that due to the potential inability of the hydroclimatic predictors in fully reflecting the anthropogenic impacts on water storage dynamics, the noises associated with hydroclimatic data, and BCNN's predictive uncertainties, the BCNN-filled WSD estimates are inevitably subject to uncertainties or even biases. For example, the GRACE WSD time series indicate that the Ganges-Brahmaputra River Basin experienced a significant water storage deficit during June-October 2018. However, the two drought indicators (i.e., SPEI-6 and SM-Z) all imply an opposite wet trend in this period, leading to biases in the BCNN estimations (Figure 5d). This limitation is expected to be alleviated with more GRACE training data available, higher-quality hydroclimatic data, the consideration of additional informative predictors related to human activities (e.g., the dataset of reservoir operations (Steyaert et al., 2022)), and future advances in deep learning.

4 Conclusions

The hydrological drought events of 2017-2018, whose duration is overlapped with the gap period (July 2017-May 2018) between the GRACE and GRACE-FO missions, are still largely unknown and poorly characterized. In this work, we bridge this research gap by reconstructing the missing TWSA fields from hydroclimatic inputs with a newly developed BCNN deep learning method. The results show that BCNN provides higher-quality reconstructions of the TWSA and water storage deficit signals to fill the data gap compared to previous studies. The reconstructed and existing signals are then utilized together to identify regions experiencing droughts and characterize the spatiotemporal evolution of water storage losses during the 2017-2018 gap. Our major new finding is that many regions in the northern mid-latitudes experienced consecutive water deficits in this period. The results show a favorable consistency with two commonly used drought indices (SPEI-6 and SM-Z), though they only partly/qualitatively reflect the water availability from respective hydrological components. At the regional scale, the droughts in six regions/river basins within the identified regions are partly/completely overlapped with the gap period and last for 13-25 months, among which the ones in Indus-Helmand-Amu-Darya River Basin (mostly in Pakistan and Afghanistan) and Iberian Peninsula are the most severe on GRACE record with the maximum deficits (-4.4 ± 0.8 cm and -7.2 ± 1.1 cm, respectively) being over 50% of the seasonal water storage variations. These droughts are resulted predominantly from continuous precipitation deficiency. The recovery rate of these droughts is sensitive to the accumulation rate of precipitation surpluses in the late- and post-drought periods, with a faster recovery when the accumulated surpluses increase rapidly and their correlation coefficient being up to 0.92.

The maximum deficits of the 2017-2018 droughts in five of the six mentioned basins occurred during the gap period and the drought duration completely overlaps with this period. Lacking awareness of these facts will lead to an underestimation of drought duration and severity. Therefore, the improved understanding toward the recent 2017-2018 hydrological droughts, together with the findings in previous studies from the perspective of meteorological/agricultural droughts, can better guide future drought prevention and mitigation policies, allows for a full characterization and a joint assessment of droughts since the GRACE mission, and further investigation of insights into the causes/impacts. Additionally, the improved TWSA reconstructions by BCNN reliably maintain the data continuity, enhance data consistency, and therefore enable a full use/analysis of time series data. The proposed BCNN method can also be flexibly extended to other hydroclimatic applications involving learning complex mappings due to its generic nature.

Data Availability Statement

The BCNN-derived TWSA, WSD, and WSDI data generated in this study are available at `https://zenodo.org/record/5336992`. The codes of BCNN are available at `https://github.com/njujinchun/BCNN4GRACE`. Other datasets used are available at the following links:

- GRACE SH: `https://podaac-tools.jpl.nasa.gov/drive/files/GeodeticsGravity`
- GRACE mascon: `http://www2.csr.utexas.edu/grace/RL06_mascons.html` (CSR), `https://earth.gsfc.nasa.gov/geo/data` (GSFC), `http://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons` (JPL)
- COST-G and GravIS: `http://gravis.gfz-potsdam.de/land`
- CPC soil moisture: `https://psl.noaa.gov/data/gridded/data.cpcsoil.html`
- CRU: `https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.05`
- ERA5-land: `https://doi.org/10.24381/cds.68d2bb30`
- FLDAS: `https://doi.org/10.5067/5NHC22T9375G`
- GHCN-CAMS: `https://psl.noaa.gov/data/gridded/data.ghcncams.html`
- GLDAS: `https://doi.org/10.5067/FOUXNLXFAZNY` (CLSM), `https://doi.org/10.5067/SXAVCZFAQLNO` (Noah), `https://doi.org/10.5067/VWTH7S6218SG` (VIC)
- GLEAM: submit a request on `https://www.gleam.eu`
- GPCP: `https://www.ncei.noaa.gov/data`
- MERRA-2: `https://doi.org/10.5067/8S35XF81C28F`
- MSWEP: submit a request on `http://www.gloh2o.org/mswep`

- MSWX: submit a request on `http://www.gloh2o.org/mswep`
- NLDAS: `https://doi.org/10.5067/NOXZSD0Z6JGD`
- PRECL: `https://psl.noaa.gov/data/gridded/data.precl.html`
- SPEI: `https://spei.csic.es/database.html`
- TerraClimate: `http://www.climatologylab.org/terraclimate.html`

Acronyms

BCNN Bayesian Convolutional Neural Network

CI Confidence Interval

GRACE(-FO) Gravity Recovery and Climate Experiment (Follow-On)

NRMSE normalized root mean squared error

NSE Nash-Sutcliffe Efficiency coefficient

R Correlation Coefficient

SM-Z Standardized Soil Moisture

SPEI-6 Six-month Standardized Precipitation Evapotranspiration Index

TWSA Terrestrial Water Storage Anomaly

WSD Water Storage Deficit

WSDI Water Storage Deficit Index

Acknowledgments

We thank the providers of the data used in our work. This work was funded by the National Key Research and Development Program of China (2018YFC1800600), National Natural Science Foundation of China (41730856, 41874095, 41977157, 42002248, 42004073), China Postdoctoral Science Foundation (2020M681550), and Jiangsu Planned Projects for Postdoctoral Research Funds (2020Z133). S.M. would like to thank Dr. Yinhao Zhu from Qualcomm AI Research and Dr. Xiaogang He from National University of Singapore for their valuable suggestions. We are also grateful to the Editors and three anonymous reviewers for their constructive comments and suggestions.

Appendix A BCNN Architecture Design and Training

The BCNN network architecture is U-shaped as shown in Figure 1. Compared to the U-Net in previous studies (Ronneberger et al., 2015; A. Y. Sun et al., 2019; Mo et al., 2022), we propose to use a modified residual-in-residual dense block (RRDB) (Wang et al., 2018) as the basic building block of BCNN for higher performance. RRDB introduces connections between non-adjacent layers to enhance information flow through networks (Huang et al., 2017) and residual learning to ease the training of deep networks (K. He et al., 2016). The combination of the two strategies can substantially strengthen the network performance in learning complex mappings (Wang et al., 2018; Mo et al., 2020). Compared to the original RRDB structure (Wang et al., 2018), we further incorporate the spatial and channel attention module (SCAM) (Fu et al., 2019) for the effective propagation of informative information through networks. More specifically, the spatial attention module outputs a $h \times w$ weight matrix assigning to the $h \times w$ pixels of the extracted feature maps to tell the network where to attend. Similarly, the channel attention module outputs n_{f} weights assigning to the n_{f} extracted feature maps to tell the network which maps to attend. The combination of these strategies in BCNN can substantially reduce the risk of overfitting and enables the network to automatically focus on informative features extracted from multi-source data for the image-to-image regression task. The RRDB and SCAM architectures are depicted in Figure A1.

The network predictions are inevitably associated with uncertainties. To quantify the predictive uncertainties, the BCNN network is trained with a SVGD (stein variational gradient descent) Bayesian inference algorithm (Q. Liu & Wang, 2016; Zhu & Zabaras, 2018) to obtain a set of parameter estimates, $\{\mathbf{w}_i\}_{i=1}^{N_S}$. The predictive uncertainties for an arbitrary input $\mathbf{x}$ can then be computed based on the N_S predictions ($\hat{\mathbf{y}}^{(i)} = \boldsymbol{\eta}(\mathbf{x}, \mathbf{w}_i)$, $i = 1, \ldots, N_S$). We set $N_S = 30$ as suggested in Zhu and Zabaras (2018). For more details regarding the SVGD Bayesian training strategy, one can refer to Q. Liu and Wang (2016) and Zhu and Zabaras (2018). The network is trained for 200 epochs with an initial learning rate of 0.0025 and a batch size of 12.

Appendix B Accuracy Improvement by Consideration of Additional Predictors

To illustrate the benefits of considering more predictors in BCNN in addition to the three commonly used ones (i.e., sTWSA, P, and T) in previous studies (e.g., A. Y. Sun et

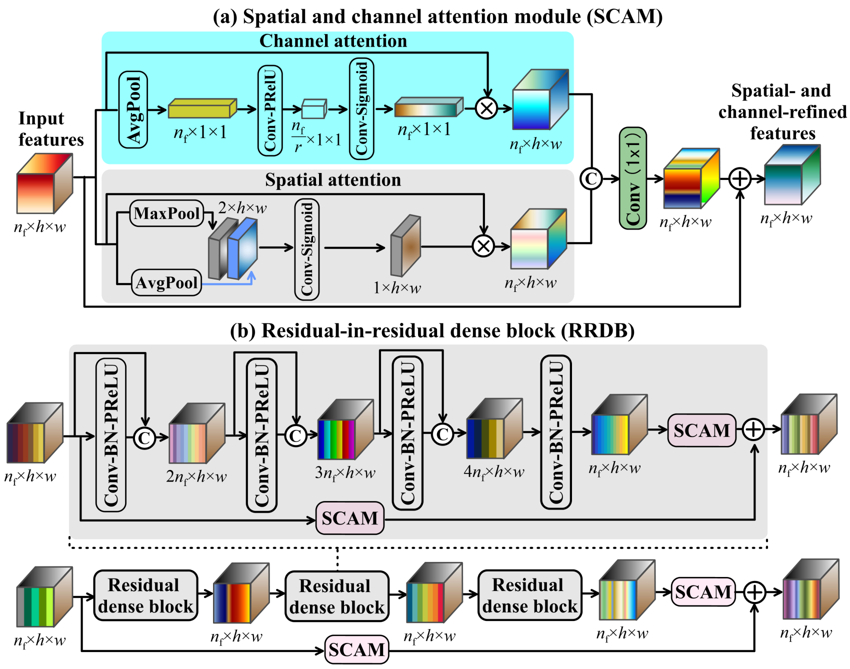

Figure A1. Diagrams of the (a) spatial and channel attention module (SCAM) and (b) residual-in-residual dense block (RRDB). $n_\mathrm{f} \times h \times w$ denotes n_f feature maps (images) with resolution $h \times w$. SCAM utilizes the outputs of average-pooling (AvgPool) and/or max-pooling (MaxPool) to produce attention maps with weight values in [0, 1] to indicate "what" and "where" to attend. $\oplus$, ⓒ, and $\otimes$ denote the addition, concatenation, and multiplication operations, respectively. Each RRDB contains three residual dense blocks, which introduce connections between non-adjacent layers to enhance information flow. Sigmoid and PReLU denote activation functions; BN: batch normalization; Conv: convolution.

al., 2019; Z. Sun et al., 2020), the testing NSE accuracy for the TWSA signals when all seven predictors and the three common predictors are respectively considered in BCNN is compared in Figure B1. It can be seen that the consideration of seven predictors achieves an accuracy improvement of 0.05∼0.1 in most regions (Figures B1c and B1d). Note that it is flexible and straightforward for BCNN to consider more predictors without requiring to modify the network architecture nor improving the training complexity. The results suggest that the commonly used three predictors are the most informative ones for TWSA reconstruction. The other four predictors are considered with the purpose of further improving the reconstruction accuracy.

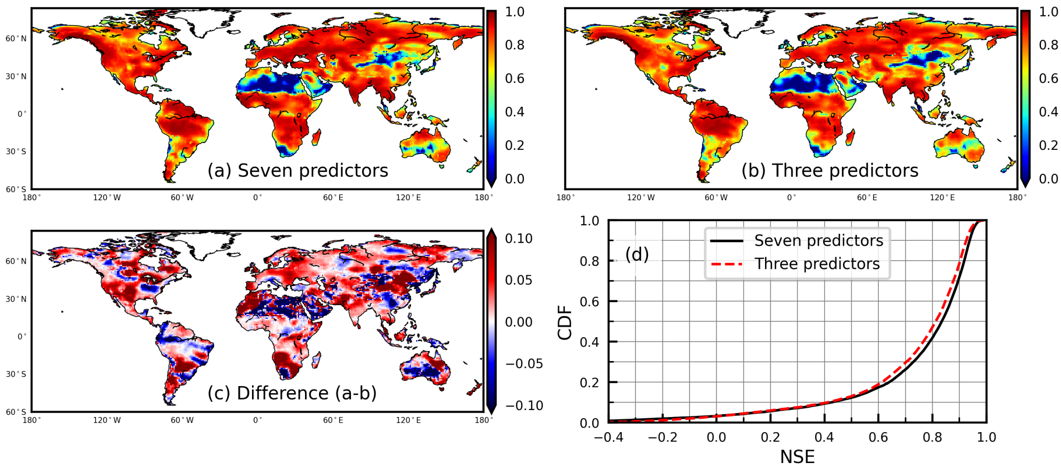

Figure B1. The testing NSE accuracy for the TWSA signals when (a) all seven predictors and (b) the commonly used three predictors (i.e., sTWSA, P, and T) are considered. (d) The accuracy differences (i.e., a-b). (d) Cumulative distribution function (CDF) of the NSE values in (a) and (b).

Appendix C This section provides some tables and figures which support the discussion of this article

Table C1. Data source candidates considered for the seven predictors, including the simulated/reanalyzed terrestrial water storage anomaly (sTWSA), cumulative water storage change (CWSC), precipitation (P), temperature (T), climatic water balance (CWB), cumulative precipitation anomaly (CPA), and cumulative CWB (CCWB). The selection criterion is the NSE or $|R|$ metric. NSE: Nash-Sutcliffe efficiency coefficient; $|R|$: absolute correlation coefficient.

Predictor	Data source candidates	Selection criterion[a]
sTWSA	ERA5-land, GLDAS v2.1 (Noah, VIC, CLSM) FLDAS, NLDAS-2 Noah, MERRA-2, ensemble[b]	NSE
CWSC	ERA5-land, GLDAS v2.1 (Noah, VIC, CLSM), FLDAS NLDAS-2 Noah, MERRA-2, TerraClimate, ensemble	NSE
P	ERA5-land, GLDAS, GPCP, MERRA-2, TerraClimate MSWEP, PRECL,CRU, ensemble	$\|R\|$
T	ERA5-land, GLDAS, FLDAS, GHCN-CAMS, CRU MSWX, ensemble	$\|R\|$
CWB	ERA5-land, GLDAS Noah, CRU, TerraClimate MSWEP (P)&GLEAM (PET), ensemble	$\|R\|$
CPA	Same as P	$\|R\|$
CCWB	Same as CWB	$\|R\|$

[a] The linear trends are removed from the time series of GRACE TWSA and predictor data before calculating the criterion.

[b] The ensemble indicates an average ensemble of all individual sources. The trends are removed before calculating the ensemble.

Table C2. Drought condition classification based on the water storage deficit index (WSDI). The classification scheme is consistent with the US Drought Monitor (Svoboda et al., 2002).

Category	Description	WSDI
WD	Near normal	(-0.5, 0.5)
D0	Abnormally dry	(-0.8, -0.5]
D1	Moderate drought	(-1.3, -0.8]
D2	Severe drought	(-1.6, -1.3]
D3	Extreme drought	(-2, -1.6]
D4	Exceptional drought	(-∞, -2]

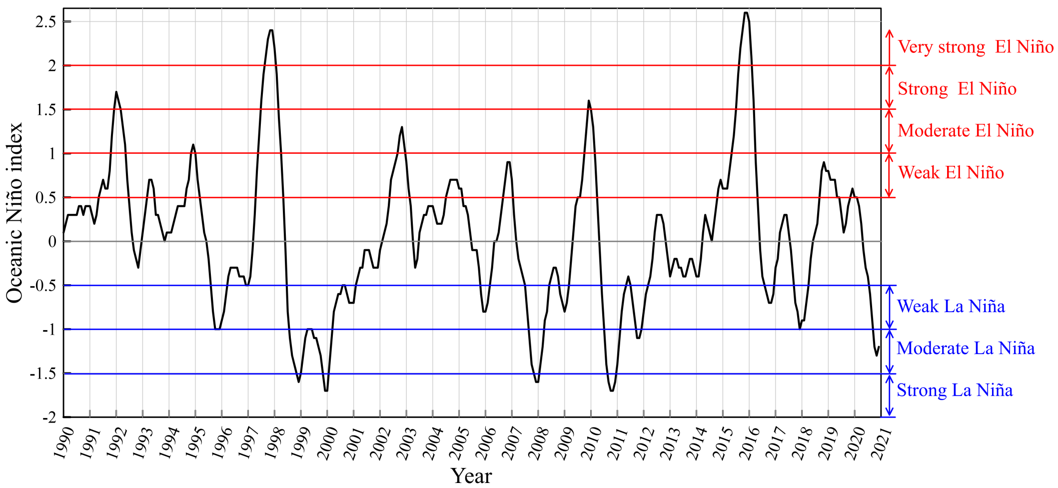

Figure C1. The time series of 3-month running mean oceanic Niño index since 1990. The data are available at `https://ggweather.com/enso/oni.htm`.

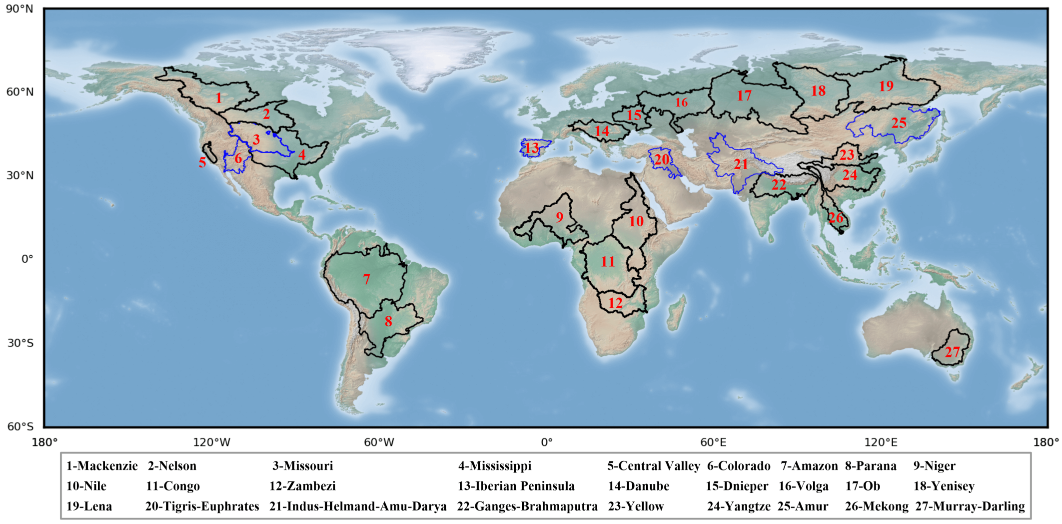

Figure C2. Regions and river basins considered for result analysis. The regions/river basins shown in blue are those experienced hydrological droughts during the 2017-2018 gap period and analyzed in section 3.2. The shapefiles are downloaded from `https://datacatalog.worldbank.org/dataset/major-river-basins-world`.

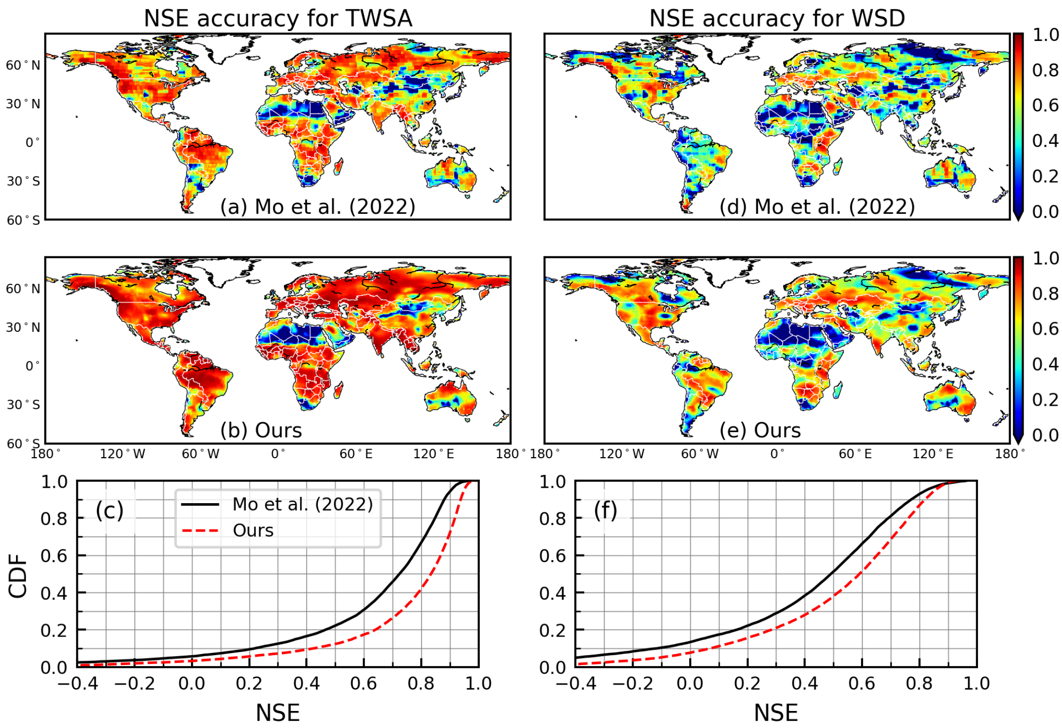

Figure C3. Comparison of the testing NSE accuracy obtained in previous study (Mo et al., 2022) and this study for the TWSA (a-c) and the water storage deficit (WSD) (d-f) signals. The major differences between Mo et al. (2022) and this study are as follows: (1) Mo et al. (2022) used only four predictors (sTWSA, CWSC, P, and T) derived from the ERA5 land dataset without utilizing the data source selection strategy; (2) the JPL GRACE mascon product was used in Mo et al. (2022), while a weighted average GRACE product is used here; (3) The BCNN method employed in this study has been improved (see Appendix A).

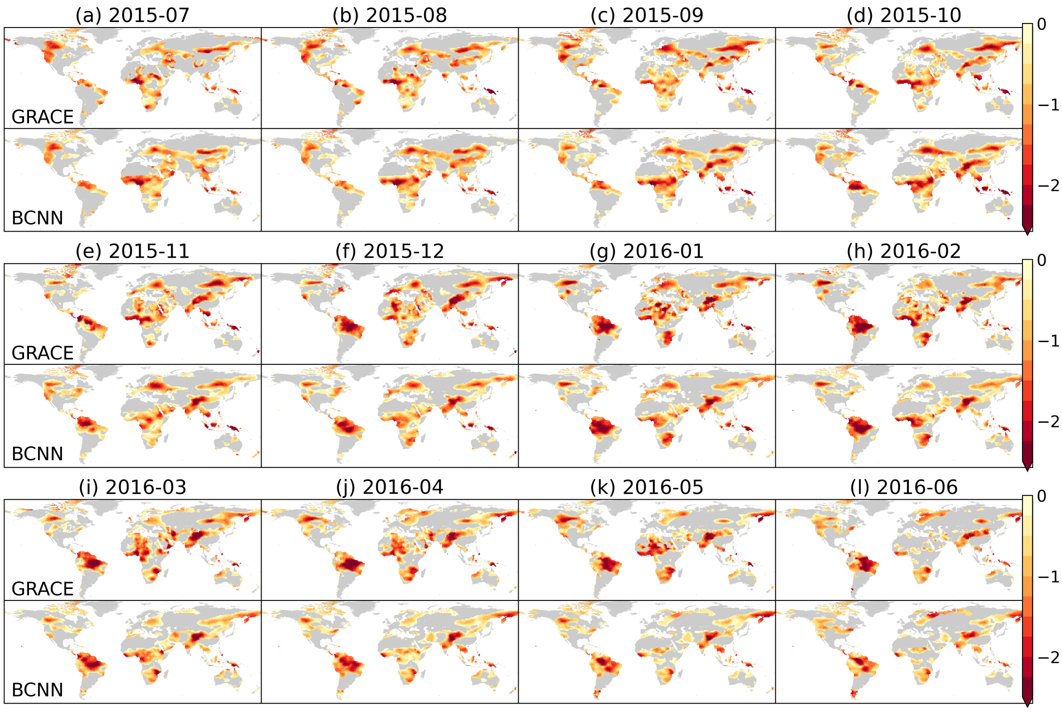

Figure C4. Comparison of the drought regions identified by the GRACE- and BCNN-derived WSDI indices during the 2015-2016 El Niño event (Figure C1). In each column, the upper and lower plots show GRACE's and BCNN's results for the same month, respectively. WSDI: water storage deficit index.

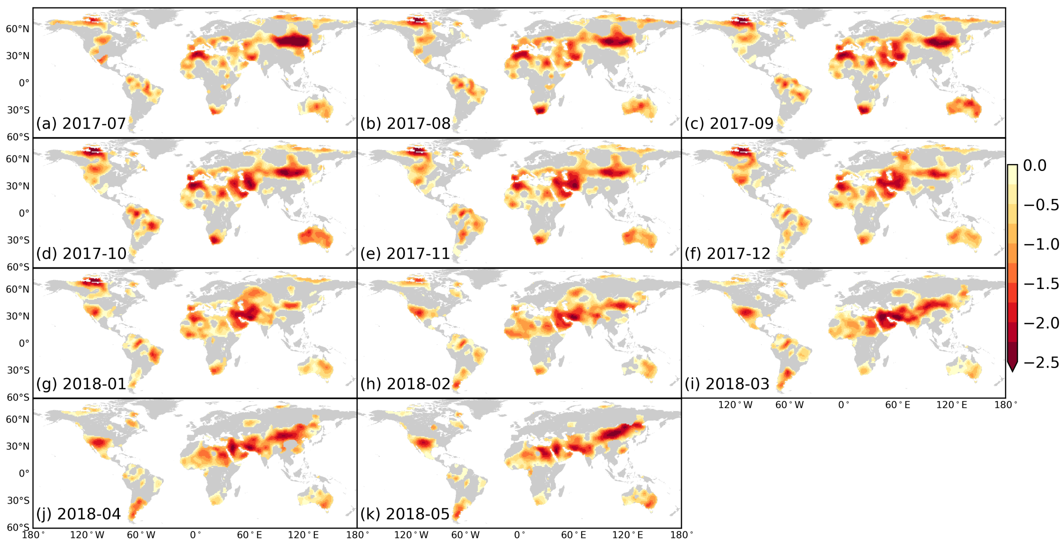

Figure C5. Drought regions identified by the SPEI-6 index during the GRACE and GRACE-FO gap (July 2017-May 2018). SPEI-6: 6-month standardized precipitation evapotranspiration index.

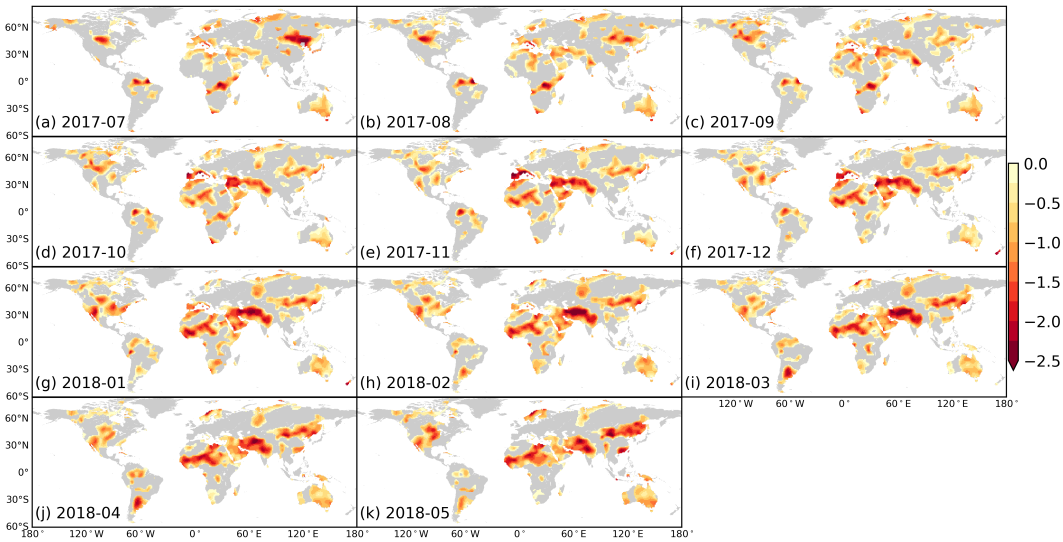

Figure C6. Drought regions identified by the SM-Z index during the GRACE and GRACE-FO gap (July 2017-May 2018). SM-Z: standardized soil moisture.

References

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., & Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, *53*(2), 452-480. doi: https://doi.org/10.1002/2014RG000456

Ali, S., Wang, Q., Liu, D., Fu, Q., Mafuzur Rahaman, M., Abrar Faiz, M., & Jehanzeb Masud Cheema, M. (2022). Estimation of spatio-temporal groundwater storage variations in the Lower Transboundary Indus Basin using GRACE satellite. *Journal of Hydrology*, *605*, 127315. doi: https://doi.org/10.1016/j.jhydrol.2021.127315

Boergens, E., Dobslaw, H., Dill, R., Thomas, M., Dahle, C., Murböck, M., & Flechtner, F. (2020). Modelling spatial covariances for terrestrial water storage variations verified with synthetic GRACE-FO data. *International Journal on Geomathematics*, *11*(1), 1-25. doi: https://doi.org/10.1007/s13137-020-00160-0

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. doi: https://doi.org/10.1029/2020GL087285

Chen, J. L., Wilson, C. R., Tapley, B. D., Longuevergne, L., Yang, Z. L., & Scanlon, B. R. (2010). Recent La Plata basin drought conditions observed by satellite gravimetry. *Journal of Geophysical Research: Atmospheres*, *115*(D22). doi: https://doi.org/10.1029/2010JD014689

Deb, P., Moradkhani, H., Abbaszadeh, P., Kiem, A. S., Engström, J., Keellings, D., & Sharma, A. (2020). Causes of the widespread 2019–2020 Australian bushfire season. *Earth's Future*, *8*(11), e2020EF001671. doi: https://doi.org/10.1029/2020EF001671

Du, J., Shi, X., Mo, S., Kang, X., & Wu, J. (2022). Deep learning based optimization under uncertainty for surfactant-enhanced DNAPL remediation in highly heterogeneous aquifers. *Journal of Hydrology*, *608*, 127639. doi: https://doi.org/10.1016/j.jhydrol.2022.127639

Eicker, A., Forootan, E., Springer, A., Longuevergne, L., & Kusche, J. (2016). Does grace see the terrestrial water cycle "intensifying"? *Journal of Geophysical Research: Atmospheres*, *121*(2), 733-745. doi: https://doi.org/10.1002/2015JD023808

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

Forootan, E., Kusche, J., Loth, I., Schuh, W.-D., Eicker, A., Awange, J., … Shum, C. K. (2014). Multivariate prediction of total water storage changes over west africa from multi-satellite data. *Surveys Geophysics*, *35*, 913–940. doi: https://doi.org/10.1007/s10712-014-9292-0

Forootan, E., Schumacher, M., Mehrnegar, N., Bezděk, A., Talpe, M. J., Farzaneh, S., … Shum, C. K. (2020). An iterative ICA-based reconstruction method to produce consistent time-variable total water storage fields using GRACE and Swarm satellite data. *Remote Sensing*, *12*(10), 1639. doi: https://doi.org/10.3390/rs12101639

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).* Long Beach, CA.

Haile, G. G., Tang, Q., Sun, S., Huang, Z., Zhang, X., & Liu, X. (2019). Droughts in East Africa: Causes, impacts and resilience. *Earth-Science Reviews*, *193*, 146-161. doi: https://doi.org/10.1016/j.earscirev.2019.04.015

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, *56*(1), 108-141. doi: https://doi.org/10.1002/2016RG000549

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Las Vegas, NV.

He, X., Estes, L., Konar, M., Tian, D., Anghileri, D., Baylis, K., … Sheffield, J. (2019). Integrated approaches to understanding and reducing drought impact on food security across scales. *Current Opinion in Environmental Sustainability*, *40*, 43-54. (System dynamics and sustainability) doi: https://doi.org/10.1016/j.cosust.2019.09.006

He, X., Feng, K., Li, X., Craft, A. B., Wada, Y., Burek, P., … Sheffield, J. (2019). Solar and wind energy enhances drought resilience and groundwater sustainability. *Nature Communications*, *10*, 1-8. doi: https://doi.org/10.1038/s41467-019-12810-5

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

Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: A reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data*, *11*(3), 1153–1170. doi: https://doi.org/10.5194/essd-11-1153-2019

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. doi: https://doi.org/10.1007/s10712-016-9367-1

Kumar, S. V., Holmes, T., Andela, N., Dharssi, I., Vinodkumar, Hain, C., ... Getirana, A. (2021). The 2019–2020 Australian drought and bushfires altered the partitioning of hydrological fluxes. *Geophysical Research Letters*, *48*(1), e2020GL091411. doi: https://doi.org/10.1029/2020GL091411

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444. doi: https://doi.org/10.1038/nature14539

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

Li, F., Kusche, J., Chao, N., Wang, Z., & Löcher, A. (2021). Long-term (1979-present) total water storage anomalies over the global land derived by reconstructing GRACE data. *Geophysical Research Letters*, *48*(8), e2021GL093492. doi: https://doi.org/10.1029/2021GL093492

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, X., Feng, X., Ciais, P., Fu, B., Hu, B., & Sun, Z. (2020). GRACE satellite-based drought index indicating increased impact of drought over major basins in China during 2002–2017. *Agricultural and Forest Meteorology*, *291*, 108057. doi: https://doi.org/10.1016/j.agrformet.2020.108057

Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., ... Longuevergne, L. (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. doi: https://doi.org/10.1016/j.rse.2017.02.011

Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N., & Save, H. (2013). GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophysical Research Letters*, *40*(13), 3395-3401. doi: https://doi.org/10.1002/grl.50655

Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., … Chen, L. (2014). Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. *Remote Sensing of Environment*, *155*, 145-160. doi: https://doi.org/10.1016/j.rse.2014.08.006

Mehrnegar, N., Jones, O., Singer, M. B., Schumacher, M., Jagdhuber, T., Scanlon, B. R., … Forootan, E. (2021). Exploring groundwater and soil water storage changes across the conus at 12.5 km resolution by a bayesian integration of grace data into w3ra. *Science of The Total Environment*, *758*, 143579. doi: https://doi.org/10.1016/j.scitotenv.2020.143579

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of Hydrology*, *391*(1), 202-216. doi: https://doi.org/10.1016/j.jhydrol.2010.07.012

Mishra, V., Thirumalai, K., Jain, S., & Aadhar, S. (2021). Unprecedented drought in South India and recent water scarcity. *Environmental Research Letters*, *16*(5), 054007. doi: https://doi.org/10.1088/1748-9326/abf289

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019). Deep autoregressive neural networks for high-dimensional inverse problems in groundwater contaminant source identification. *Water Resources Research*, *55*(5), 3856-3881. doi: https://doi.org/10.1029/2018WR024638

Mo, S., Zabaras, N., Shi, X., & Wu, J. (2020). Integration of adversarial autoencoders with residual dense convolutional networks for estimation of non-Gaussian hydraulic conductivities. *Water Resources Research*, *56*(2), e2019WR026082. doi: https://doi.org/10.1029/2019WR026082

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

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. doi: https://doi.org/10.1029/2018WR023528

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204. doi: https://doi.org/10.1038/s41586-019-0912-1

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234–241). Cham: Springer International Publishing.

Sakumura, C., Bettadpur, S., & Bruinsma, S. (2014). Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models. *Geophysical Research Letters*, *41*(5), 1389-1397. doi: https://doi.org/10.1002/2013GL058632

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

Shah, D., & Mishra, V. (2020). Drought onset and termination in India. *Journal of Geophysical Research: Atmospheres*, *125*(15), e2020JD032871. doi: https://doi.org/10.1029/2020JD032871

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558-8593. doi: https://doi.org/10.1029/2018WR022643

Steyaert, J. C., Condon, L. E., W.D. Turner, S., & Voisin, N. (2022). ResOpsUS, a dataset of historical reservoir operations in the contiguous United States. *Scientific Data*, *9*, 1-8. doi: https://doi.org/10.1038/s41597-022-01134-7

Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2020). Reconstruction of GRACE total water storage through automated machine learning. *Water Resources Research*, *57*(2), e2020WR028666. doi: https://doi.org/10.1029/2020WR028666

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. doi: https://doi.org/10.1029/2018WR023333

Sun, Z., Long, D., Yang, W., Li, X., & Pan, Y. (2020). Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resources Research*, *56*(4), e2019WR026250. doi: https://doi.org/10.1029/2019WR026250

Sun, Z., Zhu, X., Pan, Y., Zhang, J., & Liu, X. (2018). Drought evaluation using the

GRACE terrestrial water storage deficit over the Yangtze River Basin, China. *Science of The Total Environment*, *634*, 727-738. doi: https://doi.org/10.1016/j.scitotenv.2018.03.292

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... Stephens, S. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181-1190. doi: https://doi.org/10.1175/1520-0477-83.8.1181

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503-505. doi: 10.1126/science.1099192

Theron, S., Archer, E., Midgley, S., & Walker, S. (2021). Agricultural perspectives on the 2015-2018 Western Cape drought, South Africa: Characteristics and spatial variability in the core wheat growing regions. *Agricultural and Forest Meteorology*, *304-305*, 108405. doi: https://doi.org/10.1016/j.agrformet.2021.108405

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. doi: https://doi.org/10.1002/2014GL059323

van den Dool, H., Huang, J., & Fan, Y. (2003). Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *Journal of Geophysical Research: Atmospheres*, *108*(D16). doi: https://doi.org/10.1029/2002JD003114

Van Loon, A. F. (2015). Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water*, *2*(4), 359-392. doi: https://doi.org/10.1002/wat2.1085

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, *23*(7), 1696-1718. doi: https://doi.org/10.1175/2009JCLI2909.1

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... Change Loy, C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV).* Munich, Germany.

West, H., Quinn, N., & Horswell, M. (2019). Remote sensing for drought monitoring & impact assessment: Progress, past challenges and future opportunities. *Remote Sensing of Environment*, *232*, 111291. doi: https://doi.org/10.1016/j.rse.2019.111291

Wilhite, D. A., & Glantz, M. H. (1985). Understanding the drought phenomenon: The role of definitions. *Water International*, *10*(3), 111-120. doi: https://doi.org/10.1080/

02508068508686328

Xu, L., Chen, N., Moradkhani, H., Zhang, X., & Hu, C. (2020). Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis data sets. *Water Resources Research*, *56*(3), e2019WR026444. doi: https://doi.org/10.1029/2019WR026444

Yan, X., Zhang, B., Yao, Y., Yang, Y., Li, J., & Ran, Q. (2021). GRACE and land surface models reveal severe drought in eastern China in 2019. *Journal of Hydrology*, *601*, 126640. doi: https://doi.org/10.1016/j.jhydrol.2021.126640

Yi, S., & Sneeuw, N. (2021). Filling the data gaps within GRACE missions using singular spectrum analysis. *Journal of Geophysical Research: Solid Earth*, *126*(5), e2020JB021227. doi: https://doi.org/10.1029/2020JB021227

Yin, Z., Xu, Y., Zhu, X., Zhao, J., Yang, Y., & Li, J. (2021). Variations of groundwater storage in different basins of China over recent decades. *Journal of Hydrology*, *598*, 126282. doi: https://doi.org/10.1016/j.jhydrol.2021.126282

Zhao, M., A, G., Velicogna, I., & Kimball, J. S. (2017). Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes. *Journal of Climate*, *30*(16), 6297-6308. doi: https://doi.org/10.1175/JCLI-D-16-0458.1

Zhu, Y., & Zabaras, N. (2018). Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. *Journal of Computational Physics*, *366*, 415-447. doi: https://doi.org/10.1016/j.jcp.2018.04.018