Filling the gap between GRACE-and GRACE-FO-derived terrestrial water storage anomalies with Bayesian convolutional neural networks

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

[1]Nanjing University
[2]China University of Geosciences Faculty of Information Engineering
[3]Chinese Academy of Sciences
[4]Nanjing Hydraulic Research Institute

November 28, 2022

Abstract

There is an approximately one-year observation gap of terrestrial water storage anomalies (TWSAs) between the Gravity Recovery and Climate Experiment (GRACE) satellite and its successor GRACE Follow-On (GRACE-FO). This poses a challenge for water resources management, as discontinuity in the TWSA observations may introduce significant biases and uncertainties in hydrological model predictions and consequently mislead decision making. To tackle this challenge, a Bayesian convolutional neural network (BCNN) is proposed in this study to bridge this gap using climatic data as inputs. Enhanced by integrating recent advances in deep learning, BCNN can efficiently extract important features for TWSA predictions from multi-source input data. The predicted TWSAs are compared to the hydrological model outputs and three recent TWSA prediction products. Results suggest the superior performance of BCNN in bridging the gap. The extreme dry and wet events during the gap period are also successfully identified by BCNN.

Filling the gap between GRACE- and GRACE-FO-derived terrestrial water storage anomalies with Bayesian convolutional neural networks

Shaoxing Mo[1], Yulong Zhong[2], Xiaoqing Shi[1], Wei Feng[3,4], Xin Yin[5], 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]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
[4]School of Geospatial Engineering and Science, Sun Yat-Sen University, Zhuhai, China
[5]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China

Key Points:

- A Bayesian deep learning method is proposed for filling the gap between GRACE- and GRACE-FO-derived terrestrial water storage anomalies
- A state-of-the-art performance is obtained in bridging the gap at a global scale (excluding Antarctica)
- Extreme dry and wet events during the gap are successfully identified

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

Corresponding author: Wei Feng, `fengwei@whigg.ac.cn`

Abstract

There is an approximately one-year observation gap of terrestrial water storage anomalies (TWSAs) between the Gravity Recovery and Climate Experiment (GRACE) satellite and its successor GRACE Follow-On (GRACE-FO). This poses a challenge for water resources management, as discontinuity in the TWSA observations may introduce significant biases and uncertainties in hydrological model predictions and consequently mislead decision making. To tackle this challenge, a Bayesian convolutional neural network (BCNN) is proposed in this study to bridge this gap using climatic data as inputs. Enhanced by integrating recent advances in deep learning, BCNN can efficiently extract important features for TWSA predictions from multi-source input data. The predicted TWSAs are compared to the hydrological model outputs and three recent TWSA prediction products. Results suggest the superior performance of BCNN in bridging the gap. The extreme dry and wet events during the gap period are also successfully identified by BCNN.

Plain Language Summary

The remote sensing satellites Gravity Recovery and Climate Experiment (GRACE) provide valuable and accurate observations of terrestrial water storage changes at a global scale. These observations have been used widely for sustainable water resources management and understanding water cycle. However, there is an about one-year gap that the GRACE observations are missing due to a break-in period between two satellites. Filling this gap is thus of crucial significance for practical hydrological, agricultural, and ecological applications. We propose in this work a Bayesian convolutional neural network (BCNN) by leveraging recent advances in deep learning to bridge this gap based on climatic data. Results show that our BCNN achieves a state-of-the-art performance in terms of filling accuracy compared to previous studies.

1 Introduction

The Gravity Recovery and Climate Experiment (GRACE) satellite and its successor GRACE Follow-On (GRACE-FO) provide unprecedentedly accurate observations of the spatio-temporal dynamics of terrestrial water storage anomalies (TWSAs) at a global scale. These TWSA observations have been widely utilized together with hydrological models to assess water cycle, droughts and floods, the impacts of changing climate on terrestrial water storage, etc (Feng et al., 2018; Gentine et al., 2019; Rateb et al., 2020; Richey et al., 2015; Rodell et al., 2018; Soltani et al., 2021; Tapley et al., 2019). Particularly, their applications in groundwater-related studies (e.g., the groundwater depletion and land subsidence problems in North China Plain and California's Central Valley) are more extensive due to the difficult-to-observe nature of groundwater systems, substantially augmenting our knowledge toward the systems and thus benefiting groundwater resources management (e.g., Chang et al., 2020; Chen et al., 2014; Famiglietti et al., 2011; Feng et al., 2013, 2018; B. Li et al., 2019; Smith & Majumdar, 2020; Zhong et al., 2018).

However, there is an approximately one-year gap of TWSA observations between the GRACE and GRACE-FO missions. Considering that the TWSA observations are usually assimilated in the hydrological models for higher reliability (B. Li et al., 2019; Soltani et al., 2021; Yin et al., 2020; Zaitchik et al., 2008), discontinuity in the time series observations may introduce significant biases and uncertainties in model predictions and consequently mislead decision making (A. Y. Sun et al., 2020). Bridging this gap is thus of crucial importance for practical applications. There have been many efforts undertaken to predict/reconstruct GRACE TWSAs at regional or global scales using data-driven methods (e.g., Ahmed et al., 2019; Forootan et al., 2014; Forootan et al., 2020; Humphrey et al., 2017; Humphrey & Gudmundsson, 2019; Jing et al., 2020; F. Li et al., 2020; Long et al., 2014; Richter et al., 2021; A. Y. Sun et al., 2019, 2020; Z. Sun et al., 2020; Wang et al., 2021). While these studies have generally obtained relatively good performances in humid regions, the performance in

arid and semiarid regions remains relatively poor (a global aridity index map is shown in Figure A1), calling for innovative solutions. The poor performance in arid/semiarid regions is mainly due to the difficulty in modeling the long-term TWSA trends caused by anthropogenic activities (Humphrey & Gudmundsson, 2019; F. Li et al., 2020; Z. Sun et al., 2020).

Recent years have witnessed a rapid development of deep learning and its impressive performance in various applications (Gu et al., 2018; Reichstein et al., 2019; Shen, 2018). In this study, we propose a Bayesian convolutional neural network (BCNN) driven by climatic inputs to bridge the TWSA observation gap between GRACE and GRACE-FO at a global scale (excluding Antarctica). In this framework, the input climatic data and the target GRACE TWSAs are treated as images to leverage the excellent capability of convolutional neural networks (CNNs) in image processing (Gu et al., 2018; Mo, Zhu, et al., 2019; Mo, Zabaras, et al., 2019; Mo et al., 2020; A. Y. Sun et al., 2019). A. Y. Sun et al. (2019) applied CNNs for prediction of TWSAs in India and the CNN outperformed the hydrological models in providing more accurate TWSA estimates. The development of our BCNN leverages and combines the recent advances in deep learning, including the residual (He et al., 2016) and dense connection (Huang et al., 2017) modules, channel and spatial attention mechanisms (Woo et al., 2018), Bayesian training strategy (Liu & Wang, 2016; Zhu & Zabaras, 2018) and other (Gu et al., 2018). By comparing with the hydorlogical model outputs and three recent TWSA prediction products, we will show that the combination of these strategies enables the network to automatically extract informative features from multi-source driving data, offer predictive uncertainties, and consequently achieve state-of-the-art performances in filling the gap.

2 Data and Processing

2.1 GRACE TWSA Data

The monthly GRACE Mascon product released by the Jet Propulsion Laboratory (JPL) (available at `https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06_V2`), which has a spatial resolution of $0.5^\circ \times 0.5^\circ$ (Watkins et al., 2015), is used in this study. The JPL GRACE TWSA data are provided as anomalies with respect to the 2004 to 2009 average. The observations cover two periods, that is, April 2002-June 2017 (i.e. the GRACE mission) and June 2018-present (i.e. the GRACE-FO mission), with a 11-month gap in between. In addition, there are some intermittent one or two months with missing TWSA data within each mission, these missing months are interpolated using the data of neighboring months. Our aim is to fill the 11-month gap (i.e. July 2017-May 2018) for the land areas with the BCNN method. To facilitate the comparison with previous GRACE prediction studies, we resampled averagely the data to $1^\circ \times 1^\circ$ grid.

2.2 ERA5-Land Driving Data

The driving data used to predict the GRACE TWSAs are extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5-Land (ERA5L) dataset (available at `https://cds.climate.copernicus.eu`). The ERA5L dataset contains an improved version of the land component of the ERA5 climate reanalysis, making it more accurate for land applications. The data are provided at a spatial resolution of $0.1^\circ \times 0.1^\circ$ and cover a period from January 1981 to near present. The driving data contain four predictor variables, including the monthly precipitation, temperature, cumulative water storage changes (CWSCs), and ERA5L-derived TWSAs. The spatial resolution of these data is averagely resampled into to $1^\circ \times 1^\circ$ to be consistent with GRACE TWSAs.

The water storage change (WSC) is calculated as the difference between the inflow (i.e. precipitation P) and outflows (i.e., evapotranspiration ET and runoff RO) based on the

water balance:

$$\mathrm{WSC} = P - ET - RO. \tag{1}$$

The CWSC is thus written as follows:

$$\mathrm{CWSC}_t = \sum_{i=1}^{t} \mathrm{WSC}_i = \sum_{i=1}^{t} (P_i - ET_i - RO_i), \tag{2}$$

where t denotes the month index. CWSC correlates well to GRACE TWSA in most regions as shown in Figure A2. Therefore, it is used as an additional predictor of GRACE TWSAs.

The ERA5L dataset includes water storage in soil moisture (in four layers spanning from 0 to 289 cm), snow, and canopy. Thus, the ERA5L TWSAs are calculated by summing these components and then subtracting the long-term mean between 2004 and 2009 to be consistent with the GRACE TWSAs, as represented by:

$$\mathrm{TWSA}_{\mathrm{ERA5L}} = \mathrm{SMS} + \mathrm{SWS} + \mathrm{CWS} - \overline{\mathrm{TWS}}_{0409}, \tag{3}$$

where SMS, SWS, and CWS are soil moisture storage, snow water storage, and canopy water storage, respectively, $\overline{\mathrm{TWS}}_{0409}$ denotes mean of the three components during 2004-2009.

2.3 Time Series Data detrending

The GRACE TWSA time series in arid and semiarid regions often exhibits a long-term declining/rising trend caused by the human intervention and/or changing climate. This poses challenges for the TWSA prediction tasks, as the driving data (e.g., the climatic data or outputs of hydrological models) may not be able to well reflect the influences of these factors (Humphrey & Gudmundsson, 2019; F. Li et al., 2020; Z. Sun et al., 2020). For the 11-month gap (July 2017-May 2018) filling task considered here, fortunately, the GRACE TWSA data before (April 2002-June 2017) and after (June 2018-) the gap are available. Thus, we can obtain directly the long-term trend signals for the missing interval from existing data. Then we predict in the gap filling task the detrended TWSA signals instead, which are generally less challenging relative to predicting the original TWSA signals (Humphrey & Gudmundsson, 2019; F. Li et al., 2020). Mathematically, the GRACE TWSAs are decomposed via linear detrending into two components:

$$\mathrm{TWSA}_{\mathrm{GRACE}} = \mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{GRACE}} + \mathrm{trend}_{\mathrm{GRACE}}, \tag{4}$$

where $\mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{GRACE}}$ and $\mathrm{trend}_{\mathrm{GRACE}}$ are the detrended data and a linear trend term, respectively. Correspondingly, the driving data described in section 2.2 are also detrended. In the prediction task, the BCNN model learns to predict the $\mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{GRACE}}$ signals. Finally, the predictions for the original TWSAs are obtained by adding the GRACE trend:

$$\mathrm{TWSA}_{\mathrm{pred}} = \mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{pred}} + \mathrm{trend}_{\mathrm{GRACE}}. \tag{5}$$

3 Methods

3.1 BCNN Deep Learning Models

We propose a BCNN method to learn the underlying relationship between $\mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{GRACE}}$ and its predictors. Without loss of generality, here we use $\mathbf{x}$ and $\mathbf{y}$ to denote the network inputs (i.e. predictors) and outputs (i.e. $\mathrm{TWSA}^{\mathrm{detrend}}_{\mathrm{GRACE}}$), respectively. In BCNN, the inputs and outputs are both organized as images (matrices) and the learning task becomes an image regression problem:

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

where $\boldsymbol{\eta} = \boldsymbol{\eta}(\mathbf{x}, \mathbf{w})$ is a BCNN model, with $\mathbf{w}$ denoting all trainable network parameters, including the weights and biases. The inputs $\mathbf{x}$ and outputs $\mathbf{y}$ become n_x and n_y images, respectively, all with $H \times W$ pixels (grids).

The network predictions are inevitably associated with epistemic uncertainties induced by lack of training data. Failing to estimate the predictive uncertainty may lead to overconfident results. This is especially the case for the GRACE TWSA prediction task considered here, as the available training data are limited. Contrary to vanilla CNNs which treat the network parameters $\mathbf{w}$ as deterministic unknowns and thus fails to offer predictive uncertainties, in BCNN $\mathbf{w}$ are treated as random variables. Given a set of training data $\mathcal{D} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N_{\text{train}}}$, the network training is to infer the posterior distribution of $\mathbf{w}$, $p(\mathbf{w}|\mathcal{D})$. Consequently, one can obtain the prediction distribution of the target variables $\mathbf{y}$: $p(\mathbf{y}|\mathbf{w})$, $\mathbf{w} \sim p(\mathbf{w}|\mathcal{D})$, and in particular the mean $\mathbb{E}(\mathbf{y}|\mathbf{w})$ and standard deviation $\text{Std}(\mathbf{y}|\mathbf{w})$.

In BCNN, the Bayesian training strategy proposed in Zhu and Zabaras (2018) is employed. Mathematically, the BCNN model is expressed as follows:

$$\hat{\mathbf{y}} = \boldsymbol{\eta}(\mathbf{x}, \mathbf{w}) + \mathbf{n}(\mathbf{x}, \mathbf{w}), \tag{7}$$

where $\mathbf{n}(\cdot)$ is an additive Gaussian noise term modeling the aleatoric uncertainty. A variational Bayesian inference algorithm called stein variational gradient descent (SVGD) (Liu & Wang, 2016), which is similar to standard gradient descent while maintaining the particle methods' high efficiency, is utilized to estimate the posterior distribution $p(\mathbf{w}|\mathcal{D})$. In implementation, we use N_S samples of $\mathbf{w}$ to approximate the posterior distribution $p(\mathbf{w}|\mathcal{D})$. The N_S samples $\{\mathbf{w}_i\}_{i=1}^{N_S}$ are respectively optimized using the Adam optimizer (Kingma & Ba, 2015) whose gradient derives from SVGD. The predictive mean and uncertainty of BCNN for an arbitrary input $\mathbf{x}$ are then calculated based on the N_S predictions ($\hat{\mathbf{y}}^{(i)} = \boldsymbol{\eta}(\mathbf{x}, \mathbf{w}_i) + \mathbf{n}(\mathbf{x}, \mathbf{w}_i)$, $i = 1, \ldots, N_S$). For more details, one can refer to Liu and Wang (2016) and Zhu and Zabaras (2018).

3.2 BCNN Architecture Design and Training

The BCNN network architecture is depicted in Figure A3. The convolutional block attention module (CBAM) proposed in Woo et al. (2018) is used as the basic block. Given n_x images with a resolution of $H \times W$ as inputs to the network, they are passed through an alternating cascade of convolutional/transposed convolutional layers and CBAMs, each of which produces n_{f} $H^{'} \times W^{'}$ feature maps, to extract multi-scale features to finally predict n_y images for the targets. The CBAM block contains two attention modules, namely the channel and spatial attention modules (Figures A3 and A4). More specifically, the channel module outputs n_{f} weights between 0 and 1 assigning to the n_{f} feature maps to tell the network 'what' (i.e. which maps) to attend; the spatial module outputs a $H^{'} \times W^{'}$ weight matrix assigning to the $H^{'} \times W^{'}$ pixel feature maps to tell the network 'where' (i.e. which regions) to emphasize or suppress. As such, the network is able to automatically focus on important features and suppress unnecessary ones (Woo et al., 2018).

The residual connection (He et al., 2016) and dense connection (Huang et al., 2017) strategies have been shown to be effective in enhancing information flow through the network and thus substantially improving the performance. Therefore, they are implemented in our BCNN model. In particular, in the residual connection structure, the feature maps with the same shape ($n_{\text{f}} \times H^{'} \times W^{'}$) but at different layers are connected by applying element-wise addition (He et al., 2016). In the dense connection structure, the feature maps with the same size ($H^{'} \times W^{'}$) but at different layers are cascaded together and subsequently fed as inputs into the next layer (Huang et al., 2017) (Figure A3). The Mish function (Misra, 2019) is employed in BCNN as the activation function of hidden layers unless otherwise stated.

We use twelve years of monthly GRACE TWSA data from April 2002 to March 2014 (i.e. 144 months, ∼69%) to train the BCNN network, and those from April 2014 to June 2017 and June 2018 to August 2020 (i.e. 66 months, ∼31%) to test the performance. We set the number of lags for predictors to 2. That is, for month t, the inputs to BCNN are P_i, T_i, CWSC_i, and $\text{TWSA}_{\text{ERA5L},i}$ with $i = t-2, ..., t$. Thus, each sample contains $n_x = 12$ input images and $n_y = 1$ output image ($\text{TWSA}_{\text{GRACE},t}$). The region spanning

from 60°S to 84°N and 180°W to 180°E represented by a $H \times W = 144 \times 360$ image is considered. For non-land pixels (grids), the pixel values are set to a constant of 0. During network training, we use $N_S = 20$ samples of $\mathbf{w}$ in the SVGD algorithm to approximate the posterior distribution, as suggested in Zhu and Zabaras (2018). The network is trained for 200 epochs, with a mean squared error loss function quantifying the predictive accuracy, an initial learning rate of 0.0025 in the Adam optimizer and a batch size of 12. The training is performed on a single GPU (NVIDIA Tesla V100) and takes ∼80 minutes. The network performance is evaluated based on the test dataset using three commonly used metrics: correlation coefficient ($R \in [-1, 1]$), Nash-Sutcliffe efficiency coefficient (NSE$\in (-\infty, 1]$), and root mean squared error (RMSE$\in [0, +\infty)$), where R or NSE values closer to 1.0 or lower RMSE values indicate better performances.

4 Results and Discussion

4.1 Accuracy Assessment

To illustrate the performance of BCNN, the three metrics (i.e., R, NSE, and RMSE) are also computed for the ERA5L TWSAs and the hydrological model Noah-derived TWSAs ($1° \times 1°$, version 2.1) (Rodell et al., 2004).

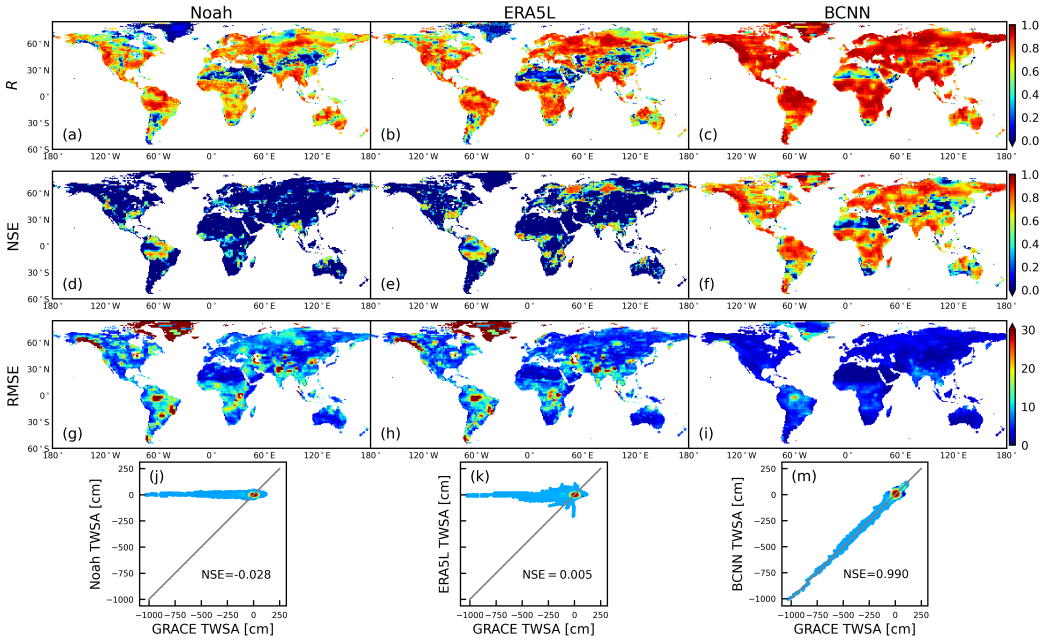

Figure 1. Spatial maps of R (row 1), NSE (row 2), and RMSE (row 3) values between the GRACE TWSAs and Noah- (column 1), ERA5L- (column 2), and BCNN-derived (column 3) TWSAs during the test period (April 2014-June 2017, June 2018-August 2020). (j-m) The density scatter plots between the GRACE and modeled TWSAs.

Figure 1 shows the spatial distribution of the accuracy metrics obtained by Noah, ERA5L, and BCNN. While Noah and ERA5L TWSAs both show relatively good correlations with GRACE TWSAs in most regions except Greenland and the extremely arid areas like Sahara, Gobi, and Arabian (Figures 1a and 1b), BCNN TWSAs exhibit clearly better correlations with GRACE TWSAs in almost all regions (Figure 1c). For the NSE metric, which measures directly the matching quality between the predicted and observed values, both Noah and ERA5L obtain relatively low values (<0) in most regions except in some

humid regions like Amazon and Southeastern United States (Figures 1d and 1e). In contrast, BCNN achieves relatively high values (>0.5) in most regions (Figure 1f). Although BCNN provides a higher accuracy than Noah and ERA5L in the desert regions, the NSE value is still low relative to other regions. This is due to the fact that the variability of GRACE TWSAs in these regions is dominated by noise (Humphrey et al., 2016). The improved performance of BCNN over Noah and ERA5L can be also illustrated by the RMSE maps (Figure 1(g-i)) and the density scatter plots between the modeled and GRACE TWSAs (Figure 1(j-m)), where our BCNN reduces significantly the RMSE errors and its predicted TWSAs display as expected a good consistency with GRACE TWSAs with an overall NSE value of 0.990, much higher than those of Noah (-0.028) and ERA5L (0.005).

The poor performance of Noah/ERA5L in predicting TWSAs is mainly due to their underestimation of the long-term trends of GRACE TWSAs (Scanlon et al., 2018). Considering the GRACE trends are employed in BCNN (section 2.3), to illustrate that the outperformance of BCNN is not only obtained by simply utilizing the GRACE trend information, but more importantly attributed to its ability to discover informative features for TWSA predictions from multi-source data, we correct the Noah and ERA5L TWSAs with the GRACE trends. That is, TWSA_m=`detrend`(TWSA_m)+$\text{trend}_{\text{GRACE}}$, where `detrend(·)` denotes the linear detrending operation, TWSA_m represents the Noah or ERA5L TWSAs. Figure A5 shows the same metrics as in Figure 1 for the corrected Noah and ERA5L TWSAs. The comparison between Figures 1(c,f,i,m) and A5 again clearly indicates BCNN's higher performance, although the consistency between the corrected Noah/ERA5L TWSAs and GRACE TWSAs has been improved significantly as expected.

Figure 2 depicts BCNN's TWSA predictions for three test months in June 2014, June 2017, and June 2020. The three months cover the early-, mid-, and late-term of the test period, with June 2017 being the last month before the missing gap. For comparison, we also show the reference GRACE TWSAs. It can be seen that BCNN successfully captures the spatial patterns of GRACE TWSAs and provides close predictions in the three months. The predicted errors in regions with high-TWSA signals (e.g., Amazon, Central Africa, South Asia, and Greenland) are generally larger compared to other regions with relatively low-TWSA signals (Figure 2(g-i)). This is consistent with the spatial distributions of BCNN's predictive uncertainties shown in Figure 2(j-m), where the regions with high-TWSA signals generally have larger predictive uncertainties. Note the predicted accuracy in these high-TWSA regions is still relatively good as the high-TWSA signals result in a high percentage accuracy. This can also be revealed by the NSE map shown in Figure 1f.

The performance of BCNN in providing reliable predictions for TWSAs is further demonstrated in Figures 2(n-y) and A7, which depict the basin/region-averaged TWSA time series derived from GRACE, Noah, ERA5L, and BCNN in different basins/regions (Locations of these basins/regions are shown in Figure A6). For completeness, we show the time series in both training and test periods. Again, the BCNN TWSAs fit obviously better with the reference GRACE TWSAs than Noah and ERA5L TWSAs. While BCNN may slightly underestimate/overestimate some peak/valley values of GRACE TWSAs during the test period, the GRACE TWSA curves are almost completely enveloped within the 95% prediction interval.

4.2 Detection of Extreme Dry and Wet Events during the Gap

The GRACE TWSA time series are effective indicators for detection of extreme dry/wet events, which cause unusual decreases/increases in the TWSA signals, and quantifying the water loss/gain during the events (Feng et al., 2018; Humphrey et al., 2016; F. Li et al., 2020). For example, the GRACE successfully identified the 2016/2017 and 2018/2019 droughts in Central Europe (Figure 2o) (Boergens et al., 2020), the droughts from 2012 to 2016 in Central Valley (Figure 2p) (Xiao et al., 2017), and the flood in Summer 2020 in Yangtze River Basin (Figure 2w) (Wei et al., 2020). As shown in Figure 2(n-y), the

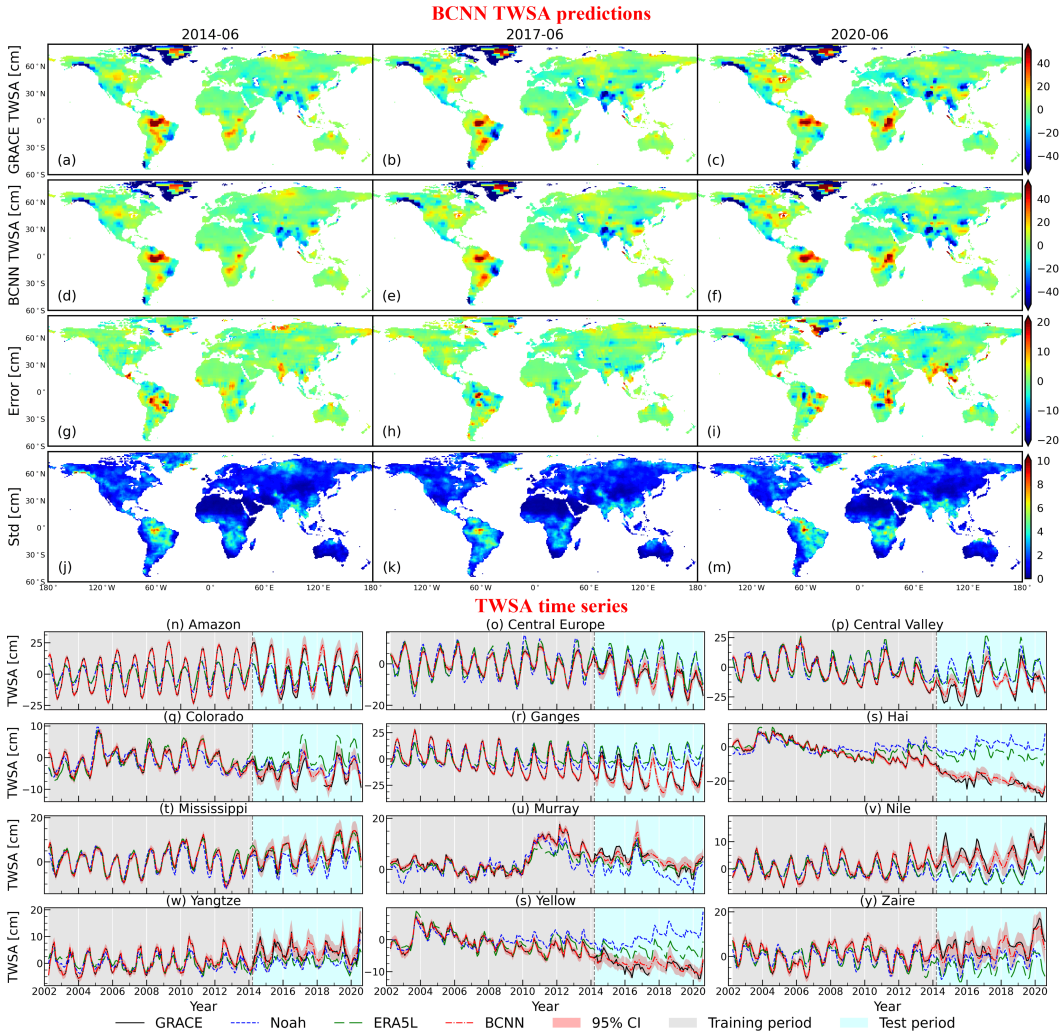

Figure 2. The GRACE (a-c) and BCNN (d-f) TWSAs in three test months (June 2014, June 2017, and June 2020). (g-i) The BCNN predicted error (i.e. $\text{TWSA}_{\text{GRACE}}$ - $\text{TWSA}_{\text{ERA5L}}$) and (j-m) standard deviation (Std). (n-y): The region/basin-averaged GRACE, Noah, ERA5L and BCNN TWSA time series in different basins/regions. The red shaded area denotes the 95% confidence interval (CI) of BCNN predictions.

BCNN TWSAs agree well with GRACE TWSAs during these extreme events, suggesting that BCNN is capable of detecting the drought- and flood-induced abnormal TWSA signals.

We further analyze BCNN's performance in detecting dry and wet events during the gap. To this end, the trend and seasonal signals are removed from the original BCNN TWSAs, which is done by fitting a linear trend via unweighted least squares, plus annual and semiannual sinusoids to the TWSA time series (Chen et al., 2014; Zhong et al., 2018). The detrended and deseasonalized TWSAs at each grid point is then standardized using the z-score formula. The resulting TWSA is denoted as $\text{sTWSA}_{\text{BCNN}}^{\text{detrend,deseason}}$ and its spatial maps in the eleven missing months are shown in Figure 3(a-k). It is observed that there are four regions (labeled with (l-o) in Figure 3h) exhibiting abnormal signals, with extremely negative and positive signals indicating dry (regions l and m) and wet events (regions n and o), respectively, where the dry/wet events in regions m and n were also reported by European State of the Climate (`https://climate.copernicus.eu/ESOTC`). To

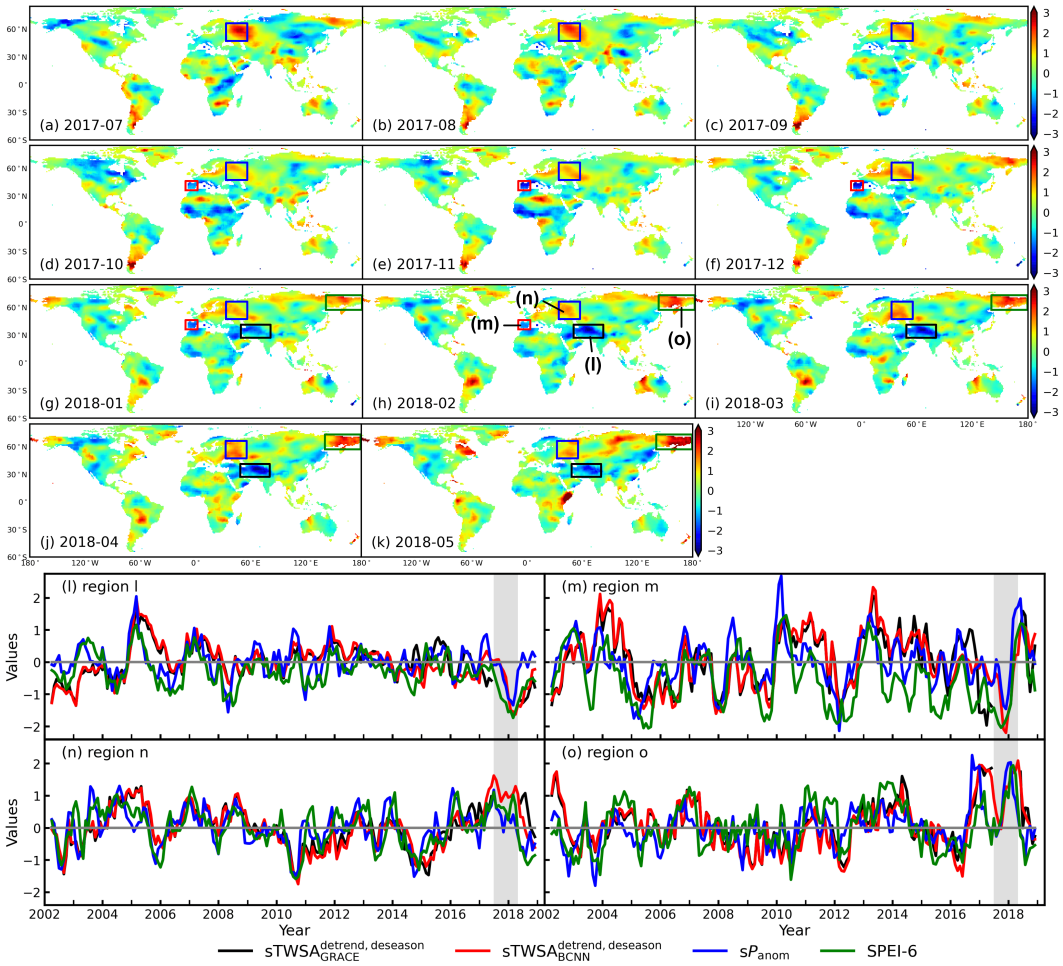

Figure 3. (a-k) The standardized signals of the detrended and deseasonalized BCNN TWSAs ($\text{sTWSA}^{\text{detrend,deseason}}_{\text{BCNN}}$) in the missing gap (July 2017-May 2018). The four regions within the rectangles (labeled l-o in h) have significant abnormal signals during this period. (l-o) The region-averaged time series of $\text{sTWSA}^{\text{detrend,deseason}}_{\text{GRACE}}$, $\text{sTWSA}^{\text{detrend,deseason}}_{\text{BCNN}}$, standardized precipitation anomalies ($\text{s}P_{\text{anom}}$), and 6-month SPEI in the four regions. The shaded area denotes the gap period.

examine this, we plot in Figure 3(l-o) the region-averaged times series of standardized precipitation anomalies ($\text{s}P_{\text{anom}}$) and 6-month standardized precipitation-evapotranspiration index (SPEI). When calculating P_{anom}, we first apply moving average on the precipitation time series ($P_t = \sum_{i=-3}^{0} P_{t+i}$) to smooth out short-term fluctuations, and then compute the anomalies for each month with respect to its respective average during 2002 and 2018. The SPEI dataset, which covers a period from 2002 to 2018, is downloaded from `https://spei.csic.es/spei_database/`. For comparison, the $\text{sTWSA}^{\text{detrend,deseason}}_{\text{GRACE}}$ time series are also displayed in the plot. It is observed that the two $\text{sTWSA}^{\text{detrend,deseason}}$ lines agree well with the $\text{s}P_{\text{anom}}$ and SPEI lines over the period 2002-2018. This demonstrates the capability of $\text{sTWSA}^{\text{detrend,deseason}}$ in detecting extreme climate events. More notably, $\text{sTWSA}^{\text{detrend,deseason}}_{\text{BCNN}}$ successfully identifies the extreme dry events in regions l and m (Figures 3l and 3m) and the extreme wet conditions in regions n and o (Figures 3n and 3o) during the 11-month gap (July 2017-May 2018).

The results presented in section 4.1 and this section indicate that BCNN can successfully captures the complex spatio-temporal behaviors of GRACE TWSAs as well as the abnormal signals caused by extreme climate conditions. It is worth noting that the test data from GRACE (April 2014-June 2017) and GRACE-FO (June 2018-Aug 2020) cover the periods before and after the missing gap (i.e. July 2017-May 2018). The good agreement between BCNN and GRACE TWSAs suggests the reliability of BCNN in bridging the gap between GRACE and GRACE-FO at a global scale.

4.3 Comparison with Previous Studies

As presented in section 1, there have been many efforts undertaken to model GRACE TWSAs using data-driven methods. Here we restrict the comparison with Humphrey and Gudmundsson (2019), F. Li et al. (2020), and Z. Sun et al. (2020) who provided publicly accessible global-scale TWSA prediction products. Note that the predicted TWSA product released by Humphrey and Gudmundsson (2019) is known as GRACE-REC. For a fair comparison, the GRACE TWSA data and the training periods used for BCNN network training are respectively the same as those employed in the three studies. The detailed descriptions of the three TWSA products are summarized in Table A1.

The comparison results are shown in Figures A8-A10, which illustrate the R, NSE, and (normalized) RMSE maps and the density scatter plots between the predicted and GRACE TWSAs. Note that the original GRACE-REC dataset provides the detrended and deseasonalized TWSAs. The trend and seasonal signals obtained from the GRACE TWSAs and Humphrey et al. (2017), respectively, have been added to the original GRACE-REC TWSAs for consistency. It can be observed that BCNN obtains better metric values in the vast majority of regions (especially in the arid and semiarid regions) and shows higher agreements with GRACE in comparison to the three previous studies. Recall that the GRACE trend information has been added to the original GRACE-REC TWSAs and was also utilized in F. Li et al. (2020) when predicting TWSAs. The results suggest the superior capability of BCNN in feature mining and thus providing reliable TWSA predictions to bridge the gap between GRACE and GRACE-FO. Two additional noteworthy advantages of BCNN over existing methods are the relatively few assumptions and pre-processing involved and its ability to handle simultaneously the global scale.

5 Conclusions

The GRACE/GRACE-FO TWSA observations, together with the hydrological models, have been vital tools for water-related studies at regional or even global scales. However, the approximately one-year gap of TWSA observations between the two GRACE missions may introduce significant biases and uncertainties in models and consequently lead to misleading predictions. In this study, we propose a deep learning-based BCNN model driven by climatic data to fill this gap. By leveraging and implementing recent advances of deep learning (e.g., the residual and dense connections, attention mechanisms, and Bayesian training strategy), the BCNN model is able to effectively and efficiently extract important information for TWSA predictions from multi-source input data. Results show that BCNN can successfully capture the complex spatio-temporal behaviors of TWSAs and identify the extreme dry/wet events during the gap. The comparisons with hydrological model outputs and previous studies further suggest that BCNN obtains a state-of-the-art performance in bridging the gap at a global scale.

The outperformance of BCNN is mainly attributed to the use of TWSA trends, which are derived from the available GRACE/GRACE-FO data before and after the gap, and its outstanding performance in feature extraction. The long-term TWSA trends are induced by anthropogenic and/or natural factors and usually challenging-to-learn (Humphrey & Gudmundsson, 2019; F. Li et al., 2020; Z. Sun et al., 2020). The utilization of this trend information makes full use of the existing data and essentially eases the learning task for

BCNN. The BCNN's robust capability in extracting key features for TWSA predictions is especially illustrated by the comparison with the TWSA prediction products in which the trend information was also employed. Note that we are concerned with bridging the gap between GRACE and GRACE-FO in the current work. For the TWSA reconstruction task, which aims to reconstruct the TWSAs for pre-2002 and is beyond the scope of this study, the trend information is unavailable. The performance of BCNN for such a task remains to be investigated.

Acknowledgments

The JPL GRACE Mascon data used in this study are available at `https://podaac.jpl.nasa.gov/dataset/TELLUS_GRAC-GRFO_MASCON_CRI_GRID_RL06_V2`; ERA5-Land data are available at `https://cds.climate.copernicus.eu`; Noah TWSA dataset is downloaded from `https://disc.gsfc.nasa.gov/`. 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), Jiangsu Planned Projects for Postdoctoral Research Funds (2020Z133), and Fundamental Research Funds for the Central Universities (020614380106). The authors acknowledge Dr. Yinhao Zhu from Qualcomm AI Research for his valuable suggestions on implementing the Bayesian training strategy. The BCNN codes and the predicted TWSA dataset generated in this work will be made available at `https://github.com/njujinchun/bcnn4grace` upon publication of this manuscript.

References

Ahmed, M., Sultan, M., Elbayoumi, T., & Tissot, P. (2019). Forecasting GRACE data over the African watersheds using artificial neural networks. *Remote Sensing*, *11*(15), 1769. doi: https://doi.org/10.3390/rs11151769

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

Chang, L.-L., Yuan, R., Gupta, H. V., Winter, C. L., & Niu, G.-Y. (2020). Why is the terrestrial water storage in dryland regions declining? A perspective based on Gravity Recovery and Climate Experiment observations and Noah land surface model with multiparameterization schemes model simulations. *Water Resources Research*, *56*(11), e2020WR027102. doi: https://doi.org/10.1029/2020WR027102

Chen, J., Li, J., Zhang, Z., & Ni, S. (2014). Long-term groundwater variations in Northwest India from satellite gravity measurements. *Global and Planetary Change*, *116*, 130-138. doi: https://doi.org/10.1016/j.gloplacha.2014.02.007

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K. J., Syed, T. H., ... Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters*, *38*(3), L03403. doi: https://doi.org/10.1029/2010GL046442

Feng, W., Shum, C. K., Zhong, M., & Pan, Y. (2018). Groundwater storage changes in china from satellite gravity: An overview. *Remote Sensing*, *10*(5), 674. doi: https://doi.org/10.3390/rs10050674

Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H.-T., & Xia, J. (2013). Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resources Research*, *49*(4), 2110-2118. doi: https://doi.org/10.1002/wrcr.20192

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 in Geophysics*, *35*(4), 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

Gentine, P., Green, J. K., Guérin, M., Humphrey, V., Seneviratne, S. I., Zhang, Y., & Zhou, S. (2019). Coupling between the terrestrial carbon and water cycles—A review. *Environmental Research Letters*, *14*(8), 083003. doi: https://doi.org/10.1088/1748-9326/ab22d6

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354-377. doi: 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. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

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

Humphrey, V., Gudmundsson, L., & Seneviratne, S. I. (2017). A global reconstruction of climate-driven subdecadal water storage variability. *Geophysical Research Letters*, *44*(5), 2300-2309. doi: https://doi.org/10.1002/2017GL072564

Jing, W., Zhao, X., Yao, L., Di, L., Yang, J., Li, Y., ... Zhou, C. (2020). Can terrestrial water storage dynamics be estimated from climate anomalies? *Earth and Space Science*, *7*(3), e2019EA000959. doi: https://doi.org/10.1029/2019EA000959

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Y. Bengio & Y. LeCun (Eds.), *3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA.*

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., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., & Lück, C. (2020). Comparison of data-driven techniques to reconstruct (1992–2002) and predict (2017–2018) GRACE-like gridded total water storage changes using climate inputs. *Water Resources Research*, *56*(5), e2019WR026551. doi: https://doi.org/10.1029/2019WR026551

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.

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

Misra, D. (2019). Mish: A self regularized non-monotonic activation function. *arXiv e-prints*, arXiv:1908.08681.

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., 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

Rateb, A., Scanlon, B. R., Pool, D. R., Sun, A., Zhang, Z., Chen, J., ... Zell, W. (2020). Comparison of groundwater storage changes from GRACE satellites with monitoring and modeling of major U.S. aquifers. *Water Resources Research*, *56*(12), e2020WR027556. doi: https://doi.org/10.1029/2020WR027556

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

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., ... Rodell, M. (2015). Quantifying renewable groundwater stress with GRACE. *Water Resources Research*, *51*(7), 5217-5238. doi: https://doi.org/10.1002/2015WR017349

Richter, H. M. P., Lück, C., Klos, A., Sideris, M. G., Rangelova, E., & Kusche, J. (2021). Reconstructing GRACE-type time-variable gravity from the Swarm satellites. *Scientific Reports*, *11*, 1117. doi: https://doi.org/10.1038/s41598-020-80752-w

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651-659. doi: https://doi.org/10.1038/s41586-018-0123-1

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

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., ... Bierkens, M. F. P. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, *115*(6), E1080–E1089. doi: https://doi.org/10.1073/pnas.1704665115

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

Smith, R. G., & Majumdar, S. (2020). Groundwater storage loss associated with land subsidence in Western United States mapped using machine learning. *Water Resources Research*, *56*(7), e2019WR026621. doi: https://doi.org/10.1029/2019WR026621

Soltani, S. S., Ataie-Ashtiani, B., & Simmons, C. T. (2021). Review of assimilating GRACE terrestrial water storage data into hydrological models: Advances, challenges and opportunities. *Earth-Science Reviews*, *213*, 103487. doi: https://doi.org/10.1016/j.earscirev.2020.103487

Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2020). Reconstruction of GRACE total water storage through automated machine learning. *Water Resources Research*, 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

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., ... Velicogna, I. (2019). Contributions of GRACE to understanding climate change.

Nature Climate Change, *9*, 358–369. doi: https://doi.org/10.1038/s41558-019-0456-2

Wang, F., Shen, Y., Chen, Q., & Wang, W. (2021). Bridging the gap between GRACE and GRACE Follow-on monthly gravity field solutions using improved multichannel singular spectrum analysis. *Journal of Hydrology*, 125972. doi: https://doi.org/10.1016/j.jhydrol.2021.125972

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. doi: https://doi.org/10.1002/2014JB011547

Wei, K., Ouyang, C., Duan, H., Li, Y., Chen, M., Ma, J., ... Zhou, S. (2020). Reflections on the catastrophic 2020 Yangtze River Basin flooding in southern China. *The Innovation*, *1*(2), 100038. doi: https://doi.org/10.1016/j.xinn.2020.100038

Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV).*

Xiao, M., Koppa, A., Mekonnen, Z., Pagán, B. R., Zhan, S., Cao, Q., ... Lettenmaier, D. P. (2017). How much groundwater did California's Central Valley lose during the 2012–2016 drought? *Geophysical Research Letters*, *44*(10), 4872-4879. doi: https://doi.org/10.1002/2017GL073333

Yin, W., Han, S.-C., Zheng, W., Yeo, I.-Y., Hu, L., Tangdamrongsub, N., & Ghobadi-Far, K. (2020). Improved water storage estimates within the North China Plain by assimilating GRACE data into the CABLE model. *Journal of Hydrology*, *590*, 125348. doi: https://doi.org/10.1016/j.jhydrol.2020.125348

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River Basin. *Journal of Hydrometeorology*, *9*(3), 535-548. doi: https://doi.org/10.1175/2007JHM951.1

Zhong, Y., Zhong, M., Feng, W., Zhang, Z., Shen, Y., & Wu, D. (2018). Groundwater depletion in the West Liaohe River Basin, China and its implications revealed by GRACE and in situ measurements. *Remote Sensing*, *10*(4), 493. doi: https://doi.org/10.3390/rs10040493

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

Appendix A Supporting Information

Table A1. Summaries of Experimental Settings in Previous Studies.

Authors	Study area	GRACE data[a]	Training period	Test period
Humphrey and Gudmundsson (2019)[b]	Global	JPL mascon RL06	-	Apr. 2014-June 2017; June 2018-July 2019
F. Li et al. (2020)[c]	Global	CSR mascon RL06[d]	Apr. 2002-June 2017	June 2018-Dec. 2018
Z. Sun et al. (2020)	26 basins	JPL mascon RL06	Apr. 2002-Jan. 2014	Feb. 2014-June 2017

[a]There might be multiple GRACE data products used in the referenced studies. Here we list the product used in the comparison.

[b]The generated data product is known as GRACE-REC and available at `https://doi.org/10.6084/m9.figshare.7670849`. The GRACE-REC product driven by the ERA5 climatic data is used for comparison.

[c]The generated data product is available at `https://github.com/strawpants/twsc_recon`.

[a]The CSR GRACE product is downloaded from `http://www2.csr.utexas.edu/grace/`.

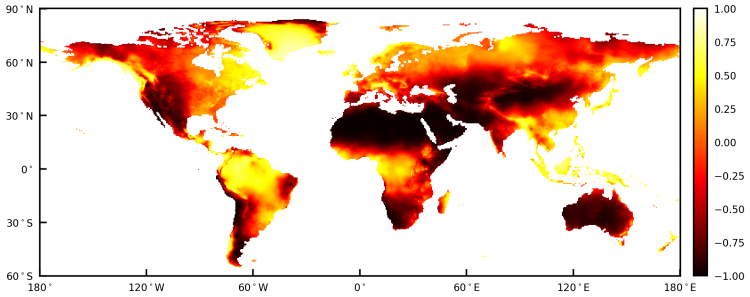

Figure A1. A global map of the aridity index, with -1 and 1 indicating the most arid and humid conditions, respectively. The data is available at `https://doi.org/10.5523/bris.16ctquxqxk46h2v61gz7drcdz3`.

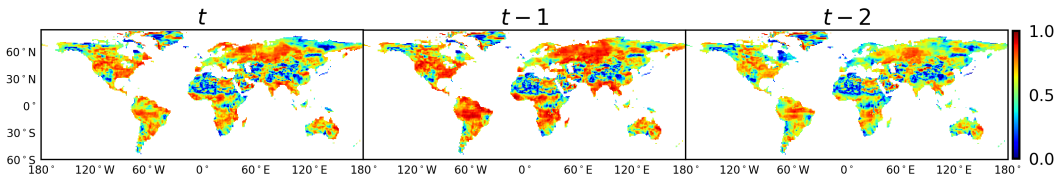

Figure A2. Spatial distribution of the absolute correlation coefficient ($|R|$) between the cumulative water storage changes (CWSCs) and GRACE TWSAs (April 2002-June 2017, June 2018-August 2020), where $(t-i)$ denotes i-month lag.

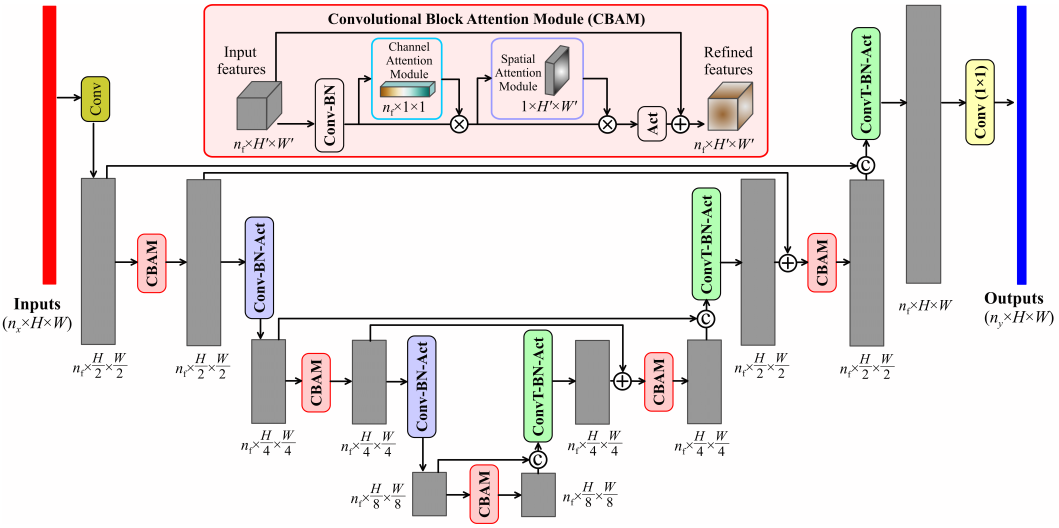

Figure A3. Illustration of the Bayesian convolutional neural network (BCNN) architecture. BCNN takes n_x images with a size of $H \times W$ as inputs and generates n_y images with the same size. It is an alternating cascade of convolutional (Conv)/transposed convolutional (ConvT) layers and convolutional block attention modules (CBAM), each of which outputs $n_\mathrm{f} = 48$ feature maps. The size of feature maps is sequentially halved in each Conv layer from $H \times W$ to $\frac{H}{8} \times \frac{W}{8}$ so as to extract multi-scale features, and then sequentially recovered to $H \times W$ using the ConvT layers. The symbols $\oplus$, ⓒ, and $\otimes$ denote the addition (i.e. residual connection), concatenation (i.e. dense connection), and multiplication (i.e. attention connection) operations, respectively. The Mish activation (Act) function (Misra, 2019) is used in the network. BN=batch normalization.

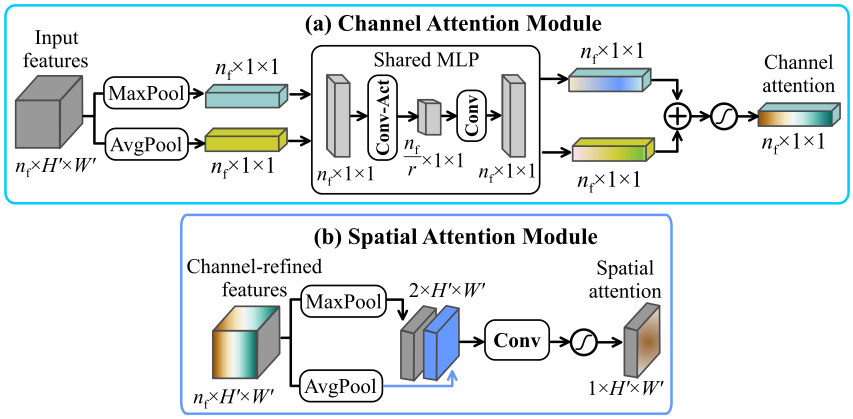

Figure A4. Diagrams of the channel and spatial attention modules. The inputs to each module are n_f feature maps with a size of $H^{'} \times W^{'}$. The channel attention module utilizes both max-pooling (MaxPool) outputs ($n_\mathrm{f} \times 1 \times 1$) and average-pooling (AvgPool) outputs ($n_\mathrm{f} \times 1 \times 1$) with a shared multi-layer perceptron (MLP) to produce a channel attention map. The spatial attention module utilizes similar two outputs (each with a shape of $1 \times H^{'} \times W^{'}$) to produce a spatial attention map. The sigmoid activation is used to guarantee the values in the attention maps are between 0 and 1. Conv=Convolution; Act=Activation.

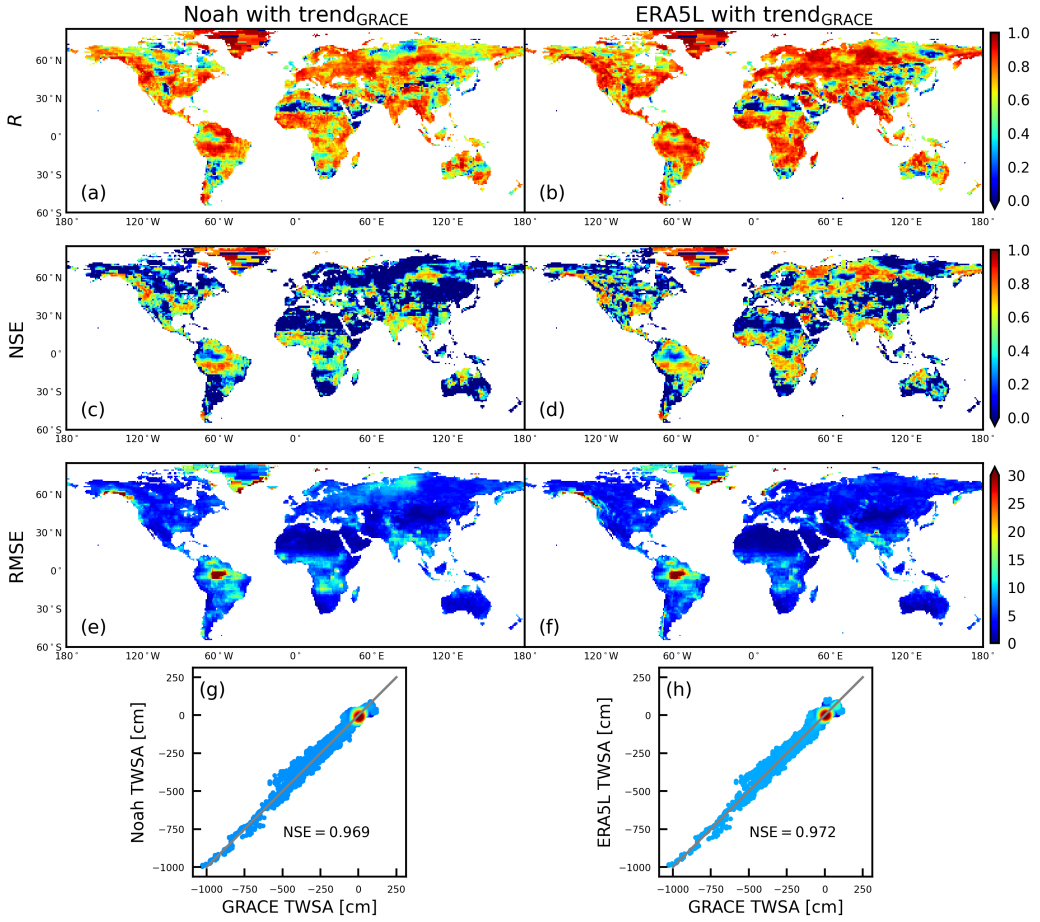

Figure A5. Spatial maps of R (row 1), NSE (row 2), and RMSE (row 3) between the GRACE TWSAs and Noah- (left) and ERA5L-derived (right) TWSAs during the test period (April 2014-June 2017, June 2018-August 2020). Row 4: The density scatter plots between the GRACE and modeled TWSAs. The Noah and ERA5L TWSAs have been corrected with GRACE TWSAs' trend (trend$_{\mathrm{GRACE}}$). That is, TWSA$_m$=`detrend`(TWSA$_m$)+trend$_{\mathrm{GRACE}}$, where `detrend`$(\cdot)$ denotes the linear detrending operation, and TWSA$_m$ represents the Noah or ERA5L TWSAs.

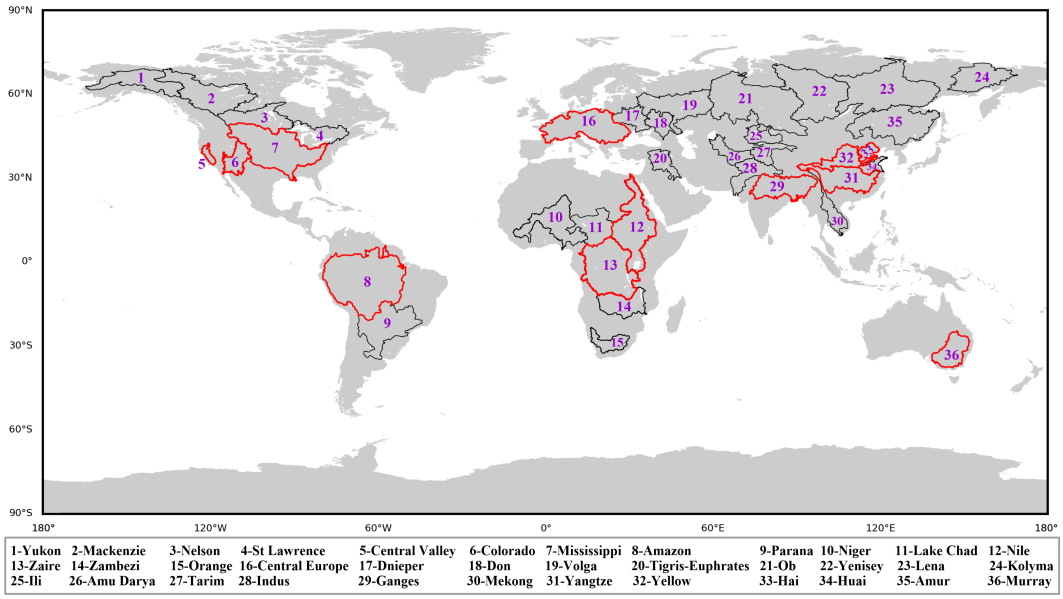

Figure A6. Locations of the regions and river basins considered for result analysis.

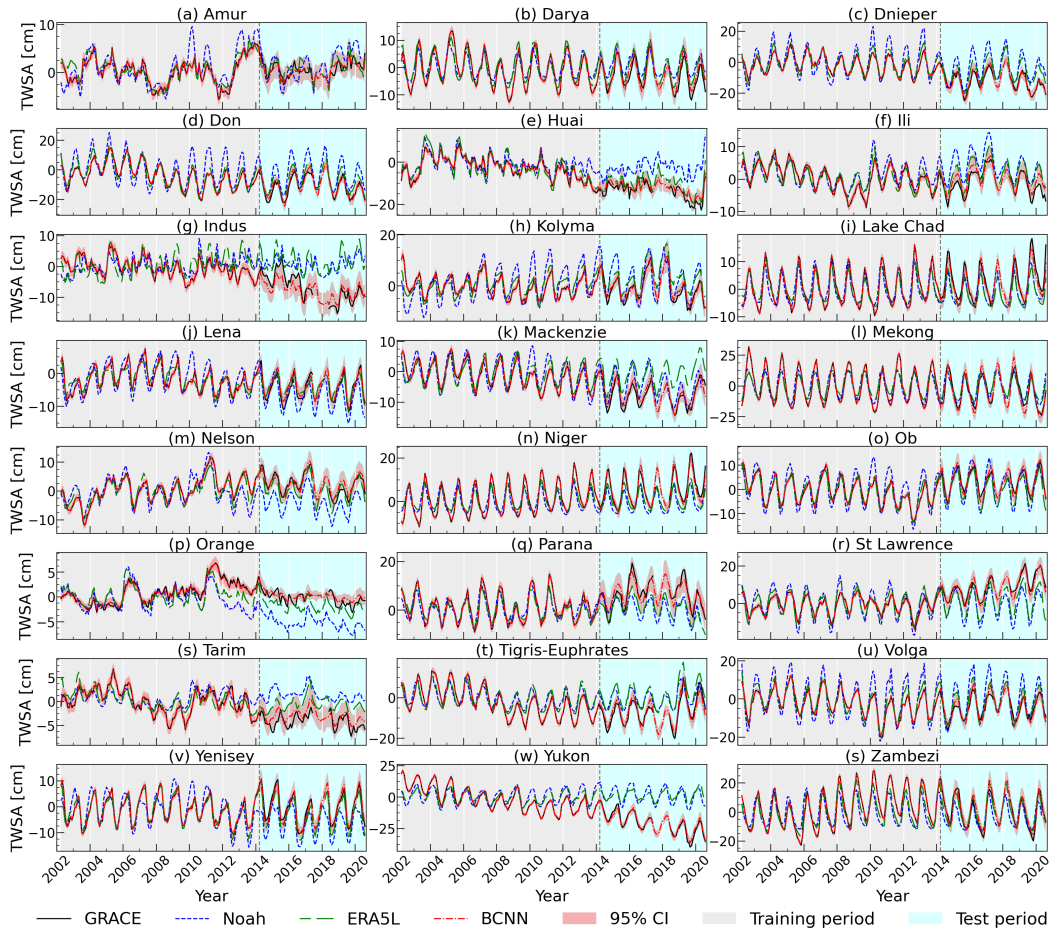

Figure A7. The basin-averaged GRACE, Noah, ERA5L, and BCNN TWSA time serials in different river basins. Shaded areas represent the 95% condence interval (CI) of BCNN predictions.

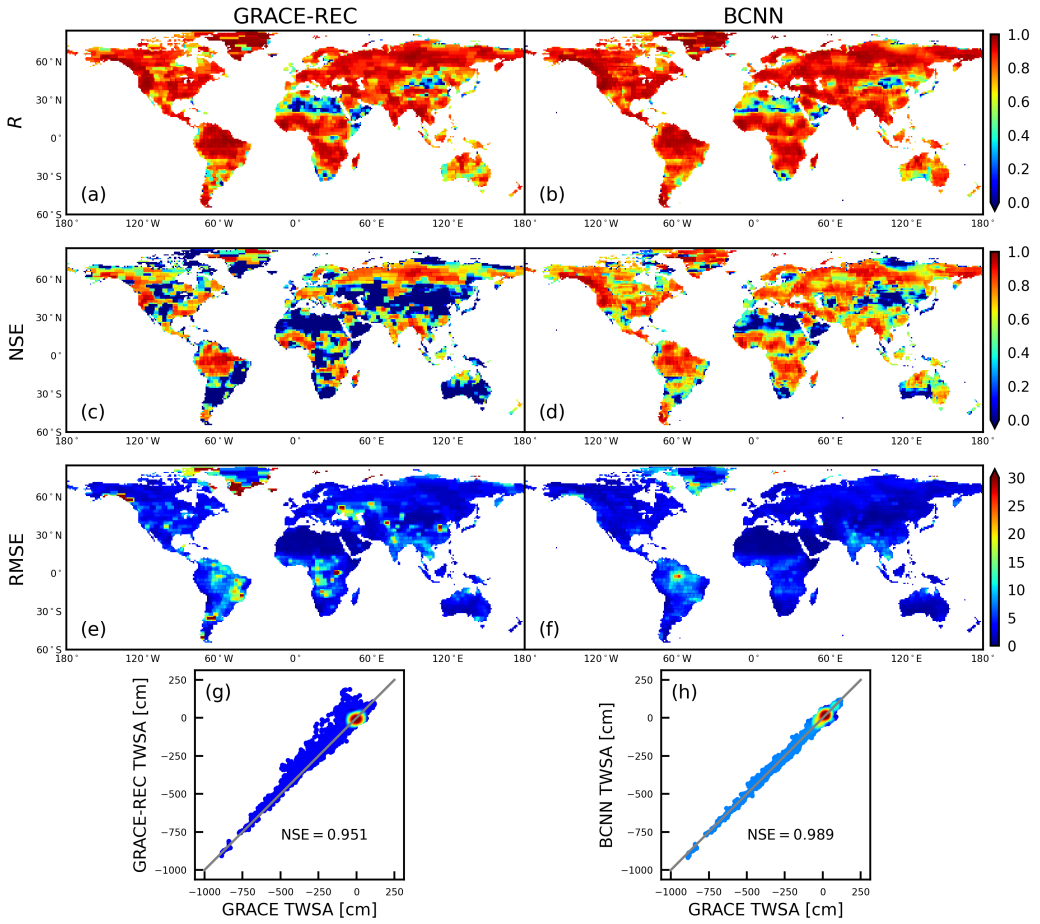

Figure A8. The predictive R (row 1), NSE (row 2), and RMSE (row 2) accuracy of GRACE-REC (left) and our BCNN (right) for the GRACE TWSAs (test period: April 2014-June 2017, June 2018-July 2019). Row 4: The density scatter plots between the GRACE and modeled TWSAs. The trending and seasonal signals extracted from GRACE TWSAs and Humphrey et al. (2017), respectively, have been added to the original GRACE-REC TWSAs for consistency.

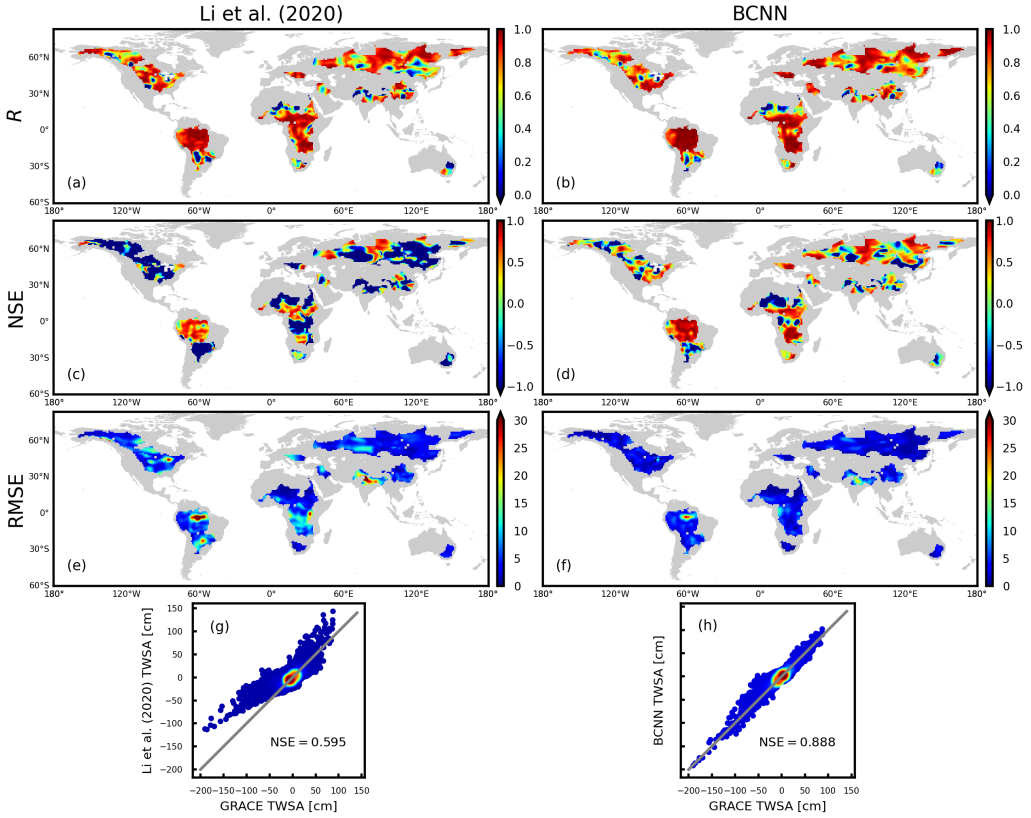

Figure A9. The predictive R (row 1), NSE (row 2), and RMSE (row 3) accuracy for the GRACE TWSAs obtained by F. Li et al. (2020) (left) and our BCNN (right). Row 4: The density scatter plots between the GRACE and modeled TWSAs. The test period is June 2018-December 2018. F. Li et al. (2020) provided the predicted TWSAs for 26 major river basins over the world.

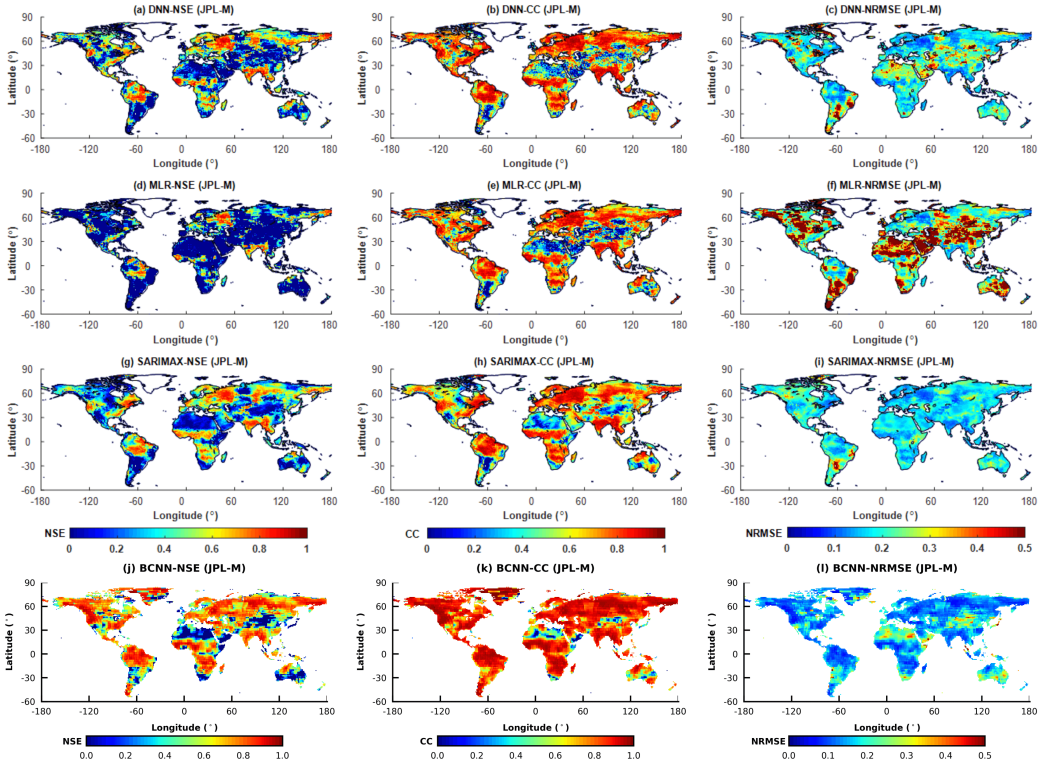

Figure A10. The predictive NSE (column 1), CC (i.e. R; column 2), and normalized RMSE (NRMSE; column 3) accuracy for the GRACE TWSAs obtained by Z. Sun et al. (2020) (rows 1-3 with deep neural network (DNN), multiple linear regression (MLR), and seasonal autoregressive integrated moving average with exogenous variables (SARIMAX) methods, respectively) and our BCNN (row 4). The test period is February 2014-June 2017. (a-i) were extracted from Figure S7 of Z. Sun et al. (2020).