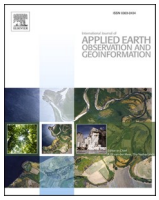
International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102404

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Reconstructing GRACE-like TWS anomalies for the Canadian landmass using deep learning and land surface model

Qiutong Yu [a], Shusen Wang [b], Hongjie He [a], Ke Yang [a], Lingfei Ma [c,*], Jonathan Li [a]

[a] Geospatial Sensing and Data Intelligence Lab, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
[b] Canada Centre for Mapping and Earth Observations, Natural Resources Canada, 560 Rochester St, Ottawa, ON K1S 4M2, Canada
[c] Engineering Research Center of State Financial Security, Ministry of Education, Central University of Finance and Economics, Beijing 102206, China

ARTICLE INFO

Keywords:
GRACE
Land Surface Model
Terrestrial Water Storage
Deep Learning

ABSTRACT

Terrestrial water storage (TWS) is an essential part of the global water cycle. Long-term information of observed and modeled TWS is fundamental to analyze water resources, meteorological extreme events (e.g., droughts and floods), and the climate change impacts. Over the past several decades, hydrologists have been applying physically-based hydrological model (GHM) and land surface model (LSM) to simulate TWS and its components (e.g., groundwater storage). However, the reliability of these physically-based models is often affected by uncertainties in climatic forcing data, model parameters, model structure, and mechanisms for physical process representations. Launched in March 2002, the Gravity Recovery and Climate Experiment (GRACE) satellite mission exclusively applies remote sensing techniques to measure the variations in TWS on a global scale. The mission length of GRACE, however, is too short to meet the requirements for analyzing long-term TWS. Therefore, lots of effort have been devoted to the reconstruction of GRACE-like TWS data for the pre-GRACE era. Data-driven methods, such as multilinear regression and machine learning, exhibit a great potential to reconstruct TWS data by integrating GRACE observations and physically-based model simulations. The advances in artificial intelligence enable adaptive learning of correlations between variables in complex spatiotemporal systems. However, the applicability of various deep learning techniques has not been adequately studied for GRACE TWS reconstruction. In this study, three deep learning-based models are developed to reconstruct the historical TWS using LSM outputs for the Canadian landmass from 1979 to 2002. The performance of the models is evaluated against the GRACE-observed TWS in 2002–2004 and 2014–2016. The trained models achieve a mean correlation coefficient of 0.96, with a mean RMSE of 53 mm. The results show that the LSM-based deep learning models significantly improve the correlations between original LSM simulations and GRACE observations.

1. Introduction

Terrestrial water storage (TWS) is referred to water storages both underneath and above the Earth's surface (Jing et al., 2020). TWS is considered as the main component of terrestrial and global hydrological process. Its dynamics under different scenarios of environmental changes including climate change and human disturbance is key to determine water resources sustainability and vulnerability (Famiglietti et al., 2011). As such, the change in TWS is considered as one of the key parameters to be studied for the assessment of hydrological cycle. The earlier methods for quantifying TWS have been mostly relied on global hydrological models (GHM) (Khaki et al. 2017; Shokri et al. 2018) or land surface models (LSM) (Kumar et al. 2017; Nie et al. 2019) which integrate some atmospheric and surface properties with in-situ observations (Tourian et al., 2018). However, the construction of these models is based on principles of physical processes of water storage, which requires large number of ground-based data that are costly, time-consuming, and restricted to a sparse set of in-situ monitoring stations. As a result, inadequate spatial and temporal coverage of ground-based observations and uncertainties in storage limit the understanding of water storage changes at large scales.

The emergence of satellite remote sensing (RS) enabled continuous monitoring of hydrological fluxes at different spatial resolutions (Becker et al., 2018). The variations in TWS change the gravity filed over a

* Corresponding author.
E-mail address: l53ma@cufe.edu.cn (L. Ma).

https://doi.org/10.1016/j.jag.2021.102404
Received 23 April 2021; Received in revised form 7 June 2021; Accepted 15 June 2021
Available online 23 June 2021

region which can be effectively detected at large scales by the Gravity Recovery and Climate Experiment (GRACE) satellite launched in March 2002. So far, GRACE and its follow-on satellites are the only RS-based method to quantify long-term TWS anomalies (TWSA) at large-scales (Heimhuber et al., 2019). The primary reason for temporal changes of the Earth's gravity field is the redistribution of water mass within thin fluid envelope of the Earth, and GRACE enables to detect tiny changes in the Earth's mass redistributions related to spatiotemporal variations of TWS at monthly time scale (Dankwa et al., 2018). The activities in producing high quality and long-term datasets for TWS normally involve the integration of various datasets from satellite observations, in-situ observations, and outputs from land surface and climate models.

Previous studies have employed the fusion of GRACE data into global hydrologic and land surface models as an attempt to enhance the model's prediction skills. A combination of Global Land Data Assimilation System (GLDAS), statistical models, and GRACE was applied by Yang et al. (2017) to estimate TWS variations within Tarim River basin from 2002 to 2015. Scanlon et al. (2018) evaluated the trends in land water storage comparing global hydrologic and water resource models, global LSMs, and GRACE data over 186 global basins. They found a large spread in the results of these models, weak matching rules between the models and the GRACE solutions, and that the global models may underestimate the trend in future climate change and human-induced water storage variations. A statistical model trained with GRACE data was applied by Humphrey and Gudmundsson (2019) to directly build past climate-driven variations of TWS from historical and meteorological datasets on daily and monthly basis.

However, the observations from the GRACE satellite have data available only since 2002, which does not meet the requirement for producing a baseline TWS information that can be used to calculate the Climate Normal which requires at least 30 years (Arguez et al., 2019). Various methods have been developed recently for the reconstruction of missing information in RS data. But most reconstruction methods are based on linear models and can only be used under limited conditions. This limitation contributes to the difficulty in handling complex surfaces and large amount of missing data (Shen et al., 2015). For some complex spatiotemporal dynamical systems like atmosphere and aquifer, traditional numerical models are often not capable to make prediction due to the lack of knowledge about the systems' inner mechanisms. On the other hand, deep learning-based methods have proven useful for making accurate predictions for complex spatiotemporal systems, by which the systems' inner mechanisms can be learned based on the historical data (Shi & Yeung, 2018).

In fact, data-driven methods such as machine learning (ML) and deep learning (DL) can serve as useful modeling techniques to quantify and simulate TWS for better prediction and deeper understanding of water cycles (Sun et al., 2019). In recent years, learning-based models have been increasingly used to predict hydrological variables by learning the correlation between the main variables and other related parameters (Sahoo et al., 2017; Hamshaw et al., 2018; Broxton et al., 2019; Kim et al., 2019; Quilty et al., 2019). For instance, Artificial Neural Network (ANN) model is one of the prevalent methods to reconstruct GRACE-like time series dataset (Long et al., 2014; Zhang et al., 2016). The fusion of ANN model into GRACE data was applied by Chen et al. (2019) to reconstruct the historical TWS record in Songhua River basin, Northern China. Mukherjee and Ramachandran (2018) applied support vector machine, ANN, and linear regression model on GRACE-derived TWSA to predict ground water level. Their finding suggests that the performance of the models improved when climatic observations are integrated with GEACE TWS data. In addition to ANN-based model, Sun et al. (2019) reconstructed GRACE TWSA data of India's landmass by integrating LSM with convolutional neural networks (CNN). The results revealed that the CNN model can effectively enhance the performance of LSMs with providing the water components. Nevertheless, previous efforts are only proven to be applicable in certain cases, which may not apply in other basins, especially those with harsh climate, arid climate, or intense human interventions (Jing et al., 2020). In different environmental scenarios, the correlations between hydrological variables (i.e., LSM simulations) and GRACE observations cannot be generalized. Moreover, only a handful of DL architectures were applied to reconstructing GRACE-like TWS, it is necessary to examine the capabilities of more algorithms and architectures such as recurrent neural network (RNN) and generative adversarial network (GAN).

The objective of this study is to develop DL-based models for reconstructing the historical terrestrial water storage datasets for Canada's landmass, based on the statistical correlations between LSM-simulated TWSA and GRACE-derived TWSA during the first GRACE mission (2002–2017), so that the existing TWS records can be extended and improved for generating the baseline TWS dataset for Canada. Three types of DL architectures (CNN, conditional GAN (cGAN), and deep convolutional autoencoder (DCAE)) are trained to learn the spatiotemporal pattern of the correlations between GRACE observations and corresponding LSM simulations. These trained models are able to predict GRACE-like TWSA using LSM simulation as inputs (i.e., without requiring observed GRACE TWS as inputs). In addition, a convolutional LSTM (ConvLSTM) model is trained to learn the temporal variations in the GRACE TWSA time series, which considers the input data as sequences and make predictions only based on temporal trends existed in GRACE data (i.e., without referencing to the LSM simulations). The results of the four models are compared in order to choose the optimal method for the final reconstruction. In the following, Section 2 describes the data used, as well as the data pre-processing methods. Section 3 presents the methodology and implementation details of model constructions. The results are presented and discussed in Section 4 and lastly, Section 5 states the main conclusions and findings.

2. Materials

2.1. GRACE TWSA data

The GRACE monthly TWSA product (RL06 spherical harmonics solution) was downloaded from the Jet Propulsion Laboratory of NASA (https://podaac.jpl.nasa.gov). The product is provided with a grid resolution of 1 degree (~110 km) at global scale. The original GRACE TWSA data measures the deviations from the mean TWS between January 2004 to December 2009 (i.e., the baseline value). In this study, the baseline value was adjusted to the mean TWS between April 2002 to December 2016 (i.e., the study period. The data was clipped by a Canadian national boundary shapefile to cover the Canada's landmass (Fig. 1), and then reprojected to Canada Lambert Conformal Conic (CanLCC) projection. Fig. 2 shows the GRACE TWS trend during the study period. The trend exhibits spatial variations over the study area, with strong negative trends in Yukon and northern Arctic region. The significant long-term water loss in these regions is presumably from

Fig. 1. Map of the study area. Canada's landmass is represented by brown color.

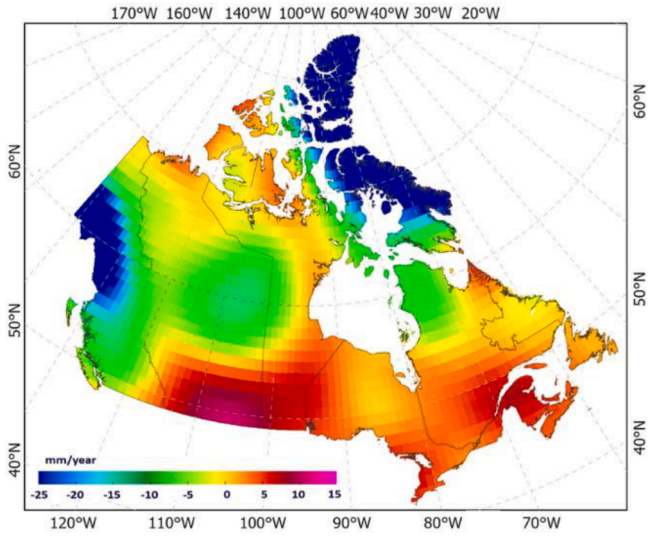


Fig. 2. Trend of GRACE TWS in the study area, calculated as the Sen's slope of the TWS over 2002–2016.

melting glaciers and permanent snow (Wang et al., 2015; Wang & Li, 2016).

2.2. EALCO-simulated TWSA data

The EALCO (Ecological Assimilation of Land and Climate Observation) model is a LSM developed by Natural Resources Canada. EALCO simulates the energy, water, and carbon dynamics by assimilating land and meteorological observation information, which can provide long-term hydrologic information with a relatively high spatial and temporal resolution (Wang et al., 2013, 2014). However, EALCO does not provide complete information brought by the GRACE satellites, such as glacial melt and human-induced water storage changes. As can be seen from Fig. 3, EALCO does not reflect the decreasing trend observed by GRACE. Therefore, the TWS data simulated by the EALCO LSM is not applicable to directly reconstruct the historic TWS dataset.

The EALCO (v4.2) daily TWS product used in this study covers the period from January 1979 to December 2016, with a spatial resolution of 5 km. EALCO TWS was calculated from simulations for soil water content, snow water equivalent and canopy water. The simulations and calibration were independent from GRACE TWS data. The EALCO TWSA was calculated by subtracting its baseline value during the same time period as the GRACE TWSA. Then the EALCO TWSA data was upscaled to 110 km resolution using the arithmetic mean to match the spatial resolution of the GRACE TWSA data. The EALCO data was processed to have the same extents and coordinate system as the GRACE data, in order to ensure the pixelwise match between EALCO and GRACE. Lastly, the daily EALCO TWSA data was averaged into month values according to the temporal coverage of GRACE TWSA. EALCO data prior to the GRACE mission (before April 5, 2020) was resampled to 30-days averages and prepared for reconstructing the historical TWS dataset. As a result, there are 158 GRACE-EALCO pairs (see Fig. 4) representing all the GRACE coverage period from April 2002 to December 2016. To be compatible with the convolutional operations during the model training, the image dimension of both datasets is clipped to 48 pixels × 48 pixels.

2.3. Train-test split and further processing

In the total of 158 GRACE-EALCO monthly data pairs, the first 20 months (April 2002–February 2004) and the last 20 months (October 2014–December 2016) are used as testing data. The rest 118 months are used for training. Moreover, data augmentation techniques (e.g., flipping, rotation) were applied to the training set to deal with the problem of insufficient amount of data. Lastly, EALCO TWSA and GRACE TWSA are separately normalized to the range of [−1, 1] for facilitating the model training process.

3. Methods

Let y denote the GRACE TWSA as the predictand variable, and X be the EALCO TWSA as the predicator. The three DL models (CNN, cGAN, and DCAE) are trained to solve a regression problem with paired training data $X = \{X_i\}_1^N$ and $y = \{y_i\}_1^N$, as stated in the formula:

$$y = f(X, p) \tag{1}$$

where N represents the total number of training samples, i is the index of training samples, and p denotes the model parameters. The trained models are able to take EALCO data as inputs and predict reconstructed GRACE-like TWSA without GRACE observations. Since the study task is essentially a regression problem, the hyperbolic tangent function (tanh) is adopted as the activation function of the final output layer and mean-square-error (MSE) is used as the global loss function for these three models.

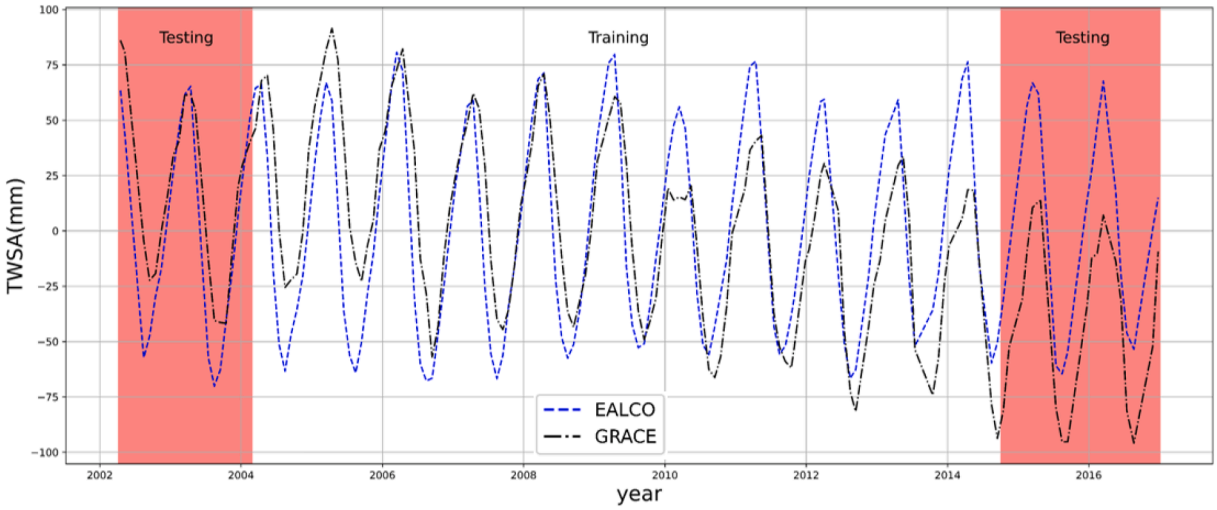


Fig. 3. Comparison of nationwide mean EALCO-simulated TWSA and mean GRACE-derived TWSA. Red background indicates the time coverage of the testing set. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

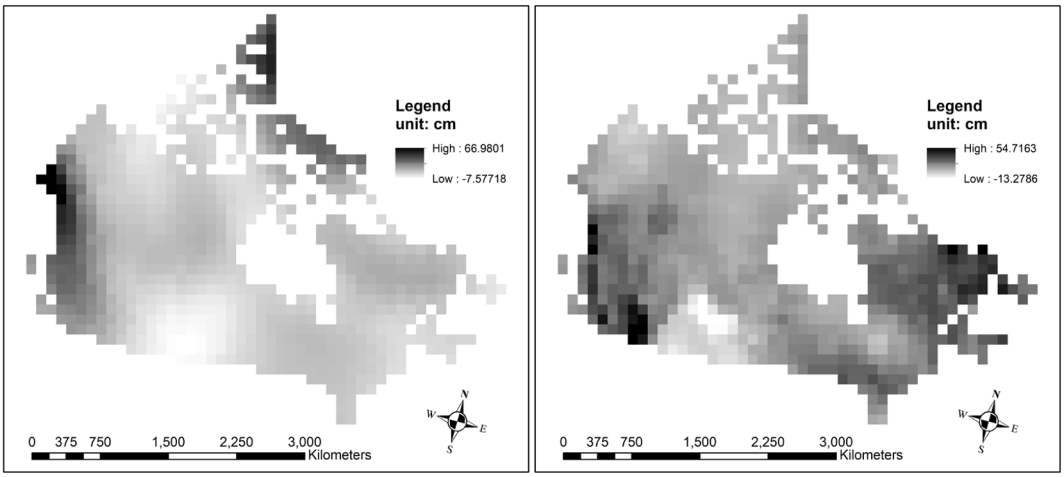


Fig. 4. GRACE TWSA (left) and upscaled EALCO TWSA (right) for April 2002 under CanLCC projection.

3.1. Squeeze-and-Excitation U-Net CNN

By interleaving a stack of functional layers such as convolutional layer, pooling layer, and fully connected layer, CNNs can extract the distinguishable and representative features from RS images in a hierarchical manner, which can be effectively applied to various RS-related image analysis tasks such as land cover classification (Al-Najjar et al., 2019), ground object detection (Miyamoto et al., 2018) and land use change detection (Cao et al., 2019).

The CNN-based model used in this study is a Squeeze-and-Excitation (SE) network (Hu et al., 2020) based on a U-Net backbone architecture (SEUNet). The SE block is a mechanism which enables the network to perform feature recalibration by using global information to selectively enhance important features and suppress secondary ones (Hu et al., 2020). The operation process of a SE block contains two parts: squeeze operation and excitation operation. The squeeze operation applies a global average pooling to aggregate feature maps across their spatial dimensions and scale each feature map to a real number which characterizes the global distribution of feature responses. The excitation operation takes the global descriptor (output of squeeze operation) and calculates weights for each feature channel, and then applies these weights to the feature maps to generate the final output of the SE block. U-Net has demonstrated excellent performance on small training datasets (Sun et al., 2019). It follows an encoder-decoder structure to conduct gradual transitions from inputs to outputs. In this study, the building blocks of original U-Net architecture were replaced with SE blocks (see Fig. 5).

3.2. Pix2Pix conditional GAN

The GAN architecture is composed of a discriminator model and a generator model. The training of the two models takes place in a synchronous and adversarial manner. The generator produces plausible images to 'deceive' the discriminator, and the discriminator identifies the plausible images. The Pix2Pix is a type of cGAN model developed by Isola et al. (2017), which the generation of the output image is conditional on the input images (see Fig. 6). It has been demonstrated on a range of image-to-image translation tasks such as converting digital maps to satellite imagery (and vice versa), black-white photo to colored photo, and object sketches to object photographs.

The Pix2Pix generator is an encoder-decoder CNN using a U-Net architecture. The model takes a source image (EALCO simulation) and generates a target image (GRACE-like prediction). The generator model is updated to minimize the loss threshold for the discriminator to mark

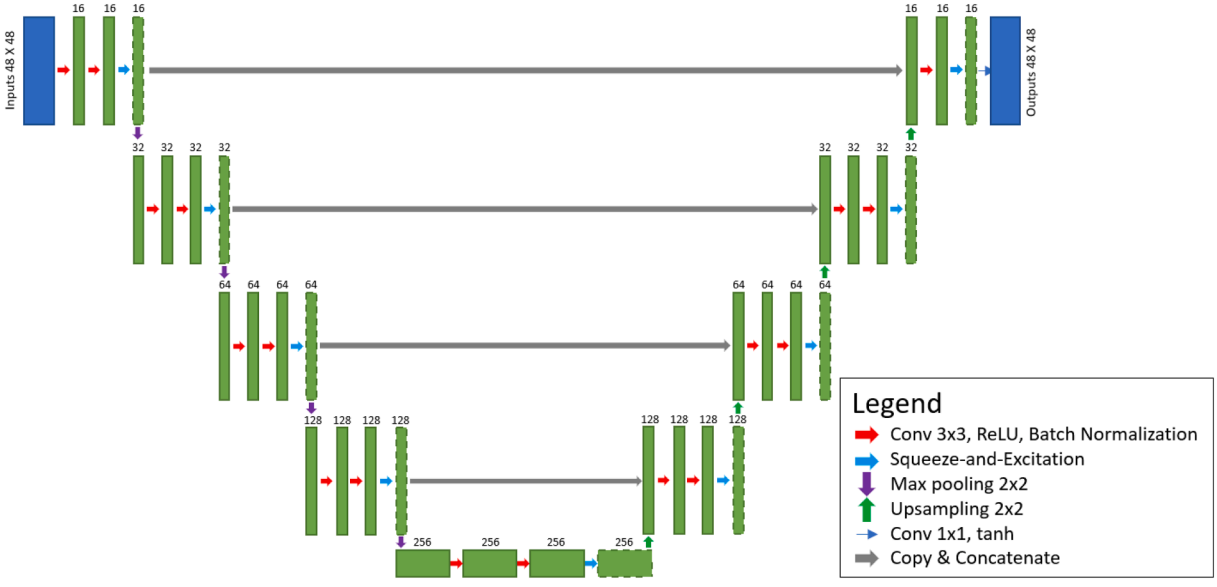


Fig. 5. Schematic diagram of the SEUNet architecture.

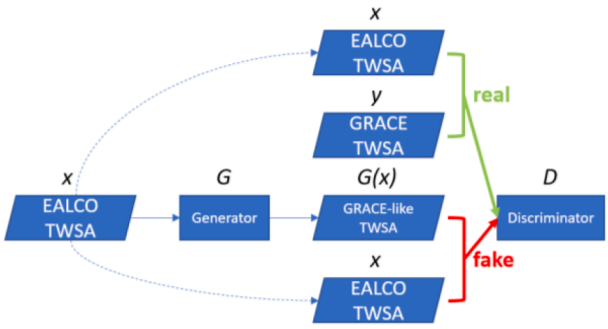


Fig. 6. Diagram of training a Pix2Pix conditional GAN to map EALCO TWSA to GRACE TWSA. The discriminator D learns to classify between fake (generated, EALCO) and real (GRACE, EALCO) tuples. Unlike an unconditional GAN, both the generator and discriminator take the predictor (EALCO) as inputs.

generated images as real. On the other hand, the discriminator is based on a PatchGAN structure that performs conditional classification based on the relationship between the model output and the number of pixels in the input image (Isola et al. 2017). It takes both the source image (GRACE observation) and the target image (GRACE-like prediction) as input and estimate the likelihood of whether the target image is real or a fake translation of the source image.

3.3. Deep convolutional autoencoder

The autoencoder (AE) is also an encoder-decoder structure by which the encoder provides compressed feature representation of the input and the decoder reconstructs the input from the representation (Azarang et al., 2019). In a DCAE model, the encoder consists of convolutional layers, and the decoder is composed of convolutional transpose layers. Comparing with original AE, DCAE is more suitable for image processing tasks such as and colorization of greyscale image restoration of damaged image, and image denoising. The diagram of the DCAE model constructed for this study is shown in Fig. 7.

The encoder contains 3 convolution layers for down-sampling by 3 × 3 kernel and ReLU activation function, thus the number of filters increases gradually from 64 to 256. The third convolutional layer is then flattened and connected a fully connected layer (FC layer). In the bottleneck, the input is converted to a dense representation (a.k.a. latent space representation) as a 128 × 128 tensor, which stores all the crucial information needed for feature detection from the original input data. This representation is connected to the second FC layer for prediction. The predicted values are reshaped to feature map dimensions, which allows the decoder to properly reconstruct the input data to a full image. Symmetrically, the decoder contains 3 convolutional transpose layers for up-sampling by 3 × 3 kernel and tanh activation function.

3.4. Encoding-Forecasting ConvLSTM

ConvLSTM was introduced by Shi et al. (2015) for the task of precipitation nowcasting based on Radar Echo images. They formulate

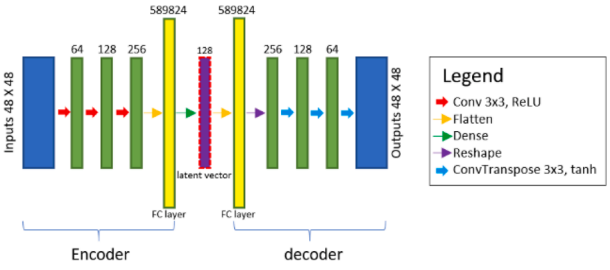


Fig. 7. Schematic diagram of the proposed DCAE architecture.

precipitation nowcasting as a spatiotemporal sequence forecasting problem that can be solved under the general sequence-to-sequence learning framework. For general-purpose sequence modeling, Fully-connected long short-term memory (FC-LSTM) has proven useful for modeling long-range temporal dependencies. FC-LSTM is an encoder-decoder forecasting model which reconstructs the input sequence and predicts the future sequence simultaneously. However, FC-LSTM does not take spatial texture into consideration, which can be only used to conduct time-series modelling on 1D tabulated data. ConvLSTM extends the FC-LSTM to have convolutional structures in each of its transitional stages, which can be directly applied on 2D image sequences. By using a convolution operator, the state of each grid/pixel in the input image is determined by its neighboring grids.

By stacking multiple ConvLSTM layers, Shi et al. (2015) developed an encoding-forecasting structure (see Fig. 8) to build an end-to-end trainable model for Radar Echo precipitation nowcasting. Their experiment results show that ConvLSTM is better than FC-LSTM in handling spatiotemporal correlations.

In this study, the Encoding-forecasting ConvLSTM model is used to compare with other three models that are constructed by the hybrid modelling approach. The inputs of the ConvLSTM model are the time series of GRACE TWSA. To achieve that, the training and testing datasets require further processing to be transformed to sequenced inputs. In the data-loading module of the model, the consecutive GRACE images were sliced with a 20-frame-wide sliding window (1 step sliding). Therefore, each sequence consists of 20 frames. The first 19 frames are for input, and the last frame is for the model to make prediction and evaluation, as shown in Fig. 9. The 19 input frames (i.e., 19 GRACE months) are set to ensure seasonal spatiotemporal trends are handled by the network. During the training process, the model updates its weights based on the prediction result of the last frame (as comparing to the ground truth of the last frame).

Additionally, the model has two modes: forecasting mode and hindcasting mode. The two modes are trained separately on different inputs. For forecasting mode, the sequenced inputs are constructed by starting from the first GRACE month (April 2002). For hindcasting mode, the data was sequenced in reversed order (starts from December 2016).

3.5. Implementation details and evaluation metrics

All models were implemented on the open-source package Keras 2 using Python 3.7. All models are trained for 100 epochs. Adam optimizer was used to train the models with a learning rate of 0.01 and batch size of 1, as recommend by Isola et al. (2017) for image-to-image translation task. All experiments were carried out on a Windows10 desktop running with sole GPU (NVIDA 2070-super, 8 Gb RAM). A detailed workflow chart for this study is presented in Fig. 10.

To evaluate the performances of the four trained models, Pearson's correlation coefficient (CC) and root-mean-square error (RMSE) between GRACE TWSA and model-predicted TWSA are calculated by the following formulas:

$$CC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(g_i - \bar{g})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(g_i - \bar{g})^2}} \quad (2)$$

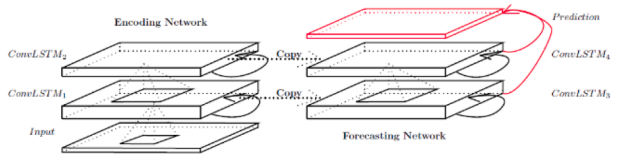


Fig. 8. Encoding-forecasting ConvLSTM architecture. *Source:* Shi et al. (2015).

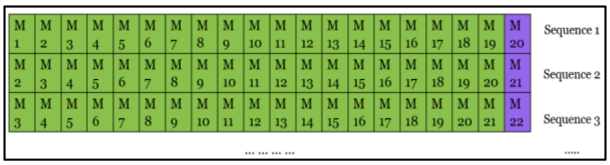


Fig. 9. Illustration of the sequence modelling. Cells in green are input frames and predicted/evaluated cells are in purple. The first sequence starts from the first month, and the second sequence starts from the second month. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - g_i)^2} \quad (3)$$

where y is the predicted TWSA, g is the GRACE-observed TWSA, n is the number of samples in the testing set. The range of CC is [−1, 1] which measures the linear correlations between predicted TWSA and GRACE-derived TWSA.

4. Results and discussion

4.1. Comparison of model-predicted TWSA

For EALCO and trained DL models, the CC and RMSE between the simulated/predicted TWSA and GRACE TWSA at both the nationwide level and pixelwise level are compared. Magnitudes and spatial distribution of pixelwise mean CC and mean RMSE of EALCO and trained DL models during the testing periods are shown in Fig. 11. It is worth noting that the metrics between EALCO and GRACE is the baseline for the performance assessment of DL models. Fig. 11a and 11f illustrates the comparison between EALCO-simulated TWSA and GRACE-observed TWSA in the testing set. It can be seen that there are large discrepancies between GRACE and EALCO in the Arctic region and Tatshenshini-Alsek provincial park, where both have spectacular glacier and icefield landscapes. As for DL-corrected TWSA, DCAE spatially outperforms other three models, and the results of SEUNet and Pix2Pix demonstrate similar spatial pattern while SEUNet slightly outperforms Pix2Pix. As shown in Fig. 11(g–j), most of uncertainties exist in the glacial and coastal areas. Specifically, the predicted TWSA has higher accuracy in inland regions than in landmass edges (including southern borders). The errors could be derived from multiple factors. For example, the broken landmass

(e.g., island, peninsula), the geological attributes of tundra and glacier, the agricultural activities in southern Ontario, and the interpolation of ocean grids. It is noticeable that ConvLSTM performs significantly poorer than other three DL models on the testing set. Even though ConvLSTM does not take EALCO TWSA with LSM-related uncertainties, its RMSE exhibits similar spatial pattern as the RMSE of EALCO TWSA, while its correlation with GRACE is generally higher than that of GRACE and EALCO. It could be caused by the model parameters and the uncertainties (or dramatical variations) of GRACE TWSA in cold regions, which makes the temporal changes of TWSA unpredictable. Overall, the three hybrid methods (Pix2Pix, SEUNet, DCAE) demonstrate satisfactory predictability in the study area, and significantly improve the TWSA modelling as comparing to original EALCO simulations (baseline) and ConvLSTM predictions.

The results of nationwide mean CC and mean RMSE are summarized in Table 1. As comparing to GRACE observations with EALCO simulations, the mean CC is 0.89 and the mean RMSE is about 105 mm. As mentioned previously, the correlation strength between physically-based LSM and GRACE is dependent on the hydrometeorology and the structure/parameterization of the LSM, which does not reflect all factors attributing to the variations in TWSA. At the nationwide level, the three hybrid methods (Pix2Pix, SEUNet, DCAE) all achieves high correlation (>0.96) with the GRACE TWSA. And the ConvLSTM modeled TWSA also has slightly higher correlations (0.93) with the GRACE TWSA than EALCO does. As for the model predictability, ConvLSTM has a mean RMSE of 89 mm, which improves the baseline by approximately 15%. Pix2Pix and SEUNet have mean RMSE of 67 mm (36% improvement over the baseline) and 64 mm (38% improvement over the baseline), respectively. DCAE is the optimal method for enhancing EALCO simulations, by which the mean CC is 0.99 and the mean RMSE is about 53 mm, resulting in 49% improvement over the baseline.

Fig. 12 plots the nationwide mean TWSA time series from four trained DL models during the GRACE mission from April 2002 to December 2016. For comparison, the EALCO-simulated TWSA (blue dash line) and GRACE-derived TWSA (black dash dot line) are also plotted. It can be seen from the figure that the fluctuations indicate the seasonal variations in TWSA. EALCO and all trained models are capable to fit the drying trends and wetting trends by capturing the seasonal variations. The ConvLSTM model tends to underestimate the dry conditions and overestimate the wet conditions, throughout the training and testing. In contrast, other three DL models achieve high accuracy during the training phases. But in the testing phases, the Pix2Pix model and the SEUNet model clearly deviate from the GRACE observations as these two models overestimate TWSA. DCAE performs well during both training and testing phases.

To sum up, the ConvLSTM model performs poorly throughout the training and testing. During the training periods, the performances of the three hybrid methods are all significantly better than the original EALCO TWSA. During the testing periods, the performances of DL models fade slightly, but the models generally still outperform original EALCO. The results further suggest that, by learning the relationship between EALCO simulations (as the predictor) and GRACE observations (as the predictand variable) in pairs, the EALCO simulation results can be significantly improved, which give a better prediction for TWSA.

The selected examples of reconstructed monthly TWSA maps for four seasons during the testing periods are demonstrated in Figs. 13–16, corresponding to the TWSA spatial distributions in January 2003, July 2003, October 2014, and April 2015, respectively. It is worth

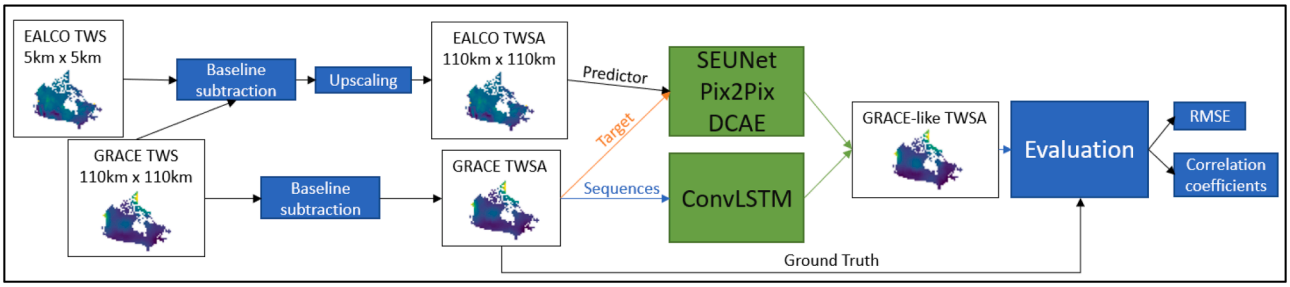


Fig. 10. Workflow diagram. EALCO-simulated and GRACE-derived TWSA were calculated by subtracting the baseline value (mean TWS between April 2002 and December 2016). The spatial resolution of EALCO TWSA was then upscaled from 5 km to 110 km to be matched with that of GRACE TWSA. SEUNet, Pix2Pix and DCAE take both EALCO and GRACE TWSA as inputs, and ConvLSTM only takes GRACE as input sequences.

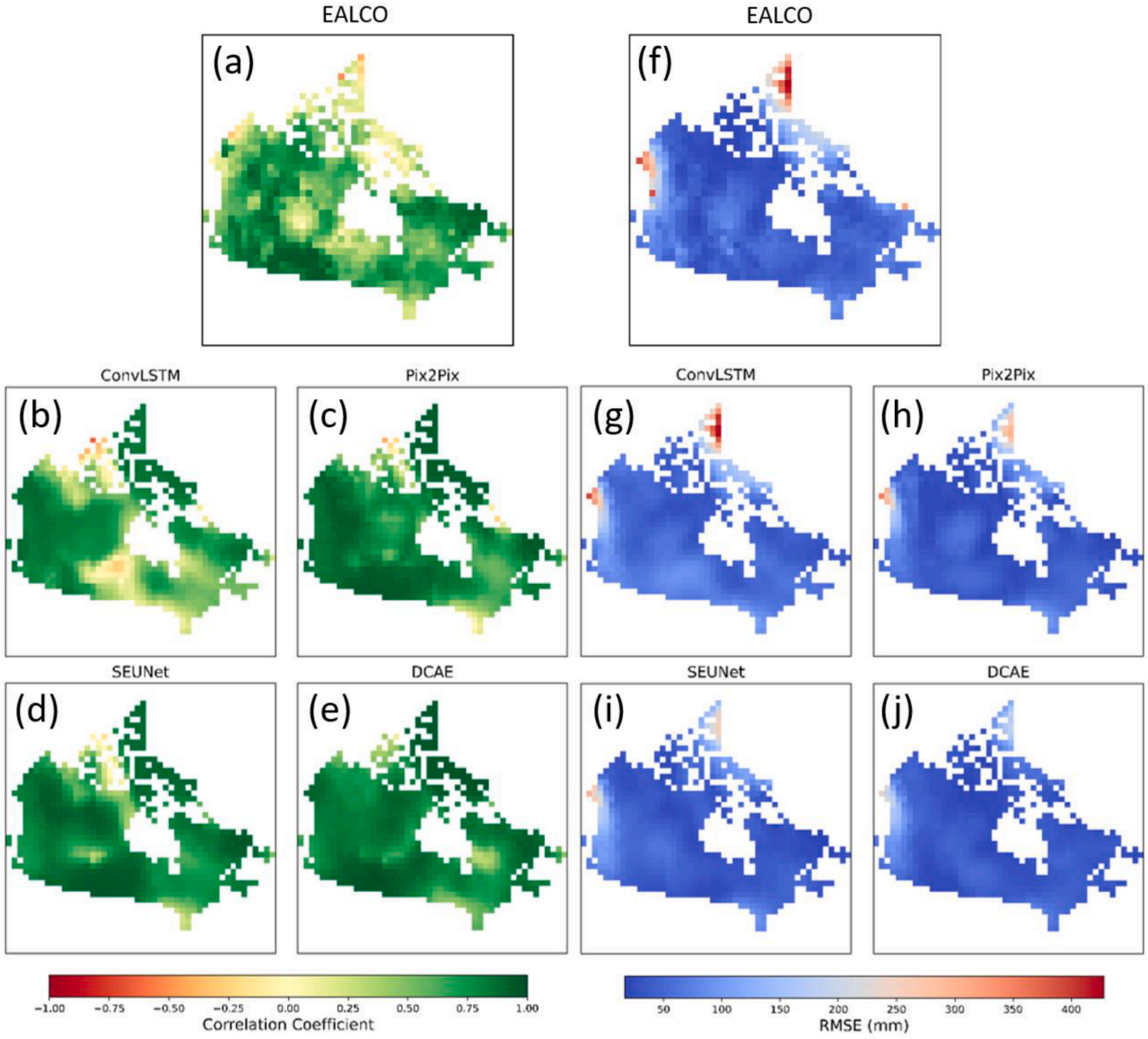


Fig. 11. Spatial distributions of pixelwise mean CC (a–e) and mean RMSE (f–j) derived from EALCO, ConvLSTM, Pix2Pix, SEUNet and DCAE.

Table 1
Comparison of nationwide mean CC and mean RMSE.

Model	# Parameter	$RMSE_{test}$ (mm)	CC_{test}
EALCO	N/A	105	0.89
ConvLSTM	1,858,049	89	0.93
Pix2Pix GAN	34,544,514	67	0.96
SEUNet	1,964,093	64	0.96
DCAE	76,125,313	**53**	**0.99**

mentioning that April and October are normally the time when most regions of Canada reach maximum and minimum TWS values (Wang et al., 2014b).

There are several factors that may cause the ConvLSTM model failed to achieve expectations. First of all, the amount of GRACE data is insufficient due to the temporal resolution and mission length of GRACE, resulting in only 158 available data samples. According to previous studies (Shi et al., 2017; Zhu et al., 2018) on applying ConvLSTM model to predict geospatial time series, the training was normally taken on more than 10,000 consecutive samples, given the nature of the study data such as precipitation and wind speed. For instance, the original Encoding-Forecasting ConvLSTM model was built to nowcast the precipitation from radar echo images which record the reflectance intensity, movement and thickness of clouds, and the time interval between each input frame is normally about 5–10 min (Shi et al., 2015). Additionally, radar echo data can be seen as a dynamic feature with predictable trajectories (the movement of clouds). On the contrary, TWS data is location-stationary in which can be seen as a static feature. Furthermore, the ConvLSTM model was expected to detect the long-term trends (e.g., deglaciation) in the TWSA data. However, the GRACE observation is technically not consecutive, which only offers the TWS measurements for certain durations. This discontinuity may also cause the model predictions to be deviated from the observations.

By incorporating physical model simulations, the three hybrid modelling approaches (Pix2Pix, SEUNet, DCAE) focus on the pixel/patch-based correlations between the EALCO TWSA and GRACE TWSA. As a result, the models' performances are not affected by the discontinuity of the GRACE observations. The trained models are able to predict TWSA for all times whenever EALCO TWSA data are available. Nevertheless, the long-term trends exist in the TWS data can affect the prediction accuracy because the model cannot detect those trends.

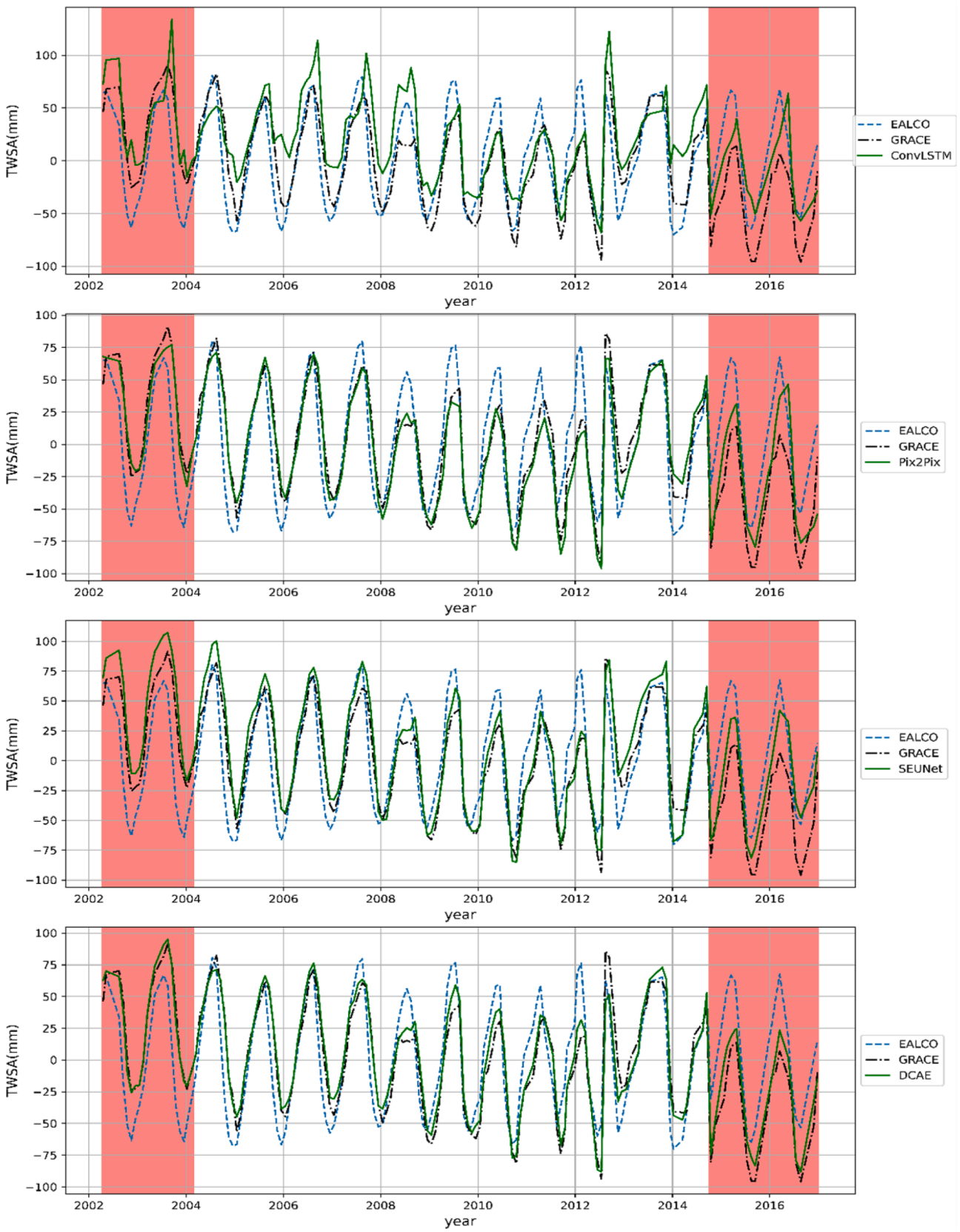


Fig. 12. Comparison of model-reconstructed GRACE-like TWSA, EALCO-simulated TWSA, and GRACE-derived TWSA during training (white background) and testing periods (red background) at nationwide level. From top to bottom: ConvLSTM, Pix2Pix, SEUNet, DCAE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

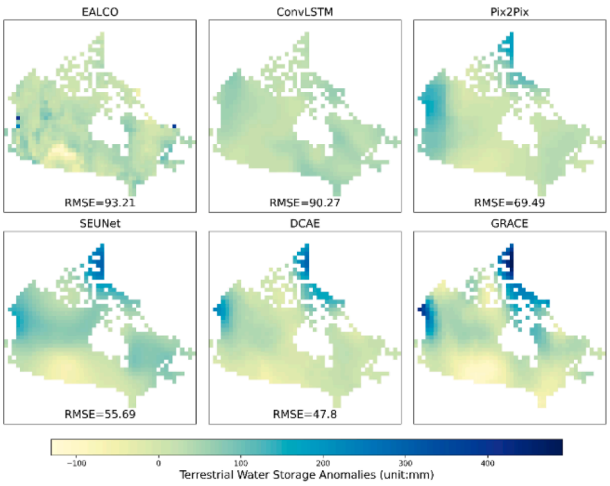


Fig. 13. Reconstructed TWSA maps for January 2003.

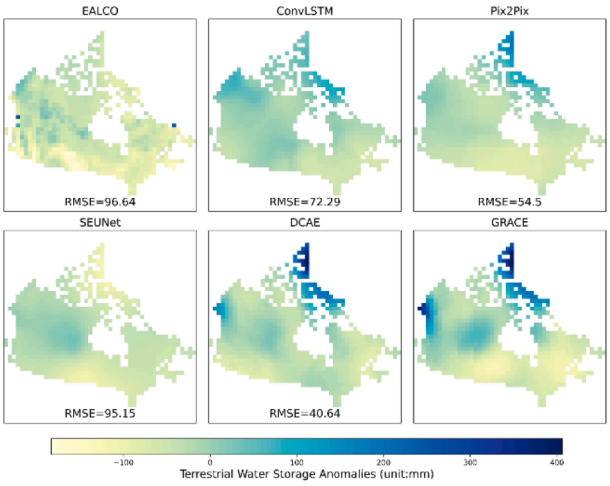


Fig. 14. Reconstructed TWSA maps for July 2003.

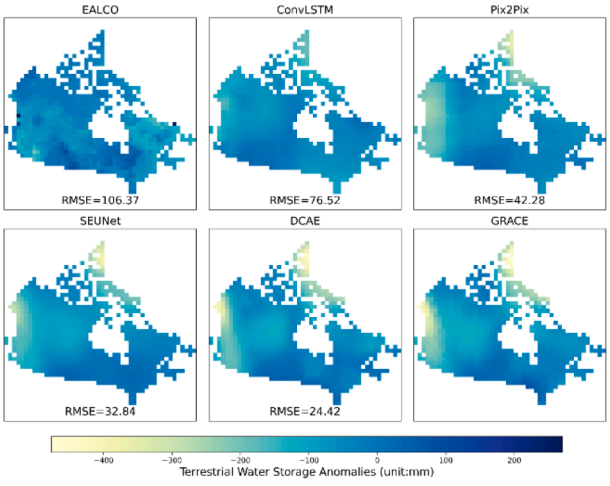


Fig. 15. Reconstructed TWSA maps for October 2014.

4.2. Reconstruction for pre-GRACE years from 1979 to 2002

Finally, the trained DCAE model was applied to reconstruct (i.e., hindcasting) the TWSA time series from January 1979 to March 2002 for the Canadian landmass. The results are demonstrated in Fig. 17, and an example pair for January 1979 is shown in Fig. 18. It can be seen that EALCO tends to overestimate dry conditions during 2002 to 2009, thus its simulations for pre-GRACE years are more likely to have overestimated values for dry conditions, and the reconstruction moderately adjusts the original EALCO simulations over dry conditions. As for post-GRACE forecasting, even though the DCAE model has been evaluated against GRACE-derived TWSA from 2014 to 2016, there is an obvious downward trend in the GRACE TWS since 2010, the model's capability for long-term forecasting remains uncertain.

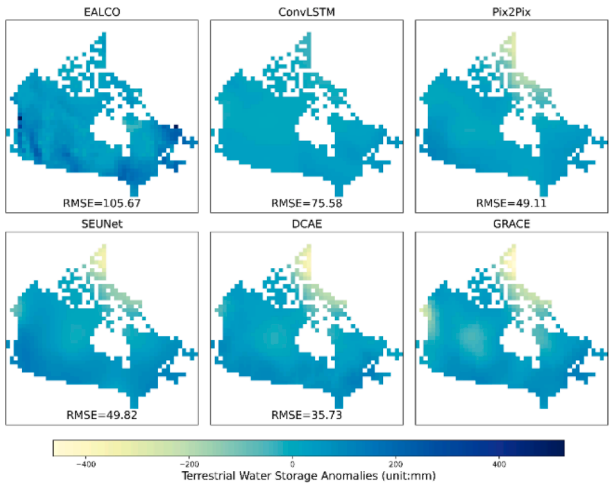


Fig. 16. Reconstructed TWSA maps for April 2015.

4.3. Limitations

It is worth noting some drawbacks in the workflow of this study, which need to be further investigated in future research. First of all, no forcing dataset (e.g., in-situ precipitation measurements) was used, which hinders this research from assessing the effects of extreme events (e.g., droughts, floods, rainstorm) on the reconstructed TWSA. Secondly, the pairing of physical simulation and GRACE observation ignores the temporal correlations (or long-term trends) existed in each TWSA time series. For future studies, it is worth to examine the applicability of pairing physical simulations sequences with corresponding GRACE observation sequences. Thirdly, only one GRACE solution product was used in this study, resulting in the overlook of uncertainties in GRACE-derived TWSA. Moreover, the TWSA reconstruction was conducted for the entire Canadian landmass, but the relationships between hydrological variables of LSM and GRACE TWS can be varying in different river basins. Last but not least, the study does not assess the interannual variability in the predicted and observed TWSA. For future studies, the input data can be de-seasonalized (i.e., detrending) to remove the seasonal trends before or after the model training.

5. Conclusions

This study applied a hybrid approach to predict GRACE-like TWSA over the Canadian landmass. Physically-based modeling and deep learning techniques were combined to train image-to-image transition models. The DL models take a pair of EALCO-GRACE samples and learn the statistical relationships between LSM-simulated TWSA and GRACE-observed TWSA in order to predict GRACE-like TWSA based on the LSM simulations. The hybrid approach was compared to a time-series prediction approach which only utilizes GRACE observations. The time-series prediction model is based on Convolutional LSTM (ConvLSTM) networks. It takes a number of temporally consecutive samples as an input sequence, and then learns the spatiotemporal trajectories exist in

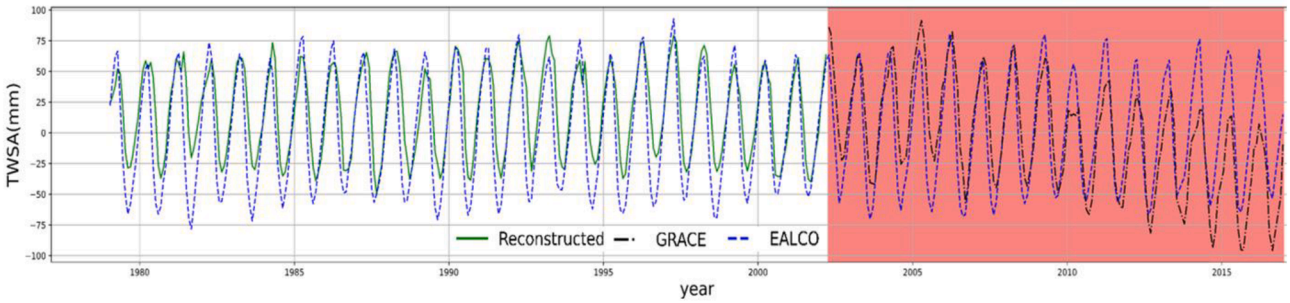


Fig. 17. DCAE-reconstructed TWSA from 1979 to 2002 (green line in white background), as comparing to GRACE TWSA from 2002 to 2016 (black dash dot line in red background). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

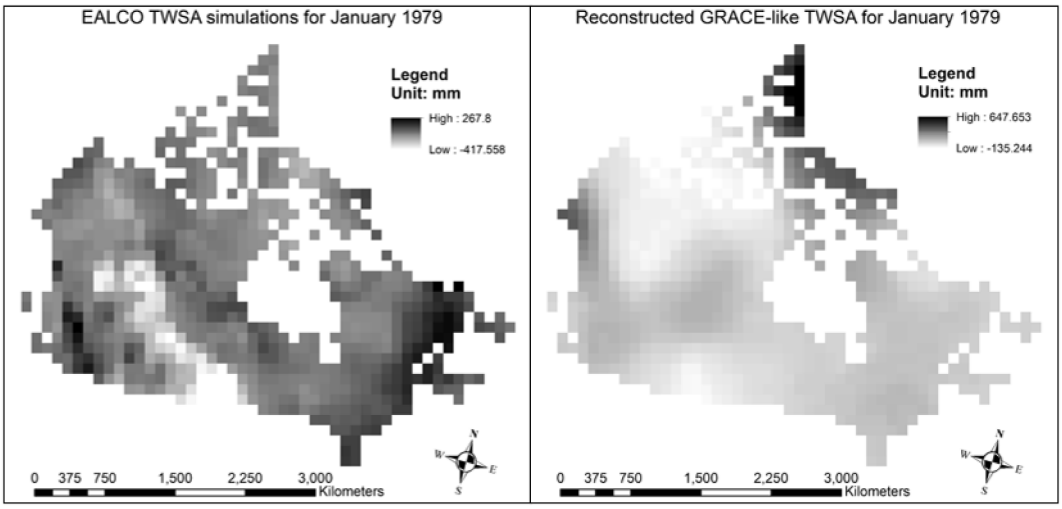


Fig. 18. DCAE-predicted TWSA for January 1979 (right), comparing to its corresponding EALCO-simulated TWSA (left).

the sequence. The trained ConvLSTM model is able to hindcast/forecast the TWSA during the time period when the GRACE observation is unavailable. The performances of these DL models were assessed by correlation coefficients and RMSE on both pixelwise level and nationwide level. The results show that the three LSM-based DL models (Pix2Pix, SEUNet, DCAE) and the ConvLSTM model are all capable to improve the original EALCO TWSA, while the LSM-based DL models significantly outperform the ConvLSTM model. By comparison, the DCAE model exhibits the optimal solution to calibrate EALCO simulations to better fit the GRACE observations, by reducing the nationwide mean RMSE from 105 mm to 53 mm. As for the spatial pattern of predicted TWSA, the LSM-based DL models perform reasonably in most of the study area, while exhibit noticeable uncertainties in dry, cold, and intensively irrigated areas.

This study indicates that deep learning techniques is a promising alternative to conventional data assimilation methods in future hydrological research. The major contribution of this study is that the feasibility of various DL network types in TWS reconstruction was investigated and examined, which provides a new train of thought for in-depth study on the application of deep learning in hydrology and other geoscientific disciplines. Future research will focus on adding climate forcing and temporal correlations to the hybrid modelling approach adopted in this study.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by the Natural Resource Canada, Contract No. 3000711130, in part by the Emerging Interdisciplinary Construction Project in Central University of Finance and Economics, and also in part by the Program for Innovation Research in Central University of Finance and Economics. We thank all the people from the Canada Center for Mapping and Earth Observations for their work in model simulation and data collection.

References

Al-Najjar, H.A.H., Kalantar, B., Pradhan, B., Saeidi, V., Halin, A.A., Ueda, N., Mansor, S., 2019. Land Cover Classification from fused DSM and UAV Images Using Convolutional Neural Networks. Remote Sens. 11 (12), 1461. https://doi.org/10.3390/rs11121461.

Arguez, A., Inamdar, A., Palecki, M.A., Schreck, C.J., Young, A.H., 2019. ENSO Normals: A New US Climate Normals Product Conditioned by ENSO Phase and Intensity and Accounting for Secular Trends. J. Appl. Meteorol. Climatol. 58 (6), 1381–1397. https://doi.org/10.1175/JAMC-D-18-0252.1.

Azarang, A., Manoochehri, H.E., Kehtarnavaz, N., 2019. Convolutional Autoencoder-Based Multispectral Image Fusion. IEEE Access 7, 35673–35683. https://doi.org/10.1109/ACCESS.2019.2905511.

Becker, M., Papa, F., Frappart, F., Alsdorf, D., Calmant, S., Da Silva, J.S., Prigent, C., Seyler, F., 2018. Satellite-based estimates of surface water dynamics in the Congo River Basin. Int. J. Appl. Earth Obs. Geoinf. 66, 196–209.

Broxton, P.D., van Leeuwen, W.J.D., Biederman, J.A., 2019. Improving Snow Water Equivalent Maps With Machine Learning of Snow Survey and Lidar Measurements. Water Resour. Res. 55 (5), 3739–3757. https://doi.org/10.1029/2018WR024146.

Cao, C., Dragicevic, S., Li, S., 2019. Land-Use Change Detection with Convolutional Neural Network Methods. Environments 6 (2), 25. https://doi.org/10.3390/environments6020025.

Chen, H., Zhang, W., Nie, N., Guo, Y., 2019. Long-term groundwater storage variations estimated in the Songhua River Basin by using GRACE products, land surface

models, and in-situ observations. Sci. Total Environ. 649, 372–387. https://doi.org/10.1016/j.scitotenv.2018.08.352.
Dankwa, S., Zheng, W., Gao, B., Li, X., 2018. Terrestrial Water Storage (TWS) Patterns Monitoring in the Amazon Basin using GRACE Observed: its Trends and Characteristics. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 768–771.
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. Geophys. Res. Lett. 38 (3) https://doi.org/10.1029/2010GL046442.
Hamshaw, S.D., Dewoolkar, M.M., Schroth, A.W., Wemple, B.C., Rizzo, D.M., 2018. A New Machine-Learning Approach for Classifying Hysteresis in Suspended-Sediment Discharge Relationships Using High-Frequency Monitoring Data. Water Resour. Res. 54 (6), 4040–4058. https://doi.org/10.1029/2017WR022238.
Heimhuber, V., Tulbure, M.G., Broich, M., Xie, Z., Hurriyet, M., 2019. The role of GRACE total water storage anomalies, streamflow and rainfall in stream salinity trends across Australia's Murray-Darling Basin during and post the Millennium Drought. Int. J. Appl. Earth Obs. Geoinf. 83, 101927.
Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372.
Humphrey, V., Gudmundsson, L., 2019. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. Earth Syst. Sci. Data 11 (3), 1153–1170. https://doi.org/10.5194/essd-11-1153-2019.
Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2017. Image-to-Image Translation with Conditional Adversarial Networks. In: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), pp. 5967–5976. doi: 10.1109/CVPR.2017.632.
Jing, W., Di, L., Zhao, X., Yao, L., Xia, X., Liu, Y., Zhou, C., 2020. A data-driven approach to generate past GRACE-like terrestrial water storage solution by calibrating the land surface model simulations. Adv. Water Resour. 143, 103683 https://doi.org/10.1016/j.advwatres.2020.103683.
Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A.I.J.M., Pattiaratchi, C., 2017. Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. Adv. Water Resour. 107, 301–316. https://doi.org/10.1016/j.advwatres.2017.07.001.
Kim, D., Yu, H., Lee, H., Beighley, E., Durand, M., Alsdorf, D.E., Hwang, E., 2019. Ensemble learning regression for estimating river discharges using satellite altimetry data: Central Congo River as a Test-bed. Remote Sens. Environ. 221, 741–755. https://doi.org/10.1016/j.rse.2018.12.010.
Kumar, S.V., Wang, S., Mocko, D.M., Peters-Lidard, C.D., Xia, Y., 2017. Similarity Assessment of Land Surface Model Outputs in the North American Land Data Assimilation System. Water Resour. Res. 53 (11), 8941–8965. https://doi.org/10.1002/2017WR020635.
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 Sens. Environ. 155, 145–160. https://doi.org/10.1016/j.rse.2014.08.006.
Miyamoto, H., Uehara, K., Murakawa, M., Sakanashi, H., Nosato, H., Kouyama, T., Nakamura, R., 2018. Object Detection in Satellite Imagery using 2-Step Convolutional Neural Networks. In: IGARSS 2018 – 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 1268–1271.
Mukherjee, A., Ramachandran, P., 2018. Prediction of GWL with the help of GRACE TWS for unevenly spaced time series data in India: Analysis of comparative performances of SVR, ANN and LRM. J. Hydrol. 558, 647–658. https://doi.org/10.1016/j.jhydrol.2018.02.005.
Nie, W., Zaitchik, B.F., Rodell, M., Kumar, S.V., Arsenault, K.R., Li, B., Getirana, A., 2019. Assimilating GRACE Into a Land Surface Model in the Presence of an Irrigation-Induced Groundwater Trend. Water Resour. Res. 55 (12), 11274–11294. https://doi.org/10.1029/2019WR025363.
Quilty, J., Adamowski, J., Boucher, M., 2019. A Stochastic Data-Driven Ensemble Forecasting Framework for Water Resources: A Case Study Using Ensemble Members Derived From a Database of Deterministic Wavelet-Based Models. Water Resour. Res. 55 (1), 175–202. https://doi.org/10.1029/2018WR023205.
Sahoo, S., Russo, T.A., Elliott, J., Foster, I., 2017. Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour. Res. 53 (5), 3878–3895. https://doi.org/10.1002/2016WR019933.
Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Schmied, H.M., 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. PNAS 115 (6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115.
Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015. Missing Information Reconstruction of Remote Sensing Data: A Technical Review. IEEE Geosci. Remote Sens. Mag. 3 (3), 61–85. https://doi.org/10.1109/MGRS.2015.2441912.
Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D., Wong, W., Woo, W., 2017. Deep learning for precipitation nowcasting: A benchmark and a new model. Adv. Neural Inf. Process. Syst. 30(NIPS 2017), 30.
Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., Woo, W., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. Adv. Neural Inf. Process. Syst. 28(NIPS 2015), 28.
Shi, X., Yeung, D., 2018. Machine Learning for Spatiotemporal Sequence Forecasting: A Survey. arXiv preprint arXiv:1808.06865.
Shokri, A., Walker, J.P., van Dijk, A.I.J.M., Pauwels, V.R.N., 2018. Performance of Different Ensemble Kalman Filter Structures to Assimilate GRACE Terrestrial Water Storage Estimates Into a High-Resolution Hydrological Model: A Synthetic Study. Water Resour. Res. 54 (11), 8931–8951. https://doi.org/10.1029/2018WR022785.
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 Resour. Res. 55 (2), 1179–1195. https://doi.org/10.1029/2018WR023333.
Tourian, M.J., Reager, J.T., Sneeuw, N., 2018. The Total Drainable Water Storage of the Amazon River Basin: A First Estimate Using GRACE. Water Resour. Res. 54 (5), 3290–3312. https://doi.org/10.1029/2017WR021674.
Wang, S., Yang, Y., Rivera, A., 2013. Spatial and seasonal variations in actual evapotranspiration over Canada's landmass. Hydrol. Earth Syst. Sci. 17, 3561–3575. https://doi.org/10.5194/hess-17-3561-2013.
Wang, S., Li, J., 2016. Terrestrial Water Storage Climatology for Canada from GRACE Satellite Observations in 2002–2014. Can. J. Remote Sens. 42 (3), 190–202. https://doi.org/10.1080/07038992.2016.1171132.
Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D.W., Sheffield, J., 2014a. Assessment of water budget for sixteen large drainage basins in Canada. J. Hydrol. 512, 1–15. https://doi.org/10.1016/j.jhydrol.2014.02.058.
Wang, S., McKenney, D.W., Shang, J., Li, J., 2014b. A national-scale assessment of long-term water budget closures for Canada's watersheds. J. Geophys. Res. 119 (14), 8712–8725. https://doi.org/10.1002/2014JD021951.
Wang, S., Huang, J., Yang, D., Pavlic, G., Li, J., 2015. Long-term water budget imbalances and error sources for cold region drainage basins. Hydrol. Process. 29, 2125–2136. https://doi.org/10.1002/hyp.10343.
Yang, P., Xia, J., Zhan, C., Qiao, Y., Wang, Y., 2017. Monitoring the spatio-temporal changes of terrestrial water storage using GRACE data in the Tarim River basin between 2002 and 2015. Sci. Total Environ. 595, 218–228. https://doi.org/10.1016/j.scitotenv.2017.03.268.
Zhang, D., Zhang, Q., Werner, A.D., Liu, X., 2016. GRACE-Based Hydrological Drought Evaluation of the Yangtze River Basin, China. J. Hydrometeorol. 17 (3), 811–828. https://doi.org/10.1175/JHM-D-15-0084.1.
Zhu, Q., Chen, J., Zhu, L., Duan, X., Liu, Y., 2018. Wind Speed Prediction with Spatio-Temporal Correlation: A Deep Learning Approach. Energies 11 (4), 705. https://doi.org/10.3390/en11040705.