Integration of Satellite Data, Physically-based Model, and Deep Neural Networks for Historical Terrestrial Water Storage Reconstruction

by

Qiutong Yu

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Geography

Waterloo, Ontario, Canada, 2021

Author’s Declaration

This thesis consists of material all of which I authored or co-authored, a Statement of Contributions is included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists of four chapters all of which I have been the lead author. While my colleagues Hongjie He and Dr. Ke Yang assisted with the data preprocessing, Dr. Shusen Wang from Nature Resources Canada proposed the initial concept of this study, and my supervisor, Professor Dr. Jonathan Li, provided comments and edits on Chapters 1, 3 and 4, they have taken a more collaborative role as coauthor on Chapter 3. As the lead author of Chapter 3, I conceptualized the experiment design, conducted the coding for model constructions, carried out data analysis, and created figures and tables.

Abstract

Terrestrial water storage (TWS) is an essential part of the global water cycle. Long-term monitoring of observed and modeled TWS is fundamental to analyze droughts, floods, and other meteorological extreme events caused by the effects of climate change on the hydrological cycle. Over the past several decades, hydrologists have been applying physically-based global hydrological model (GHM) and land surface model (LSM) to simulate TWS and the water components (e.g., groundwater storage) composing TWS. 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 has been devoted to the reconstruction of GRACE-like TWS data during the pre-GRACE era. Data-driven methods, such as multilinear regression and machine learning, exhibit a great potential to improve TWS assessments by integrating GRACE observations and physically-based simulations. The advances in artificial intelligence enable adaptive learning of correlations between variables in complex spatiotemporal systems. As for GRACE reconstruction, the applicability of various deep learning techniques has not been well studied previously. Thus, in this study, three deep learning-based models are developed based on the LSM-simulated TWS, to reconstruct the historical TWS in the Canadian landmass from 1979 to 2002. The performance of the models is evaluated against the GRACE-observed TWS anomalies from 2002 to 2004, and 2014 to 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 match between original LSM simulations and GRACE observations.

Acknowledgements

First of all, I would like to extend my sincere gratitude to my supervisor, Professor Dr. Jonathan Li, for giving me the opportunity to do research and providing invaluable guidance throughout my graduate study. I am deeply grateful of his instructive advice and useful suggestions for the completion of this thesis.

I would also like to sincerely thank my thesis committee numbers, Dr. Michael Chapman, Professor at the Department of Civil Engineering, Ryerson University, and Dr. Bryan Tolson, Professor at the Department of Civil and Environmental Engineering, University of Waterloo, for serving as my thesis examining committee, and for providing insightful suggestions.

My special thanks go to Dr. Shusen Wang, Research Scientist at the Canada Centre for Mapping and Earth Observations, Natural Resources Canada for many constructive discussion and comments. Thanks for providing funding and data to support my thesis work.

I also would like to thank all the members in Geospatial Sensing and Data Intelligence Lab at the University of Waterloo, who shared their valuable knowledge and insight with me during group meetings. Special thanks to Hongjie He and Dr. Ke Yang, for providing advices and helps to finalize this thesis. Mr. Alan Anthony and the staff at Centre for Mapping, Analysis & Design (MAD), Faculty of Environment, University of Waterloo, are acknowledged for their necessary technical supports.

Furthermore, I would like to thank all of my friends in the University of Waterloo, for their friendships and support, especially in such a tough time of COVID-19 pandemic.

Last but most important, I am extremely grateful to my parents and other family members for their love, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my fiancée, Yingfang, for her love, understanding, prayers and continuing support to complete my research works.

March 10, 2021

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AE	Autoencoders
AI	Artificial Intelligence
ANN	Artificial Neural Network
CanLCC	Canada Lambert Conformal Conic projection
CC	Pearson's correlation coefficient
CCMEO	The Canada Centre for Mapping and Earth Observation
CE	Cumulative Effects
cGAN	conditional Generative Adversarial Network
CLSM	Catchment Land Surface Model
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-term Memory
DCAE	Deep Convolutional Autoencoder
DL	Deep Learning
DNN	Deep Neural Network
EALCO	Ecological Assimilation of Land and Climate Observation
EO	Earth observation
EO4CE	Earth Observation Baseline Data for Cumulative Effects
EWH	Equivalent Water Height
FC-LSTM	Fully-connected long short-term memory
GAN	Generative Adversarial Network
GHM	Global Hydrological Model
GLDAS	Global Land Data Assimilation System

GRACE	Gravity Recovery and Climate Experiment
GWS	Groundwater storage
LSM	Land Surface Model
LSTM	Long Short-term Memory
ML	Machine Learning
MSE	Mean Square Error
NRCan	Natural Resources Canada
PDF	Probability Distribution Function
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SE	Squeeze-and-Excitation
SHC	Spherical Harmonic Coefficient
SWE	Snow Water Equivalent
tanh	hyperbolic tangent function
TWS	Terrestrial Water Storage
TWSA	Terrestrial Water Storage Anomalies

Chapter 1

Introduction

1.1. Motivation

Environmental instability may prove to be the greatest challenge that human being will face over the next several decades. This instability is primarily resulting from a combination of abrupt change drivers including human activity, forest fires, and floods, as well as gradual change drivers such as climate change and pollution (Janzen et al., 2020). Effective environmental management requires a demonstrated understanding of how these change drivers are impacting the status and trends, as well as how the landmass is changing. To ensure that any region in Canada is able to effectively manage the problem of environmental instability, the Canada Centre for Mapping and Earth Observation (CCMEO) of Natural Resources Canada (NRCan) has implemented a five-year project 'Earth Observation Baseline Data for Cumulative Effects' (EO4CE) which directly contributes to the Government of Canada's Cumulative Effects (CE) assessments of environmental change and human activities. The EO4CE project aims to develop Earth Observation (EO) -based baseline datasets of a wide range of status and trends variables and improve capacity to conduct regional impact assessments through the development of baseline data on a national scale (Janzen et al., 2020). The production of these datasets will be operationalized to support current CE assessments, future assessments, and ongoing monitoring.

Terrestrial Water Storage (TWS) includes all the water components in terrestrial ecosystem, which represents water resources availability. Its dynamics under the different scenarios of environmental change and human disturbance is key to determine water resources sustainability and vulnerability. As such, TWS information is critical in CE assessments and it is one of the key parameters to be studied in the EO4CE project. The activities in producing high quality and long-term datasets for TWS involve the integration of various datasets from remote sensing, in situ observations, and simulated outputs from land surface and hydrological models.

The emergence of satellite remote sensing enabled continuous monitoring over hydrological fluxes at different spatial resolutions. NASA launched the Gravity Recovery and Climate Experiment (GRACE) mission in March 2002, offering monthly TWS anomalies (TWSA) (i.e., deviations from a long-term mean) measurements at global scale. The operating mechanism

of GRACE data acquisition is that the satellites detect the change in gravity field over a region. The primary reason for changes in 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). GRACE observations have been widely adopted to understand the temporal trends in TWS variations at regional and global scales (Andersen et al., 2005; Rodell et al., 2009; Famiglietti et al., 2011; Frappart et al., 2013; Voss et al., 2013; Shamsudduha et al., 2017; Lezzaik et al., 2018;). However, the observations of TWS from the GRACE satellite has data available for only 15 years (2002-2017), which does not meet the requirement for producing the baseline TWS information that can be used to calculate the Climate Normal which requires at least 30 years (Arguez et al., 2019). Therefore, the extension of the GRACE-observed TWS dataset for Canada's Landmass is an important process for the delivery of EO4CE project.

Hydrological fluxes simulated by physically-based Land Surface Models (LSM) have become a widely used data resources for analyzing environmental changes and water resources management (Jing et al., 2020). The Ecological Assimilation of Land and Climate Observation (EALCO) model developed by NRCan is a LSM which simulates the energy, water, and carbon dynamics by utilizing land and meteorological observation information, which can provide hydrologic information with a relatively high spatial and temporal resolution (Wang et al., 2014b). In addition, EALCO is able to produce TWS simulations for long-term time span, as it is forced by long-term climatic data. However, there are notable discrepancies found in TWS trends between GRACE satellite observations and EALCO model simulations. Because the development of LSM is based on mathematical representations of physical processes that dominate the flow and storage of water in space and time (Sun et al., 2019). And the construction of LSM can be restricted to a sparse set of in situ measurements. Consequently, deficient spatiotemporal coverage of in situ observations and parameter uncertainties made the model's understanding of long-term trends in water storage limited. Despite the data availability over longer period, the TWS estimates from EALCO cannot be directly adopted as the historical TWS dataset for the EO4CE project.

The GRACE TWS data have been widely used to adjust, calibrate, and assimilate the TWS simulations from LSMs (Lo et al., 2010; Houborg et al., 2012; van Dijk et al., 2014; Schumacher et al., 2016; Khaki et al., 2017), for tuning model parameters as well as improving the model's

predictive performance. Previous studies have indicated that there is a linear or polynomial correlation between GRACE-observed TWSA and LSM-simulated TWSA in the region of Amazon basin (Nie et al., 2015; Humphrey et al., 2017). This suggests that these calibrated physical models are applicable in extending the time span of GRACE TWS (2002 - 2017) to that of these models. Additionally, adjusted LSMs can be used to bridge the data gap for the 1-year missing data between GRACE and its successor, GRACE-FO (launched in 2018), by substitution with LSM simulations to fill gaps in GRACE-derived TWS.

Nevertheless, previous works regarding LSM-GRACE fusion are only proven to be applicable in certain areas, which may not apply in other basins, especially those with dry, cold climatic conditions and intense human interventions (Jing et al., 2020). In different environmental scenarios, the linear correlation between LSM simulations and GRACE observations can be either poor or strong. Moreover, the GRACE-LSM assimilation could be confined by intrinsic uncertainties in parameterization and architecture of LSMs, as well as the extrinsic uncertainties from the assimilation methodology adopted (Scanlon et al., 2018).

Recent studies have been focusing on using machine learning (ML) -based methods to link LSM-simulated TWS with GRACE TWS for hindcasting and forecasting TWS over the periods when GRACE TWS observation is not available (Sun et al., 2019; Sun et al., 2020; Jing et al., 2020). Machine learning-based methods are data-driven (i.e., black box models). Pure data-driven models are only suitable for studies with sufficient number of observations, but without comprehensive understanding of the underlying physical processes (Sun et al., 2019). For instance, a regression model between in-situ runoff observations and GRACE observations. However, in reality the variation of TWS is resulting from complex and sophisticated hydrologic processes, and the physical principles of these processes cannot be overlooked as reconstructing GRACE TWS. Therefore, it is desirable to adopt a hybrid approach by incorporating the information represented in physically-based LSM, with the advances in data-driven machine learning models. Moreover, only a handful of deep learning algorithms were applied to reconstructing GRACE TWS, it is necessary to examine the capabilities of more algorithms and architectures due to the rapid evolvement of deep learning techniques.

1.2. Objective of the study

This thesis aims to develop deep learning (DL)-based models and construct the historical TWS dataset for Canada's landmass. The main concept is to calibrate EALCO-simulated TWSA by learning the spatiotemporal patterns of matchings between EALCO simulations and GRACE satellite observations using deep learning techniques. Once trained and validated, these deep learning-based models are able to predict GRACE-like TWS with using LSM TWS simulation as inputs (i.e., without requiring observed GRACE TWS as inputs), so that the existing TWS records can be extended for generating the baseline historical TWS datasets in Canada. The hybrid approach is examined on three DL models built on different network architectures. The predictive performance of each model is assessed and compared to a pure data-driven method (i.e., only using GRACE TWS data), as well as the LSM-simulated TWSA, thereby proposing an optimal predictor for TWS data reconstruction. This research directly contributes to the delivery of the CCMEO cumulative effects project. Also, it is expected to improve the water modelling research in CCMEO and collaborates with other researchers in Canada.

1.3. Structure of the thesis

This thesis is structured in accordance with the University of Waterloo manuscript-option format which a manuscript article is presented in a standalone chapter and prepared to be submitted for publication. Chapter 1 introduces the motivation and research objectives of this study. Chapter 2 reviews previous studies regarding monitoring/quantifying of TWS for hydrological research, assimilation of GRACE TWS data into physically-based model simulations, reconstruction of GRACE TWS data, and applications of ML and DL techniques in hydrology and earth system science. Chapter 3 presents the manuscript entitled "Reconstructing GRACE-like TWS Anomalies in the Canadian Landmass Using Deep Learning and Land Surface Model", which details the methodology, results, and findings of this study. Finally, Chapter 4 summarizes the contributions and limitations of this study, as well as recommendations for future work.

Chapter 2

Related Studies

This chapter is divided into three sections designed to review previous studies relevant to this research. Section 2.1 presents the methodologies for monitoring and quantifying TWS, as well as the significance of TWS in hydrometeorology. Section 2.2 reviews studies regarding data fusion of GRACE-derived TWS, climate datasets, and physically-based LSM or global hydrological models (GHM), aimed at hindcasting historical TWS data, forecasting future TWS variations, downscaling the GRACE data to finer spatiotemporal resolutions, or filling the observation gaps between GRACE and GRACE-FO missions. Lastly, Section 2.3 presents some applications of ML and DL in general geoscientific studies, with a focus on hydrology.

2.1. Assessment of terrestrial water storage

Terrestrial water storage (TWS) can be defined as the sum of water stored above and below Earth's surface (Rodell et al., 2018), which is given by the following equation:

$$TWS = SWS + SMS + SnWS + GWS + CWS \tag{2.1}$$

where SWS stands for surface water, SMS is surface and root soil moisture, SnWS is water stored in snow and ice, GWS is groundwater, and CWS is water stored in vegetations (also known as canopy water).

Variations in TWS are often caused by exchanges with oceans and the atmosphere through the mass fluxes in hydrological processes (e.g., evaporation and groundwater flow) (Ni et al., 2017), as well as human activities (e.g., groundwater withdrawal for irrigation) (Shamsudduha & Panda, 2019). It is important to monitor and quantify TWS variations as it reflects water availability, hydrological extremes, and human impacts on the water cycle (Famiglietti, 2004), thus helping further research in environmental studies, as well as efficient water resource management and policies (Fu et al., 2015). For instance, Ni et al. (2017) analyzed the intercorrelations of TWS changes and El Nino events at global scale, based on a combination of multiple data sources such as LSM simulations, GRACE observations, and precipitation record.

Typical methods to estimate the TWS include in situ observations, land surface models, and remote sensing observations. In situ surface measurements of TWS are essentially limited over

large areas because normally the process is costly and labor-intensive. As a result, the monitoring and analysis of TWS variations have been hindered for large-scale areas (Yang et al., 2013). Although some components of TWS, such as soil moisture, snow water equivalent (SWE), groundwater, can be measured in situ, those measurements are generally local, which lacks an observational basis for estimating these components at large scales (Serreze et al. 1999; Robock et al. 2000; Alley et al. 2002; Tang et al., 2010). As a result, the need for better observations of TWS and its components is increasingly recognized over the recent years.

2.1.1. Physically-based modelling

Physically-based hydrological models have been widely applied to investigate spatial distribution and trends of TWS at global scale, and to have a better understanding of climate dynamics. Model estimates is commonly produced by the simulation of land surface fluxes such as precipitation, evapotranspiration, air temperature, solar longwave radiation (Tang et al., 2010; Ma et al., 2017). Global land surface models (LSMs) and global hydrological water resource models (GHWRMs) are two most commonly used types of global hydrological models (GHM) (Bierkens, 2015).

By definition, LSMs are models that simulate mass fluxes between the land surface and the atmosphere (e.g., energy flux, water flux, and carbon flux), while GHWRMs focus on the earth water cycle and its subdomains in order to deal with global water availability concerns (Zhang et al., 2017). Hence, one of the major differences between LSMs and GHWRMs is that LSMs are more physically-based as integrating carbon, energy, and water balances. On the other hands, LSMs may not accurately simulate variations in TWS because of their emphasis on fluxes (Milly et al., 2010). Additionally, GHWRMs take human water consumption into consideration, whereas most LSMs do not (Scanlon et al., 2018). Nevertheless, LSMs have been widely applied to analyzing the spatiotemporal variations in TWS as well as modelling hydrological variables that composed the TWS.

Rapid development of LSM has resulted in improvement of existing physical process representation, with addition of new processes and functionalities. Several studies have conducted comparison of LSMs performance on simulating water cycles. The Water Model Intercomparison Project evaluated five GHMs and six LSMs developed between 1994 and 2010 (Haddeland et al., 2011). The project reveals large differences in simulated evapotranspiration and runoff. The

variations are ranging from 42,000 to 85,000 km^3/year. Schewe et al. (2014) suggest that the results of these intercomparison projects indicate that there are considerable uncertainties in included GHMs and LSMs, which also underscore the necessity to adopt multiple models for studying fluxes and water storage. Moreover, these projects have primarily focus on particular schemes such as evapotranspiration, runoff, or soil moisture content, which does not meet the requirement for a comprehensive TWS assessment (Zhang et al., 2017).

The scheme of LSM structures have developed from simplex to complex over the years. This 'evolution' is based on the need to integrate more feedback mechanisms between various earth subsystems, as well as more sophisticated characterization of physical processes (Cai et al., 2014). Therefore, for these evolved LSMs which have more complex structure, more complicated evaluation standard is needed, which required miscellaneous observational datasets to assess their performances (Luo et al., 2012). However, conventional tests have been mainly carried out at smaller scales in order to have better measurements for in situ observations. For large scale LSM simulation, it is necessary to assess the model performance under a variety of natural conditions or different environmental scenarios (Ma et al., 2017). For instance, Cai et al. (2014) presented a comparative study to examine the performances of some widely-used global LSMs (VICM, CLM, Noah, Noah-MP) in water balance simulations over the contiguous United States. The results show that Noah-MP outperforms other models regarding simulating soil moisture and TWS. However, the comparison does not give a comprehensive evaluation of the model performance in the entire study region. Because the climates, ecosystems, and human activity level of each river basin within a large-scale region could be noticeably different. Therefore, the transferability and flexibility of a hydrological model must be elucidated by assessing basins consisting of significantly varying, complicated natural conditions (Gupta et al., 2014). For example, in a study evaluating the capability of Noah-MP, Ma et al. (2017) selected 18 test regions across the entire contiguous United States. Their results show that Noah-MP performs poorly on simulating the monthly TWSA in regions that severely affected by either agricultural production or extreme climatic events, which caused sudden change in the local water storage. The authors further suggest that Noah-MP needs to have additional modules focusing on human-induced disturbances and reservoir water storage change, in order to improve its performance over certain river basins.

It is worth mentioning that, for a given study area with identical meteorological forcing data, there still are large differences between the TWS simulations from different LSMs. The uncertainties in the physical process representation of each model make it difficult to determine a general-purpose LSM for any regions. Therefore, the selection of model for simulating TWS should consider the hydrometeorological regimes of the study area.

Table 2-1. Comparison of some commonly used LSMs in simulating TWS

Model Name	CLM4.0	Noah-MP	CLSM-F2.5	EALCO (v4.2)
Paper	Lawrence et al. (2011)	Niu et al. (2011)	Koster et al. (2000)	Wang et al. (2014a)
Spatial resolution in North America	0.125° in NLDAS grid	0.125° in NLDAS grid	0.125° in NLDAS grid	5km
Temporal resolution	1 hour	1 hour	1 hour	0.5 hour
Number of snow layers	5	3	3	Dynamic snow layering
Dynamical vegetation	Yes	Yes	No	Yes
Human-induced disturbance	No	No	No	No

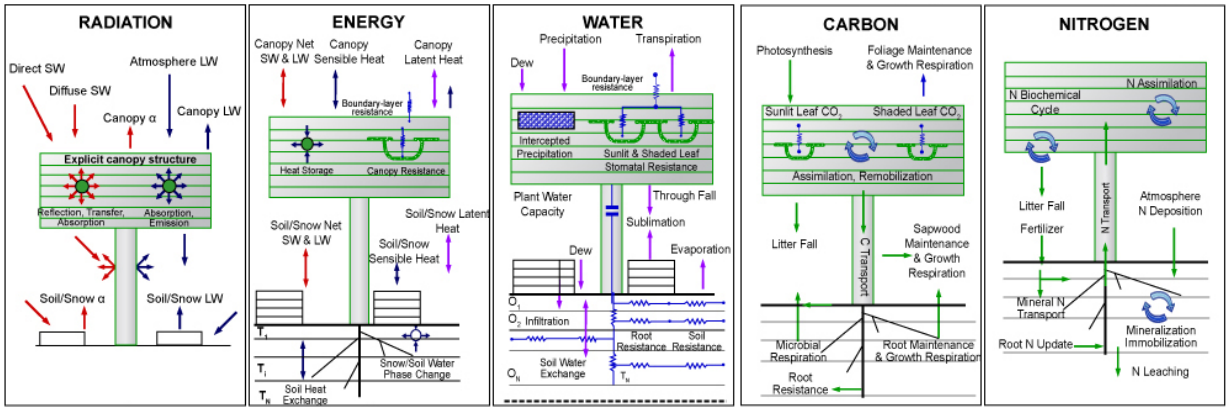

Figure 2-1. Schematic diagram of EALCO model. Source: Janzen et al. (2020)

EALCO is an EO-based LSM developed by NRCan for simulating the physical processes existing in the terrestrial water cycles (Wang, 2005). The model incorporates five major modules (as shown in Figure 2-1) designed for modelling land surface radiation transfer, energy balance, water dynamics, carbon cycle, and nitrogen biogeochemical cycles, respectively (Wang, 2008).

The module of water cycle is dynamically correlated with the other four modules in order to coordinate the feedback mechanism of atmosphere and ecosystem (e.g., vegetation) in the water cycle. EALCO estimates the TWS variations by simulating a variety of physical processes with equivalent weight in terrestrial water cycle, which are listed as follow:

- Canopy evapotranspiration by leaf transpiration and canopy evaporation
- Plant root water uptake and plant water storage.
- Soil evaporation and snow sublimation at ground surface.
- Soil water transfers and soil water-groundwater interactions.
- Snow processes which use dynamic snow layering schemes for snow accumulation and thaw simulations.
- Sublimation from intercepted rain, snow, dew, or frost.
- Open water surface evaporation from lakes, reservoirs, and rivers.

The EALCO LSM incorporates various mechanisms specifically designed for simulating water and energy fluxes in cold region, so that it can be better applied in Canadian regions (Zhong et al., 2020). Its performance has been evaluated in several LSM/GHM intercomparison studies (Widlowski et al., 2011; Medlyn et al., 2015). Moreover, EALCO is able to provide the terrestrial TWS dynamics at relatively high spatiotemporal resolutions (Janzen et al., 2020). To sum up, the robustness of EALCO is reliable as shown in previous works, thus EALCO is expected to be effectively applicable in this thesis research.

2.1.2. Satellite observations

As mentioned previously, one noticeable disadvantage of existing GHM and LSM is the lack of comprehensive modeling of all components in the water cycle. For instance, most of commonly used physical models tend to underestimate the TWS changes caused by human-induced and climatic impacts (Jiang et al., 2020). Tang et al. (2010) also argued that the main limitation in using model-simulated TWS is in a given river basin, the effects of artificial reservoirs and irrigation water withdrawals, are not represented in most LSMs.

In recent years, satellite remote sensing (RS) has become a viable tool to investigate the spatiotemporal pattern of various hydrological variables at local or global scales. A number of satellite-borne instruments have been used for monitoring water storage changes. For instance, Jiang et al. (2020) implemented an independent component analysis of Global Navigation Satellite

System (GNSS) observations to investigate sudden TWS changes related to hydrometeorological extremes. GNSS remotely detects near-real-time TWS changes after significant hydrological events by measuring instantaneous water loading signals with millimeter-level precision, which provides useful constraints for operational hydrological monitoring of daily TWS variations. However, continuous GNSS-based tracking of TWS is greatly limited by station density and GNSS data processing procedures, which is not suitable for long-term TWS assessment. Lee et al. (2015) quantified the surface water storage changes in large scale forested floodplains using the L-band Synthetic Aperture Radar (SAR) instrument and Envisat altimetry measurements. In this paper, SAR images were used to extract the flooded extents, and satellite radar altimetry measures the water level. The variations in surface water storage were estimated by integrating the extracted flooded extents and water level measurements. Moreover, Yuan et al. (2017) applied Interferometric SAR and satellite altimetry measurements to map the spatial pattern of water depth and water volumes in floodplains. The two-dimensional (2D) maps were then used to estimate the absolute water storages over the study areas. It is worth mentioning that radar-based assessments rely on the SAR backscattering coefficients to estimate water level changes. However, the preferred L-band wavelength SAR data is only provided by a few satellites, which results in time gaps when observational data from these satellites were not available. Furthermore, the capability of radars and radiometers can be adversely affected by atmospheric and near-surface phenomena (Rodell et al., 2009).

The GRACE satellite mission was launched in March 2002, which was a cooperative project between NASA and German Aerospace Center (DLR) for monitoring the anomalies in Earth's gravity field (Rodell et al., 2009). The mission was carried out by two identical satellites operating in a coorbiting manner with 500 km altitude, each following the other with a distance of approximately 220 km. The satellites emit microwaves to monitor their separation distance, which changes as the satellites pass through gravity anomalies (Tapley et al., 2004). For example, when the first satellite reaches a positive gravity anomaly, its attraction increases, so the distance between the satellites increases. And when the second satellite reaches the gravity anomaly, its attraction also increases, thereby reducing the inter-satellite distance. Consequently, GRACE is able to demonstrate the mass changes in the earth surface in an aggregate form but cannot distinguish the components. The inter-satellite distance measurements are used to calculate the temporal changes in the gravity field after removing the effects of non-gravitational accelerations

as detected by onboard accelerometers (Tapley et al., 2004). These gravity changes are related to the redistribution of mass, which mostly caused by the circulation of water through lands, oceans, and atmosphere. After removing the atmospheric and oceanic effects, the monthly gravity variations obtained from GRACE can be inverted for global estimates of vertically integrated terrestrial water storage. And its accuracy increases as the spatial scale increases (Swenson & Wahr, 2002; Swenson et al., 2006). Comparing with SAR and other satellite instruments, GRACE has the major advantage that it is capable to detect changes in groundwater storage underneath the land surface (Rodell et al., 2009).

There is now substantial research adopting GRACE data to assess TWS variability, both on land and in the oceans (Jiang et al., 2014). For instance, Rodell et al. (2018) conduct an observation-based assessment of how global water landscape response to human impacts and climate changes, by quantifying trends in TWS observed by GRACE. In this study, the authors quantified 34 trends in GRACE-observed TWS from 2002 to 2016, and their findings show that seasonal variability, overexploitation of groundwater, climate change are the major driving factors of these trends. Bonsor et al. (2018) analyzed GRACE TWS data to estimate the interannual TWS variations in basins where in-situ observations are scarce. In a review study, Jiang et al. (2014) summarized a variety of hydrological applications of GRACE data including TWS change monitoring, drought analysis, flood analysis, hydrological components evaluation, and glacier mass balance detection.

Nevertheless, GRACE observations are not the optimal dataset for real-time water storage monitoring due to the latency period of two to three months, monthly temporal resolution, and data gaps (Jiang et al., 2020). Nevertheless, it has significantly facilitated the evaluations of GHM and LSM at the large scales (Niu et al., 2007; Cai et al., 2014; Ni et al., 2017; Zhang et al., 2017; Ma et al., 2017; Sliwainska et al., 2018; Scanlon et al., 2018; Scanlon et al., 2019). Although there is often a mismatch between the spatial resolution of GRACE product and that of corresponding GHM or LSM, the GRACE-derived TWS data can be downscaled to finer spatial resolution by applying proper post-processing procedures.

The post-processing of raw GRACE data results in different levels of data. The Level-0 data is the raw result of telemetry data reception. The Level-1 data is the output after converting binary encoded Level 0 data to engineering units, and to quantities used in Level-2 processing.

Level-2 data is presented in the form of near-monthly potential spherical harmonic coefficients with their full error variance-covariance matrix (Tapley et al., 2004). The Level-3 GRACE data is derived from Level-2 solutions to depict the Equivalent Water Height (EWH) to a normally geographic coordinate system (Landerer & Swenson, 2012). The Level-3 GRACE TWS products are available in two different types of solutions, spherical harmonic coefficient (SHC) solutions and mascon solutions, respectively (example products shown in Figure 2-2). Monthly TWSA product of both solutions can be downloaded from the website of NASA Jet Propulsion Laboratory (JPL) (https://grace.jpl.nasa.gov/data/monthly-mass-grids), in gridded format. The SHC solution product contains gridded TWS estimates derived from the Level-2 dataset through de-stripe and smoothing (Swenson & Wahr, 2006). On the other hands, the mascon solutions apply mass concentration blocks (i.e., mascon) to parameterize the gravity field (Watkins et al., 2015).

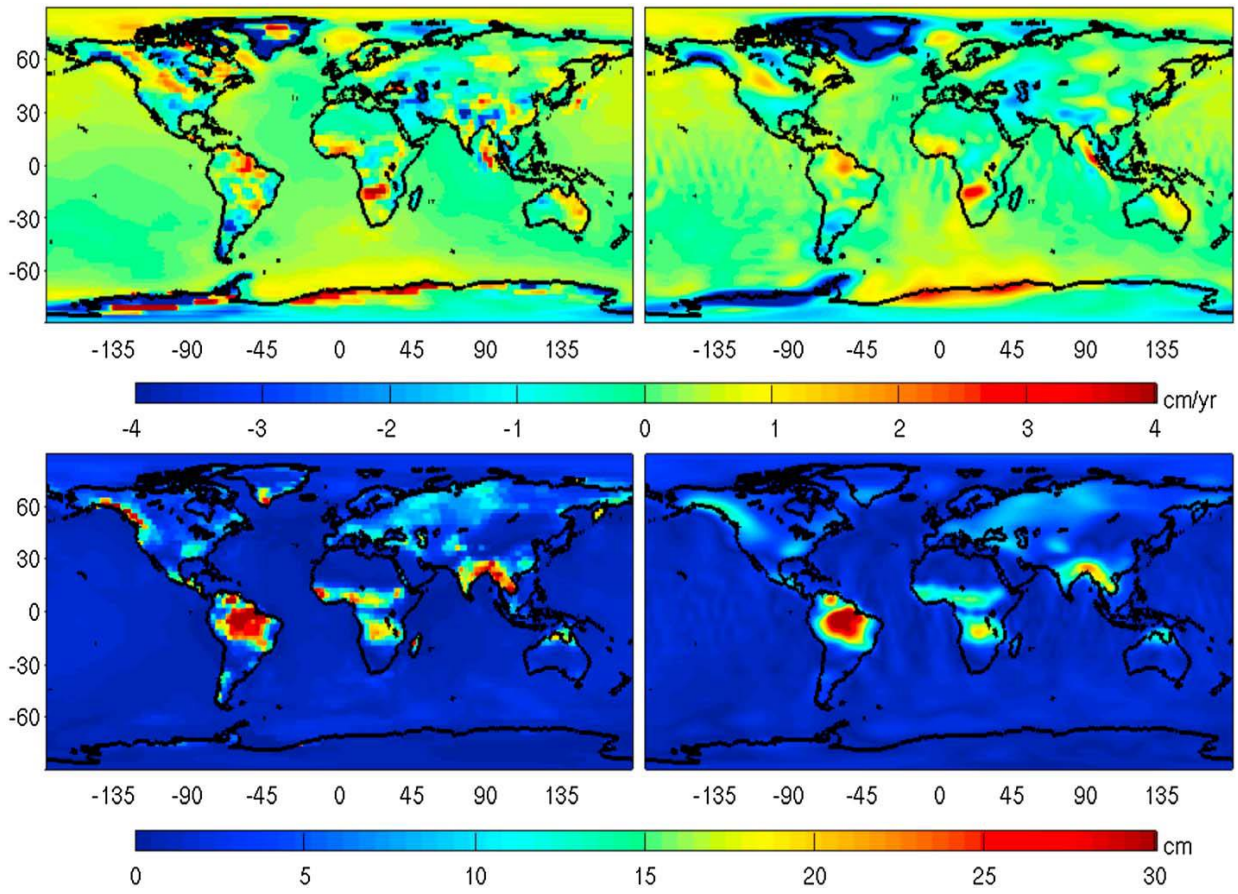

Figure 2-2. The 2003–2013 TWS trend (top) and annual amplitude (bottom) for JPL RL05M mascon solution (left) and JPL RL05 spherical harmonic solution DS300 (right), expressed in cm/year and centimeter of EWH, respectively. Source: Watkins et al. (2015)

The two GRACE solutions mainly differ in the adopted basis functions. The selection of solution products should be based on the research objective. For instance, the mascon solution contains information of geophysical location while SHCs lack localization information. However, each mascon estimate brings uncertainties resulting from leakage errors (i.e., errors in coastlines, which are intrinsic to the mascon basis functions) and solution errors (i.e., errors in the parameterization process) (Sun et al., 2020). Fortunately, these errors could be reduced or even removed during the data inversion process, by modelling the mascon's localization as priori information (Zhong et al., 2020).

2.2. Data fusion of GRACE, physically-based models, and climatic data

2.2.1. Assimilating GRACE observations into model simulations

The inaccuracies and uncertainties in GHM and LSM often limited their reliability. These drawbacks are results from oversimplified representation of hydrometeorological processes and the errors of climate forcing data (Soltani et al., 2021). GRACE satellite observations have facilitated novel approaches to calibrate and adjust GHM and LSM. One solution is to improve the model performance through changes in model architectures and hyperparameters (Lo et al., 2010), where GRACE data is used as a reference data. Another solution is to adopt data assimilation techniques, by using advanced statistical methods based on prior and posterior information, to directly calibrate and adjust the outputs of models, as well as reducing the systematic deficiencies in model. As a result, the simulation is able to fit better to the corresponding GRACE observation (van Dijk et al., 2014; Schumacher et al., 2016).

Considerable effort in recent years has been devoted to assimilating GRACE TWS into physically-based hydrological models to improve the model simulation of various TWS components. At each assimilation step whenever a new observation is available, the assimilation scheme applies error corrections based on the discrepancies between model simulations and GRACE observations. Apte et al. (2008) indicated that the accuracy of simulations and observations determine the correction level to be applied to the models, and correspondingly weight of observations. The authors adopt a Bayesian perspective by suggesting that assimilation updates the probability distribution function (PDF) of each state variable of the model in the presence of observation. However, hydrological models are high-dimensional which is either non-Gaussian or nonlinear. The computational burden would be large when the posterior PDF is not

analytically derivable for the state variables (Vrugt et al., 2013). Therefore, techniques of sequential filtering or variational filtering are applied to numerically solved the Bayesian estimation (see Figure 2-3) (Subramanian et al., 2012).

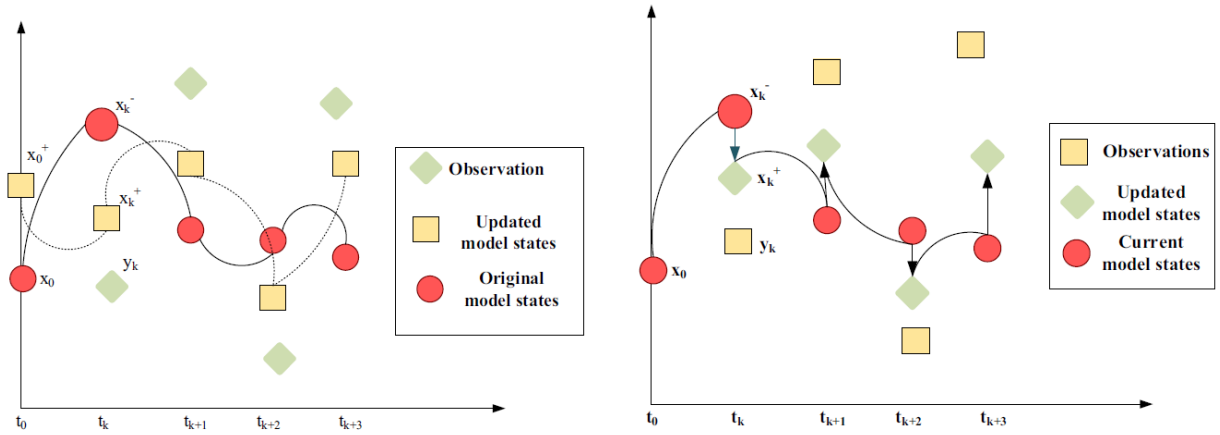

Figure 2-3. Scheme of variational filtering (left) and scheme of sequential filtering (right). Source: Soltani et al., 2021

For instance, van Dijk et al. (2014) conducted a global water cycle analysis that integrates GRACE-observed TWS, satellite water level altimetry, and simulations from five GHMs. Their approach was to combine GHM-simulated water balance components with several ancillary data sources (e.g., lake water level), to generate a combination of prior estimates of monthly TWS variations. The data assimilation was carried out by sequentially merging the ensembled model estimates and GRACE observations. However, the results contain great uncertainties occur in regions where glacier mass loss occurs. The authors suggested that the uncertainties could be caused by the error in original GRACE TWS data over glaciated land area. Schumacher et al. (2016) systematically assess the effects of spatially correlated errors in GRACE TWSA products as assimilating into WaterGAP GHM, by introducing the GRACE data in various spatial scales. They noticed that higher spatial resolution of TWSA observations would increase the impact of GRACE error correlation on the assimilation results, particularly in basins that are elongated in north–south direction and in basins where there is a large difference between model-simulated TWSA and GRACE-derived TWSA.

In addition to GRACE data errors, model uncertainties could result from model parameters, model structure, surface meteorological forcing, and unmodeled processes such as human

activities (Zhang et al., 2017). In other words, the modeling remains imperfect as the models still need further improvements on simulating all physical processes of water cycle and the intercorrelations between different water components. Girotto et al. (2017) investigate the modeling errors as assimilating GRACE TWS data into Catchment Land Surface Model (CLSM) over India. They noticed that GRACE is able to detect the TWS depletion caused by human-induced groundwater extraction, while the model failed to demonstrate such reduction as simulating representations of groundwater. Thus, the assimilation introduces a strong negative trend in simulated groundwater in the particular regions. However, the assimilation also introduces a negative trend in simulated evapotranspiration, which is unrealistic because evapotranspiration is likely enhanced by irrigation, but the model does not simulate irrigation. It shows that the assimilation is a double-edged sword as it can degrade certain water components while improving the others.

The interaction between physically-based models and GRACE data is a mutual process since GRACE data is able to improve the modeling performance by facilitating parameter tuning and uncertainty reduction, and in which the models provide a baseline and reference for data fusion and interpolation (Rajabi et al., 2018). Previous studies have shown that GRACE-derived TWS data have the capability of improving physically-based simulations, while there are some obstacles when assimilating GRACE observations into GHM or LSM. One of the impediments is that GRACE observations and LSM simulations have huge discrepancies in spatial and temporal resolutions (Soltani et al., 2021). In fact, the hydrological applications of GRACE-derived TWS are often limited due to their monthly temporal resolution and coarse spatial resolution (~110 km), and the distinguishing of different water storage components (Girotto et al., 2016). Fortunately, the fusion of GRACE TWS and model-simulated TWS have the potential to partition the vertically integrated GRACE observations into disintegrated surface and subsurface water components. Schumacher et al. (2018) suggested that the integration of GRACE data and GHM/LSM does not only improve simulation of seasonality and trend of TWS, but also it improves the simulation of individual water storage components. For instance, Girotto et al. (2016) proposes an assimilation system to integrate gridded GRACE-TWS observations and LSM, for specifically improving groundwater and soil moisture estimates.

Moreover, the fusion techniques can be used to downscale GRACE-TWS information to finer spatial and temporal scales (Rahaman et al., 2019). In fact, downscaling has been an important role in research related to satellite remote sensing. Yin et al. (2018) indicated that the downscaling methods can be categorized into dynamic downscaling and statistical downscaling. Dynamic downscaling is to construct a regional numerical model at higher spatial resolution, and the model can be applied at smaller scales. Statistical downscaling is to establish the relationship between large-scale variables and data from small-scale observations. Dynamic downscaling has been widely adopted to downscale GRACE TWS through assimilating into LSMs (Zaitchik et al., 2008; Houborg et al., 2012; Sahoo et al., 2013), but such applications often require extensive computing time as well as complex data sources, and many of the above-mentioned procedures depend on the selected LSM, some of which are lacking surface or groundwater components (Sahour et al., 2020). On the contrary, statistical downscaling is relatively simple to implement. During recent years, a variety of statistical methods, such as multivariate regression and artificial neural networks (ANN), have been applied to downscale GRACE data to high resolution dataset (Yin et al., 2018; Seyoum et al., 2019; Sahour et al., 2020; Zhong et al., 2020).

2.2.2. Reconstruction/Extension of GRACE TWS datasets using climatic variables

The temporal coverage is often referred as the major shortcoming of GRACE-derived TWS data because the time span of available GRACE observations is less than 20 years (April 2002 to the present), along with a 1-year data gap between the two GRACE missions (July 2017 to May 2018), which greatly limits GRACE applications for long-term and consistent hydrological studies. Classical approaches to hindcast historical TWS would either rely on physically-based models or basin-scale water balance calculations (Mueller et al., 2011). However, the estimation may be limited by uncertainties in model parameters and structures. Therefore, a number of studies have dedicated to the subject of data-driven reconstruction of GRACE TWS dataset by constructing empirical relationships between GRACE TWS and related climatic and hydrological variables (Nie et al., 2015; Humphrey et al., 2017; Tang et al., 2017; Yin et al., 2019; Li et al., 2020; Sohoulande et al., 2020). For instance, Nie et al. (2015) reconstructed monthly and annual TWS time series over the Amazon Basin from 1948 to 2012 by integrating GRACE TWSA data, with multiple hydrological variables derived from Global Land Data Assimilation System (GLDAS) products, including monthly precipitation, evapotranspiration, surface runoff, and subsurface runoff. Humphrey et al. (2017) presented a global TWS data reconstruction for the period 1985 to

2015 using a statistical model based on precipitation and temperature, but the reconstructed product neglects temporal variability and trends. Tang et al. (2017) developed a water balance model for reconstructing annual TWS change and groundwater storage (GWS) change during 1980 to 2015. The model is based on Budyko equation which describe the correlations between mean annual precipitation, mean annual evapotranspiration, and runoff. The model parameters were estimated by minimizing the root-mean-square error (RMSE) between the Budyko-incorporated TWS and GRACE-derived TWS. Sohoulande et al. (2020) proposed a unified predictive framework to retrieve monthly TWSA by incorporating climatic and hydrological variables as considering the impacts of lag signals. The framework consists of two stages operating on gridded data. The first stage applies principal component analysis on multiple meteorologic variables represented in time-series form. In the second stage, a multivariate regression is applied on the principal components' scores. The authors noted that for certain locations the prediction of TWS is limited because the framework is not comprehensive enough as it only relies on climate forcing.

2.3. ML/DL applications in geoscientific and hydrological studies

The advances in data acquisition techniques, such as satellite remote sensing and crowdsourcing tools, have led to the substantial growth of the volume and variety of geospatial data (Yuan et al., 2020). With the abundant data sources and enhanced computational power, many machine learning (ML) techniques have made great contribution in geoscientific studies (Reichstein et al., 2019). Deep learning (DL) is a subbranch of ML focusing on large-size and deep ANNs. It has gained remarkable results in modelling spatiotemporal data (i.e., combining spatial learning and sequence learning) in the fields of computer vision and time-series analysis, which provides opportunities for novel methods in environmental monitoring from satellite imagery data (Shen, 2018). Reichstein et al. (2019) argued that classical ML approaches may not be optimal for modelling spatiotemporal systems (e.g., water cycles). On the other hands, DL is able to gain further understanding of these complex systems by automated feature extraction. Therefore, they suggested that DL should be used for analyzing the system behavior dominated by spatiotemporal context, as well as performing spatial and temporal interpolation (i.e., predictive ability).

2.3.1. Common DL network architectures in geoscientific studies

This section summarizes a selection of the cornerstone neural network architectures that commonly used by papers related to geoscience and spatiotemporal data.

Autoencoders

Autoencoders (AE) are neural networks (or elements of deep neural networks) that are used to reduce the dimensionality of datasets and then reproduce the inputs, so that the output will be a close representation of the input (Hinton & Salakhutdinov, 2006). Autoencoding refers to a unsupervised data compression algorithm that learned automatically from data samples. AE is divided into two parts, namely encoder and decoder, and each part can be regarded as several hidden layers between an input layer and an output layer. The performance of the network is measured by how close the outputs approximate the inputs after training. AE are implemented in an unsupervised manner to generate a representation feature of the dataset within the hidden layer neurons in the bottleneck, also called the latent vector (see Figure 2-4a), the trained weights are optimized as generating these features. On the output side, the network needs to reconstruct the inputs with condensed information (Shen, 2018). Zhang et al. (2019) proposed an unsupervised RS image segmentation method based on a dual autoencoder network, which shows satisfactory results on water information extraction.

Convolutional neural network

Convolutional neural network (CNN) is composed of a stack of basic units that shrink in width from input to output (see Figure 2-4b). Each basic unit could be one of the operational layers including convolutional layer, pooling layer, and activation layer (Yuan et al., 2020). A convolutional layer applies multiple cascaded convolution kernels to extract intricate knowledge from the input. For example, a convolutional layer that processes an image can extract positional information from images, such as understanding spatial autocorrelation. The pooling layer can obtain the dimensionally-reduced data from the input data through max-pooling or average-pooling operations. CNNs have been widely applied to deal with imagery analytical tasks such as missing data prediction (Das & Ghosh, 2017), cloud removal (Zhang et al., 2018), and land surface temperature reconstruction (Wu et al., 2019), due to its strong nonlinear representation capability, which acquired state-of-the-art results. For instance, Pan et al. (2019) introduced a CNN-based model to predict daily precipitation and tested it at 14 monitoring sites across the contiguous United States and proved that, if provided with sufficient data, the precipitation estimates from the CNN method are better than the reanalysis precipitation products and statistical downscaling products. In computer vision, CNN have shown its significance for the analysis of RS imagery.

Recurrent neural networks and Long short-term memory networks

Recurrent neural network (RNN) includes a recurrent layer which the neurons within the layer could be connected to each other. This mechanism enables RNNs to carry information learned within a neuron to the next neuron in the same layer. Therefore, RNN is convenient when the input data is in sequential format such as a time-series data or a text. RNN has become an important model for analyzing time-series changes in spatiotemporal systems due to its excellent performance in short sequence processing. Spatiotemporal systems are common in the real-world, such as traffic flow, diffusion of air pollutants, and regional precipitation. Hindcasting or forecasting the past/future states of these spatiotemporal systems based on the available observations is a significant and challenging problem. For some complex spatiotemporal dynamical systems like atmosphere and aquifer, traditional numerical models are not capable to make prediction due to the lack of knowledge about the systems' inner mechanisms. In these complex situations, RNN-based methods have been proven useful for making accurate predictions, by which the systems' inner mechanisms can be learned based on the historical data (Zheng et al., 2015). For instance, Freeman et al. (2018) applied the RNN model to nowcast air pollution with time series data from air monitoring stations in Kuwait. In general, RNN shows advantage in handling short sequence problems.

However, RNN perform relatively poor on processing time series data over a long period (Yuan et al., 2020). As a variant of RNN, Long Short-Term Memory (LSTM) networks have been widely adopted for problems with long time sequence. A LSTM layer is composed by a number of memory blocks that are recurrently connected (see Figure 2-4c). Each memory block contains memory cells and gate units (input gate, forget gate, output gate). Gates are neural net layer with trainable weights, which enables optional transfer of information within the memory block. During the training process, the input gate determines if the inputs are significant to be saved in the memory cell, the forget gate determines if which past state memory should be reserved, and the output gate decides how the saved memory is used to produce the output. This mechanism allows the LSTM network to remember information from the long past while neglecting nonessential information (Shen, 2018). Unlike RNN, LSTM neurons produce two different values yielded by a series of activations and operations. You et al. (2017) developed a LSTM-based model for crop yield prediction in which enables real-time forecasting. Reddy and Prasad (2018) applied LSTM

and ANN on satellite-derived vegetation index to forecast changes in vegetation cover. Yang et al. (2018) built an end-to-end LSTM network to conduct forecast of seawater surface temperature. One convolutional layer was added to the network so that not both temporal and spatial relationships of the time series data are included. In addition, Fang et al. (2017) and Fang et al. (2019) estimated the long-term soil moisture by LSTM, which exhibits enhanced transferability capability as comparing with autoregressive models and traditional linear regression.

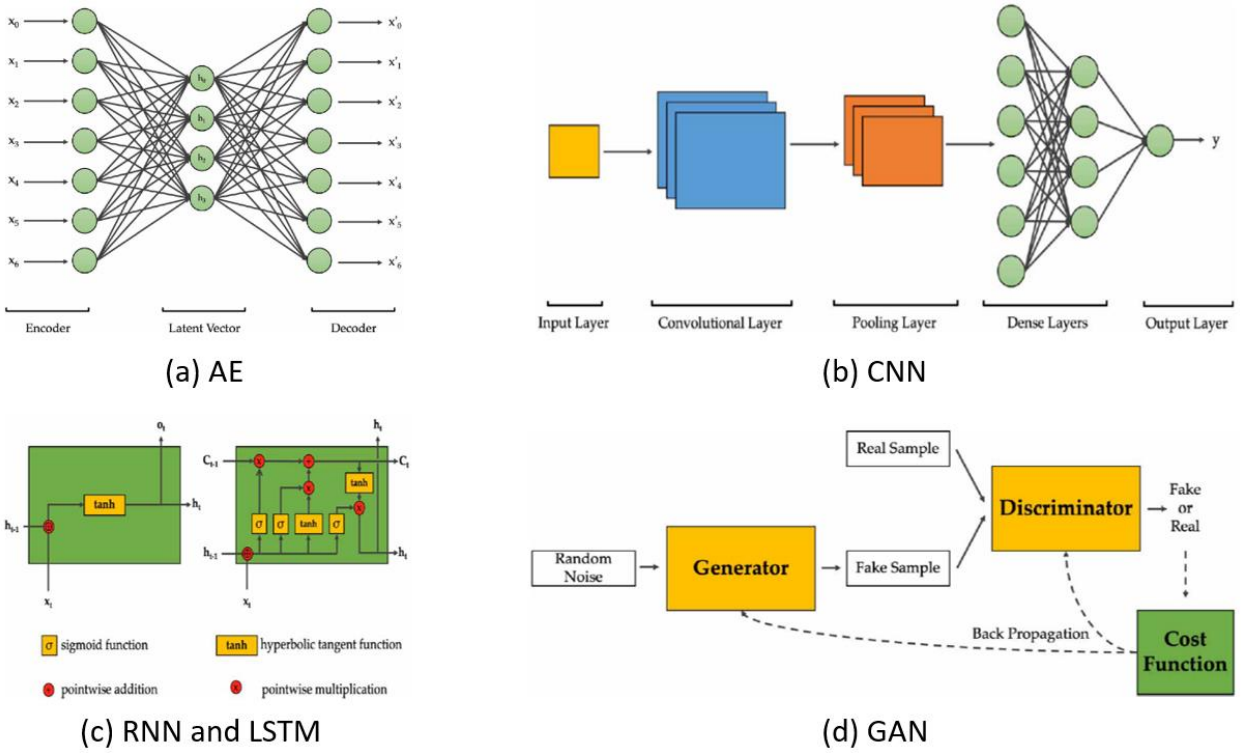

Figure 2-4. Architectures of components of four types of DNN structures: (a) Autoencoder, (b) CNN, (c) RNN and LSTM, and (d) GAN. Source: Sit et al. (2020)

Generative adversarial networks

Generative adversarial networks (GAN) were motivated by the need to model high-dimensional and multimodal distributions (Goodfellow et al., 2014). It consists of two separately running neural networks, which are generator and discriminator, respectively (see Figure2-4d). The two networks are competing with each other during the training. The generator aims to generate fake examples out of the input dataset while the discriminator aims to determine whether the generated example is fake or not. As they both intend to 'fool' each other, it causes them to

improve their capability of generating fake outputs and discriminating fake from real. GANs are capable of learning translation tasks such as super-resolution (Lei et al., 2020) or image-to-image translation (Isola et al., 2017), which can be applied in RS imagery processing and analysis.

2.3.2. DL opportunities in hydrology and water resource

Table 2-2. Conventional methods and DL-based methods for geoscientific tasks. Source: Reichstein et al. (2019)

Analytic task	Scientific task	Conventional methods	Potential DL methods	Related papers
Classification and anomaly detection	Land use and change detection	Pixel-by-pixel spectral classification	CNNs	Cao et al., 2019
	Find extreme weather patterns	Multivariate, threshold-based detection	Supervised and semi-supervised CNNs	Racah et al., 2017
Regression	Predict vegetation properties from atmospheric conditions	Semi-empirical algorithms (temperature sums, water deficits)	RNNs	Minh et al., 2018
	Predict fluxes from atmospheric conditions	Random forests, kernel methods, feedforward neural networks	RNNs, LSTM networks	Asanjan et al., 2018
	Predict river runoff in ungauged catchments	Process models or statistical models with hand-designed topographic features	Combination of CNNs and RNNs	Yuan et al., 2018
State prediction	Downscaling and bias-correcting forecasts	Dynamic modelling and statistical approaches	CNNs, conditional generative adversarial networks (cGAN)	Vandal et al., 2017
	Precipitation nowcasting	Physical modelling with data assimilation	Convolutional LSTM (ConvLSTM)	Shi et al., 2015
	Seasonal forecasts	Physical modelling with initial conditions from data	ConvLSTM	Mu et al., 2019
	Transport modelling	Physical modelling of transport	Combination of physical modelling and CNNs	de Bezenac et al., 2019

Unlike some other disciplines, the applications of DL have not been widely adopted in solving problems in earth system science, especially in the field of hydrology. Previous works regarding DL applications in hydrology can be categorized into three main types, first is the extraction of hydrometeorological information from gridded data (e.g., RS imagery); Second is the dynamic modeling of hydrologic variables observed/measured by sensor networks; and third is the learning and prediction of intricate data distributions (Shen et al., 2018). Nevertheless, in fields related to earth system science such as RS and GIS, application of DL has progressed rapidly because it has done exceedingly well in extracting information from raw observations and measurements. Reichstein et al. (2019) summarized the comparisons between conventional approaches and DL approaches to a variety of geoscientific tasks, as shown in Table 2-2.

Due to the fact that hydrological modelling relies on sequential datasets, many of previous studies focus on sequence prediction tasks and regression tasks. And studies based on imagery data (or 2D maps) tend to work on matrix prediction tasks. Therefore, CNNs and LSTMs are commonly used in DL-based hydrologic research, while GANs have not been widely adopted in this filed (Sit et al., 2020). The selection of network architecture also depends on the type of subdomain. For instance, some studies focus on groundwater modelling with ancillary data on 2D maps, such as CO2 saturation field map (Mo et al. 2019), water balance map (Sun et al., 2019), and hydraulic conductivity field map (Zhou et al., 2020). As a result, these studies used CNNs to construct the models. On the other hands, some studies focused on capabilities of surface water prediction, such as water temperature and water level. For instance, Xiao et al. (2019) developed a ConvLSTM model to predict the temperature of seawater surface based on satellite-derived temperature data for 36 years. The results showed that the ConvLSTM model significantly outperforms two different LSTM models for short-term and mid-term prediction. Qi et al. (2019) forecasted daily reservoir inflow by integrating the results from the LSTM models and decomposed inflow data.

2.3.3. GRACE data processing using ML and DL

During the past few years, applying ML/DL technologies on GRACE data have gained a lot of attention. Most of these previous works focus on two tasks: the spatial downscaling of GRACE data, and the reconstruction/extension of GRACE-derived dataset. As mentioned previously, the application of GRACE at local scales has been greatly limited due to the coarse

spatial resolution. Seyoum et al. (2019) implemented a tree-based model based on Boosted Regression Tree, to downscale GRACE-derived TWSA into high-resolution map of GWS anomalies in a glacial aquifer. Rahaman et al. (2019) adopted random forest, an unsupervised ML method, to integrate GRACE data with other hydrological variables simulated by the Noah LSM. Their proposed method downscaled the spatial resolution of GRACE-derived GWS anomalies in Northern High Plains aquifer from 1 degree to 0.25 degrees. Similarly, Chen et al. (2019) developed a random forest model for downscaling GRACE-derived GWS in northeast China. Sahour et al. (2020) applied multilayer perceptron to extract relationships between coarse resolution GRACE observations (110 km x 110 km) and various hydrologic variables (e.g., precipitation, snow cover, streamflow, temperature, soil moisture and evapotranspiration). The extracted relationships and fine-resolution dataset of these variables are used to predict monthly TWS data for the test site, with a downscaled spatial resolution of approximately 14 km x 14 km.

As for the reconstruction or extension of GRACE datasets, the utilization of ML and DL methods has been shown to be useful by a number of recently published studies. In a study by Yang et al. (2018), the authors used random forest, support vector machines, and ANN methods to hindcast over 50 years of GRACE-like TWSA time series in large-scale river basins located in northwestern China, based on GRACE-derived TWSA and hydrological variables from GLDAS. The results showed that the random forest outperform other methods. Ahmed et al. (2019) forecasted GRACE TWS on 10 major African watersheds and predicted drought events. A recurrent dynamic ANN model was constructed to investigate the nonlinear relationships between GRACE TWS data and related hydrological variables. Sun et al. (2019) applied three CNN models which originally designed for semantic segmentation, to learn the spatiotemporal patterns of mismatch between GRACE-derived TWSA and TWSA simulated by Noah LSM in India's landmass. The trained models are capable to predict TWSA at any timestamp given the corresponding Noah simulations. Jing et al. (2020) conducted a case study to hindcast historical TWSA at the Nile River basin by developing a ML-based reconstruction model incorporating random forest algorithm and a spatially moving window structure. The model was then used to calibrate LSM-simulated historical TWSA from GLDAS based on GRACE-derived TWSA. It is noticeable that previous efforts have demonstrated the potential of ML/DL in reconstructing GRACE-like data at basin and local scales. However only a handful of ML/DL architectures were applied, it is worth to examine the capabilities of more architectures.

Chapter 3

Reconstructing GRACE-like TWS Anomalies in the Canadian Landmass Using Deep Learning and Land Surface Model

3.1. Introduction

Terrestrial water storage (TWS) is referred to water volumes both underneath and above the Earth's surface (Jing et al., 2020). TWS is considered as the main component of terrestrial and worldwide hydrological process, which its dynamics under different scenarios of environmental changes including climate change and human disturbance is key to determine water resources sustainability and vulnerability (Famiglietti, 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 relied either 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 coefficients limit the understanding of water storage changes at large scales.

The emergence of satellite remote sensing (RS) enabled continuous monitoring over hydrological fluxes at different spatial resolutions. The variations in TWS change the gravity filed over a 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 satellite is probably the only RS-based method to quantify long-term TWS anomalies (TWSA) at large-scales. 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 Peng et al. (2017) to measure 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 and 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 for only about 15 years, 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). Nowadays, various methods have been developed for the reconstruction of missing information in RS data for different problems. 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 not capable to make prediction due to the lack of knowledge about the systems' inner mechanisms. On the contrary, 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 with 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 3.2 describes the data used, as well as the data pre-processing methods. Section 3.3 presents the methodology and implementation details of model constructions. The results are

presented and discussed in Section 3.4. And lastly, Section 3.5 states the main conclusions and findings.

3.2. Data and data preprocessing

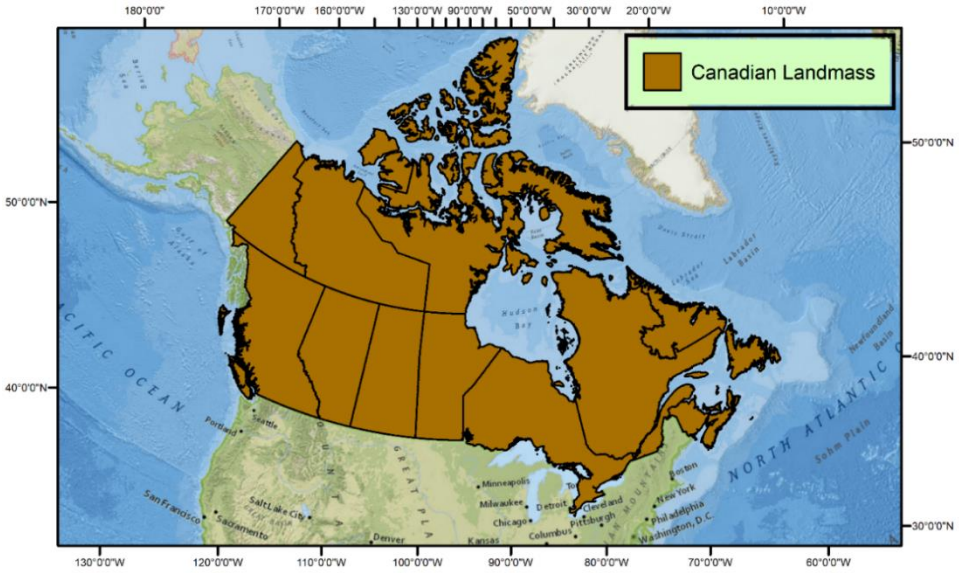

Figure 3-1. Map of the study area. Canada's landmass is represented by brown color

3.2.1. GRACE TWSA data

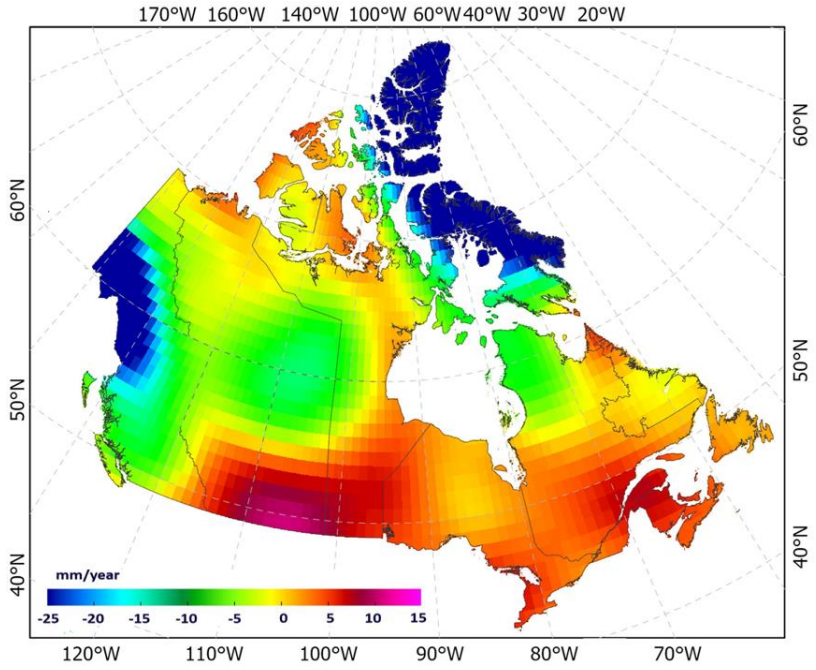

Figure 3-2. Trend of GRACE TWS in the study area, calculated as the linear slope of the TWS over 2002-2016.

The GRACE monthly TWSA product (RL06 spherical harmonics solution) was downloaded from the Jet Propulsion Laboratory (https://podaac.jpl.nasa.gov) of NASA. The product is provided with a grid resolution of 1 degree (~110km) 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), so that GRACE TWSA and EALCO TWSA data can be comparable to each other (Zhong et al., 2020). The data was clipped by a Canadian national boundary shapefile to cover the Canada's landmass, and then reprojected to Canada Lambert Conformal Conic (CanLCC) projection. Figure 3-2 shows the GRACE TWS trend during the study period. The trend exhibits values varying across the locations, and there are strong negative trends in Yukon and Arctic region. The significant long-term water loss in these regions is presumably resulting from melting glaciers and permafrost (Wang & Li, 2016).

3.2.2. 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 hydrologic information with a relatively high spatial and temporal resolution (Wang, 2005). However, EALCO cannot provide complete information brought by the GRACE satellites, as it may underestimate the trends caused by climate change (e.g., glacial ablation) and human-induced water storage changes. As can be seen from Figure 3-3, EALCO does not reflect the decreasing trend indicated by GRACE. Therefore, the TWS data simulated by the EALCO LSM is not applicable to directly reconstruct the historic TWS dataset.

The EALCO daily TWS product (v4.2) used in this study cover the period from January 1979 to December 2016, with a spatial resolution of 5km. 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 the same baseline value 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 same extents and coordinate system as the GRACE data, in order to ensure that pixelwise match between EALCO and GRACE. Lastly,

the daily EALCO TWSA data was averaged accordingly to match 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 Figure 3-4) representing all the GRACE coverage phases 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 x 48 pixels.

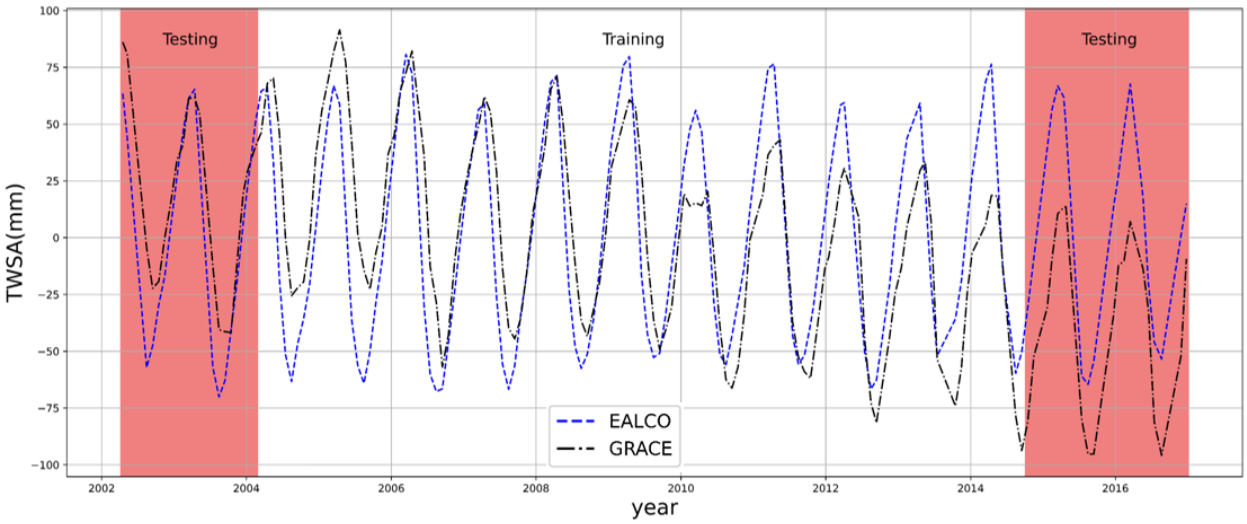

Figure 3-3. Comparison of nationwide mean EALCO-simulated TWSA and mean GRACE-derived TWSA. Red background indicates the time coverage of the testing set.

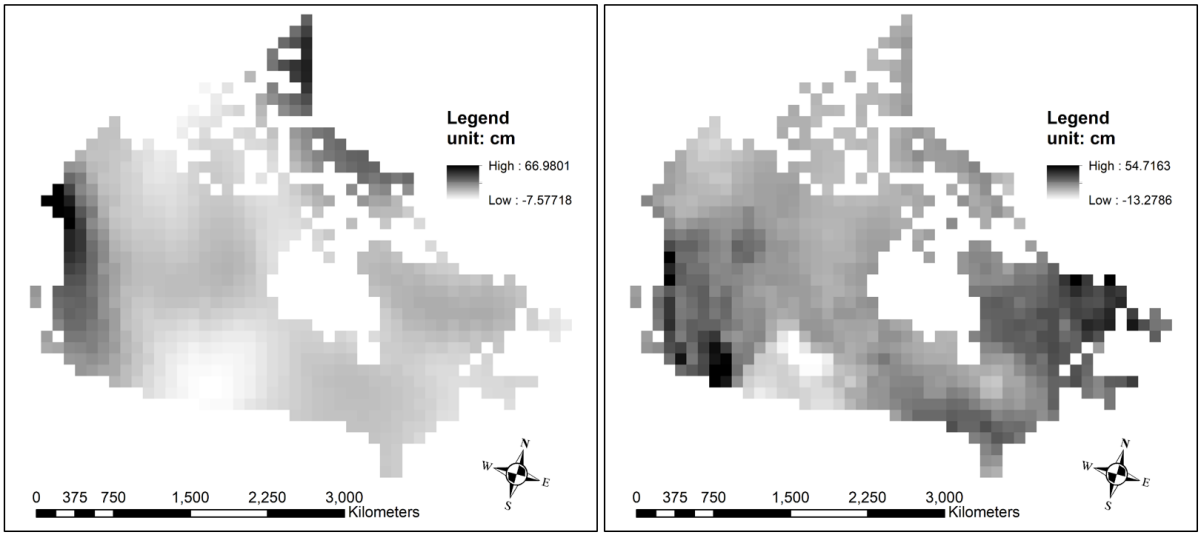

Figure 3-4. GRACE TWSA (left) and upscaled EALCO TWSA (right) for April 2002 under CanLCC projection

3.2.3. Train-test split and further processing

In total there are 158 GRACE-EALCO monthly data pairs, corresponding to the 158 GRACE coverage months. Given its sequenced time-series form, the first 20 months (April 2002 to February 2004) and the last 20 months (October 2014 to 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.

3.3. 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 $\mathbf{X} = \{X_i\}_1^N$ and $\mathbf{y} = \{y_i\}_1^N$, as stated in the formula:

$$y = f(X, p) \tag{3.1}$$

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.

3.3.1. 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 Figure 3-5).

3.3.2. 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 Figure 3-6). 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.

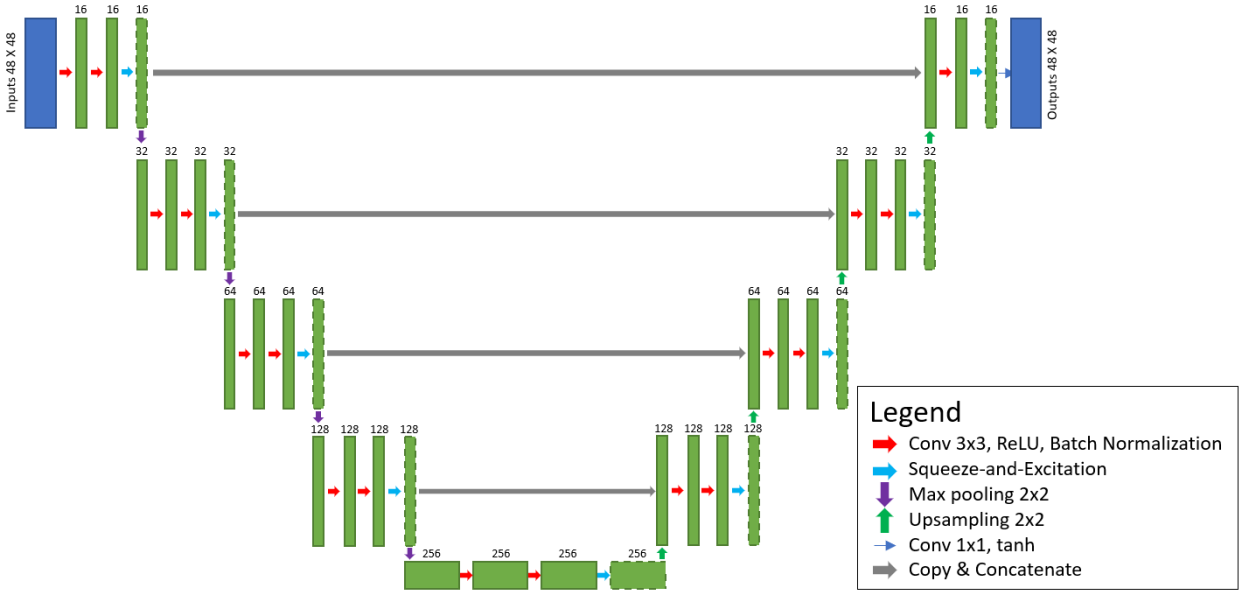

Figure 3-5. Schematic diagram of the SEUNet architecture

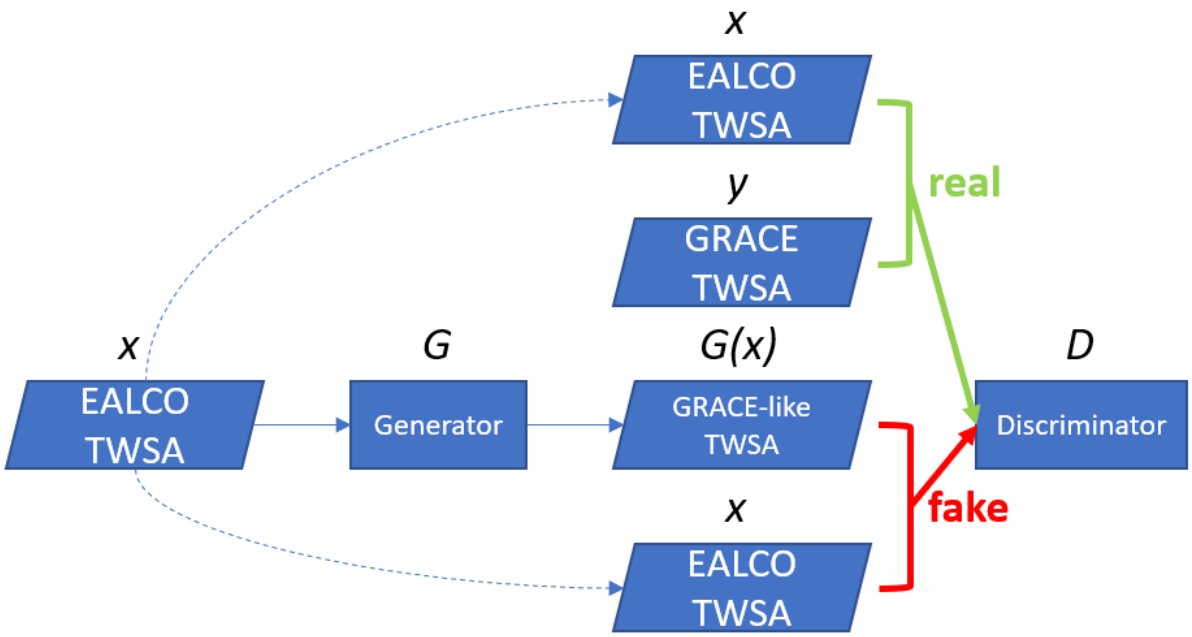

Figure 3-6. Diagram of training a Pix2Pix conditional GAN to map EALCO TWSA to GRACE TWSA. The discriminator D learns to classify between fake (generated, EALCO) and real (GRACE, EALCO) tuples. Unlike an unconditional GAN, both the generator and discriminator take the predictor (EALCO) as inputs.

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 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.

3.3.3. 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 Figure 3-7.

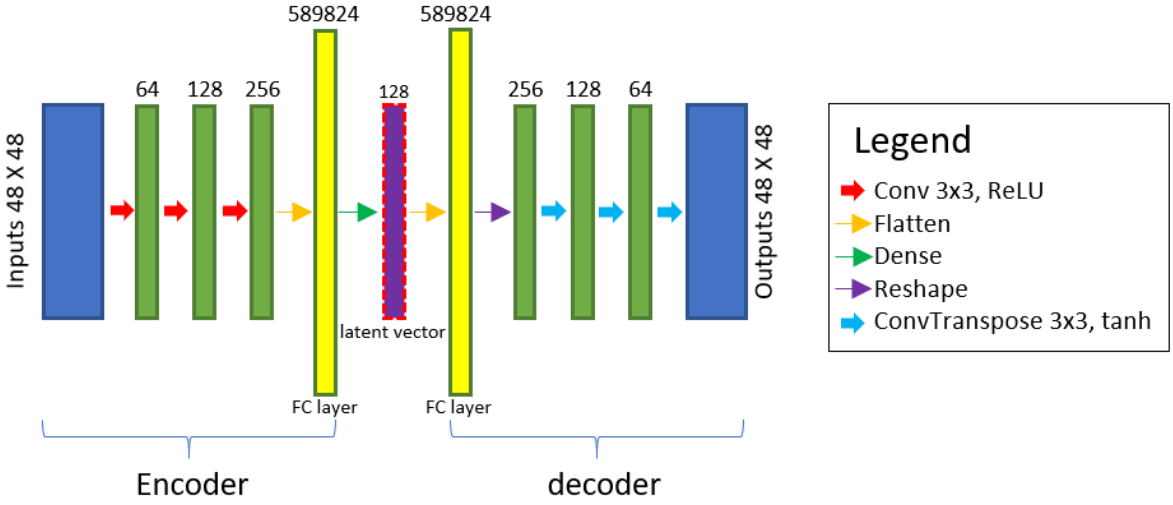

Figure 3-7. Schematic diagram of the proposed DCAE architecture

The encoder contains 3 convolution layers for down-sampling by 3x3 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 128x128 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 3x3 kernel and tanh activation function.

3.3.4. Encoding-Forecasting ConvLSTM

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

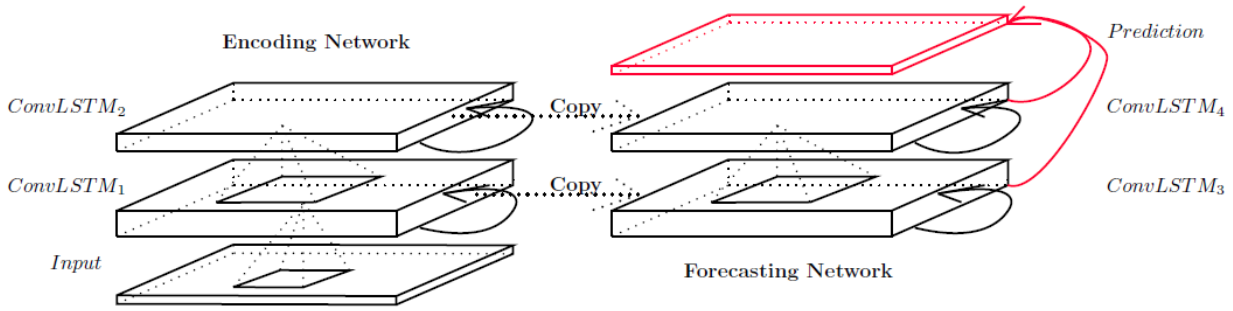

Figure 3-8. Encoding-forecasting ConvLSTM architecture. Source: Shi et al. (2015)

By stacking multiple ConvLSTM layers, Shi et al. (2015) developed an encoding-forecasting structure (see Figure 3-8) 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.

M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20
M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21
M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	M 22
...	...	...	...																

Sequence 1
Sequence 2
Sequence 3
.....

Figure 3-9. 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.

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 Figure 3-9. 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).

3.3.5. Implementation details and Evaluation metrics

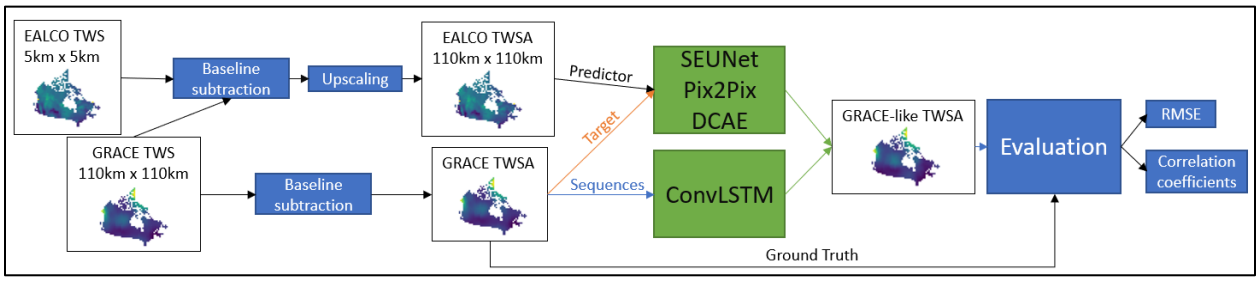

Figure 3-10. Workflow diagram. EALCO-simulated and GRACE-derived TWSA were calculated by subtracting the baseline value (mean TWS between April 2002 and December 2016). The spatial resolution of EALCO TWSA was then upscaled from 5km to 110km to be matched with that of GRACE TWSA. SEUNet, Pix2Pix and DCAE take both EALCO and GRACE TWSA as inputs, and ConvLSTM only takes GRACE as input sequences.

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, 8Gb RAM). A detailed workflow chart for this study is presented in Figure 3-10.

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:

$$CC = \sum_{i=1}^{n}(y_i - \bar{y})\,(g_i - \bar{g}) / \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(g_i - \bar{g})^2} \quad (3.2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - g_i)^2} \quad (3.3)$$

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.

3.4. Results and Discussion

3.4.1. 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 Figure 3-11. It is worth noting that the metrics between EALCO and GRACE is the baseline for the performance assessment of DL models. Figures 3-11a and 3-11f 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 Figure 3-11(g, h, i, 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.

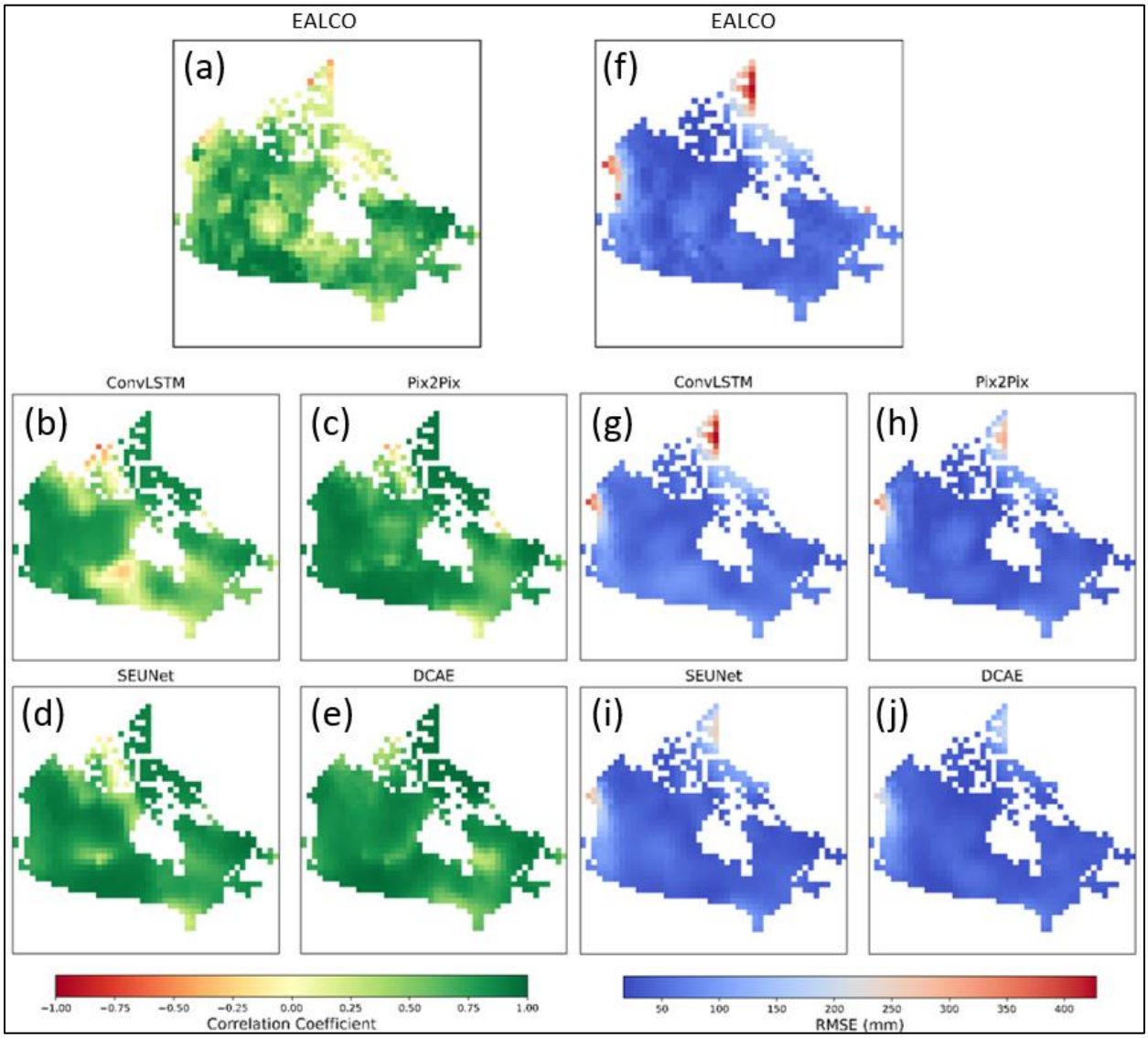

Figure 3-11. Spatial distributions of pixelwise mean CC (a, b, c, d, e) and mean RMSE (f, g, h, i, j) derived from EALCO, ConvLSTM, Pix2Pix, SEUNet and DCAE

The results of nationwide mean CC and mean RMSE are summarized in Table 3-1. 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.

Table 3-1. 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**

Figure 3-12 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 Figures 3-13, 3-14, 3-15, and 3-16, corresponding to the TWSA spatial distributions in January 2003, July 2003, October 2014, and April 2015, respectively. It is worth 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 to 10 minutes (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.

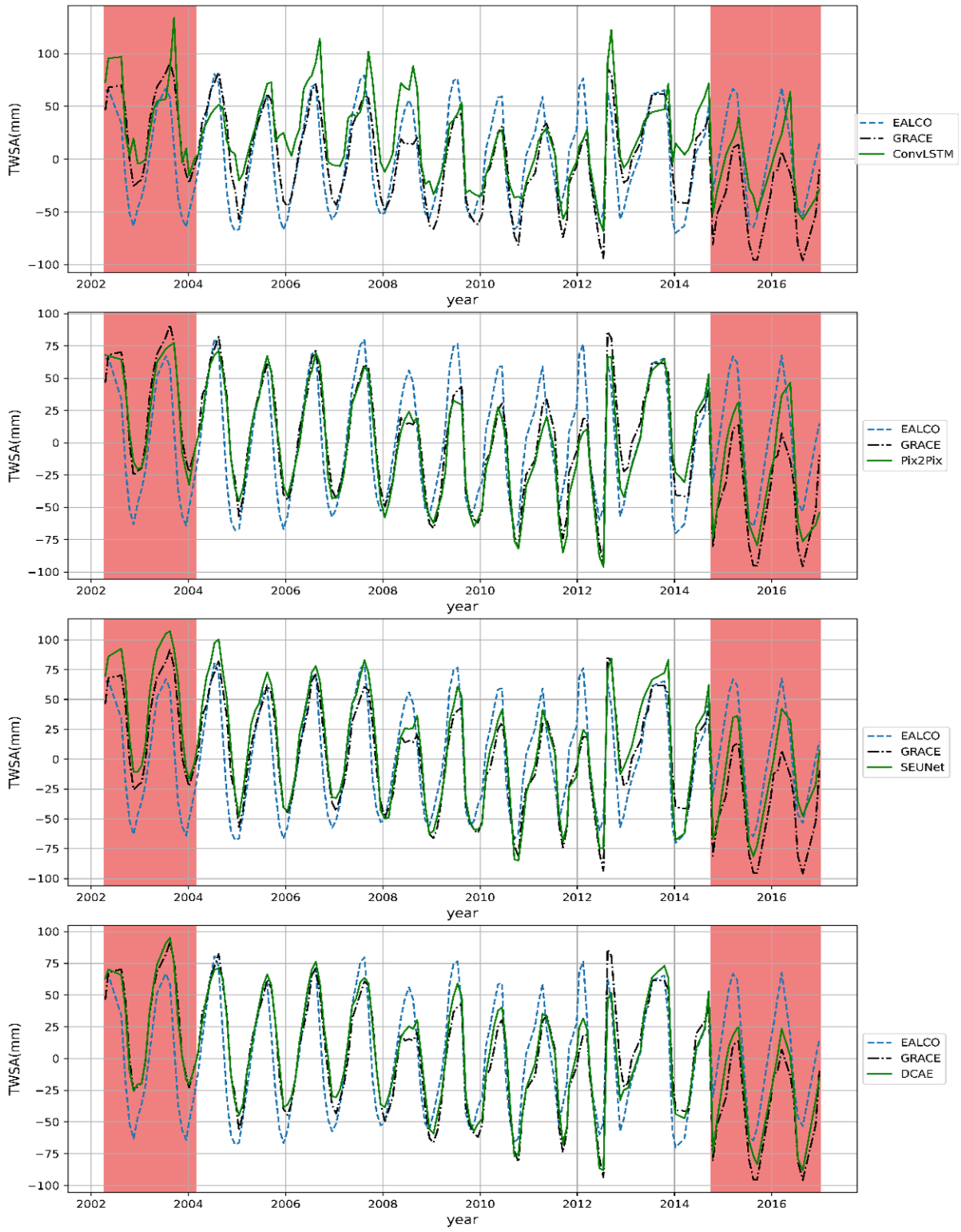

Figure 3-12. Comparison of model-reconstructed GRACE-like TWSA, EALCO-simulated TWSA, and GRACE-derived TWSA during training (white background) and testing periods (red background) at nationwide level. From top to bottom: ConvLSTM, Pix2Pix, SEUNet, DCAE

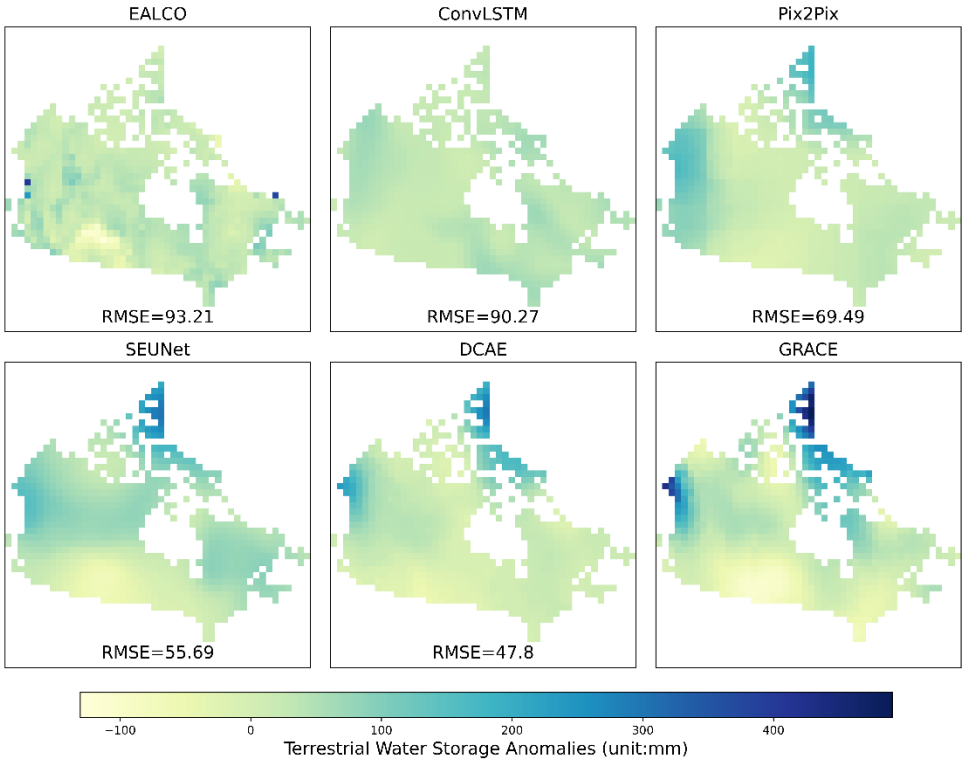

Figure 3-13. Reconstructed TWSA maps for January 2003

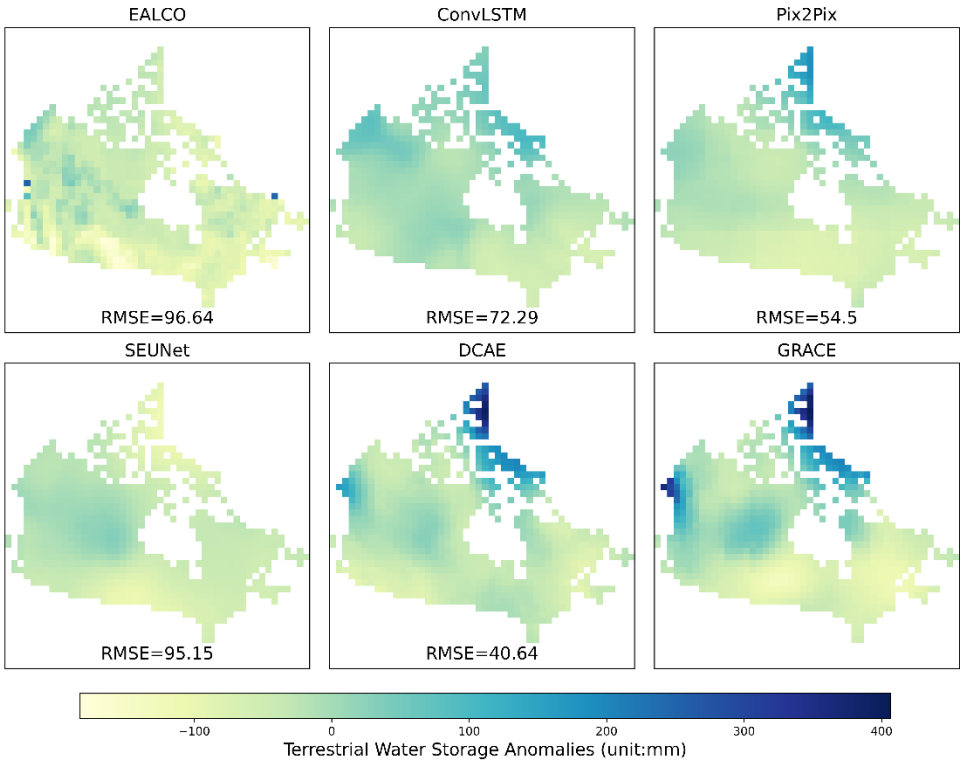

Figure 3-14. Reconstructed TWSA maps for July 2003

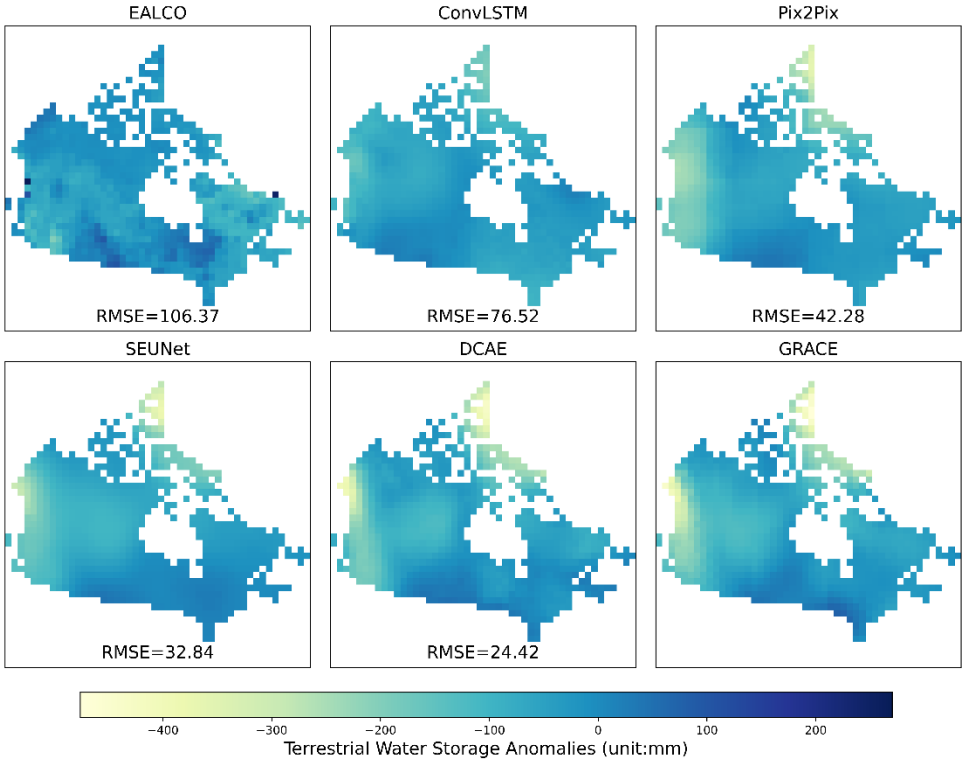

Figure 3-15. Reconstructed TWSA maps for October 2014

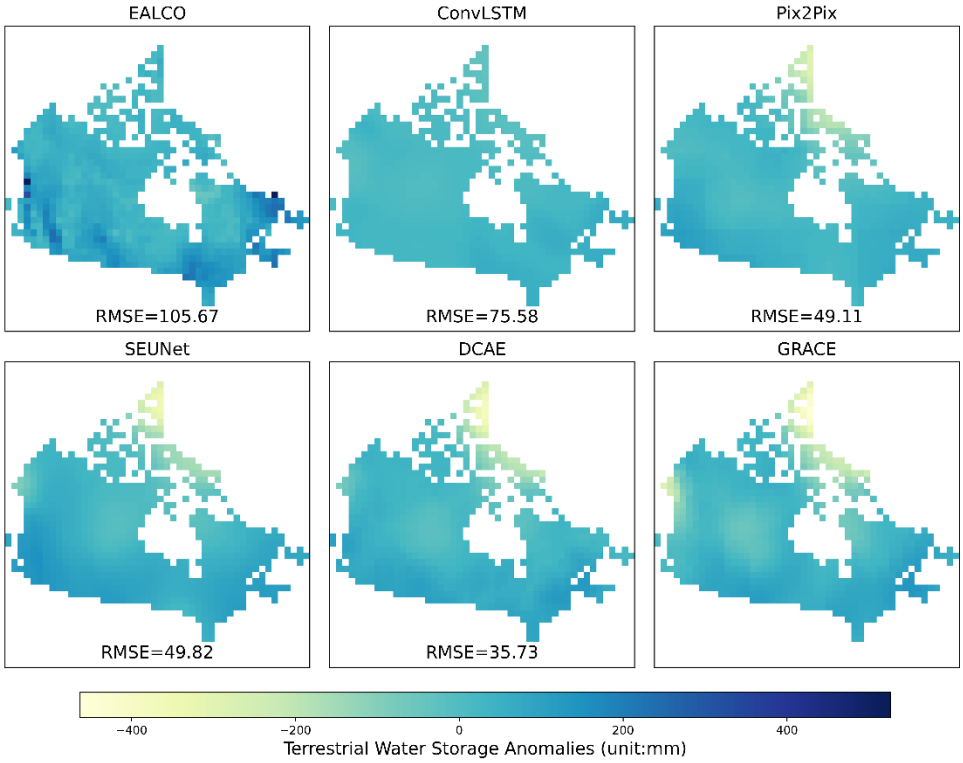

Figure 3-16. Reconstructed TWSA maps for April 2015

3.4.2. 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 Figure 3-17. 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.

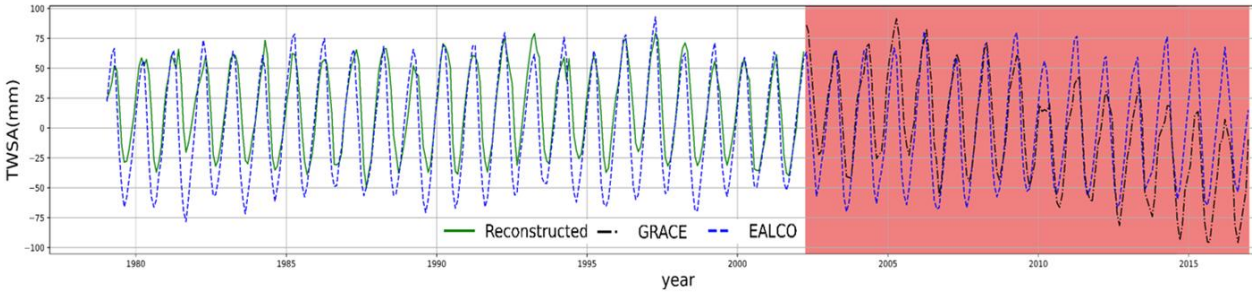

Figure 3-17. DCAE-reconstructed TWSA from 1979 to 2002 (green line in white background), as comparing to GRACE TWSA from 2002 to 2016 (black dash dot line in red background).

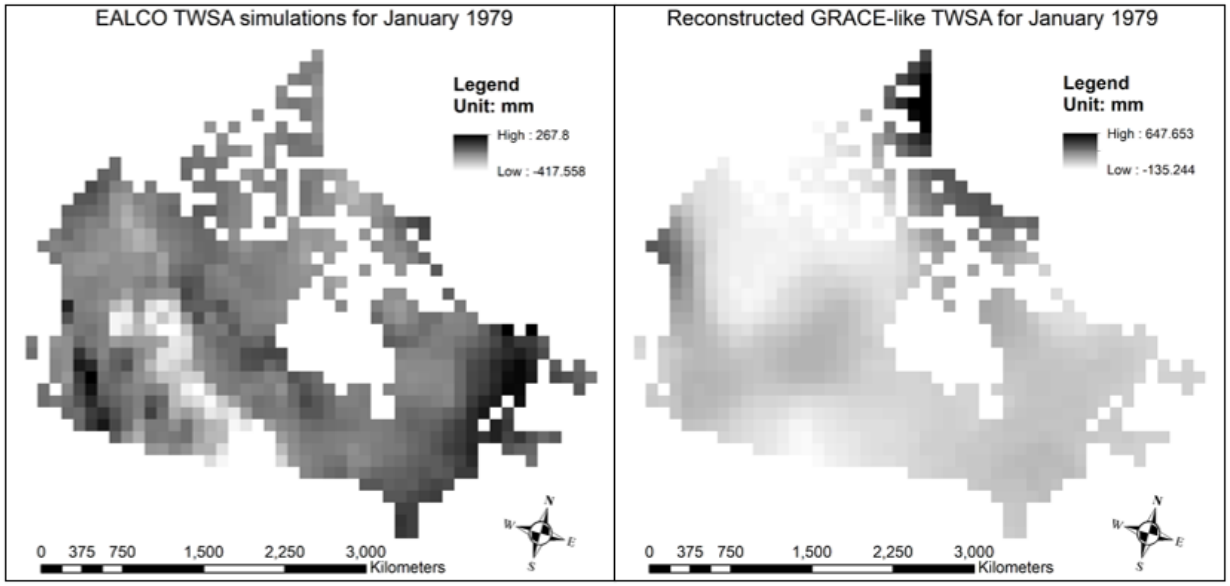

Figure 3-18. DCAE-predicted TWSA for January 1979 (right), comparing to its corresponding EALCO-simulated TWSA (left)

3.4.3. 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. The lack of climatic forcing data 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 (the ground truth). 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.

3.5. 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 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 104.6 mm to 53.11 mm. As for the spatial patter 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.

Chapter 4

Conclusions and Recommendations for Future Research

4.1. Thesis Conclusions

Long-term Terrestrial Water Storage (TWS) monitoring is crucial for assessing the cumulative effect from environmental change and human activities. Physically-based models, such as global hydrological model (GHM) and land surface model (LSM), are increasingly applied for simulating the variations in TWS. However, the applications of simulated TWS are often hindered by the uncertainties in climatic forcing, parameterization, and physical process representations (e.g., missing information for groundwater), especially when a fine spatiotemporal resolution is required. The Gravity Recovery and Climate Experiment (GRACE) dual-satellite was the very first remote sensing mission to monitor temporal changes in TWS, which provides new insights for understanding the global water cycle. However, the available TWS observation data from GRACE satellites is insufficient due to its short mission length, which is not qualified to construct a baseline TWS dataset for calculating the Climate Normal over a period of 30 years. Previous works have applied GRACE-derived TWS to calibrate the TWS simulation results from GHM/LSM by using data assimilation techniques. Due to the large discrepancies and weak matching links between model simulations and satellite observations, conventional statistical methods such as linear regression, normally require abundant knowledge of basin characteristics (i.e., more predictors) that might not be available at all time.

In recent years, several researchers dedicated to developing new data fusion methods based on machine learning (ML) and deep learning (DL), for GRACE and physical model simulations. DL techniques are capable to extract and learn complex relationships between the variables in dynamic spatiotemporal systems. The learned relationships (either linear or nonlinear) can be used to make prediction with computational robustness while prior knowledge of underlying physical processes is not required.

This thesis presented a comparative study of the performance of various DL network architectures in the task of TWS anomalies (TWSA) reconstruction from GRACE observations. In this study, four deep learning-based models to reconstruct GRACE-like TWSA for Canada's Landmass. The first model is based on ConvLSTM networks, which considers the input data as

time series sequences. The ConvLSTM model only takes the GRACE time series to learn the temporal trends of observed TWSA, and then hindcast/forecast TWSA for time periods beyond the GRACE mission. And other three models (Pix2Pix conditional GAN, SEUNet CNN, deep convolutional autoencoder) are based on a hybrid modelling approach aimed at learning the correlations between LSM-simulated TWSA and GRACE-derived TWSA, both in gridded format, by which each input raster (EALCO simulations) is mapped to a corresponding output raster (GRACE observations). Results show that three LSM-based DL models significantly outperform the ConvLSTM model. Hence, the knowledges acquired from decades of hydrological modelling is worthwhile for reconstructing TWA, which should not be neglected. Additionally, the hybrid modelling approach significantly improves the original LSM-simulated TWS by increasing its correlations with GRACE-observed TWS.

To sum up, this study demonstrated the potential of combining physical modelling and DL algorithms to perform TWS reconstruction at a large scale. By comparing the performance of the three LSM-based DL models, the deep convolutional autoencoder (DCAE) provides the most promising results, which improves the original LSM simulations by about 49%. Thus, the DCAE model is applied to construct the historical TWS baseline dataset for the Canadian landmass.

4.2. Recommendations for Future Research

Despite its capability to extend the GRACE TWS dataset with improvements over original EALCO LSM simulation, there are some primary limitations of the proposed hybrid modelling approach, as listed below:

- No use of meteorological and ecological forcing data such as precipitation, air temperature, normalized difference vegetation index. These variables can be used as additional predictors for estimating TWSA. The lack of climatic forcing data also hinders this research from assessing the effects of extreme events (e.g., droughts, floods, rainstorm) on the reconstructed TWSA.
- The hybrid method does not explore the temporal correlations. TWS variations can have time lag effect as well as long-term trends. Even though the ConvLSTM model does not produce satisfactory results in this study, the applicability of time-series modelling in TWS reconstruction is worth further investigations.

- The performances of the DL models can be further assessed on different GRACE solutions, and the uncertainties in GRACE-derived TWS should be taken into consideration. Capturing GRACE observations does not mean getting closer to the 'actual' TWS.
- Due to the absence of GRACE-FO TWS data, this study does not evaluate the applicability of the purposed DCAE model on long-term forecasting of TWS.
- The study is conducted for the entire Canadian landmass as a whole, which neglects the environmental differences between river basins.
- The study does not deseasonalize the input data to assess the interannual variability in the predicted and observed TWSA.

For future research, some sophisticated preprocessing techniques, such as sequencing and detrending, can be employed to investigate the temporal variations of TWS. Meanwhile, additional forcing datasets can be combined with TWS data to enhance the predictability of the DL models. And the performances of the models can be tested against TWS derived from GRACE-FO observations. Furthermore, instead of conducting a nationwide modelling and analysis, a basin-based study can be carried out for improving the models' performances in specific regions such as the Arctic and Southern Ontario.

References

Alley, W., Healy, R., LaBaugh, J., & Reilly, T. (2002). Hydrology - Flow and storage in groundwater systems. *Science*, 296(5575), 1985-1990. doi:10.1126/science.1067123

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 Sensing*, 11(12), 1461. doi:10.3390/rs11121461

Andersen, O. B., Seneviratne, S. I., Hinderer, J., & Viterbo, P. (2005). GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters*, 32(18), 1-4. doi:10.1029/2005GL023574

Apte, A., Jones, C. K. R. T., & Stuart, A. M. (2008). A Bayesian approach to Lagrangian data assimilation. Tellus Series A - *Dynamic Meteorology and Oceanography*, 60(2), 336-347. doi:10.1111/j.1600-0870.2007.00295.x

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. *Journal of Applied Meteorology and Climatology*, 58(6), 1381-1397. doi:10.1175/JAMC-D-18-0252.1

Asanjan, A. A., Yang, T., Hsu, K., Sorooshian, S., Lin, J., & Peng, Q. (2018). Short-Term Precipitation Forecast Based on the PERSIANN System and LSTM Recurrent Neural Networks. *Journal of Geophysical Research-Atmospheres*, 123(22), 12543-12563. doi:10.1029/2018JD028375

Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22, 106-114. doi:10.1016/j.jag.2012.04.012

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

Banerjee, C., Sharma, A., & D, N. K. (2021). Decline in terrestrial water recharge with increasing global temperatures. *Science of the Total Environment*, 764 doi:10.1016/j.scitotenv.2020.142913

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51(7), 4923-4947. doi:10.1002/2015WR017173

Bonsor, H. C., Shamsudduha, M., Marchant, B. P., MacDonald, A. M., & Taylor, R. G. (2018). Seasonal and decadal groundwater changes in African sedimentary aquifers estimated using GRACE products and LSMs. *Remote Sensing*, 10(6) doi:10.3390/rs10060904

Bring, A., Fedorova, I., Dibike, Y., Hinzman, L., Mard, J., Mernild, S. H., . . . Woo, M. (2016). Arctic terrestrial hydrology: A synthesis of processes, regional effects, and research challenges. *Journal of Geophysical Research - Biogeosciences*, 121(3), 621-649. doi:10.1002/2015JG003131

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 Resources Research*, 55(5), 3739-3757. doi:10.1029/2018WR024146

Cai, X., Yang, Z. -., Xia, Y., Huang, M., Wei, H., Leung, L. R., & Ek, M. B. (2014). Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over conus using the NLDAS test bed. *Journal of Geophysical Research*, 119(24), 13,751-13,770. doi:10.1002/2014JD022113

Cao, C., Dragicevic, S., & Li, S. (2019). Land-Use Change Detection with Convolutional Neural Network Methods. *Environments*, 6(2), 25. doi: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. *Science of the Total Environment*, 649, 372-387. doi:10.1016/j.scitotenv.2018.08.352

Chen, L., He, Q., Liu, K., Li, J., & Jing, C. (2019). Downscaling of GRACE-Derived Groundwater Storage Based on the Random Forest Model. *Remote Sensing*, 11(24), 2979. doi:10.3390/rs11242979

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.

IGARSS 2018 - 2018 *IEEE International Geoscience and Remote Sensing Symposium*, 768-771.

Das, M., & Ghosh, S. K. (2017). A Deep-Learning-Based Forecasting Ensemble to Predict Missing Data for Remote Sensing Analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12), 5228-5236. doi:10.1109/JSTARS.2017.2760202

de Bezenac, E., Pajot, A., & Gallinari, P. (2019). Deep learning for physical processes: incorporating prior scientific knowledge. *Journal of Statistical Mechanics-Theory and Experiment*, 2019(12), 124009. doi:10.1088/1742-5468/ab3195

Famiglietti, J. S. (2004). Remote sensing of terrestrial water storage, soil moisture and surface waters. *Geophysical Monograph Series*, 197-207. doi:10.1029/150gm16

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) doi:10.1029/2010GL046442

Fang, K., Pan, M., & Shen, C. (2019). The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(4), 2221-2233. doi:10.1109/TGRS.2018.2872131

Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental US Using a Deep Learning Neural Network. *Geophysical Research Letters*, 44(21), 11030-11039. doi:10.1002/2017GL075619

Fang, Z., Jia, T., Chen, Q., Xu, M., Yuan, X., & Wu, C. (2018). Laser stripe image denoising using convolutional autoencoder. *Results in Physics*, 11, 96-104. doi:10.1016/j.rinp.2018.08.023

Frappart, F., Ramillien, G., & Ronchail, J. (2013). Changes in terrestrial water storage versus rainfall and discharges in the Amazon basin. *International Journal of Climatology*, 33(14), 3029-3046. doi:10.1002/joc.3647

Freeman, B. S., Taylor, G., Gharabaghi, B., & The, J. (2018). Forecasting air quality time series using deep learning. *Journal of the Air & Waste Management Association*, 68(8), 866-886. doi:10.1080/10962247.2018.1459956

Fu, Y., Argus, D. F., & Landerer, F. W. (2015). GPS as an independent measurement to estimate terrestrial water storage variations in Washington and Oregon. *Journal of Geophysical Research: Solid Earth*, 120(1), 552-566. doi:10.1002/2014JB011415

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., & Rodell, M. (2016). Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research*, 52(5), 4164-4183. doi:10.1002/2015WR018417

Girotto, M., De Lannoy, G. J. M., Reichle, R. H., Rodell, M., Draper, C., Bhanja, S. N., & Mukherjee, A. (2017). Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophysical Research Letters*, 44(9), 4107-4115. doi:10.1002/2017GL072994

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, 27, 2672-2680.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463-477. doi:10.5194/hess-18-463-2014

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voss, F., Arnell, N. W., . . . Yeh, P. (2011). Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results. *Journal of Hydrometeorology*, 12(5), 869-884. doi:10.1175/2011JHM1324.1

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 Resources Research*, 54(6), 4040-4058. doi:10.1029/2017WR022238

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507. doi:10.1126/science.1127647

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial

water storage observations. *Water Resources Research*, 48, W07525. doi:10.1029/2011WR011291

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

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2020). Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8), 2011-2023. doi:10.1109/TPAMI.2019.2913372

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:10.1002/2017GL072564

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:10.5194/essd-11-1153-2019

Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 5967-5976. doi:10.1109/CVPR.2017.632

Jiang, D., Wang, J., Huang, Y., Zhou, K., Ding, X., & Fu, J. (2014). The Review of GRACE Data Applications in Terrestrial Hydrology Monitoring. *Advances in Meteorology*, 2014, 725131. doi:10.1155/2014/725131

Jiang, Z., Hsu, Y. -., Yuan, L., & Huang, D. (2021). Monitoring time-varying terrestrial water storage changes using daily GNSS measurements in Yunnan, southwest China. *Remote Sensing of Environment*, 254 doi:10.1016/j.rse.2020.112249

Jing, W., Zhang, P., Zhao, X., Yang, Y., Jiang, H., Xu, J., . . . Li, Y. (2020). Extending GRACE terrestrial water storage anomalies by combining the random forest regression and a spatially moving window structure. *Journal of Hydrology*, 590 doi:10.1016/j.jhydrol.2020.125239

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. *Advances in Water Resources*, 143, 103683. doi: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. *Advances in Water Resources*, 107, 301-316. doi: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 Sensing of Environment*, 221, 741-755. doi:10.1016/j.rse.2018.12.010

Koster, R., Suarez, M., Ducharne, A., Stieglitz, M., & Kumar, P. (2000). A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *Journal of Geophysical Research - Atmospheres*, 105(D20), 24809-24822. doi:10.1029/2000JD900327

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 Resources Research*, 53(11), 8941-8965. doi:10.1002/2017WR020635

Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48, W04531. doi:10.1029/2011WR011453

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., . . . Slater, A. G. (2011). Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. *Journal of Advances in Modeling Earth Systems*, 3, M03001. doi:10.1029/2011MS000045

Lee, H., Durand, M., Jung, H. C., Alsdorf, D., Shum, C. K., & Sheng, Y. (2010). Characterization of surface water storage changes in Arctic lakes using simulated SWOT measurements.

International Journal of Remote Sensing, 31(14), 3931-3953. doi:10.1080/01431161.2010.483494

Lee, H., Jung, H. C., Yuan, T., Beighley, R. E., & Duan, J. (2015). Controls of Terrestrial Water Storage Changes Over the Central Congo Basin Determined by Integrating PALSAR ScanSAR, Envisat Altimetry, and GRACE Data. *Remote Sensing of the Terrestrial Water Cycle*, 206, 117-129.

Lei, S., Shi, Z., & Zou, Z. (2020). Coupled Adversarial Training for Remote Sensing Image Super-Resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5), 3633-3643. doi:10.1109/TGRS.2019.2959020

Lezzaik, K., & Milewski, A. (2018). A quantitative assessment of groundwater resources in the Middle East and North Africa region. *Hydrogeology Journal*, 26(1), 251-266. doi:10.1007/s10040-017-1646-5

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) doi:10.1029/2019WR026551

Lo, M., Famiglietti, J. S., Yeh, P. J. -., & Syed, T. H. (2010). Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data. *Water Resources Research*, 46, W05517. doi:10.1029/2009WR007855

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:10.1016/j.rse.2014.08.006

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., . . . Zhou, X. H. (2012). A framework for benchmarking land models. *Biogeosciences*, 9(10), 3857-3874. doi:10.5194/bg-9-3857-2012

Ma, N., Niu, G. -., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A Systematic Evaluation of Noah-MP in Simulating Land-Atmosphere Energy, Water, and Carbon Exchanges Over

the Continental United States. *Journal of Geophysical Research: Atmospheres*, 122(22), 12,245-12,268. doi:10.1002/2017JD027597

Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., . . . Norby, R. J. (2015). Using ecosystem experiments to improve vegetation models. *Nature Climate Change*, 5(6), 528-534. doi:10.1038/nclimate2621

Milly, P. C., Cazenave, A., Famiglietti, J. S., Gornitz, V., Laval, K., Lettenmaier, D. P., . . . Wilson, C. R. (2010). Terrestrial water-storage contributions to sea-level rise and variability. *Understanding Sea-Level Rise and Variability*, 226-255. doi:10.1002/9781444323276.ch8

Minh, D. H. T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., & Maurel, P. (2018). Deep Recurrent Neural Networks for Winter Vegetation Quality Mapping via Multitemporal SAR Sentinel-1. *IEEE Geoscience and Remote Sensing Letters*, 15(3), 464-468. doi:10.1109/LGRS.2018.2794581

Miro, M. E., & Famiglietti, J. S. (2018). Downscaling GRACE Remote Sensing Datasets to High-Resolution Groundwater Storage Change Maps of California's Central Valley. *Remote Sensing*, 10(1), 143. doi:10.3390/rs10010143

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. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 1268-1271.

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:10.1029/2018WR024638

Mu, B., Li, J., Yuan, S., Luo, X., & Dai, G. (2019). NAO Index Prediction using LSTM and ConvLSTM Networks Coupled with Discrete Wavelet Transform. *2019 International Joint Conference on Neural Networks (IJCNN)*, DOI: 10.1109/IJCNN.2019.8851968.

Mueller, B., Hirschi, M., & Seneviratne, S. I. (2011). New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data. *Hydrological Processes*, 25(7), 996-1008. doi:10.1002/hyp.7652

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. *Journal of Hydrology*, 558, 647-658. doi:10.1016/j.jhydrol.2018.02.005

Janzen, D., Bourgon, J., Brisco, B., Canisius, F., Chen, W., Choma, G., . . . Zhang, Y. (2020). EO baseline data for cumulative EFFECTS, year-end report (FY 2019/20). doi:10.4095/326159

Ni, S., Chen, J., Wilson, C. R., Li, J., Hu, X., & Fu, R. (2018). Global Terrestrial Water Storage Changes and Connections to ENSO Events. *Surveys in Geophysics*, 39(1) doi:10.1007/s10712-017-9421-7

Nie, N., Zhang, W., Zhang, Z., Guo, H., & Ishwaran, N. (2016). Reconstructed Terrestrial Water Storage Change (ΔTWS) from 1948 to 2012 over the Amazon Basin with the Latest GRACE and GLDAS Products. *Water Resources Management*, 30(1), 279-294. doi:10.1007/s11269-015-1161-1

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 Resources Research*, 55(12), 11274-11294. doi:10.1029/2019WR025363

Niu, G. -., Yang, Z. -., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. *Journal of Geophysical Research Atmospheres*, 112(7) doi:10.1029/2006JD007522

Niu, G., Yang, Z., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., . . . Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research - Atmospheres*, 116, D12109. doi:10.1029/2010JD015139

Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving Precipitation Estimation Using Convolutional Neural Network. *Water Resources Research*, 55(3), 2301-2321. doi:10.1029/2018WR024090

Qi, Y., Zhou, Z., Yang, L., Quan, Y., & Miao, Q. (2019). A Decomposition-Ensemble Learning Model Based on LSTM Neural Network for Daily Reservoir Inflow Forecasting. *Water Resources Management*, 33(12), 4123-4139. doi:10.1007/s11269-019-02345-1

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 Resources Research*, 55(1), 175-202. doi:10.1029/2018WR023205

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, & Pa, C. (2017). ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Advances in Neural Information Processing Systems 30 (Nips 2017)*, 30

Rahaman, M. M., Thakur, B., Kalra, A., Li, R., & Maheshwari, P. (2019). Estimating High-Resolution Groundwater Storage from GRACE: A Random Forest Approach. *Environments*, 6(6), 63. doi:10.3390/environments6060063

Rajabi, M. M., Ataie-Ashtiani, B., & Simmons, C. T. (2018). Model-data interaction in groundwater studies: Review of methods, applications and future directions. *Journal of Hydrology*, 567, 457-477. doi:10.1016/j.jhydrol.2018.09.053

Reddy, D. S., & Prasad, P. R. C. (2018). Prediction of vegetation dynamics using NDVI time series data and LSTM. *Modeling Earth Systems and Environment*, 4(1), 409-419. doi:10.1007/s40808-018-0431-3

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:10.1038/s41586-019-0912-1

Robock, A., Vinnikov, K., Srinivasan, G., Entin, J., Hollinger, S., Speranskaya, N., . . . Namkhai, A. (2000). The Global Soil Moisture Data Bank. *Bulletin of the American Meteorological Society*, 81(6), 1281-1299. doi:10.1175/1520-0477(2000)081<1281:TGSMDB>2.3.CO;2

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

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, 460(7258), 999-1002. doi:10.1038/nature08238

Sahoo, A. K., De Lannoy, G. J. M., Reichle, R. H., & Houser, P. R. (2013). Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. *Advances in Water Resources*, 52, 19-33. doi:10.1016/j.advwatres.2012.08.007

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 Resources Research*, 53(5), 3878-3895. doi:10.1002/2016WR019933

Sahour, H., Sultan, M., Vazifedan, M., Abdelmohsen, K., Karki, S., Yellich, J. A., . . . Elbayoumi, T. M. (2020). Statistical Applications to Downscale GRACE-Derived Terrestrial Water Storage Data and to Fill Temporal Gaps. *Remote Sensing*, 12(3), 533. doi:10.3390/rs12030533

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. *Proceedings of the National Academy of Sciences of the United States of America*, 115(6), E1080-E1089. doi:10.1073/pnas.1704665115

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., . . . Kabat, P. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9), 3245-3250. doi:10.1073/pnas.1222460110

Schumacher, M., Forootan, E., van Dijk, A. I. J. M., Mueller Schmied, H., Crosbie, R. S., Kusche, J., & Doell, P. (2018). Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model. *Remote Sensing of Environment*, 204, 212-228. doi:10.1016/j.rse.2017.10.029

Schumacher, M., Kusche, J., & Doll, P. (2016). A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. *Journal of Geodesy*, 90(6), 537-559. doi:10.1007/s00190-016-0892-y

Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, 35(7), 2145-2160. doi:10.1029/1999WR900090

Seyoum, W. M., Kwon, D., & Milewski, A. M. (2019). Downscaling GRACE TWSA Data into High-Resolution Groundwater Level Anomaly Using Machine Learning-Based Models in a Glacial Aquifer System. *Remote Sensing*, 11(7), 824. doi:10.3390/rs11070824

Shamsudduha, M., & Panda, D. K. (2019). Spatio-temporal changes in terrestrial water storage in the Himalayan river basins and risks to water security in the region: A review. *International Journal of Disaster Risk Reduction*, 35 doi:10.1016/j.ijdrr.2019.101068

Shamsudduha, M., Taylor, R. G., Jones, D., Longuevergne, L., Owor, M., & Tindimugaya, C. (2017). Recent changes in terrestrial water storage in the Upper Nile Basin: An evaluation of commonly used gridded GRACE products. *Hydrology and Earth System Sciences*, 21(9), 4533-4549. doi:10.5194/hess-21-4533-2017

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:10.1029/2018WR022643

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 Geoscience and Remote Sensing Magazine*, 3(3), 61-85. doi: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. *Advances in Neural Information Processing Systems 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. *Advances in Neural Information Processing Systems 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 Resources Research*, 54(11), 8931-8951. doi:10.1029/2018WR022785

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, 82(12), 2635-2670. doi:10.2166/wst.2020.369

Śliwińska, J., Wińska, M., & Nastula, J. (2019). Terrestrial water storage variations and their effect on polar motion. *Acta Geophysica*, 67(1), 17-39. doi:10.1007/s11600-018-0227-x

Sohoulande, C. D. D., Martin, J., Szogi, A., & Stone, K. (2020). Climate-driven prediction of land water storage anomalies: An outlook for water resources monitoring across the conterminous United States. *Journal of Hydrology*, 588 doi:10.1016/j.jhydrol.2020.125053

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 doi:10.1016/j.earscirev.2020.103487

Subramanian, A. C., Hoteit, I., Cornuelle, B., Miller, A. J., & Song, H. (2012). Linear versus Nonlinear Filtering with Scale-Selective Corrections for Balanced Dynamics in a Simple Atmospheric Model. *Journal of the Atmospheric Sciences*, 69(11), 3405-3419. doi:10.1175/JAS-D-11-0332.1

Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2021). Reconstruction of GRACE Total Water Storage Through Automated Machine Learning. *Water Resources Research*, 57(2), e2020WR028666. doi:https://doi.org/10.1029/2020WR028666

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining Physically Based Modeling and Deep Learning for Fusing GRACE Satellite Data: Can We Learn From Mismatch? *Water Resources Research*, 55(2), 1179-1195. doi: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) doi:10.1029/2019WR026250

Swenson, S., & Wahr, J. (2002). Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research - Solid Earth*, 107(B9), 2193. doi:10.1029/2001JB000576

Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters*, 33(8), L08402. doi:10.1029/2005GL025285

Swenson, S., Yeh, P. J. -., Wahr, J., & Famiglietti, J. (2006). A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois. *Geophysical Research Letters*, 33(16) doi:10.1029/2006GL026962

Tang, Q., Gao, H., Yeh, P., Oki, T., Su, F., & Lettenmaier, D. P. (2010). Dynamics of Terrestrial Water Storage Change from Satellite and Surface Observations and Modeling. *Journal of Hydrometeorology*, 11(1), 156-170. doi:10.1175/2009JHM1152.1

Tang, Y., Hooshyar, M., Zhu, T., Ringler, C., Sun, A. Y., Long, D., & Wang, D. (2017). Reconstructing annual groundwater storage changes in a large-scale irrigation region using GRACE data and Budyko model. *Journal of Hydrology*, 551, 397-406. doi:10.1016/j.jhydrol.2017.06.021

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

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 Resources Research*, 54(5), 3290-3312. doi:10.1029/2017WR021674

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., & Ganguly, A. R. (2017). DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3097983.3098004

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

Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, 49(2), 904-914. doi:10.1002/wrcr.20078

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., & Schoups, G. (2013). Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water Resources*, 51, 457-478. doi:10.1016/j.advwatres.2012.04.002

Wang, S. (2008). Simulation of evapotranspiration and its response to plant water and CO2 transfer dynamics. *Journal of Hydrometeorology*, 9(3), 426-443. doi:10.1175/2007JHM918.1

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. *Journal of Hydrology*, 512, 1-15. doi:10.1016/j.jhydrol.2014.02.058

Wang, S., & Li, J. (2016). Terrestrial Water Storage Climatology for Canada from GRACE Satellite Observations in 2002-2014. *Canadian Journal of Remote Sensing*, 42(3), 190-202. doi:10.1080/07038992.2016.1171132

Wang, S., McKenney, D. W., Shang, J., & Li, J. (2014b). A national-scale assessment of long-term water budget closures for Canada's watersheds. *Journal of Geophysical Research*, 119(14), 8712-8725. doi:10.1002/2014JD021951

Wang, S. (2005). Dynamics of surface albedo of a boreal forest and its simulation. *Ecological Modelling*, 183(4), 477-494. doi:10.1016/j.ecolmodel.2004.10.001

Wang, S. (2008). Simulation of evapotranspiration and its response to plant water and CO2 transfer dynamics. *Journal of Hydrometeorology*, 9(3), 426-443. doi:10.1175/2007JHM918.1

Watkins, M. M., Wiese, D. N., Yuan, D., 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:10.1002/2014JB011547

Widlowski, J. -., Pinty, B., Clerici, M., Dai, Y., De Kauwe, M., de Ridder, K., . . . Yuan, H. (2011). RAMI4PILPS: An intercomparison of formulations for the partitioning of solar radiation in land surface models. *Journal of Geophysical Research - Biogeosciences*, 116, G02019. doi:10.1029/2010JG001511

Wiese, D. N., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research*, 52(9), 7490-7502. doi:10.1002/2016WR019344

Wu, P., Yin, Z., Yang, H., Wu, Y., & Ma, X. (2019). Reconstructing Geostationary Satellite Land Surface Temperature Imagery Based on a Multiscale Feature Connected Convolutional Neural Network. *Remote Sensing*, 11(3), 300. doi:10.3390/rs11030300

Xiao, C., Chen, N., Hu, C., Wang, K., Xu, Z., Cai, Y., . . . Gong, J. (2019). A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, 120, 104502. doi:10.1016/j.envsoft.2019.104502

Yang Peng, Xia Jun, Zhan Chesheng, Qiao Yunfeng, & Wang Yueling. (2017). Monitoring the spatio-temporal changes of terrestrial water storage using GRACE data in the Tarim River basin between 2002 and 2015. *Science of the Total Environment*, 595, 218-228. doi:10.1016/j.scitotenv.2017.03.268

Yang, P., Xia, J., Zhan, C., & Wang, T. (2018). Reconstruction of terrestrial water storage anomalies in Northwest China during 1948–2002 using GRACE and GLDAS products. *Hydrology Research*, 49(5), 1594-1607. doi:10.2166/nh.2018.074

Yang, T., Wang, C., Chen, Y., Chen, X., & Yu, Z. (2015). Climate change and water storage variability over an arid endorheic region. *Journal of Hydrology*, 529(P1), 330-339. doi:10.1016/j.jhydrol.2015.07.051

Yang, T., Wang, C., Yu, Z., & Xu, F. (2013). Characterization of spatio-temporal patterns for various GRACE- and GLDAS-born estimates for changes of global terrestrial water storage. *Global and Planetary Change*, 109, 30-37. doi:10.1016/j.gloplacha.2013.07.005

Yang, Y., Dong, J., Sun, X., Lima, E., Mu, Q., & Wang, X. (2018). A CFCC-LSTM Model for Sea Surface Temperature Prediction. *IEEE Geoscience and Remote Sensing Letters*, 15(2), 207-211. doi:10.1109/LGRS.2017.2780843

Yin, W., Hu, L., Han, S., Zhang, M., & Teng, Y. (2019). Reconstructing terrestrial water STORAGE variations from 1980 to 2015 in the Beishan area of China. *Geofluids*, 2019, 1-13. doi:10.1155/2019/3874742

Yin, W., Hu, L., Zhang, M., Wang, J., & Han, S. (2018). Statistical Downscaling of GRACE-Derived Groundwater Storage Using ET Data in the North China Plain. *Journal of Geophysical Research - Atmospheres*, 123(11), 5973-5987. doi:10.1029/2017JD027468

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, 4559–4566.

Yuan, X., Chen, C., Lei, X., Yuan, Y., & Adnan, R. M. (2018). Monthly runoff forecasting based on LSTM-ALO model. *Stochastic Environmental Research and Risk Assessment*, 32(8), 2199-2212. doi:10.1007/s00477-018-1560-y

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., . . . Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241, 111716. doi:10.1016/j.rse.2020.111716

Yuan, T., Lee, H., Jung, H. C., Aierken, A., Beighley, E., Alsdorf, D. E., . . . Kim, D. (2017). Absolute water storages in the Congo River floodplains from integration of InSAR and satellite radar altimetry. *Remote Sensing of Environment*, 201, 57-72. doi:10.1016/j.rse.2017.09.003

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:10.1175/2007JHM951.1

Zhang, D., Zhang, Q., Werner, A. D., & Liu, X. (2016). GRACE-Based Hydrological Drought Evaluation of the Yangtze River Basin, China. *Journal of Hydrometeorology*, 17(3), 811-828. doi:10.1175/JHM-D-15-0084.1

Zhang, L., Dobslaw, H., Stacke, T., Guentner, A., Dill, R., & Thomas, M. (2017). Validation of terrestrial water storage variations as simulated by different global numerical models with GRACE satellite observations. *Hydrology and Earth System Sciences*, 21(2), 821-837. doi:10.5194/hess-21-821-2017

Zhang, Q., Yuan, Q., Zeng, C., Li, X., & Wei, Y. (2018). Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial-Temporal-Spectral Deep Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8), 4274-4288. doi:10.1109/TGRS.2018.2810208

Zhang, R., Yu, L., Tian, S., & Lv, Y. (2019). Unsupervised remote sensing image segmentation based on a dual autoencoder. *Journal of Applied Remote Sensing*, 13(3), 038501. doi:10.1117/1.JRS.13.038501

Zheng, Y., Yi, X., Li, M., Li, R., Shan, Z., Chang, E., … Li, T. (2015). Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. doi:10.1145/2783258.2788573

Zhong, D., Wang, S., & Li, J. (2021). A Self-Calibration Variance-Component Model for Spatial Downscaling of GRACE Observations Using Land Surface Model Outputs. *Water Resources Research*, 57(1), e2020WR028944. https://doi.org/10.1029/2020WR028944

Zhou, Z., Shi, L., & Zha, Y. (2020). Seeing macro-dispersivity from hydraulic conductivity field with convolutional neural network. *Advances in Water Resources*, 138, 103545. doi:10.1016/j.advwatres.2020.103545

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. doi:10.3390/en11040705