Generating Long-term GRACE-like Total Water Storage Change Products using Conditional Generative Adversarial Networks

Jielong Wang[1,2],Yunzhong Shen[1],Joseph Awange[2]

[1]College of Surveying and Geo-informatics, Tongji University,China - (wangjielong, yzshen)@tongji.edu.cn
[2] School of Earth and Planetary Sciences, Spatial Sciences Discipline, Curtin University, Australia
- (jielong.wang, j.awange)@curtin.edu.au

Keywords: GRACE, Total Water Storage, Deep Learning, Generative Adversarial Networks, Convolutional Neural Networks.

Abstract

Since 2002, the Gravity Recovery And Climate Experiment (GRACE) and its Follow-On (GRACE-FO, hereafter GRACE) missions have offered global observations of total water storage (TWS). However, the relatively short record of GRACE data poses a significant challenge for researchers to investigate the full range and long-term variability in TWS. In this study, we present RecGAN, a novel Conditional Generative Adversarial Network (CGAN) comprising a RecNet generator and pixel discriminator. Our approach aims to generate long-term GRACE-like TWS observations by calibrating the WaterGAP Global Hydrology Model (WGHM). The generator is trained to produce observations conforming to the distribution of GRACE data, while the discriminator is trained to assess whether each generated pixel resembles GRACE data. Our results show that RecGAN effectively enhances the consistency between GRACE observations and WGHM-derived TWS changes, achieving improved correlation coefficients, Nash-Sutcliffe Efficiency, and Normalized Root-Mean-Square Error. In addition, RecGAN is robust to different GRACE mascon data, crop sizes used during the training period, and hydrological models targeted for calibration. This study illustrates a promising application of employing CGANs to fine-tune the WGHM output to match GRACE observations. This approach enables the generation of long-term TWS change datasets, which are invaluable for evaluating long-term water storage fluctuations, allocating water resources, and forecasting future hydrological extremes.

1. Introduction

Since 2002, the Gravity Recovery And Climate Experiment (GRACE) and its Follow-On (GRACE-FO, hereafter referred to as GRACE unless explicitly stated otherwise) missions have provided valuable observations for monitoring global as well as regional total water storage (TWS) changes (Awange et al., 2016). TWS changes reflect surface water, snow, ice, soil moisture, and groundwater changes, influenced by both anthropogenic activities and climate variability (Rodell et al., 2018; Rodell and Li, 2023; Rodell and Reager, 2023). GRACE-derived TWS data have been extensively used to explore climate changes and variability (e.g., Awange et al., 2013), estimate hydrological flux, and detect trends of human-induced groundwater depletion (Agutu et al., 2019; Tapley et al., 2019). However, the limited GRACE data, spanning roughly 20 years, has hampered its use in investigating the complete range and long-term variability in TWS, impeding our ability to evaluate long-term water storage changes needed to enhance water resource management practices.

Before the GRACE era, TWS retrievals relied on hydrological models or in-situ water level measurements (Huang et al., 2013). Nevertheless, these methods have inherent limitations. For instance, hydrological models generally inadequately represent certain water components (e.g., groundwater) due to data availability or uncertainty (Scanlon et al., 2016), while in-situ observations lack the same spatial resolution as GRACE (Li et al., 2021, 2020). Therefore, numerous studies have sought to reconstruct historical TWS data by learning its empirical relationship with various explanatory variables (e.g., Humphrey and Gudmundsson, 2019; Li et al., 2021; Wang et al., 2023; Yin et al., 2023). Wang et al. (2023) proposed a deep learning model called RecNet to reconstruct past centenary TWS changes over the Yangtze River Basin, China. Using a set of machine learning models (e.g., random forest and neural networks), Yin et al. (2023) prolonged GRACE-derived TWS changes to 1940. Jing et al. (2020) reconstructed the Nile River Basin's TWS data from 1979 to 2013 using a bias-corrected approach. Specifically, they applied two ensemble learning algorithms, namely random forest and the extreme gradient boost, to align the output of the land surface model with GRACE.

WaterGAP Global Hydroloy Model (WGHM) is a global hydrological model that quantifies human use of groundwater and surface water, as well as water flows and storage (Müller Schmied et al., 2021). Since 1996, it has been constantly improved and showed improved consistency with GRACE-derived observations (Müller Schmied et al., 2021). However, a relatively large discrepancy still exists because of uncertain climatic forcings and unconsidered hydrological processes in WGHM's simulation (Scanlon et al., 2016).

In this study, we propose a conditional generative adversarial network (CGAN) to calibrate WGHM-derived TWS changes based on GRACE observations. GANs, initially developed by Goodfellow et al. (2014), consist of two neural networks: the Generator and Discriminator. The Generator's objective is to learn to generate a fake sample distribution to deceive the Discriminator, while the Discriminator aims to learn to distinguish between real and fake distribution generated by the Generator. Since their inception, GANs have achieved impressive performance across various tasks, including image processing, video generation and prediction, and spatiotemporal series forecasts (Dash et al., 2024). GAN represents unsupervised learning, where the model learns to generate data without explicit supervision. In contrast, CGAN is supervised learning where the model is conditioned on additional information, typically labels or context, to generate more targeted outputs (Mirza and Osindero, 2014). Therefore, we design a novel CGAN architecture, expecting that our Generator can generate long-term GRACE-like TWS data by learning the relationship between GRACE-derived TWS changes and those simulated by WGHM.

2. Datasets

2.1 GRACE and GRACE-FO Data

GRACE is a twin satellite mission jointly implemented by US NASA and the German Aerospace Centre (Chen, 2020). Since being launched in 2002, GRACE has enabled unprecedented observations of Earth's gravity field via tracking the inter-satellite range and range rate. At a monthly scale, changes in this field primarily reflect large-scale mass transport and redistribution within the atmosphere, hydrosphere, ocean, cryosphere, and solid Earth (Rodell and Reager, 2023). GRACE was decommissioned in November 2017, while GRACE-FO was launched in 2018 to continue the scientific tasks of the GRACE mission.

This study employs the gridded monthly global water storage/height anomalies from the Jet Propulsion Laboratory (JPL) Mascons with the Coastal Resolution Improvement (CRI) filter. Compared to conventional spherical-harmonic solutions, this dataset has a higher signal-to-noise ratio due to a priori information derived from near-global geophysical models (NASA/JPL, 2019). The GRACE data from April 2002 to June 2017 and GRACE-FO data from June 2018 to December 2019 are used here, both of which are presented as anomalies relative to the 2004-2009 time-mean baseline. Except for the data gap between GRACE and GRACE-FO, missing months are filled using linear interpolation using the mean values determined in the months before and after the missing month. We also test the robustness of the choice of mascon datasets, including CSR (Save et al., 2016) and GSFC mascon solutions (Loomis et al., 2021).

2.2 WGHM Data

Based on the intricate water balance theory, WGHM is one of the most widely utilized models for assessing both human-induced and natural water storage variability. It simulates daily water flows and ten water components, such as lakes, reservoirs, and groundwater (Müller Schmied et al., 2021). Nevertheless, due to errors in climate forcings and incomplete realism of process representations, WGHM-derived TWS changes are considered less reliable than those derived from GRACE. For this study, the simulated TWS data from WaterGAP 2.2d for 1901-2019 are used here, with a spatial resolution of 0.5 degrees. We also use the TWS data from the Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004) to check the robustness of our main results.

2.3 Precipitation and Temperature Data

While precipitation and temperature have been considered the inputs in WGHM's simulation, we utilize them as the context to help train our CGAN. Monthly precipitation and temperature data from 1901 to 2019 are obtained from Climatic Research Unit Time Series (CRU TS v. 4.07) high-resolution (0.5 degrees or finer) gridded datasets. This dataset is created by interpolating monthly climate anomalies from extensive networks of weather station observations (Harris et al., 2020).

3. Methods

3.1 RecGAN

We introduce a deep generative model called RecGAN (Figure 1) to calibrate WGHM-derived TWS changes. The generator G of RecGAN is RecNet developed by Wang et al. (2023), while the discriminator D is a pixel discriminator proposed by Isola et al. (2017). The goal of G is to learn a mapping from observed data x (WGHM data) and context vector z (precipitation and temperature) to the target y (GRACE data), $G : \{x, z\} \rightarrow y$. The generator G is trained to produce outputs (TWS data) that cannot be distinguished from “real” images (GRACE data) by an adversarially trained discriminator, D, which is trained to recognize whether the data is from G or from GRACE. The objective of RecGAN can be expressed as

$$\begin{aligned}\mathcal{L}_{RecGAN} =& \mathbb{E}_{x,y}[logD(x,y)]+\\ & \mathbb{E}_{x,z}[log(1-D(x,G(x,z)))],\end{aligned} \tag{1}$$

where G tries to minimize this objective function, while D tries to maximize it. RecNet has demonstrated its reconstructive capability in the Yangtze River Basin by Wang et al. (2023), whereas the pixel-discriminator is used to determine whether each pixel generated by G is ‘similar’ to GRACE.

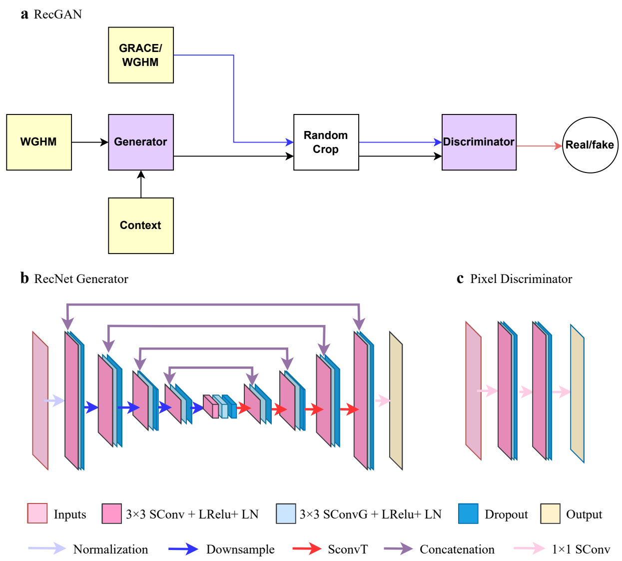

Figure 1. (a) The architecture of RecGAN consists of (b) RecNet generator and (c) pixel discriminator. The arrows indicate the different operations performed. SConv for spectral normalization convolution; SConvG, spectral normalization grouped convolution; LN, layer normalization; SConvT, spectral normalization transposed convolution.

3.2 Training Details

Considering the water accumulation time-lag effect, the precipitation and temperature data from the current month and the previous 11 months are used as the context variables, and the WGHM data from the current month is used as the observed variable. Both the input and target data are resampled into 1-degree grid, and during the training and validation periods, a random crop size of 64x64 is employed. These choices allow RecGAN to fit within memory. Furthermore, the Greenland and

Antarctic regions are excluded due to the relatively high uncertainty associated with ice sheets and glaciers.

Training GANs is widely acknowledged as challenging due to their adversarial nature (Ravuri et al., 2021). Following the study of Zhang et al. (2019), we use spectral normalization in the generator and discriminator. In addition, the two-timescale update rule is applied specifically to address slow learning in regularized discriminators. We adopt the Wasserstein distance with gradient penalty as the loss function (Liu et al., 2019). Instead of batch normalization use in RecNet (Wang et al., 2023), layer normalization is used in RecGAN.

3.3 Evaluation Metrics

Since our primary objective is generating the historical TWS changes by adjusting WGHM TWS data, we attempt to test RecGAN's reconstructive capability. The latter part of GRACE data from November 2004 to December 2019 is used as the training (70%) and validation sets (15%), whereas the earlier part (15%) from April 2002 to October 2004 is left as the testing set. The performance of RecGAN is evaluated using the correlation coefficient (CC), normalized root-mean-square error (NRMSE), and Nash-Sutcliffe efficiency (NSE). CC measures how future outcomes are likely to be predicted by the model. NSE is used to quantify the predictive skill of the model relative to the mean of observations, with a negative NSE value indicating the mean observed value is a better predictor than the forecasting model, while NRMSE describes the global fitness of the predictive model (Wang et al., 2024). These metrics are calculated as follows:

$$\begin{aligned} CC &= \frac{\sum_{t=1}^{n}(o_t-\bar{o})(p_t-\bar{p})}{\sqrt{\sum_{t=1}^{n}(o_t-\bar{o})^2(p_t-\bar{p})^2}}; \\ NSE &= 1-\frac{\sum_{t=1}^{n}(o_t-p_t)^2}{\sum_{t=1}^{n}(p_t-\bar{p})^2}; \\ NRMSE &= \sqrt{\frac{\sum_{t=1}^{n}(o_t-p_t)^2}{n}}, \end{aligned} \tag{2}$$

where o and p are the observed and predicted values, respectively; the overbar denotes mean values; n is the number of target data for testing.

4. Results

4.1 Comparing RecGAN with WGHM

To validate whether the reconstruction generated by RecGAN exhibits improved consistency with GRACE observations when compared to WGHM, we compute the CC and NSE values between GRACE observations and RecGAN's reconstruction or WGHM during the testing period (i.e., from April 2002 to October 2004). This method treats the GRACE data as the ground truth.

As shown in Figure 2a, the original output of WGHM data displays satisfactory CC values with GRACE observations, while relatively low CC values are generally observed in arid regions (e.g., Africa and Australia). This is likely attributed to limited precipitation, leading to weak TWS changes signal. However, after calibrating WGHM-derived data, RecGAN's reconstruction reveals improved CC values compared to WGHM (Figure 2d). This is also evident in Figure 2j, which shows the cumulative distributions of the CC values obtained by RecGAN outperforms WGHM.

In regard to the NSE values, the reconstruction from RecGAN also enhances the consistency between WGHM and GRACE observations, especially evident in the western hemisphere (Figures 2b and 2e). Notably, RecGAN achieves positive NSE values across most Canadian regions, where negative NSE values are observed in the comparison between WGHM and GRACE observations. The cumulative distributions of the NSE values further indicate RecGAN's superior performance (Figure 2k).

As for the NRMSE values, significant discrepancies between WGHM and GRACE-derived TWS changes are observed in the northern Africa and northwestern China (Figure 2c). Nevertheless, RecGAN effectively reduces these discrepancies (Figure 2f). This improvement is further highlighted by the cumulative distribution of the NRMSE values (Figure 2l), indicating lower NRMSE in RecGAN compared to WGHM.

We also compare RecGAN's performance with RecNet, a deep learning model developed by Wang et al. (2023). Figures 2g-i indicate generally consistent performance between RecGAN and RecNet. Specifically, RecGAN and RecNet demonstrate similar CC values with GRACE observations (Figure 2j). However, RecGAN yields relatively low NRMSE and high NSE values compared to RecNet (Figures 2k-l). This is probably attributed to the adversarial structure of RecGAN, given the identical RecNet structure and parameters applied in these two methods.

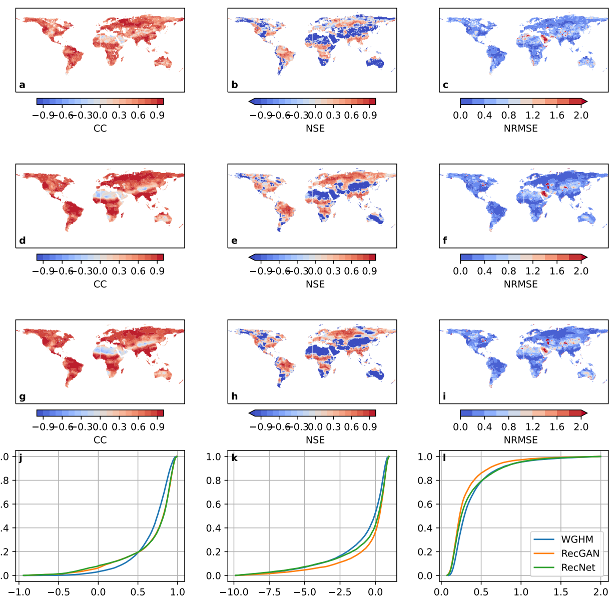

Figure 2. Evaluation of the reconstruction during the testing period 2002-2004. Correlation coefficients (CC), Nash-Sutcliffe Efficiencies (NSE), and Normalized root-mean-square error (NRMSE) of WGHM (a-c), RecGAN's reconstruction (d-f), and RecNet's reconstruction when compared to GRACE observations. (j-l) Empirical cumulative distribution of these performance metrics.

4.2 TWS Change Series over Example Basins

To illustrate the potential improvement brought by RecGAN on the initial WGHM output, we select four river basins: Mackenzie, Gobi Interior, Nelson, and Nile, as representative examples. It is evident that RecGAN can improve the consistency between GRACE observations and WGHM (Figure 3). Specifically, the

CC values are increased by 13.8%, 162.5%, 15.1%, and 10.3% in the Mackenzie, Gobi Interior, Nelson, and Nile basins, respectively. Similar increases in the NSE values are also observed. For example, the NSE value between GRACE observations and WGHM increases from 0.17 to 0.46 in the Nelson River basin (Figure 3c) while increasing from 0.36 to 0.54 in the Nile River basin (Figure 3d). In addition, the NRMSE values in these basins decrease to varying extents.

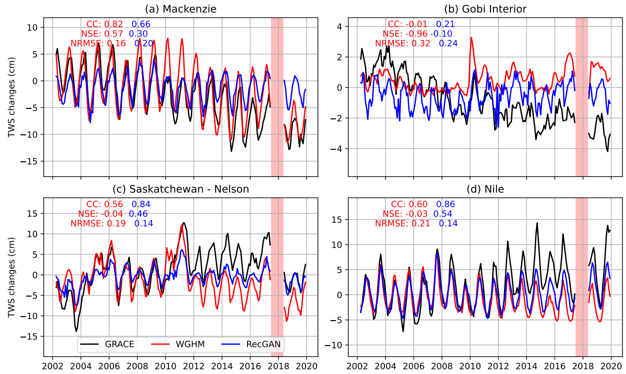

Figure 3. Time series of TWS changes derived from GRACE, WGHM, and RecGAN over the Mackenzie, Gobi Interior, Nelson, and Nile basins. The metrics in red and blue denote the performance of WGHM and RecGAN, respectively.

4.3 Robustness Check

Our main results are derived using the CRU precipitation and temperature data as the inputs and the JPL mascons data as the target. In this section, we test the robustness of RecGAN's performance to different mascon solutions, hydrological models, and crop sizes.

As shown in the first and second rows of Figure 4, RecGANs trained with crop sizes of 32x32 and 96x96 display generally consistent results, indicating RecGAN's robustness to crop size variations. We also train RecGAN using mascon data from CSR and GSFC. The results, depicted in the third and fourth rows of Figure 4, reveal RecGAN's robustness to distinct mascon solutions.

Finally, RecGAN is trained to calibrate GLDAS-derived TWS changes instead of those from WGHM. Similar performance is observed (the last row of Figure 4) compared to models trained using WGHM-derived TWS changes (Figure 2), suggesting that RecGAN has the potential to improve alternative hydrological models.

5. Discussions

GANs facilitate a wide variety of applications, including image generation and manipulation (Dash et al., 2024). They embody adversarial learning for image-to-image translation, with the goal of translating an input image from one domain (e.g., WGHM) to another domain (e.g., GRACE), given input-output image pairs as training data (Wang et al., 2018). This concept aligns seamlessly with our primary objective: transforming WGHM-derived TWS changes into GRACE-like observations. Therefore, we design RecGAN to fulfill this task. Our results have demonstrated RecGAN's ability to generate GRACE-like observations based on WGHM outputs. Nevertheless, several key issues need to be highlighted.

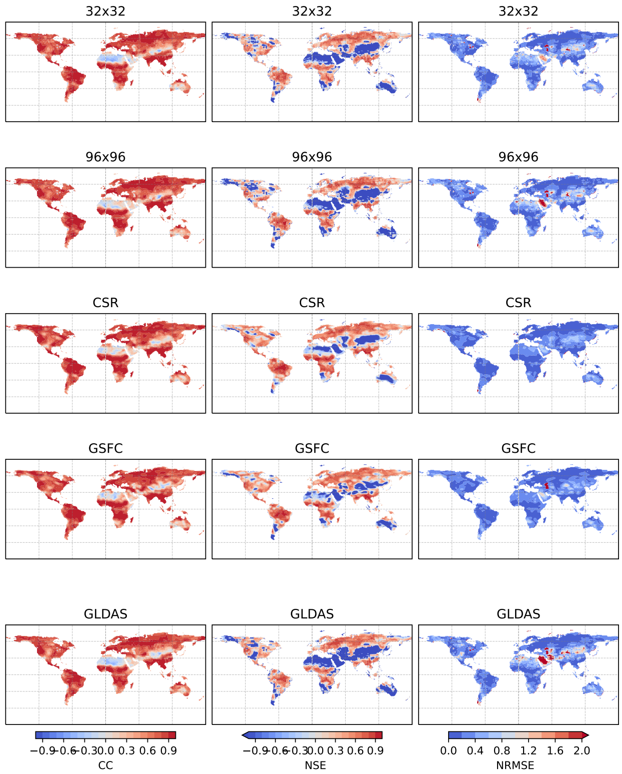

Figure 4. Robustness of main results. The CC, NSE, and NRMSE values during the testing period between GRACE observations and RecGAN trained using crop sizes of 32x32, 96x96, CSR mascon data, GSFC mascon data, and GLDAS data.

First, it's crucial to acknowledge that RecGAN's performance is undoubtedly influenced by the size of the dataset used for training. As of the time of writing, there are about 200 monthly GRACE observations available. It is anticipated that better RecGAN performance can be achieved with an increase in the quantity of GRACE observations. To mitigate this issue, data augmentation techniques can be applied. For example, Wang et al. (2024) proposed a noise-augmented technique to augment training datasets, leading to improved deep learning performance. Furthermore, GANs have proven to be effective tools for generating synthetic ”training data” (Sandfort et al., 2019). These techniques will be explored in our future works.

While RecGAN shows potential for improving consistency between WGHM and GRACE observations, it exhibits relatively poor performance in arid regions, such as the interior of Africa and Australia (Figure 2). This is likely due to weak TWS signals resulting from limited precipitation and high evapotranspiration. Moreover, TWS changes in arid regions (e.g., Gobi Interior) may have many small and unstable trends caused by irregular precipitation, leading to high uncertainties in these areas (Figure 3b). It is important to note that these findings assume that GRACE observations serve as the ground truth. Large lakes (e.g., Lake Victoria in Africa), reservoirs, and glaciers may also impact the calibration, but comprehensively studying their influence requires additional effort and exploration.

We chose to calibrate WGHM outputs because this model incorporates simulations of human-induced water use and groundwater dynamics. Additionally, WGHM generally exhibits a high correlation with GRACE observations. We utilize a linear least squares method to decompose TWS changes into trend, annual, and semi-annual components. We observe moderate discrepancies in the trends between WGHM and GRACE (Figure 5), indicating challenges in simulating this component, which is likely attributed to human activities, such as groundwater abstraction and dam construction (Rodell et al., 2018). Human-induced TWS changes have often been under-reconstructed or unreconstructed in previous studies due to the absence of long-term observations specifically reflecting human activities (Wang et al., 2023). Potential improvements in RecGAN performance can be anticipated if appropriate human factors or their proxies are incorporated as context variables.

Finally, this work demonstrates a potential application of using RecGAN to generate GRACE-like observations, indicating the efficacy of deep generative models in addressing remote sensing challenges. While acknowledging the limitations inherent in this study, we anticipate exploring avenues for enhancing RecGAN's performance in our future research endeavors.

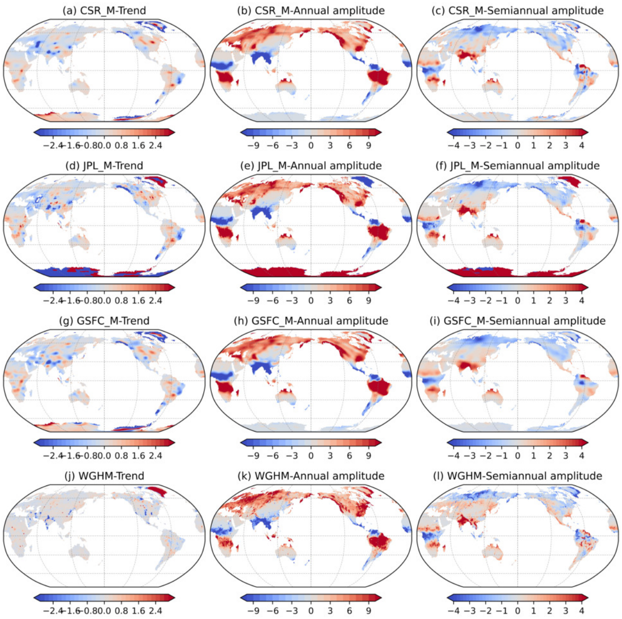

Figure 5. Trend, annual amplitude, and semiannual amplitude derived from CSR Mascons, JPL Mascons, GSFC Mascons, and WGHM.

6. Conclusion

In this study, we proposed a deep generative model called RecGAN, which consists of a RecNet generator and pixel discriminator, to generate GRACE-like TWS observations by calibrating those derived from WGHM. Compared to the initial WGHM output, RecGAN effectively improves consistency with GRACE observations, achieving higher CC and NSE values, along with lower NRMSE values. Our main results are robust to different mascon solutions, crop sizes, and hydrological models.

Deep generative models, such as GANs, have undergone rapid evolution in recent years. Their intrinsic adversarial generation capabilities have enabled the resolution of numerous problems previously deemed unsolvable, much like the study presented here. We anticipate further refining and leveraging this methodology to address a broader array of environmental-related scientific challenges.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 42274005) and the China Scholarship Council (Grant No. 202306260241). Curtin University, School of Earth and Planetary Sciences is also appreciated for hosting Jielong Wang during the period of this study.

References

Agutu, N., Awange, J., Ndehedehe, C., Kirimi, F., Kuhn, M., 2019. GRACE-derived groundwater changes over Greater Horn of Africa: Temporal variability and the potential for irrigated agriculture. *Science of The Total Environment*, 693, 133467. doi.org/10.1016/j.scitotenv.2019.07.273.

Awange, J., Khandu, Schumacher, M., Forootan, E., Heck, B., 2016. Exploring hydro-meteorological drought patterns over the Greater Horn of Africa (1979–2014) using remote sensing and reanalysis products. *Advances in Water Resources*, 94, 45–59. doi.org/10.1016/j.advwatres.2016.04.005.

Awange, J. L., Anyah, R., Agola, N., Forootan, E., Omondi, P., 2013. Potential impacts of climate and environmental change on the stored water of Lake Victoria Basin and economic implications: Impacts of Climate Change on Waters of Lake Victoria. *Water Resources Research*, 49(12), 8160–8173. doi.org/10.1002/2013WR014350.

Chen, J. L., 2020. *Tracking Earth's Water in Motion from Satellite Gravity Observations*. Springer International Publishing, Cham, 1–7.

Dash, A., Ye, J., Wang, G., 2024. A Review of Generative Adversarial Networks (GANs) and Its Applications in a Wide Variety of Disciplines: From Medical to Remote Sensing. *IEEE Access*, 12, 18330–18357. doi.org/10.1109/ACCESS.2023.3346273.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27. doi.org/10.1007/978-3-658-40442-0$_9$.

Harris, I., Osborn, T. J., Jones, P., Lister, D., 2020. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific Data*, 7(1), 109. doi.org/10.1038/s41597-020-0453-3.

Huang, Y., Salama, M. S., Krol, M. S., Van Der Velde, R., Hoekstra, A. Y., Zhou, Y., Su, Z., 2013. Analysis of long-term terrestrial water storage variations in the Yangtze River basin. *Hydrology and Earth System Sciences*, 17(5), 1985–2000. doi.org/10.5194/hess-17-1985-2013.

Humphrey, V., Gudmundsson, L., 2019. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. doi.org/10.5194/essd-11-1153-2019.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Honolulu, HI, 5967–5976.

Jing, W., Di, L., Zhao, X., Yao, L., Xia, X., Liu, Y., Yang, J., Li, 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.org/10.1016/j.advwatres.2020.103683.

Li, F., Kusche, J., Chao, N., Wang, Z., Löcher, A., 2021. Long-Term (1979-Present) Total Water Storage Anomalies Over the Global Land Derived by Reconstructing GRACE Data. *Geophysical Research Letters*, 48(8), e2021GL093492. doi.org/10.1029/2021GL093492.

Li, F., Kusche, J., Rietbroek, R., Wang, Z., Forootan, E., Schulze, K., Lück, C., 2020. Comparison of Data-Driven Techniques to Reconstruct (1992–2002) and Predict (2017–2018) GRACE-Like Gridded Total Water Storage Changes Using Climate Inputs. *Water Resources Research*, 56(5), e2019WR026551. doi.org/10.1029/2019WR026551.

Liu, H., Gu, X., Samaras, D., 2019. Wasserstein GAN With Quadratic Transport Cost. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Seoul, Korea (South), 4831–4840.

Loomis, B. D., Felikson, D., Sabaka, T. J., Medley, B., 2021. High-Spatial-Resolution Mass Rates From GRACE and GRACE-FO: Global and Ice Sheet Analyses. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB023024. doi.org/10.1029/2021JB023024.

Mirza, M., Osindero, S., 2014. Conditional Generative Adversarial Nets. arXiv:1411.1784 [cs, stat].

Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., Döll, P., 2021. The global water resources and use model WaterGAP v2.2d: model description and evaluation. *Geoscientific Model Development*, 14(2), 1037–1079. doi.org/10.5194/gmd-14-1037-2021.

NASA/JPL, 2019. Jpl grace and grace-fo mascon ocean, ice, and hydrology equivalent water height coastal resolution improvement (cri) filtered release 06 version 02.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S., Madge, S., Prudden, R., Mandhane, A., Clark, A., Brock, A., Simonyan, K., Hadsell, R., Robinson, N., Clancy, E., Arribas, A., Mohamed, S., 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878), 672–677. doi.org/10.1038/s41586-021-03854-z.

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

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

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

Rodell, M., Reager, J. T., 2023. Water cycle science enabled by the GRACE and GRACE-FO satellite missions. *Nature Water*, 1, 47–49. doi.org/10.1038/s44221-022-00005-0.

Sandfort, V., Yan, K., Pickhardt, P. J., Summers, R. M., 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*, 9(1), 16884. doi.org/10.1038/s41598-019-52737-x.

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

Scanlon, B. R., Zhang, Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., Longuevergne, L., Chen, J., 2016. Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, 52(12), 9412–9429. doi.org/10.1002/2016WR019494.

Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Sasgen, I., Famiglietti, J. S., Landerer, F. W., Chambers, D. P., Reager, J. T., Gardner, A. S., Save, H., Ivins, E. R., Swenson, S. C., Boening, C., Dahle, C., Wiese, D. N., Dobslaw, H., Tamisiea, M. E., Velicogna, I., 2019. Contributions of GRACE to understanding climate change. *Nature Climate Change*, 9(5), 358–369. doi.org/10.1038/s41558-019-0456-2.

Wang, J., Shen, Y., Awange, J. L., Yang, L., 2023. A deep learning model for reconstructing centenary water storage changes in the Yangtze River Basin. *Science of The Total Environment*, 905, 167030. doi.org/10.1016/j.scitotenv.2023.167030.

Wang, J., Yang, L., Shen, Y., Chen, Q., 2024. Total water storage anomalies reconstruction using noise-augmented u-shaped network: A case study in the Yangtze River Basin. *Computers & Geosciences*, 183, 105498. doi.org/10.1016/j.cageo.2023.105498.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2018. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, UT, USA, 8798–8807.

Yin, J., Slater, L. J., Khouakhi, A., Yu, L., Liu, P., Li, F., Pokhrel, Y., Gentine, P., 2023. GTWS-MLrec: global terrestrial water storage reconstruction by machine learning from 1940 to present. *Earth System Science Data*, 15(12), 5597–5615. doi.org/10.5194/essd-15-5597-2023.

Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. *International conference on machine learning*, PMLR, 7354–7363.