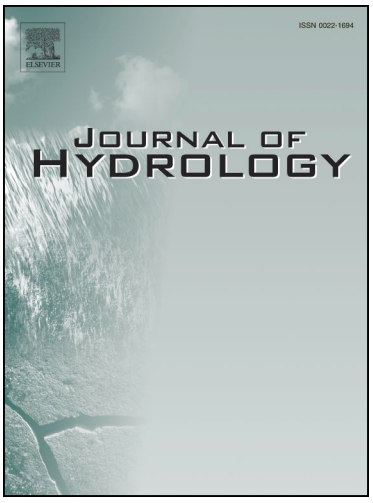
Machine learning assessment of hydrological model performance under localized water storage changes through downscaling

Journal Pre-proofs



Machine learning assessment of hydrological model performance under localized water storage changes through downscaling

Ikechukwu Kalu, Christopher E. Ndehedehe, Vagner G. Ferreira, Mark J. Kennard

PII: S0022-1694(23)01539-1
DOI: https://doi.org/10.1016/j.jhydrol.2023.130597
Reference: HYDROL 130597

To appear in: *Journal of Hydrology*

Received Date: 4 September 2023
Revised Date: 13 November 2023
Accepted Date: 19 November 2023

Please cite this article as: Kalu, I., Ndehedehe, C.E., Ferreira, V.G., Kennard, M.J., Machine learning assessment of hydrological model performance under localized water storage changes through downscaling, *Journal of Hydrology* (2023), doi: https://doi.org/10.1016/j.jhydrol.2023.130597

Machine learning assessment of hydrological model performance under localized water storage changes through downscaling

Ikechukwu Kalu[a,b], Christopher E. Ndehedehe[a,b], Vagner G. Ferreira[c], Mark J. Kennard[a,b]

[a]*School of Environment & Science, Griffith University, Nathan, QLD 4111, Australia.*

[b]*Australian Rivers Institute, Griffith University, Nathan, QLD 4111, Australia.*

[c]*School of Earth Sciences and Engineering, Hohai University, Nanjing, China*

* Corresponding author.
Email address: ikechukwu.kalu@griffithuni.edu.au (ORCID: 0000-0003-1708-4354)

Abstract

The coarse spatial resolution of the Gravity Recovery and Climate Experiment (GRACE) data has limited its application in the management of local-scale water resources. To address this limitation, we developed a new downscaling approach using predictors from regional and global hydrological models for a 15-year period (2002-2017) and tested it in the northern Great Artesian Basin, Australia. We used four different machine learning algorithms (support vector machine, partial least squares, gaussian process and random forest) to downscale the original GRACE estimate of 0.5° to a spatial grain size of 0.1° (global) and 0.05° (regional). This was based on precipitation, evapotranspiration and runoff estimates from the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) and Australian Water Outlook (AWO) hydrological models, respectively. The downscaled products were validated using 42 in-situ precipitation observations spread across the test region. We further evaluated which of the downscaled products best mimicked local-scale hydrology using a range of statistical metrics. Our results showed that regional hydrological models best characterized the dynamics of local scale hydrology (rainfall v. downscaled product), and the gaussian process regression algorithm made the best predictions for both models. The correlation coefficients for the raw values varied from 0.45 to 0.49 while that of the standardized values varied from 0.46 to 0.52 with the random forest model providing the best fitting for the regional-based products. The regional downscaling approach employed in this study may be readily integrated into local water resources planning programs.

1. Introduction

The emerging assessment of catchment-scale water resources under a changing climate has intensified the need to bridge the spatial scale gap between coarse GRACE- Terrestrial Water Storage (TWS) estimates and local scale hydro-climatic variables through downscaling (Atkinson, 2013; Sachindra et al., 2018). Presently, the two major methods for downscaling are dynamic and statistical downscaling (Yin et al., 2018). While dynamic downscaling techniques are used in building regional numerical models at finer spatial resolutions, statistical downscaling applies historical observations to derive relationships between large-scale climate factors and local-scale hydrological variables (Ndehedehe, 2022a). This provides the capacity to flexibly build models and incorporate high-resolution modelling factors to improve the spatial resolution of the response variable. These high-resolution modelling factors can be

acquired from either regional or global hydrological models. The entire statistical downscaling process is aimed at providing finer details of a hydrological estimate (in our case, GRACE-TWS), which best depicts the local scale hydrology of the region of interest. It is thus pertinent to compare the outputs from regional and global hydrological models to improve understanding of which provides the best estimate for local-scale terrestrial water storage assessments. Wilby et al. (2004) sub-divided statistical downscaling approaches into three categories: weather classification-based approaches, weather generators and regression-based approaches. Of the three, the latter has gained the most popularity and application due to its simplicity. The regression techniques widely used in statistical downscaling include multi-linear regression, generalized linear models, artificial neural networks, support vector machines, partial least squares regression etc (Sachindra et al., 2014; Beecham et al., 2014; Ahmed et al., 2015; Goly et al., 2014; Vishwakarma et al., 2021). Due to their learning and prediction capabilities, they are often referred to as machine learning methods.

Machine learning methods like regression and neural networks can effectively solve non-linear problems and simulate complex hydrological processes. This makes them well-suited for downscaling applications (Tao et al., 2023). For example, Vishwakarma et al., (2021) exploited dominant common statistical modes between high resolution datasets from global hydrological models to improve the spatial resolution of global GRACE-TWS using partial least squares regression. The efficacy of their downscaled products for the 160 catchments explored in their study was evaluated by using their individual conservation of mass. Using an empirical regression model, Ning et al. (2014) developed an integrated downscaled product of GRACE-TWS at 0.25º scale from global hydrological models. The downscaled product was validated against in-situ groundwater levels over the Yunnan province, China. Apart from global hydrological models, the work undertaken by Miro and Famiglietti (2018) revealed the potential of regional hydrological models in machine learning (artificial neural network) downscaling of GRACE products. Using high resolution groundwater storage changes from the San Joaquin valley in California, their work demonstrated the flexibility of the method and datasets in accurately reproducing local groundwater trends and its heterogenous behaviour. Furthermore, He et al. (2021) developed three convoluted neural network (CNN) - based frameworks to downscale the GRACE-TWS from 110km to 10km using simulations from the Ecological Assimilation of Land and Climate Observation (EALCO) regional hydrological model and compared the outputs of the CNN-based methods with the empirical linear regression-based downscaling method. These studies explored the use of high-resolution global and regional hydrological models in their downscaling approach. They also demonstrate the benefits (e.g., simplicity, flexibility etc.) associated with each machine learning technique in advancing knowledge on the downscaling of GRACE products. However, there is still uncertainty as to how downscaled products mimic local scale hydrological fluxes and patterns (catchment-based land water storage changes) to enable us determine which hydrological model and machine learning method best suits the downscaling operation. Very few studies have employed a hybrid approach that combines the strengths of both global and regional hydrological models in downscaling (e.g., Seyoum et al., 2019). Such an integrated approach can potentially provide more robust and accurate downscaling results, especially when dealing with regions that are subject to diverse hydrological conditions and require a comprehensive understanding of both large-scale and small-scale dynamics. However, for the purpose of this study, we limit our experiments to global and regional hydrological models.

Global hydrological models (GHMs) have emerged as a critical tool in analysing complex, large-scale hydrological systems and predicting the distribution, availability, and movement of water resources globally. Their broad coverage provides them with a consistent framework for

hydrological simulations across different regions and enables them to capture large-scale climate patterns and trends. This has made them very popular in statistical downscaling operations of GRACE-ΔTWS estimates (e.g., Vishwakarma et al., 2021; Ning et al., 2014; Yin et al., 2018). Regional hydrological models (RHMs) on the other hand provide detailed, high-resolution simulations at a regional or local scale. They capture fine-scale hydrological processes, topography, land use and other factors that influence water availability and movement. Their incorporation of local data and observations makes it easier to represent the unique characteristics of specific regions leading to a more realistic representation of regional hydrological processes. Global and regional hydrological models share some similar characteristics but also have distinguishing features. These features can include varying hydrological scales, focus areas, specific purposes and land surface modelling modes (i.e., climate inputs and post processing techniques). In this study, our emphasis is to examine the effects of these characteristics in land water storage assessments by downscaling with high resolution predictors from both model types. Regional hydrological downscaling produces higher spatial resolution and provides detailed designs aimed at capturing local hydrological processes. However, their downscaling is strongly dependent on the lateral boundary conditions and the techniques used to constrain the high-resolution predictors to the coarse scale predictand. The large-scale errors inherent in the coarse scale predictands tends to be retained in the downscaled model, thus increasing the risk of overfitting and uncertainty amplification (Pielke & Wilby, 2012). On the other hand, global hydrological models which best characterizes large-geographical extents maintains a consistent modelling approach with GRACE-TWS and ensures uniformity in the analysis of TWS variations leading to a better quantification and propagation of uncertainties in the model output (Güntner, 2008). To have a better understanding of our downscaling process and final products, we applied four well-known regression models on both the global and regional downscaling models. To estimate the performance of our downscaled products over local-scale hydrology, we compared the downscaled estimate from each method and model to in-situ precipitation estimates over the study region and period. The importance of precipitation in validation was warranted as several studies have reported its role as the most dominant forcing mechanism responsible for the spatio-temporal variations of GRACE TWS in tropical and temperate regions (e.g., Song et al., 2015; Rodell et al., 2018 Meng et al., 2019; Yin et al., 2020; Kalu et al., 2021, 2022).

The key objective is to evaluate which hydrological model and machine learning method provides the most accurate downscaled water storage estimates compared to local-scale hydrological flux changes. To achieve this, we employed high spatial-resolution datasets from both global and regional hydrological models to perform a statistical downscaling of GRACE-TWS using various regression-based approaches. The downscaled products from both models were evaluated against in-situ based precipitation patterns over the study region to provide insights as to which model best characterizes local-scale hydrology. By local-scale hydrology, we mean the dynamics of land water storage changes induced by hydrological flux variables over a small catchment. We also make suggestions on the preferred model and method for regional ΔTWS downscaling operations.

2. Study region and data

The Carpentaria Basin (CB) is located in northern Australia and situated at the northern part of the Great Artesian Basin (Fig. 1) with an estimated area of 300,000 km^2. Its climate is influenced by monsoonal patterns of tropical northern Australia resulting in distinct wet and dry seasons. Our choice of the CB as a suitable test region was to provide a more comprehensive understanding of freshwater dynamics, its availability, and the impacts of

climate variability across a wet-dry climate zone. This knowledge is vital for sustainable water management, ecosystem preservation, and mitigating water-related challenges in diverse environments.

The climatic condition of the Carpentaria Basin is hot and humid with two seasons per year. The dry season runs from May to October and is characterized by lower rainfall and higher temperatures. During this period, water sources may become more limited, and some areas may experience reduced water availability, fire risks, and drought conditions (Kaushik et al., 2021). On the other hand, the wet season occurs from November to April. During this time, the region receives heavy rainfall due to the monsoon winds and the convergence of moist air masses. These rains are crucial for replenishing water bodies, supporting vegetation growth and maintaining ecological balance. Almost all the rains are compressed into two or three months, and during this period many low-lying areas are flooded. A large portion of the CB comprises floodplains, with very low relief (Fig. 1), resulting in significant tidal inundation during the wet season. More than 80% of the annual rainfall results in large, inundated areas of these floodplains, much of which recede relatively rapidly, in some downstream areas (Ndehedehe et al., 2021)

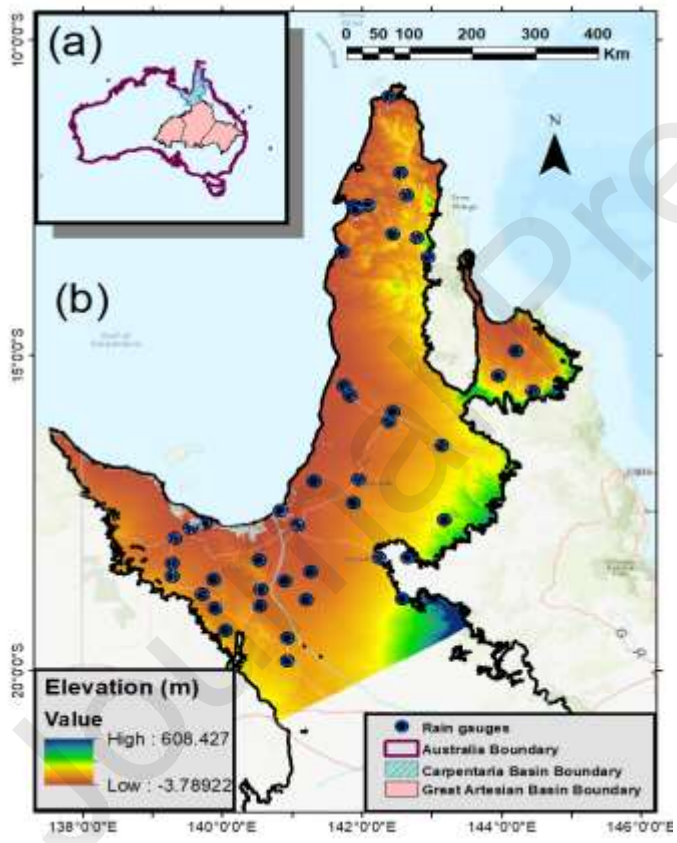


Fig 1. Study area map showing (a) the situation of the CB in Australia and (b) the topography of the test region in the northern Great Artesian Basin, Australia and the location of the rain gauges used for validation and local-scale assessment. (The ground surface topography data used for this figure was gotten from the Australian Smoothed Digital Elevation Model which has been smoothed to reduce noise and improve the representation of surface shape).

In recent years, the CB, like many other regions around the world has been affected by the impacts of climate change (CSIRO, 2019). Changes in precipitation patterns, increased

temperatures, and more intense weather events can influence the basin's hydrology, affecting water availability and the productivity of freshwater habitats by limiting flows, which supply nutrients and supports the movement of fisheries from floodplain environments to marine systems (Ndehedehe et al., 2021). Using the CB as a study region, our analysis provides more insight into the climate and hydrology of the region to ensure sustainability of water resources, protect ecosystem services and support the livelihood of local communities.

2.1 GRACE data

For this study, we used GRACE level 3 mascon products from the centre for space research (CSR) of the University of Texas, Austin (Save et al., 2016) to quantify changes in terrestrial water storage. The GRACE level 3 mascon products used in this study covered a total period of 183 months, spanning from April 2002 to June 2017. Data gaps for months within our study period were resolved using cubic interpolation for cases where more than two consecutive solutions were missing. For single missing months (for e.g., June 2003, Jan 2011, May 2012, Oct 2012), their values were interpolated by averaging the values of 2 months before and after the month with missing data (e.g., Long et al., 2015; Yin et al., 2018). It is important to note that GRACE dataset at 1.0 º, 0.5 º and even 0.25º represent spatially interpolated quantities of a coarser GRACE sample. This means that these available high resolution GRACE products do not contain any new physical information at a spatial scale better than the native resolution (*see* Luthcke et al., 2013; Watkins et al., 2015; Vishwakarma et al., 2021). Therefore, the choice of which GRACE sample was used for the experiment may not directly affect the results of our regression-based downscaling operation which aims at the introduction of new hydrological variables to GRACE-TWS at higher resolutions. For our study, we used the filtered and processed samples of the nominal GRACE dataset, which was provided at a 0.5º × 0.5º grid cell for our experiment.

2.2 Global model (FLDAS data)

The Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) is a collaborative data assimilation project involving the U.S. Geological Survey (USGS), the University of California Santa Barbara (UCSB) Climate Hazards Group (CHG), NASA Goddard Space Flight Centre (GSFC) and the Earth Resources Observation and Science (EROS) centre (Khorrami et al., 2023). By integrating remotely sensed, modelled, and in-situ observations, FLDAS simulates different hydrological variables at a global scale. Our hydrological variables of interest were precipitation, evapotranspiration and runoff datasets from the Noah 3.6.1 model simulations with a spatial resolution of 0.10 degrees. The simulations were forced by a combination of the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) data and Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) 6-hourly rainfall data to produce high resolution datasets. These high-resolution predictors were extracted from the FLDAS-Noah and integrated into our downscaling algorithms to produce finer-scale resolutions of ΔTWS over our test region.

2.3 Regional model (AWO data)

The Australian Water Outlook (AWO) is a regional water balance model based on the landscape component of the Australian Water Resource Assessment system (AWRA-L). The model is a one-dimensional grid-based water balance model that has a semi-distributed representation of the soil, groundwater, and surface water stores. The AWO system incorporates a wide range of climate inputs, downscaling techniques, post processing and assimilation of near real time satellite soil moisture states as inputs to the AWRA-L model (Van Dijk, 2010; Frost & Shokri,

2021), to provide a consistent set of hydrological outputs at 0.05° grids across Australia. Our high-resolution predictor datasets of interest were precipitation, evapotranspiration, and runoff datasets. These were used to perform a statistical downscaling of ΔTWS over our test region, which was validated using in-situ precipitation data (section 2.4).

2.4 In-situ data

In this study, monthly precipitation values of 42-gauge stations in the CB from April 2002 to June 2017 were collected from the Australian Bureau of Meteorology (BOM). This data represents the local scale hydro-climatology of our test region. The rain gauges are almost uniformly distributed over our test region, though with a concentration to the lower part of our study area, and their locations are shown in Fig 1. The rain gauge time series generated from this dataset is the arithmetic average of all the gauges within our test region used in our experiment. Overall, the in-situ datasets for the 42 gauges are continuous during our study period, and the gauges with missing monthly datasets were spatially interpolated using the same technique from the GRACE data gaps. Table 1 details the information regarding the characteristics and source of the acquired datasets used in our study.

Table 1. Characteristics and sources of the dataset used in our processing.

Type	Parameter	Mission	Resolution
Satellite (a)	ΔTWS	GRACE	0.5º × 0.5º
Global Model (b)	Precipitation Evapotranspiration Runoff	FLDAS (Noah)	0.1º × 0.1º
Regional Model (c)	Precipitation Evapotranspiration Runoff	AWRA-L (AWO)	0.05º × 0.05º
In-situ (d)	Gauge precipitation	Observations	-

(a) https://www2.csr.utexas.edu/grace/RL06_mascons.html

(b) https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_C_GL_M_001/summary?keywords=FLDAS_NOAH

(c) https://awo.bom.gov.au/products

(d) http://www.bom.gov.au/climate/data/index.shtml

3. Methods

3.1 Regression-based downscaling approach

A primary premise of the spatial downscaling method is that spatial relationships established between GRACE-ΔTWS and hydrological variables at a coarse spatial scale can be used for finer scale ΔTWS estimations using high-resolution predictors (Arshad et al., 2022; Khorrami et al., 2023). The specific steps of our regression-based downscaling approach are summarized in Fig. 2 and explained below.

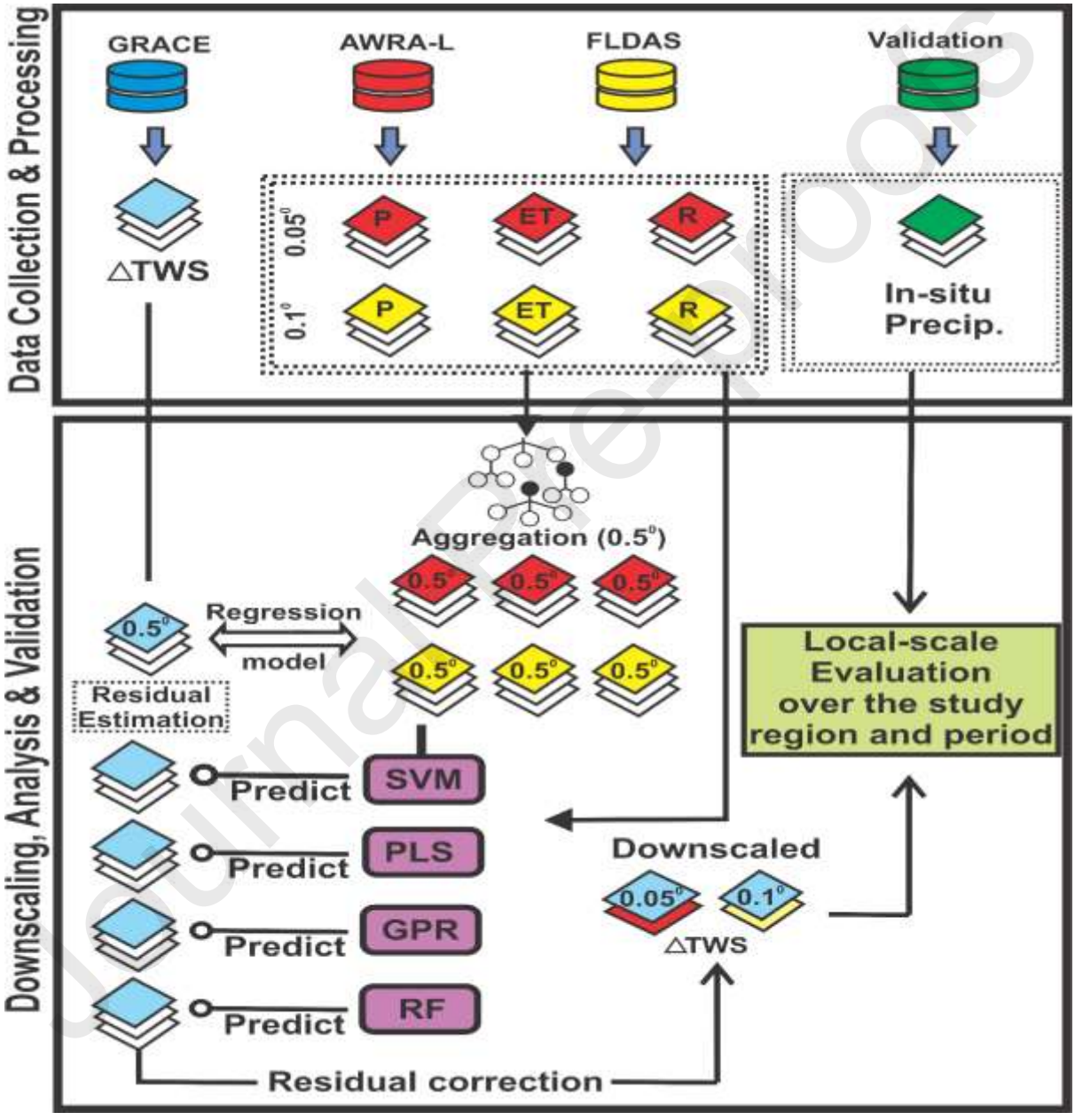


Fig 2. Schematic flow of the analysis used in this study.

i. *Data resampling and aggregation*: All high-resolution predictors from both models (precipitation, evapotranspiration, and runoff) were spatially resampled to coarse resolution of the original GRACE samples by pixel aggregation. This was necessary to achieve consistency in the spatial grain size of the predictors and response variables.

ii. *Machine learning*: Four different machine learning models (*see* section 3.2) were built to establish a regression-based relationship between the original GRACE-TWS and the predictors from the global and regional models, respectively. These models were used to predict TWS at 0.5º.

iii. *Model residuals*: The predicted TWS was subtracted from the original GRACE TWS (0.5º × 0.5º) to obtain the model residuals for the respective learning models (ε).

iv. *Evaluation*: The predictive capabilities of each individual machine learning method was analysed and the residuals were subsequently interpolated to 0.10º × 0.10º for the global and 0.05º × 0.05º for the regional hydrological model predictors using ordinary kriging approach (Duan & Li, 2016; Arshad et al., 2022).

v. *Downscaling*: Downscaled ΔTWS estimates were obtained using high-resolution hydrological predictors from both global (0.10º × 0.10º) and regional (0.05º × 0.05º) models in conjunction with the residuals in step iv. The steps used in the prediction for each machine learning model are further detailed in section 3.2.

3.2 Machine learning models

In this study, we evaluated the capabilities of four different machine learning models in our regression-based downscaling, namely: support vector machines (SVM), partial least squares regression (PLR), gaussian process regression (GPR), and random forest (RF). Their formulations, statistical properties and how they were developed for our downscaling process are detailed below.

3.2.1 Support vector machines

The support vector machine (SVM) is a supervised learning technique proficient in linear and non-linear regressions. Its principle is based on the method of structural risk minimization and statistical learning theory (Kalu et al., 2023; Roy & Chakraborty, 2023). The SVM regression is considered a nonparametric technique because it relies on kernel functions. In SVM regression, the set of training data includes predictor variables and observed response values. The goal is to find a function *f(x)* that deviates from y_n by a value no greater than ε for each training point x, and at the same time as flat as possible.

Suppose we have a set of training data where x_n is a multivariate set of N observations with observed response values y_n; To find the linear function $f(x) = x'\beta + b$ and make sure it is as flat as possible, we must find *f(x)* with the minimal norm value ($\beta'\beta$). This is expressed as a convex optimization problem to minimize $J(\beta) = 0.5 \times \beta'\beta$ subject to all residuals having a value less than ε (Vapnik, 1999)

It is possible that no such function *f(x)* exists to satisfy these constraints for all points, which warrants the introduction of slack variables (ξ_i, ξ_i^*) for each point to deal with their respective infeasible constraints (eqn. S3-S4) (Smola & Schölkopf, 2004).

3.2.2 Partial least square regression

The partial least square regression (PLR) model is a two-edged multivariate tool as it combines features from multi linear regression and principal component analysis (Ndehedehe & Ferreira, 2020). It decomposes the mean data of X (predictor variable) and Y (target variable) while

attempting to maximize the covariance between the two variables whilst complying with certain normalization and orthogonality constraints (e.g., Ndehedehe & Ferreira, 2020).

Consider a set of data $\{(x_i, y_i); i = 1,2,3, \dots, n\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, where n is the number of samples (in our case 183). The centered data matrices X and Y are decomposed as

$$X_{n\times p} = t_{n\times 1}P'_{p\times 1} + E_{n\times p}, Y_{n\times q} = u_{n\times 1}q'_{q\times 1} + F_{n\times q'} \quad (1)$$

where n is the corresponding dependent samples, t and u are latent vectors for the n observations, p and q are the loading vectors, while E and F are the residual vectors. PLR maximizes the square covariance between the latent vectors (t and u) to generate the projection vectors w and h as t=Xw and u=Yh (*see* eqns. S7-S10) (Yan et al., 2013; Chen et al., 2018).

3.2.3 Gaussian process regression

The Gaussian process regression (GPR) is a supervised non-parametric kernel-based probabilistic model. It is a machine learning method based on the Bayes theory and statistical learning theory (Kalu et al., 2023).

Consider a set of data $\{(x_i, y_i); i = 1,2,3, \dots, n\}$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, drawn from an indefinite distribution. The GPR algorithm attempts to predict the value of the target variable y_{pred} based on a linear regression of the form

$$y = x^T\beta + \varepsilon \quad (2)$$

where, $\varepsilon \sim N(0, \sigma^2)$. The priori distribution of the observation y and the joint prior distribution of the observed y value and the predicted value, y_{pred} can be obtained by

$$y \sim N(0, K(X, X) + \sigma_n^2 I_n) \quad (3)$$

$$\begin{bmatrix} y \\ y_{pred} \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I_n & K(X, x_{pred}) \\ K(x_{pred}, X) & K(x_{pred}, x_{pred}) \end{bmatrix}\right) = N\left(0, \begin{bmatrix} K & K_*^T \\ K_* & K_{**} \end{bmatrix}\right) \quad (4)$$

where, $K(x, x)$ represents a symmetric positive definitive covariance matrix, whose elements estimate the association between x_i and x_j through the chosen kernel function; $K(x_*, x) = K(x_*, x)^T$ depicts the covariance matrix between the test sets, x_{pred} and the training set x; $K(x_*, x_*)$ depicts the covariance matrix of the test set itself; I_n represents an n-by-n unit matrix.

3.2.4 Random forest

Random forest (RF) integrates ensemble learning methods that combines multiple decision trees to create a robust model. Each tree handles a subset of data to make a collective prediction using bootstrapping for subsample selections, and randomness in decision making (Heung et al., 2014). In RF regression, an extra function is utilized in combining the decision of individual trees (Eqn. 5).

Consider a set of data for bootstrap aggregating $X = x_1, x_2, x_3 \dots x_n$ having a target of $Y = y_1, y_2, y_3 \dots y_n$, the bagging of sub-sample repeated N times to fit the trees.

$$\widehat{rF} = \frac{1}{N}\sum_{i=1}^{N} f_i(x') \quad (5)$$

where, $f_i(x')$ are the random forest trees and $\widehat{rF}$ depicts the prediction of an unknown point x' using the bootstrapped dataset.

3.3 Evaluation metrics

The performance of our machine learning and downscaled products were evaluated using four statistical metrics which include spearman's rank correlation (R), Nash-Sutcliff Efficiency (NSE), root mean square error (RMSE), and mean absolute error (MAE) (Rahman et al., 2023) (Eqns. S11-S14). It is important to point out that the use of original coarse GRACE data to validate the downscaled higher resolution data can be problematic. This is because the original GRACE data represents large scale regional trends, so it may not fully capture localized dynamics shown in the downscaled data. Therefore, the direct comparison to GRACE could underestimate errors in the downscaled products. To circumvent this, we performed a spatial aggregation of our downscaled products to allow for a direct comparison. We took spatial averages over larger areas closer to the GRACE resolution to provide a better and proficient method of consistency check. To further check for statistical inference, we resampled the downscaled products using bootstrapping. This process entails creating multiple simulated datasets for the SVM, PLR, GPR and RF algorithms by randomly sampling observations from the downscaled estimates with replacements. The simulated products were also used to estimate the properties of our statistical model and gauge the performance of our learning algorithms (Li et al., 2010; Arsenault et al., 2018).

3.4 Wavelet coherence

Wavelet coherence analysis is particularly well-suited for studying relationships between time series data that exhibit non-stationary and time-varying behaviour, making it a valuable tool for examining the association between hydrological variables. We used it to assess the local-scale evaluation of the various downscaled products (from different models and machine learning algorithms) over our test region and within the time-frequency space. Wavelet coherence attempts to determine the relationship between the variables (in-situ precipitation v. downscaled TWS) by estimating the correlation between them that varies between 0 and 1. The coefficient of wavelet coherence (Grinsted et al., 2004; Ali et al., 2022) between the two sets of time series can be denoted as

$$R^2(s) = \frac{|S(s^{-1}W_n^{XY}(s))|^2}{S\left(s^{-1}|W_n^X(s)|^2\right)\cdot S(s^{-1}|W_n^Y(s)|^2)} \quad (6)$$

where $R^2(s)$ is the coherence coefficient minimum and maximum coherence at 0 and 1, and $W_n^{XY}(s)$ is the cross wavelet transforms between series two series (in our case, in-situ precipitation and all the downscaled TWS products), S is the smoothing operator and is represented as

$$S(W) = S_{scale}(S_{time}(W_n(s))) \quad (7)$$

where, S_{scale} refers to smoothing along the wavelet scale axis and S_{time} smoothing in time. For our experiment, we examined the wavelet coherence at a 5% significance level.

To perform appropriate analysis and consider domain knowledge to interpret the underlying patterns of the in-situ precipitation and downscaled products, we transformed their time series into a series of percentiles. The transformed time series have been shown to react 'more linearly' to climate datasets (*see* Grinsted et al., 2004). The need to interpret the underlying

patterns more efficiently warranted our use of the continuous wavelet transform (CWT) to extract a low signal-to-noise ration signal between the in-situ estimates and downscaled GRACE-TWS. The CWT is a popular tool for assessing localized intermittent oscillations in time series observations. However, it is mostly proficient in the examination of two time series that are linked in some way. It can be very useful in investigating if regions in time frequency intervals with large common power have a consistent phase interaction and suggests causality between the time series observations.

3.5 Regression model and local scale evaluation/validation

The variables used for our predictions are representative of the dominant hydrological fluxes of the region. Precipitation, evapotranspiration and runoff form the water budget equation which illustrates the water exchange between the ocean, land and atmosphere. They collectively regulate the movement, distribution, and availability of water within a region, shaping its hydrological cycle and influencing both natural ecosystems and anthropogenic activities.

These variables and the GRACE-TWS were set as input and target variables, respectively for each of the four algorithms explored in our experiment. The variables which were originally in fine resolution were aggregated to the resolution of the GRACE-TWS to maintain consistency between the spatial grains of the input and target variables. This enabled us to build a functional regression model based on the four learning algorithms. These regression models were used to predict GRACE-TWS both at the original scale and at a finer scale using the input variables at their original resolutions. It was important to predict GRACE TWS at the original scale to estimate the residuals between the observed and predicted GRACE TWS for each algorithm and add them back to the predictions of the high-resolution GRACE TWS. Adding back the residual after the entire downscaling process is a critical step to enhancing the accuracy and spatial representation of the downscaled results. By incorporating residual information, the downscaled estimates better capture finer-scale variability and localized features that are important for understanding and managing terrestrial water storage dynamics at regional scales. However, we attempted to evaluate the capabilities of the regression models in the prediction of high resolution ΔTWS estimates based on the independent parameters of the regional and global models before adding back the residuals. This is because the predicted values without the residuals represent the systematic trends and relationships captured by the individual learning models. By providing values based on the overall patterns and trends present in the original GRACE data, we can efficiently gauge the performance of each learning model before adding the residuals back to them.

One of the primary goals of statistical downscaling of GRACE-TWS is to provide more detailed and localized information about changes in water storage within a specific region. It therefore follows that the task of validating any downscaled product must be through a local-scale hydrological metric. For our study, we validated our downscaled products and assessed their closeness to local-scale hydrological variability using in-situ precipitation observations of 42 rain gauges scattered across our test region. In-situ precipitation observations provided essential data to quantify rainfall input (which is regarded as the primary variable in the water budget), model hydrological processes, validate models and manage water resources within a catchment. Our choice of in-situ precipitation over satellite precipitation is aimed at assessing information for accurate local-scale water storage assessments which has been observed to be the most dominant contributor of GRACE-ΔTWS (Ndehedehe et al., 2021; Shah & Mishra, 2021; Kalu et al., 2022).

Finally, with the use of the bootstrapping resampling technique, we performed a probabilistic validation of the downscaled products against multiple simulated datasets which were randomly sampled from the respective regression models with replacement. This provided us the opportunity to carefully consider the uncertainty estimates of our regression-based downscaled products.

3.6 Global versus Regional model downscaling

GHMs ensure uniformity in modelling approaches and parameterization, which can be advantageous for comparative studies and global climate change analysis. This allows researchers to assess how changing climate patterns might affect water availability, distribution and related phenomena on a broad scale. We tried to understand how these numerous benefits of the GHM impacts on local-scale downscaling by using different regression-based algorithms (Fig. 3). And given that GHMs offer a reference point for evaluating the performance of regional models, comparing outputs from regional and global models could help identify areas where regional models exceed or provides a better characterization of local scale hydrology better than their global counterparts.

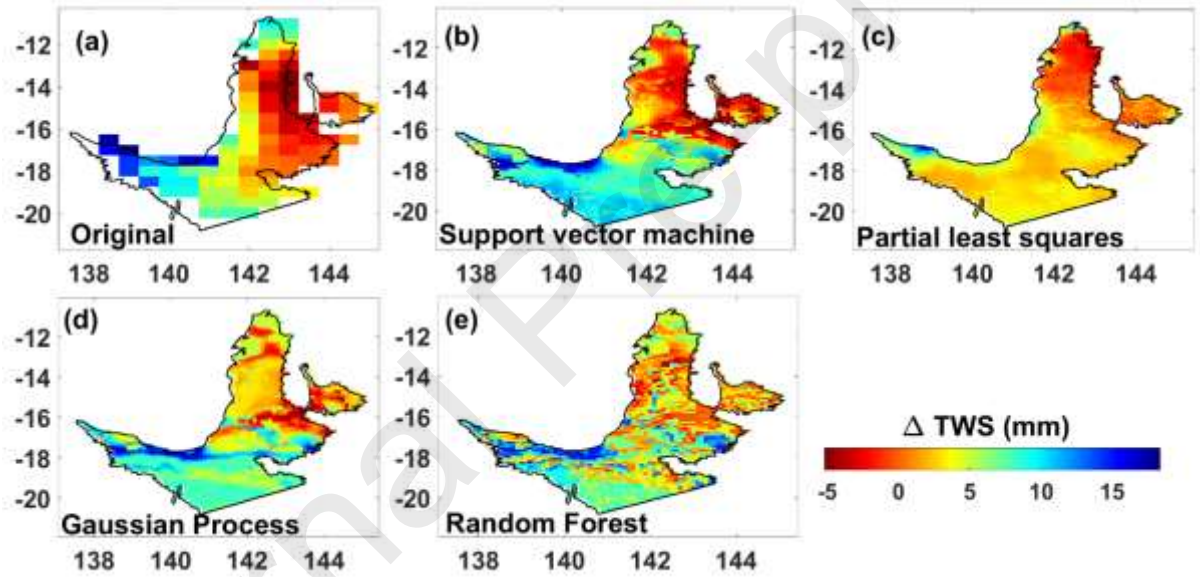


Fig 3. Regression-based downscaling at 0.1° × 0.1° based on global hydrological models (FLDAS). These observations are the mean of monthly GRACE -TWS observations from April 2002 to June 2017.

Surprisingly, we have more downscaling operations use GHMs than RHMs in their operations. Even though this may be due to the easy availability of GHMs, it is important to point out the relevance of RHMs in downscaling operations particularly in crucial areas with complex terrains, diverse land cover, or specific hydrological challenges.

We tested the same regression models used for the GHMs in RHMs (Fig. 4). This helped us to verify which hydrological model provides a more detailed assessment of local management practices on water resources.

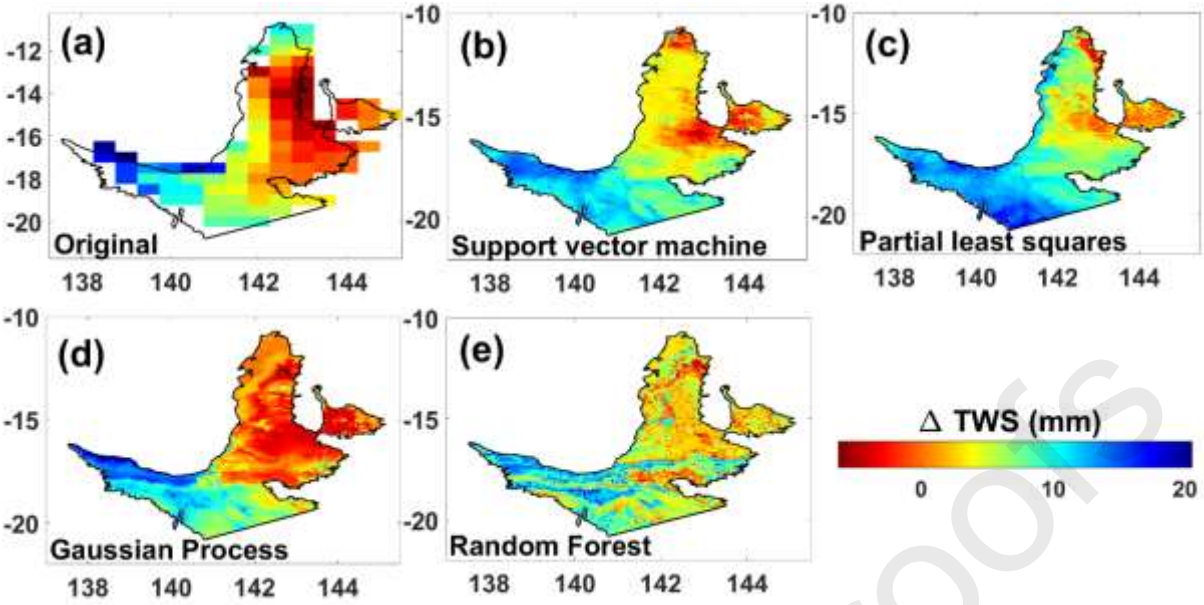


Fig 4. Regression-based downscaling at 0.05º × 0.05º based on regional hydrological (AWO) models. These observations are the mean of monthly GRACE -TWS observations from April 2002 to June 2017.

4. Results and Discussion

4.1 Performance of the regression models

To gauge the performance of the regression models, it was necessary to compare the original and downscaled ΔTWS estimates (Figs. 5, 6).

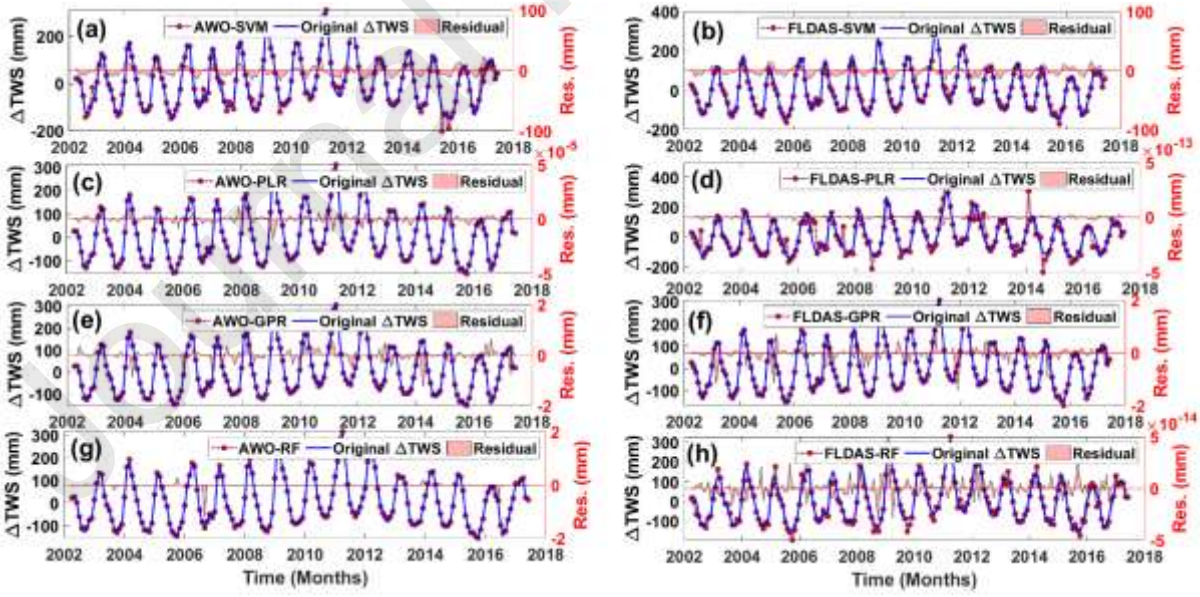


Fig 5. Time series of the original and downscaled GRACE-TWS for each algorithm before adding back the residuals. The regional and global hydrological models which are represented in panels a,c,e,g and b,d,f,h, respectively, are used to downscale GRACE TWS. The right ordinate (coloured red) in each pane represents the residuals which depicts the differences between the original and predicted values.

For the regional based predictions, the gaussian process regression showed exceptionally high agreement (R = 0.99, NSE = 0.99), and the lowest MAE (0.80mm) and RMSE (1.44mm), respectively. This is followed by the partial least square regression model with MAE (1.22mm) and RMSE (1.77mm) slightly higher than the gaussian process regression model. The random forest also maintained a very good fit with MAE and RMSE values of 3.58mm and 4.80mm, respectively, while the support vector machine, which provided a very good fitting of (R = 0.98, NSE = 0.97, RMSE = 17.04mm, and MAE = 10.67mm) has the lowest fitting amongst the four models used for the regional-based predictions (Fig 6a-d). The global based predictions tended to be slightly less accurate than the regional based predictions. The gaussian process regression also indicated the highest agreement between original and downscaled products (R = 0.99, NSE 0.99), and the lowest MAE (4.63mm) and RMSE (7.20mm), respectively. The support vector machine also showed considerably strong agreement (R = 0.98, NSE = 0.97), with MAE and RMSE values of 12.11mm 15.49mm, respectively. The random forest regression model also produced a good fitting index of R (0.98mm), NSE (0.95), MAE (17.72mm) and RMSE (23.69mm), while the partial least square regression had the lowest fitting index of R (0.91), NSE (0.81), MAE (26.79mm) and RMSE (44.71mm) amongst the four models used for the global-based predictions (Fig. 6e-h). The varied performance of these learning algorithms under different scenarios (regional and global predictors) shows that the inherent differences between the regional and global datasets significantly impacts on the performance and behaviour of the machine learning models and hydrological model properties such as data distribution, scale discrepancy and spatial variability are key in maintaining accuracy and consistency in regression models.

Overall, the downscaled results were consistent with the original GRACE-ΔTWS observations (Fig. 5). Finer spatial patterns were discernible with the increased spatial resolution for all the machine learning methods compared to the original GRACE-ΔTWS observations (Fig. 3 and 4). We however observed an increase in the range between the maximum and minimum downscaled values with the regional-based downscaled TWS values having a wider range (Table 2). On average, the range of all the downscaled estimates increased slightly (Table 2), thereby widening the contrast between the data and the highlighting spatial patterns.

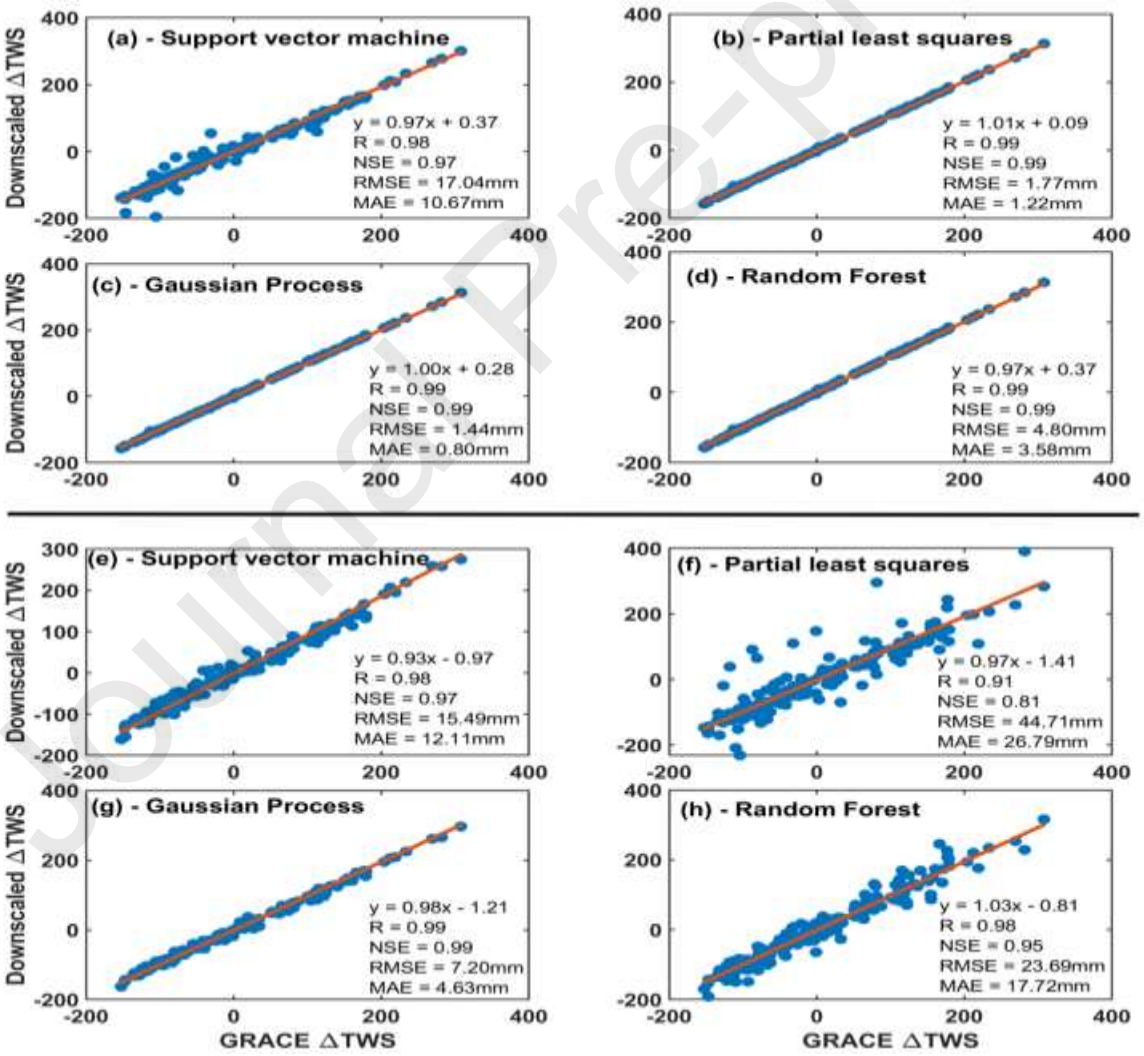
Table 2. summary attributes of the original and downscaled ΔTWS estimates

ΔTWS	**Min (*mm*)**	**Max (*mm*)**	**Average (*mm*)**	**SD (*mm*)**	**CV (*unitless*)**
GRACE-CSR	-196.17	328.20	4.81	108.56	18.84
AWO-SVM	-202.29	311.31	5.66	101.02	17.85
AWO-PLR	-155.83	310.14	5.61	104.05	18.56
AWO-GPR	-153.62	308.13	5.72	103.47	18.09

AWO-RF	-156.67	314.73	6.21	104.64	16.86
FLDAS-SVM	-171.54	295.49	4.15	97.42	23.49
FLDAS-PLR	-237.58	405.01	3.89	109.39	28.15
FLDAS-GPR	-167.21	305.50	4.27	101.43	23.77
FLDAS-RF	-196.17	328.20	4.81	108.56	22.55

469 Note: SD = Standard deviation, CV = coefficient of variation

470



471

472 Fig 6. Accuracy estimation of the machine learning models after adding the residuals back to
473 the high-res. predictions. The upper panel (a - d) represents regional based predictions from the
474 AWO, while the lower panel (e - h) represents global based predictions from the FLDAS model.

The x-axis shows the TWS derived from GRACE and the y-axis shows the downscaled TWS from the four machine learning methods for both regional (Fig. 6a-d) and global (Fig. 6e-f) based variables.

4.2 Validation and local-scale assessments

Precipitation plays a key role in driving terrestrial water storage by providing the primary source of water that replenishes various reservoirs within the earth's hydrological cycle (Khandu et al., 2016; Fatolazadeh & Goïta, 2021). When precipitation occurs, it infiltrates the soil, replenishing soil moisture and groundwater reservoirs. Additionally, it contributes to the filling of lakes, rivers, and other surface water bodies. Therefore, monitoring precipitation patterns is essential for understanding the availability and distribution of water resources in a region as they tell the most relatable story about the water storage dynamics in a localized unit. This is also evidenced in their action as the key variable in the water budget equation (Lehmann et al., 2022) in comparison to the other hydrological flux variables (ET and runoff). Their role in driving localized water storage dynamics is depicted in the strong relationship noticed between TWS and precipitation as shown in Figure 7a. Their relationship, however, is affected by lags which vary for different regions. We observed a two-month lag between precipitation and TWS for our entire study period which was partitioned into a four-year epoch (Fig. 7b, c, d, e) for a more detailed assessment. Our partitioning shows that the 2-month lag experienced over our study period was more consistent and strongest between 2002 to 2005 (Fig. 7b).

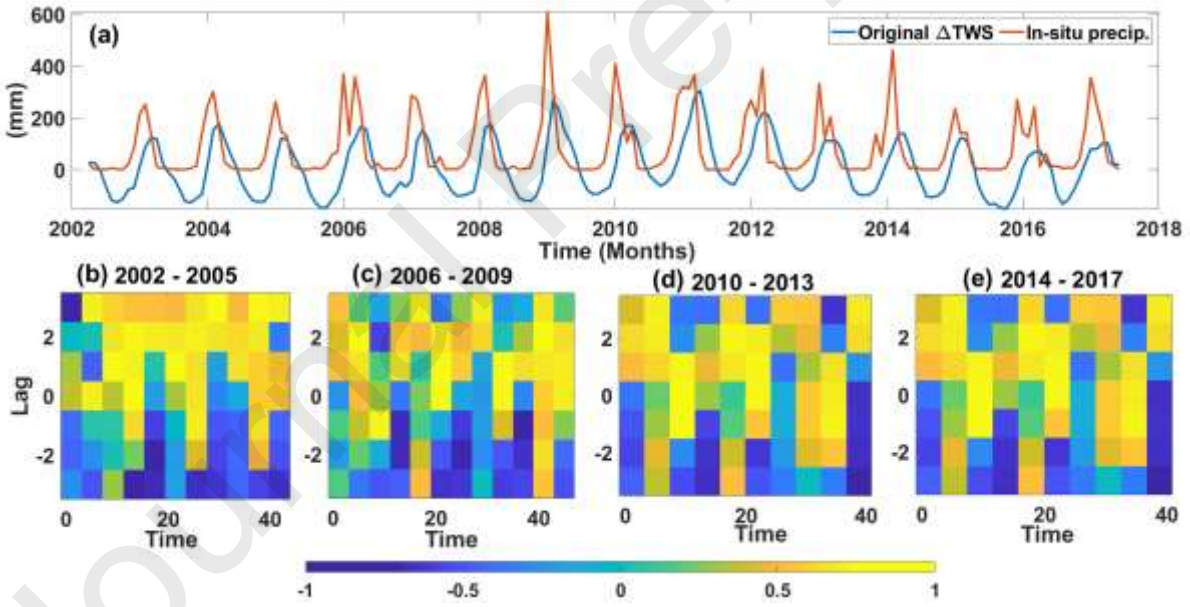


Fig 7. Windowed cross correlation between monthly precipitation and TWS signals up to a pre-defined lag of 3 months. Precipitation drives terrestrial water storage and therefore provides a key metric (via in-situ observations) for local scale water storage assessments. The lag between precipitation and TWS was assessed over four-year epochs (Figs. 7b, c, d, e) partitioned from the study period. The warm yellow grids represent a lag of > +2 months while the cool blue grids represent a lag of < -2 months.

It is important to point out that lag effects between precipitation and TWS are affected by factors such as varying soil types, land cover, geological characteristics etc. The CB which is characterized by a mix of sediments, sands, clays and gravels witnesses a diffuse percolation of precipitation as the soil is often porous, thereby allowing rainwater to infiltrate the ground.

This helped to ensure the minimum lag period of two months between precipitation and terrestrial water storage as is the case with other regions having similar hydrogeological frameworks (e.g., Ndehedehe & Ferreira, 2020; Kalu et al., 2021). The heterogenous nature of the hydrogeological formation of the CB results in the varying percolation rates in different parts of the basin which translates to inconsistent lag period between precipitation and TWS as shown in figure 7. The most inconsistent lag period was recorded between 2006 to 2009 (Fig. 7c). This may not be totally unrelated to the heavy rains in 2009 and 2010 that terminated the Australian millennium drought which started in 1997 (Ndehedehe, 2022b), thus disrupting the 10-year drought-induced hydrological cycle. This shows that the highly variable nature of precipitation in time and space gives us the best representation of terrestrial water storage dynamics over all scales.

4.2.1 Hydrological flux variable (In-situ precipitation) as a metric for local-scale water assessments

In-situ precipitation datasets was used as an index for local-scale water assessments by observing the similarity in trend between the time series of the in-situ precipitation and the GRACE TWS datasets over our study period (Fig. 7). This also provided another avenue to test which of the machine learning models best mimics the trend. This is important because it is expected that the downscaled estimates should represent regional scale features and trends not present in the original estimates. Between April 2002 and July 2017, the difference between the long-term linear trends of the in-situ precipitation. and GRACE TWS was +0.09mm/month, where each had a positive trend of +1.36 and +1.45, respectively. The GPR and the RF for both the regional and global datasets best captured these trends in both short and long term (Table 3). The linear trends of the different downscaled products and the in-situ precipitation values are summarized in Table 3.

Table 3. Comparison of in-situ precipitation and ΔTWS linear trends at different spatial resolutions over our test region. The rain gauges depict the localized hydrological patterns, while the GRACE-CSR, FLDAS and AWO models represents the satellite, global-downscaled and regional-downscaled ΔTWS estimates whose trends are evaluated against the in-situ precipitation patterns.

Resolution	**Precip. / ΔTWS**	**Data Source**	**Trend (mm/month)**		
			2002 - 2009	**2010 - 2017**	**2002 - 2017**
-	Rain gauges	In-situ	+5.35	-7.72	+1.36
0.5°	GRACE-CSR	Satellite	+4.07	-14.92	+1.45
0.1°	FLDAS-SVM	Global model	0.00	0.00	0.00
	FLDAS-PLR		0.00	0.00	0.00
	FLDAS-GPR		+3.94	-15	+1.71
	FLDAS-RF		+3.51	-16.02	+2.18

0.05°	AWO-SVM	Regional model	0.00	0.00	0.00
	AWO-PLR		0.00	0.00	0.00
	AWO-GPR		+4.00	-14.83	+1.44
	AWO-RF		+3.85	-14.96	+1.33

Figure 8 demonstrates the correlation between the averaged in-situ precipitation gauges and all the regression-based variants of the downscaled GRACE-TWS for both regional and global hydrological models. it is important to note that the in-situ precipitation observations used in this exercise were heterogeneously distributed with most readings coming from the lower regions. However, they maintained reasonable uncertainty and deemed suitable for our analysis (Table S1). The global model prediction of the partial least square regression (Fig 8d) maintained the strongest correlation between the downscaled and in-situ precipitation observations. Overall, we observed that the signals of the global hydrological models performed slightly better than those of the regional hydrological models in characterizing local-scale water storage changes by their higher correlations with the in-situ estimates (Fig. 8b, d ,f).

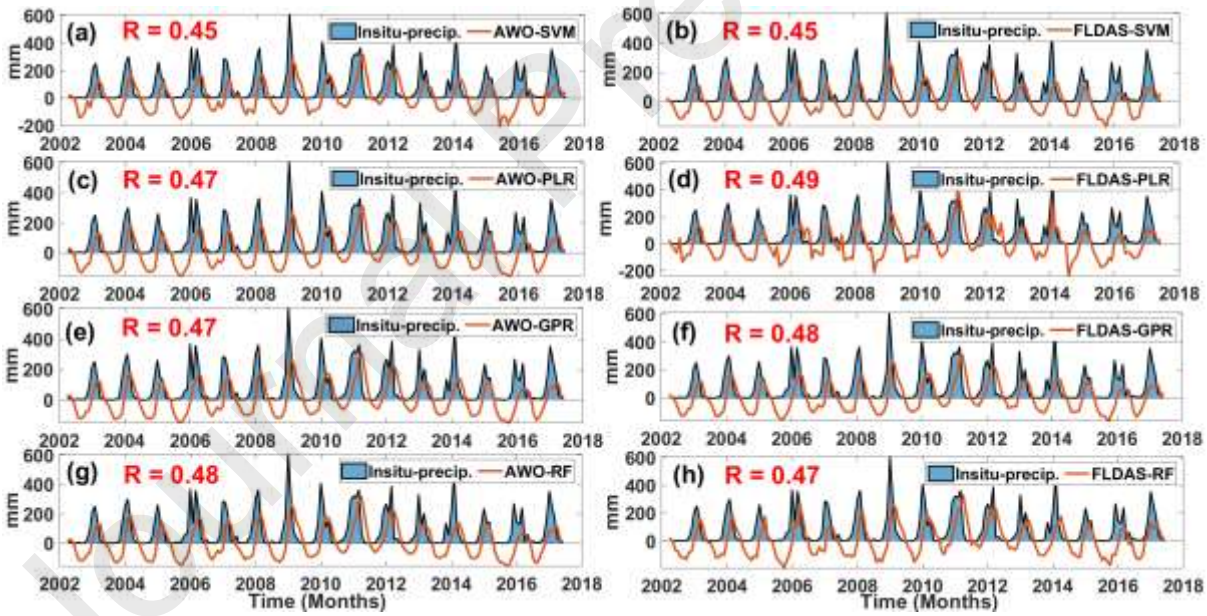


Fig 8. Raw assessment of the collinearity between the downscaled products and the in-situ estimates used for our local-scale assessment. Figures a, c, e, g and b, d, f, h represents the relationship between the in-situ precipitation patterns and the downscaled products from the regional and global hydrological models, respectively.

To better understand the signals at different spatial scales, we utilized the CWPS in identifying the contributions of short- and long-term fluctuations in the time series of the in-situ precipitation and downscaled estimates. The power is given by the colour. The colour ranges from blue (low power) to yellow (high power). The yellow colours represent areas of high power. Figure 9 provides insight to the time-frequency content of the in-situ precipitation

estimates and all the downscaled products at a 5% confidence level. Our result captured a dominant frequency band corresponding to seasonal cycles, semi-annual oscillations, and even shorter-term variations. For example, a noticeable power spike in 2006 is noticed in the CWPS plot for precipitation (Fig. 9a), and the downscaled product that best mimicked this spike was the FLDAS-PLR (Fig. 9g). This agrees with the result in Fig 8d where the PLR produced the strongest correlation with the in-situ precipitation estimates. The high frequency power spike noticed in Fig. 9a is representative of climate actions (such as El Nino Southern Oscillation - ENSO) in regulating precipitation patterns. The spikes noticed in the downscaled products (Fig. 9b-i) are often linked to extreme weather events, such as intense rainfall or drought and the CWPS technique provides a robust way to detect such events and study their frequency characteristics.

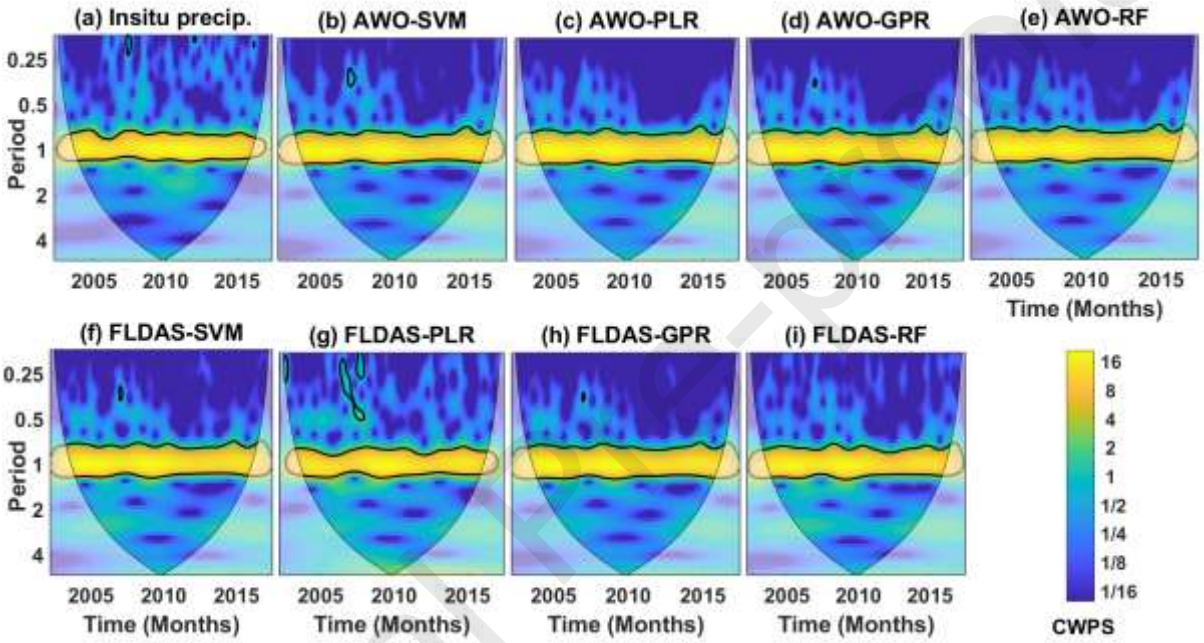
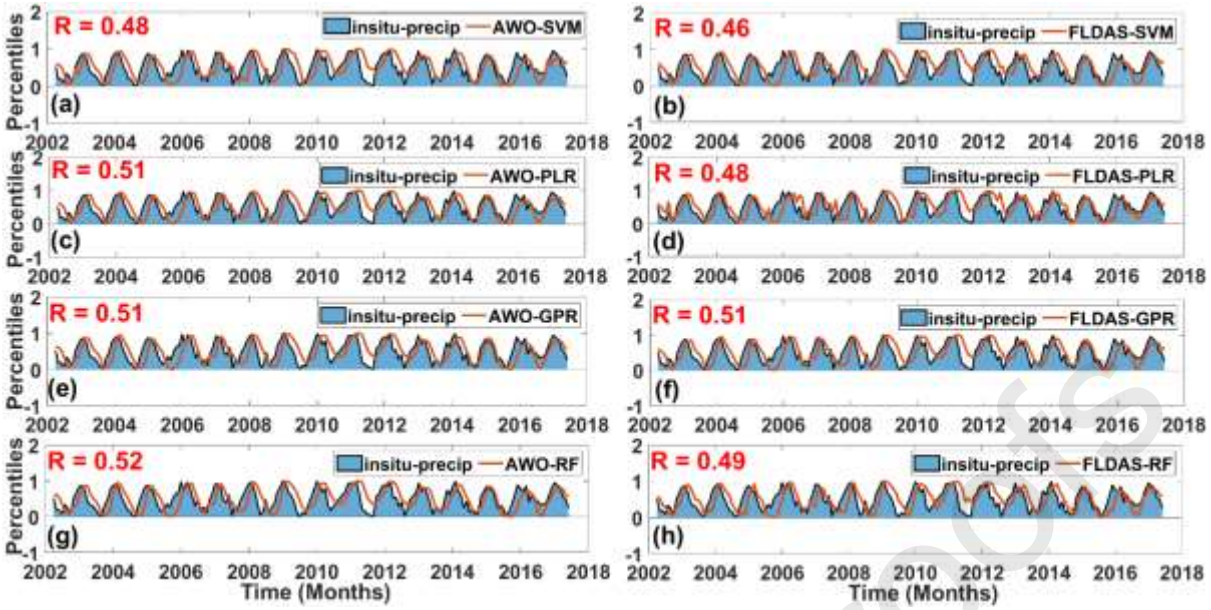


Fig 9. The continuous wavelet power spectrum of the in-situ observations (a), regional-based downscaled estimates (b - e) and the global-based downscaled estimates (f - i). The thick black contour represents the 5% significance level against the noise and cone of influence. NB: CWPS means continuous wavelet power spectrum.

4.2.2 Local scale examination and uncertainty assessment of the downscaled products

The local scale examination of the downscaled products is important to depict which hydrological model best characterizes the hydrology of small catchments. This examination was done by investigating the connection and relationship between the standardized time series in the light of the expected causality links.



577

578 Fig 10. The standardized time series of in-situ precipitation and the region-based (a,c,e,g) and
579 global-based (b,d,f,h) downscaled products. The time series have been transformed into a series
580 of percentiles as shown in the y-axis.

581 We observed a stronger correlation between the standardized in-situ observations and the
582 downscaled products in Fig. 10 than we did in the raw time series in Fig 8. The standardization
583 shown in Fig 10 was achieved by transforming the probability density function (pdf) of the
584 time series observations into a record of percentiles (in terms of its cumulative distribution
585 function) thus forcing the pdf to be rectangular. This has the impact of limiting the width of
586 intermittent oscillations and enhancing model fitting and improving predictive accuracy. The
587 time series observations in Fig. 10 reveals patterns, trends and relationships which were not
588 evident in the time series of Fig 8. Results from the transformed datasets revealed that for the
589 region-based downscaling, the random forest prediction produced the best fitting with R=0.52
590 (Fig. 10g). The gaussian process and partial least squares regression both produced a good
591 fitting index of 0.51, while the support vector machine produced a fitting index of R = 0.48.
592 Whereas, for the global-based downscaling, the gaussian process regression produced the best
593 fitting with R=0.51 followed by the random forest regression, partial least squares regression
594 and the support vector machines with R values of 0.49, 0.48, and 0.46, respectively. This result
595 is in line with accuracy estimation of the machine learning models shown in Fig. 6.

596 Our experiment therefore shows that region-based predictors represent a better characterization
597 and depiction of large-scale hydrological dynamics over their global counterparts. This was
598 further represented in the wavelet coherence plot in Fig. 11. Regions are evident in the time
599 frequency domain depicted in Fig. 11 where the in-situ precipitation and the downscaled time
600 series co-vary but may not necessarily have high power (Adeyeri & Ishola, 2021). The hot
601 colours (yellow) depict regions with a significant coherence, while the cooler colours (blue)
602 represent a lesser degree of coherence between the in-situ precipitation and downscaled
603 estimates. High-frequency powers of 0.8 and 1 are pronounced over the entire plots showing a
604 significant relationship between the standardized precipitation and downscaled TWS series.
605 The span of their periodicities varies for the various machine learning models. It is important
606 to point out that the area of a time frequency plot above the 5% significance level is not an
607 accurate depiction of causality. Even if the scales have been correctly weighted for the
608 averaging, it is still a possibility that the in-situ precipitation and the corresponding downscaled

estimate be perfectly correlated at a particular scale while the area of significant correlation is less than the 5% significance level. Notwithstanding, our results from Fig. 11 suggests that the in-situ precipitation mirrors the corresponding downscaled products which means that our significance level is extensive and suitable for our experiment.

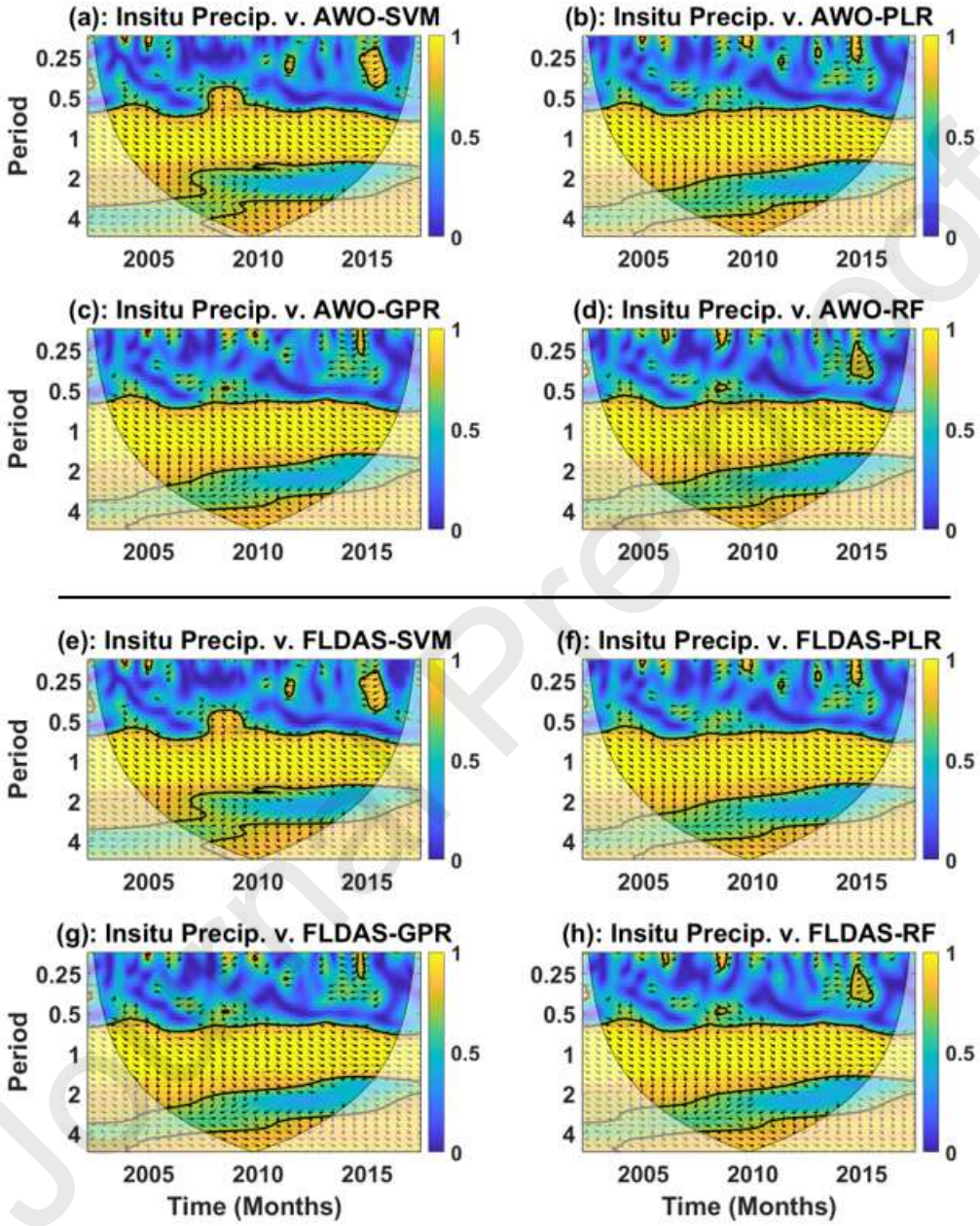


Fig 11. Squared wavelet coherence between the standardized estimates of the insitu precipitation and the region-based (a-d) and global-based (e-h) downscaled products. The 5% significance level against the noise is shown as the polygon contours.

For the regional based models, the SVM showed relatively high uncertainties towards the northern part of the test region (Fig. 12a). These uncertainties are attributed to the strong influence and processes of multi-annual water storage which has been documented to be prevalent in the area (Kaushik et al., 2021). These processes have resulted to complex

properties of the CB as it relates to exchange of fluxes and water storage. For the global based models, the SVM showed a relatively low R^2 value towards its western end, and this scenario was consistent in the GPR and RF algorithms for both regional and global hydrological models (Fig. 12c, d, e, g, h). These low R^2 values could imply strong interactions with non-climatic factors such as land-use and anthropogenic influences. Another plausible explanation could be that the regression models (i.e., SVM, GPR and RF) are not suitable for modelling the location unlike the partial least square regression (Fig. 12b, f) which maintained a strong R^2 value across the entire test region for both regional and global hydrological models. Overall, the four regression models show a good fit and can be relied upon for the statistical downscaling operation of hydrologically dynamic regions.

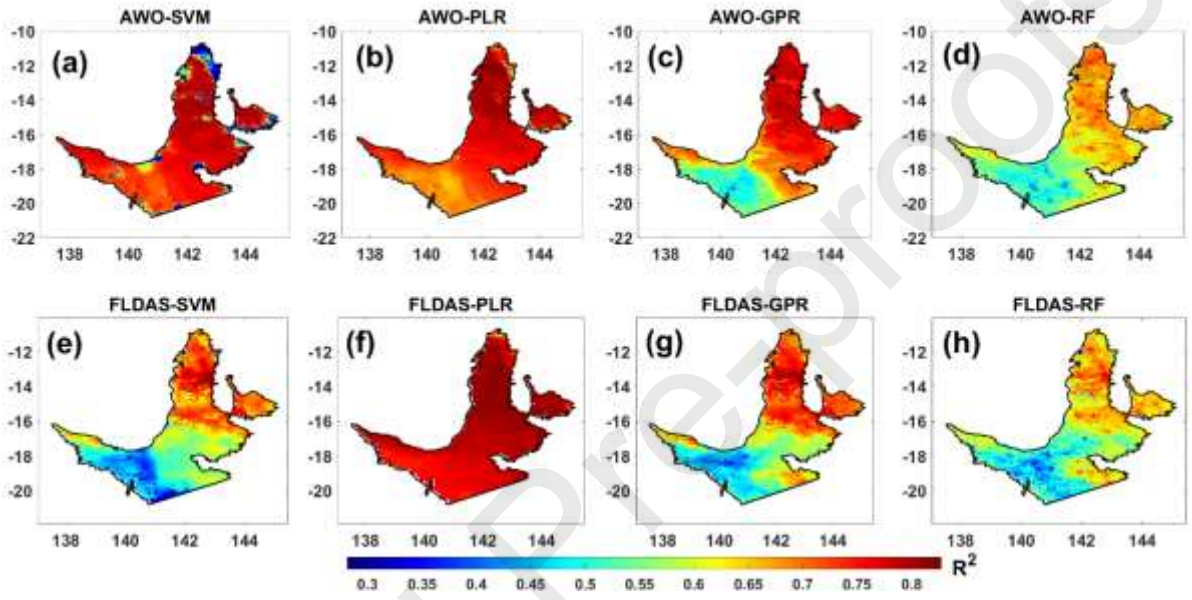
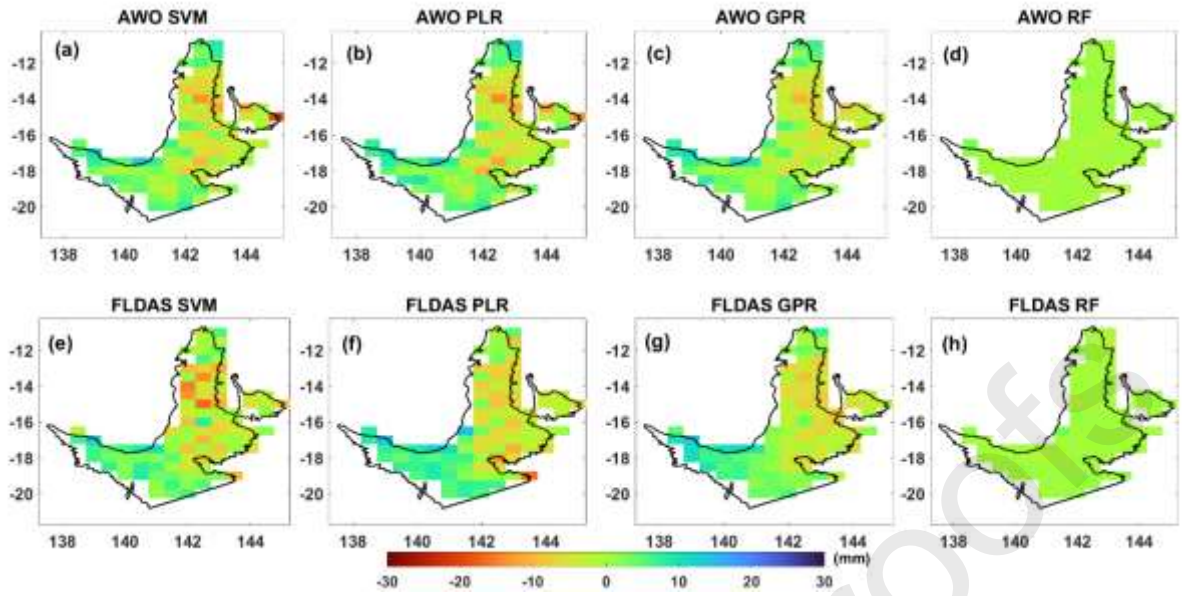


Fig 12. Uncertainty assessment of the downscaled products using bootstrapping. Areas of relatively high uncertainty correspond with lower R^2 values and vice versa.

4.2.3 Residual correction

Residual correction is an important step in statistical downscaling operations. Its primary purpose is to account for finer-scale variations that the initial downscaling model (i.e., predicted TWS at high resolution) may not capture (Fig. 13). Even though the initial downscaling process aims to improve the representation of regional-scale features, it is still composed of large-scale patterns which may not clearly depict fine-scale details of the downscaled product. However, adding back residuals fine-tunes the downscaled estimates by adjusting for unmodeled fine-scale variability and ensures that the final downscaled datasets represents both the regional and local conditions accurately.



644

645 Fig 13. Spatial representation of the residuals between the original and predicted GRACE TWS
646 estimates for the SVM (a,e), PLR (b,f), GPR (c,g), and RF (d,h) machine learning models
647 applied in our study. These residuals account for the complex variabilities that may not be
648 captured in the downscaling operation and must be added back to the predictions to correct
649 systematic bias and ensure a more realistic representation of our downscaled products.

650 The relatively high uncertainties observed in the SVM model (Fig. 12a, e) is made more distinct
651 in Fig. 13a,e showing that the SVM was not able to effectively capture the complex
652 hydrological variabilities exhibited in the study region. The GPR and RF methods had the least
653 residuals across the entire study region for both regional and global models which resulted in
654 the relatively high accuracy achieved in their downscaling and machine learning predictions
655 processes (Fig. 5, 6). This shows that the probabilistic predictions and ensemble learning
656 method of the GPR and RF, respectively, are effective in the modelling and capturing of
657 complex hydrological signals evident in the CB compared to other machine learning methods.

658 4.3 Limitations and future directions

659 Results from our analysis may be affected due to the different sources of uncertainties involved.
660 One of them lie in the inherent processing errors of the original GRACE estimates which was
661 circumvented using mascon solution data, which appears superior to the spherical harmonics
662 data (Aryal & Zhu, 2020; Khorrami et al., 2023). Moreover, our technique of filling the missing
663 monthly data gaps using a linear interpolation approach may induce the associated uncertainties
664 of the GRACE data considering the non-linearity of climate datasets. But given its ease in
665 application and prevalent utilization within the hydrologic community, we reconstructed the
666 missing months using a linear interpolation method. Another issue applies to the uncertainties
667 of the global and regional hydrological models used to extract high-resolution predictors for
668 our downscaling operation (Table S1). Cao et al. (2015) suggests that a key solution in
669 diminishing the uncertainties posed by hydrological model outputs is to use the ensemble mean
670 of several models with the same characteristics. However, due to the lack of many hydrological
671 models simulated at 0.1° and 0.05° spatial resolution for the global and regional models,
672 respectively, we only applied the FLDAS and AWO datasets in this current study.

673 Another issue lies in the interpolation of the residuals to match the spatial grain size of the
674 high-resolution predictors. Examination of residual in statistical downscaling operations help
675 bridge the gap between large-scale climate model outputs and fine-scale observations, capture
676 local variability, maintain statistical relationships, and provide a means for uncertainty
677 assessments. It is important we explore more technical capabilities besides interpolation in
678 treating these residuals as their inclusion to the predicted components ensures accuracy,
679 consistency and makes the downscaled values representative of the original observations.

680 It is important to note that climate actions and anthropogenic influences could introduce further
681 uncertainties to the local-scale assessments of downscaled products. This is due to factors such
682 as urbanization and land-use changes which could easily introduce biases into in-situ
683 precipitation measurements resulting to a modification of local precipitation patterns which
684 may not reflect the true water storage changes of the test region. Therefore, for other
685 experiments that wish to replicate our study, it is important to account for these influences,
686 consider data quality and representativeness, and assess the broader implications of human-
687 induced changes on the hydrological cycle of the test region.

688 **5. Conclusions**

689 The core of downscaling operations lies in the assimilation of additional high-resolution
690 information to improve the estimates of the target variable in the characterization of catchment-
691 scale operations. The task of downscaling is important because water decision makers and
692 managers are usually concerned with water storage changes at scales much smaller than the
693 resolution of GRACE data. It follows therefore that, to realize the full potential of GRACE in
694 hydrology, monthly GRACE-TWS should be downscaled to meet the requirements of local-
695 scale water resource management. This study explored two sets of high-resolution predictors
696 from global (FLDAS) and regional (AWO) hydrological models to downscale GRACE-TWS.
697 Our choice of the different model types was to ascertain which gives a better assessment of
698 hydrological dynamics on a local scale. To achieve this, we validated the downscaled products
699 with in-situ precipitation datasets scatter across the test region. The validation assessment
700 provided insights to which hydrological model best characterizes the dynamics of local-scale
701 hydrology as well as the method. We explored four different machine learning methods in our
702 downscaling operation and tested their prediction capabilities on both models.

703 The machine learning methods (SVM, PLR, GPR, RF) are used to downscale GRACE-TWS
704 using a regression-based approach from the spatial resolution of 0.5º to 0.1º and 0.05º for the
705 global and regional hydrological model estimates, respectively. It was important to measure
706 how well each regression model performed in (i) establishing a functional regression model
707 between predicted and observed TWS estimates, and (ii) predicting high resolution GRACE
708 TWS using the model parameters. To achieve this, we gauged the performance of each learning
709 using statistical metrics prior to adding back the residuals. This was warranted because the
710 predicted values without the residuals represent the systematic trend and relationships captured
711 by the individual learning models and provides a framework to gauge their unalloyed
712 performances. The GPR made the best prediction across board with RMSE and MAE values
713 of 1.44mm, 0.80mm for the regional model and 7.20mm, 4.63mm for the global model
714 estimates. This gives credence to the probabilistic framework of the GPR in model data
715 assessments and its ability to capture spatial correlations and temporal dependencies prominent
716 in hydrological datasets.

717 Overall, the findings of our study reports that the region-based downscaled TWS using the
718 random forest method provided the best characterization of local scale hydrology over our test

region. This was because it maintained the most significant correlation to the in-situ precipitation values. To get the most out of this relationship, we transformed the probability density function of the time series for the in-situ precipitation and all the downscaled estimates into a record of percentiles. This helped force the pdf into a rectangular form thus limiting the impact of intermittent oscillations dominant in hydrological variables and improving their respective predictive accuracies. Another advantage of this operation is that it has not outliers. Generally, our downscaling algorithm is efficient and reliable over out test region, where the primary assumptions were satisfied. This low-cost approach provides insights to which model and method best suits downscaling over regions with significant hydrological dynamics. This study basically focused on the technique and datasets used for statistical downscaling, and the use of a more robust, continuous and comprehensive in-situ precipitation network can be further investigated to combat the biases posed by land-use changes and urbanization in the validation of local-scale water resources in future research.

Acknowledgements

The authors are grateful to the Australian Bureau of Meteorology (BOM), NASA, and the Joint Research Centre for all the data (e.g., FLDAS, AWO, GRACE and rainfall) used in this study. Ikechukwu Kalu received funding from Griffith University Postgraduate Research Scholarships. Christopher E. Ndehedehe is supported by the Australian Research Council Discovery Early Career Researcher Award (DE230101327) for the project, *Assessing the impacts of drought and water extraction on groundwater resources in Australia.*

References

Adeyeri, O. E., & Ishola, K. A. (2021). Variability and Trends of Actual Evapotranspiration over West Africa: The Role of Environmental Drivers. *Agricultural and Forest Meteorology*, *308*, 108574.

Ahmed, K., Shahid, S., Haroon, S. B., & Xiao-Jun, W. (2015). Multilayer perceptron neural network for downscaling rainfall in arid region: a case study of Baluchistan, Pakistan. *Journal of Earth System Science*, *124*, 1325-1341.

Ali, S., Liu, D., Fu, Q., Cheema, M. J. M., Pal, S. C., Arshad, A., ... & Zhang, L. (2022). Constructing high-resolution groundwater drought at spatio-temporal scale using GRACE satellite data based on machine learning in the Indus Basin. *Journal of Hydrology*, *612*, 128295.

Arsenault, R., Brissette, F., & Martel, J. L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of hydrology*, *566*, 346-362.

Arshad, A., Mirchi, A., Samimi, M., & Ahmad, B. (2022). Combining downscaled-GRACE data with SWAT to improve the estimation of groundwater storage and depletion variations in the Irrigated Indus Basin (IIB). *Science of the Total Environment*, *838*, 156044.

Aryal, Y., & Zhu, J. (2020). Multimodel ensemble projection of meteorological drought scenarios and connection with climate based on spectral analysis. *International Journal of Climatology*, *40*(7), 3360-3379.

Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *22*, 106-114.

760 Beecham, S., Rashid, M., & Chowdhury, R. K. (2014). Statistical downscaling of multi-site
761 daily rainfall in a South Australian catchment using a Generalized Linear
762 Model. *International journal of climatology*, *34*(14), 3654-3670.

763 Cao, Y., Nan, Z., & Cheng, G. (2015). GRACE gravity satellite observations of terrestrial
764 water storage changes for drought characterization in the arid land of northwestern
765 China. *Remote Sensing*, *7*(1), 1021-1047.

766 Chen, H., Sun, Y., Gao, J., Hu, Y., & Yin, B. (2018). Solving partial least squares regression
767 via manifold optimization approaches. *IEEE transactions on neural networks and learning*
768 *systems*, *30*(2), 588-600.

769 CSIRO (2019). Water in the Gulf of Carpentaria Drainage Division: Summary of a report to
770 the Australian Government from the CSIRO Northern Australia Sustainable Yields Project.
771 https://publications.csiro.au/rpr/download?pid=procite:1c140d75-b30e-491b-b2f2-
772 c83ed659cbbd&dsid=DS1

773 Duan, S. B., & Li, Z. L. (2016). Spatial downscaling of MODIS land surface temperatures
774 using geographically weighted regression: Case study in northern China. *IEEE Transactions*
775 *on Geoscience and Remote Sensing*, *54*(11), 6458-6469.

776 Fatolazadeh, F., & Goïta, K. (2021). Mapping terrestrial water storage changes in Canada
777 using GRACE and GRACE-FO. *Science of The Total Environment*, *779*, 146435.

778 Frost, A. J., and Shokri, A., (2021). The Australian Landscape Water Balance model (AWRA-
779 L v7). Technical Description of the Australian Water Resources Assessment Landscape model
780 version 7. https://awo.bom.gov.au/assets/notes/publications/AWRA-
781 Lv7_Model_Description_Report.pdf.

782 Goly, A., Teegavarapu, R. S., & Mondal, A. (2014). Development and evaluation of statistical
783 downscaling models for monthly precipitation. *Earth Interactions*, *18*(18), 1-28.

784 Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform
785 and wavelet coherence to geophysical time series. *Nonlinear processes in*
786 *geophysics*, *11*(5/6), 561-566.

787 Güntner, A. (2008). Improvement of global hydrological models using GRACE data. *Surveys*
788 *in geophysics*, *29*, 375-397.

789 He, H., Yang, K., Wang, S., Petrosians, H. A., Liu, M., Li, J., ... & Li, J. (2021). Deep
790 learning approaches to spatial downscaling of GRACE Terrestrial Water Storage Products
791 using EALCO Model over Canada. *Canadian Journal of Remote Sensing*, *47*(4), 657-675.

792 Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping
793 at a regional-scale: A Random Forest approach. *Geoderma*, *214*, 141-154.

794 Kalu, I., Ndehedehe, C. E., Okwuashi, O., & Eyoh, A. E. (2021). Assessing freshwater
795 changes over Southern and Central Africa (2002–2017). *Remote Sensing*, *13*(13), 2543.

796 Kalu, I., Ndehedehe, C. E., Okwuashi, O., Eyoh, A. E., & Ferreira, V. G. (2022). An
797 assimilated deep learning approach to identify the influence of global climate on hydrological
798 fluxes. *Journal of Hydrology*, *614*, 128498.

Kalu, I., Ndehedehe, C. E., Okwuashi, O., Eyoh, A. E., & Ferreira, V. G. (2023). Reconstructing terrestrial water storage anomalies using convolution-based support vector machine. *Journal of Hydrology: Regional Studies*, *46*, 101326.

Kaushik, P. R., Ndehedehe, C. E., Burrows, R. M., Noll, M. R., & Kennard, M. J. (2021). Assessing Changes in Terrestrial Water Storage Components over the Great Artesian Basin Using Satellite Observations. *Remote Sensing*, *13*(21), 4458.

Khandu, Forootan, E., Schumacher, M., Awange, J. L., & Mueller Schmied, H. (2016). Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the G anges-B rahmaputra-M eghna River Basin. *Water Resources Research*, *52*(3), 2240-2258.

Khorrami, B., Pirasteh, S., Ali, S., Sahin, O. G., & Vaheddoost, B. (2023). Statistical Downscaling of GRACE TWSA Estimates to a 1-km Spatial Resolution for a Local-scale Surveillance of Flooding Potential. *Journal of Hydrology*, 129929.

Lehmann, F., Vishwakarma, B. D., & Bamber, J. (2022). How well are we able to close the water budget at the global scale?. *Hydrology and earth system sciences*, *26*(1), 35-54.

Li, Z., Shao, Q., Xu, Z., & Cai, X. (2010). Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. *Journal of Hydrology*, *385*(1-4), 76-83.

Long, D., Yang, Y., Wada, Y., Hong, Y., Liang, W., Chen, Y., ... & Chen, L. (2015). Deriving scaling factors using a global hydrological model to restore GRACE total water storage changes for China's Yangtze River Basin. *Remote Sensing of Environment*, *168*, 177-193.

Luthcke, S. B., Sabaka, T. J., Loomis, B. D., Arendt, A. A., McCarthy, J. J., & Camp, J. (2013). Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology*, *59*(216), 613-631.

Meng, F., Su, F., Li, Y., & Tong, K. (2019). Changes in terrestrial water storage during 2003–2014 and possible causes in Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *124*(6), 2909-2931.

Milewski, A. M., Thomas, M. B., Seyoum, W. M., & Rasmussen, T. C. (2019). Spatial downscaling of GRACE TWSA data to identify spatiotemporal groundwater level trends in the Upper Floridan Aquifer, Georgia, USA. *remote sensing*, *11*(23), 2756.

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.

Ndehedehe, C. (2022a). Satellite Observations of Terrestrial Water Storage. In: Satellite Remote Sensing of Terrestrial Hydrology. Springer, Cham. https://doi.org/10.1007/978-3-030-99577-5_14

Ndehedehe, C. (2022b). Drought Events. In: Satellite Remote Sensing of Terrestrial Hydrology. Springer, Cham. https://doi.org/10.1007/978-3-030-99577-5_11

Ndehedehe, C. E., & Ferreira, V. G. (2020). Assessing land water storage dynamics over South America. *Journal of Hydrology*, *580*, 124339.

Ndehedehe, C. E., Ferreira, V. G., Agutu, N. O., Onojeghuo, A. O., Okwuashi, O., Kassahun, H. T., & Dewan, A. (2021). What if the rains do not come?. *Journal of Hydrology*, *595*, 126040.

Ndehedehe, C. E., Onojeghuo, A. O., Stewart-Koster, B., Bunn, S. E., & Ferreira, V. G. (2021). Upstream flows drive the productivity of floodplain ecosystems in tropical Queensland. *Ecological indicators*, *125*, 107546.

Ning, S., Ishidaira, H., & Wang, J. (2014). Statistical downscaling of GRACE-derived terrestrial water storage using satellite and GLDAS products. Proceedings of the civil society B1 (Hydraulic Engineering), *70*(4), I_133-I_138.

Pielke Sr, R. A., & Wilby, R. L. (2012). Regional climate downscaling: What's the point?. *Eos, Transactions American Geophysical Union*, *93*(5), 52-53.

Rahman, M., Tumon, M. S. H., Islam, M. M., Chen, N., Pham, Q. B., Ullah, K., ... & Dewan, A. (2023). Could climate change exacerbate droughts in Bangladesh in the future?. *Journal of Hydrology*, *625*, 130096.

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

Roy, A., & Chakraborty, S. (2023). Support vector machine in structural reliability analysis: A review. *Reliability Engineering & System Safety*, 109126.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Shahid, S., & Perera, B. J. C. (2018). Statistical downscaling of precipitation using machine learning techniques. *Atmospheric research*, *212*, 240-258.

Sachindra, D. A., Huang, F., Barton, A. F., & Perera, B. J. C. (2014). Multi-model ensemble approach for statistically downscaling general circulation model outputs to precipitation. *Quarterly Journal of the Royal Meteorological Society*, *140*(681), 1161-1178.

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.

Shah, D., & Mishra, V. (2021). Strong influence of changes in terrestrial water storage on flood potential in India. *Journal of Geophysical Research: Atmospheres*, *126*(1), e2020JD033566.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*, 199-222.

Song, C., Ke, L., Huang, B., & Richards, K. S. (2015). Can mountain glacier melting explains the GRACE-observed mass loss in the southeast Tibetan Plateau: From a climate perspective?. *Global and Planetary Change*, *124*, 1-9.

Tao, H., Al-Sulttani, A. H., Salih, S. Q., Mohammed, M. K., Khan, M. A., Beyaztas, B. H., ... & Yaseen, Z. M. (2023). Development of high-resolution gridded data for water availability identification through GRACE data downscaling: Development of machine learning models. *Atmospheric Research*, *291*, 106815.

Van Dijk, A. (2010). The Australian Water Resources Assessment System. Landscape Model (version 0.5). Technical Report 3. CSIRO. Australia. https://awo.bom.gov.au/assets/notes/publications/Van_Dijk_AWRA05_TechReport3.pdf

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Vishwakarma, B. D., Zhang, J., & Sneeuw, N. (2021). Downscaling GRACE total water storage change using partial least squares regression. *Scientific data*, *8*(1), 95.

Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth*, *120*(4), 2648-2671.

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., & Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA*, *27*.

Yan, B., Fang, N. F., Zhang, P. C., & Shi, Z. H. (2013). Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression. *Journal of Hydrology*, *484*, 26-37.

Yin, W., Hu, L., Zhang, M., Wang, J., & Han, S. C. (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.

Yin, W., Hu, L., Zheng, W., Jiao, J. J., Han, S. C., & Zhang, M. (2020). Assessing underground water exchange between regions using GRACE data. *Journal of Geophysical Research: Atmospheres*, *125*(17), e2020JD032570.

- RHMs better characterizes the dynamics of local scale hydrology more than GHMs
- Probabilistic framework of the GPR improves its ability in hydrological modelling.
- In-situ precip. provides a key metric for local-scale water storage assessments.
- Regression downscaling effectively enhances GRACE-TWS res. for local evaluations

CRediT authorship contribution statement

Ikechukwu Kalu: Writing – original draft, Writing – review & editing, Methodology, Formal analysis, Conceptualization, Investigation, Software. **Christopher E. Ndehedehe:** Conceptualization, Supervision, Software, Investigation, Formal analysis, Methodology, Writing – original draft, Writing –

review & editing. **Vagner G. Ferreira:** Supervision, Writing – original draft, Writing – review & editing, Formal analysis. **Mark J. Kennard:** Conceptualization, Supervision, Software, Writing – review & editing.

Abstract

The coarse spatial resolution of the Gravity Recovery and Climate Experiment (GRACE) data has limited its application in the management of local-scale water resources. To address this limitation, we developed a new downscaling approach using predictors from regional and global hydrological models for a 15-year period (2002-2017) and tested it in the northern Great Artesian Basin, Australia. We used four different machine learning algorithms (support vector machine, partial least squares, gaussian process and random forest) to downscale the original GRACE estimate of 0.5° to a spatial grain size of 0.1° (global) and 0.05° (regional). This was based on precipitation, evapotranspiration and runoff estimates from the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) and Australian Water Outlook (AWO) hydrological models, respectively. The downscaled products were validated using 42 in-situ precipitation observations spread across the test region. We further evaluated which of the downscaled products best mimicked local-scale hydrology using a range of statistical metrics. Our results showed that regional hydrological models best characterized the dynamics of local scale hydrology (rainfall v. downscaled product), and the gaussian process regression algorithm made the best predictions for both models. The correlation coefficients for the raw values varied from 0.45 to 0.49 while that of the standardized values varied from 0.46 to 0.52 with the random forest model providing the best fitting for the regional-based products. The regional downscaling approach employed in this study may be readily integrated into local water resources planning programs.