Journal of Contemporary Researchin the Built Environment ISSN: 2636-4468
Volume 9, Number 1, March 2025 © Department of Building, University of Uyo
DOI: https://doi.org/10.68128/jocrebe.2025.49icwzn2

RECONSTRUCTION OF GRACE-LIKE TWSA IN SUB-SAHARAN AFRICA USING GRACE, GLDAS PRODUCTS AND GLOBAL TELECONNECTION PATTERNS

John O. Anwana[1] and Aniekan E. Eyoh[2]
[1&2]Department of Geoinformatics and Surveying, University of Uyo, Akwa Ibom State, Nigeria

ABSTRACT

Purpose: Changes in Terrestrial Water Storage (TWS) are strategic and essential to evaluate water resources, climatic extreme impact and agricultural production. Presently, products from Gravity Recovery and Climate Experiment (GRACE) twin satellites are distinctive method of examining regional to global hydrological fluxes. However, short timeframe (2002 – till date) restricts their use in prolonged climatologically and hydrological projects. Secondly, the occasional temporal gaps in the GRACE products hinder continuity in time series analysis. Progress in artificial intelligence make possible customized learning of inter-relationship between variables in intricate spatio-temporal systems. Nonetheless, the relevance of many deep learning procedures has neither been sufficiently learned for GRACE TWS reconstruction.
Design/methodology/approach: The study develops a reconstruction model for GRACE TWS anomalies (TWSA) based on climate dataset from Land surface model (GLDAS) products. The climate dataset was further optimized by statistical technique of Independent Component Analysis, Principal component analysis and spatio-temporal decomposition technique. The study adopted Deep Belief Network (DBN) deep learning and C4.5 data mining models. Long-term TWSA of sub-Saharan Africa from 2002 to 2022 was simulated and implemented on the above mentioned models and results compared with GRACE-derived TWSA observations. Validation metrics were used to assess the robustness of reconstruction.
Findings: Trained DBN model achieve a range correlation coefficient between 0.789 and 0.92 which was better than that of C4.5 and a range RMSE between 10.98 and 20.19 mm and mean absolute error range between 8.31 and 15.81 lower than that of C4.5 indicating a good agreement with observed GRACE TWSA.
Research limitations/Implications: The limitation and implication is the variation of increased bias and percolation rectification blunders and changes in mountainous and coastal sub-regions.
Practical implications: It is recommended that advance deep learning prediction model be adopted for modeling TWS reconstruction.
Originality/value – The outcome shows that DBN model outperforms C4.5 model. Therefore, this study provides an alternative method required in generating Terrestrial Water Storage signals using climate dataset.

Keywords: Terrestrial water Storage Anomalies (TWSA), GRACE, Deep Belief Network (DBN), GADAS, Sub-Saharan Africa.

[2] Email: aniekaneyoh@uniuyo.edu.ng

1. INTRODUCTION

Terrestrial water storage (TWS) is the summation of water found on land surface, underground surface, tree canopies and root zone, rivers and snow (Rodell *et al*., 2004; Li *et al*, 2012; Long *et al*., 2017; Jing *et al*., 2020). TWS constitute a major component of regional hydrological fluxes. Its fluctuations when examined through complex environmental variables such as climate change and anthropogenic factors is important in determining water resources availability and sustenance (Famiglietti *et al.*, 2011). Therefore, for efficient and timely assessment of regional water cycle modulation, changes in TWS signals are key parameters to be applied. The Gravity Recovery and Climate Experiment (GRACE) twin satellites are used to derive estimates of TWS changes (Landerer and Swenson, 2012; Syed *et al*., 2008), for hydrology studies, drought and flood detection evaluation studies (Chen *et al*., 2009; Frappart *et al*., 2012; Long *et al*., 2014; Forootan *et al*., 2019), and for constraining water storage models (Khaki *et al*., 2017; Richler *et al.*, 2008). Long-term variation information of the TWS is advantageous for complex hydrological assessment and prognosis (such as drought and flood. The technique is of great importance for managing water resources, regional water cycle and regional climatology (Yang *et al*., 2017; Chandan and Nagesh, 2018). GRACE-derived products are of great benefit to water resources studies, essentially as a consequence of direct provision of TWS changes estimations (Nie *et al*., 2016; Scanlon *et al*., 2012; Mo *et al*., 2016). However, applying TWS products from GRACE in hydrological processes are limited by short period of availability (2002-present) and noticeable temporal-based missing data gaps. Gywali *et al*. (2022) reported on significant temporal missing data gaps of about 20 months within GRACE products, missing data gap of about 11 months between GRACE and GRACE Follow-On products and within GRACE-FO products recorded 2 months of temporal missing data gaps. These inherent temporal missing data gaps make GRACE-derived TWS products strenuous to evaluate.

The increase availability of climate data set through remote sensing methods, has facilitated the reconstruction of long-term GRACE-like TWS products through models' simulations. However, complexity and non-linearity of climate data greatly hinders the capability of most hydrological models to simulate accurate conditions (Jing *et al.*, 2020). To address these challenges, an innovative approach of data optimization by statistical technique was incorporated into this study. Reconstruction of TWS with climate based data using machine learning algorithm has witnessed increasing concern in the last few years (Becker *et al*., 2011; Long *et al*., 2014, Nie *et al*., 2015, Yang *et al*., 2018; Gywali *et al*., 2022). Combining data-driven machine learning models into the TWSA reconstruction is advantageous compared to the uncertainty found in physical model and ineffectiveness noticed in statistical conventional methods (Kalu *et al*., 2022 Ndehedehe., 2023; Yang *et al*., 2020).

1.1. Objectives of the Study

Considering the background above the specific objectives are to:

i. Examine influence of climate parameters on climatology of sub-Saharan Africa.

ii. Recreate GRACE-Like TWS of sub-Saharan Africa based on significant hydrological parameters.

iii. Determine the geographical locations (hotspot) in sub-Saharan Africa most prone to the impact of extreme TWS variability.

iv. Evaluate the accuracy of the TWS reconstruction over sub-Saharan Africa based on the C4.5 and Deep Belief Network (DBN) reconstruction models.

Aim at reconstructing Terrestrial Water Storage (TWS) over sub-Saharan Africa using C4.5 data mining model and Deep Belief Network deep learning model.

1.2. Research Questions:

This research was guided by the following vital scientific questions:

i. What are the climate parameters that influence climatology in study region?

ii. What are the significant hydrological parameters in recreating TWS?

iii. What regions (hotspots) of sub-Saharan Africa are more prone to extreme TWS variability?

iv. How can C4.5 data mining and DBN deep learning approaches effectively model and recreate TWS over sub-Saharan Africa?

These questions were answered by analysing the climate input as independent variables, regional hydrological model and TWS variations from GRACE satellite observations as dependent variable.

The computational ability of DBN reconstruction methods with its inherent capabilities is currently being employed in hydrological studies. On the other hand, C4.5 data mining methods possesses the capabilities of interpretability, reconstruction of missing data and simplicity, yet tends to over fit on datasets with many features. Hence, employing data optimization techniques like PCA and ICA enhances the model's performance (Agana *et al*., 2018; Gyaneshwar *et al*., 2023; Kalu *et al.*, 2022; Liu *et al.*, 2021; Ndehedehe., 2023; Yang *et al*., 2020; Sun *et al*., 2019).

2. Literature Review

2.1. Area of the Study

Sub-Saharan Africa which is commonly referred to as black Africa is geographically located between latitude 27° 25´N and 34° 5´S and longitude 51° 28´E and 17°39´W. Sub-Sahara Africa is made up of all African Countries located on the southern part of Sahara (United Nation report, 2021; Stockdale, 2017). The region is reported to be most vulnerable to impacts of climate change because of its geographical location being surrounded by major Oceans and the Sahel in the north. The area is prone to climatic extreme events such as flood and drought and has low capacity for adaptation (Stockdale, 2017).



Figure 1: Map showing Study Area

Source: PBS Learning Media and updated by the researcher (2023)

Figure 1 shows study area with boundaries marked red. One third of the world drought occurs in this region and the frequency of storms and flood is growing fastest in this region. The annual average rainfall over sub-Saharan Africa ranges from 151mm to 3200mm and the annual average temperature range of 15˚C to 33.9˚C in summer (Nwaerema, *et al.*, 2019; Stockdale, 2017). Sub-Saharan Africa has tropical and sub-tropical climate, bordering Atlantic Ocean in the west and Indian Ocean in the East. These surrounding oceans are primary producers of changing climate drivers, as well as large-scale convective systems which stimulate extreme hydro-climatic conditions and influences meteorological processes in sub-regional and global hydrological system. These global climate teleconnections and the large-scale ocean-land-atmosphere interactions over the African continent have been known to bring about intense climate events (Ndehedehe *et al*., 2017) and are known to greatly influence TWS variability.

2.1. Data Sources

Secondary data utilized for this study were processed satellite remote sensing data collected from America National Oceanic and Atmospheric Administration (NOAA) and Gravity Recovery and Climate Experiment (GRACE) derived TWS from University of Texas, Centre for Space Research, complimented by gauges-based in-situ field observations from meteorological stations over sub-Saharan Africa, coordinated and generated by credible international research laboratories as recorded in Table 1 below.

Table 1: Sources of research data

Data	Unit	Sources
GRACE derived TWS	mm	University of Texas centre for space research
Precipitation	mm	Global Land Data Assimilation System
Evapotranspiration	mm	Global Land Data Assimilation System
Runoff	mm	Global Land Data Assimilation System
Land Surface Temperature	kelvin	Global Land Data Assimilation System
Soil Moisture	mm	Global Land Data Assimilation System
El Nino Southern Atlantic Oscillation (ENSO)	standardize	NOAA
North Atlantic Oscillation (NAO)	standardize	NOAA
Atlantic Multidecadal Oscillation (AMO)	standardize	NOAA
Indian Ocean Dipole (IOD)	standardize	NOAA
Nino 1+2	standardize	NOAA
Nino 3	standardize	NOAA
Nino 3.4	standardize	NOAA
Nino 4	standardize	NOAA
Pacific Decadal Oscillation (PDO)	standardize	NOAA
Tropical North Atlantic (TNA	standardize	NOAA
Tropical South Atlantic (TSA)	standardize	NOAA
Atlantic Meridional Mode (AMM)	standardize	NOAA

2.1.1 GRACE Derived TWS

TWS observations from GRACE are used in validating TWS simulated from recreation models for the study period. Observation data from three production laboratories: Jet Propulsion Laboratory (JPL), Center for Space Research (CSR) and German Research Center for Geosciences (GFZ) (Landerer and Swenson, 2012), were used including two mascon products from JPL and CSR; mascon products are known to improves capturing of

TWS estimates in various regions (Scanlon et al., 2016). The long-term changes were also computed for observed monthly TWS of (2002-2022) by subtracting the mean of TWS of (2002-2009) to obtained terrestrial water storage anomalies (TWSA). Data unit is in mm/month, Figure 2 shows the spatial pattern of bi-annual GRACE signal during the study period.

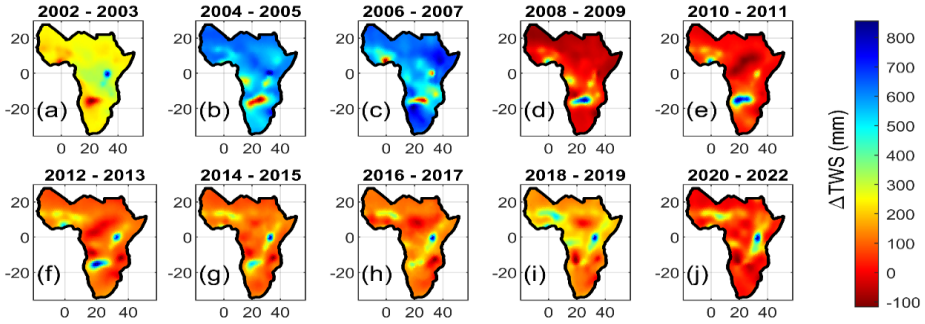


Figure 2: GRACE TWSA signals in the study region

(Gridded maps are in degrees)

2.1.2 Hydrological Parameters

Global Land Data Assimilation System (GLDAS) combines multiple Land Surface Models (LSMs) with large amount of observation-based data, facilitated by the Land Information System. GLDAS drives four Land Surface Models (LSMs): Mosaic, Noah, CLM, and VIC. Averaging of the 3-hourly products aggregated to Monthly products. GLDAS provide high-resolution estimates of land surface conditions and fluxes (Rodell *et al.*, 2004).

Currently, products with spatial resolution of both 0.25° and 1° are being provided. The four models boast of 1.0° data range from 1979 to present, while the 0.25° data cover 2000 to present only for the Noah model. GLDAS has two versions of products, version 1 (GLDAS-1) and the more recent version 2 (GLDAS-2). GLDAS-2 is climatologically more reliable than GLDAS-1. For this study, GLDAS products were interpolated geographically to become comparable and were also averaged to redistribute the inconsistencies inherent in the predictor datasets used for this study. The following monthly GLDAS climate products were used in the study: Precipitation, Evapotranspiration, Runoff, including Land surface temperature and Soil Moisture. Figure 3 depict the spatial patterns of above mentioned hydro-climate parameters while Figure 4 shows the temporal evolution of same and Figure 5 shows the decomposed hydro-climate predictors variables, integrated with empirical orthogonal function (EOF).

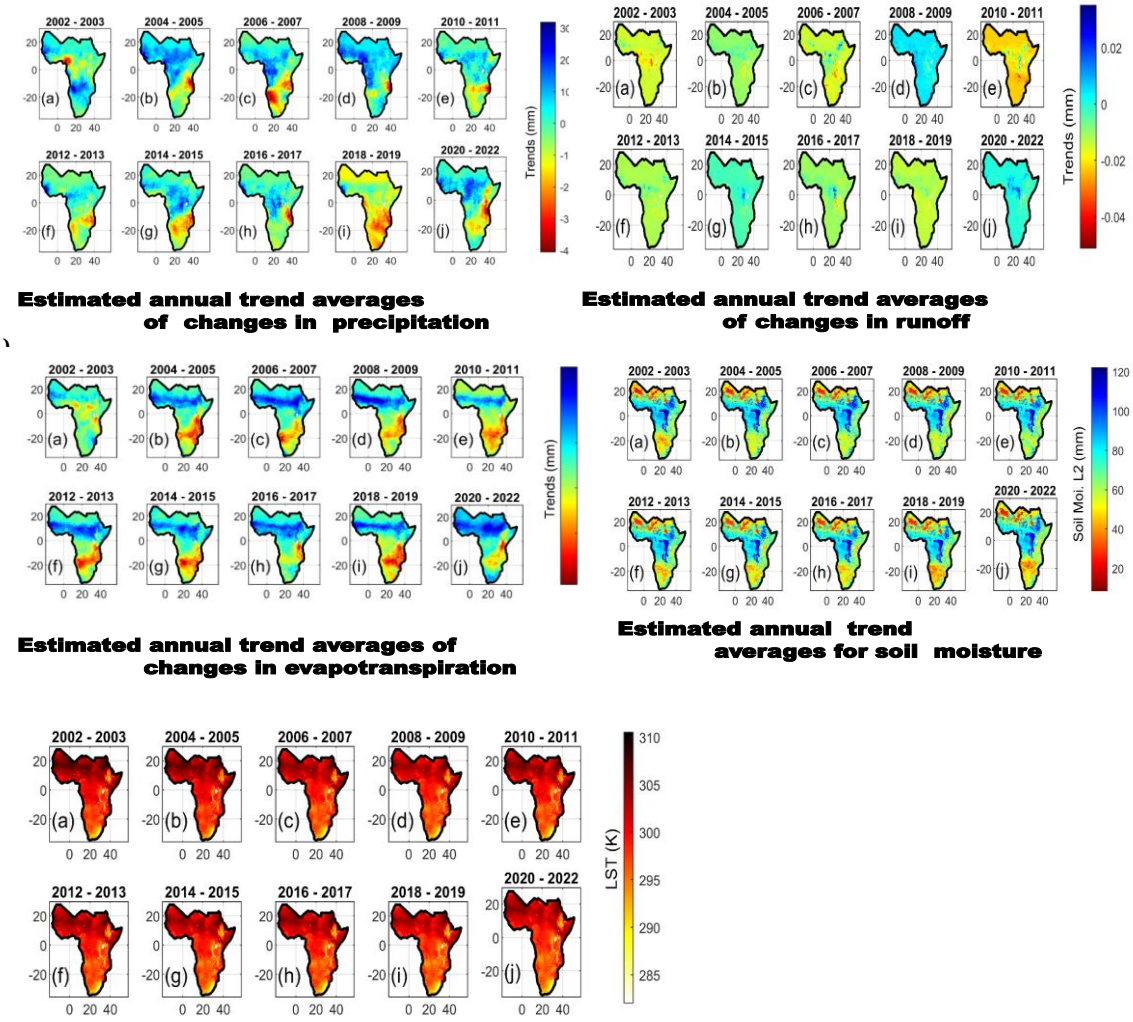


Figure 3: GLDAS Climate data Products (Spatial Pattern)
(The gridded maps are in degrees)

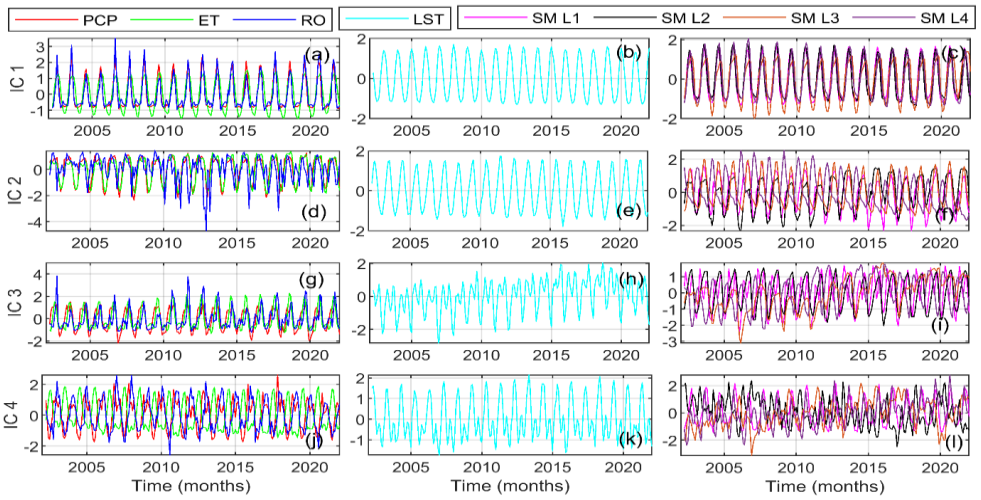


Figure 4: Time series of Precipitation, Evapotranspiration, Runoff, (a), (d), (g), (j), LST (b), (e), (h), (k) and soil moisture components (c), (f), (i), (l) (Temporal Pattern)

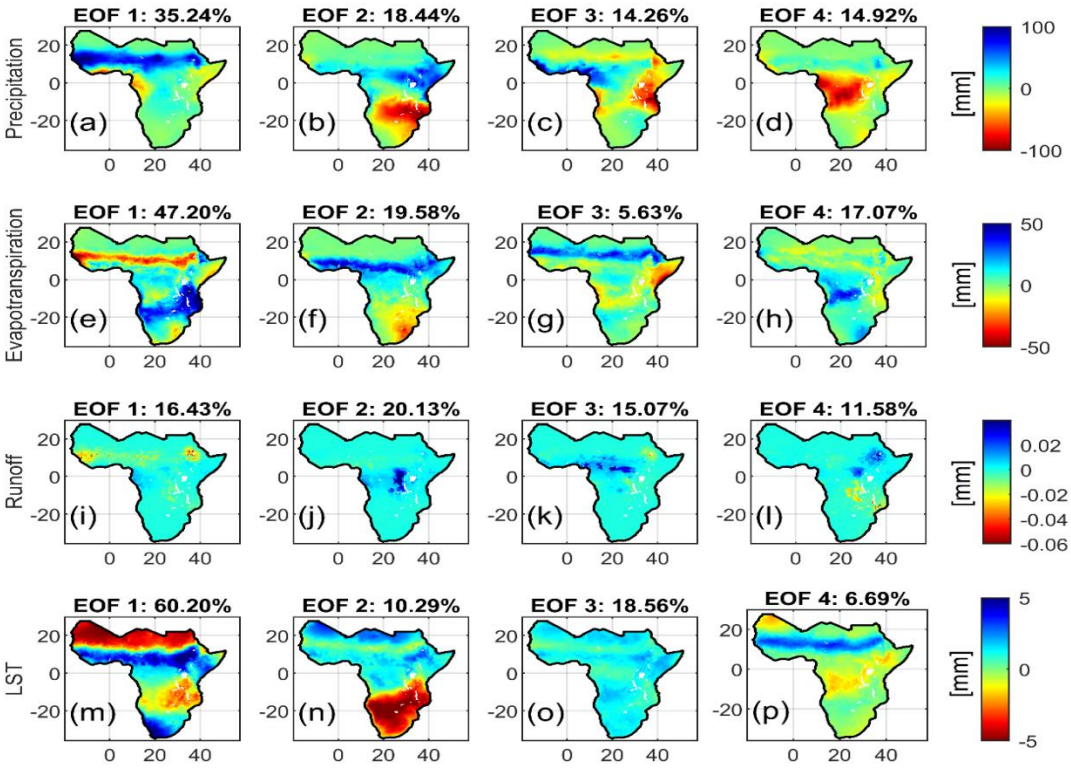


Figure 5: ICA decomposition of the key water budget terms and land surface temperature changes showing spatial patterns of variations

(The gridded maps are in degrees).

3. METHODOLOGY

3.1. Technique of Data Analysis

Two technique of statistical analysis procedures; Principal and Independent component analysis aimed at data dimensionality reduction, noise removal and interpretability upgrade were employed to facilitate improved GRACE-Like Reconstruction accuracy.

i. Principal Component Analysis (PCA): More commonly applied statistical procedure for data investigation and examination (Jolliffe, 2002). PCA pertain to lineage of data compression methods and is extremely useful when data examine are very large (that is multivariable), big (that is repeated measurements per variable), and thoroughly be in agreement or correlated. Goal of the exercise is to identify and reduced set of features that constitute the original data in a decreased-dimensional subset with a negligible loss of information. Some of the merits of working with a reduced dataset instead of the original high-dimensional data includes: (i) Potential to apprehend data in 2D or 3D, (ii) Reduced storage space, (iii) Elimination of co-linearity and (iv) Reduction of noise.

Being a statistically dimension reduction procedure, PCA transform data into a new, lower-dimensional subspace from the original observations to perhaps non-correlated variables referred to as principal components. Particularly, the original data are extrapolated into a new coordinate system such that the first principal axis agree with the direction along which data vary the most; second principal axis, corresponds to the direction along which data vary the most after the first direction, and so on.

The first principal component is considered to contain highest amount of information with respect to the initial parameters, if the aggregate input rate exceeds 85% (Liu and Wang, 2011).

ii. Independent Component Analysis (ICA): This is a stochastic procedure of learning a linear function of a stochastic variable. Its major difference to other statistical methods like the PCA is in the supposition of non- Gaussianity which promote recognition of original underlying components in contrast to classical methods. The aim is to identify and segment components that are eminently independent and non-Gaussian. ICA strives to identify the originality of components using simple supposition of respective statistical resources. Notwithstanding other methods, the fundamental processes are assumed to be independent of each other, a situation that is practical if there is an agreement with defined physical procedure. However, what differentiate ICA from PCA in variables analysis is that it applies non-Gaussian structure of the dataset which is important for regaining the underlying structure of the components that generated the data. It is also important to note that ICA is an unsupervised procedure that takes input data in the structure of a single data matrix by not considering the preferred output of the system. ICA is therefore data-driven method of data analysis such that it can simply measure some occurrence without scheming different investigatory conditions.

Fundamentally, ICA break down given data matrix into variety of stochastically Independent root indicators including a matrix fusion as exemplified in Ndehedehe et al. (2017) and Kalu et al. (2022).

3.2. Multi-Linear Regression Analysis (MLRA)

This is a statistical method used to pattern the relationship between predictand and predictor's parameters. In this particular approach, least squares application is required and has been extensively applied in climate science and hydrology, specifically to explain the potential relationship between major hydrological state parameters (Diatta and Fink, 2014). For a better appreciation of the seasonal, inter and intra-seasonal changes in the time series data, for example; GRACE derived TWS and all the 17 predictor variables imposed on a define grid point. Regression model of the format below was appropriated to the time series `data:

$$X(t) = \beta_0 + \beta_1 t + \beta_2 \sin(2\pi t) + \beta_3 \cos(2\pi t) + \beta_4 \sin(4\pi t) + \beta_5 \cos(4\pi t) + \varepsilon \quad \ldots Equation\ 1$$

Where X(t) represents the predictand (GRACE derived TWS) by time t, β_0 is offset constant, β_1 is linearity tendency, β_2 and β_3 take care of annual signals and β_4, β_5 take care of semi-annual signals. Symbol ε is the bias which represents numerical difference between model outputs and observations. Harmonic components such as annual amplitude and semi-annual amplitude were estimated using least squares as below:

$$Annual\ Amplitude = \sqrt{(\beta_2)^2 + (\beta_3)^2} \quad \ldots Equation\ 2$$

$$Semi\ Annual\ Amplitude = \sqrt{(\beta_4)^2 + (\beta_5)^2} \quad \ldots Equation\ 3$$

Equation 4 gives the formula for root-mean square error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}(x_{obs} - x_{sim})^2} \quad \ldots Equation\ 4$$

Where, x_{obs} and x_{sim} are observed and predicted values for the regression model respectively for a total of n months. Coefficient of multiple determinations was computed to evaluate model's accuracy. Furthermore, application of MLRA is expected to provide understanding on the level to which reconstruction of TWS against all 17 independent variables can be depended upon.

3.3. The Deep-Belief-Network Reconstruction

This deep learning data-driven model was used for reconstruction of Terrestrial Water Storage Anomalies by integrating with two sets of analyzed predictor's dataset from Principal and Independent component analysis respectively;

Integration with Principal Component Analysis: Integration of principal component analysed dataset into deep belief networks for regression analysis and reconstruction of GRACE TWSA involve a systematic approach to extract relevant features and enhance model performance using dimensionality reduction techniques. Unlike traditional neural networks, DBN's provided a powerful framework for feature learning and representation, thus potentially improving on prediction accuracy. The study leverage PCA to convert the original 17 input variables to a reduced set of principal components which were then used as input for DBN model.

The integration of PCA into DBN framework allows for the efficient depiction of standard data in a reduced-dimensional space while preserving essential information. Individual principal component is a linear synthesis of the original input variables, providing a compact yet informative description of the dataset. By feeding these principal components into the DBN model, it aims to encapsulate the underlying structure and relationships in the data more efficiently.

Integration with Independent Component Analysis: In the integration of independent component analysis (ICA) into the Deep Belief Network (DBN) for reconstruction of Terrestrial Water Storage (TWS), the exercise embarks on a distinct approach tailored to the unique characteristics of DBNs. While DBNs offer a powerful framework for feature learning and representation, the incorporation of ICA aims to enhance the capability of reconstruction model in discerning the fundamental structure and relationships of the data more effectively.

Initially, the ICA-DBN integrated model extracts the analysed elements of the initial 17 input variables employing statistical technique of ICA. These techniques capture the salient features of the initial datasets. Notably, discordant components, representing the least effective information are identified during training and subsequently removed from further consideration.

3.4. C4.5 Reconstruction

C4.5 decision tree regression data mining model was used for reconstruction of Terrestrial Water Storage Anomalies by integrating with two sets of analysed predictor's dataset, this analysis was implemented with statistical principal and independent component analysis tools respectively;

3.5. Integration with Principal Component Analysis (PCA)

Integrating principal component analysed dataset into C4.5 model for reconstructing GRACE TWSA involves a distinct approach tailored to the unique characterization of decision trees. Classification and regression tasks are more efficiently implemented by using decision tree regression technique and are specifically design to produce a model that determine the value of a target variable by learning simple decision rules such as "if-then-else". a piece-wise repeated approximation to approximate complex relationships. This approach enables C4.5 to discern underling data structures and relationships more efficiently and ultimately improves on reconstruction capabilities for GRACE TWSA.

3.6. Integration with Independent Component Analysis (ICA)

Integration of ICA data into C4.5 model for reconstruction of GRACE TWSA presupposes adopting a nuanced approach that considers the unique characteristics of C4.5 while leveraging benefits of dimensionality reduction through ICA. C4.5 decision tree algorithm offers interpretability and simplicity yet tends to over-fit on datasets with many features. Hence, employing dimensionality reduction techniques like PCA, ICA, for feature selection beforehand can enhance the model's performance.

Initially, the integrated ICA-C4.5 model extracts the independently constructed components of the original 17 predictors using the ICA technique. These components

encapsulate distinct features of the dataset, aiding in better understanding and interpretation of the underlying patterns. Additionally, noise IC's representing the least effective information, are identified, and eliminated during training to mitigate over-fitting issues.

3.7. Training and testing for Validation

To compute the optimum PC-DBN, PC-C4.5, IC-DBN, and IC-C4.5 reconstructions, the first 249 modes of GRACE-TWS were inserted together with 249 modes of predictor variables. The 249 modes are representative of April 2002 to December 2022 (that is, the study period). Training steps of deep learning routines followed synchronization between time series dataset representing predictor's variables and predictand variable. This was then followed by running the DBN and C4.5 simulation steps for all dominant modes. Higher orders of PC's and IC's were however not considered while integrating them into the DBN and C4.5 routines.

The reconstruction of GRACE-Like TWS routines was done with 70% of datasets used for training and 30% used for testing and ascertaining the accuracy of the reconstructions.

4. PRESENTATIONAND DISCUSSION OF RESULTS

4.1. Evaluation of Reconstructed GRACE-Like TWS anomalies with GRACE-Derived Observations

Figures 6 to 9 depicts contrast of reconstructed GRACE-Like simulation and Observation data from GRACE throughout sub-Saharan Africa (SSA) using DBN and C4.5 model framework respectively for both IC and PC modes. Nonetheless certain variation between the reconstructed GRACE-Like TWS and GRACE-derived TWS anomalies are noticeable as consequence of increased bias and percolation rectification blunders (Chaudhari *et al*., 2018; Felfelani *et al*., 2017; Longuevergne *et al*., 2010) and changes in mountainous and coastal sub-regions (Nie *et al*., 2016; Seager and Henderson, 2013; Sebastian *et al.*, 2016). Tables 2 and 3 present validation results using various metrics such as resubstitution-loss, bias, correlation coefficient (R), mean absolute error (MAE) and root mean square error (RMSE),

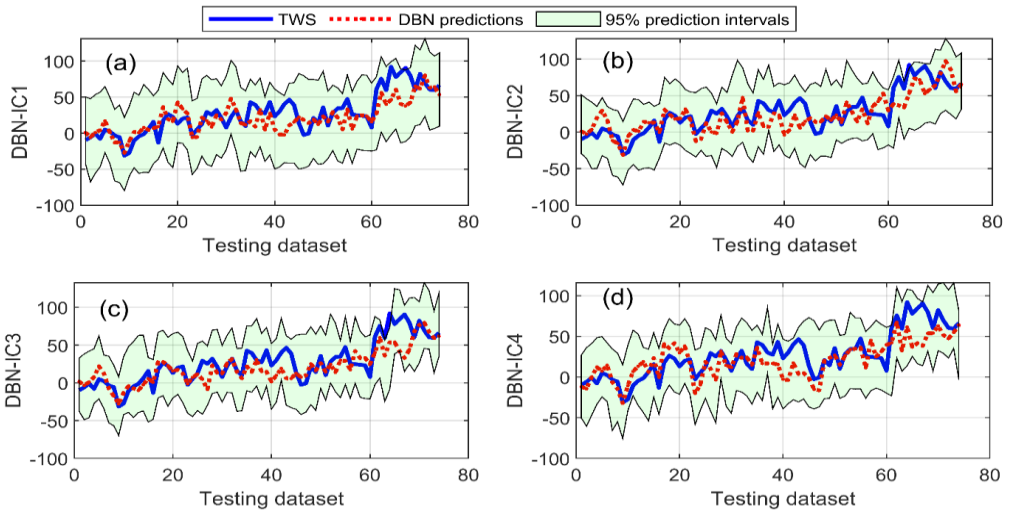


Figure 6: GRACE-Like TWS Anomalies Reconstruction using the IC-DBN framework

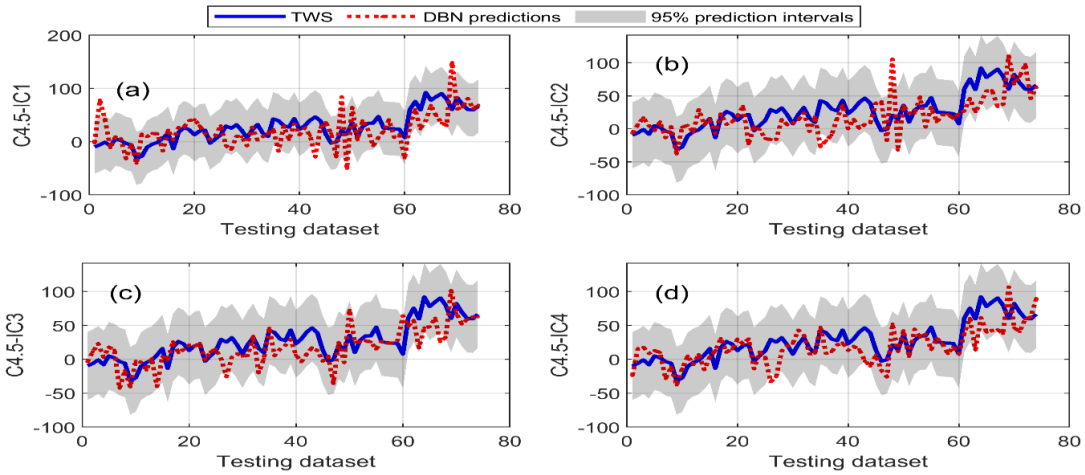


Figure 7: TWSA prediction using the IC-C4.5 framework

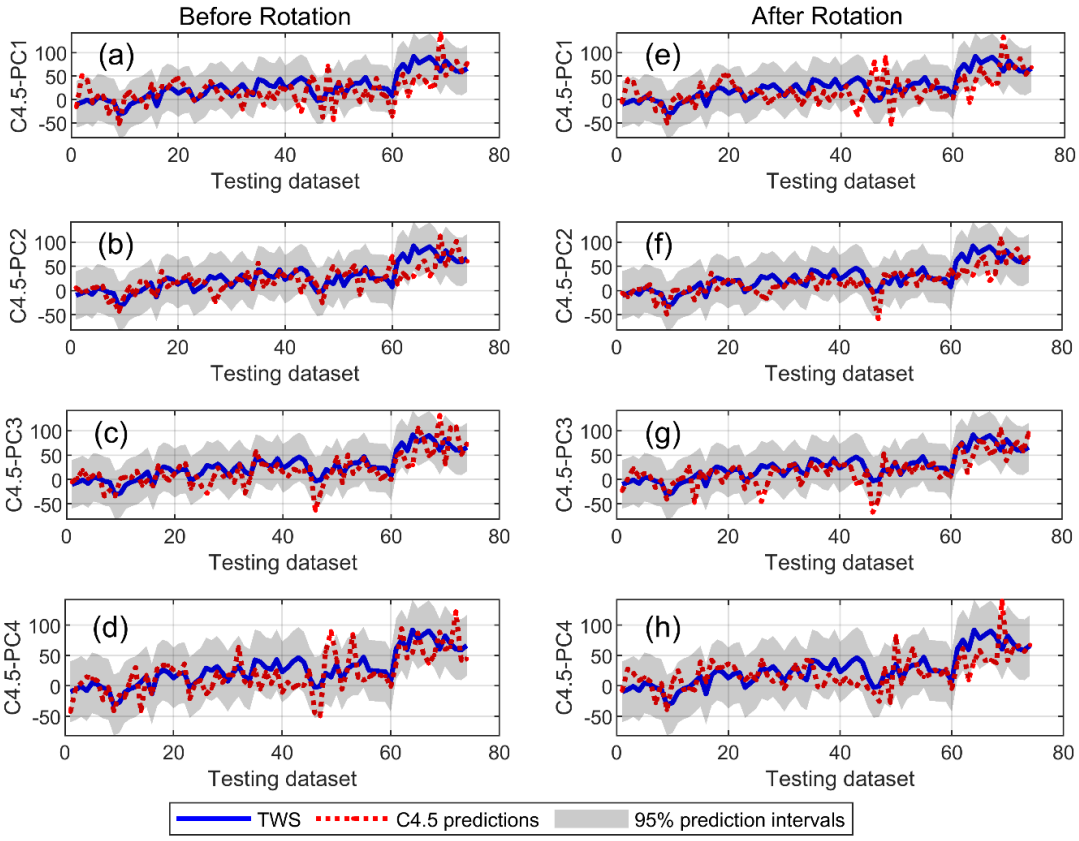


Figure 8: TWSA prediction using the PC-C4.5 framework

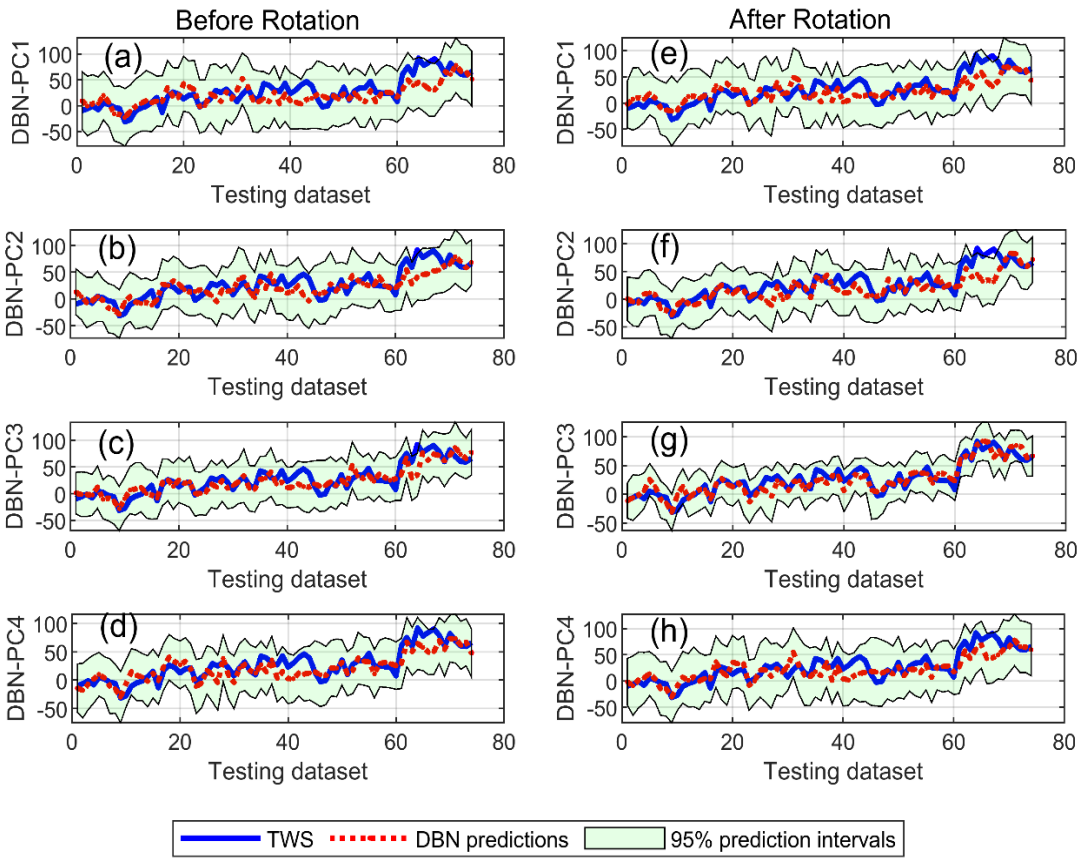


Figure 9: GRACE-TWSA prediction using the PC-DBN framework

Table 2: Evaluation metrics of the C4.5 decision tree regression model for Reconstruction of GRACE-Like TWS.

Training Phase			**Testing Phase**		
Resub.	**Loss**	**Bias**	**RMSE** (mm)	**R**	**MAE** (mm)
PC1-C4.5*	4.1700	8.2426	29.8796	0.5721	22.8056
PC2-C4.5*	4.2555	6.6751	24.0938	0.6277	18.1961
PC3-C4.5*	4.2329	17.0149	24.1828	0.7389	18.4590
PC4-C4.5*	4.1474	12.8466	26.2411	0.6804	20.5717
PC1-C4.5	4.1121	12.9442	30.7740	0.4824	22.9752
PC2-C4.5	4.0870	07.6798	22.7124	0.7028	16.7966
PC3-C4.5	4.0456	20.3247	23.4209	0.7316	16.3831
PC4-C4.5	4.2368	09.8404	25.0004	0.6087	19.2631
IC1- C4.5	4.1933	01.9038	30.3311	0.5121	22.7759
IC2- C4.5	4.1584	06.7127	27.7208	0.5711	21.2360
IC3- C4.5	4.2282	05.1763	23.4722	0.6906	17.8438
IC4- C4.5	4.0788	14.9781	23.0145	0.6880	17.7575
Raw	4.1054	08.4021	28.6077	0.6204	21.5535

Table 3: Evaluation metrics of the Deep Belief Network for Reconstruction of GRACE-Like TWS.

Training Phase			**Testing Phase**		
Resub.	**Loss**	**Bias**	**RMSE** (mm)	**R**	**MAE** (mm)
PC1-DBN*	34.002	13.5349	20.089	0.7277	15.6454
PC2-DBN*	47.1715	10.5397	17.8123	0.7921	13.5400
PC3-DBN*	16.1397	07.4814	12.3355	0.8641	0.8728
PC4-DBN*	43.3696	09.9916	15.4940	0.8523	12.1881
PC1-DBN	14.7245	07.4814	18.4506	0.7910	14.4601
PC2-DBN	55.1883	10.6570	18.1703	0.7985	13.2626
PC3-DBN	05.5212	04.7185	10.9822	0.9229	08.3131
PC4-DBN	30.3595	09.5532	16.9805	0.8137	13.4990
IC1-DBN	55.2810	11.4953	20.1965	0.7095	15.8140
IC2-DBN	28.6356	08.4235	16.8554	0.8262	13.0368
IC3-DBN	14.7517	07.0594	17.7567	0.8068	12.5342
IC4-DBN	29.2856	08.8785	17.1567	0.8075	13.3105
Raw	4.9651	05.1283	16.8742	0.8111	12.1121

4.2. Discussion Based on Study Objectives

4.2.1 Influence of Climate Parameters on Climatology

To understand how each climate variables influence land water storage dynamics and climate change throughout the study period, analysis showed that precipitation, evapotranspiration, runoff, land surface temperatures, and soil moisture layers were the most dominating TWS impacts across Central, Eastern and parts of Southern Africa. El Nino events leads to diminished TWS levels across the region especially over West, Central and East Africa (Forootan *et al.,* 2014). Conversely, La Nina events bolstered TWS. El-Nino and La-Nina events altered seasonal trade wind direction, sea surface temperature and sea level pressure of Pacific Ocean, these impacts are transferred to long distant ecosystem by teleconnections pattern. Atlantic multi--decadal oscillation (AMO) and North Atlantic oscillation (NAO) influences precipitation patterns over sub-Saharan Africa, with negative NAO phases contributing to drier conditions and reduced TWS, while positive AMO phases enhances rainfall and elevate TWS levels. Indian Ocean dipole (IOD), characterized by temperature anomalies in Indian Ocean, plays a crucial role in modulating rainfall variability across the study region, and is observed to significantly affect TWS signals near eastern Africa.

4.2.2 Most significant hydrological parameters

From the decomposition Analysis, Precipitation has been found to have significant positive correlations with all other hydrological parameters. It suggests that the main parameter that influences modelling and recreation processes of TWS over sub-Saharan Africa is Precipitation. This finding collaborates with the finding reported in Nie *et al*., (2016).

Analysis further shows that Hydrological Parameters is majorly significant on yearly temporal scale over sub- Saharan Africa region in tandem with TWSA variations, this finding also agrees with the outcome of study over Amazon basin by Nie *et al.,* (2016) and Yang *et al.,* (2018) in reconstructing terrestrial water storage in North East China using GRACE and hydrological dataset from Global Land Data Assimilation System (GLDAS).

4.2.3 Geographical Locations (Hotspot) in sub-Saharan Africa Most Prone to the Impact of Extreme TWS Variability

Sub-Saharan Africa is characterized by diverse ecosystems, socio-economic disparities, and a heavy reliance on natural resources for livelihoods. However, those characteristics render the region defenceless against influence of climate change and this was explored based on its response to observed terrestrial water storage changes from GRACE satellite. Among the notable regions facing significant challenges includes; River

Basins and surrounding countries that are particularly susceptible to climate change and extremes TWS variability resulting in incessant flood and drought events, the coastal Countries (Ayandele *et al.,* 2023; Nhantumbo *et al.,* 2022), Sahel region, Horn of Africa region including parts of southern Africa region (World meteorological report, 2019).

4.2.4 Evaluation of the accuracy of GRACE-Like TWS Reconstruction Based on C4.5 and DBN models

In this study, two distinct machine learning algorithms were employed, DBN and C4.5 decision tree, to reconstruct GRACE-Like TWSA. The predictive accomplishments of these models were appraised by applying various validation metrics together with root mean square error (RMSE), correlation coefficient (R) and mean absolute error (MAE), The DBN model validation result is presented in Table 3, while that of C4.5 is presented in Table 2. The RMSE of C4.5 model range between 22.7124 and 30.3311 implying a moderate agreement with GRACE -TWSA observations compared to DBN model RMSE range between 10.9822 and 20.1965, reflecting a good agreement with GRACE -TWSA observations. The Correlation coefficient (R) of C4.5 model range between 0.4824 and 0.7389 reflecting a moderate linear agreement with GRACE- TWSA observations compared to DBN model Correlation coefficient (R) range between 0.789 and 0.9221 implying a good linear agreement with GRACE- TWSA observations.

The results obtained through the application of Deep Belief Network model corroborated strongly with past studies in reconstructing TWSA over other regions of the World using related machine leaning algorithms including Random Forest (RF), Extreme Gradient Boost (XGB), Convolution Neural Network (CNN) and Support Vector Machine (SVM) Algorithms respectively (Becker *et al*., 2018; Gyawali *et al.,* 2022; Jing *et al*., 2020; Yu et al., 2021 and Yang *et al*., 2018).

5. CONCLUSION

Rapidly changing dynamics of terrestrial water cycle caused by increasing climate actions trigger complex processes that make conventional methods limited in quantifying the extent of such changes on sub-Saharan Africa's land water cycle. This study applied deep learning and data mining approach to recreate GRACE-like TWSA over sub-Saharan Africa. Seventeen monthly climate parameters from 2002 to 2022 corresponding to GRACE-Derived observations were trained to learn the statistical relationships between DBN/C4.5 simulated TWSA and GRACE-observed TWSA after statistical optimization processes by principal component analysis and independent component analysis respectively. The models were able to hind-cast the TWSA during the study period. Performances of these models were assessed by correlation coefficients, RMSE and mean absolute error (MAE) on regional scale. The results show that DBN models outperform C4.5 model. By comparison, DBN model exhibits optimal solution to properly fit GRACE observations, boasting RMSE of C4.5 model range between 22.7124 and 30.3311 mm to between 10.9822 and 20.1965 mm. As for the spatial pattern of predicted TWSA, the Multi-linear regression model performs reasonably in most of the study area, while it exhibits noticeable uncertainties in coastal and mountainous areas.

This study indicates that DBN deep learning technique is a promising alternative to conventional methods in future hydrological evaluations, it contributes significantly to climate science by enhancing the understanding of TWS variability and climate change impacts in sub-Saharan Africa. Through the novel integration of principal component analysis (PCA) and independent component analysis (ICA) with deep learning algorithms like Deep Belief Networks (DBN) and C4.5 regression, the study advances research ability with respect to assessing influence of hydrological parameters and global climate indices on the region's climatology and most importantly, the feasibility of DBN and C4.5 Models in TWS reconstruction was investigated.

REFERENCES

Agana, N.A. and Homaifar, A. (2018). EMD-based predictive deep belief network for time series prediction: an application to drought forecasting. *Hydrology*, 5 (1):18

Ayandale, A., Oluwanranti, A., Ayandale, O., Borderan, M., Sterly, I., Sokdaporak, P.,Jegede, M., Weldemariam, L. and Ayide, A. (2022). Extreme Climate Events in sub-Saharan Africa: A call for Improving Agricultural Technology Transfer. Climate services, 27, 100311. DOI:10.1016/j.cliser.100311

Becker M., Papa, F., Frappart, F., Alsdorf, D., Calmant, S., Da Silva, J.S., Prigent, C., Seyler, F. (2018). Satellite-based estimates of surface water dynamics in the Congo River Basin. *International Journal on Application of Earth Observation Geoinformation*. 66, 196-209.

Choudhori, S., Felfelani, F., Shin, S., and Pokhrel, Y. (2018). Climate and Anthropogenic Contributions to the Desiccation of the Second Largest Saline Lake in the Twentieth Century. *Journal of Hydrology*.560, 342–353. https://doi.org/10.1016/ j.jhydrol..03.034.

Chen, J. L., Wilson, C. R., Tapley, B. D., Yang, Z. L., and Niu, G. Y. (2005). Drought event in the Amazon River basin as measured by GRACE and estimated by climate models, *Journal of Geophysical Research-Solid Earth*. 114, 1–9. https://doi.org/ 10.1029/2008JB006056, 2009.

Diatta, S and Fink, A. H. (2014). Statistical relationship between remote climate indices and West African monsoon variability. *International Journal of Climatology*, *34*(12): 3348-3367.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., & Schmied, H. M. (2014). Calibration/data assimilation approach for integrating GRACE data into the Water Gap Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surveys in Geophysics*, **35**(6), 1285–1309. **https://doi.org/10.1007/s10712-014-9309-8**

Eyoh, A., Anwana, J., Ettang, I. and Ekpa, A. (2023). Multi-decadal Hydrologic Change and Variability in the Amazon River Basin: Assessing the TWS and NDVI Components in South America. *Journal of the Faculty of Environmental Studies, University of Uyo, JED*, 18(2), 183-203.

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

Felfelani, F., Wada, Y., Longuevergne, L. and Pokhrel, Y. N. (2017). Natural and human-induced terrestrial water storage change: A global analysis using hydrological models and GRACE, *Journal of Hydrol*ogy, 553, 105– 118, https://doi.org/10.1016 /j.jhydrol.2017.07.048,

Forootan, E., Kusche, J., Loth, I., Schuh, W. D., Eicker, A., Awange, J. and Shum, C. K. (2014). Multivariate prediction of total water storage changes over West Africa from multi-satellite data, *Surveys in Geophysics*, *35*(4), 913-940.

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

Gyaneshwar, A., Mishra, A., Chadha, U. and Srinidivase, K. A. (2023) . Contemporary Review on Deep Learning Model for Drought Prediction. *Sustainability*, 15(7): 6160, DOI 10.3390/su15076160

Gyawali, B., Ahmed, M., Murgulet, D and Wiesse, D. (2022). Filling temporal gaps within and between GRACE and GRACE-Fo terrestrial water storage records: an innovative approach, *Journal of Remote Sensing* 14(7): 1565.https://doi.org/10.3390/rs/1471565 (Retrieved on 23rd August, 2022)

Jing, Wenlong., Zhao, Xiaodan., Yao, Ling., Di, Liping., Yang, Ji., Li, Yong., Guo Liying and Zhou. Chenghu. (2024). Can Terrestial Water Storage Dynamics be Estimated From Climate Anomalies? *Earth and Space Science*s. 3, 1-19. http://doi .org/10.1029/2019EA000959, 2020. (Retrieved 28th January,

Jolliffe, I. T. (2002). Principal Component Analysis, 2nd Edition , Springer, New York.

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

Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., van Dijk, A.I.J.M and Pattiaratchi, C. (2017). Assessing sequential data assimilation techniques for integrating GRACE data into hydrological model *Advance Water Resources*, 107. 301-316.

Landerer, F. W. and Swenson, S. C. (2012), Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resources Research*, 48, 1 https://doi.org/10. 1029 /2011WR011453.

Li, F.; Kusche, J.; Rietbroek, R.; Wang, Z., Forootan, E., Schulze, K. and 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.

Liu, H. and Wang, J. (2011). Integrating independent component analysis and principal component analysis with neural network to predict Chinese stock market. *Mathematical Problems in Engineering*, *2011*(1), 382659

Liu, W., Yu, H., Yang, L., Yin, Z., Zhu, M. and Wen, X. (2021). Deep learning-based predictive framework for ground water level forecast in arid irrigated areas. *Water*. *13(18*). https://doi.org/10.3390/ w13182558. (Retrieved 20th April, 2023)

Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N. and Save, H. (2012). GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophysical Research Letters*, *40*(13), 3395-3401.

Long, D.; Shen, Y.; Sun, A.; Hong, Y.; Longuevergne, L.; Yang, Y.; Li, B.; Chen, L. **201**4), Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. *Remote Sensing of Environment. 155*, 145–160.

Longuevergne, L., Scanlon, B. R. and Wilson, C. R. (2010). GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains aquifer, USA, *Water Resources Res*earch, 46, 1– 15, https://doi.org/ 10.1029/2009WR008564,

Mo, S., Zhong, Y. and Shi, X. (2021) Filling the gaps between GRACE and GRACE-FO-derived terrestrial water storage anomalies with Bayesian convolution neural networks, ESS open archive. DOI:10.1002/essour.10505916.,(Retrieved 3rd September, 2023)

Ndehedehe, C. E. (2023). Hydro-Climatic Extremes in the Anthropocene, *Springer Climate*, https://doi.org/10.1007/978-3-031-37727-3_8 279.

Ndehedehe, C. E., Awange, J. L., Agutu, N. O.and Okwuashi, O. (2018). Changes in hydro meteorological conditions over tropical West Africa (1980–2015) and links to global climate. *Global and planetary change*. 16(*2*): 321-341.

Ndehedehe, C. E., Awange, J. L., Kuhn, M., Agutu, N. O and Fukuda, Y. (2017). Climate teleconnections influence on West Africa's terrestrial water storage. *Hydrological Processes*, 31(18): 3206-3224.

Nhantumbo, B., Dade, O and Ghosi, F. (2023). Sea level rise and climate change impact on African coastal system and cities. *Researchgate.net.* DOI:10.5772/intechopen .113083, (Retrieved on 22nd August, 2023)

Nie, N., Zhang, W., Zhang, Z., Guo, H. and Ishwaran, H. (2016). Reconstruction of Terrestrial Water Storage Change From 1948-2012 over Amazon Basin with Latest

GRACE and GLDAS Products. *Water Resource Management*, 30, 279-294. DOI 10.1007/s11269-1161-1. (Retrieved 20th August, 2023)

Nwaerema, P., Vincent, O. N., Amadou, C. and Atuma, I. M. (2019). Spatial Assessment of Land Surface Temperature and Emissivity in the Tropical Littoral City of Port Harcourt, Nigeria. *International Journal of Environment and Climate*, 9(2), 88-103 Reichler, T. and Kim, J. (2008). How well do coupled models simulate today's climate? *Bulletin of the American Meteorological Society*, *89*(3): 303-312.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., and Toll, D. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381-394.

Seager, R. and Henderson, N. (2013). Diagnostic computation of moisture budgets in the ERA-Interim reanalysis with reference to analysis of CMIP-archived atmospheric model data. *Journal Climate*, 26, 7876–7901, doi:10.1175/JCLI-D-13-00018.1.

Sebastian, D.E., Pathak A. and Ghosh, S. (2016). Use of atmospheric budget to reduce uncertainty in estimated water availability over South Asia from different re-analysis, s*cience Report,* 6 (1), 1-10

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

Stockdale, N. (2017). Sub-Saharan Africa World Geography: Understanding a changing world, https://worldgeography.abc-clio.com/ (Retrieved on 4th August, 2023)

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A. and Zhong, Z. (2019).Combining Physically Based Modeling and Deep Learning for fusing GRACE Satellite Data: Can We Learn From Mismatch?, *Water Resources Res*earch, 55, 1179–1195, doi:10.1029/2018WR023333.

Syed, T. H., Famiglietti, J. S.and Chambers, D. P. (2008). GRACE-based estimates of terrestrial fresh water discharge from basin to continental scales. *Journal of Hydrometeorology*, *10*(1), 22-40,

Van Loon, A. F. and Laaha, G. (2015). Hydrological drought severity explained by climate and catchment characteristics, *Journal of Hydrol*ogy, 526, 3–14, https://doi.org/10.1016/j.jhydrol.2014.10.059,

Yan, J., Goa, Y., Yu, Y., Xu, H. and Xu, Z. A. (2020). Prediction Model Based Deep Belief Network and Least Square SVR applied to Cross Water Quality. *Water,* 12(7) 19-29. DOI. org/10.3.390/w/3390/10171929

Yang, P., Xia, J., Zhan, C. and Wang, T. (2018). Reconstruction of terrestrial water storage anomalies in northwest China during 1948-2002, using GRACE and GLDAS products. *Hydrology Research* 49, 1594-1607. https//doi.org/10.9166/nh.2018.074

Yang, P., Xia, J., Zhan, C., Qiao, Y., Wang, Y. (2017). Monitoring the spatio-temporal changes of terrestrial water storage using GRACE data in the Tarim River basin between 2002 and 2015 *Science Total Environment, 595 (2017), 218-228, 10.1016/j.scitotenv. 03.268*

Yin, W.; Hu, L.; Han, S.-C.; Zhang, M.; Teng, Y. (2019). Reconstructing Terrestrial Water Storage Variations from 1980 to 2015 in the Beishan Area of China. *Geofluid*, 2019(1), 3874742. http//doi.org/10.1155/2019/3874742

Yu, Q., Wang, S., He, H., Yang, K., Ma, L. and Li, J. (2021). Reconstructing GRACE-like TWS anomalies for the Canadian landmass using deep learning and land surface model. *International Journal of Applied Earth Observation and Geoinformation*, *102*: 102404. Htts:// doi.org/10.1016/j-jag.2021.102404

Willmott, C.J., Robeson, S.M. and Matsuura, K. 2005). Advantages of Mean Absolute Error over Root Mean Square Error in Assessing Average Model Performance. *Climate Research*, 3, 79-82. https://doiorg/10.3354/cr 030079.

World Meteorological Organization (WMO). (2019). Implementation plan of the Global Framework for Climate Service. https://library.wmo.int/doc_num.php?explnum_id=4028. Geneva. (Retrieved 5th January, 2024)

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