DOWNSCALING GRACE MISSIONS GROUNDWATER STORAGE ESTIMATES USING A MACHINE LEARNING MODEL APPROACH FOR CLIMATE CHANGE MONITORING IN THE UPPER ZAMBEZI CATCHMENT

By

Christopher Shilengwe

(B.Eng UNZA)

A thesis submitted to the University of Zambia in fulfilment of the requirement for the Master of Science in Integrated Water Resource Management

The University of Zambia,

School of Mines, Department of Geology,

Integrated Water Resources Management Centre,

Lusaka.

November, 2024

COPYRIGHT

DECLARATION

I, Christopher Shilengwe, hereby declare that this thesis is my original work and that is has not previously been submitted for a degree, diploma or other qualification at this or any other University. Any incorporated published work or material from other works has been properly referenced and acknowledged.

Signature of author: …………………………………………..

Date: ……………………………………….

CERTIFICATE OF APPROVAL

This thesis of Christopher Shilengwe has been approved as partial fulfillment of the requirements for the Degree of Master of Science in Integrated Water Resources Management by the University of Zambia.

1 ………………………………………………………………………… (Chairperson)

Signature ………………………………………… Date …………………………………

2 ………………………………………………………………………… (Examiner)

Signature ………………………………………… Date …………………………………

3 ………………………………………………………………………… (Examiner)

Signature ………………………………………… Date …………………………………

4 ………………………………………………………………… (External Examiner)

Signature ………………………………………… Date …………………………………

5 ………………………………………………………………………… (Supervisor)

Signature ………………………………………… Date …………………………………

ABSTRACT

The Upper Zambezi Catchment (UZC) of Sub-Saharan Africa, is among the areas most susceptible to the effects of climate change. Significant challenges for water management are brought on by rising temperatures, shifting rainfall patterns and an increase in the frequency of droughts. In this area, groundwater is an essential resource for ecosystems and human needs. Effective water management, particularly in the context of climate change, depends on groundwater storage (GWS) monitoring. By measuring gravity anomalies, the GRACE and GRACE-FO satellite missions have completely changed how terrestrial water storage (TWS) is estimated globally. However, the use of GWS data for local and regional water management is restricted because of the coarse spatial resolution of these missions.

The primary objective of this research was to use open-source remotely sensed data and machine learning (ML) to downscale the GRACE/GRACE-FO GWS estimates to a finer spatial resolution (5 km). Several hydrometeorological variables, including evapotranspiration (ET), land surface temperature (LST), normalised difference vegetation index (NDVI), enhanced vegetation index (EVI), soil moisture and rainfall, were integrated during the downscaling process using the Extreme Gradient Boosting (XGBoost) and Random Forest (RF) algorithms. This method was unique in that it used a time-series split for cross-validation and Z-score normalisation. This made it possible to evaluate the model's performance with confidence.

The findings showed that RF regularly performed better than XGBoost in downscaling GWS estimates, yielding more precise predictions because of its capacity to control intricate correlations between variables and prevent overfitting. The correlation coefficients between the coarse input data and the downscaled data in the Barotse sub-catchment for Nash-Sutcliffe efficiency (NSE) were 0.81 (XGBoost) and 0.89 (RF). Segmenting the UZC into smaller sub-catchments improved model accuracy by enabling the analysis to focus on smaller regions with more homogeneous hydrogeological characteristics. In regions with unconfined aquifers, the downscaling greatly enhanced the GWS estimates. The model, however had challenges in areas with confined aquifers, which usually have low hydraulic connectivity and slow recharge rates.

One of the study's main findings was that groundwater recharge is greatly decreased below a rainfall threshold of 614.39 millimetres during the wet season. This threshold is essential for anticipating periods of low recharge and planning for water scarcity. The study also showed the extent to which the downscaled GRACE/GRACE-FO data captured drought conditions in groundwater, especially during El Niño years (2015, 2016, 2018, 2019 and 2023), underscoring the missions' potential for drought monitoring. The study found that, on average, GWS decreased throughout the UZC between 2009 and 2023, with anomalies varying from -433 millimetres to +264 millimetres.

This research underscores the potential of machine learning in enhancing the spatial resolution and accuracy of GWS estimates, making GRACE/GRACE-FO data more useful for regional water management. These findings have significant implications for managing groundwater resources in the UZC and similar regions experiencing climate variability and increasing water demand.

DEDICATION

To my parents, Christopher Mwandu Shilengwe Snr., Margaret Kaluba Shilengwe,

To my sisters, Diana Shilengwe, Nsama Shilengwe, Chansa Shilengwe and Tamika Shilengwe.

To my late sister, Tionge Shilengwe.

To Esther Mwape.

.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Kawawa Banda, for his invaluable guidance, support and encouragement throughout this research journey. His insightful feedback and expertise have been instrumental in shaping this work and I am truly grateful for his unwavering commitment to my academic growth.

I am also profoundly thankful to the Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) for their generous financial support, which made this research possible. Their commitment to fostering scientific research in the region has been a cornerstone of my academic development. I extend my heartfelt appreciation to the UNZA IWRM Centre for providing a conducive environment for research and learning. The resources and opportunities provided by the Centre have been invaluable in the successful completion of this thesis.

My sincere thanks go to my mentor, Dr. Faustin Banda, whose guidance and wisdom have been a source of inspiration throughout my academic and professional journey. His mentorship has been critical in navigating the challenges of this research, and I am deeply grateful for his support. I would also like to acknowledge my supervisor at the National Remote Sensing Centre, Mr. Michael Katongo Phiri**,** for his understanding and support during this process. His flexibility and encouragement have allowed me to balance my professional responsibilities with my academic pursuits.

Finally, I would like to thank my family, friends and colleagues for their continuous support and encouragement, which have sustained me through the ups and downs of this research journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ACRONYMS AND ABBREVIATIONS

AMSL	Above Mean Sea Level
ANN	Artificial Neural Network
CLSM	Catchment Land Surface Model
CSR	Center for Space Research
CART	Classification and Regression Tree
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station Data
CORDEX-SA	Coordinated Regional Climate Downscaling Experiment-South Asia
DWRD	Department of Water Resource Development
EVI	Enhanced Vegetation Index
ET	Evapotranspiration
XGBoost	Extreme Gradient Boosting
FLDAS	FEWS NET Land Data Assimilation System
GFZ	German Research Centre for Geosciences
GWS	Groundwater Storage
GRACE-FO	GRACE Follow-On
GRACE	Gravity Recovery and Climate Experiment
GW-SW	Groundwater Surface Water
JPL	Jet Propulsion Laboratory
LST	Land Surface Temperature
LEW	Lumped Elementary Watershed
ML	Machine Learning
MAE	Mean Absolute Error
MODIS	Moderate Resolution Imaging Spectroradiometer
NSE	Nash-Sutcliffe Model Efficiency Coefficient
NASA	National Aeronautics and Space Administration
NRSC	National Remote Sensing Centre
NDVI	Normalised Difference Vegetative Index
NRMSE	Normalised Root Mean Square Error
PFIT	Permutation Feature Importance Technique

RCM	Regional Climate Model
RF	Random Forests
RNNs	Recursive Neural Networks
RMSE	Root Mean Square Error
SASSCAL	Southern African Science Service Centre for Climate Change and Adaptive Land Management
STL	Seasonal Trend Decomposition using Loess
SWRMS	Spatial Water Resource Monitoring Systems
SPEI	Standardised Precipitation Evapotranspiration Index
TWS	Terrestrial Water Storage
TWSA	Terrestrial Water Storage Anomaly
UNZA	The University of Zambia
UZC	Upper Zambezi Catchment
WaPOR	Water Productivity Open-access Portal
WARMA	Water Resources Management Authority
WS	Water Storage
ZMW	Zambian Kwacha
ZMD	Zambia Meteorological Department

CHAPTER ONE: INTRODUCTION

1.1 Introduction

Climate change continues to have an impact on both natural environments and human activities in Sub-Saharan Africa. Forecasts for the region point to a persistent trend of rising temperatures, especially in the interior subtropics with more frequent extreme heat waves, increasing aridity and changing patterns of rainfall, with Southern Africa experiencing significantly less precipitation than other regions (Engelbrecht et al., 2015; Field and Barros, 2014). It is anticipated that climate change will have a significant impact on water storage, affecting groundwater and surface water systems (Mupangwa et al., 2023).

Ecosystems and human needs depend on Terrestrial Water Storage (TWS), which includes water in lakes, rivers, streams, soil and groundwater (Miro and Famiglietti, 2018). The conservation of mass based on the hydrological water budget, a resource-intensive procedure, was the standard method for calculating TWS (Amdar et al., 2024). The Gravity Recovery and Climate Experiment (GRACE) and its successor mission GRACE Follow-On (GRACE-FO), which were launched in 2002 and 2018, respectively, revolutionised this field by providing remote sensing capabilities to monitor TWS using a geodetic approach (Ferreira et al., 2023). The GRACE/GRACE-FO missions findings estimated gravity disturbances that were correlated with the presence of hydrological features (Humphrey et al., 2023).

Aquifers store Groundwater Storage (GWS), which is essential for human use and accounts for a sizeable amount of TWS. Measurements of GWS at the national and continental levels have been improved by the GRACE/GRACE-FO missions (Hadavi et al., 2024). However, this dataset has limited applicability for local and regional management due to its coarse spatial resolution (Yin et al., 2022).

Transforming large-scale, coarse datasets into smaller finer data is known as downscaling, the statistical and dynamic approaches are the two main downscaling techniques (Wang et al., 2024). To create high-resolution regional models that capture the physical relationships between variables, the dynamic approach leverages physically based global models. While highly precise, this approach demands significant computational resources and time (Zuo et al., 2021). In contrast, the statistical method enhances spatial resolution by utilising correlations between large and small scale data. Statistical downscaling methods like regression, machine learning (ML) and classification have been used to generate higher-resolution GWS estimates (Fatolazadeh et al., 2022).

1.2 Problem Statement

The study region has seen a rapid population rise over the past decade, which has put a strain on water supplies, leading to the construction of dams and reservoirs that interfere with bodies like rivers and streams (Banda, et al., 2023a). It has been noted that the study region lacks infrastructure that can monitor and assess water storage changes over a time series and this has limited the study and assessment of the impacts of climate change on groundwater storage (Samuel et al., 2023). The GRACE/GRACE-FO Mission provides a solution to fill this data gap by providing derived GWS estimates at global scales (Kumar et al., 2024). Groundwater storage estimate datasets such as those derived from GRACE/GRACE-FO have been widely used for the quantification and estimation of the existence of groundwater storage. However the dataset is coarse in spatial resolution and does not suffice for mapping of groundwater at small scales (Khorrami et al., 2023). Use of a multitude of observation wells proves to be daunting due to the significant financial resources needed.

1.3 Objectives

The main objective of this study was to downscale groundwater storage estimates from GRACE/GRACE-FO mission to a higher resolution dataset by using a ML model based approach using open-source remotely sensed data. To achieve this the specific objectives below were applied.

1. Implement a ML model to downscale groundwater storage estimates using remote sensing to 5 km spatial resolution.
2. Assess the performance of ML algorithms in downscaling groundwater storage estimates.
3. Assess and validate the accuracy and reliability of the downscaled GRACE/GRACE-FO data obtained in Objective (2)
4. Evaluate and interpret the effectiveness of GRACE/GRACE-FO data in detecting and characterising groundwater drought conditions

1.4 Research Questions

The following research questions were developed in order to be in line with the specific goals and objectives of the study.

1. How can machine learning models be applied to downscale GRACE/GRACE-FO groundwater storage estimates to achieve a spatial resolution of 5 km?
2. Which machine learning algorithms are most effective in accurately downscaling groundwater storage estimates from GRACE/GRACE-FO data?
3. What is the accuracy and reliability of the downscaled GRACE/GRACE-FO groundwater storage estimates when validated against ground-truth data and other remote sensing datasets?
4. What extent can GRACE/GRACE-FO data accurately detect and characterise groundwater drought conditions, and what are the limitations of this approach?

1.5 Significance of Study

- The study quantified groundwater storage for the Upper Zambezi Catchment (UZC).
- The quantified groundwater storage will facilitate assessment of water resources at local scales which has not been previously possible.

CHAPTER TWO: LITERATURE REVIEW

The literature review carried out for the research is highlighted in this section.

2.1 Challenges of Water Management in Southern Africa

Sub-Saharan Africa has grappled with the adverse effects of climate on water resources which have far reaching effects where the food security and socio-economic well-being of the region is concerned (Brenya et al., 2024). The Zambezi River Basin in particular has been categorised as one of the river basins on the continent to be severely affected by climate change. This has been seen by the increase in the prevalence of droughts which impact the river flows in basins. In terms of run-off variability, the Zambezi River Basin exhibits highly sensitive variability with fluctuations in precipitation (Beilfuss, 2012).

Groundwater storage forms a crucial aspect of water security and supply in the Southern African region as it is relied on heavily for human consumption, industry and agriculture (Zeng et al., 2024). The advent of climate change has consequentially put the groundwater in the region at risk, this coupled together with urban sprawl and urbanisation. Therefore, the protection of groundwater and its sustainable management therein by the employing of strategies that spearhead conservation policies drawn from stakeholder engagements is of critical importance (Gaffoor, 2017).

The region has neither prioritised nor given adequate attention to groundwater in conversations that relate to water resource management (Hughes et al., 2023). The oversight consequentially diminishes the pivotal role that groundwater plays in the well-being of the populous. In order to sustainably manage water resources, there needs to be a clear understanding and comprehension of the role that groundwater plays (Braune and Xu, 2010).

Technology advancements have led to the development of Spatial Water Resource Monitoring Systems (SWRMS) which have their base in remote sensing technologies (Chen et al., 2024). These systems have enabled the enhanced understanding of variables of the hydrological water budget (evapotranspiration, precipitation, temperature, run-off, river flow and others) over a time series which also includes predictive modelling for future scenarios (Sokeng et al., 2024). These advances have brought about a means to understand and quantify water resources and how they are affected by climate change and urbanisation. A number of these systems have been successful in the quantifying of groundwater at a regional level which has enabled increased conversations and policy adjustments in the management of water resources (Dube et al., 2023).

2.2 Impacts of Climate Change on Groundwater Resources in Sub-Saharan Africa

Climate change poses a significant threat to groundwater resources in Sub-Saharan Africa, exacerbating existing challenges related to water scarcity and management (Seka et al., 2024). Rising temperatures and changing precipitation patterns are expected to alter the recharge rates of aquifers, leading to increased variability in groundwater availability (Taylor et al., 2013). This is particularly concerning in regions that rely heavily on groundwater for drinking water and irrigation, such as the UZC.

Studies have shown that the frequency and intensity of droughts are likely to increase in Sub-Saharan Africa, further stressing groundwater resources (Visvalingam et al., 2024). In addition, sea-level rise and saltwater intrusion pose risks to coastal aquifers, which are vital sources of freshwater for many communities (Field and Barros, 2014). Effective groundwater management in the context of climate change requires not only improved monitoring and modelling but also the integration of climate adaptation strategies into water resource management plans (Chaminé et al., 2022).

2.3 Water Resources in the Upper Zambezi Catchment

Winsemius et al. (2006) undertook a study on the terrestrial water storage changes derived from GRACE/GRACE-FO over a time series in comparison with outputs from a calibrated hydrological model, the Lumped Elementary Watershed (LEW) of the Upper Zambezi. However, the authors noted temporal inconsistencies between the hydrological model and TWS data. They attributed this discrepancy to a GRACE/GRACE-FO defect, as the LEW model had been validated with independent training data. These irregularities are partly due to the limited spatial resolution of GRACE/GRACE-FO, resulting in signals that originate from outside the target area. These signals occur due to the heterogeneity of local hydrogeological conditions in neighbouring pixels. Improvement of the spatial-temporal resolution of GRACE/GRACE-FO through regional downscaling techniques were proposed using local basis functions such as high resolution variables of hydrology.

A fundamental aspect of understanding water storage at the catchment level is the precise mapping and evaluation of groundwater resources (Texier et al., 2024). Groundwater is affected by surface and near surface processes such as infiltration and evapotranspiration (ET) and therefore the enhanced understanding of groundwater surface water (GW-SW) interactions is necessary (Seka et al., 2024). The Barotse Flood Plain which is a sub-basin of the UZC was investigated with the use of remote sensing and hydro-chemical analysis. In their study, Banda et al, (2023a) noted that the region exhibits a strong relationship between groundwater and vegetation as is observed by the Normalised Difference Vegetative Index (NDVI) and therefore vegetation in the region is highly sensitive to groundwater changes. This is due to the fact that the region has a near surface water table that is less than 10 metres and therefore is easily accessible by vegetation.

Groundwater quality assessments have been scarcely carried out in developing regions in comparison to groundwater storage which has hampered water resource management in said regions (Tang et al., 2022). Aquatic ecosystem health and water quality can be assessed through the use of benthic invertebrates, which are sensitive to pollution and environmental changes. In their research, Banda et al. (2023b) studied the relationship between water quality and benthic macro invertebrates in the Barotse Floodplain, a wetland that is heavily dependent on groundwater in Zambia, using a water quality survey and canonical correspondence analysis.

2.4 Data Scarcity in Hydrological Studies and Its Implications

Data scarcity remains a significant challenge in hydrological studies, particularly in developing regions like Sub-Saharan Africa (Dube et al., 2023). The lack of consistent and reliable ground-based observations hinders the accurate assessment of water resources, making it difficult to manage them effectively (Hannah et al., 2011). This scarcity is particularly acute in rural and remote areas, where infrastructure for data collection is often lacking.

The implications of data scarcity are far-reaching, affecting everything from the calibration of hydrological models to the validation of remotely sensed products (Nilsson and Nielsen, 2024). For example, without sufficient in-situ measurements, it is challenging to validate satellite-derived estimates of groundwater storage or soil moisture, leading to uncertainties in the estimates (Dembélé and Zwart, 2016). Consequently, there is a pressing need for innovative approaches to address these data gaps, such as the integration of citizen science initiatives or the leveraging of open source satellite based sensors for data collection (Buytaert et al., 2014).

2.5 Working With Data Scarcity: The Role of Remotely Sensed Data

In regions where ground-based data is scarce, remotely sensed data has become an invaluable resource for hydrological studies (Carneiro et al., 2023). Satellite observations provide continuous, large-scale coverage, making them essential for monitoring hydrological variables such as precipitation, evapotranspiration and groundwater storage (Brunner et al., 2007). Remotely sensed data, when combined with modelling techniques, can help fill critical gaps in areas with limited or no ground-based observations.

However, working with remotely sensed data in the context of data scarcity presents its own challenges (Gong et al., 2024). The accuracy of satellite-derived products can be affected by factors such as sensor resolution, atmospheric interference and the heterogeneity of the landscape (Karnieli et al., 2010). As a result, it is crucial to validate these products using ground-based data that is available or to apply data assimilation techniques that can improve the accuracy of the estimates (Tangdamrongsub et al., 2015).

2.6 Correlation of Hydrometeorological Variables with Hydrogeological Parameters

Understanding the correlation between hydrometeorological variables, such as precipitation, temperature and evapotranspiration and hydrogeological parameters, like groundwater levels and soil moisture, is essential for effective water resource management (Roy et al., 2024). These correlations are crucial for predicting groundwater recharge rates, assessing drought impacts and managing agricultural water use. Recent studies underscore the importance of these relationships, particularly in the context of climate change, which significantly impacts groundwater levels and soil moisture dynamics (Ahmed et al., 2022; Jeihouni et al., 2019). For instance, Haji-Aghajany et al. (2023) explored how variations in precipitation and temperature influence groundwater levels in regions affected by subsidence, demonstrating that changes in these meteorological indicators can lead to significant alterations in groundwater resources.

Malik et al. (2021) examined the effect of groundwater levels on soil moisture, soil temperature and land surface temperature. Their research underscores the importance of understanding how groundwater fluctuations impact soil moisture dynamics, which in turn affect land surface temperatures.

Moreover, the study by Pablos et al. (2016) highlights the use of remote sensing technologies to monitor soil moisture and land surface temperature dynamics. This research emphasises the value of remote sensing in capturing the complex interactions between hydrometeorological and hydrogeological parameters, providing critical data that can enhance the accuracy of hydrological models and inform water resource management strategies.

2.7 Challenges in the Use of GRACE/GRACE-FO Data for Hydrological Applications

Despite the advancements in data assimilation techniques, significant challenges remain in applying these methods to GRACE/GRACE-FO data for hydrological applications. One of the primary challenges is the coarse spatial resolution of GRACE/GRACE-FO data, which limits its direct applicability to local-scale studies (Xu et al., 2023). The inherent noise in GRACE/GRACE-FO observations, caused by factors such as atmospheric mass variations and tidal effects, further complicates its effectiveness in hydrological studies (Wahr et al., 2004). Downscaling is a technique that improves the spatial or temporal resolution of data, allowing coarse or large-scale information to be applied to more detailed, small-scale contexts (Impollonia et al., 2024). In environmental sciences and remote sensing, downscaling takes global or regional data, such as climate or satellite-derived measurements and refines it to provide insights at a finer spatial scale (Li et al., 2024).

Broadly, two categories of downscaling methods exist which are the statistical and dynamic downscaling methods (Leroux et al., 2024). In the case of statistical downscaling, statistical relationships between coarse global and regional high resolution climatic variables are utilised to generate higher-resolution regional climate information. On the other hand; dynamic downscaling employs the principles of meteorological theory and physical models through numerical simulations (Zhang et al., 2023). While statistical downscaling imposes a relatively lower computational burden, there remains uncertainty about the transferability of relationships from the current climate to future climatic conditions. Dynamic downscaling, while demanding more substantial computational resources and generating a greater volume of output, is advantageous in that the underlying universal dynamic and physical theories can be applied to both present and future climatic conditions (Diro et al., 2012).

Another challenge in the use of GRACE/GRACE-FO data lies in the downscaling process, specifically in the selection of appropriate high-resolution datasets to serve as predictors. Commonly used variables such as NDVI, land surface temperature (LST) and soil moisture may vary in effectiveness depending on the region and specific hydrological processes involved (Ullah et al., 2023). This is particularly relevant in regions like Sub-Saharan Africa, where variability in climatic conditions and limited ground-based observations complicate the validation of downscaled products (Tao et al., 2023)

Researchers have explored hybrid models, combining statistical and dynamic approaches, to improve groundwater estimates. ML methods have been effective in enhancing spatial resolution, yet the computational demands of these models remain a significant barrier to broader application (Satizabal-Alarc et al., 2023).

2.8 Dynamic Downscaling of Hydrological Variables

Thorough investigations into the repercussions of climate change necessitate regional climate data with finer spatial resolution. As a result, the application of dynamic downscaling methods has been employed to refine these datasets (Taniguchi, 2016).

In the research by Jayasankar et al. (2021) it was emphasised that employing high-resolution Regional Climate Model (RCM) simulations offers a valuable approach for obtaining realistic climate change projection data. The study specifically utilised a high-resolution dynamic downscaling framework (CCSM4-WRF) focused on India. To illustrate the benefits of this enhanced resolution, a comparison was drawn between the outcomes of CCSM4-WRF which is a high resolution model simulation at a resolution of 9 km based on India and RCM simulations at a coarser resolution of 50 km from the Coordinated Regional Climate Downscaling Experiment-South Asia (CORDEX-SA). The primary variable under investigation in the study was precipitation and the assessment found that the larger part of the CORDEX-SA noted a negative bias when it comes to the precipitation prevalence in the region.

Hanssen-Bauer et al. (2003) conducted a study using the ECHAM4/OPYC3 climate model scenario to examine climate change effects in Norway. They employed both empirical and dynamical downscaling methods to project temperature and precipitation changes. The empirical models used large-scale temperature for local variations and combined temperature and sea-level pressure for precipitation. The dynamical model, HIRHAM is a Regional Climate Model that is built on HIRLAM with ECHAM4/OPYC3 optimisation and biases. Both methods projected an annual mean temperature increase in Norway from 1980–1999 to 2030–2049. Despite minor discrepancies, the study revealed general agreement between the two approaches.

Brown et al. (2008) extensively reviewed a number of downscaling models in Africa. They found that dynamic downscaling is valuable for capturing intricate terrain features and changing land use effects on regional climate in areas where these factors are significant. Statistical downscaling offers a simpler alternative that requires less computing resources, suitable for estimating specific variables at distinct points for sector models or decision-making. However, it may sacrifice a comprehensive understanding of regional climate dynamics, potentially leading to overly confident estimates. The trade off in the decision between the use of either a dynamic or statistical downscaling approach must take into account the computer processing resources needed as well as the level of generalisation and biases of the chosen approach (Wang et al., 2024).

2.9 Statistical Downscaling of GRACE/GRACE-FO Groundwater Storage Estimates

Numerous studies have been conducted in the downscaling of groundwater storage estimates by a statistical method. The common theme is the use of high resolution hydrological variables that are used to establish relationships with the coarse dataset. Yin et al. (2018) used various evapotranspiration datasets to downscale GRACE/GRACE-FO mission’s groundwater storage estimates by use of a correlation algorithm. In principle, the downscaling of GRACE/GRACE-FO data is based on the understanding that changes in gravity anomalies correspond to movement and changes in the hydrosphere. The algorithm correlated the prevalence of instances of ET with groundwater thickness due to there being a strong linkage between GWS and ET in the study area. One of the drawbacks of this method is that it used one hydrological variable to downscale GWS estimates which may not apply to other regions with differing hydrogeological characteristics.

Milewski et al. (2019) carried out spatial downscaling of GRACE/GREACE-FO Terrestrial Water Storage Anomaly (TWSA) for the purpose of understanding of groundwater changes in the Upper Florida aquifer. The method employed various high resolution hydrological variables including, NDVI, LST, soil moisture and precipitation. The target resolution was 5 kilometres and this necessitated the need to resample the inputs to this value. A key takeaway from the study is the carrying out of a preliminary correlation assessment to determine the weighting and importance of each input variable in the model using Pearson's correlation coefficient carried out between the predictors (hydrological variables) and dependent variable (GRACE/GRACE-FO TWSA).

The overall framework of statistical downscaling is as follows:

i. Isolate and prepare GRACE/GRACE-FO TWS for analysis and resample it to the target resolution (Vishwakarma et al., 2021);
ii. Decide on the hydrological variables that will serve as the high resolution inputs in the downscaling process; and
iii. Carrying out a sampling of whether the downscaling will be based on a ML model or an empirical regression analysis (Chen et al., 2019).

2.10 The Role of Machine Learning in Hydrological Studies

ML has emerged as a powerful tool in hydrological studies, particularly in the context of climate change where the demand for accurate and timely data is increasingly critical. ML algorithms can process vast amounts of data from various sources, including satellite imagery, in-situ measurements and climate models, to generate insights that would be difficult to obtain through traditional methods (Shortridge et al., 2016).

In the realm of hydrology, ML has been applied to predict stream flow, rainfall and groundwater levels with high accuracy. For instance, Kratzert et al. (2019) demonstrated that deep learning models could outperform traditional hydrological models in predicting daily stream flow across multiple basins. Similarly, Mosavi et al. (2018) reviewed the application of ML techniques in flood forecasting, highlighting their potential to improve early warning systems.

Moreover, ML algorithms such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) have been widely used in downscaling coarse-resolution data, such as

GRACE/GRACE-FO satellite observations, to generate high-resolution estimates of groundwater storage (Khan et al., 2024). These models are particularly effective in capturing non-linear relationships between hydrological variables, making them ideal for complex environmental systems where multiple factors interact (Chen et al., 2021a).

2.11 Machine Learning Based Statistical Downscaling of Groundwater Storage Estimates

The use of empirical methods in the downscaling of hydrological variables has its limitations, leading to the widespread adoption of ML methods. The ability of ML algorithms to solve non-linear regression problems and replicate intricate hydrological processes has made them highly sought-after in GRACE/GRACE-FO data downscaling research (Yamini and Manjula, 2023). In particular, the ML spatial downscaling proposes the combination of multiple ML algorithms (Chen et al., 2021b). Following the downscaling process, it is essential to assess the consistency of TWS changes before and after downscaling through the use of error parameters such as root mean square error and mean absolute error amongst others (Faraji et al., 2024). The validation of downscaled TWS and GWS is dependent on the prevalence of observation wells in the selected study area as their consistency can be used for this purpose. In the absence of such observation wells, other methods, such as conservation of mass or correlation with a hydrological variable may be applied (Zhang et al., 2021).

Fu et al. (2022) utilised ML to generate spatiotemporally continuous Standardised Precipitation Evapotranspiration Index (SPEI) data for South Western China from 1901-2018. Their study revealed that downscaled SPEI using four Classification and Regression Tree (CART) approaches can provide valid local drought information, with the extra trees approach demonstrating the best performance. The proposed approach has the potential to enhance long-term drought monitoring in areas without in-situ data.

Sahour, (2020) applied a ML downscaling on GRACE/GRACE-FO TWS using Lower Peninsula of Michigan as a study area. The preliminary analysis carried out was the identification of clusters in the study area with similar geophysical signals that have comparable pixel values. The variables of the hydrological water budget (precipitation, NDVI, LST, ET and soil moisture) that were used were split by using the 80:20 ratio between training and validation. Thereafter, the variables together with the GRACE/GRACE-FO TWS were passed to the chosen algorithms (Artificial

Neural Network (ANN), XGBoost and multivariate regression) for downscaling. The error analysis was carried-out using the R-squared (R^2), the Nash-Sutcliffe Model Efficiency Coefficient (NSE) and the Normalised Root Mean Square Error (NRMSE). It was noted that XGBoost outperformed the other two algorithms in the downscaling process. Observation wells distributed around the study area were used for validation of the downscaled GRACE/GRACE-FO TWS.

2.12 Validation of Remotely Sensed Data with Ground-Based Observations

Remote sensing technologies, such as those provided by the GRACE/GRACE-FO missions, have revolutionised the monitoring of groundwater resources on a global scale. However, the integration of remote sensing data with ground-based observations is crucial for enhancing the accuracy and reliability of hydrological models (Famiglietti et al., 2011). Ground-based measurements, such as those from observation wells, provide critical validation for remote sensing estimates and help to reduce uncertainties in model predictions.

Coelho et al. (2017) validated groundwater data in California's Central Valley by integrating satellite-based observations with field-measured water balance methods, enhancing the accuracy of storage estimates by combining remote and local data sources. Fallatah et al. (2019) demonstrated in Brazil's semi-arid regions that integrating cloud-computing tools with remote sensing and in-situ data can significantly improve the validation of groundwater recharge models, especially in data-scarce areas. Lucas et al. (2015) validated remote sensing data for the Guarani Aquifer System by comparing it with field measurements, particularly focusing on groundwater recharge in outcrop zones to ensure accurate estimation.

Moreover, the development of data assimilation techniques, which merge satellite observations with model outputs and ground measurements, has shown promise in enhancing the accuracy of hydrological predictions (Hilker et al., 2012). These techniques allow for the continuous updating of model states with real-time data, improving the responsiveness of water resource management to changing environmental conditions (Lakshmivarahan and Stensrud, 2009).

2.13 Future Directions in Groundwater Research Leveraging Machine Learning and Remote Sensing

The future of groundwater research in the context of climate change will likely see an increased reliance on ML and remote sensing technologies. As the availability of high-resolution datasets continues to grow, so too will the opportunities for developing dedicated models that can capture the complexities of groundwater systems (Tao et al., 2022b).

Emerging areas of research include the use of deep learning algorithms for time-series analysis of groundwater data, which could provide more accurate predictions of groundwater trends under different climate scenarios (Kumar et al., 2023). Additionally, the integration of remotely sensed data with socio-economic indicators offers a promising avenue for understanding the broader impacts of groundwater depletion on communities and ecosystems (McCabe et al., 2017).

Furthermore, the development of global-scale groundwater models, supported by advances in computing power and data availability, holds the potential to transform our understanding of groundwater dynamics in response to climate change (Richey et al., 2015). These models could provide critical insights for policymakers and water managers, enabling more informed decision-making in the face of growing water scarcity.

2.14 Research Gap

The literature on ML-based spatial downscaling of GRACE/GRACE-FO data in the Sub-Saharan region revealed several research gaps. Firstly, there is a noticeable lack of studies utilising ML techniques for downscaling GRACE/GRACE-FO data in the UZC, despite the potential to address local hydrological challenges more effectively. Second, most existing studies prefer using the original GRACE/GRACE-FO datasets for downscaling, often overlooking the benefits of incorporating Global Land Data Assimilation System (GLDAS) derived groundwater data, which could improve the accuracy of models by capturing regional hydrological dynamics more precisely. Additionally, there is a gap in the methodology used for model training and validation; most studies do not employ a time series split to separate training from validation datasets, potentially leading to overfitting and less reliable predictions. Finally, the normalisation of datasets before inputting them into ML models for validation is not commonly practiced, despite its importance for enhancing model performance and ensuring better downscaling results.

CHAPTER THREE: DESCRIPTION OF THE STUDY AREA

This chapter details the study areas', location, physical and socio-economic characteristics.

3.1 Location of the Upper Zambezi Catchment

The Greater Zambezi River Basin, situated in Sub-Saharan Africa, is comprised of three main sub-basins: the Lower Zambezi; Middle Zambezi; and lastly Upper Zambezi which this study focuses on (Samuel et al., 2024; Dube and Nhamo, 2023). The study area spans across the borders of western Zambia, eastern Angola, northern Botswana, western Zimbabwe and eastern Namibia, covering an area of approximately 515,000 km^2. Geographically, it is located between latitudes 10°57'17"S and 18°58'26"S, and longitudes 18°21'29"E and 26°25'22"E as depicted in Figure 1 (Beilfuss, 2012).

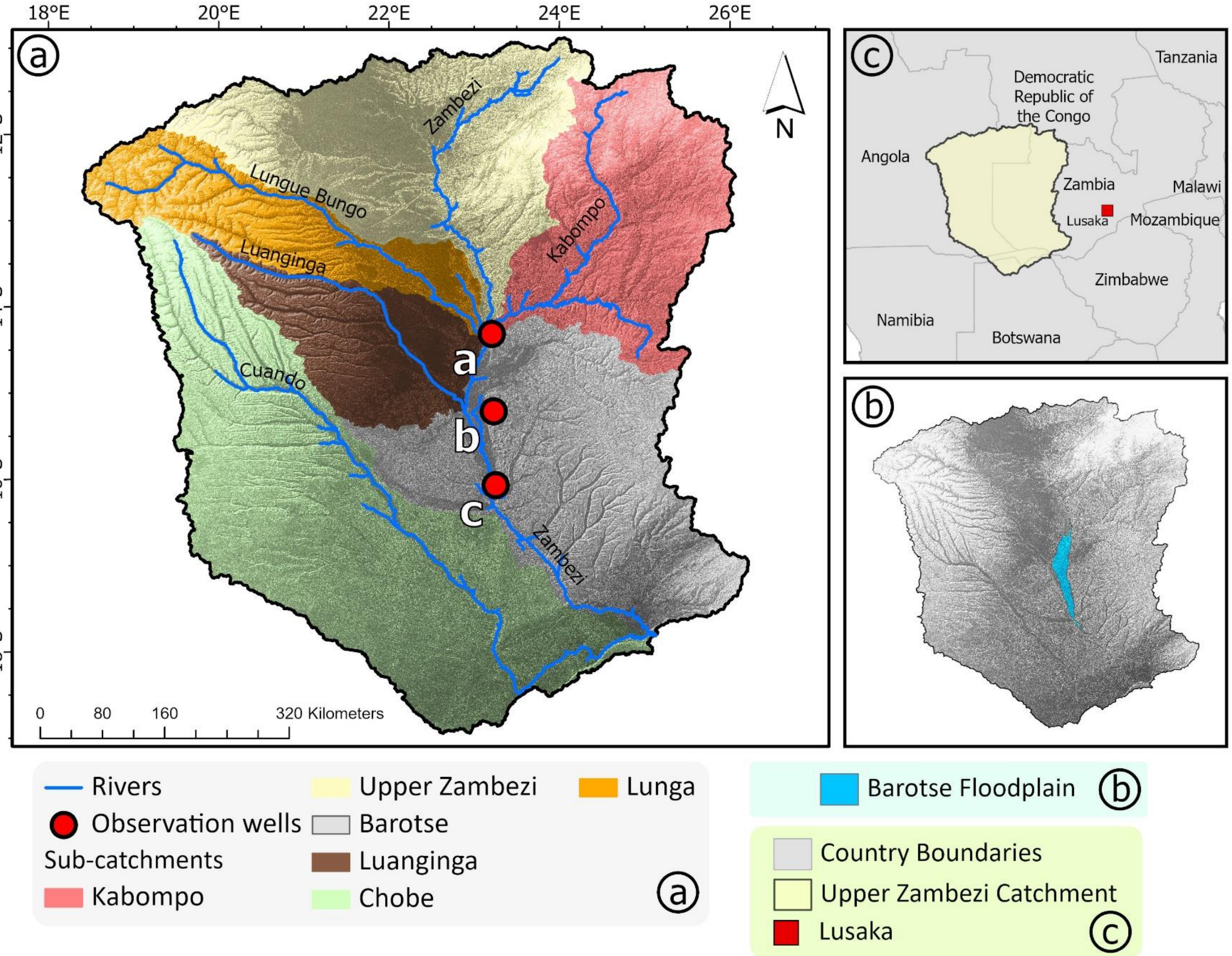


Figure 1: Location map of the UZC showing the: (a) rivers, observation wells (a – Lukulu, b – Mongu and c – Senanga) used in the study and sub-catchments, (b) location of Ramsar sites and (c) location of the UZC in Southern Africa, country boundaries and the location of Lusaka

3.2 Physical Characteristics

The physical characteristics of the study area covers the, hydrological catchments, geology observation wells, hydrology and topography, vegetation, socio-economics and population.

3.2.1 Hydrological Catchments

The UZC, which is an essential part of the larger Zambezi River Basin, is composed of six major sub-catchments: Kabompo; Luanginga; Lunga; Chobe; Barotse; and Upper Zambezi (Beilfuss, 2012). These sub-catchments play a crucial role in the hydrological dynamics of the UZC, each contributing to the overall water flow, sediment transport and ecological balance within the basin. The main river of the UZC is the Zambezi, which is fed by several tributaries, including the Lungue-Bungo, Kabompo, Luanginga and Cuando/Chobe Rivers, each contributing to the basin's hydrological network (Banda et al., 2023c; Hughes et al., 2023; Mendelsohn, 2022). The varying climatic and geographical characteristics across these sub-catchments influence the hydrological processes, making this region vital for water resource management and environmental conservation efforts. The distribution and layout of these sub-catchments are illustrated in Figure 1 (Lautze et al., 2017; Chirashi, 2014).

3.2.2 Geology

The hydrology and landscape of the UCZ are greatly impacted by its complex geology (Figure 2). A variety of metamorphic formations, sedimentary deposits and volcanic rocks can be found in the region, these features allow for water infiltration, storage and aquifer recharge (Bäumle et al., 2018). In the northern areas, metamorphic rocks such as schist, gneiss and quartzite create stable terrains that serve as aquitards, preventing excessive drawdown in deeper aquifers while limiting groundwater movement (Persits et al., 1997).

Sand, calcrete and bioclastic formations make up the majority of the UZC and are essential for groundwater recharge (Elorza and Recio, 2023). Extrusive volcanic rocks can be seen in the Southern UZC, especially in the south-eastern regions (Rubio-Arellano et al., 2024). Although they have a smaller impact on groundwater storage, these younger geological formations have an impact on water flow and landscape shape.

The sandstone formations that are dispersed throughout the Barotse Flood Plain, especially in the vicinity of Lukulu, Mongu and Senanga, serve as vital aquifers that maintain water availability throughout the catchment and further support groundwater storage because of their high porosity (Jagert et al., 2023; Banda et al., 2019). The Floodplain supports subsurface aquifers by permitting substantial water infiltration due to its high permeability. The sandy nature of the floodplain makes it an important area for subsurface recharge and surface water retention, supporting groundwater-dependent local ecosystems and communities (Bäumle et al., 2018; Balon and Coche, 1974).

3.2.3 Observation Wells

The study utilised observation well data from three strategically located wells in the Lukulu, Mongu, and Senanga districts, representing upstream, midstream, and downstream locations along the Zambezi River within the Barotse sub-catchment of the UZC as depicted in Figure 1. These wells, situated on sandstone geological formations, were specifically selected due to their capacity to provide consistent and reliable time series water level data, with daily measurements recorded from December 2019 to October 2020 (Banda et al., 2019). The selected wells, which exhibit a near-surface water table at depths of less than 10 meters, are integral to understanding groundwater dynamics in this region. Their strategic placement across different sections of the floodplain allows for a comprehensive analysis of water level variations influenced by both geological formations and hydrological processes. This spatial coverage is critical for accurately assessing groundwater trends and potential impacts within the Zambezi River's extensive floodplain system (Banda et al., 2023c).

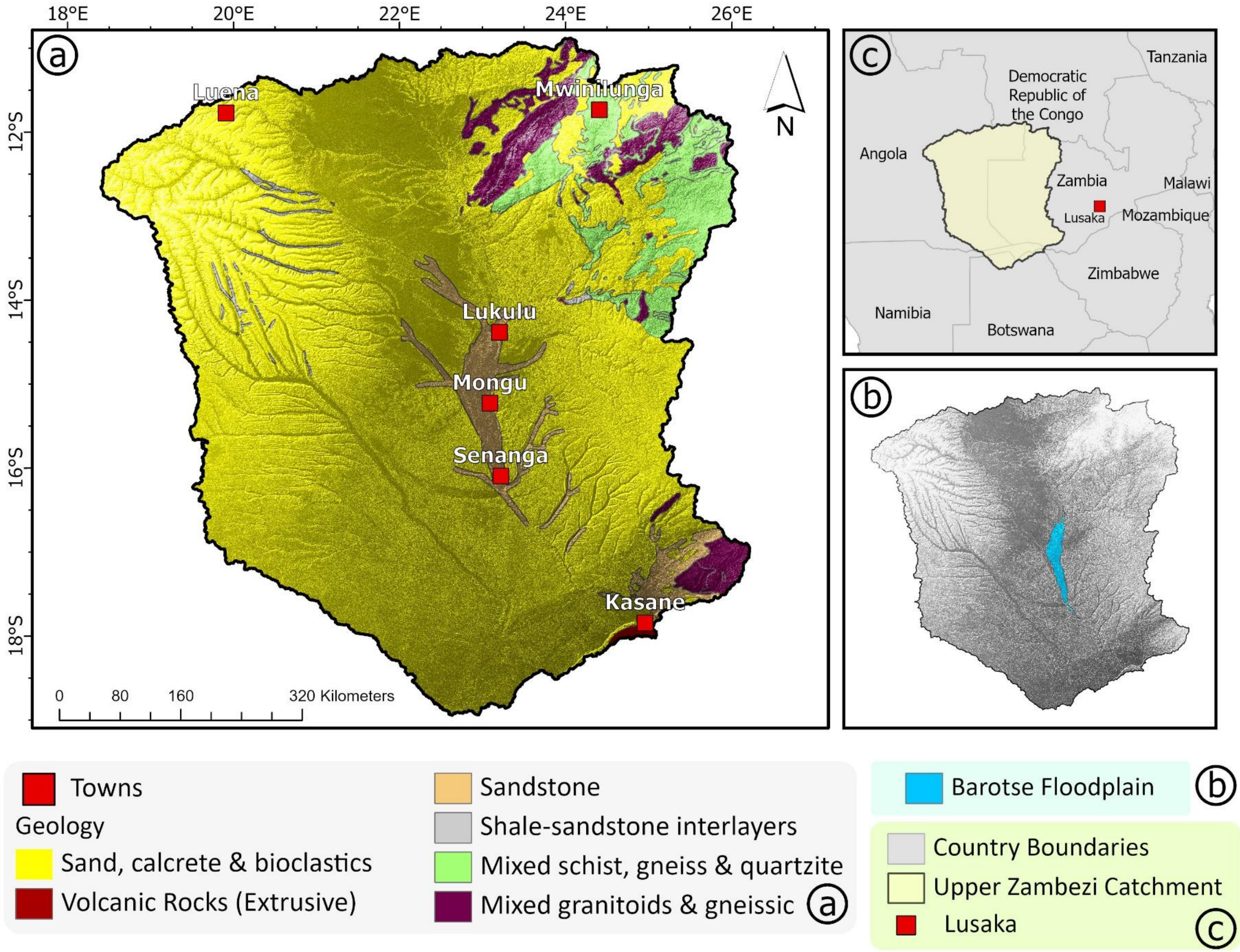


Figure 2: Location map of the Upper Zambezi Catchment, showing the: (a) location of towns and geology, (b) location of Ramsar sites and (c) location of the UZC in Southern Africa, country boundaries and the location of Lusaka

3.2.4 Hydrology and Topography

The UZC, with an average annual rainfall of 1,000 millimetres, exhibits considerable variability in precipitation across its vast area (Hughes et al., 2023). This region is also characterised by its abundant Kalahari sands, which play a crucial role in the local hydrological processes, particularly in groundwater recharge and soil moisture retention. The terrain within the catchment ranges from 909 to 1,677 metres above mean sea level (AMSL), providing a diverse landscape that influences the flow patterns of the major rivers, including the Zambezi and Chobe, which are of significant hydrological interest (Banda et al., 2023a).

Moreover, the region is home to an important Ramsar site, the Barotse Floodplain. The Barotse Floodplain, a vast wetland system, plays a critical role in flood management and supports a rich biodiversity, making it a site of international importance (Ramsar, 2004). The Ramsar site is vital for the conservation of wetlands and the sustainable management of water resources in the UZC (Mosimane et al., 2017; Ramsar, 2004).

3.2.5 Vegetation

The vegetation within the UZC is highly diverse, reflecting the region's varying climatic, hydrological and soil conditions. Predominantly, the area is covered by Miombo woodlands, which are characterised by species such as *Brachystegia* spp., *Julbernardia* spp., and *Isoberlinia* spp (Chomba et al., 2022; Zimba et al., 2020). These woodlands are well-adapted to the seasonal rainfall patterns and the sandy soils of the region (Timberlake and Chidumayo, 2011). In addition to the Miombo woodlands, the region also includes significant areas of grasslands, particularly within the floodplains of the Barotse and along the river valleys. These grasslands play a crucial role in supporting local agriculture and wildlife, providing grazing grounds for livestock and habitat for a wide range of fauna (Chomba et al., 2022). The riparian zones along the Zambezi and Cuando/Chobe rivers are characterised by dense riverine forests, which are important for maintaining the ecological balance of the waterways and providing resources for the local communities (Lautze et al., 2017).

3.2.6 Socio-economics

The socio-economic landscape of the UZC is largely shaped by traditional agricultural practices and the use of natural resources. The majority of the population relies on subsistence farming, with key crops including maize, millet and sorghum (Kiesel et al., 2022; Tirivarombo and Hughes, 2011). Livestock farming, particularly cattle rearing, is also a crucial part of the local economy, providing both food security and income for rural households (Hulme and Arntzen, 1996). Fishing, timber harvesting and the collection of non-timber forest products are also significant livelihood activities, particularly in areas close to rivers and forests (FAO, 2018).

In addition to traditional livelihoods, tourism is an increasingly important economic activity in the region, particularly in areas surrounding the Barotse Flood Plain (Ramsar, 2004). The tourism sector provides employment opportunities and has led to the development of infrastructure such as lodges and transport services. However, economic development is uneven, with disparities existing between different areas, often based on the availability of natural resources and proximity to tourist attractions (World Bank, 2019).

Earnings and employment data on the Zambian side of the study region indicate significant disparities based on sector and location. The average monthly earnings for paid employees on the Zambian portion are estimated at ZMW 4,125, with men generally earning more than women (ZamStats, 2021). Where workers in agriculture, forestry and fisheries sectors earn the least, averaging ZMW 2,073 per month. Formal employment accounts for approximately 26.8% of the employed population, with the remainder engaged in informal or household-based activities (ZamStats, 2021).

3.2.7 Population

The population distribution across the UZC is shaped by various factors, including terrain, water resource availability and socio-economic activities. The study area, which spans across several regions in Zambia, Angola and Namibia, is predominantly rural, with higher population densities in fertile regions and areas with access to water bodies like the Zambezi River (Petrovic and Sherland, 2023; ZamStats, 2022).

In Zambia's portion of the UZC, particularly in the Western Province, which includes the Barotse Floodplain, has a relatively low population density of approximately 10.8 persons per square kilometre (Breton et al., 2024). Southern Province, particularly in areas around the Zambezi River, support higher population densities due to fertile soils and better access to water, leading to more intensive agricultural activities (ZamStats, 2022). In the North-Western Province, population growth is rapid, driven by mining activities and related economic opportunities in mining districts like Kalumbila and Solwezi (ZamStats, 2022).

These provinces have experienced varying rates of population growth, with North-Western Province recording the highest growth at 74.7% between 2010 and 2022, reflecting the region's expanding economic activities (Arnzen et al., 2024; ZamStats, 2021).

CHAPTER FOUR: METHODOLOGY

This chapter details the methods applied in this research. It outlines the study design, data sources, downscaling method and analysis.

4.1 Study Design

The study implemented a ML-based statistical downscaling approach to enhance the spatial resolution of the GRACE/GRACE-FO GWS data from GLDAS, to a fine 5 km scale. The GWS dataset that is provided by GLDAS is referred to as the GLDAS-GRACE GWS dataset in this study. This approach involved pre-processing to standardise the spatial and temporal resolution of input hydrometeorological variables like precipitation, soil moisture and evapotranspiration. The datasets were normalised using the Z-score method and feature engineering was applied to capture complex hydrological interactions. Downscaling was performed using RF and XGBoost models, and the outputs were validated with groundwater level measurements from observation wells. The entire process was implemented in Python, with cartographic work in ArcGIS Pro. Hypothesis testing ensured the statistical significance of the results. This is summarised in Figure 3.

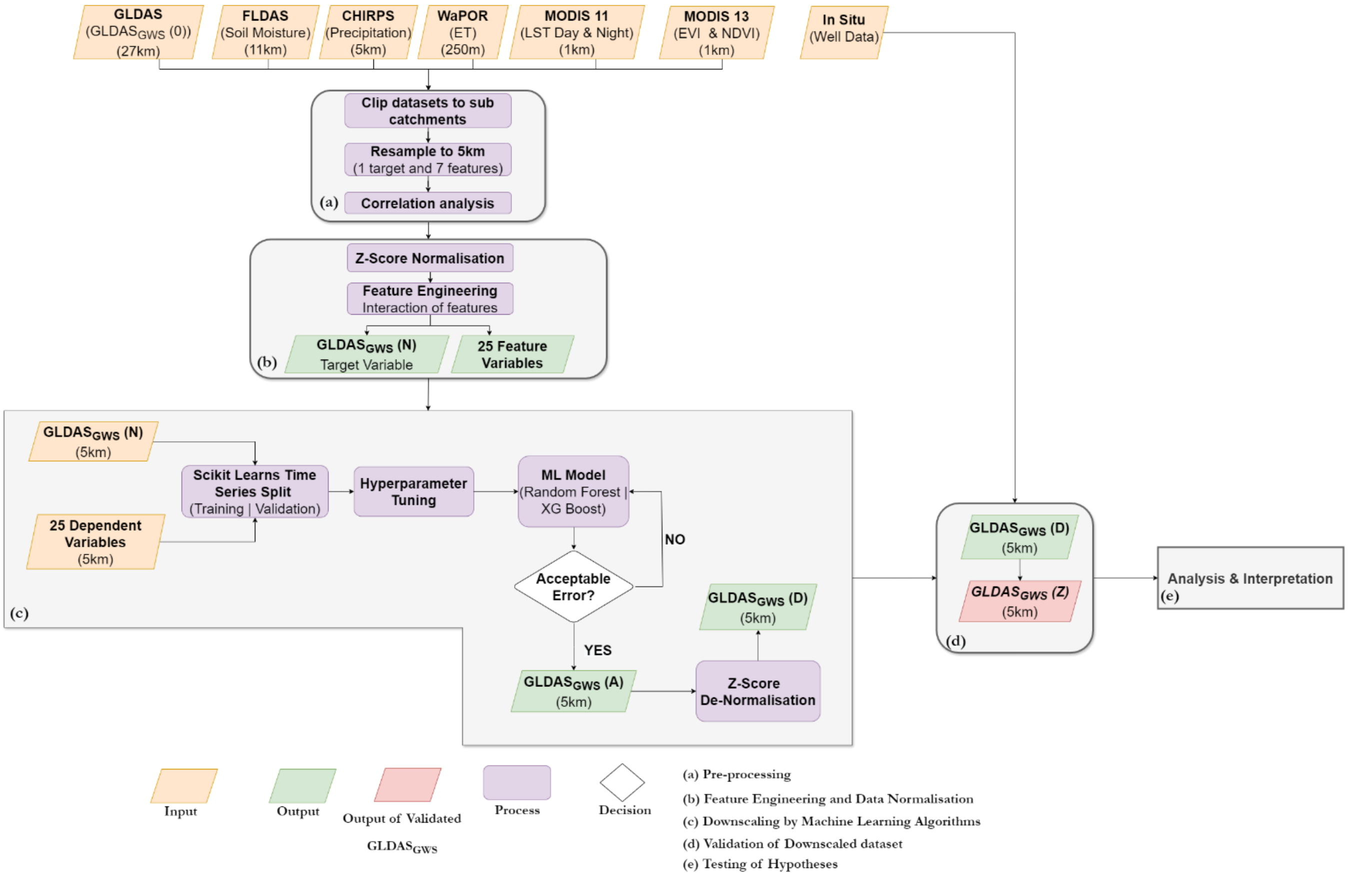


Figure 3: Flowchart of the downscaling and validation process, where: $GLDAS_{GWS}(0)$ is the GLDAS-GRACE GWS at its 27 km spatial resolution; $GLDAS_{GWS}(N)$ is the normalised GLDAS-GRACE GWS at 5 km spatial resolution; $GLDAS_{GWS}(A)$ is the normalised downscaled GLDAS-GRACE GWS at 5 km spatial resolution that was not validated; $GLDAS_{GWS}(D)$ is the de-normalised downscaled GLDAS-GRACE GWS at 5 km spatial resolution that was not validated; and $GLDAS_{GWS}$ (Z) is the de-normalised downscaled GLDAS-GRACE GWS at 5 km spatial resolution that was validated.

4.2 Primary Data

The primary datasets utilised in the research consisted of hydrometeorological variables obtained from remotely sensed sources that were used as inputs for the downscaling process. Their specifications and sources are in Table 1.

Table 1: Description of datasets in the study

Variable	Product name	Time period	Format	Spatial resolution	Temporal resolution	Reference	Website
LST - Day	MODIS	2009-2023	Raster	1 km	Daily	(Wan et al., 2021)	https://lpdaac.usgs.gov/products/mod11a1v061/
LST - Night	MODIS	2009-2023	Raster	1 km	Daily	(Wan et al., 2021)	https://lpdaac.usgs.gov/products/mod11a1v061/
NDVI	MODIS	2009-2023	Raster	1 km	16 Days	(Didan, 2021)	https://lpdaac.usgs.gov/products/mod13a2v061/
EVI	MODIS	2009-2023	Raster	1 km	16 Days	(Didan, 2021)	https://lpdaac.usgs.gov/products/mod13a2v061/
ET	WaPOR	2009-2023	Raster	250 m	10 Days	(FAO, 2020)	https://data.apps.fao.org/wapor/?lang=en
Precipitation	CHIRPS	2009-2023	Raster	5 km	Monthly	(Funk et al., 2015)	https://chc.ucsb.edu/data/chirps
Groundwater	GLDAS	2009-2023	Raster	27 km	Monthly	(Li and Rodell, 2015; Rodell et al., 2004)	https://disc.gsfc.nasa.gov/datasets/GLDAS_CLSM025_DA1_D_2.2/summary
Soil moisture	FLDAS	2009-2023	Raster	11 km	Monthly	(McNally et al., 2022)	https://disc.gsfc.nasa.gov/datasets/FLDAS_NOAH01_C_GL_M_001/summary
Well Data	Local Water Authorities	2018-2020	Point	-	Daily	-	-

4.2.1 Gravity Recovery and Climate Change Experiment (GRACE/GRACE-FO)

The GRACE/GRACE-FO missions have been pivotal in Earth observation, offering valuable data on changes in ice sheets, water storage and tectonic movements. The mission's data is processed by key institutions like, the German Research Centre for Geosciences (GFZ), The National Aeronautics and Space Administrations (NASA), Jet Propulsion Laboratory (JPL), and the University of Texas' Center for Space Research (CSR), each providing unique gravity field solutions. In this study, high-resolution hydrometeorological data was integrated with the CSR derived dataset from GLDAS to enhance localised analysis of terrestrial water storage changes (Tapley et al., 2019; Landerer and Swenson, 2012).

4.2.2 Global Land Data Assimilation System (GLDAS)

The Global Land Data Assimilation System integrates land surface modelling with data from satellites and ground-based observations to provide accurate estimates of land surface states and fluxes. The system, particularly the Catchment Land Surface Model (CLSM) Version 2.2, operates at a spatial resolution of 0.25° (27 kilometres) and generates detailed data on various elements such as soil moisture, snow water equivalent and canopy water storage (Li and Rodell, 2015; Rodell et al., 2004). By using GRACE/GRACE-FO observations, GLDAS effectively estimates global groundwater variability, which is vital for understanding groundwater storage changes at local scales. The integration of GRACE/GRACE-FO CSR solutions into the GLDAS dataset allows for the creation of a dataset termed in this study as GLDAS-GRACE GWS, which provides enhanced groundwater storage estimates crucial for environmental monitoring and water resource management. This study utilises the GWS dataset provided by GLDAS for the downscaling process in the UZC.

Satellite data from programs like GRACE/GRACE-FO, which track variations in TWS, can be incorporated into the water budget equation. The TWS, which is made up of soil moisture, groundwater, surface water and other components, can be remotely estimated owing to these satellite observations. Differentiating the contributions from each component of TWS is crucial for successfully separating changes in GWS from TWS. The water budget equation can be rearranged to represent TWS as the total of these hydrological elements (Wu et al., 2019). GLDAS derives GWS from GRACE/GRACE-FO TWS by modelling and subtracting non-groundwater components, isolating GWS as the residual (Rodell et al., 2004):

$$\Delta GWS = \Delta TWS - \Delta SW - \Delta SM - \Delta IW \qquad \text{Equation 1}$$

where:

(ΔGWS) is the change in groundwater storage;

(ΔTWS) is the change in Terrestrial Water Storage measured by GRACE/GRACE-FO

(ΔSW) is the change in surface water storage;

(ΔSM) is the change in the soil moisture; and

(ΔIW) is the change in Interception Water.

4.2.3 FEWS NET Land Data Assimilation (FLDAS)

The FEWS NET Land Data Assimilation System (FLDAS) Model L4 Global Monthly offers globally optimised data on land surface states and fluxes, which are critical for understanding hydrogeological processes like groundwater recharge (McNally et al., 2022). This study leverages soil moisture data from FLDAS at depths of 100 to 200 centimetres below ground, as it provides a more consistent and reliable indicator of groundwater presence. The deeper soil moisture measurements reflect long-term trends and subsurface processes, making them invaluable for accurate predictions of groundwater storage and availability.

4.2.4 Moderate Resolution Imaging Spectroradiometer (MODIS)

Daily LST data, NDVI and Enhanced Vegetative Index (EVI) are all provided by the NASA Terra satellite sensor Moderate Resolution Imaging Spectroradiometer (MODIS). These datasets are available as the MODIS MOD11A1 and MOD13A2 version 6 datasets, respectively (Didan, 2021; Wan et al., 2021). Understanding surface energy balance, ecosystem health and the consequences of climate change depend heavily on LST data. Green vegetation and biomass assessments require the use of the NDVI and EVI, which are updated every 16 days at a resolution of 1 kilometre. Regions with near-surface groundwater can be identified with the help of both NDVI and EVI, which take into account biomass productivity and atmospheric conditions as well as canopy background.

4.2.5 Water Productivity Open-access Portal (WaPOR)

The Food and Agriculture Organisation (FAO) of the United Nations developed the Water Productivity Open-access Portal (WaPOR), which provides ET datasets for the Near East and Africa. WaPOR version 2.2 uses remote sensing technology to provide comprehensive insights into water usage and stress in various landscapes, assisting with agricultural planning and effective water management techniques (FAO, 2020). Because it allows for more informed decision-making regarding water productivity, this tool is particularly useful in arid and semi-arid regions where water resources are scarce.

4.2.6 Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS)

The Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) is a moderate resolution precipitation dataset available from the Climate Hazards Centre at the University of California, Santa Barbara. A precipitation time series at a spatial resolution of 5 kilometre is produced by CHIRPS, which combines satellite imagery and data from in-situ stations (Funk et al., 2015). This rainfall time series is especially helpful for trend analysis and drought monitoring. This collection, which covers the years 1981 to the present, offers important long-term precipitation data for a range of climatic studies.

4.3 Secondary Data

The secondary data for the study were water levels obtained from observation wells in the study region.

4.3.1 Observation Well Data

Validation of groundwater storage estimates in the UZC is crucial to ensure the accuracy and reliability of the modelled data (Gleeson et al., 2011). For this purpose, the Barotse catchment, a key sub-basin within the UZC, was selected as the validation region. This area was specifically chosen because it houses observation wells in Lukulu, Mongu, and Senanga, which provide direct and continuous groundwater level measurements. These wells measure groundwater levels or head in metres, indicating the height of the water column above a reference point. This head measurement is essential as it directly reflects groundwater availability, providing insight into seasonal variations and long-term trends in the sub-basin (Fan et al., 2013). These wells, located at strategic points in the Barotse, were essential in capturing the spatial and temporal variability of groundwater within the sub-basin. The data from these wells was critical for comparing and validating the downscaled GRACE/GRACE-FO groundwater storage estimates, ensuring that the model outputs accurately reflect the real-world groundwater conditions within the UZC.

4.4 Downscaling Method

This sub-section details the downscaling methodology used in the study.

4.4.1 Pre-processing

In order to ensure that all input variables were consistent in both spatial and temporal resolutions, adjustments were made prior to passing the data to the ML models. The input variables were the dependent variable, GLDAS-GRACE GWS and the independent variables included key hydrological factors like precipitation, soil moisture and evapotranspiration, chosen to represent water balance dynamics (Rzepecka and Birylo, 2020; Wu et al., 2019). The 5 kilometre spatial resolution was achieved by resampling high-resolution data and applying bilinear interpolation for lower-resolution datasets, which preserved spatial patterns. Additionally, all temporal datasets were aggregated into monthly means to match the temporal structure of the GWS dataset, ensuring alignment across all variables. Furthermore, the datasets were clipped to six sub-catchments within the UZC to enable sub-catchment-specific downscaling.

In this study, the datasets for the ML models were assessed to ensure significant predictive power. A correlation analysis was first conducted between groundwater storage (GLDAS-GRACE GWS) and various hydrometeorological variables for the period from January 2009 to December 2023. To refine the data, Seasonal Trend Decomposition using Loess (STL) was applied, isolating the long-term trend from seasonal fluctuations and residual anomaly (Ouyang et al., 2021). The trend components were then used to compute the Pearson, Kendall Tau and Spearman correlation coefficients, which provided insight into the relationships between groundwater storage and the input variables.

Spearman’s rank correlation coefficient was utilised to measure the strength and direction of the ranked variable association, while Kendall-tau examined the concordance and discordance of pairs of observations. Pearson's correlation coefficient assessed the linear relationships between the variables. These correlation metrics were key to ensuring that only relevant and predictive variables were inputted into the ML models, allowing for a more accurate downscaling process (Teng and Chen, 2024; Puth et al., 2015).

These can be calculated using the equations below,

Spearman's rank correlation coefficient (ρ):

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \qquad \text{Equation 2}$$

where:

d_i is the difference between the ranks of corresponding values; and
n is the number of observations.

Kendall tau (τ):

$$\tau = \frac{(C-D)}{\frac{1}{2}n(n-1)} \qquad \text{Equation 3}$$

where:

C is the number of concordant pairs;
D is the number of discordant pairs; and
n is the number of observations.

Pearson's correlation coefficient (r):

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \qquad \text{Equation 4}$$

where:

x_i and y_i are the individual sample points; and
$\bar{x}$ and $\bar{y}$ are the means of the x and y variables, respectively.

4.4.2 Feature Engineering and Data Normalisation

In ML, feature engineering and data normalisation are critical for improving model accuracy. In this study, Z-score normalisation was applied to standardise the variables, reducing the impact of outliers and ensuring consistency. Interaction terms between hydrometeorological variables, such as precipitation and evapotranspiration, were also generated to capture key relationships within the hydrological cycle, enhancing predictive accuracy (Kuhn and Johnson, 2019). Additionally, the datasets were clipped to six sub-catchments in the UZC to ensure the downscaling was tailored to the unique hydrological conditions of each area.

4.4.3 Downscaling by Machine Learning Algorithms

RF and XGBoost algorithms were employed to downscale GLDAS-GRACE GWS data to a finer spatial resolution. The input datasets incorporated a variety of hydrometeorological variables, alongside their interaction terms, which were essential in capturing complex hydrological processes. Prior to model implementation, all variables underwent Z-score normalisation, a process which ensures consistency in the data by standardising the range and distribution of input features, thereby improving model accuracy and reducing the impact of outliers. This approach was critical in enhancing the spatial precision of GWS estimates for the study area.

4.4.3.1 Random Forest

Breiman (2001) developed the random forests algorithm, an ensemble ML algorithm meant to lessen the overfitting that comes with individual decision trees. To do this, during the training phase, several decision trees are built, each based on a bootstrap sample of the dataset, and randomness is introduced by choosing a subset of features at each tree split. When bagging and random feature selection are combined, there is increased model diversity, which lowers overfitting and produces results that are more broadly applicable.

The RF algorithm is essential to this study's downscaling of GRACE/GRACE-FO groundwater storage data. Of particular significance is the method's capacity to simulate intricate relationships between groundwater storage and hydrometeorological input variables. RF models are ideally suited to improve the spatial resolution of GRACE/GRACE-FO derived groundwater estimates because they capture the nonlinear relationships and spatial heterogeneity within the data (Chen et al., 2019). An additional benefit of RF is its built-in permutation feature importance technique (PFIT), which highlights the most important factors affecting groundwater storage and improves the model's interpretability. Furthermore, RF efficiently handles missing data, guaranteeing reliable predictions even in cases where the input data is noisy or incomplete.

4.4.3.2 Extreme Gradient Boosting (XGBoost)

Chen & Guestrin, (2016) introduced extreme gradient boosting, a gradient boosting based algorithm that is highly efficient and well-known for its speed and accuracy when working with large datasets. Unlike random forests, which build trees independently, XGBoost builds trees sequentially, with each tree aiming to correct the errors of the previous one. By resolving the shortcomings of its predecessors, this iterative approach greatly improves model performance. In addition, XGBoost includes anti-overfitting strategies like regularisation by the L1 (Lasso) and L2 (Ridge), which aid in managing model complexity. Moreover, it minimises the chance of overfitting by pruning superfluous branches during the training phase to get rid of parts that do not add much to the model's predictive power. XGBoost also has the benefit of parallel processing, which significantly cuts down on the amount of time needed for training. The algorithm has been effectively utilised for the downscaling of groundwater storage data from GRACE/GRACE-FO, producing outcomes that closely correspond with in-situ measurements (Ali et al., 2023; Zhang et al., 2021).

4.4.3.3 Training and Evaluation of Downscaling Model

In order to implement the ML based downscaling, the GLDAS-GRACE GWS data was the dependent variable, and the independent variables consisted of 25 features, including seven hydrometeorological variables and eighteen interacted terms derived from these variables. To make the datasets compatible with ML models, they were restructuring into 2D arrays. ‘TimeSeriesSplit’ from scikit-learn was utilised during cross-validation to maintain the temporal dependencies in the data. In order to prevent data leakage, a critical step in handling time series data as this method makes sure that the training set always comes before the test set (Hyndman, 2018).

A randomised search cross validation (Randomised Search CV) technique was used to optimise hyperparameters for the RF and XGBoost models. This approach is more effective than a thorough grid search because it takes random samples from a pre-defined hyperparameter space (Zhou et al., 2022). The optimal parameters for the XGBoost model were determined by conducting a sevenfold cross-validation with 700 iterations. These included a subsample rate of 0.7, 1,000 estimators, a maximum tree depth of 11, a learning rate of 0.005 and regularisation constraints

(Zhou et al., 2022). After 350 iterations, the RF model's ideal parameters were determined to be 100 trees, as well as changes to the minimum sample split, leaf size and bootstrap sampling (Breiman, 2001).

Model performance was assessed using R^2, Mean Absolute Error (MAE), NSE, and Root Mean Square Error (RMSE). R^2 indicates the proportion of variance explained by the model, MAE provides the average magnitude of prediction errors, NSE evaluates the predictive power of hydrological models, and RMSE measures the average prediction error magnitude, reflecting the model's accuracy (Willmott and Matsuura, 2005). These metrics can be calculated as below.

$$R^2 = 1 - \frac{\Sigma\,(y_i - \hat{y}_i)^2}{\Sigma\,(y_i - \bar{y})^2} \qquad \text{Equation 5}$$

$$MAE = \frac{1}{n}\Sigma|y_i - \hat{y}_i| \qquad \text{Equation 6}$$

$$\text{NSE} = 1 - \frac{\Sigma\,(y_i - \hat{y}_i)^2}{\Sigma\,(y_i - \bar{y})^2} \qquad \text{Equation 7}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(\hat{y}_i - \hat{y}_i)^2} \qquad \text{Equation 8}$$

where:

y_i is the observed value;
$\hat{y}_i$ is the predicted value;
$\bar{y}$ is the mean of the observed values; and
n is the number of observations.

4.4.4 Validation of Downscaled Groundwater Estimates

Given the discrepancies that frequently occur between in-situ measurements and remotely sensed data, validating groundwater storage estimates derived from GRACE/GRACE-FO data is essential (Tariq et al., 2023). It is necessary to convert water level readings from ground-based observation wells into water storage (WS) anomalies as part of this validation process, which calls for an understanding of the aquifer's unique properties (Scanlon et al., 2012). For the transformation in this study, observation well data from the Barotse Flood Plain's Kalahari unconfined aquifer were utilised. A key equation was used to translate the water level data into water storage anomalies. This equation took into account the physical characteristics of the aquifer, such as porosity and specific yield, and allowed for a more precise comparison with estimates derived from GRACE/GRACE-FO (Nenweli et al., 2024).

$$\Delta GWS, unconfined = Sy * \Delta h \qquad \text{Equation 9}$$

where:

(ΔGWS) is the groundwater storage derived from the specific yield value and the water level;

(Sy) is the specific yield of the aquifer; and

(Δh) is the water level in the observation well.

4.4.5 Testing of Hypotheses

In this study, the link between wet season rainfall and groundwater recharge was investigated using the GWS data from 2009 to 2023, across three groundwater storage datasets: GLDAS-GRACE; XGBoost; and RF. The wet season, defined from October to May, was critical for identifying rainfall thresholds that significantly impact groundwater recharge. By applying a percentile thresholding method, the 10th and 90th percentiles of groundwater storage volumes were computed to determine the bounds of normal recharge, as induced by the cumulative rainfall in the wet season (Zhang et al., 2024; Shilengwe et al., 2023; Wu et al., 2019). El Niño years (2015, 2016, 2018, 2019 and 2023), known for causing anomalies, were excluded, ensuring the analysis focuses on typical wet season conditions. For each percentile, corresponding rainfall values were identified, shedding light on the relationship between rainfall and recharge.

CHAPTER FIVE: RESULTS AND DISCUSSION

This chapter interprets the results of the study and discusses them. The downscaling and subsequent analyses were conducted at the sub-catchment level to avoid leakage and allow the ML models to account for the heterogeneous hydrogeology in the UZC as reported by Shilengwe et al. (2024).

5.1 Correlation of Hydrometeorological Variables with Groundwater Storage

A correlation analysis was carried out to evaluate the hydrometeorological variables' association with the GLDAS-GRACE GWS dataset and to direct the selection of pertinent inputs for the downscaling models. This step was crucial to ensure that the downscaling process only included the variables that had the strongest correlation with groundwater storage. The analysis offered a thorough grasp of the linear and nonlinear relationships that affect groundwater dynamics by assessing important variables. This foundation increased the precision of groundwater storage predictions at finer spatial resolutions (Table 2).

In the Barotse region, NDVI and groundwater storage showed the strongest correlation, with a strong positive relationship (Pearson: 0.5669), suggesting a strong relationship between groundwater availability and vegetation health. Given that deep-rooted plants access subsurface moisture, vegetation in many ecosystems reflects groundwater dynamics, which is consistent with findings that link NDVI to groundwater storage (Blank et al., 2023; Martínez-de la Torre and Miguez-Macho, 2019). When compared to NDVI, the EVI showed the weakest correlation (Pearson: 0.2912), which might indicate that this index is less sensitive to changes in groundwater similar to what was found in the study by Arast et al. (2020).

There was a significant inverse relationship (Pearson: -0.6514) between the LST-Night and groundwater storage in the Chobe region. This relationship is seen in semi-arid areas where warmer temperatures can increase evaporation and decrease groundwater recharge which is in line with research in similar regions by Malik et al. (2021). Groundwater dynamics in this area were not closely linked to vegetation health as determined by EVI, as evidenced by the weakest correlation (Pearson: 0.1280) (Rashid et al., 2024).

Soil moisture showed the strongest positive correlation (Pearson: 0.5658) for the Kabompo region, indicating how important soil moisture is in regulating groundwater levels. Comparable results have been noted in other locations where soil moisture has a significant impact on groundwater recharge, especially in water-limited environments (Arast et al., 2020). The weakest correlation (Pearson: -0.1871) was found between LST night and GWS, indicating that groundwater storage is not significantly impacted by night time temperature changes.

Soil moisture was once again found to be the most significant factor associated with groundwater storage in the Luanginga region (Pearson: 0.7610). Given that soil moisture promotes groundwater system infiltration and recharge, this strong correlation highlights the fundamental role soil moisture plays in determining groundwater levels, which is a common finding in similar research (Rashid et al., 2024). On the other hand, EVI emphasises on changes in surface vegetation that might not directly correlate to groundwater variations, it showed the weakest correlation (Pearson: 0.0525) in this region, which is consistent with other regions where this index may not adequately capture groundwater dynamics.

The idea that soil moisture is a key component of groundwater recharge was further supported by the fact that it was also the most significant variable for Lunga (Pearson: 0.7660). In regions with seasonal rainfall, where soil moisture serves as a buffer that replenishes groundwater reserves during dry spells, the relationship between soil moisture and groundwater is particularly strong (Rashid et al., 2024). The weakest correlation was found between LST daytime surface temperatures (Pearson: -0.0293), suggesting that groundwater levels in this area are not greatly impacted by daytime surface temperatures.

Furthermore, ET showed the strongest positive association with groundwater storage in the Upper Zambezi (Pearson: 0.3582). This relationship brings to light the interaction between groundwater and evapotranspiration, as higher ET rates can decrease groundwater recharge by increasing atmospheric water loss. Numerous investigations have supported this relationship, demonstrating that evapotranspiration can have a major effect on groundwater dynamics, especially in areas with a lot of vegetation cover (Zhu et al., 2015). The EVI once again showed the weakest correlation (Pearson: -0.2764), indicating that this vegetation index has a limited ability to explain groundwater variability in this area.

When using XGBoost and RF models to predict groundwater storage from GLDAS-GRACE data, the results of the PFIT (Appendix 1) present consistent findings across all regions. Rainfall and soil moisture were found to be the most important predictors across all regions by XGBoost, which is in line with previous research showing that precipitation has a major impact on groundwater dynamics, especially in semi-arid regions (Ali et al., 2023; Tao et al., 2022a). This intense emphasis on soil moisture and rainfall highlights how effectively XGBoost uses the most predictive variables for estimates of groundwater storage.

On the other hand, the RF model displayed a more evenly distributed distribution of significance across a larger number of features, such as ET, LST and soil moisture. This more comprehensive technique enabled RF to take into consideration the complex nature of hydrological processes, offering a thorough comprehension of groundwater variability (Madani et al., 2022; Tyralis et al., 2019). The ability of RF to incorporate a variety of environmental variables is demonstrated by the inclusion of extra factors like LST and NDVI in RF models (Pham et al., 2022).

Table 2: Correlations between GLDAS-GRACE GWS dataset with input hydrometeorological variables

Region	Correlation metric	Correlation results	Dataset						
			Precipitation	ET	LST - Day	NDVI	EVI	LST - Night	Soil Moisture
Barotse	Pearson	Correlation	0.5395	0.3315	-0.3105	0.5669	0.2912	-0.5715	0.5031
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kendall Tau	Correlation	0.3944	0.2494	-0.2184	0.3891	0.2001	-0.3804	0.3886
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spearman	Correlation	0.5662	0.3385	-0.3051	0.5603	0.3015	-0.5267	0.5274
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Chobe	Pearson	Correlation	0.6022	0.3124	-0.2701	0.3290	0.1280	-0.6514	0.5269
		p-value	0.0000	0.0000	0.0002	0.0000	0.0869	0.0000	0.0000
	Kendall Tau	Correlation	0.3808	0.0803	-0.0809	0.1281	-0.0386	-0.4581	0.3511
		p-value	0.0000	0.1094	0.1067	0.0107	0.4416	0.0000	0.0000
	Spearman	Correlation	0.5577	0.1338	-0.1169	0.2018	0.0075	-0.6486	0.5210
		p-value	0.0000	0.0733	0.1183	0.0066	0.9203	0.0000	0.0000
Kabompo	Pearson	Correlation	0.4052	0.2159	-0.3917	0.4244	0.5171	-0.1871	0.5658
		p-value	0.0000	0.0036	0.0000	0.0000	0.0000	0.0119	0.0000
	Kendall Tau	Correlation	0.2334	0.0906	-0.2742	0.2961	0.3485	-0.1388	0.3714
		p-value	0.0000	0.0709	0.0000	0.0000	0.0000	0.0057	0.0000
	Spearman	Correlation	0.3407	0.2119	-0.4246	0.4460	0.4817	-0.2058	0.5272
		p-value	0.0000	0.0043	0.0000	0.0000	0.0000	0.0056	0.0000
Luanginga	Pearson	Correlation	0.5201	0.4649	-0.2563	0.5134	0.0525	-0.4882	0.7610
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Kendall Tau	Correlation	0.3194	0.2484	-0.0564	0.2549	-0.0951	-0.2708	0.5433
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spearman	Correlation	0.4790	0.3740	-0.0575	0.3738	-0.1084	-0.4104	0.7656
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Lunga	Pearson	Correlation	0.4866	0.3309	-0.0293	0.1685	-0.0768	-0.3402	0.7660
		p-value	0.0000	0.0000	0.6958	0.0238	0.3058	0.0000	0.0000
	Kendall Tau	Correlation	0.3392	0.1830	0.0859	-0.0742	-0.1908	-0.2178	0.5734
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Spearman	Correlation	0.4949	0.2963	0.1173	-0.1145	-0.2359	-0.3123	0.7480
		p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Upper Zambezi	Pearson	Correlation	0.0858	0.3582	-0.0946	0.0782	-0.2764	-0.0159	0.3197
		p-value	0.2522	0.0000	0.2063	0.2968	0.0002	0.8319	0.0000
	Kendall Tau	Correlation	0.0526	0.2403	-0.0098	-0.0288	-0.0941	-0.0554	0.2257
		p-value	0.2941	0.0000	0.8450	0.5659	0.0607	0.2698	0.0000
	Spearman	Correlation	0.0985	0.3777	-0.0084	-0.0438	-0.1362	-0.0904	0.3264
		p-value	0.1882	0.0000	0.9111	0.5597	0.0682	0.2273	0.0000

5.2 Performance of Machine Learning Models in Downscaling

The effectiveness of XGBoost and RF in downscaling GWS data was assessed for a number of sub-catchments in the UZC, revealing key insights into their strengths and weaknesses. Across smaller regions like Luanginga and Lunga, RF consistently produced better results in terms of, lower errors (MAE as well as RMSE), smoother scatter plots and higher correlation values (R^2), as depicted in, Figure 4, Table 3, Appendices 2 and 3. The key to RF's effectiveness in these situations was its ability to combine several variables and minimise overfitting by averaging decision trees. This is consistent with research that shows how well RF handles complex environmental data, guaranteeing prediction stability and lower variance (Tyralis et al., 2019; Breiman, 2001).

On the other hand, scatter plots showed irregular patterns in larger sub-catchments such as Chobe, especially with XGBoost. These discontinuities imply that model performance declines with increasing sub-catchment size, possibly as a result of data leakage errors or the heterogeneity of hydrological processes over larger regions (Rodriguez-Galiano et al., 2012; Rodriguez et al., 2010). Although the larger area increased prediction variability, as evidenced by the higher error metrics and scatter plot deviation from the fit line, RF still outperformed XGBoost.

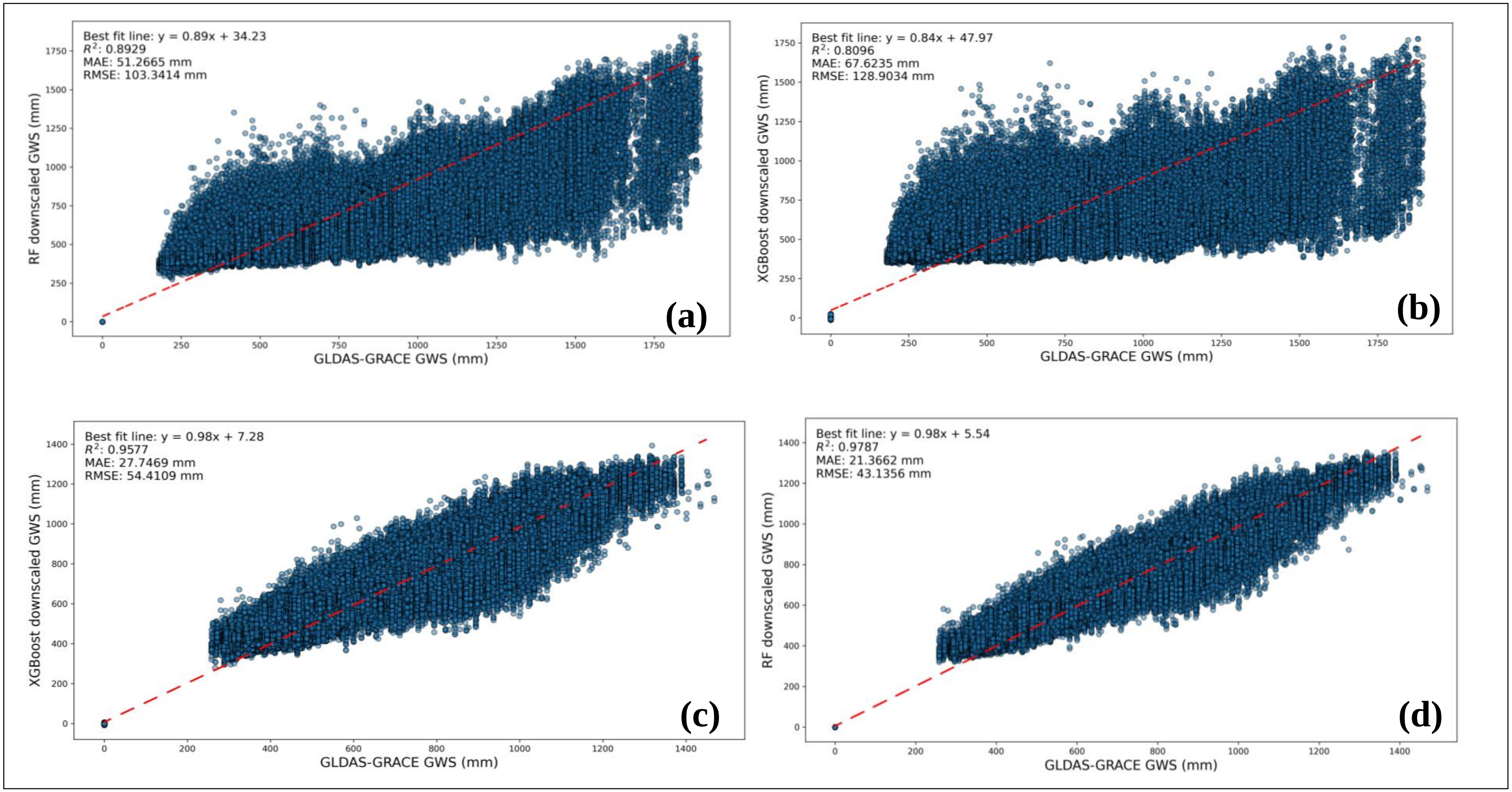


Figure 4: The GLDAS-GRACE GWS and the results from the two downscaling models are shown in scatter plots, where: (a) and (b) represent the scatter plots for the model in the Barotse for the RF and XGBoost models respectively; (c) and (d) represent the scatter plots for the model in the Luanginga for the RF and XGBoost models respectively

Table 3: The two downscaling models' error metrics demonstrating the difference in their predictive power

Sub-catchment	Barotse		Chobe		Kabompo		Luanginga		Lunga		Upper Zambezi	
Metric	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**
MAE (mm)	51.2665	67.6235	28.3004	38.7938	92.8631	131.1274	21.3662	27.7469	11.4688	19.5782	38.3953	63.6605
RMSE (mm)	103.3414	128.9034	62.0562	78.3993	171.7332	232.1686	43.1356	54.4109	28.9766	45.5584	73.1869	117.5097
R^2	0.8929	0.8096	0.9369	0.894	0.9018	0.7808	0.9787	0.9577	0.9913	0.9743	0.9803	0.9391
NSE	0.8929	0.8096	0.9369	0.894	0.9018	0.7808	0.9787	0.9577	0.9913	0.9743	0.9803	0.9391
Scatter Plot	strong positive correlation		discontinuous prediction variance		strong positive correlation		excellent positive correlation		excellent positive correlation		excellent positive correlation	
Area (km^2)	106,732		154,743		72,523		41,931		46,853		93,443	

5.3 Spatiotemporal Improvement of Groundwater Storage Estimates

The findings demonstrate how downscaling GLDAS-GRACE GWS data from 27 kilometres to 5 kilometres using XGBoost and RF models significantly improves the spatial resolution of the dataset. In contrast to the original coarse dataset, the downscaled products provide more localised and detailed spatial patterns, as shown by the visual comparison of the datasets in Figure 5. By capturing finer details of hydrological processes that are not visible in the coarser GLDAS-GRACE GWS data, the downscaled models' improved spatial resolution makes it possible to depict groundwater storage anomalies throughout the study area more clearly which is in line with the findings in the studies by Satizabal-Alarc et al. (2023) and Sahour (2020).

In this downscaling process, the RF and XGBoost models both make use of distinct regression capabilities. RF provides a more balanced and comprehensive integration of environmental factors into the predictions by distributing the importance across a wider range of variables, such as LST, NDVI and ET, whereas XGBoost tends to concentrate more on a small number of highly influential variables (Wu et al., 2019; Yin et al., 2018). These differences are visible in the improved spatial resolution, where XGBoost concentrates on attaining greater accuracy in particular significant regions, whereas RF models frequently offer more fine-grained spatial detail as a result of their variable integration.

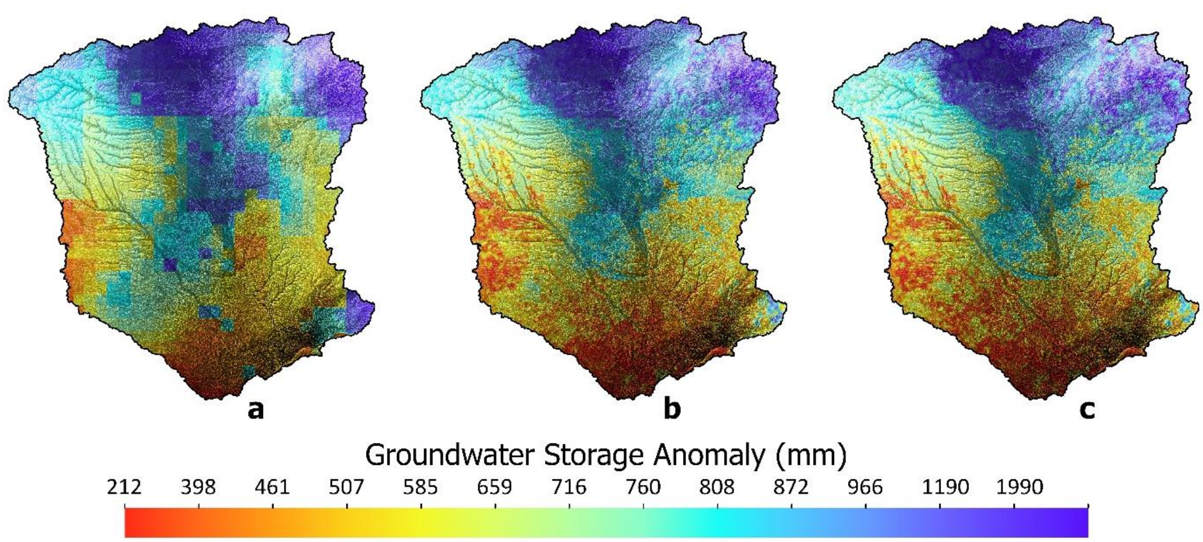


Figure 5: The spatial improvement between the GLDAS-GRACE GWS and the two downscaled datasets, datasets from the January 2015 study in the UZC. Where: (a) GLDAS-GRACE GWS at

27 km spatial resolution; (b) downscaled GWS from the XGBoost model at 5 km spatial resolution; and (c) downscaled GWS from the RF model at 5 km spatial resolution

5.4 Validation of Downscaled Groundwater Storage Estimates

The validation of the downscaled GLDAS-GRACE GWS dataset was performed using three observation wells located in the Barotse sub-catchment, which allowed for an evaluation of the groundwater fluctuations and capabilities of the RF and XGBoost models as reported by Shilengwe et al. (2024). By extending the findings to infer results across the additional five sub-catchments (Figure 6), which share similar hydrogeological characteristics, this validation is being integrated into the larger study (Brutsaert, 2023). Although this method contributes to a more comprehensive understanding of model performance, it has drawbacks, chief among them being the absence of direct observational data in the other regions. Bhanja et al. (2016) found that comparable hydrogeological conditions enable comparable validation across nearby sub-catchments, but this assumption adds uncertainty, particularly in light of possible localised variations in the dynamics of groundwater storage.

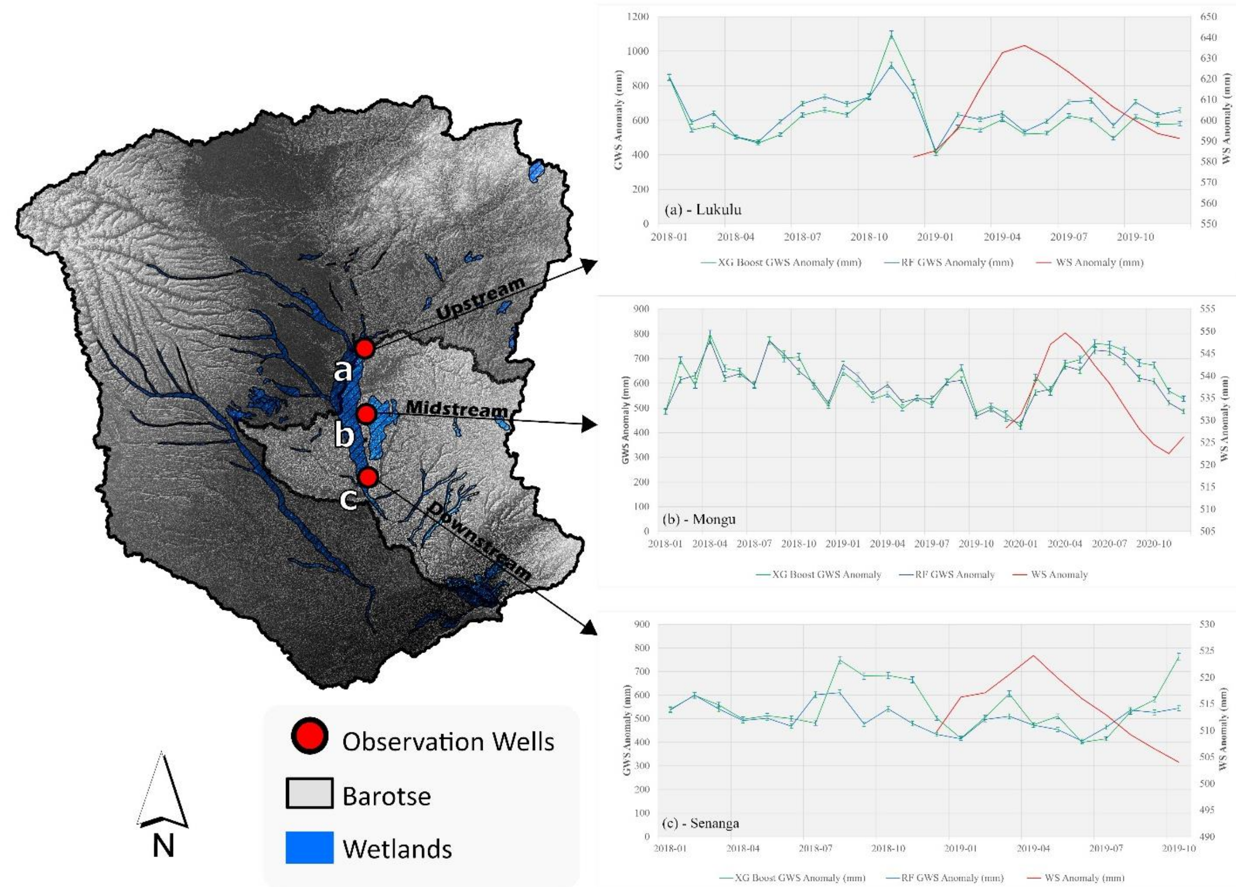


Figure 6: The time series of the RF and XGBoost GWS anomalies in the Barotse sub-catchment were plotted on the primary axis, while the observed water storage anomalies from well data were plotted on the secondary axis. These plots were generated for observation wells in the following locations, (a) Lukulu district hospital in Lukulu, (b) Department of Water Resource Development (DWRD) in Mongu, and (c) Lukanda primary school in Senanga

It was observed that the XGBoost and RF models, employed for downscaling GWS estimates exhibited varying degrees of accuracy in predicting groundwater fluctuations (Table 4).

At DWRD in Mongu, the XGBoost model demonstrated limited efficacy in accurately modelling groundwater dynamics, whereas the RF model exhibited superior performance. This suggests that the RF model is more reliable in this mid- stream region, which aligns with other studies in similar regions (Nafouanti et al., 2023). Similar outcomes were observed up stream at Lukulu hospital, where the RF model again outperformed the XGBoost model. This indicates that the RF model can effectively predict groundwater variations even in upstream areas (Adib et al., 2023). Conversely, at Senanga's Lukanda primary school, both models encountered challenges in accurately modelling groundwater behaviour. This could be attributed to the presence of artesian wells, which are characterised by pressurised groundwater storage, leading to delayed responses to surface recharge events (Dan et al., 2021).

The downscaling of GLDAS-GRACE GWS estimates had significant limitations, despite its promising outcomes. The dependence on the quality and spatial resolution of the input hydrometeorological data is one of the main obstacles. The downscaled outputs can be severely distorted by discrepancies or errors in these input datasets, which can spread throughout the model (Famiglietti and Rodell, 2013).

Furthermore, locations with complex hydrogeological conditions such as confined aquifers with disconnected recharge zones and lagged responses to surface conditions are less amenable to the models' effectiveness (Feng et al., 2018). Model performance is also impacted by the temporal resolution of the input data in that it restricts the models' ability to precisely depict short-term groundwater shifts (Serravalle Reis Rodrigues et al., 2023). Additionally, due to the short-term temporal coverage of the observation well data, the validation of downscaled GWS estimates especially in this study was restricted, which could have resulted in exaggerated correlation results in some cases (Bennett, 2024). This emphasises that in order to improve model validation and further hone the downscaling process, more well data and long-term monitoring are required (Ghorbani et al., 2013; Kusche et al., 2009).

Table 4: Correlation analysis between downscaled GLDAS-GRACE groundwater storage estimates and observed well data represented as water storage anomalies in the Barotse sub-catchment

Location	Model	Pearson		Kendall Tau		Spearman		Interpretation
		Correlation	p-value	Correlation	p-value	Correlation	p-value	
DWRD, Mongu	XGBoost	-1	0	-0.8563	0.0207	-0.8847	0.0081	Strong negative linear relationship (Pearson), Strong negative monotonic relationship (Kendall and Spearman)
	RF	1	0	0.6667	0.0634	0.7283	0.0635	Strong positive linear relationship (Pearson), Moderate positive monotonic relationship (Kendall and Spearman)
Lukanda Primary School, Senanga	XGBoost	0.6008	0.1537	0.2222	0.5133	0.3846	0.3943	Moderate positive linear relationship (not significant, Pearson), Weak positive monotonic relationship (not significant, Kendall and Spearman)
	RF	-0.3301	0.4697	-0.7379	0.0265	-0.8975	0.0061	Weak negative linear relationship (not significant, Pearson), Strong negative monotonic relationship (Kendall and Spearman)

Lukulu Hospital, Lukulu	XGBoost	1	0	0.8563	0.0207	0.8847	0.0081	Strong positive linear relationship (Pearson), Strong positive monotonic relationship (Kendall and Spearman)
	RF	1	0	1	0.0089	1	0	Strong positive linear relationship (Pearson), Strong positive monotonic relationship (Kendall and Spearman)

5.5 Impact of Drought on Groundwater Resources

In this study, GWS data from three different models was analysed, namely: GLDAS-GRACE; XGBoost; and RF. The analysis examined the impact of El Niño years from 2009 to 2023, specifically 2015, 2016, 2018, 2019 and 2023. Precipitation thresholds that were determined by calculating the 10th and 90th percentiles of GWS volume across all models are depicted in Table 5. The XGBoost model had a 10th percentile value of 351.20 billion m³, equivalent to 592.99 millimetres of rainfall. The RF and GLDAS-GRACE datasets showed similar percentile values (Taylor et al., 2013; Rodell et al., 2009)

GWS volumes and rainfall exceeded the 90th percentile thresholds in 2009, 2010, 2013 and 2017, according to all the datasets. In 2009, XGBoost recorded 430.77 billion m³ of GWS, which equates to 1045.02 millimetres cumulative rainfall in the wet season. Higher rainfall in non-El Niño years may promote groundwater recharge, as El Niño events were not present. This finding is consistent with previous research that has highlighted the link between consistent rainfall and stable groundwater recharge (Tanachaichoksirikun and Seeboonruang, 2019; Jasechko and Taylor, 2015).

Table 5: The amount of precipitation availability-based GWS thresholds, with the 90th percentile indicating the upper-bound threshold and the 10th percentile indicating the lower-bound threshold

	GLDAS-GRACE GWS		**RF GWS**		**XGBoost GWS**	
	GWS Volume (billion m³)	**Rainfall (mm)**	**GWS Volume (billion m³)**	**Rainfall (mm)**	**GWS Volume (billion m³)**	**Rainfall (mm)**
Upper Threshold	439.24	776.82	441.55	882.83	439.2	882.83
Lower Threshold	350.74	614.39	352.77	614.39	351.2	614.39

GWS volumes consistently fell below the 10th percentile thresholds, and reduced rainfall was a consistent feature in El Niño years. For instance, the XGBoost dataset shows that GWS volumes were only 375.58 billion m³ in 2015, indicating less recharge, even though there was 886.89 millimetres of rainfall. According to earlier research, El Niño-related climate patterns disrupt regular rainfall, which reduces groundwater recharge (Cuthbert et al., 2019; Trenberth, 2011). This trend persisted in other El Niño years (Figure 7).

A distinct pattern emerges from the analysis, El Niño years are characterised by lower than normal rainfall, which reduces groundwater recharge. On the other hand, years like 2009 and 2010 that were not El Niño were linked to more rainfall and recharge. Even in cases where rainfall exceeded anticipated thresholds, the differences between GWS and rainfall imply that variables other than rainfall, such as surface runoff, evapotranspiration and changes in land use, are important determinants of recharge rates (Noori et al., 2023; Lerner, 2020; Healy, 2010). This highlights the necessity of detailed modelling techniques and high-resolution data to capture the intricacies of groundwater recharge, particularly during times of climatic variability (Taylor et al., 2013; Rodell et al., 2009).

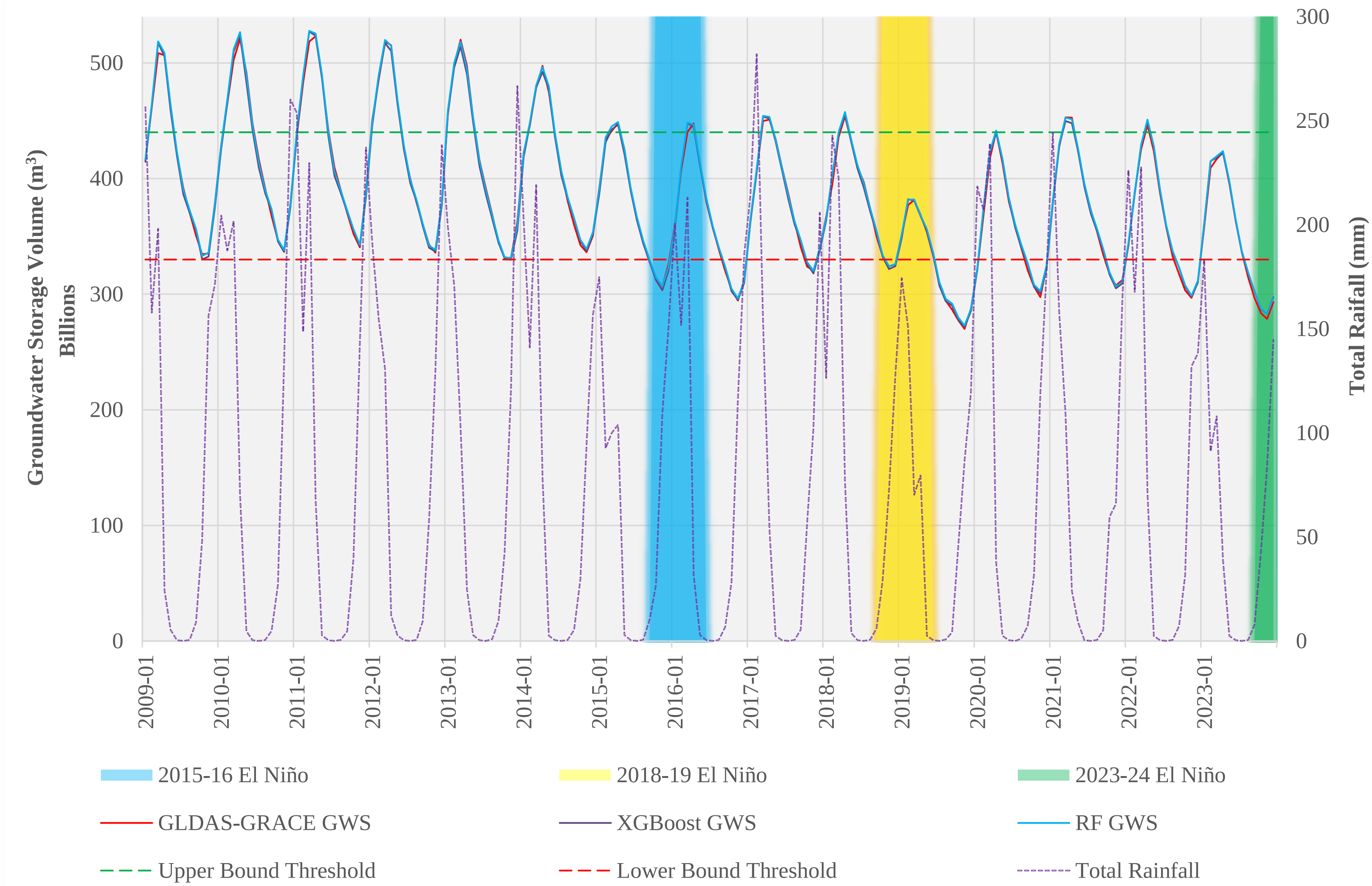


Figure 7: The primary vertical axis shows the GLDAS-GRACE GWS volume and the predicted GWS volume from the XGBoost and RF models, while the secondary vertical axis shows the total monthly CHIRPS rainfall (01-2009 to 12-2023)

5.6 Spatiotemporal Variation of Groundwater Storage

Analysis of the spatiotemporal changes in GWS from 2009 to 2023 reveals both notable gains and losses in the UZC (Figures 8 and 9). The south-western and southern portions of the study area, around the Barotse and Chobe, where anomalies range from -100 millimetres to -433 millimetres, are where GWS losses are most prevalent (Singh and Saravanan, 2020). Regular groundwater depletion occurred in these areas, most likely as a result of, excessive extraction for domestic and agricultural purposes, groundwater outflow and less frequent rainfall during El Niño years, which aligns with studies in similar regions (Ndehedehe, 2022). Increased surface runoff and decreased infiltration during times of intense but brief rainfall are additional factors that may have contributed to this depletion (Singh and Raju, 2020).

However, as can be observed in the north-eastern portion of the UZC, the regions around the Kabompo and Upper Zambezi sub-catchments exhibit notable GWS gains, with anomalies up to +264 millimetres. These grains are probably the result of more rainfall and better groundwater recharge conditions (Nazari et al., 2023). Better land management techniques or less groundwater extraction may have contributed to these regions with higher GWS gains, which would have led to a steady rise in storage over time (Sokneth et al., 2022). Different geological and hydrological characteristics, which regulate how water percolates and is stored in the subsurface, could also have contributed to the variations in recovery and depletion rates throughout the study area (Solihuddin et al., 2024).

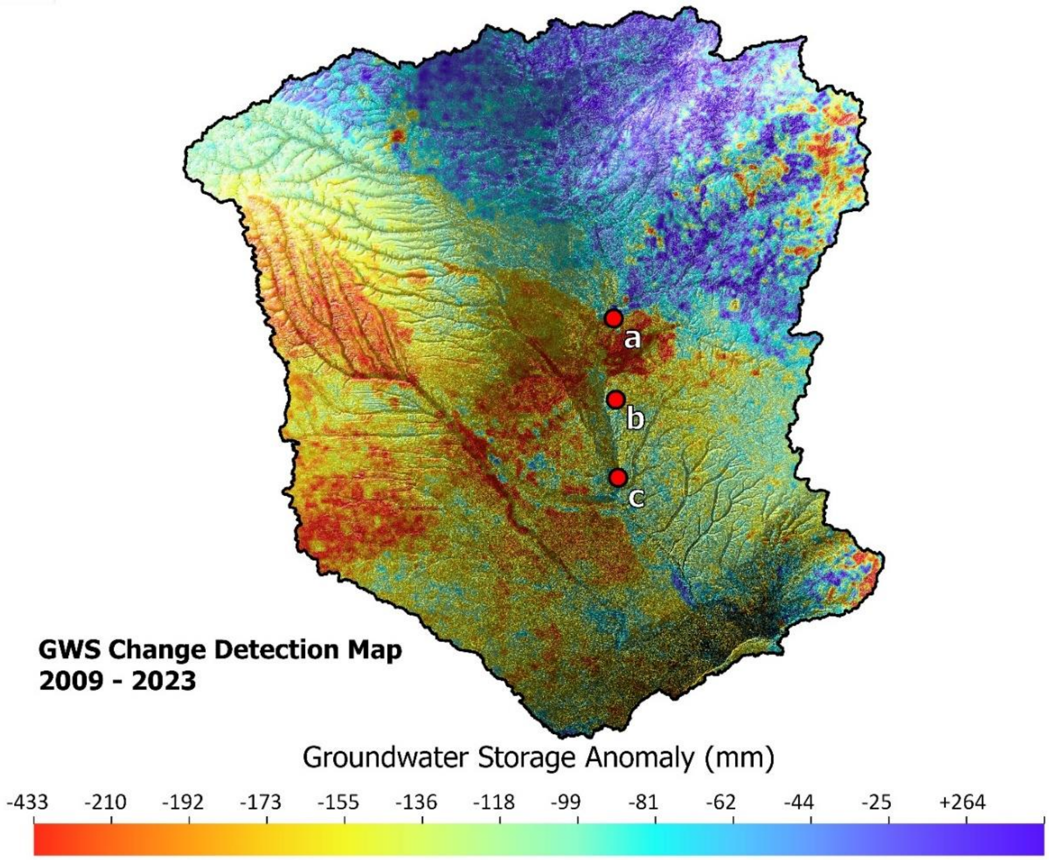


Figure 8: The RF GWS multi-year mean change detection map from 2009 to 2023 in the UZC, where (a) is the Lukulu district hospital in Lukulu, (b) the DWRD in Mongu and (c) the Lukanda primary school in Senanga

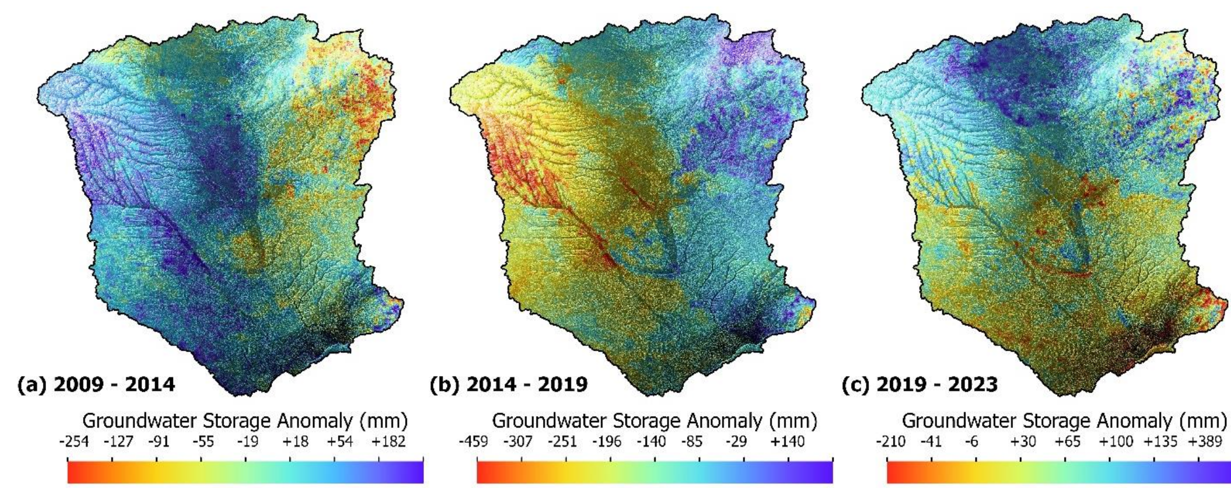


Figure 9: The RF GWS multi-year mean change detection maps in the UZC, for the time periods, (a) 2009 to 2014, (b) 2014 to 2019 and (c) 2019 to 2023

5.7 Long-Term Trend of Groundwater Storage

Significant changes in response to El Niño years are revealed by the trends in GWS volume across the three datasets (GLDAS-GRACE, XGBoost and RF) as depicted in Figures 10 and 11. The influence of decreased rainfall and increased drought conditions linked to El Niño events was reflected by the consistent decrease of GWS volumes across all datasets during these periods (Malik and Ford, 2024; Jeihouni et al., 2019). For instance, all three datasets indicate that a considerable drop in GWS levels occurred in 2015 and 2016 as a result of the drought that worsened throughout Southern Africa, including Zambia. This is consistent with the results of multiple studies that emphasise how El Niño intensifies droughts in Southern Africa, which in turn impacts groundwater recharge (Fortnam et al., 2021). This pattern was also observed in 2018 and 2019, further confirming the correlation between El Niño episodes and diminished groundwater storage (Siderius et al., 2018)

The decline in GWS trends starting in 2015 was a component of a larger regional drought pattern in which longer dry spells and less rainfall resulted in less groundwater recharge, especially in Southern African nations (Goddard and Gershunov, 2020). For instance, during the El Niño that occurred in 2015–2016, a significant portion of Southern Africa saw a decrease in lake levels and river runoff, which had an adverse effect on groundwater levels (Santoso et al., 2017). According to the data, GWS levels sharply decreased during these years, which is indicative of greater hydrological stress in the area. This pattern is consistent with hydrological research that highlights how less precipitation during El Niño years causes less water to infiltrate into aquifers, leading to a long-term reduction in groundwater supplies (Wu et al., 2018).

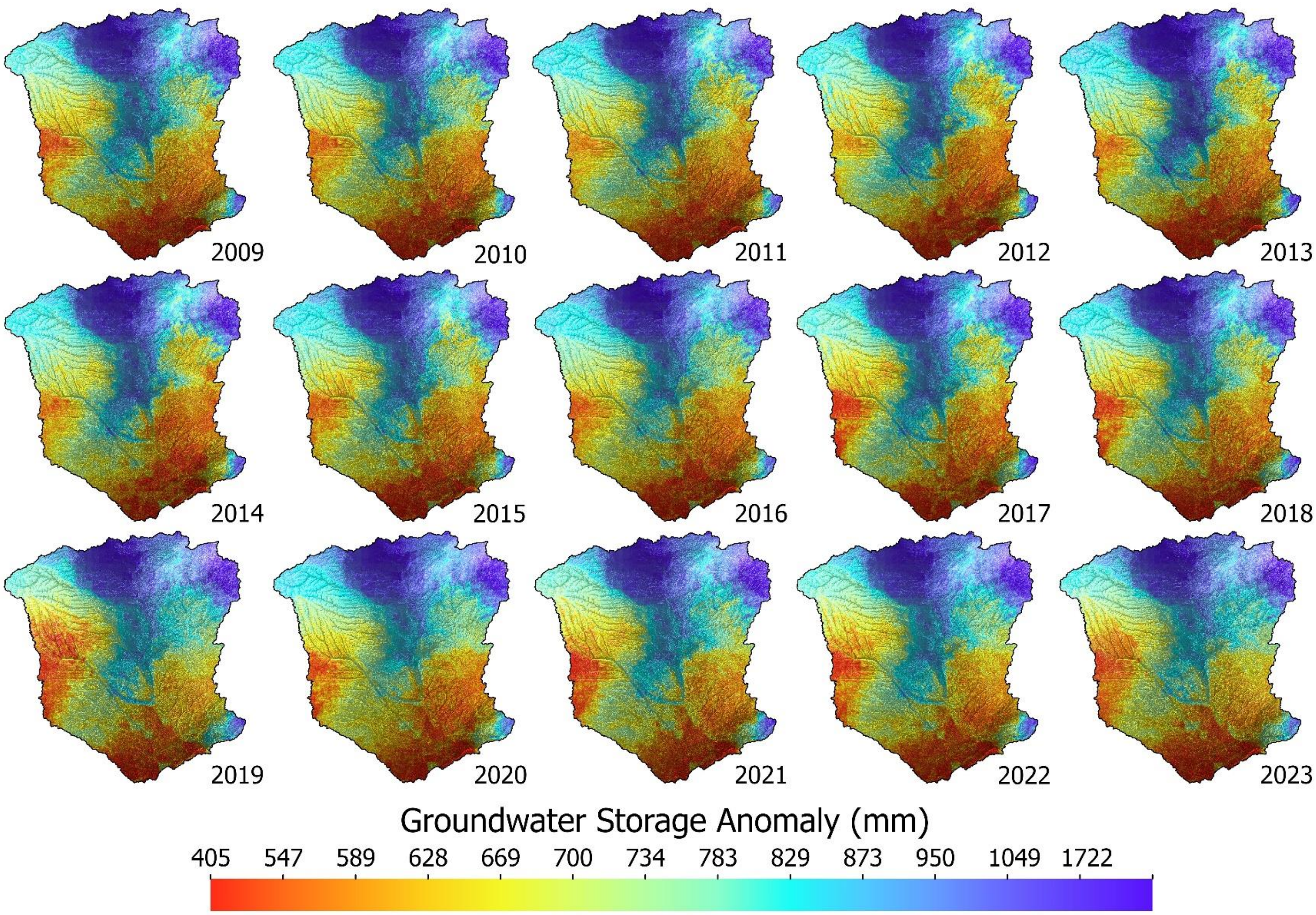


Figure 10: The RF GWS dataset's multi-year spatiotemporal variation, represented by the yearly means from 2009 to 2023

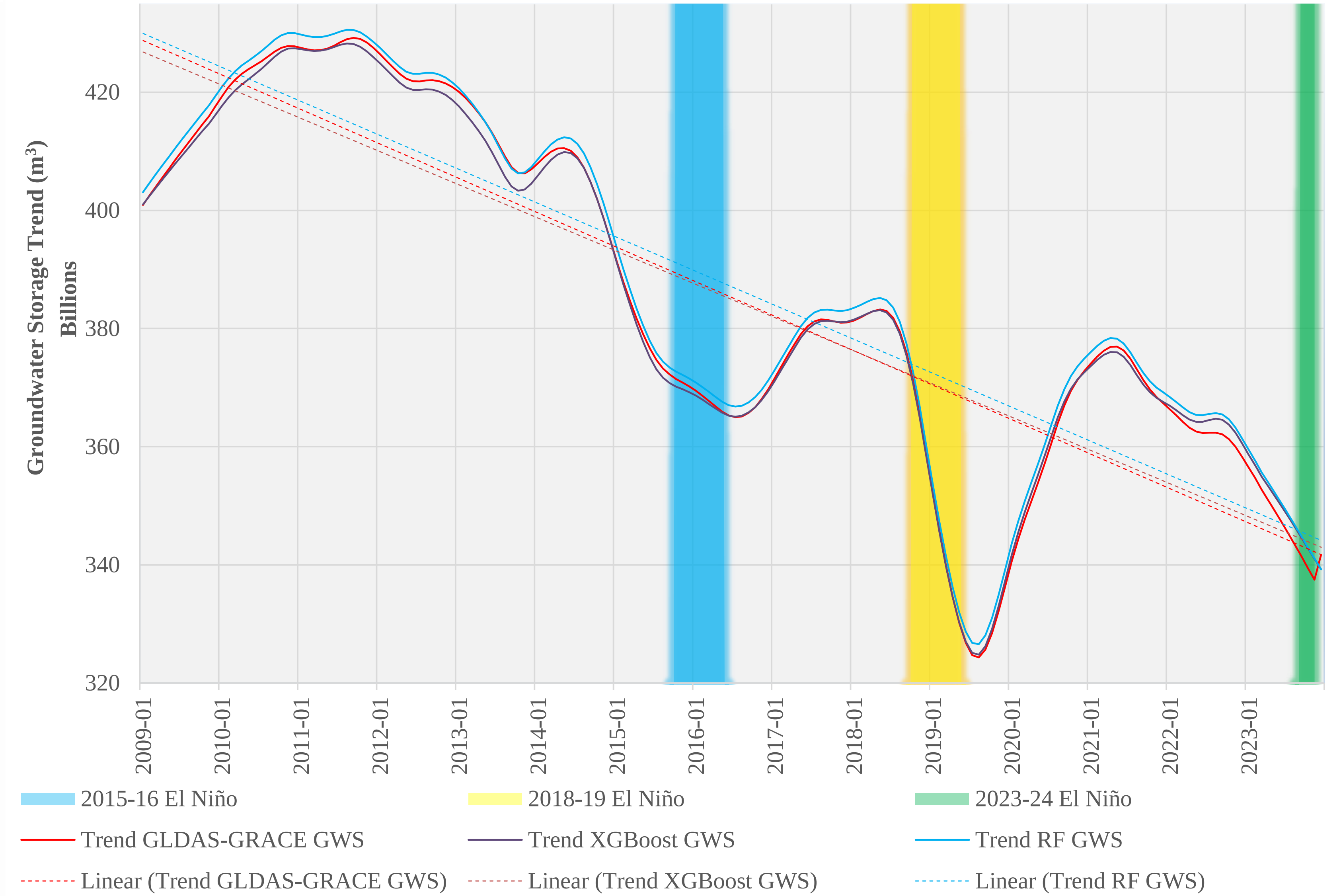


Figure 11: STL decomposition reveals trend components for GLDAS-GRACE GWS, XGBoost GWS and RF GWS models on the primary vertical axis from 2009 to 2023

After 2016, there was a slow recovery in GWS volumes, with some slight stabilisation in 2017. However, the levels never recovered to those of prior years, like 2010. Relatively lower than expected groundwater recharge persisted in the post-2016 period, which could be explained by ongoing climate fluctuations, changes in land use and ongoing regional droughts made worse by El Niño (Amare et al., 2024). These circumstances highlight the intricate connection between terrestrial hydrological processes and atmospheric phenomena like El Niño, where long-term groundwater storage patterns are influenced by surface water conditions and rainfall variability (Sanchez-Matos et al., 2024; Montaud et al., 2022).

CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS

This chapter presents the study's findings and suggestions in accordance with the objectives highlighted in Chapter 1.

6.1 Conclusions

The following were the conclusions of the study.

(i) According to the correlation analysis, there were regional variations in the relationship between GLDAS-GRACE GWS and the input hydrometeorological variables. The primary drivers of this variation are differences in the hydrogeology. Rainfall and groundwater recharge, for example were more strongly correlated in areas with permeable substrates and high soil moisture retention. On the other hand, correlations were weaker in regions with confined aquifers or impervious surfaces. This emphasises how crucial it is to take local hydrogeological conditions into account when simulating groundwater storage. In these areas, additional input parameters unique to the local geology may be necessary for accurate GWS predictions.

(ii) RF consistently performed better than XGBoost in every sub-catchment when randomised search CV was used for hyperparameter tuning during the downscaling. Furthermore, RF was able to control complex relationships between variables and prevent overfitting. Because of its improved ability to capture even the smallest spatial variations in GWS, it is a good choice for downscaling tasks that involve a range of environmental factors. XGBoost's feature importance, on the other hand, focused primarily with soil moisture and rainfall, which might have caused it to ignore other important factors. But because RF showed a more uniform focus on all input variables, it was able to take into account of a wider variety of factors that affect GWS, producing more thorough and precise downscaling results.

(iii) The UZC was divided into smaller sub-catchments, which greatly improved GWS downscaling accuracy. Smaller sub-catchments performed better than larger ones during the downscaling process, making this improvement especially noticeable. Because the smaller areas decreased the heterogeneity in geological and hydrological conditions, the machine learning models could be calibrated more precisely. Larger sub-catchments, like

Chobe, displayed some limitations, perhaps as a result of greater spatial variability that was more difficult for the models to account for.

(iv) The downscaling of GRACE/GRACE-FO using machine learning estimates offered reliable groundwater data at a 5 kilometres resolution. For areas with little groundwater monitoring, where localised water resource management relies on higher-resolution data, this improved accuracy is essential. The downscaling offered important spatial insights while preserving GRACE/GRACE-FO temporal patterns. GWS predictions were less accurate in areas with confined aquifers due to slow recharge rates and restricted hydraulic connectivity. This underlines the shortcomings of the current model in such hydrogeological regions and raises the possibility that more geophysical data or alternative models may be required to enhance predictions in these areas.

(v) One important finding was that groundwater recharge drastically declines below a rainfall threshold of 614.39 millimetres during the wet season. This threshold is essential for water managers to predict periods of low recharge and alleviate water scarcity. Furthermore, groundwater droughts were well-represented in the downscaled GRACE/GRACE-FO data, especially in El Niño years (2015, 2016, 2018, 2019 and 2023). Reduced precipitation and decreased groundwater storage during these years showed how useful GRACE/GRACE-FO data is for drought monitoring. Lastly, the study found that groundwater storage in the UZC decreased overall between 2009 and 2023, with anomalies ranging from -433 millimetres to +264 millimetres. This indicates significant spatial variability and underscores the difficulties in maintaining water resources in the face of rising demand and climate variability.

6.2 Recommendations

The study's findings lead to the following suggestions being given to the relevant parties.

(i) The downscaled GRACE/GRACE-FO data should be supplemented by larger in-situ groundwater monitoring networks, which Water Resources Management Authority (WARMA) and the DWRD under the Ministry of Water Development and Sanitation should prioritise funding. Particularly in areas with little monitoring, as this will enhance validation efforts and provide more precise, localised insights into groundwater dynamics.

(ii) WARMA and the Zambia Meteorological Department (ZMD) should keep a careful eye on seasonal rainfall patterns because 614.39 millimetres of rainfall is the threshold for sufficient groundwater recharge in the UZC. To reduce the likelihood of water shortages during periods of predicted low recharge, early warning systems and proactive water conservation measures should be put in place.

(iii) In regions where aquifers are confined, the downscaling model performed poorly. In these areas, where traditional models struggle to capture complex groundwater dynamics and slow recharge rates, it is advised that the Southern African Science Service Centre for Climate Change and Adaptive Land Management (SASSCAL) and the University of Zambia (UNZA) develop alternative geophysical data or more specialised models be created to enhance predictions.

(iv) Long-term plans for sustainable water use are required, as evidenced by the overall decrease in groundwater storage that was seen between 2009 and 2023. In areas where groundwater supplies are severely depleting, WARMA should promote conservation measures, curtail over-extraction and put frameworks into place that safeguard these resources.

(v) National Remote Sensing Centre (NRSC), UNZA and WARMA should develop regional downscaling techniques that can be used in other hydrogeological contexts since the model's performance was enhanced by dividing the study area into sub-catchments. Better planning for water resources in a variety of geographical areas and more accurate estimates of groundwater storage will be made possible by this.

(vi) The downscaling of groundwater storage estimates could be greatly enhanced by recurrent neural networks (RNNs), which are better at identifying patterns and temporal dependencies. The NRSC can contribute expertise in satellite data, SASSCAL can provide funding and technical support and UNZA can conduct research on deep learning techniques to advance the development and use of RNNs. Long-term predictions of groundwater dynamics may become more precise with this method, particularly in areas with intricate hydrological processes.

REFERENCES

Adib, A., Moradi, A., Lotfirad, M., Azizipour, M., Liaghat, A., 2023. Investigating the relationship between meteorological, hydrological and groundwater resource droughts under the influence of upstream dam reservoir effects. Model. Earth Syst. Environ. 9, 3609–3619. https://doi.org/10.1007/s40808-023-01710-9

Ahmed, A.N., Yafouz, A., Birima, A.H., Kisi, O., Huang, Y.F., Sherif, M., Sefelnasr, A., El-Shafie, A., 2022. Water level prediction using various machine learning algorithms: a case study of Durian Tunggal river, Malaysia. Engineering Applications of Computational Fluid Mechanics 16, 422–440. https://doi.org/10.1080/19942060.2021.2019128

Ali, S., Khorrami, B., Jehanzaib, M., Tariq, A., Ajmal, M., Arshad, A., Shafeeque, M., Dilawar, A., Basit, I., Zhang, L., 2023. Spatial downscaling of GRACE data based on XGBoost model for improved understanding of hydrological droughts in the Indus Basin Irrigation System (IBIS). Remote Sensing 15, 873.

Amare, Z.Y., Geremew, B.B., Kebede, N.M., Amera, S.G., 2024. Impacts of El Niño-Southern oscillation on rainfall amount and anticipated humanitarian impact. Environ Dev Sustain. https://doi.org/10.1007/s10668-024-04476-x

Amdar, N., Seyoum, S., Al-Bakri, J., Rutten, M., Jewitt, G., Mul, M., 2024. Developing a water budget for the Amman-Zarqa basin using water accounting plus and the pixel-based soil water balance model. Model. Earth Syst. Environ. https://doi.org/10.1007/s40808-024-02159-0

Arast, M., Ranjbar, A., Mousavi, S.H., Abdollahi, K., 2020. Assessment of the Relationship Between NDVI-Based Actual Evapotranspiration by SEBS. Iran J Sci Technol Trans Sci 44, 1051–1062. https://doi.org/10.1007/s40995-020-00895-3

Arnzen, A., Wagman, J., Chishya, C., Orange, E., Eisele, T.P., Yukich, J., Ashton, R.A., Chanda, J., Sakala, J., Chanda, B., Muyabe, R., Kaniki, T., Mwenya, M., Mwaanga, G., Eaton, W.T., Mancuso, B., Mungo, A., Mburu, M.M., Bubala, N., Hagwamuna, A., Simulundu, E., Saili, K., Miller, J.M., Silumbe, K., Hamainza, B., Ngulube, W., Moonga, H., Chirwa, J., Burkot, T.R., Slutsker, L., Littrell, M., 2024. Characteristics of the Western Province, Zambia, trial site for evaluation of attractive targeted sugar baits for malaria vector control. Malar J 23, 153. https://doi.org/10.1186/s12936-024-04985-0

Balon, E.K., Coche, A.G., 1974. Geology and Soils, in: Balon, E.K., Coche, A.G. (Eds.), Lake Kariba: A Man-Made Tropical Ecosystem in Central Africa. Springer Netherlands, Dordrecht, pp. 18–24. https://doi.org/10.1007/978-94-010-2334-4_2

Banda, A.M., Banda, K., Sakala, E., Chomba, M., Nyambe, I.A., 2023. Assessment of land use change in the wetland of Barotse Floodplain, Zambezi River Sub-Basin, Zambia. Nat Hazards 115, 1193–1211. https://doi.org/10.1007/s11069-022-05589-0

Banda, K., Mulema, M., Chomba, I., Chomba, M., Levy, J., Nyambe, I., 2023a. Investigating groundwater and surface water interactions using remote sensing, hydrochemistry, and stable isotopes in the Barotse Floodplain, Zambia. Geology, Ecology, and Landscapes 1–16.

Banda, K., Ngwenya, V., Mulema, M., Chomba, I., Chomba, M., Nyambe, I., 2023b. Influence of water quality on benthic macroinvertebrates in a groundwater-dependent wetland. Frontiers in Water 5, 1177724.

Banda, K.E., Mwandira, W., Jakobsen, R., Ogola, J., Nyambe, I., Larsen, F., 2019. Mechanism of salinity change and hydrogeochemical evolution of groundwater in the Machile-Zambezi

Basin, South-western Zambia. Journal of African Earth Sciences 153, 72–82. https://doi.org/10.1016/j.jafrearsci.2019.02.022

Bäumle, R., Himmelsbach, T., Noell, U., 2018. Hydrogeology and geochemistry of a tectonically controlled, deep-seated and semi-fossil aquifer in the Zambezi Region (Namibia). Hydrogeology Journal 3, 885–914. https://doi.org/10.1007/s10040-018-1896-x

Beilfuss, R., 2012. A risky climate for southern African hydro: assessing hydrological risks and consequences for Zambezi River Basin dams.

Bennett, G., 2024. Analysis of methods used to validate remote sensing and GIS-based groundwater potential maps in the last two decades: A review. Geosystems and Geoenvironment 3, 100245. https://doi.org/10.1016/j.geogeo.2023.100245

Bhanja, S.N., Mukherjee, A., Saha, D., Velicogna, I., Famiglietti, J.S., 2016. Validation of GRACE based groundwater storage anomaly using in-situ groundwater level measurements in India. Journal of Hydrology 543, 729–738. https://doi.org/10.1016/j.jhydrol.2016.10.042

Blank, D., Eicker, A., Jensen, L., Güntner, A., 2023. A global analysis of water storage variations from remotely sensed soil moisture and daily satellite gravimetry. Hydrology and Earth System Sciences 27, 2413–2435. https://doi.org/10.5194/hess-27-2413-2023

Braune, E., Xu, Y., 2010. The role of ground water in Sub-Saharan Africa. Groundwater 48, 229–238.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brenya, R., Jiang, Y., Sampene, A.K., Zhu, J., 2024. Food security in sub-Sahara Africa: Exploring the nexus between nutrition, innovation, circular economy, and climate change. Journal of Cleaner Production 438, 140805. https://doi.org/10.1016/j.jclepro.2024.140805

Breton, G., Barham, L., Mudenda, G., Soodyall, H., Schlebusch, C.M., Jakobsson, M., 2024. BaTwa populations from Zambia retain ancestry of past hunter-gatherer groups. Nat Commun 15, 7307. https://doi.org/10.1038/s41467-024-50733-y

Brown, C., Greene, A.M., Block, P.J., Giannini, A., 2008. Review of downscaling methodologies for Africa climate applications.

Brunner, P., Hendricks Franssen, H.-J., Kgotlhang, L., Bauer-Gottwein, P., Kinzelbach, W., 2007. How can remote sensing contribute in groundwater modeling? Hydrogeol J 15, 5–18. https://doi.org/10.1007/s10040-006-0127-z

Brutsaert, W., 2023. Hydrology. Cambridge university press.

Buytaert, W., Zulkafli, Z., Grainger, S., Acosta, L., Alemie, T.C., Bastiaensen, J., De Bièvre, B., Bhusal, J., Clark, J., Dewulf, A., Foggin, M., Hannah, D.M., Hergarten, C., Isaeva, A., Karpouzoglou, T., Pandeya, B., Paudel, D., Sharma, K., Steenhuis, T., Tilahun, S., Van Hecken, G., Zhumanova, M., 2014. Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management, and sustainable development. Front. Earth Sci. 2. https://doi.org/10.3389/feart.2014.00026

Carneiro, F.M., Filho, A.L. de B., Ferreira, F.M., Junior, G. de F.S., Brandão, Z.N., da Silva, R.P., Shiratsuchi, L.S., 2023. Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. Smart Agricultural Technology 5, 100292. https://doi.org/10.1016/j.atech.2023.100292

Chaminé, H.I., Afonso, M.J., Barbieri, M., 2022. Advances in Urban Groundwater and Sustainable Water Resources Management and Planning: Insights for Improved Designs with Nature, Hazards, and Society. Water 14, 3347. https://doi.org/10.3390/w14203347

Chen, H., Yang, X., Bian, D., Xiang, W., 2024. A novel composite cloud model-based three-stage evaluation for the spatial equilibrium of water resources in China. Journal of Cleaner Production 456, 142356. https://doi.org/10.1016/j.jclepro.2024.142356

Chen, L., He, Q., Liu, K., Li, J., Jing, C., 2019. Downscaling of GRACE-derived groundwater storage based on the random forest model. Remote Sensing 11, 2979.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Z., Zheng, W., Yin, W., Li, X., Zhang, G., Zhang, J., 2021a. Improving the spatial resolution of GRACE-derived terrestrial water storage changes in small areas using the machine learning spatial downscaling method. Remote Sensing 13, 4760.

Chen, Z., Zheng, W., Yin, W., Li, X., Zhang, G., Zhang, J., 2021b. Improving the spatial resolution of GRACE-derived terrestrial water storage changes in small areas using the machine learning spatial downscaling method. Remote Sensing 13, 4760.

Chirashi, M.E., 2014. Zambezi River-sea route for the SADC Region: A study on how to extend navigation on the Zambezi River. Zimbabwe Journal of Science and Technology 9, 1–20.

Chomba, I.C., Banda, K.E., Winsemius, H.C., Eunice, M., Sichingabula, H.M., Nyambe, I.A., 2022. Integrated hydrologic-hydrodynamic inundation modeling in a groundwater dependent tropical floodplain. Journal of Human, Earth, and Future 3, 237–246.

Coelho, V.H.R., Montenegro, S., Almeida, C.N., Silva, B.B., Oliveira, L.M., Gusmão, A.C.V., Freitas, E.S., Montenegro, A.A.A., 2017. Alluvial groundwater recharge estimation in semi-arid environment using remotely sensed data. Journal of Hydrology 548, 1–15. https://doi.org/10.1016/j.jhydrol.2017.02.054

Cuthbert, M.O., Taylor, R.G., Favreau, G., Todd, M.C., Shamsudduha, M., Villholth, K.G., MacDonald, A.M., Scanlon, B.R., Kotchoni, D.O.V., Vouillamoz, J.-M., Lawson, F.M.A., Adjomayi, P.A., Kashaigili, J., Seddon, D., Sorensen, J.P.R., Ebrahim, G.Y., Owor, M., Nyenje, P.M., Nazoumou, Y., Goni, I., Ousmane, B.I., Sibanda, T., Ascott, M.J., Macdonald, D.M.J., Agyekum, W., Koussoubé, Y., Wanke, H., Kim, H., Wada, Y., Lo, M.-H., Oki, T., Kukuric, N., 2019. Observed controls on resilience of groundwater to climate variability in sub-Saharan Africa. Nature 572, 230–234. https://doi.org/10.1038/s41586-019-1441-7

Dan, L., Shi, C., Zhang, J., Wei, L., Yang, L., 2021. Application of Xgboost Algorithm Based on Machine Learning in Reservoir Prediction of Offshore Oilfield, in: Lin, J. (Ed.), Proceedings of the International Field Exploration and Development Conference 2020. Springer, Singapore, pp. 3157–3167. https://doi.org/10.1007/978-981-16-0761-5_296

Dembélé, M., Zwart, S.J., 2016. Evaluation and comparison of satellite-based rainfall products in Burkina Faso, West Africa. International Journal of Remote Sensing 37, 3995–4014. https://doi.org/10.1080/01431161.2016.1207258

Didan, K., 2021. MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V061. NASA EOSDIS Land Processes DAAC.

Diro, G.T., Tompkins, A.M., Bi, X., 2012. Dynamical downscaling of ECMWF Ensemble seasonal forecasts over East Africa with RegCM3. Journal of Geophysical Research: Atmospheres 117.

Dube, K., Nhamo, G., 2023. Evaluating climate Change’s impact on hydroelectricity in the Zambezi river basin. Heliyon 9, e23235. https://doi.org/10.1016/j.heliyon.2023.e23235

Dube, Thandekile, Dube, Timothy, Marambanyika, T., 2023. A review of wetland vulnerability assessment and monitoring in semi-arid environments of sub-Saharan Africa. Physics and Chemistry of the Earth, Parts A/B/C 132, 103473. https://doi.org/10.1016/j.pce.2023.103473

Elorza, J., Recio, C., 2023. Contemporary beachrock formation at Arrigunaga beach (Bizkaia, Spain) on anthropogenic slag. J Iber Geol 49, 189–213. https://doi.org/10.1007/s41513-023-00211-z

Engelbrecht, F., Adegoke, J., Bopape, M.-J., Naidoo, M., Garland, R., Thatcher, M., McGregor, J., Katzfey, J., Werner, M., Ichoku, C., Gatebe, C., 2015. Projections of rapidly rising surface temperatures over Africa under low mitigation. Environ. Res. Lett. 10, 085004. https://doi.org/10.1088/1748-9326/10/8/085004

Fallatah, O.A., Ahmed, M., Cardace, D., Boving, T., Akanda, A.S., 2019. Assessment of modern recharge to arid region aquifers using an integrated geophysical, geochemical, and remote sensing approach. Journal of Hydrology 569, 600–611. https://doi.org/10.1016/j.jhydrol.2018.09.061

Famiglietti, J.S., Lo, M., Ho, S.L., Bethune, J., Anderson, K.J., Syed, T.H., Swenson, S.C., de Linage, C.R., Rodell, M., 2011. Satellites measure recent rates of groundwater depletion in California's Central Valley. Geophysical Research Letters 38. https://doi.org/10.1029/2010GL046442

Famiglietti, J.S., Rodell, M., 2013. Water in the Balance. Science. https://doi.org/10.1126/science.1236460

Fan, Y., Li, H., Miguez-Macho, G., 2013. Global Patterns of Groundwater Table Depth. Science 339, 940–943. https://doi.org/10.1126/science.1229881

FAO, 2020. WaPOR V2 Database Methodology. Remote Sensing for Water Productivity Technical Report: Methodology Series. FAO, Rome.

FAO, F., 2018. Food and agriculture organization of the United Nations. Rome, URL: http://faostat. fao. org 403–403.

Faraji, Z., Kaviani, A., Khosravi, L., 2024. Downscaling the MODIS land surface temperature using a trapezial concept applied to the MODIS and sentinel 2 images. Environ Monit Assess 196, 665. https://doi.org/10.1007/s10661-024-12800-1

Fatolazadeh, F., Eshagh, M., Goïta, K., 2022. New spectro-spatial downscaling approach for terrestrial and groundwater storage variations estimated by GRACE models. Journal of Hydrology 615, 128635.

Feng, W., Shum, C.K., Zhong, M., Pan, Y., 2018. Groundwater Storage Changes in China from Satellite Gravity: An Overview. Remote Sensing 10, 674. https://doi.org/10.3390/rs10050674

Ferreira, V., Yong, B., Montecino, H., Ndehedehe, C.E., Seitz, K., Kutterer, H., Yang, K., 2023. Estimating GRACE terrestrial water storage anomaly using an improved point mass solution. Sci Data 10, 234. https://doi.org/10.1038/s41597-023-02122-1

Field, C.B., Barros, V.R., 2014. Climate Change 2014 – Impacts, Adaptation and Vulnerability: Regional Aspects. Cambridge University Press.

Fortnam, M., Atkins, M., Brown, K., Chaigneau, T., Frouws, A., Gwaro, K., Huxham, M., Kairo, J., Kimeli, A., Kirui, B., Sheen, K., 2021. Multiple impact pathways of the 2015–2016 El Niño in coastal Kenya. Ambio 50, 174–189. https://doi.org/10.1007/s13280-020-01321-z

Fu, R., Chen, R., Wang, Changjing, Chen, X., Gu, H., Wang, Cong, Xu, B., Liu, G., Yin, G., 2022. Generating High-Resolution and Long-Term SPEI Dataset over Southwest China through Downscaling EEAD Product by Machine Learning. Remote Sensing 14, 1662.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific data 2, 1–21.

Gaffoor, Z., 2017. Using geostatistical-hydrogeological approach to develop groundwater monitoring system in South Western Karoo, South Africa.

Ghorbani, M.A., Khatibi, R., Hosseini, B., Bilgili, M., 2013. Relative importance of parameters affecting wind speed prediction using artificial neural networks. Theoretical and Applied Climatology 114, 107–114.

Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H.H., Manning, A.H., Van Beek, L.P.H., Jellinek, A.M., 2011. Mapping permeability over the surface of the Earth: MAPPING GLOBAL PERMEABILITY. Geophys. Res. Lett. 38, n/a-n/a. https://doi.org/10.1029/2010GL045565

Goddard, L., Gershunov, A., 2020. Impact of El Niño on Weather and Climate Extremes, in: El Niño Southern Oscillation in a Changing Climate. American Geophysical Union (AGU), pp. 361–375. https://doi.org/10.1002/9781119548164.ch16

Gong, Z., Ge, W., Guo, J., Liu, J., 2024. Satellite remote sensing of vegetation phenology: Progress, challenges, and opportunities. ISPRS Journal of Photogrammetry and Remote Sensing 217, 149–164. https://doi.org/10.1016/j.isprsjprs.2024.08.011

Hadavi, D., Mousavi, S.M., Rahimzadegan, M., 2024. An intercomparison of the groundwater level estimations by GRACE and GRACE-FO satellites and groundwater modeling in Iran. Acta Geophys. 72, 3609–3629. https://doi.org/10.1007/s11600-024-01308-4

Haji-Aghajany, S., Amerian, Y., Amiri-Simkooei, A., 2023. Impact of Climate Change Parameters on Groundwater Level: Implications for Two Subsidence Regions in Iran Using Geodetic Observations and Artificial Neural Networks (ANN). Remote Sensing 15, 1555. https://doi.org/10.3390/rs15061555

Hannah, D.M., Demuth, S., van Lanen, H.A.J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., Tallaksen, L.M., 2011. Large-scale river flow archives: importance, current status and future needs. Hydrological Processes 25, 1191–1200. https://doi.org/10.1002/hyp.7794

Hanssen-Bauer, I., Førland, E.J., Haugen, J.E., Tveito, O.E., 2003. Temperature and precipitation scenarios for Norway: comparison of results from dynamical and empirical downscaling. Climate Research 25, 15–27.

Healy, R.W., 2010. Estimating groundwater recharge. Cambridge university press.

Hilker, T., Hall, F.G., Tucker, C.J., Coops, N.C., Black, T.A., Nichol, C.J., Sellers, P.J., Barr, A., Hollinger, D.Y., Munger, J.W., 2012. Data assimilation of photosynthetic light-use efficiency using multi-angular satellite data: II Model implementation and validation. Remote Sensing of Environment 121, 287–300. https://doi.org/10.1016/j.rse.2012.02.008

Hughes, D.A., Read, L., Jeuland, M., Kapangaziwiri, E., Elkurdy, M., Lambl, D., Hale, E., Opperman, J.J., 2023. Insights from a comparison of two hydrological modelling approaches in the Kwando (Cuando) River and the western tributaries of the Zambezi River basin. Journal of Hydrology: Regional Studies 48, 101482. https://doi.org/10.1016/j.ejrh.2023.101482

Hulme, M., Arntzen, J., 1996. Climate change and Southern Africa: an exploration of some potential impacts and implications in the SADC region. Climatic Research Unit, University of East Anglia Norwich, UK.

Humphrey, V., Rodell, M., Eicker, A., 2023. Using Satellite-Based Terrestrial Water Storage Data: A Review. Surv Geophys. https://doi.org/10.1007/s10712-022-09754-9

Hyndman, R.J., 2018. Forecasting: principles and practice. OTexts.

Impollonia, G., Croci, M., Amaducci, S., 2024. Upscaling and downscaling approaches for early season rice yield prediction using Sentinel-2 and machine learning for precision nitrogen fertilisation. Computers and Electronics in Agriculture 227, 109603. https://doi.org/10.1016/j.compag.2024.109603

Jagert, F., Immenhauser, A., Wohnlich, S., 2023. Hydrogeochemical modelling of groundwater in a fractured Carboniferous sandstone aquifer. Grundwasser - Zeitschrift der Fachsektion Hydrogeologie 28, 225–242. https://doi.org/10.1007/s00767-023-00549-2

Jasechko, S., Taylor, R.G., 2015. Intensive rainfall recharges tropical groundwaters. Environ. Res. Lett. 10, 124015. https://doi.org/10.1088/1748-9326/10/12/124015

Jayasankar, C.B., Rajendran, K., Sajani, S., 2021. Does increasing the spatial resolution in dynamical downscaling impact climate change projection of Indian summer monsoon, population and GDP? Theoretical and Applied Climatology 145, 441–453.

Jeihouni, E., Mohammadi, M., Eslamian, S., Zareian, M.J., 2019. Potential impacts of climate change on groundwater level through hybrid soft-computing methods: a case study—Shabestar Plain, Iran. Environ Monit Assess 191, 620. https://doi.org/10.1007/s10661-019-7784-6

Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A., 2010. Use of NDVI and Land Surface Temperature for Drought Assessment: Merits and Limitations. https://doi.org/10.1175/2009JCLI2900.1

Khan, F., Liou, Y.-A., Spöck, G., Wang, X., Ali, S., 2024. Assessing the impacts of temperature extremes on agriculture yield and projecting future extremes using machine learning and deep learning approaches with CMIP6 data. International Journal of Applied Earth Observation and Geoinformation 132, 104071. https://doi.org/10.1016/j.jag.2024.104071

Khorrami, B., Ali, S., Gündüz, O., 2023. Investigating the Local-scale Fluctuations of Groundwater Storage by Using Downscaled GRACE/GRACE-FO JPL Mascon Product Based on Machine Learning (ML) Algorithm. Water Resour Manage 37, 3439–3456. https://doi.org/10.1007/s11269-023-03509-w

Kiesel, C., Dannenberg, P., Hulke, C., Kairu, J., Revilla Diez, J., Sandhage-Hofmann, A., 2022. An argument for place-based policies: The importance of local agro-economic, political and environmental conditions for agricultural policies exemplified by the Zambezi region, Namibia. Environmental Science & Policy 129, 137–149. https://doi.org/10.1016/j.envsci.2021.12.012

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A.K., Hochreiter, S., Nearing, G.S., 2019. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. Water Resources Research 55, 11344–11354. https://doi.org/10.1029/2019WR026065

Kuhn, M., Johnson, K., 2019. Feature engineering and selection: A practical approach for predictive models. Chapman and Hall/CRC.

Kumar, B., Atey, K., Singh, B.B., Chattopadhyay, R., Acharya, N., Singh, M., Nanjundiah, R.S., Rao, S.A., 2023. On the modern deep learning approaches for precipitation downscaling. Earth Sci Inform 16, 1459–1472. https://doi.org/10.1007/s12145-023-00970-4

Kumar, M., Soni, P., Swargiary, D., 2024. Monitoring Terrestrial Water Storage Using GRACE/GRACE-FO Data over India: A Review. Iran J Sci Technol Trans Civ Eng. https://doi.org/10.1007/s40996-024-01623-2

Kusche, J., Schmidt, R., Petrovic, S., Rietbroek, R., 2009. Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. Journal of geodesy 83, 903–913.

Lakshmivarahan, S., Stensrud, D., 2009. Ensemble Kalman filter. Control Systems, IEEE 29, 34–46. https://doi.org/10.1109/MCS.2009.932225

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resources Research 48, 2011WR011453. https://doi.org/10.1029/2011WR011453

Lautze, J., Phiri, Z., Smakhtin, V., Saruchera, D., 2017. The Zambezi river basin. Routledge.

Lerner, D.N., 2020. Groundwater recharge, in: Geochemical Processes, Weathering and Groundwater Recharge in Catchments. CRC Press, pp. 109–150.

Leroux, M.-D., Bonnardot, F., Somot, S., Alias, A., Kotomangazafy, S., Ridhoine, A.-O.S., Veerabadren, P., Amélie, V., 2024. Developing climate services for vulnerable islands in the Southwest Indian Ocean: A combined statistical and dynamical CMIP6 downscaling approach for climate change assessment. Climate Services 34, 100491. https://doi.org/10.1016/j.cliser.2024.100491

Li, B., Rodell, M., 2015. Evaluation of a model-based groundwater drought indicator in the conterminous US. Journal of Hydrology 526, 78–88.

Li, S., Xu, C., Su, M., Lu, W., Chen, Q., Huang, Q., Teng, Y., 2024. Downscaling of environmental indicators: A review. Science of The Total Environment 916, 170251. https://doi.org/10.1016/j.scitotenv.2024.170251

Lucas, M., Oliveira, P.T., Melo, D., Wendland, E., 2015. Evaluation of remotely sensed data for estimating recharge to an outcrop zone of the Guarani Aquifer System (South America). Hydrogeology Journal 23, 1–9. https://doi.org/10.1007/s10040-015-1246-1

Madani, A., Hagage, M., Elbeih, S.F., 2022. Random Forest and Logistic Regression algorithms for prediction of groundwater contamination using ammonia concentration. Arab J Geosci 15, 1619. https://doi.org/10.1007/s12517-022-10872-2

Malik, I.H., Ford, J.D., 2024. Monitoring climate change vulnerability in the Himalayas. Ambio. https://doi.org/10.1007/s13280-024-02066-9

Malik, M.S., Shukla, J.P., Mishra, S., 2021. Effect of Groundwater Level on Soil Moisture, Soil Temperature and Surface Temperature. J Indian Soc Remote Sens 49, 2143–2161. https://doi.org/10.1007/s12524-021-01379-6

Martínez-de la Torre, A., Miguez-Macho, G., 2019. Groundwater influence on soil moisture memory and land–atmosphere fluxes in the Iberian Peninsula. Hydrology and Earth System Sciences 23, 4909–4932. https://doi.org/10.5194/hess-23-4909-2019

McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E.C., Franz, T.E., Shi, J., Gao, H., Wood, E.F., 2017. The future of Earth observation in hydrology. Hydrology and Earth System Sciences 21, 3879–3914. https://doi.org/10.5194/hess-21-3879-2017

McNally, A., Jacob, J., Arsenault, K., Slinski, K., Sarmiento, D.P., Hoell, A., Pervez, S., Rowland, J., Budde, M., Kumar, S., 2022. A Central Asia hydrologic monitoring dataset for food and water security applications in Afghanistan. Earth System Science Data 14, 3115–3135.

Mendelsohn, J., 2022. The Angolan Catchments of Northern Botswana's Major Rivers: The Cubango, Cuito, Cuando and Zambezi Rivers, in: Eckardt, F.D. (Ed.), Landscapes and Landforms of Botswana. Springer International Publishing, Cham, pp. 15–36. https://doi.org/10.1007/978-3-030-86102-5_2

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, 2756.

Miro, M., Famiglietti, J., 2018. Downscaling GRACE Remote Sensing Datasets to High-Resolution Groundwater Storage Change Maps of California's Central Valley. Remote Sensing 10, 143. https://doi.org/10.3390/rs10010143

Montaud, J.-M., Dávalos, J., Pécastaing, N., 2022. Socioeconomic Risks of Extreme El Niño Event-Related Road Damages in Peru. Environ Model Assess 27, 831–851. https://doi.org/10.1007/s10666-022-09830-9

Mosavi, A., Ozturk, P., Chau, K., 2018. Flood Prediction Using Machine Learning Models: Literature Review. Water 10, 1536. https://doi.org/10.3390/w10111536

Mosimane, K., Struyf, E., Gondwe, M.J., Frings, P., Van, P.D., Schaller, J., Wolski, P., Schoelynck, J., Conley, D.J., Murray, -Hudson Mike, 2017. Variability in chemistry of surface and soil waters of an evapotranspiration-dominated flood-pulsed wetland : solute processing in the Okavango Delta, Botswana. Water SA 43, 104–115. https://doi.org/10.4314/wsa.v43i1.13

Mupangwa, W., Chipindu, L., Ncube, B., Mkuhlani, S., Nhantumbo, N., Masvaya, E., Ngwira, A., Moeletsi, M., Nyagumbo, I., Liben, F., 2023a. Temporal Changes in Minimum and Maximum Temperatures at Selected Locations of Southern Africa. Climate 11, 84. https://doi.org/10.3390/cli11040084

Mupangwa, W., Chipindu, L., Ncube, B., Mkuhlani, S., Nhantumbo, N., Masvaya, E., Ngwira, A., Moeletsi, M., Nyagumbo, I., Liben, F., 2023b. Temporal Changes in Minimum and Maximum Temperatures at Selected Locations of Southern Africa. Climate 11, 84. https://doi.org/10.3390/cli11040084

Nafouanti, M.B., Li, J., Nyakilla, E.E., Mwakipunda, G.C., Mulashani, A., 2023. A novel hybrid random forest linear model approach for forecasting groundwater fluoride contamination. Environ Sci Pollut Res 30, 50661–50674. https://doi.org/10.1007/s11356-023-25886-w

Nazari, A., Zaryab, A., Ahmadi, A., 2023. Estimation of groundwater storage change in the Helmand River Basin (Afghanistan) using GRACE satellite data. Earth Sci Inform 16, 579–589. https://doi.org/10.1007/s12145-022-00899-0

Ndehedehe, C., 2022. Satellite Observations of Terrestrial Water Storage, in: Ndehedehe, C. (Ed.), Satellite Remote Sensing of Terrestrial Hydrology. Springer International Publishing, Cham, pp. 331–386. https://doi.org/10.1007/978-3-030-99577-5_14

Nenweli, R., Watson, A., Brookfield, A., Münch, Z., Chow, R., 2024. Is groundwater running out in the Western Cape, South Africa? Evaluating GRACE data to assess groundwater storage during droughts. Journal of Hydrology: Regional Studies 52, 101699.

Nilsson, B., Nielsen, K., 2024. Validation of Sentinel-6MF based lake levels – An assessment with in situ data and other satellite altimetry data. Advances in Space Research 73, 5806–5821. https://doi.org/10.1016/j.asr.2024.04.006

Noori, R., Maghrebi, M., Jessen, S., Bateni, S.M., Heggy, E., Javadi, S., Noury, M., Pistre, S., Abolfathi, S., AghaKouchak, A., 2023. Decline in Iran's groundwater recharge. Nature Communications 14, 6674.

Ouyang, Z., Ravier, P., Jabloun, M., 2021. STL Decomposition of Time Series Can Benefit Forecasting Done by Statistical Methods but Not by Machine Learning Ones. Engineering Proceedings 5, 42. https://doi.org/10.3390/engproc2021005042

Pablos, M., Martínez-Fernández, J., Piles, M., Sánchez, N., Vall-llossera, M., Camps, A., 2016. Multi-Temporal Evaluation of Soil Moisture and Land Surface Temperature Dynamics Using in Situ and Satellite Observations. Remote Sensing 8, 587. https://doi.org/10.3390/rs8070587

Persits, F.M., Ahlbrandt, T.S., Tuttle, M.L., Charpentier, R.R., Brownfield, M.E., Takahashi, K.I., 1997. Maps showing geology, oil and gas fields and geological provinces of Africa (Report No. 97- 470A), Open-File Report. Reston, VA. https://doi.org/10.3133/ofr97470A

Petrovic, S., Sherland, M., 2023. Zambia, in: Petrovic, S. (Ed.), World Energy Handbook. Springer International Publishing, Cham, pp. 123–136. https://doi.org/10.1007/978-3-031-31625-8_15

Pham, Q.B., Kumar, M., Di Nunno, F., Elbeltagi, A., Granata, F., Islam, A.R.Md.T., Talukdar, S., Nguyen, X.C., Ahmed, A.N., Anh, D.T., 2022. Groundwater level prediction using machine learning algorithms in a drought-prone area. Neural Comput & Applic 34, 10751–10773. https://doi.org/10.1007/s00521-022-07009-7

Puth, M.-T., Neuhäuser, M., Ruxton, G.D., 2015. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. Animal Behaviour 102, 77–84. https://doi.org/10.1016/j.anbehav.2015.01.010

Ramsar, C., 2004. The list of wetlands of international importance. RAMSAR Secretariat: Gland, Switzerland.

Rashid, K.J., Tuli, R.D., Nasher, N.M.R., Akter, T., Karim, K.H.R., Hasan, Md.M., Talha, M., Chowdhury, S.I.A., Musharrat, M., 2024. Greening and browning trend with physio-climatic drivers in Chattogram division, Bangladesh. Environ Dev Sustain. https://doi.org/10.1007/s10668-024-05505-5

Richey, A.S., Thomas, B.F., Lo, M.-H., Reager, J.T., Famiglietti, J.S., Voss, K., Swenson, S., Rodell, M., 2015. Quantifying renewable groundwater stress with GRACE. Water Resources Research 51, 5217–5238. https://doi.org/10.1002/2015WR017349

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., 2004. The Global Land Data Assimilation System in: Bulletin of the American Meteorological Society Volume 85 Issue 3 (2004) [WWW Document]. URL https://journals.ametsoc.org/view/journals/bams/85/3/bams-85-3-381.xml (accessed 6.7.24).

Rodell, M., Velicogna, I., Famiglietti, J.S., 2009. Satellite-based estimates of groundwater depletion in India. Nature 460, 999–1002. https://doi.org/10.1038/nature08238

Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 569–575. https://doi.org/10.1109/TPAMI.2009.187

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Roy, S., Rao, C.M., Abioui, M., 2024. Evaluation of non-stationary spatial relationship between meteorological-environmental parameters and PM2.5. Advances in Space Research 73, 4106–4124. https://doi.org/10.1016/j.asr.2024.01.009

Rubio-Arellano, A.B., Morán-Ramírez, J., Ramos-Leal, J.A., 2024. Identification of natural water sources and anthropogenic contamination applying the VISHMOD methodology in a volcano-sedimentary aquifer. Model. Earth Syst. Environ. 10, 4845–4861. https://doi.org/10.1007/s40808-024-02032-0

Rzepecka, Z., Birylo, M., 2020. Groundwater Storage Changes Derived from GRACE and GLDAS on Smaller River Basins—A Case Study in Poland. Geosciences 10, 124. https://doi.org/10.3390/geosciences10040124

Sahour, H., 2020. Statistical Downscaling Techniques to Enhance the Spatial Resolution of the Grace Satellite Data and to Fill Temporal Gaps. Western Michigan University.

Samuel, S., Dosio, A., Mphale, K., Faka, D.N., Wiston, M., 2023. Comparison of multimodel ensembles of global and regional climate models projections for extreme precipitation over four major river basins in southern Africa— assessment of the historical simulations. Climatic Change 176, 57. https://doi.org/10.1007/s10584-023-03530-3

Samuel, S., Mengistu Tsidu, G., Dosio, A., Mphale, K., 2024. Exploring the influence of improved horizontal resolution on extreme precipitation in Southern Africa major river basins: insights from CMIP6 HighResMIP simulations. Clim Dyn 62, 8099–8120. https://doi.org/10.1007/s00382-024-07325-9

Sanchez-Matos, J., Vázquez-Rowe, I., Kahhat, R., 2024. AWARE characterization factors in Peru encompassing El Niño and climate change events: does increased water availability guarantee less water scarcity? Int J Life Cycle Assess. https://doi.org/10.1007/s11367-024-02369-9

Santoso, A., Mcphaden, M.J., Cai, W., 2017. The Defining Characteristics of ENSO Extremes and the Strong 2015/2016 El Niño. Reviews of Geophysics 55, 1079–1129. https://doi.org/10.1002/2017RG000560

Satizabal-Alarc, D.A., Suhogusoff, A., Ferrari, L.C.K.M.F., 2023. Characterization of groundwater storage changes in the Amazon River Basin based on downscaling of GRACE/GRACE-FO data with machine learning models [WWW Document]. URL https://www.researchgate.net/publication/375964151_Characterization_of_groundwater_storage_changes_in_the_Amazon_River_Basin_based_on_downscaling_of_GRACEGRACE-FO_data_with_machine_learning_models (accessed 6.5.24).

Scanlon, B.R., Longuevergne, L., Long, D., 2012. Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. Water Resources Research 48. https://doi.org/10.1029/2011WR011312

Seka, A.M., Guo, H., Zhang, J., Han, J., Bayable, E., Ayele, G.T., Workneh, H.T., Bayouli, O.T., Muhirwa, F., Reda, K.W., 2024. Evaluating the future total water storage change and hydrological drought under climate change over lake basins, East Africa. Journal of Cleaner Production 447, 141552. https://doi.org/10.1016/j.jclepro.2024.141552

Serravalle Reis Rodrigues, V.H., de Melo Barros Junior, P.R., dos Santos Marinho, E.B., Lima de Jesus Silva, J.L., 2023. Wavelet gated multiformer for groundwater time series forecasting. Sci Rep 13, 12726. https://doi.org/10.1038/s41598-023-39688-0

Shilengwe, C., Banda, K., Nyambe, I., 2024. Machine learning downscaling of GRACE/GRACE-FO data to capture spatial-temporal drought effects on groundwater storage at a local scale under data-scarcity. Environ Syst Res 13, 38. https://doi.org/10.1186/s40068-024-00368-1

Shilengwe, C., Nyimbili, P.H., Msendo, R., Banda, F., Mukupa, W., Erden, T., 2023. Synthetic Aperture Radar and Optical Sensor Techniques Using Google Earth Engine For Flood Monitoring and Damage Assessment – A Case Study of Mumbwa District, Zambia. zictjournal 7, 7–15. https://doi.org/10.33260/zictjournal.v7i1.122

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., 2016. Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences 20, 2611–2628. https://doi.org/10.5194/hess-20-2611-2016

Siderius, C., Gannon, K.E., Ndiyoi, M., Opere, A., Batisani, N., Olago, D., Pardoe, J., Conway, D., 2018. Hydrological Response and Complex Impact Pathways of the 2015/2016 El Niño in Eastern and Southern Africa. Earth's Future 6, 2–22. https://doi.org/10.1002/2017EF000680

Singh, A., Raju, A., 2020. Application of Grace Satellite Data for Assessment of Groundwater Resources in Central Ganga Alluvial Plain, Northern India, in: Shukla, V., Kumar, N. (Eds.), Environmental Concerns and Sustainable Development: Volume 1: Air, Water and Energy Resources. Springer, Singapore, pp. 153–162. https://doi.org/10.1007/978-981-13-5889-0_7

Singh, L., Saravanan, S., 2020. Satellite-derived GRACE groundwater storage variation in complex aquifer system in India. Sustain. Water Resour. Manag. 6, 43. https://doi.org/10.1007/s40899-020-00399-3

Sokeng, V.-C.J., Oulare, S., Kouamé, K.F., Mertens, B., Kone, T., Catry, T., Pillot, B., Ouattara, P.E., Kone, D., Sow, M., 2024. Cloud computing and spatial hydrology for monitoring the Buyo and Kossou reservoirs in Côte d'Ivoire. Remote Sensing Applications: Society and Environment 36, 101353. https://doi.org/10.1016/j.rsase.2024.101353

Sokneth, L., Mohanasundaram, S., Shrestha, S., Babel, M.S., Virdis, S.G.P., 2022. Evaluating aquifer stress and resilience with GRACE information at different spatial scales in Cambodia. Hydrogeol J 30, 2359–2377. https://doi.org/10.1007/s10040-022-02570-w

Solihuddin, T., Prihantono, J., Daulat, A., Bramawanto, R., Risandi, J., Kepel, T.L., Heriati, A., Mustikasari, E., 2024. Assessment of a small island's groundwater resilience under the pressure of anthropogenic and natural stresses on Tunda Island, Indonesia. Environ Earth Sci 83, 380. https://doi.org/10.1007/s12665-024-11649-4

Tanachaichoksirikun, P., Seeboonruang, U., 2019. Effect of Climate Change on Groundwater Age of Thailand's Lower Chao Phraya Basin. IOP Conf. Ser.: Mater. Sci. Eng. 639, 012032. https://doi.org/10.1088/1757-899X/639/1/012032

Tang, W., Pei, Y., Zheng, H., Zhao, Y., Shu, L., Zhang, H., 2022. Twenty years of China's water pollution control: Experiences and challenges. Chemosphere 295, 133875. https://doi.org/10.1016/j.chemosphere.2022.133875

Tangdamrongsub, N., Steele-Dunne, S.C., Gunter, B.C., Ditmar, P.G., Weerts, A.H., 2015. Data assimilation of GRACE terrestrial water storage estimates into a regional hydrological model of the Rhine River basin. Hydrology and Earth System Sciences 19, 2079–2100. https://doi.org/10.5194/hess-19-2079-2015

Taniguchi, K., 2016. Future changes in precipitation and water resources for Kanto Region in Japan after application of pseudo global warming method and dynamical downscaling. Journal of Hydrology: Regional Studies 8, 287–303.

Tao, H., Al-Sulttani, A.H., Salih, S.Q., Mohammed, M.K.A., Khan, M.A., Beyaztas, B.H., Ali, M., Elsayed, S., Shahid, S., 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. https://doi.org/10.1016/j.atmosres.2023.106815

Tao, H., Awadh, S.M., Salih, S.Q., Shafik, S.S., Yaseen, Z.M., 2022a. Integration of extreme gradient boosting feature selection approach with machine learning models: application of weather relative humidity prediction. Neural Comput & Applic 34, 515–533. https://doi.org/10.1007/s00521-021-06362-3

Tao, H., Hameed, M.M., Marhoon, H.A., Zounemat-Kermani, M., Heddam, S., Kim, S., Sulaiman, S.O., Tan, M.L., Sa'adi, Z., Mehr, A.D., Allawi, M.F., Abba, S.I., Zain, J.M., Falah, M.W., Jamei, M., Bokde, N.D., Bayatvarkeshi, M., Al-Mukhtar, M., Bhagat, S.K., Tiyasha, T., Khedher, K.M., Al-Ansari, N., Shahid, S., Yaseen, Z.M., 2022b. Groundwater level prediction using machine learning models: A comprehensive review. Neurocomputing 489, 271–308. https://doi.org/10.1016/j.neucom.2022.03.014

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

Tariq, A., Ali, S., Basit, I., Jamil, A., Farmonov, N., Khorrami, B., Khan, M.M., Sadri, S., Baloch, M.Y.J., Islam, F., 2023. Terrestrial and groundwater storage characteristics and their quantification in the Chitral (Pakistan) and Kabul (Afghanistan) river basins using GRACE/GRACE-FO satellite data. Groundwater for Sustainable Development 23, 100990.

Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2013. Ground water and climate change. Nature Clim Change 3, 322–329. https://doi.org/10.1038/nclimate1744

Teng, T.-P., Chen, W.-J., 2024. Using Pearson correlation coefficient as a performance indicator in the compensation algorithm of asynchronous temperature-humidity sensor pair. Case Studies in Thermal Engineering 53, 103924. https://doi.org/10.1016/j.csite.2023.103924

Texier, J., Gonçalvès, J., Stieglitz, T., Vallet-Coulomb, C., Labille, J., Marc, V., Poulain, A., Dussouillez, P., 2024. Groundwater-surface water exchanges in an alluvial plain in southern France subjected to pumping: A coupled multitracer and modeling approach. Journal of Hydrology: Regional Studies 56, 101995. https://doi.org/10.1016/j.ejrh.2024.101995

Timberlake, J., Chidumayo, E., 2011. Miombo ecoregion vision report. Occasional Publications in Biodiversity 20.

Tirivarombo, S., Hughes, D.A., 2011. Regional droughts and food security relationships in the Zambezi River Basin. Physics and Chemistry of the Earth, Parts A/B/C, 11th WaterNet/WARFSA/GWP-SA Symposium: IWRM for National and Regional Integration through Science, Policy and Practice 36, 977–983. https://doi.org/10.1016/j.pce.2011.08.003

Trenberth, K.E., 2011. Changes in precipitation with climate change. Climate Research 47, 123–138. https://doi.org/10.3354/cr00953

Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. Water 11, 910. https://doi.org/10.3390/w11050910

Ullah, W., Ahmad, K., Ullah, S., Tahir, A.A., Javed, M.F., Nazir, A., Abbasi, A.M., Aziz, M., Mohamed, A., 2023. Analysis of the relationship among land surface temperature (LST), land use land cover (LULC), and normalized difference vegetation index (NDVI) with topographic elements in the lower Himalayan region. Heliyon 9, e13322. https://doi.org/10.1016/j.heliyon.2023.e13322

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

Visvalingam, G., Krishnaraj, S., Andiyappan, R.K., Kamalapathy, R., Datchanamourthy, S.V., Lagudu, S., 2024. Understanding the impact of climate-induced sea level rise on groundwater inundation in a low-lying coastal area: A numerical simulation in Southeast India. Regional Studies in Marine Science 70, 103401. https://doi.org/10.1016/j.rsma.2024.103401

Wahr, J., Swenson, S., Zlotnicki, V., Velicogna, I., 2004. Time-variable gravity from GRACE: First results. Geophysical Research Letters 31. https://doi.org/10.1029/2004GL019779

Wan, Z., Hook, S., Hulley, G., 2021. MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061. 2021, distributed by NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD11A1.061

Wang, H., Yin, S., Yue, T., Chen, X., Chen, D., 2024. Developing a dynamic-statistical downscaling framework for wind speed prediction for the Beijing 2022 Winter olympics. Clim Dyn 62, 7345–7363. https://doi.org/10.1007/s00382-024-07282-3

Wang, Y., Shi, H., Yang, X., Jiang, Y., Wu, Y., Shui, J., Liu, Y., Guo, M., Li, L., 2024. Spatial downscaling of SMAP soil moisture to high resolution using machine learning over China's Loess Plateau. CATENA 247, 108492. https://doi.org/10.1016/j.catena.2024.108492

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research 30, 79–82.

Winsemius, H.C., Savenije, H.H.G., Van De Giesen, N.C., Van Den Hurk, B., Zapreeva, E.A., Klees, R., 2006. Assessment of Gravity Recovery and Climate Experiment (GRACE) temporal signature over the upper Zambezi. Water Resources Research 42.

World Bank, 2019. World development report 2020: Trading for development in the age of global value chains. The World Bank.

Wu, Q., Si, B., He, H., Wu, P., 2019. Determining Regional-Scale Groundwater Recharge with GRACE and GLDAS. Remote Sensing 11, 154. https://doi.org/10.3390/rs11020154

Wu, Y.-K., Hong, C.-C., Chen, C.-T., 2018. Distinct Effects of the Two Strong El Niño Events in 2015–2016 and 1997–1998 on the Western North Pacific Monsoon and Tropical Cyclone Activity: Role of Subtropical Eastern North Pacific Warm SSTA. Journal of Geophysical Research: Oceans 123, 3603–3618. https://doi.org/10.1002/2018JC013798

Xu, L., Ferris, D., Huggins, X., Wong, J.S., Mohan, C., Sadri, S., Chandanpurkar, H.A., Sanyal, P., Famiglietti, J.S., 2023. From coarse resolution to practical solution: GRACE as a science communication and policymaking tool for sustainable groundwater management. Journal of Hydrology 623, 129845. https://doi.org/10.1016/j.jhydrol.2023.129845

Yamini, P.R., Manjula, R., 2023. Enhanced streamflow prediction using SWAT’s influential parameters: a comparative analysis of PCA-MLR and XGBoost models. Earth Sci Inform 16, 4053–4076. https://doi.org/10.1007/s12145-023-01139-9

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, 5973–5987.

Yin, W., Zhang, G., Han, S.-C., Yeo, I.-Y., Zhang, M., 2022. Improving the resolution of GRACE-based water storage estimates based on machine learning downscaling schemes. Journal of Hydrology 613, 128447. https://doi.org/10.1016/j.jhydrol.2022.128447

ZamStats, 2022. 2022 Census of Population and Housing: Preliminary Report. ZamStats, Lusaka, Zambia.

ZamStats, 2021. 2021 Labour Force Survey Report. ZamStats, Lusaka, Zambia.

Zeng, R., Zhang, Z., Zhao, S., Su, R., Wei, Z., Wang, X., Long, Z., Ma, J., Chen, G., Meng, X., 2024. Effects of human and tectonic activities on groundwater in the upper Yellow River terraces of the loess Plateau. Journal of Hydrology 645, 132279. https://doi.org/10.1016/j.jhydrol.2024.132279

Zhang, J., Liu, K., Wang, M., 2021. Downscaling Groundwater Storage Data in China to a 1-km Resolution Using Machine Learning Methods. Remote Sens. 2021, 13, 523.

Zhang, J., Lu, Z., Li, C., Lei, G., Yu, Z., Li, K., 2024. Estimation of groundwater-level changes based on GRACE satellite and GLDAS assimilation data in the Songnen Plain, China. Hydrogeol J 32, 1495–1509. https://doi.org/10.1007/s10040-024-02815-w

Zhang, M., Guo, Z.-Y., Dong, G.-T., Tan, J.-G., 2023. Projected heat wave increasing trends over China based on combined dynamical and multiple statistical downscaling methods. Advances in Climate Change Research 14, 758–767. https://doi.org/10.1016/j.accre.2023.09.001

Zhou, F., Fan, H., Liu, Y., Ye, Y., Diao, H., Wang, Z., Rached, R., Tu, Y., Davio, E., 2022. Application of xgboost algorithm in rate of penetration prediction with accuracy, in: International Petroleum Technology Conference. IPTC, p. D012S111R002.

Zhu, L., Gong, H., Dai, Z., Xu, T., Su, X., 2015. An integrated assessment of the impact of precipitation and groundwater on vegetation growth in arid and semiarid areas. Environ Earth Sci 74, 5009–5021. https://doi.org/10.1007/s12665-015-4513-5

Zimba, H., Coenders-Gerrits, M., Kawawa, B., Savenije, H., Nyambe, I., Winsemius, H., 2020. Variations in canopy cover and its relationship with canopy water and temperature in the miombo woodland based on satellite data. Hydrology 7, 58.

Zuo, J., Xu, J., Chen, Y., Li, W., 2021. Downscaling simulation of groundwater storage in the Tarim River basin in northwest China based on GRACE data. Physics and Chemistry of the Earth, Parts A/B/C 123, 103042. https://doi.org/10.1016/j.pce.2021.103042

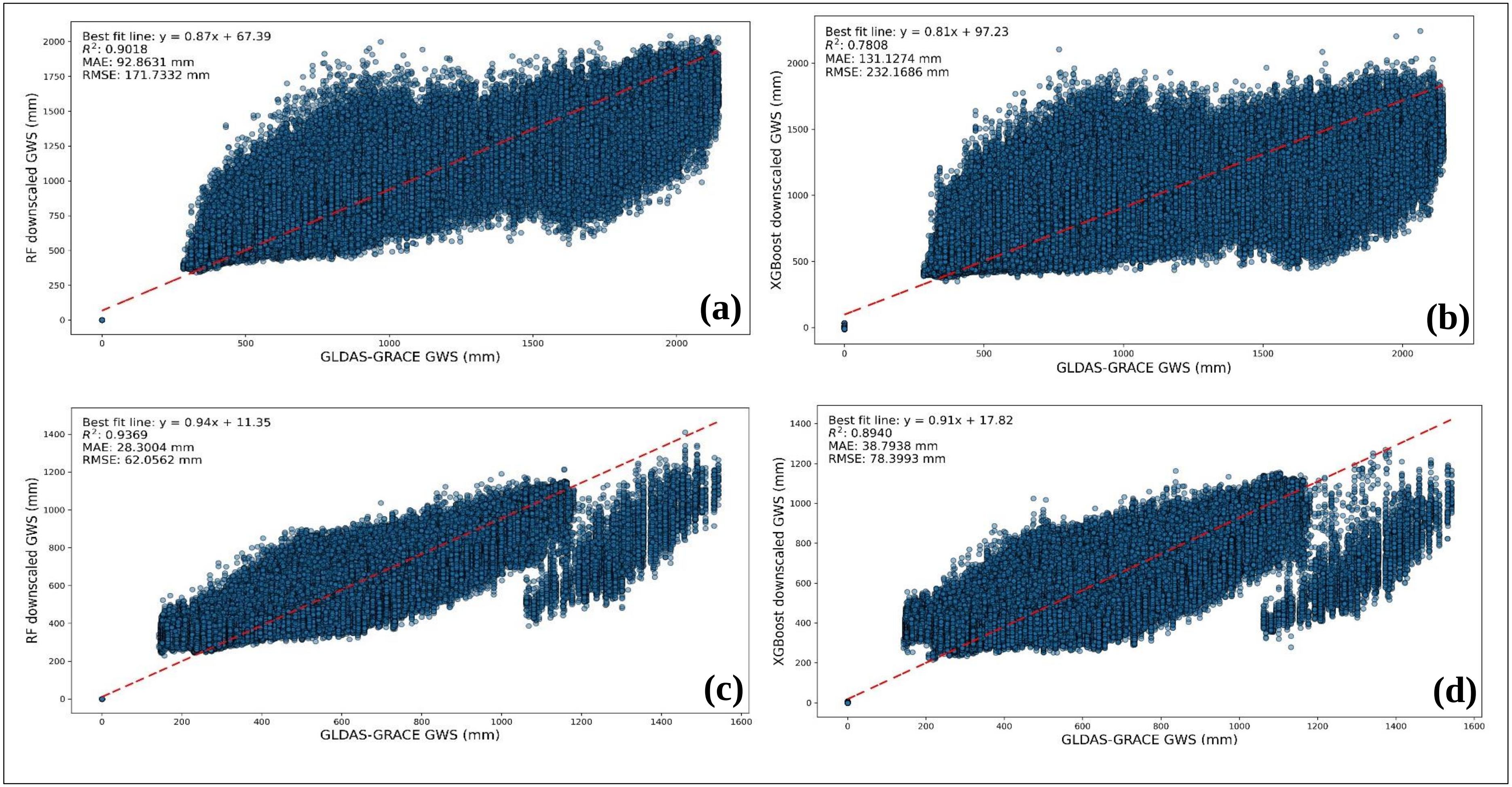
APPENDICES

Appendix 1: PFIT of input variables for the six sub-catchments of the UZC

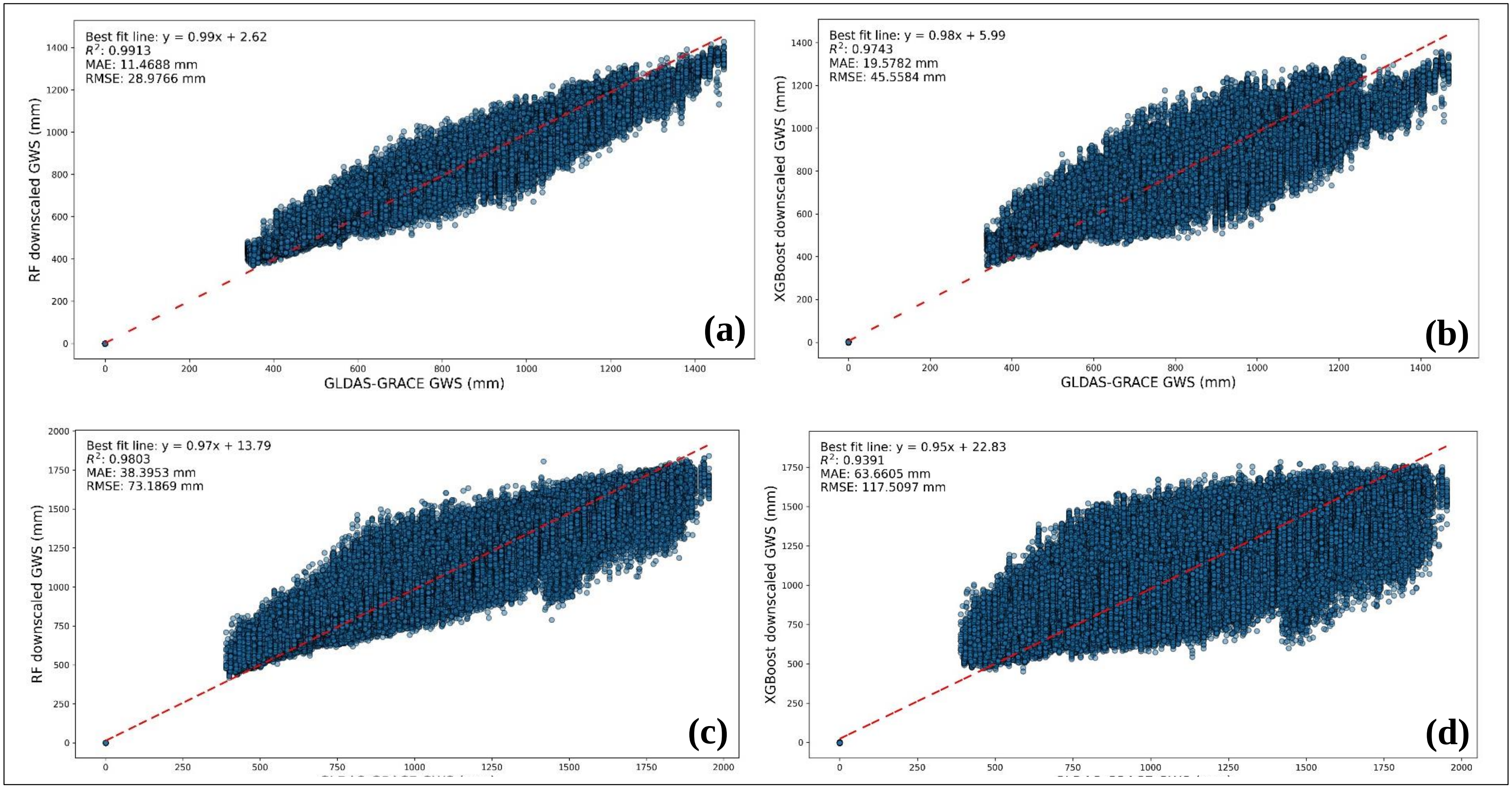
	PFIT											
	Barotse		**Chobe**		**Kabompo**		**Luanginga**		**Lunga**		**Upper Zambezi**	
Feature	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**	**RF**	**XGBoost**
Rainfall	0.0601	0.6409	0.0958	0.0051	0.1236	0.2266	0.1056	0.0040	0.0800	0.6592	0.0758	0.2909
ET	0.0946	0.0051	0.0772	0.0225	0.1450	0.0165	0.1696	0.0809	0.1387	0.0018	0.1673	0.0065
LST	0.0775	0.0234	0.0683	0.0010	0.0677	0.0323	0.1051	0.3244	0.0890	0.0005	0.1127	0.0293
NDVI	0.1567	0.1824	0.1443	0.6710	0.1138	0.0402	0.1757	0.0418	0.2193	0.0113	0.1081	0.0156
EVI	0.1914	0.0378	0.1058	0.2466	0.0967	0.0045	0.1298	0.0613	0.1498	0.0016	0.2020	0.0272
LSTNight	0.0900	0.0048	0.0318	0.0018	0.0232	0.0072	0.0269	0.0019	0.0100	0.0006	0.0508	0.0018
SoilMoisture	0.1249	0.0154	0.2417	0.0048	0.1598	0.4640	0.1514	0.4598	0.1300	0.3138	0.1247	0.5842
Rainfall_ET	0.0030	0.0024	0.0015	0.0006	0.0072	0.0091	0.0007	0.0019	0.0005	0.0007	0.0016	0.0022
Rainfall_LST	0.0044	0.0026	0.0022	0.0008	0.0073	0.0057	0.0164	0.0009	0.0093	0.0004	0.0119	0.0022
Rainfall_NDVI	0.0036	0.0025	0.0023	0.0014	0.0081	0.0075	0.0016	0.0010	0.0005	0.0004	0.0023	0.0024
Rainfall_EVI	0.0035	0.0024	0.0020	0.0010	0.0067	0.0060	0.0012	0.0010	0.0004	0.0004	0.0021	0.0012
Rainfall_LSTNight	0.0043	0.0036	0.0019	0.0008	0.0079	0.0081	0.0014	0.0010	0.0005	0.0005	0.0174	0.0013
Rainfall_SoilMoisture	0.0046	0.0055	0.0020	0.0011	0.0101	0.0112	0.0085	0.0018	0.0005	0.0005	0.0026	0.0042
ET_LST	0.0065	0.0032	0.0015	0.0006	0.0258	0.0141	0.0007	0.0006	0.0088	0.0005	0.0103	0.0009
ET_NDVI	0.0075	0.0075	0.0016	0.0006	0.0089	0.0309	0.0036	0.0022	0.0009	0.0015	0.0083	0.0012
ET_EVI	0.0256	0.0025	0.0016	0.0014	0.0085	0.0136	0.0033	0.0010	0.0016	0.0011	0.0061	0.0085
ET_LSTNight	0.0034	0.0026	0.0092	0.0008	0.0110	0.0191	0.0008	0.0010	0.0004	0.0004	0.0017	0.0014

ET_SoilMoisture	0.0058	0.0029	0.0014	0.0007	0.0133	0.0090	0.0006	0.0008	0.0008	0.0004	0.0016	0.0011
LST_EVI	0.0118	0.0033	0.0776	0.0271	0.0445	0.0092	0.0089	0.0009	0.0493	0.0006	0.0043	0.0021
LST_SoilMoisture	0.0438	0.0050	0.0216	0.0025	0.0364	0.0223	0.0289	0.0019	0.0103	0.0007	0.0362	0.0036
NDVI_LSTNight	0.0123	0.0255	0.0208	0.0010	0.0114	0.0065	0.0094	0.0009	0.0106	0.0004	0.0220	0.0026
NDVI_SoilMoisture	0.0188	0.0041	0.0500	0.0024	0.0209	0.0091	0.0057	0.0033	0.0298	0.0006	0.0041	0.0020
EVI_LSTNight	0.0076	0.0085	0.0132	0.0016	0.0068	0.0070	0.0013	0.0010	0.0006	0.0004	0.0035	0.0020
EVI_SoilMoisture	0.0177	0.0025	0.0206	0.0012	0.0193	0.0067	0.0026	0.0009	0.0483	0.0005	0.0040	0.0035
LSTNight_SoilMoisture	0.0206	0.0039	0.0041	0.0016	0.0161	0.0135	0.0405	0.0038	0.0100	0.0014	0.0186	0.0023

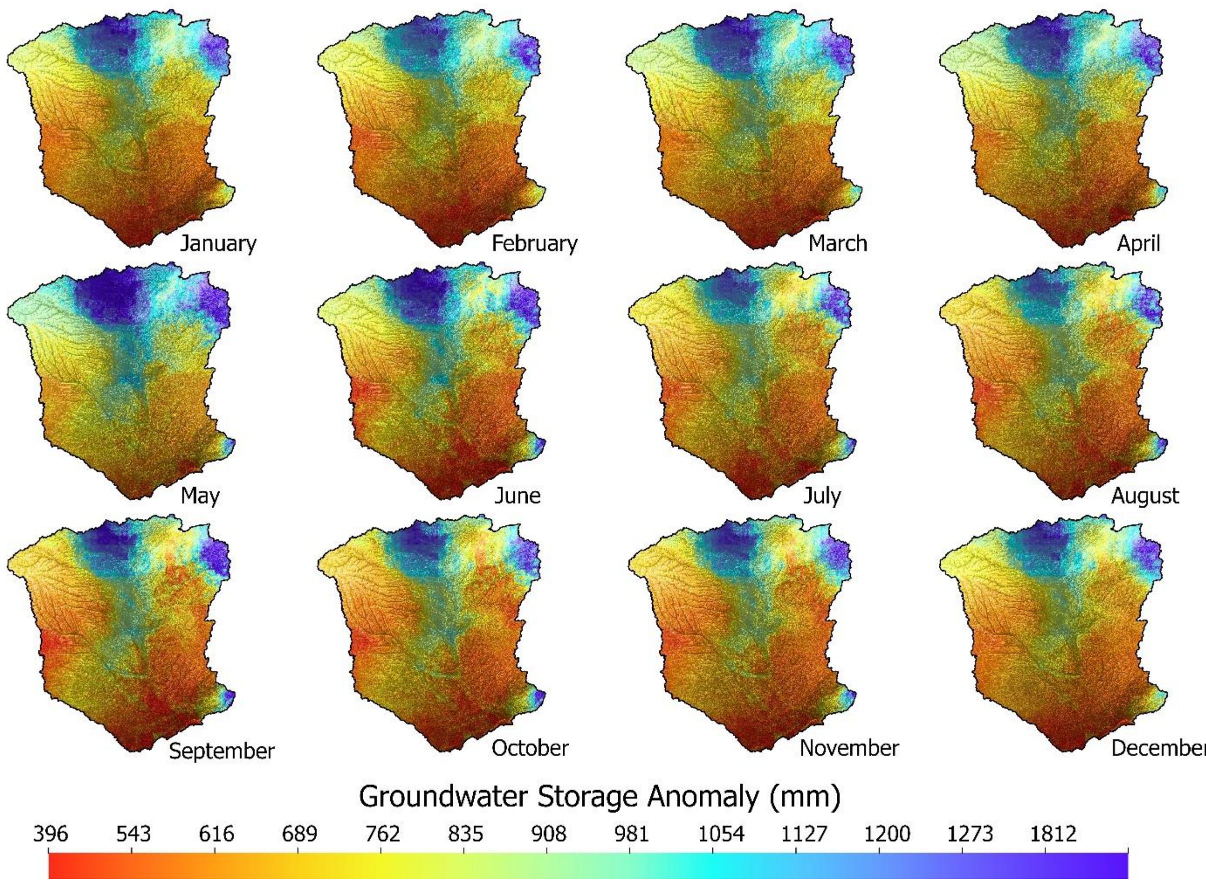
Appendix 2: The GLDAS-GRACE GWS and the results from the two downscaling models are shown in scatter plots, where: (a) and (b) represent the scatter plots for the model in the Kabompo for the RF and XGBoost models respectively; (c) and (d) represent the scatter plots for the model in the Chobe for the RF and XGBoost models respectively



Appendix 3: The GLDAS-GRACE GWS and the results from the two downscaling models are shown in scatter plots, where: (a) and (b) represent the scatter plots for the model in the Lunga for the RF and XGBoost models respectively; (c) and (d) represent the scatter plots for the model in the Upper Zambezi for the RF and XGBoost models respectively



Appendix 4: The RF GWS dataset's monthly spatiotemporal variation, represented by the monthly means from 2009 to 2023



Appendix 5: Research paper published in Environmental Systems Research by Springer Nature on Machine learning downscaling of GRACE/GRACE-FO data to capture spatial-temporal drought effects on groundwater storage at a local scale under data-scarcity

Shilengwe *et al. Environmental Systems Research* (2024) 13:38
https://doi.org/10.1186/s40068-024-00368-1

Environmental Systems Research

RESEARCH Open Access

Machine learning downscaling of GRACE/GRACE-FO data to capture spatial-temporal drought effects on groundwater storage at a local scale under data-scarcity

Christopher Shilengwe[1,2*], Kawawa Banda[2] and Imasiku Nyambe[2]

Abstract
The continued threat from climate change and human impacts on water resources demands high-resolution and continuous hydrological data accessibility for predicting trends and availability. This study proposes a novel threefold downscaling method based on machine learning (ML) which integrates: data normalization; interaction of hydrometeorological variables; and the application of a time series split for cross-validation that produces a high spatial resolution groundwater storage anomaly (GWSA) dataset from the Gravity Recovery and Climate Experiment (GRACE) and its successor mission, GRACE Follow-On (GRACE-FO). In the study, the relationship between the terrestrial water storage anomaly (TWSA) from GRACE and other land surface and hydrometeorological variables (e.g., vegetation coverage, land surface temperature, precipitation, and in situ groundwater level data) is leveraged to downscale the GWSA. The predicted downscaled GWSA datasets were tested using monthly in situ groundwater level observations, and the results showed that the model satisfactorily reproduced the spatial and temporal variations in the GWSA in the study area, with Nash-Sutcliffe efficiency (NSE) correlation coefficient values of 0.8674 (random forest) and 0.7909 (XGBoost), respectively. Evapotranspiration was the most influential predictor variable in the random forest model, whereas it was rainfall in the XGBoost model. In particular, the random forest model excelled in aligning closely with the observed groundwater storage patterns, as evidenced by its high positive correlations and lower error metrics (Mean Absolute Error (MAE) of 54.78 mm; R-squared (R^2) of 0.8674). The downscaled 5 km GWSA data (based on random forest) showed a decreasing trend in storage associated with variability in the rainfall pattern. An increase in drought severity during El Niño lengthened the full recovery time of groundwater based on historical storage trends. Furthermore, the time lag between the occurrence of precipitation and recharge was likely controlled by the drought intensity and the spatial recharge characteristics of the aquifer. Projected increases in drought severity could further increase groundwater recovery times in response to droughts in a changing climate, resetting storage to a new tipping condition. Therefore, climate change adaptation strategies must recognise that less groundwater will be available to supplement the surface water supply during droughts.