MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

Université Felix Houphouët-Boigny

UNITE DE FORMATION ET DE RECHERCHE
SCIENCES DES STRUCTURES DE LA
MATIERE ET DE TECHNOLOGIE

RÉPUBLIQUE DE CÔTE D'IVOIRE
UNION - DISCIPLINE - TRAVAIL

Nº 836

INTERNATIONAL MASTER PROGRAM
IN RENEWABLE ENERGY AND GREEN HYDROGEN

SPECIALITY: Georesources (Water/Wind) And Green Hydrogen Technology

MASTER THESIS

Topic:

MACHINE LEARNING-BASED ASSESSMENT OF REGIONAL GROUNDWATER RECHARGE VARIABILITY AND ITS IMPLICATIONS FOR SUSTAINABLE WATER SUPPLY IN RENEWABLE ENERGY PROJECTS ACROSS AFRICA

Presented on September 25, 2025, by:

Afis AKIBODE

JURY:

Mme. ZORO Georgina Emma H.	**President**	Associate Professor, UFHB
M. KOFFI Aka Stéphane	**Examiner**	Assistant Professor, UFBH
M. KACOU Modeste Huberson Ahiba	**Supervisor**	Senior Researcher, UFHB
M. Harrie-Jan Hendricks-Franssen	**Co-Supervisor**	Professor, Jülich FZ - IBG 3
M. Bamidele Oloruntoba	**Co-Supervisor**	PHD, Jülich FZ - IBG 3

Academic year: 2024-2025

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to the West African Science Service Center on Climate Change and Adapted Land Use **(WASCAL)** for granting me the opportunity to be part of this remarkable program, and to the German Federal Ministry of Education and Research **(BMBF)** for their generous financial support that made this journey possible.

I extend my sincere appreciation to the President of Félix Houphouët-Boigny University, Côte d'Ivoire, **Prof. BALLO Zié**, for kindly accepting and accommodating me during the third and fourth semesters of my master's program.

My heartfelt thanks go to **Dr. Kouassi K. Edouard**, Directors of WASCAL Côte d'Ivoire, for their continuous support and encouragement throughout this program.

I wish to acknowledge **Prof. N'golo A. Koné**, Deputy Director of WASCAL Côte d'Ivoire, and **Dr. Wanignon Ferdinand Fassinou**, Coordinator of the H2 Programme, for their invaluable leadership.

I would like to thank the Directors of WASCAL Niger, **Prof. Rabani Adamou,** for their support during the first year of the program.

I would like to express my sincere gratitude to **Prof. Harry Vereecken,** Director of the Institute of Bio and Geosciences - Agrosphere (IBG-3) for accepting me into his institute for my internship.

My sincere thanks go to my supervisors: **Dr. Kacou Modeste,** for his time and useful contributions; **Prof Harrie-Jan Hendricks-Franssen,** for his precious insights into this work; and **M. Bamidele Oloruntoba,** who was much more than a mentor.

I would like to extend my sincere thanks to **Dr Bagher Bayat** for his valuable contribution to this work.

I would also like to thank the jury members for their outstanding job during the defense. The president, **Dr ZORO Georgina Emma,** and the examiner, **Dr KOFFI Aka Stéphane.**

To the entire **WASCAL community** across Niger, Côte d'Ivoire, and Germany, I am grateful for the collective effort that shaped my academic and personal growth during these four training semesters.

ABSTRACT

Water is a precious resource in Africa. The sustainable management of this resource is important for our everyday lives, as well as for renewable energy applications such as green hydrogen production. Especially in desert regions, this water is mainly groundwater. This study assesses how land surface characteristics affect groundwater recharge across Africa, in addition to the dominant impact of precipitation. A combination of correlation analysis, principal component analysis (PCA), and machine learning models (Random Forest and Gradient Boosting) interpreted with Shapley additive exPlanations (SHAP) was applied across eight African regions. These methods were used to analyse the Community Land Model (CLM) simulated datasets for different African land surface characteristics. The findings of the analysis indicate that precipitation is the main factor affecting variations in recharge. However, its effectiveness is significantly impacted by factors such as soil depth, slope, vegetation, organic matter, and soil texture. Three dominant recharge regimes were identified. These are runoff-limited regimes, evapotranspiration-limited regimes, and precipitation-constrained regimes (e.g., Sahara and Mediterranean). The runoff-limited regimes are characterized by shallow soils and steep terrain that restrict infiltration despite high rainfall. The evapotranspiration-limited regimes are characterized by vegetation and organic-rich soils that drive moisture losses. The precipitation-limited regimes are characterized by low rainfall that dominates recharge regardless of land characteristics. Random Forest models outperformed Gradient Boosting in predictive accuracy (R^2 up to 0.98), and SHAP analysis provided quantifications of variable importance. These findings highlight the critical role of land-surface heterogeneity in shaping groundwater availability and its implications for water-energy planning. In renewable energy strategies that involve groundwater, policies should consider recharge variability. They should also manage soil and slopes in runoff-limited regions. And they should assess the balance between the water needs for energy production and competing demands such as drinking water and environmental needs. The study emphasises the importance of adapting region-specific approaches to groundwater management. These are needed to support Africa's renewable energy transition.

Keywords: Groundwater recharge, Land surface characteristics, Machine learning, Africa, Water-Energy nexus.

RÉSUMÉ

L’eau est une ressource précieuse en Afrique. La gestion durable de cette ressource est importante pour notre vie quotidienne, mais aussi pour des applications en énergie renouvelable telles que la production d’hydrogène vert. Particulièrement dans les régions désertiques, cette eau est l'eau souterraine. Cette étude évalue comment les caractéristiques de la surface terrestre affectent la recharge des eaux souterraines à travers l’Afrique, en plus de l’impact dominant des précipitations. La combinaison d’analyse de corrélation, de composantes principales et de modèle d’apprentissage automatique (Random Forest et Gradient Boosting) interprétée par SHAP (Shapley Additive exPlanations) a été appliquée à travers huit régions d’Afrique. Ces méthodes sont utilisées pour analyser des données simulées par CLM (Community Land Model), représentant les caractéristiques de la surface terrestre. Les résultats des analises indiquent que les variations de la recharge des eaux souterraines sont plus impactées par la précipitation. Cependant, les facteurs tels que la profondeur du sol, la pente, la végétation, la matière organique et la texture du sol influencent aussi la variation de la recharge. Trois régimes de recharge sont identifiés. Il s'agit des régimes limités par le ruissellement, des régimes limités par l'évapotranspiration et des régimes limités par les précipitations (par exemple, le Sahara et la Méditerranée). Les régimes limités par le ruissellement se caractérisent par des sols peu profonds et un terrain en pente qui limitent l'infiltration malgré des précipitations élevées. Les régimes limités par l'évapotranspiration se caractérisent par une végétation et des sols riches en matière organique favorisant les pertes d'humidité. Les régimes limités par les précipitations se caractérisent par des précipitations faibles qui dominent la recharge, quelles que soient les caractéristiques du terrain. Les modèles Random Forest ont surpassé le Gradient Boosting en termes de précision prédictive (R^2 jusqu'à 0,98), et l'analyse SHAP a fourni une interprétation robuste de l'importance des variables. Ces résultats soulignent le rôle essentiel de l'hétérogénéité de la surface terrestre dans la détermination de la disponibilité des eaux souterraines et ses implications pour la planification de l'eau et de l'énergie. Les politiques devraient intégrer la variabilité de la recharge dans les stratégies en matière d'énergies renouvelables, donner la priorité à la gestion des sols et des pentes dans les régions où la recharge est limité par le ruissellement, et évaluer soigneusement l'équilibre entre l'utilisation de l'eau pour la production d'énergie et les demandes concurrentes telles que l'eau potable, l'agriculture et les écosystèmes dans les environnements arides. L'étude souligne la nécessité d'adopter des approches spécifiques à chaque région en matière de gestion des eaux souterraines afin de soutenir le développement durable et la transition vers les énergies renouvelables en Afrique.

Mots-clés: recharge des eaux souterraines, caractéristiques de la surface terrestre, apprentissage automatique, Afrique, nexus eau-énergie.

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

ADLMs	: Advanced Deep Learning Models
ANNs	: Artificial Neural Networks
CAF	: Central Africa
CEAF	: Central East Africa
CLM	: Community Land Model
CSP	: Concentrated Solar Power
ET	: Evapotranspiration
ET_R	: Evapotranspiration ratio
HEC-HMS	: Hydrologic Engineering Center's Hydrologic Modeling System
IRENA	: International Renewable Energy Agency
IRRIG	: Irrigation
GB	: Gradient Boosting
GWR	: Groundwater recharge
LAI	: Leaf Area Index
LMM	: Linear Mixed Model
MED	: Mediterranean
ML	: Machine Learning
NEAF	: Northeast Africa
PCA	: Principal Component Analysis
PCT_CLAY	: Percentage of clay
PCT_SAND	: Percentage of sand
PDP	: Partial Dependence Plots

PFI	: Permutation Feature Importance
PRECIP	: Precipitation
PV	: Photovoltaic
RF	: Random Forest
RMSE	: Root Mean Squared Error
RO_R	: Runoff ratio
SAH	: Sahara
SDG	: Sustainable Development Goal
SEAF	: Southeast Africa
SHAP	: SHapley Additive exPlanations
SVMs	: Support Vector Machines
SWAF	: Southwest Africa
SWAT	: Soil and Water Assessment Tool
UN	: United Nation
UNESCO	: United Nations Educational, Scientific and Cultural Organization
WAF	: West Africa
XAI	: Explainable Artificial Intelligence
ZBEDROCK	: Soil depth

LIST OF FIGURES

LIST OF TABLES

GENERAL INTRODUCTION

GENERAL INTRODUCTION

In many regions of the African continent, especially in arid and semi-arid areas, where surface water availability is limited, groundwater is a vital source of water for domestic, agricultural, and industrial use, as well as for ecosystems (Gebreslassie et al., 2025; MacDonald et al., 2012). Groundwater recharge (GWR), defined as the process by which water percolates from the surface to replenish underground aquifers, is therefore crucial for the sustainable management of water resources (Gebreslassie et al., 2025). GWR however, is a complex and spatially heterogeneous process influenced by both climatic variables and land surface characteristics (Fu et al., 2019).

Traditionally, studies on GWR have focused on dominant climatic drivers such as precipitation and potential evapotranspiration (Atawneh et al., 2021). While these are undeniably important, recent research suggests that land surface features have a significant impact on how water percolates and accumulates underground (Fu et al., 2019; Toure et al., 2024). Understanding these interactions is especially critical in the African context, where land conditions vary widely across regions (Oloruntoba et al., 2025) and data scarcity (Akpoti et al., 2024) presents additional challenges to water resource planning.

Concurrently, the growing shift toward renewable energy technologies, including concentrated solar power (CSP), hydropower, and green hydrogen production, intensifies the demand for sustainable water supply. These technologies depend heavily on reliable water resources, making groundwater management an essential pillar of Africa's renewable energy transition (IRENA, 2021; Winkler et al., 2025). For hydropower, aquifer-fed baseflow helps sustain dry-season river discharge, ensuring continuous electricity generation (Bardsley, 1995). In the case of green hydrogen, electrolysis requires around 9 liters of pure water per kilogram of hydrogen produced (Scholz, 2024), meaning that large-scale deployment will place significant pressure on existing water resources. Concentrated solar power plants, such as Morocco's Noor complex, also highlight the water-energy balance: initial wet-cooling designs consumed millions of cubic meters of water annually, before later phases adopted dry-cooling to reduce demand (Fares & Abderafi, 2018; Ersoy et al., 2022).

Despite increasing recognition of the importance of groundwater for sustainable development (Adom et al., 2022; Anghileri et al., 2024; Biazar et al., 2025), there remains limited understanding of how non-climatic factors, specifically land surface characteristics, influence GWR across diverse African regions. Moreover, traditional hydrological modeling approaches may not fully capture the complex, nonlinear relationships governing GWR variability across space. There is therefore a need in addition to known traditional methods, a data-driven, region-specific analyses that leverage the potential of machine learning (ML) and explainable artificial intelligence (XAI) to better understand the drivers of groundwater recharge and their implications for sustainability (Jung et al., 2024; Maity et al., 2024; Siabi et al., 2022).

The main objective of this study is to assess how land surface characteristics influence groundwater recharge across different regions of Africa, alongside the primary factors of precipitation and potential evapotranspiration. Specifically, the study aims to identify which land surface traits affect GWR variability across Africa and to evaluate the roles of precipitation and land surface features in this variability. Furthermore, it seeks to quantify how regional land surface characteristics contribute to estimated GWR using machine learning and feature importance analysis. Finally, the study examines how variations in Africa's land surface impact the amount of water that can be safely used in renewable energy projects, such as hydropower and green hydrogen production. It also provides recommendations for policymakers.

The work combines hydrology, data science, and energy planning. It provides policy-relevant recommendations to optimize the use of groundwater in renewable energy projects across Africa. These outcomes are expected to contribute to the sustainable resources management and align with Sustainable Development Goal (SDG) 6, Clean Water and Sanitation, and SDG 7, Affordable and Clean Energy.

This master’s thesis is divided into three main chapters: the chapter 1 provides a comprehensive review of existing literature. Then, the chapter 2 describes the materials and methods employed in this study and the chapter 3 focuses on presenting the results obtained and discussing their implication.

CHAPTER 1: LITERATURE REVIEW

CHAPTER 1: LITERATURE REVIEW

This chapter provides a comprehensive review of existing literature on groundwater recharge in Africa, the impact of land surface characteristics on GWR, the use of machine learning for assessment, and the nexus between groundwater and renewable energy development. The review identifies gaps and sets the foundation for this study's approach to improving groundwater resource planning through advanced modeling and data analytics.

1.1. Groundwater Recharge and Land Surface Interactions

1.1.1. Concept and Mechanisms of Groundwater Recharge (GWR)

Groundwater recharge (GWR) is a fundamental hydrological process through which water percolates from the surface to replenish underground aquifers (Gebreslassie et al., 2025). It plays a critical role in maintaining water availability for agriculture, domestic use, and industrial activities, especially in regions with limited surface water resources (Sishodia et al., 2018). Recharge can be classified into natural and artificial types. The former occurs through precipitation infiltration, river seepage, or diffuse flow, while the latter involves managed aquifer recharge techniques (Figure 1.1.).

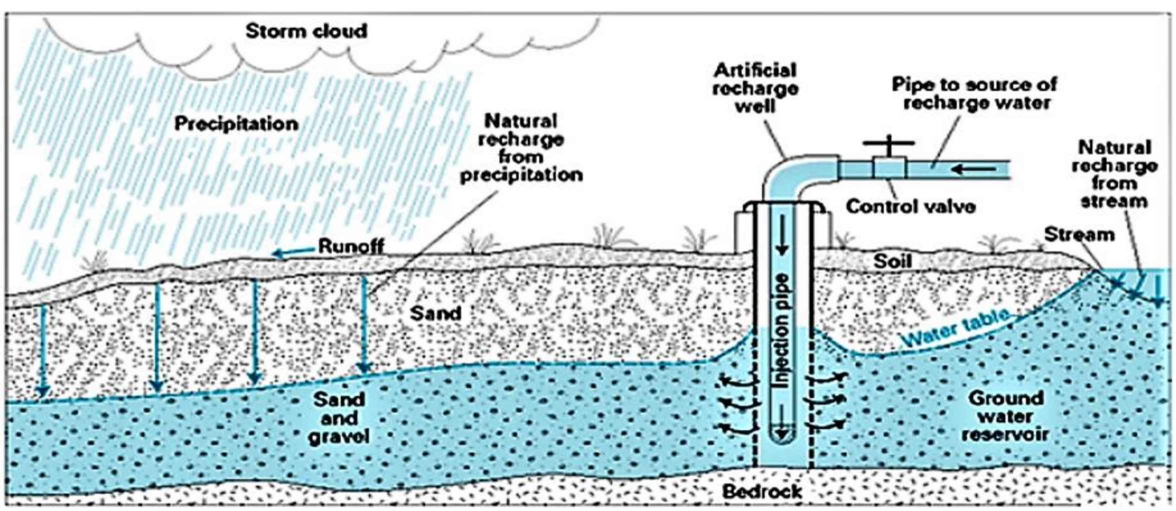

Figure 1.1. Groundwater recharge processes

Source: https://sswm.info/water-nutrient-cycle/water-sources/hardwares/precipitation-harvesting/subsurface-groundwater-recharge-

GWR is closely linked to the water cycle and is influenced by several factors, including soil type, land use, vegetation cover, and climatic conditions (Scanlon et al., 2012).

1.1.2. Role of Land Surface Characteristics in GWR Variability

Beyond climatic factors like rainfall and evapotranspiration, the rate and spatial distribution of GWR are strongly controlled by land surface and soil properties (Fu et al., 2019). Soil texture and structure affect infiltration and water-holding capacity (Basset et al., 2023; Franzluebbers, 2002). Land cover features, such as vegetation density and root depth, influence evapotranspiration and water retention (Alam, 2017). Topography and slope control runoff generation and the spatial redistribution of infiltrated water (Toure et al., 2024).

1.1.3. Challenges in GWR Estimation Across Africa

The estimation of groundwater recharge is challenging due to the lack of reliable data, the differences in estimation methods, and the influence of both climatic and non-climatic factors (Gebru et al., 2024; Wang et al., 2010). Unlike surface water, recharge cannot be directly observed at large scales; it must be inferred from indirect measurements or models. Common methods include water balance calculations, analysis of groundwater level fluctuations, tracer techniques (like isotopes and chloride profiles), and modeling (Gebru et al., 2024; Wang et al., 2010). Water balance approaches quantified recharge as the residual of precipitation after accounting for other fluxes. Within this framework, recharge can be calculated by subtracting actual evapotranspiration and surface runoff over a defined period. The general water balance equation can be expressed as Equation 1 (Wang et al., 2010):

$$R = P - R - ET \pm \Delta S \qquad \textit{Equation 1}$$

In this equation, R is recharge, P is precipitation, ET is evapotranspiration, Q is surface runoff and ΔS is the change in soil or groundwater storage. Tracer methods quantified groundwater recharge by tracking the movement of natural or applied chemical, isotopic, or gaseous markers in the subsurface (Scanlon, 2010). The chloride mass balance method is one of the most used tracer techniques to estimate groundwater recharge. This approach is based on the conservative nature of chloride, which is supplied naturally through precipitation and atmospheric deposition. As precipitation (P) infiltrates the soil, water is lost through evapotranspiration, but chloride remains and becomes more concentrated in the recharge (R) water. By comparing the chloride concentration in precipitation (C_p) and its concentration in groundwater (C_{gw}), recharge is estimate using Equation 2 (Scanlon et al., 2002):

$$R = \frac{P.C_p}{C_{gw}} \qquad \textit{Equation 2}$$

Although tracer methods are valuable in arid and semi-arid areas, they are not suitable for estimating regional groundwater recharge (Gebru et al., 2024; Wang et al., 2010). Modelling approaches such as MODFLOW and ParFlow are governed by hydrodynamic mechanisms. These models physically describe the movement of water in three dimensions, including in both saturated and unsaturated zones. They incorporate water balance processes, but the physical mechanism of evapotranspiration processes is not usually described explicitly (Tian et al., 2012). The selection of the method used mostly depends on the data available (Atawneh et al., 2021; Gebru et al., 2024). Spatially, GWR is influenced by soil type and land attributes. It is temporally also impacted by interannual weather variability (Fu et al., 2019). The integration of climate and land surface data is thus crucial for GWR assessment.

1.2. Machine Learning for Hydrology and GWR Analysis

1.2.1 Traditional vs. Machine Learning Approaches in Hydrology

Traditional models have long been used to simulate hydrological processes. These approaches rely on physics-based methods that require extensive calibration and domain-specific knowledge (Biazar et al., 2025). Some examples include catchment-scale models such as the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) and the Soil and Water Assessment Tool (SWAT) model (Biazar et al., 2025). Traditional models offer the advantage of providing physically interpretable outputs and the flexibility to simulate multiple hydrological variables (Devia et al., 2015). However, these models require extensive calibration, since many parameters cannot be measured directly. This often leads to multiple parameters sets providing similar results, which damages confidence in the robustness of the model (Beven, 2006). According to Du & Pechlivanidis (2025), traditional models demonstrate inaccuracies of processes due to their inability to represent complex and non-linear hydrometeorological processes. This limits their effectiveness in local conditions. Machine learning (ML) models like neural networks and decision trees are data-driven. These models focus on identifying statistical correlations between input and output variables. They learn and capture nonlinearities and interactions directly from data without the need for a physical prescription (Biazar et al., 2025). Nevertheless, machine learning models are frequently criticized for their limited interpretability (Nearing et al., 2021).

1.2.2. Key Machine Learning Techniques for Water Resource Modeling

A variety of machine learning algorithms have been applied in groundwater studies, each with strengths and limitations depending on the problem type and available data. Popular algorithms include Artificial Neural Networks (ANNs), Tree-based Ensembles, Support Vector Machines (SVMs), and Advanced Deep Learning Models (ADLMs) (Biazar et al., 2025; Pazola et al., 2024; Toure et al., 2024). For spatial mapping tasks, like predicting recharge or groundwater potential across a region from environmental attributes, ensemble tree methods are more popular. Random Forest (RF) and Gradient Boosting Machines (GBM) are frequently used. They handle nonlinear relationships and interactions well (Anand et al., 2025; Gómez-Escalonilla et al., 2022), and are relatively robust to overfitting, while also providing measures of feature importance. For example, Pazola et al. (2024) used a Random Forest (RF) model to generate a high-resolution recharge map for the African continent, and RF showed best performance compared to a previous continent-wide estimate derived from a linear mixed model (LMM). However, RF models can be computationally expensive when dealing with very large datasets (Genuer et al., 2017), while GB models are more sensitive to noise (Xiang et al., 2020). For time-series forecasting problems, such as predicting groundwater level fluctuations or recharge over time, neural networks and deep learning models are common. In Ghana, for example, Siabi et al. (2022) used artificial neural networks to predict groundwater recharge based on historical data from 1960 to 2018 (58 years). Despite their importance, artificial neural networks respond slowly to gradient-based learning processes and require repeated parameter tuning (Mosavi et al., 2018). Overall, the choice of model is dependent on a number of factors, including the predictive accuracy, the importance of interpretability, and the availability of the computational process.

1.2.3. Explainable AI for Feature Importance Analysis in GWR Studies

Despite the power of ML models, a major limitation is their lack of interpretability. Explainable Artificial Intelligence (XAI) methods such as partial dependence plots (PDP), SHAP (SHapley Additive exPlanations), and permutation feature importance (PFI) have been developed to interpret complex ML models (Holzinger et al., 2020). SHAP is an XAI method that assigns each feature a contribution value toward prediction, grounded in cooperative game theory (Lundberg & Lee, 2017). It allows researchers to assess not only which features are most influential across the model but also how feature importance varies with context and interactions.

1.3. Groundwater and Renewable Energy Nexus in Africa

1.3.1. Water Needs in Renewable Energy Projects

Water plays a critical role in several renewable energy projects, serving both as a resource and a constraint in energy system planning. The sustainable management of food, energy, and water is closely interlinked, and it is essential to evaluate these connections (UNESCO, 2018). In hydropower, water is the primary energy source, with energy generated from kinetic and potential energy of flowing or stored water. While hydropower dams rely on surface river flow, a significant portion of dry-season river baseflow is fed by groundwater discharge. In other words, aquifers slowly draining into rivers help maintain streamflow during periods with no rain, which is crucial for year-round hydropower generation (Bardsley, 1995). The role of groundwater is even more direct in the green hydrogen sector. Green hydrogen production involves electrolyzing water using renewable electricity (solar/wind) to generate hydrogen fuel (IRENA, 2021). The reaction requires pure water as an input, typically around 9 liters of water per kilogram of hydrogen produced (Scholz, 2024). This water demand is substantial when scaled to industrial production. In arid regions, groundwater becomes an obvious candidate to supply electrolysis plants. However, reliance on groundwater raises sustainability flags: a recent evaluation found that only a small fraction (on the order of 16%) of the technical hydrogen production potential in Sub-Saharan Africa could be met by sustainable groundwater yields (Winkler et al., 2025). Another sector is solar energy, particularly concentrated solar power (CSP) plants and large photovoltaic (PV) farms. Utility-scale CSP plants (such as Morocco’s Noor Uranate complex) use mirrors to concentrate sunlight and heat a working fluid to drive turbines. Traditional CSP designs use wet cooling for the power cycle, consuming significant water for cooling towers. For example, the Noor I plant used wet cooling and consumed about 1.8 million m^3 of water per year (Fares & Abderafi, 2018). In arid environments, such water demand is problematic. Morocco mitigated this by switching later CSP phases (Noor II and III) to dry cooling (Ersoy et al., 2022). Nonetheless, CSP plants still need some water for mirror cleaning to maintain efficiency in dusty desert air. Groundwater can be an important source for these needs, especially if surface water allocations are limited. Integrating water availability assessments, including GWR estimates, into energy planning is therefore crucial.

1.3.2. Challenges in Integrated Water-Energy Planning in Africa

The effective implementation of water-energy planning in Africa is hindered by multiple challenges. These challenges include institutional and political barriers (Adom et al., 2022; Anghileri et al., 2024; Nhamo et al., 2018), which limits the coordination and execution of integrated strategies. According to Donkor & Wolde (2022), the separation of water and energy management across various departments leads to inadequate resource allocation and a lack of integrated planning. Also, the lack of reliable data on energy and water resources represents a substantial obstacle to making evidence-based decisions (UN-Water/Africa, 2000). These issues are made worse by financial limitations on the continent (Chigozie Ani et al., 2024). Furthermore, the existence of numerous transboundary river and lake basins introduces an additional layer of intricacy, necessitating the maintenance of sustained regional collaboration and diplomatic coordination (UN-Water/Africa, 2000). Combined, these challenges reflect the urgent need for strengthened institutional capacity, improved data systems, enhanced funding mechanisms, and more harmonized governance to support the advancement of integrated water-energy planning in Africa.

1.3.3. Policy and Sustainability Considerations

Governments must enhance adaptation strategies. They should implement capacity building and awareness campaigns on the water-energy nexus (Donkor & Wolde, 2022). The mobilization of financial resources continues to represent a considerable challenge; nevertheless, innovative financing strategies, such as blended finance approaches, have the potential to unlock both public and private capital for infrastructure development (Leigland et al., 2016; Tonkonogy et al., 2018). Transparent outputs from machine learning tools can support integrated planning further by identifying vulnerable areas. These will guide efficient decisions towards water and energy management.

PARTIAL CONCLUSION

This chapter has reviewed the current knowledge on groundwater recharge and its interactions with land surface characteristics. It emphasizes the importance of machine learning as a powerful tool in hydrology. The connection between groundwater and renewable energy development in Africa is also mentioned.

CHAPTER 2: MATERIALS AND METHODS

CHAPTER 2: MATERIALS AND METHODS

This chapter outlines the study area and the datasets, analytical tools, and software employed to process and analyze the data.

2.1. Study Area

This research focuses on the African continent. Given its diverse land surface characteristics and climate conditions, the continent is subdivided into eight regions, following the climate regionalization framework proposed by Oloruntoba et al. (2025). These regions include the Mediterranean (MED), the Sahara (SAH), West Africa (WAF), Central Africa (CAF), Northeast Africa (NEAF), Central East Africa (CEAF), Southeast Africa (SEAF), and Southwest Africa (SWAF). This subdivision enables a more region-specific investigation of groundwater recharge, helping to minimize the masking effects that a continental-scale analysis would impose. Figure 2.1 highlights the spatial extent of each region.

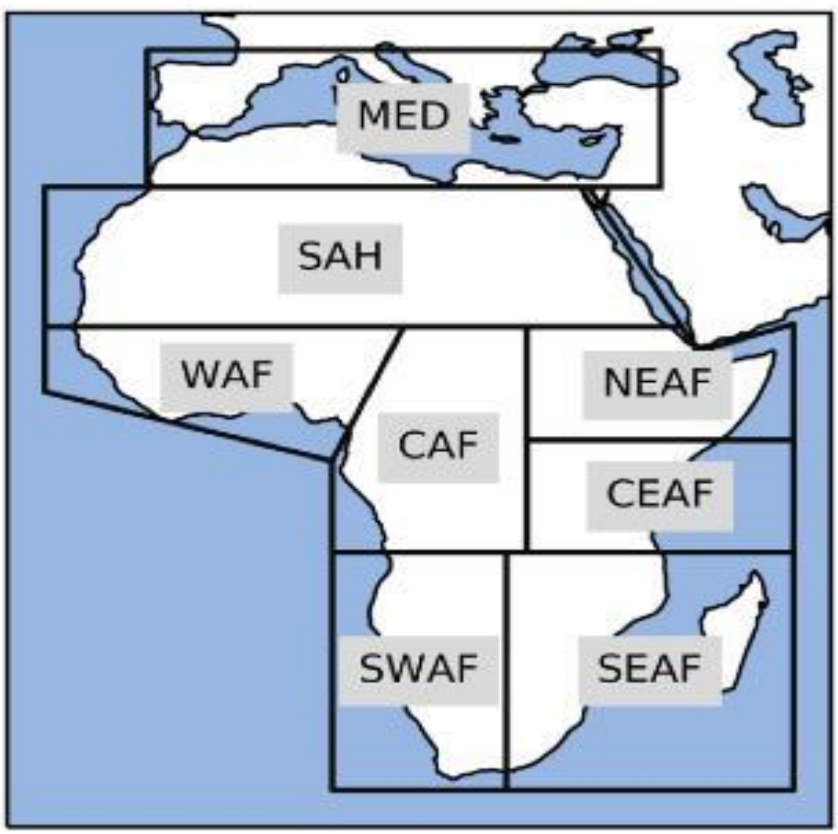

Figure 2.1. Classification of the African regions according to Oloruntoba et al. (2025)

2.2. Data Sources

2.2.1. Hydrology and Land Surface Data

Hydrological and land-surface variables were obtained from the Community Land Model (CLM), version 5 (CLM5), which provides physically based simulations of land-atmosphere interactions at the continental scale (CTSM, 2020; Lawrence et al., 2019). The CLM was run using the Global Soil Wetness Project, version 3 (GSWP3) atmospheric forcing dataset, which supplies precipitation and other meteorological drivers (Kim, 2017). The key variables selected for this study are precipitation (PRECIP, mm/year), evapotranspiration (ET, mm/year), runoff (RUNOFF, mm/year), percentage of Clay (PCT_CLAY, %), percentage of Sand (PCT_SAND, %), soil depth (ZBEDROCK, m), mean topographic slope (SLOPE, degrees), Organic Matter Density (ORGANIC, kg/m^3), and Leaf Area Index (LAI) representing vegetation density.

2.2.2. Groundwater Recharge Data

The Groundwater recharge data were obtained using the general water balance approach, Equation 3 (Bayat et al., 2023).

$$\mathrm{GWR} = (\mathrm{PRECIP} + \mathrm{IRRIG}) - (\mathrm{ET} + \mathrm{RUNOFF}) \qquad \textit{Equation 3}$$

where:

- PRECIP = Precipitation (mm/year)
- IRRIG = Irrigation water input (mm/year)
- ET = Evapotranspiration (mm/year)
- RUNOFF = Surface runoff (mm/year)

As Irrigation is negligible compared to other variables, we consider the determination of GWR to be as Equation 4:

$$\mathrm{GWR} = \mathrm{PRECIP} - (\mathrm{ET} + \mathrm{RUNOFF}) \qquad \textit{Equation 4}$$

2.2.3. Derived Hydrological Ratios

Two dimensionless indices were calculated to evaluate water partitioning:

- Evapotranspiration Ratio (ET_R):

Equation 5 was used to determine the fraction of precipitation lost to evapotranspiration.

$$ET_R = \frac{ET}{PRECIP} \qquad \textit{Equation 5}$$

- Runoff Ratio (RO_R):

We used Equation 6 to calculate the fraction of rainfall that leaves as surface runoff.

$$RO_R = \frac{RUNOFF}{PRECIP} \qquad \textit{Equation 6}$$

2.3. Data Processing and Analysis

2.3.1. Methods, Techniques, And Software

This study uses a dataset provided in NetCDF (Network Common Data Form), a widely adopted format for storing multidimensional scientific data. Several materials are employed to process and analyse these data effectively:

- **NetCDF Operator (NCO):**

  NCO is a suite of command-line programs specifically designed for manipulating and analyzing NetCDF files (Zender, 2008). In this study, NCO was used for initial data manipulation tasks, including a spatial view of each variable, as well as the head of each dataset (showing general information about each variable: Name, Dimension, Unit, etc.).

- **Climate Data Operator (CDO):**

  CDO is another powerful set of command-line tools, originally developed to process and analyse data produced by climate and numerical weather prediction models (Kaspar et al., 2010). We employed CDO for more advanced manipulation, such as subsetting, regridding, temporal aggregation, and merging.

- **Python:**

  Python is currently the fastest-growing programming language in the world due to its ease of use, quick learning curve, and numerous high-quality packages for data science and machine learning (Vallat, 2018). In this study, Python is predominantly used for data analysis and visualisation. The key libraries used are **Xarray** (for handling multidimensional arrays), **Pandas** (for data manipulation), **Matplotlib** and **Seaborn** (for plotting), **Scikit-learn** (for machine learning applications), **NumPy** (for efficient numerical computations and array operations), and **Cartopy** (for geospatial data visualisation).

- **Quantum Geographic Information System (QGIS):**

  QGIS is a free, open-source geographic information system that provides tools for visualising, analysing, and mapping spatial data (Elakkiya & Sankarganesh, 2023). QGIS is used in this study to extract and generate region-specific datasets corresponding to eight predefined regions from the larger continental-scale dataset. This enables us to target spatial analysis, which is an important step in our process.

2.3.3. Data Analysis

Four analytical approaches have been adopted:

- **Spatial Distribution Analysis:**

  We visualized the spatial patterns of GWR, precipitation, and land surface variables to gain an idea of how each variable is distributed and how they are spatially associated across the continent.

- **Correlation Analysis:**

  We examined the relationships between land surface characteristics and both Evapotranspiration ratio (ET_R) and Runoff ratio (RO_R) for each region using the Pearson correlation coefficient (Equation 7) (Asuero et al., 2006).

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad \textit{Equation 7}$$

Where:

- x_i, y_i = paired values of two variables at ith grid cell,
- $\bar{x}, \bar{y}$ = mean values of x and y,
- n = number of obvervation.

The coefficient fluctuates within the range of -1 to +1, where values proximate to +1 denote a robust positive linear relationship, values proximate to -1 denote a robust negative linear relationship, and values proximate to 0 denote an absence of significant linear correlation.

❖ **Principal Component Analysis:**

Principal Component Analysis (PCA) is a dimensionality-reduction technique that transforms a set of correlated predictors into a smaller set of uncorrelated variables known as principal components (PCs) (Hasan & Abdulazeez, 2021). The method identifies dominant gradients in multivariate data by decomposing the covariance matrix into eigenvalues and eigenvectors (Equation 8):

$$Z = XW \qquad \textit{Equation 8}$$

Where:

- X = standardized data matrix of predictors (eig, Soil depth, slope, LAI)
- W = matrix of eigenvectors (principal component loadings)
- Z = transformed data (principal component scores)

The eigenvalues quantify the variance explained by each PC, while the eigenvectors define directions of maximum variability in the data. Typically, the first few PCs capture the majority of total variance, highlighting the most influential gradients (Jollife & Cadima, 2016).In this study, PCA helps summarize land-surface characteristics data into a smaller number of interpretable gradients.

❖ Machine Learning Models

Two ensemble machine learning (ML) models were used to quantify the predictive influence of land-surface characteristics on groundwater recharge. They are Random Forest (RF) and Gradient Boosting (GB). Both rely on decision trees but differ in how they aggregate information.

- **Random Forest (RF)**

RF constructs an ensemble of decision trees using bootstrap sampling (bagging) and random feature selection at each node. Each tree predicts independently, and the overall prediction is the average of all trees. A single decision tree can be expressed as Equation 9 (Anand et al., 2025):

$$f(x) = \sum_{j=1}^{T} w_j \cdot h_j(x) \qquad \textit{Equation 9}$$

Where:

- $f(x)$ = the predicted value for an input vector x,
- T = the total number of trees,
- w_j = the weight associated with tree j,
- $h_j(x)$ = the prediction from tree j.

By aggregating across many randomized trees, RF reduces variance and is robust to overfitting. Its strength lies in handling nonlinear relationships and complex feature interactions, which are common in hydrological and environmental data.

- **Gradient Boosting (GB)**

GB builds trees sequentially, where each new tree corrects the errors of the previous model (Equation 10). GB optimizes performance step-by-step (Zhang & Haghani, 2015).

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \qquad \textit{Equation 10}$$

Where:

- $F_{m-1}(x)$ = the model from the previous step,
- h_m(x) = the weak learner,
- γ_m = the learning rate controlling the contribution of each tree.

- **Performance Evaluation Parameters**

Performance was evaluated using Root Mean Square Error (RMSE) and Coefficient of Determination (R^2) (Xing et al., 2019):

RMSE determines the average magnitude of errors between the predicted and the experimental value (Equation 11). It is expressed in the same units as the target variable, making it simple to interpret in practical terms. A lower RMSE suggests a higher degree of accuracy, while a higher RMSE indicates weaker model performance.

R^2 evaluates the proportion of variance in the experimental data that is explained by the model (Equation 12). It ranges from 0 to 1. Values closer to 1 represent a better fit, while values closer to 0 indicate a lower fit.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y_i} - y_i)^2} \qquad \textit{Equation 11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y_i} - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad \textit{Equation 12}$$

Where:

- n = the number of samples used,
- y_i = the experimental value at the i^{th} sample,
- $\hat{y_i}$ = the predicted value of the i^{th} sample,
- $\bar{y}$ = the average value of the entire sample.

These metrics assess prediction accuracy (RMSE) and explanatory power (R^2) of the models in representing groundwater recharge variability.

- **SHAP Analysis**

SHAP (Shapley Additive exPlanations) is grounded in cooperative game theory and assigns each feature a contribution value toward prediction (Lundberg & Lee, 2017).

$$f(x) = \emptyset_0 + \sum_{i=!}^{M} \emptyset_i \qquad \textit{Equation 13}$$

Where:

- $f(x)$ = model prediction,
- $\emptyset_0$ = mean prediction across all samples,
- $\emptyset_i$ = contribution of feature I,
- M = number of features.

PARTIAL CONCLUSION

The combination of CLM datasets, advanced geospatial tools, and machine learning techniques provides a comprehensive framework for assessing groundwater recharge drivers across Africa. By integrating both statistical and AI-based approaches, this methodology captures spatial variability while enhancing interpretability, offering a strong foundation for the results presented in the next chapter.

CHAPTER 3: RESULTS AND DISCUSSION

CHAPTER 3: RESULTS AND DISCUSSION

This chapter presents the results of the statistical and machine learning analyses performed to investigate the spatial variability of groundwater recharge (GWR) and its controlling factors across Africa. By combining correlation analysis, principal component analysis (PCA), and SHAP-based machine learning interpretation, the chapter explores how land-surface characteristics impact GWR, and how these interactions vary across regions. The implications of these findings for sustainable water supply in renewable energy projects are also discussed.

3.1. Spatial Variability of GWR, Precipitation, and Land-Surface Characteristics

The first stage of the analysis aimed to understand the spatial distribution of groundwater recharge and its correlation with land surface characteristics across Africa. Figure 3.1 shows the spatial distribution of GWR, Precipitation, and land surface characteristics over Africa.

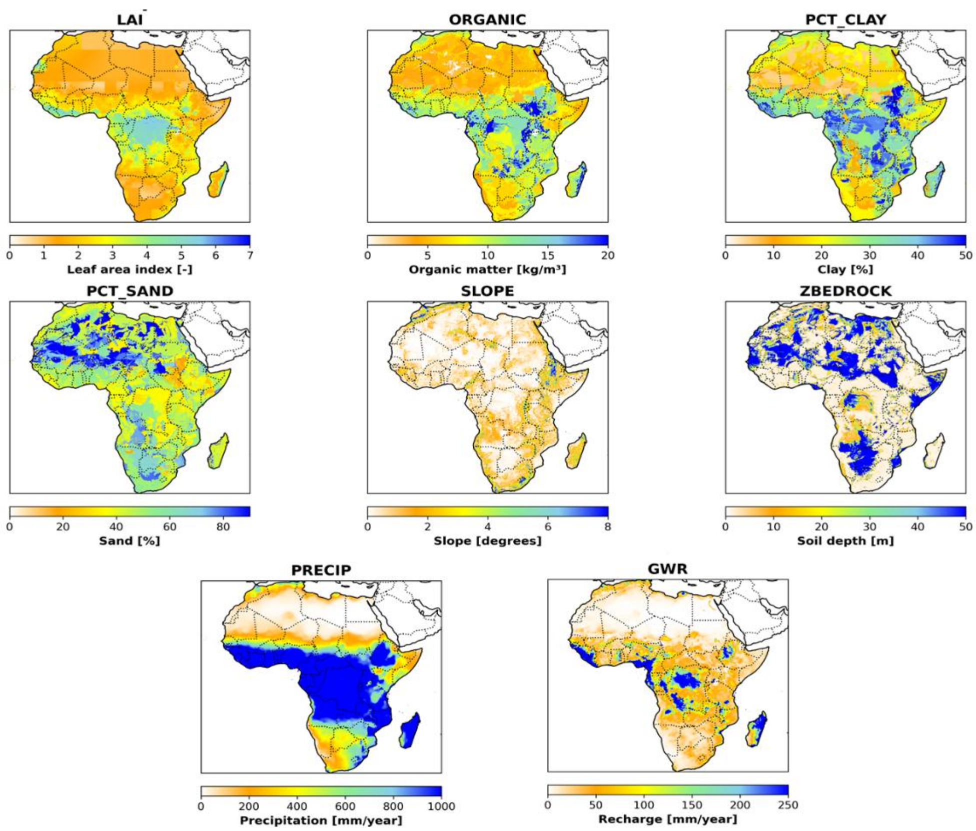

Figure 3.1. Spatial distribution of GWR, Precipitation, and land surface characteristics over Africa.

The spatial distribution shows that recharge is impacted by not only precipitation but also land surface characteristics. These characteristics are the proportion of sand (PCT_SAND) and clay (PCT_CLAY), vegetation density, represented by Leaf Area index (LAI), organic matter (ORGANIC), slope (SLOP), and soil thickness (ZBEDROCK). The interactions between these factors are non-linear and dependent on each region. There is then a need for a regional approach for the analysis.

3.2. Regional Correlation Analysis

We examined the correlation between each land-surface characteristic and precipitation partitioning (Evapotranspiration Ratio and the Runoff Ratio) to investigate the influence of individual land-surface characteristics on groundwater recharge.

3.2.1. Factors Controlling Evapotranspiration Ratio (ET_R)

Several consistent patterns are evident in the relationships between land-surface characteristics and the evapotranspiration ratio (Figure 3.2). Soil depth (ZBEDROCK) shows a positive association with ET_R across most regions, for example, in Central Africa (0.37), Southwest Africa (0.60), and Southeast Africa (0.55). This suggests that soil depth has a significant and general impact on the proportion of rainfall that is lost through evapotranspiration. Vegetation cover shows negative correlations with ET_R. In Central Africa (-0.35), Southeast Africa (-0.31), and West Africa (-0.17), higher vegetation density is associated with lower evaporative fractions. This likely reflects the ability of vegetated areas to retain soil moisture. Soil texture shows region-specific impact with ET_R. Clay fractions are negatively related to ET_R in CEAF, SWAF, and SEAF, with correlation values ranging from -0.23 to -0.38. This pattern indicates that finer-textured soils are generally associated with lower evaporative losses. Conversely, sandy soils show weak to moderate positive associations in SWAF and SEAF (+0.28 to +0.29). These results suggest that sandy textures may facilitate enhanced infiltration and deeper soil-water availability, which can lead to sustained evapotranspiration. Terrain slope is negatively correlated with ET_R in most regions, such as SWAF (-0.46) and SEAF (-0.34). Steeper terrain, therefore, tends to coincide with lower evaporative fractions, likely due to faster redistribution of precipitation as surface flow, which reduces the residence time of water available for evapotranspiration. Organic matter also shows low to moderate negative correlation across the regions, MED (0.00), SAH (0.01), CEAF (-0.30). This suggests that organic matter has less influence on ER_R.

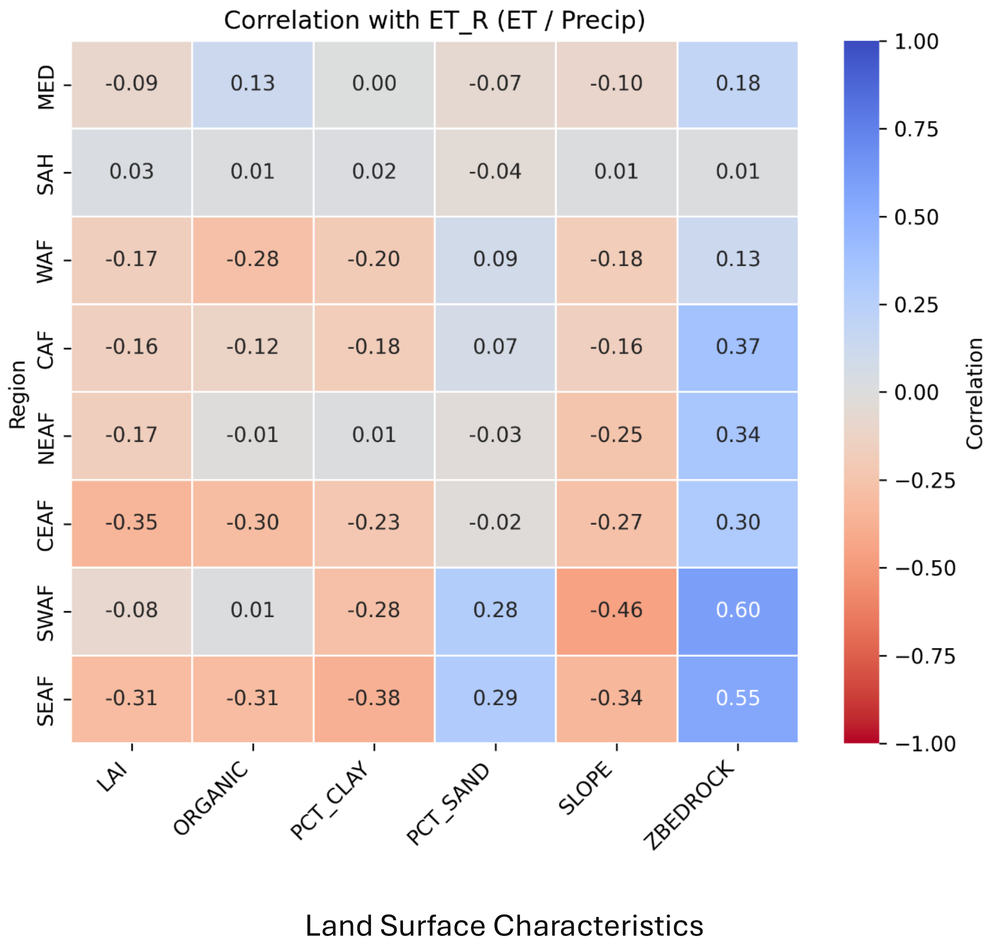

Figure 3.2. Correlation between land surface characteristics and ET_R in eight African regions

3.2.2. Factors Controlling Runoff Ratio (RO_R)

Figure 3.3 highlights the relationships between land-surface characteristics and the fraction of rainfall that is converted into surface runoff. Among these, slope emerges as the most pronounced control, with positive correlations observed across most regions, for example, in Northeast Africa (+0.50), Southwest Africa (+0.61), and Southeast Africa (+0.40). These results indicate that steeper terrain systematically coincides with areas where a larger share of precipitation is transformed into runoff. Soil depth also plays a critical role, with shallow soils (low ZBEDROCK values) strongly associated with higher runoff ratios. This is evidenced by robust negative correlations between soil depth and RO_R in several regions, such as NEAF (-0.77), West Africa (-0.72), and Central Africa (-0.74). Soil texture exerts region-specific influences on RO_R. In WAF, CAF, and SEAF, higher clay content corresponds with elevated runoff ratios (0.26 to 0.43), consistent with the lower permeability of finer-textured soils that restrict infiltration. By contrast, sand tends to display negative correlations in most regions, indicating that coarser textures, which promote infiltration, are associated with proportionally lower runoff fractions. Vegetation and organic matter also show positive correlations with RO_R in certain regions, most notably in WAF where LAI (+0.43) and organic matter (+0.38) coincide with higher runoff fractions.

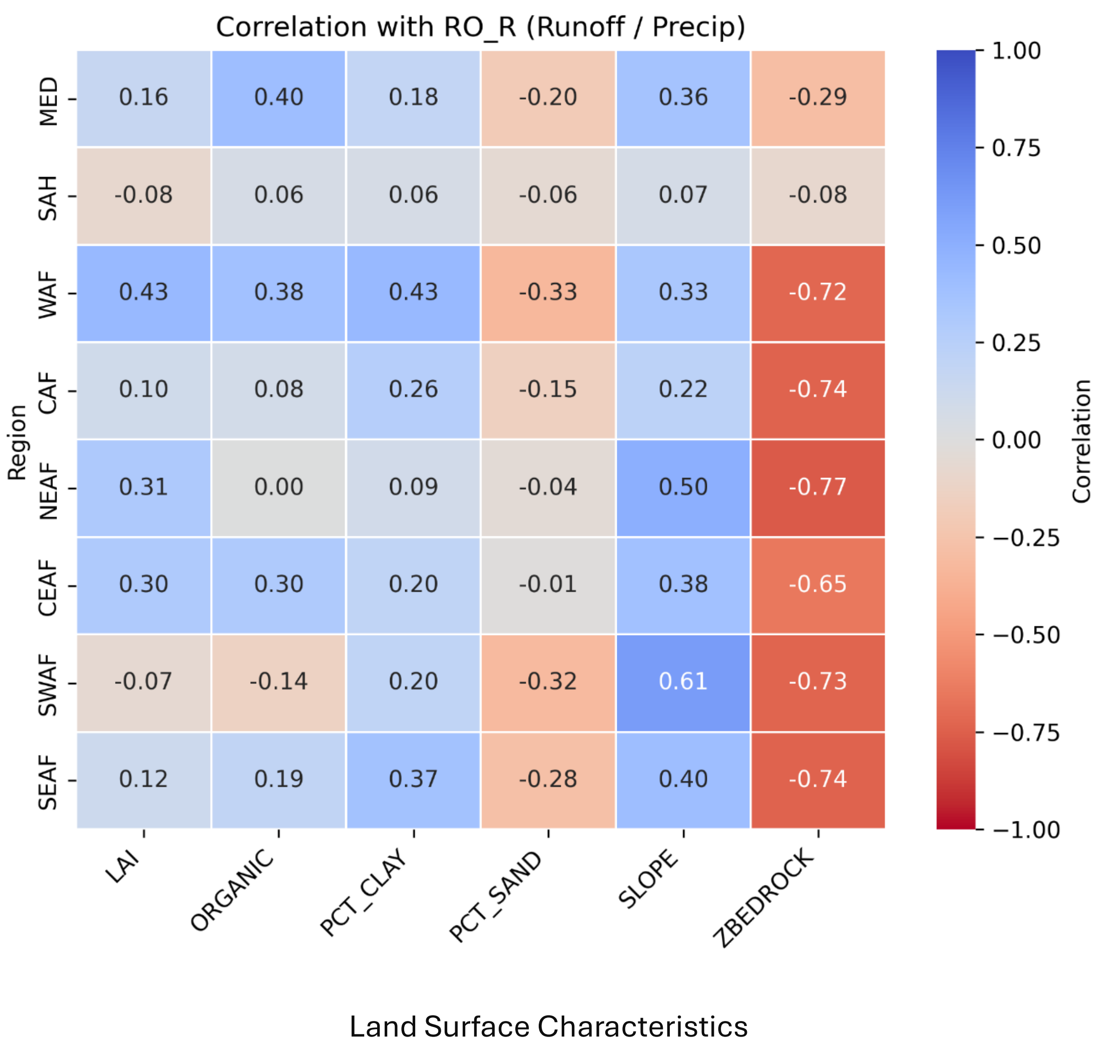

Figure 3.3. Correlation between land surface characteristics and RO_R in eight African regions

3.3. Principal Component Analysis: Dominant Gradients Influencing Recharge Partitioning

❖ **Mediterranean (MED) region**

In the Mediterranean region (Figure 3.4), the first principal component (PC1) is dominated by organic matter (0.50), clay content (0.47), and is negatively associated with sand (-0.42) and soil depth (-0.29). Given this grouping, PC1 is interpreted as a Soil Texture Gradient.

The second component (PC2) is positively dominated by soil depth (0.58) and clay content (0.41), while sand (-0.47) and slope (-0.39) have a more negative association. As such, PC2 is interpreted as a Terrain-Drainage gradient.

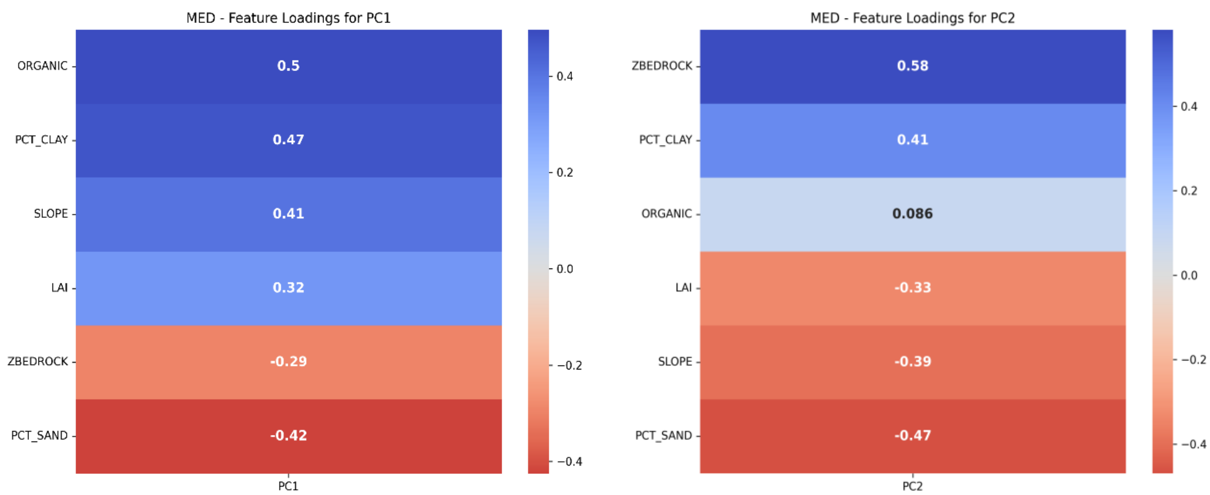

Figure 3.4. PCA feature loadings for the Mediterranean region, showing PC1 as a Soil Texture Gradient and PC2 as a Terrain-Drainage Gradient

❖ **Sahara (SAH) region**

In the Sahara (Figure 3.5), PC1 shows strong positive loadings for clay (0.60) and organic matter (0.54), and a strong negative loading for sand (-0.57). This defines a Soil Texture Gradient.

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for soil depth (0.70) and negative loadings for slope (-0.69).

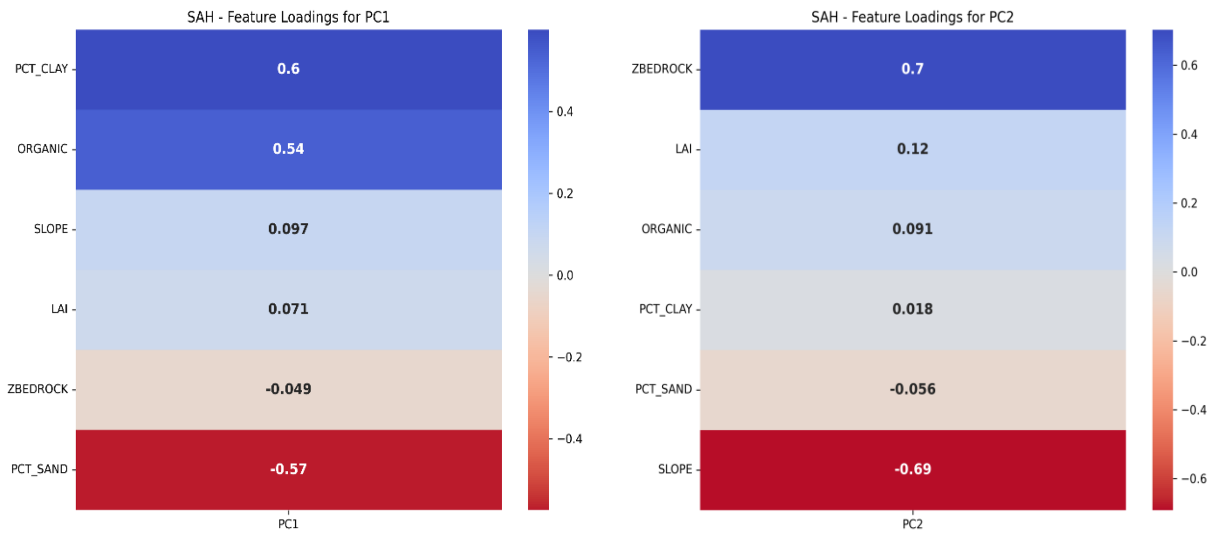

Figure 3.5. PCA feature loadings for the Sahara region, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

❖ West Africa (WAF)

In West Africa (Figure 3.6), PC1 captures a clear Soil Texture Gradient, with strong positive loadings for clay (0.52), organic matter (0.48), and vegetation (LAI: 0.40), contrasted against a strong negative loading for sand (-0.50).

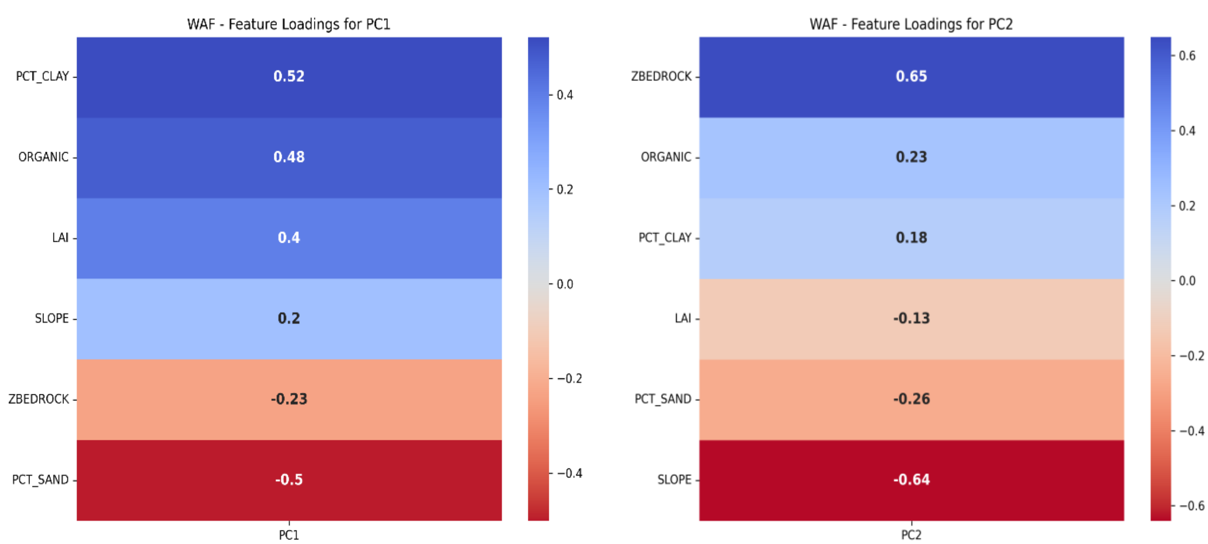

Figure 3.6. PCA feature loadings for the West Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for soil depth (0.65) and negative loadings for slope (-0.64).

❖ **Central Africa (CAF)**

For Central Africa (Figure 3.7), PC1 represents a Soil Texture Gradient, with strong positive loadings for clay content (0.57), organic matter (0.49), and vegetation (LAI: 0.34), opposed by a strong negative loading for sand (-0.55).

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for soil depth (0.75) and negative loadings for slope (-0.62).

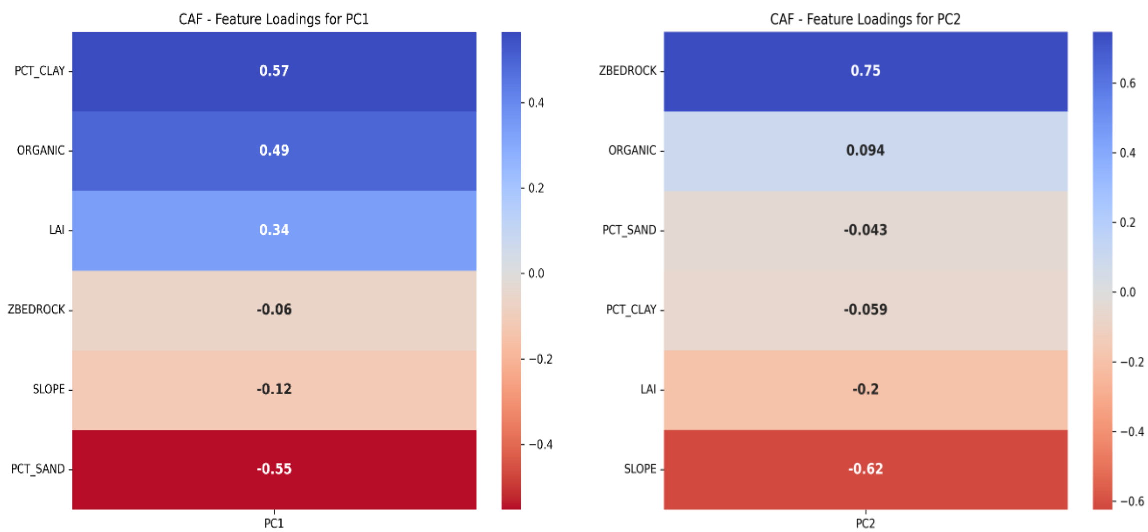

Figure 3.7. PCA feature loadings for the Central Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

❖ **Northeast Africa (NEAF)**

In Northeast Africa (Figure 3.8), PC1 aligns with a Soil Texture Gradient, with strong positive contributions from clay (0.59), organic matter (0.52), and vegetation (LAI: 0.33), opposed by a strong negative loading from sand (-0.51).

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for soil depth (0.70) and negative loadings for slope (-0.64).

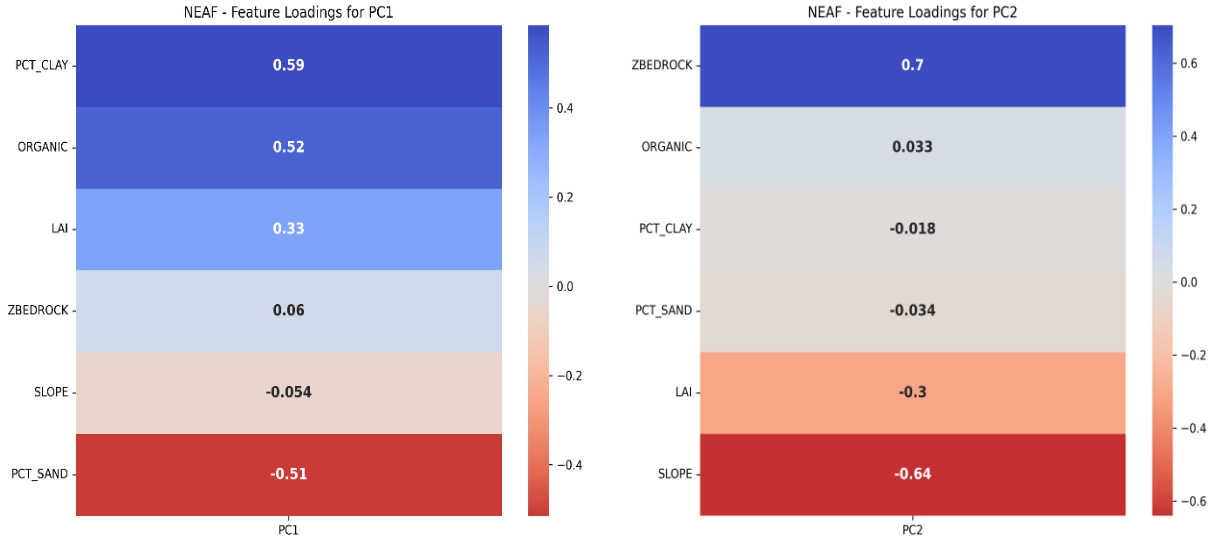

Figure 3.8. PCA feature loadings for the Northeast Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

❖ Central East Africa (CEAF)

In CEAF (Figure 3.9), PC1 reflects a Soil Texture Gradient, dominated by high positive loadings for clay (0.54), organic matter (0.49), and vegetation (LAI: 0.37), opposed by a strong negative loading for sand (-0.48).

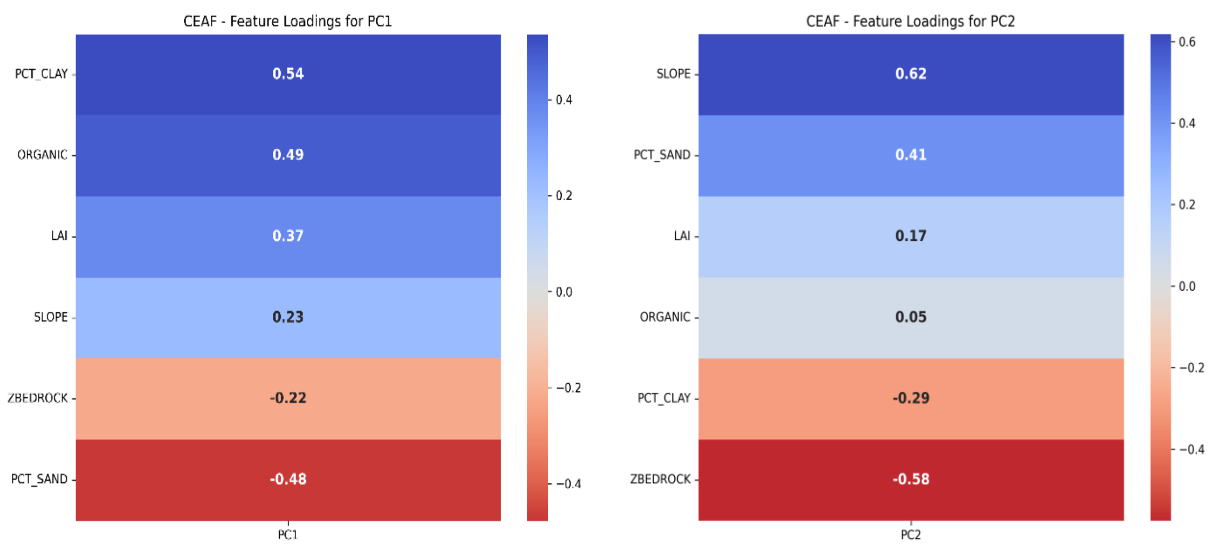

Figure 3.9. PCA feature loadings for the Central East Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for slope (0.62) and negative loadings for soil depth (-0.58).

- **Southwest Africa (SWAF)**

In SWAF (Figure 3.10), PC1 is strongly aligned with a Soil Texture Gradient, with high positive loadings for clay (0.59), organic matter (0.42), and LAI (0.27), contrasted with strong negative loading for sand (-0.54).

PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loading for slope (0.63), opposed by strong negative contributions for soil depth (-0.63) and organic matter (-0.40).

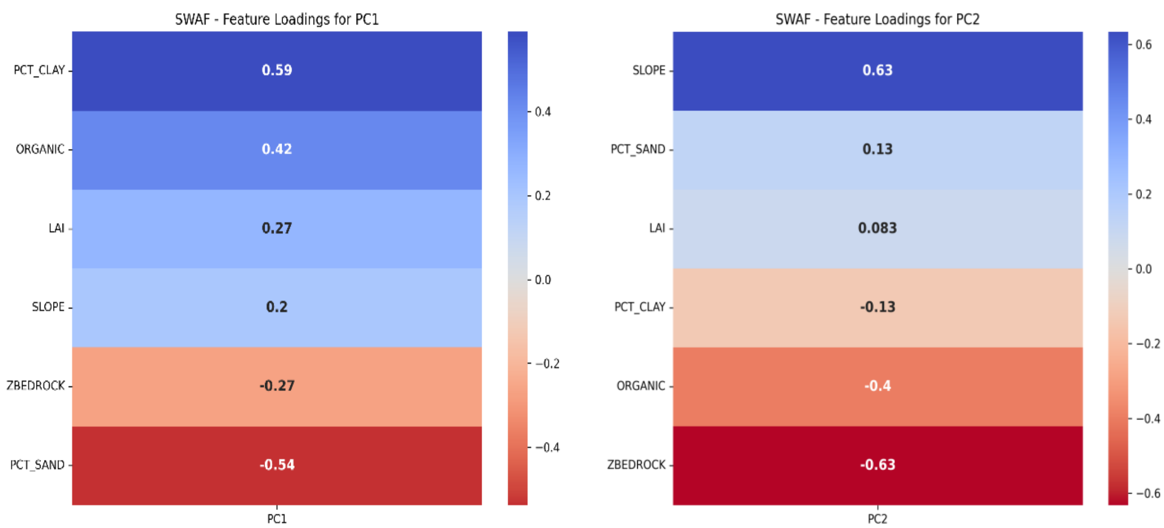

Figure 3.10. PCA feature loadings for the Southwest Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

- **Southeast Africa (SEAF)**

In SEAF (Figure 3.11), PC1 reflects a Soil Texture Gradient, with strong positive loadings for clay (0.55), organic matter (0.45), and LAI (0.27), opposed by strong negative loadings for sand (-0.54). PC2 represents an Infiltration versus Drainage Gradient, dominated by positive loadings for slope (0.71). and negative loadings for soil depth (-0.58).

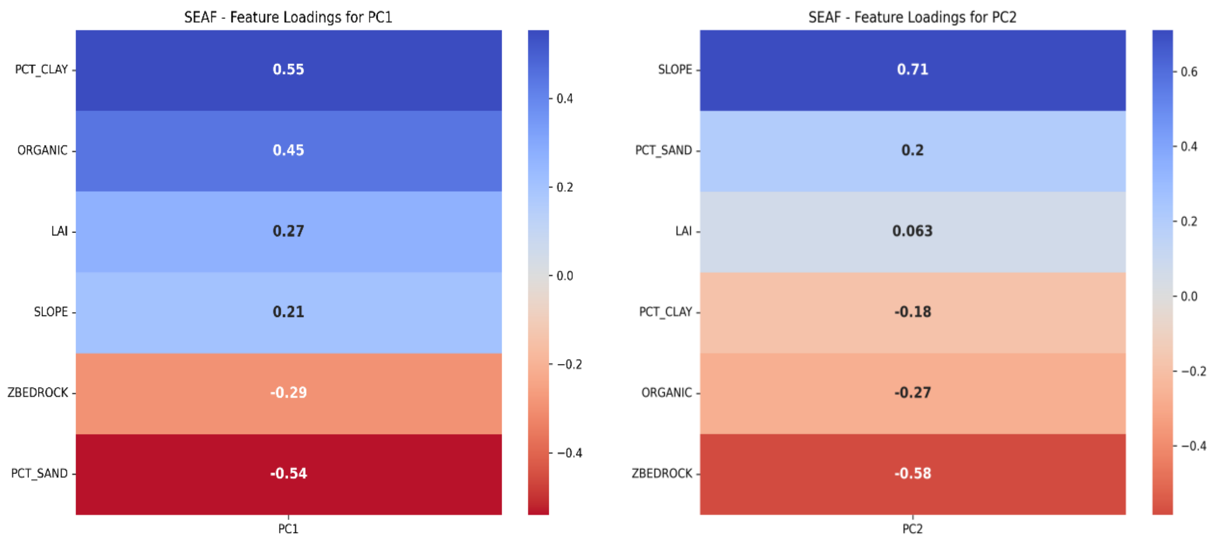

Figure 3.11. PCA feature loadings for the Southeast Africa, showing PC1 as a Soil Texture Gradient and PC2 as an Infiltration versus Drainage Gradient

❖ **Summary of PCA Findings**

Across the eight African regions, PCA consistently revealed two dominant gradients: (i) a Soil Texture Gradient (PC1), capturing the joint variability of clay, sand, organic matter, and vegetation, and (ii) an Infiltration versus Drainage Gradient (PC2), dominated by soil depth and slope. These gradients highlight that recharge efficiency is shaped by soil, vegetation and terrain interactions beyond precipitation alone. While PCA provides valuable dimension reduction and identification of covarying features, it does not quantify the magnitude of each factor's influence on recharge. To address this limitation and capture potential nonlinearities, machine learning models with SHAP interpretation were subsequently applied.

3.4. Machine-learning performances and SHAP-based interpretation

3.4.1. Model performances (Random Forest and Gradient Boosting)

The performance of machine learning models, specifically Random Forest (RF) and Gradient Boosting (GB), was evaluated using two key metrics (Table 3.1): Root Mean Squared Error (RMSE) and the coefficient of determination (R^2). Across all regions, RF models consistently outperformed GB models in both metrics. The superiority of RF was most evident in regions like

West Africa and South West Africa, where the models achieved R^2 values of 0.98 and 0.97, respectively. These high scores indicate that RF models could explain more than 97% of the observed variability in recharge. Furthermore, in low-recharge environments like the Sahara, RF produced very low RMSE values (as low as 2.33 mm/year), highlighting its robustness and predictive accuracy even in data-scarce or climatically extreme settings.

Even though Gradient Boosting models exhibited marginally diminished accuracy in comparison to Random Forest, this does not signify inadequate performance. GB still demonstrated strong predictive ability in several regions, particularly Southeast Africa (SEAF) and West Africa (WAF), achieving R^2 values above 0.90 in these areas. This difference arises from the two algorithms' distinct approaches (refer to Chapter 2).

Table 3.1. Model Performance Across Different Regions

Models / Regions	Model Performances			
	Random Forest (RF)		**Gradient Boosting (GB)**	
	R^2	**RMSE**	**R^2**	**RMSE**
MED	0.92	10.84	0.88	13.36
SAH	0.95	2.33	0.79	4.84
WAF	0.98	32.24	0.95	44.05
CAF	0.95	26.16	0.91	35.27
NEAF	0.90	16.82	0.81	23.47
CEAF	0.88	19.94	0.80	25.87
SWAF	0.97	9.44	0.93	13.69
SEAF	0.95	30.49	0.94	36.32

3.4.2. SHAP Feature Importance Analysis

- **Mediterranean Region**

In the Mediterranean region, SHAP analysis confirms that, in both the Random Forest (RF) and Gradient Boosting (GB) models (Figure 3.12), precipitation (PRECIP) is by far the most influential predictor of groundwater recharge. In the RF analysis, precipitation has an average SHAP of around 30 mm per year. This suggests that high precipitation significantly increases predicted recharge.

Soil depth is shown as the most important land surface feature. In both models, deeper soils are associated with positive SHAP values, while shallow soils produce negative SHAP values of up to -25 mm/year, particularly in the RF model. This suggests that, in the Mediterranean region, shallow soils are associated with lower recharge predictions. This is likely due to their connection with higher runoff ratios and limited infiltration. Other variables play moderate roles. Slope produces mostly negative SHAP values at higher angles. Organic matter is associated with slight downward shifts in predicted recharge, which potentially reflects locations where retained soil moisture is lost through evapotranspiration or runoff. Texture variables (clay and sand fractions) and LAI (vegetation) show relatively small effects (average SHAP contributions < 5 mm/year) and do not consistently shift recharge predictions in one direction, suggesting that in the Mediterranean climate these variables alone do not strongly regulate the predicted recharge fraction.

Figure 3.12. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in the Mediterranean region, using RF (top) and GB (bottom) models.

❖ **Sahara Region**

In the Sahara region (Figure 3.13), SHAP analysis again identifies precipitation as the primary predictor of groundwater recharge, though with considerably lower average SHAP magnitudes compared to the Mediterranean region (~6 mm/year for Random Forest, and ~3.5 mm/year for Gradient Boosting). This reflects the overall limited recharge potential under arid climatic conditions. Among land surface variables, clay content (PCT_CLAY) shows the strongest modelled influence. Higher clay fractions consistently shift predictions downwards, indicating that locations with finer-textured soils tend to be associated with reduced recharge predictions, most likely because such soils limit infiltration and promote localized surface runoff. Sand content (PCT_SAND) has a weaker and mixed influence, sometimes shifting predictions upward at high sand values in the RF model, suggesting that locally coarser soils may correspond to slightly higher infiltration potential. Other features, including LAI, soil depth (ZBEDROCK), organic matter, and slope produce only small SHAP shifts (mostly <1 mm/year on average) and remain centred near zero. This implies that, within the Sahara, recharge is almost entirely controlled by the scarcity and variability of precipitation itself, with land-surface differences playing only a minor role in the fraction of rainfall that becomes recharge.

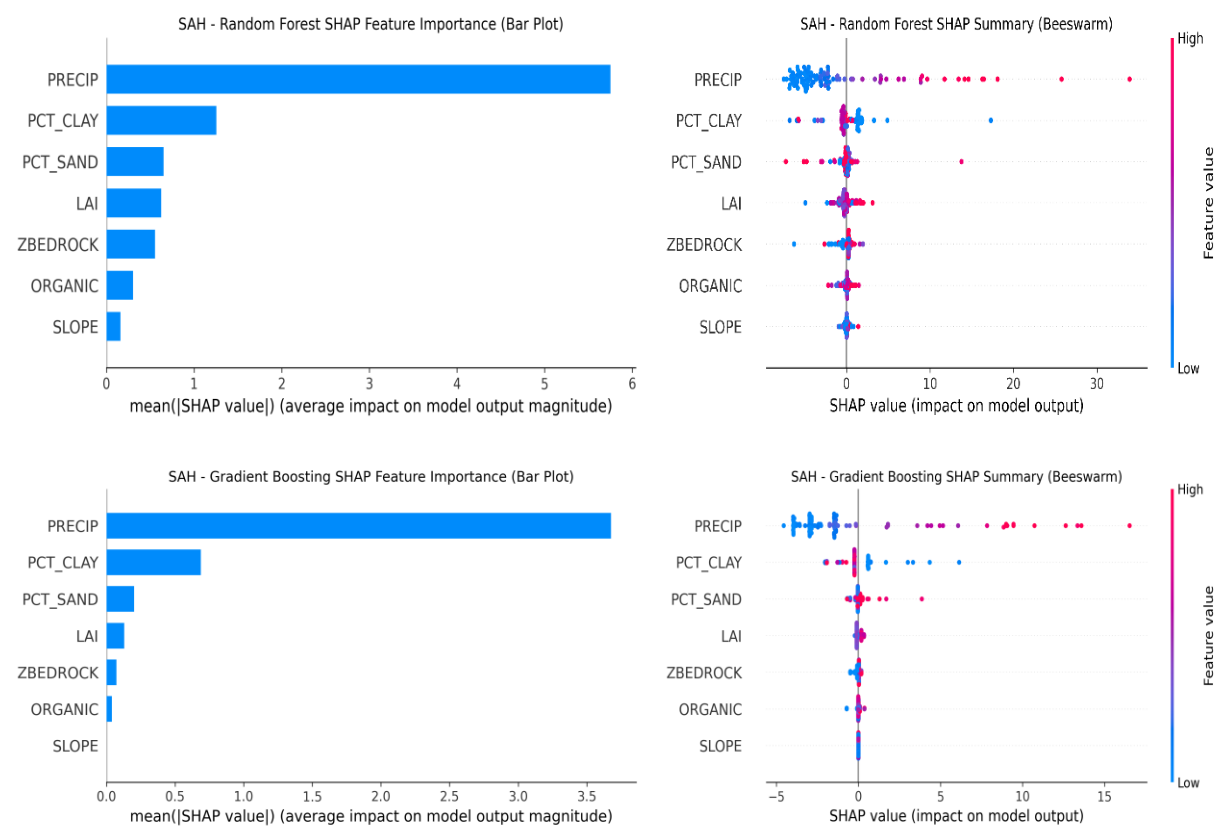

Figure 3.13. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in the Sahara region, using RF (top) and GB (bottom) models.

❖ West Africa Regions

In West Africa, SHAP analysis reveals a particularly strong dependence of recharge predictions on precipitation, with mean SHAP magnitudes exceeding 120 mm/year in the Random Forest model and 70 mm/year in the Gradient Boosting model (Figure 3.14). This emphasises that recharge potential in WAF is highly responsive to rainfall anomalies. The most influential land-surface variable in both models is soil depth (ZBEDROCK). Deeper soils are associated with large positive SHAP shifts (up to +100 mm/year in extreme RF cases), whereas shallow soils can shift predicted recharge downwards by approximately -100 mm/year. These strong model responses indicate that, within West Africa, variations in soil thickness align closely with differences in infiltration opportunity and storage capacity. Vegetation (LAI) and slope have secondary influences. High LAI tends to depress predicted recharge due to its association with stronger evapotranspiration demand. Soil texture shows less influence on recharge prediction.

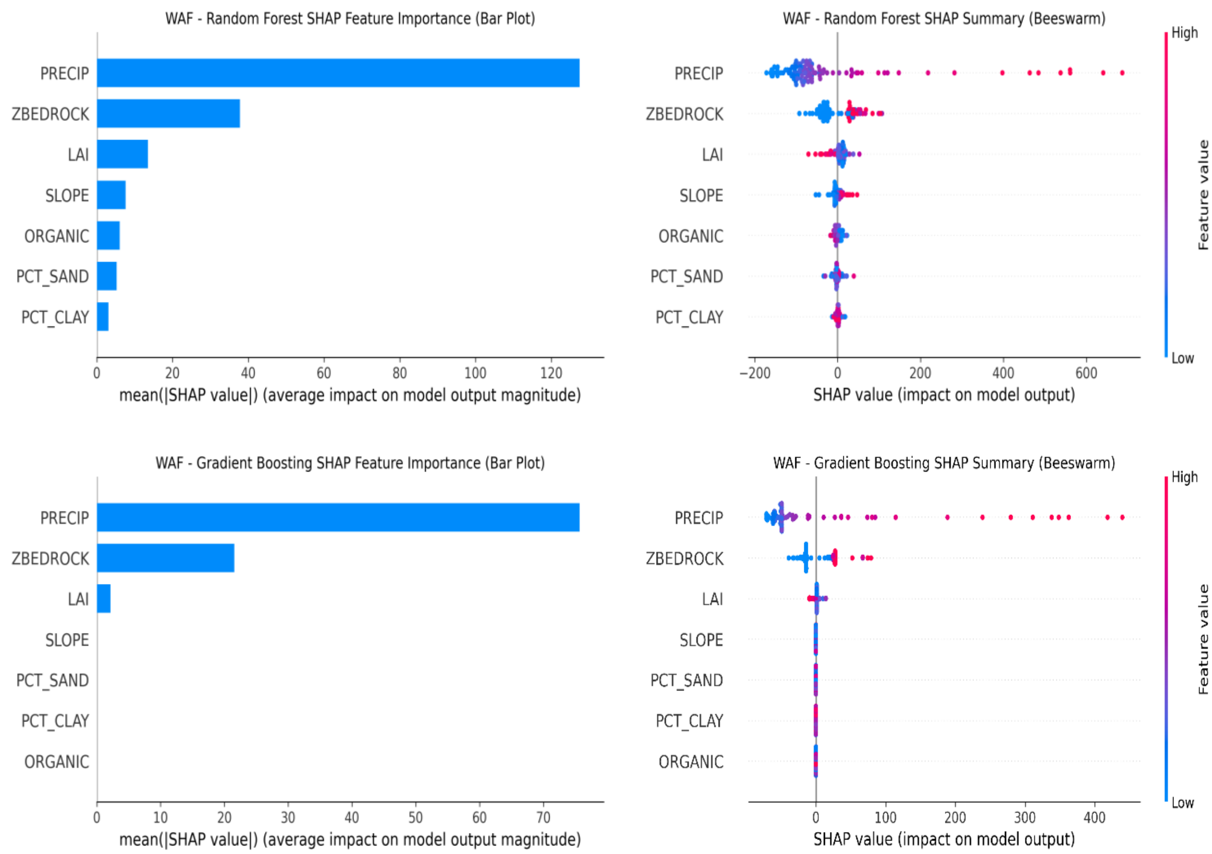

Figure 3.14. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in West Africa, using RF (top) and GB (bottom) models.

❖ Central Africa Region

In Central Africa, SHAP analysis reveals that the recharge regime is strongly influenced by precipitation. Average SHAP magnitudes exceed 70 mm/year in Random Forest models and are approximately 35 mm/year under Gradient Boosting (Figure 3.15). These high values reflect the variability in rainfall and the strong hydrological response typical of humid tropical zones. In Central Africa, SHAP analysis reveals that the recharge regime is strongly influenced by precipitation. Average SHAP magnitudes exceed 70 mm/year in Random Forest models and are approximately 35 mm/year under Gradient Boosting (see Figure 3.15). These high values reflect the variability in rainfall and the strong hydrological response typical of humid tropical zones. Among land surface features, soil depth is the dominant modulator of predicted recharge. Deeper soils shift the model outputs upward (up to 150 mm/year in RF), while shallow soils show negative

SHAP shifts (approximately -100 mm/year). This suggests that infiltration opportunity is a major determinant of recharge efficiency in CAF. Vegetation presents a modest downward influence on recharge (mean SHAP < 20 mm/year). Similarly, organic matter content, clay, and slope have minor impacts compared to precipitation and soil thickness.

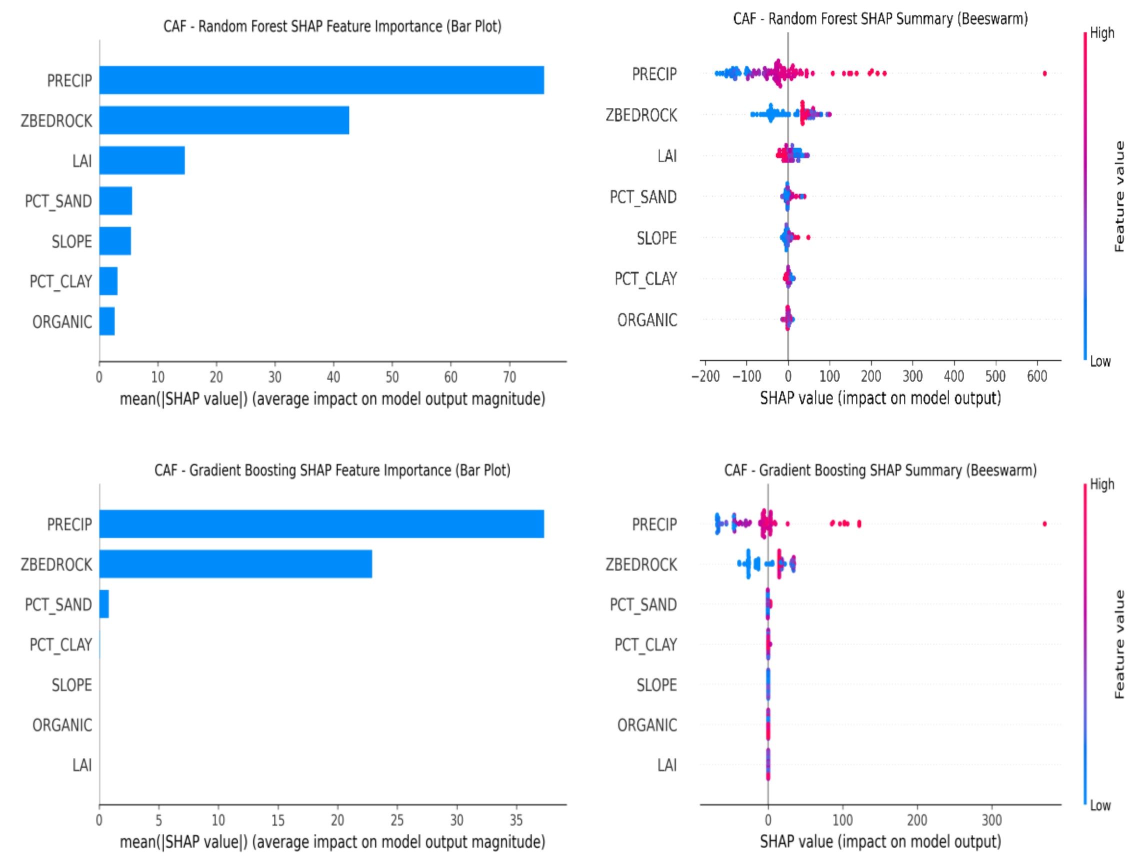

Figure 3.15. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in Central Africa, using RF (top) and GB (bottom) models.

❖ Northeast Africa Region

In Northeast Africa, precipitation remains the most influential predictor of groundwater recharge in both Random Forest and Gradient Boosting models, with mean SHAP values of around 35 mm/year and 18 mm/year, respectively (Figure 3.16). This again highlights that recharge potential in NEAF is primarily governed by rainfall variability. Interestingly, compared to other regions, vegetation (LAI) emerges as the dominant non-precipitation modulator, producing average SHAP magnitudes of approximately 7 mm/year and shifting model predictions downward in the RF model. Higher LAI values tend to coincide with negative SHAP values, indicating that locations with denser vegetation are associated with reduced recharge predictions, most likely due to elevated evapotranspiration losses. Soil depth (ZBEDROCK) plays a secondary but positive role (mean SHAP ~6 mm/year in RF), suggesting that deeper soils are associated with a greater portion of rainfall contributing to recharge. Organic matter and soil texture (PCT_CLAY and PCT_SAND) also show moderate influence (mean SHAP ~2-5 mm/year), with higher organic and clay content generally decreasing recharge predictions. Slope has relatively small average contributions (< 2 mm/year).

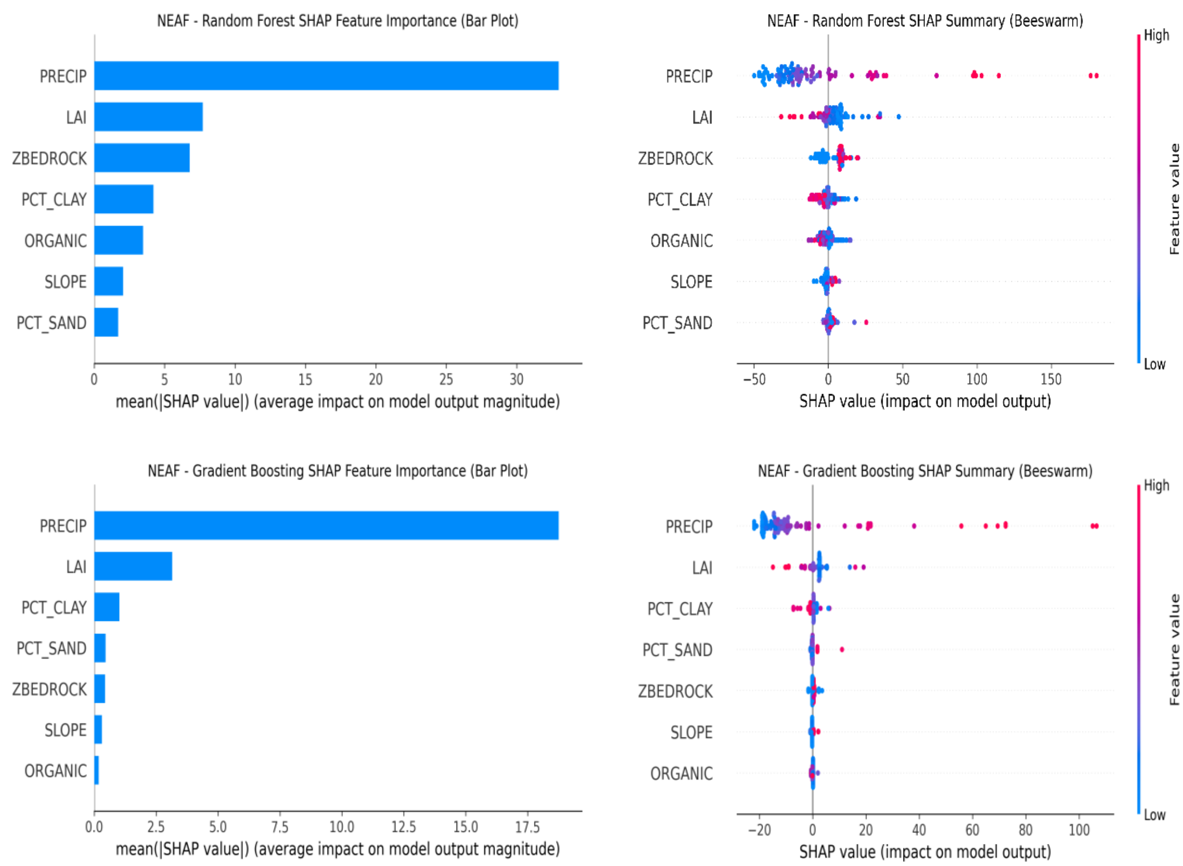

Figure 3.16. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in Northeast Africa, using RF (top) and GB (bottom) models.

❖ **Central East Africa Region**

In Central East Africa, precipitation remains the dominant driver of recharge predictions, with average SHAP values of approximately 30 mm/year under the Random Forest model and 15 mm/year using Gradient Boosting (Figure 3.17). This reflects a high dependence of recharge on rainfall variability in this moderately wet region. Soil depth (ZBEDROCK) is the largest non-climatic modulator, with an average SHAP impact of about 16 mm/year in RF and 6 mm/year in GB. Deeper soils are aligned with positive SHAP shifts, while shallow profiles correspond with decreases of up to -40 mm/year, consistent with runoff-limited recharge in areas of restricted infiltration capacity. Among the remaining variables, vegetation, clay content and organic matter exhibit modest SHAP magnitudes (~5 mm/year) with a general trend for higher vegetation, clay or organic content to shift predictions slightly downward. This aligns with a pattern in which water retention and evapotranspiration demand reduce recharge efficiency in CEAF. Slope and sand content show minimal influence (<5 mm/year on average), indicating that within this region, terrain and coarse-texture effects are secondary once soil depth is accounted for.

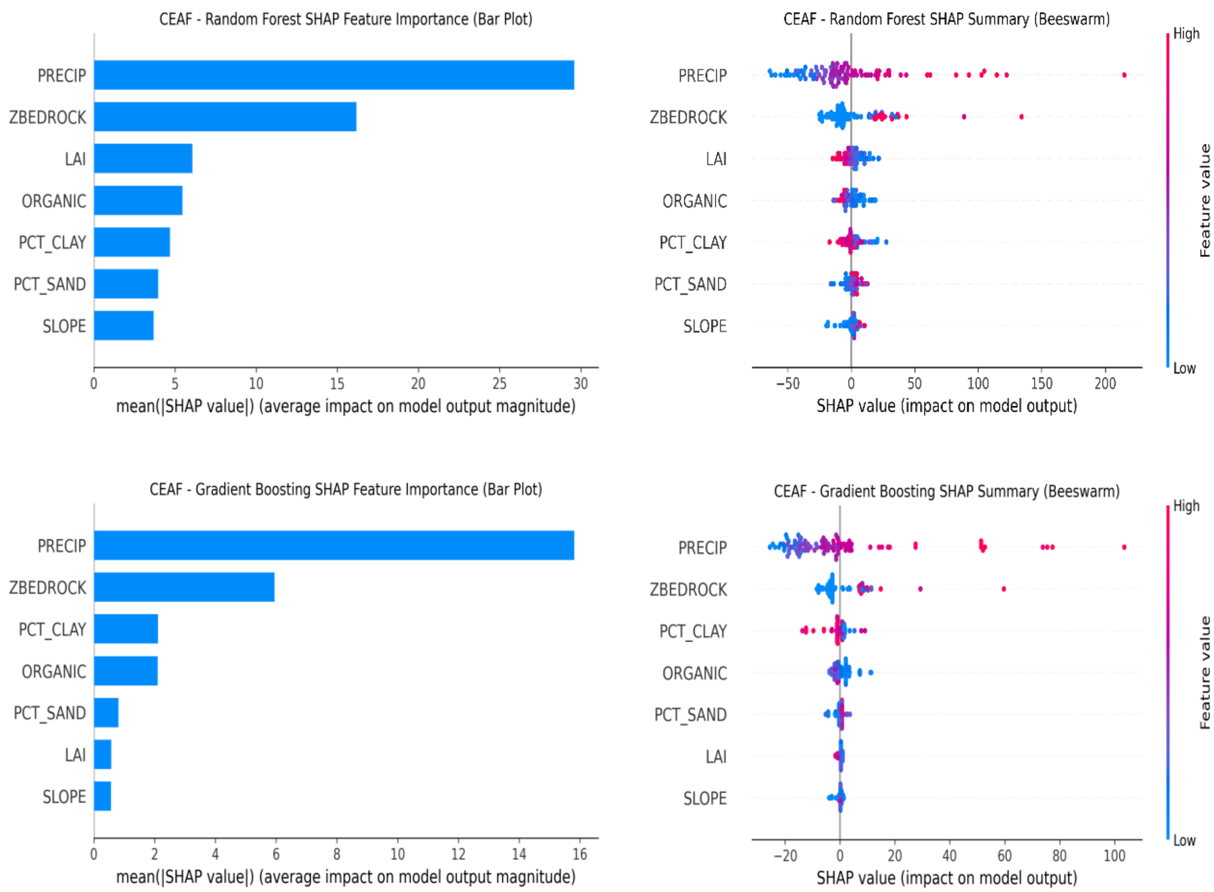

Figure 3.17. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in Central East Africa, using RF (top) and GB (bottom) models.

❖ **Southwest Africa Region**

In Southwest Africa, recharge prediction remains predominantly controlled by precipitation, which contributes the highest SHAP magnitudes in both the Random Forest (~35 mm/year) and Gradient Boosting (~20 mm/year) models (Figure 3.18). Nevertheless, vegetation (LAI) emerges as a notable land surface modulator with mean SHAP impacts of ~10 mm/year (RF) and ~5 mm/year (GB), confirming a stronger linkage between recharge variability and evapotranspiration regulation in this region compared to others. At higher LAI values, SHAP values tend to shift recharge predictions downward (sometimes by ~20 mm/year), implying that areas with dense vegetation have reduced effective recharge given the same precipitation input, consistent with elevated ET demands. Soil depth (ZBEDROCK) plays a secondary but positive role (mean SHAP ~5 mm/year in RF), suggesting that deeper soils are associated with a greater portion of rainfall contributing to recharge. Conversely, slope, texture parameters (PCT_SAND, PCT_CLAY), and organic matter have minimal average SHAP contributions (< 3 mm/year), indicating that runoff-related controls are less dominant in SWAF.

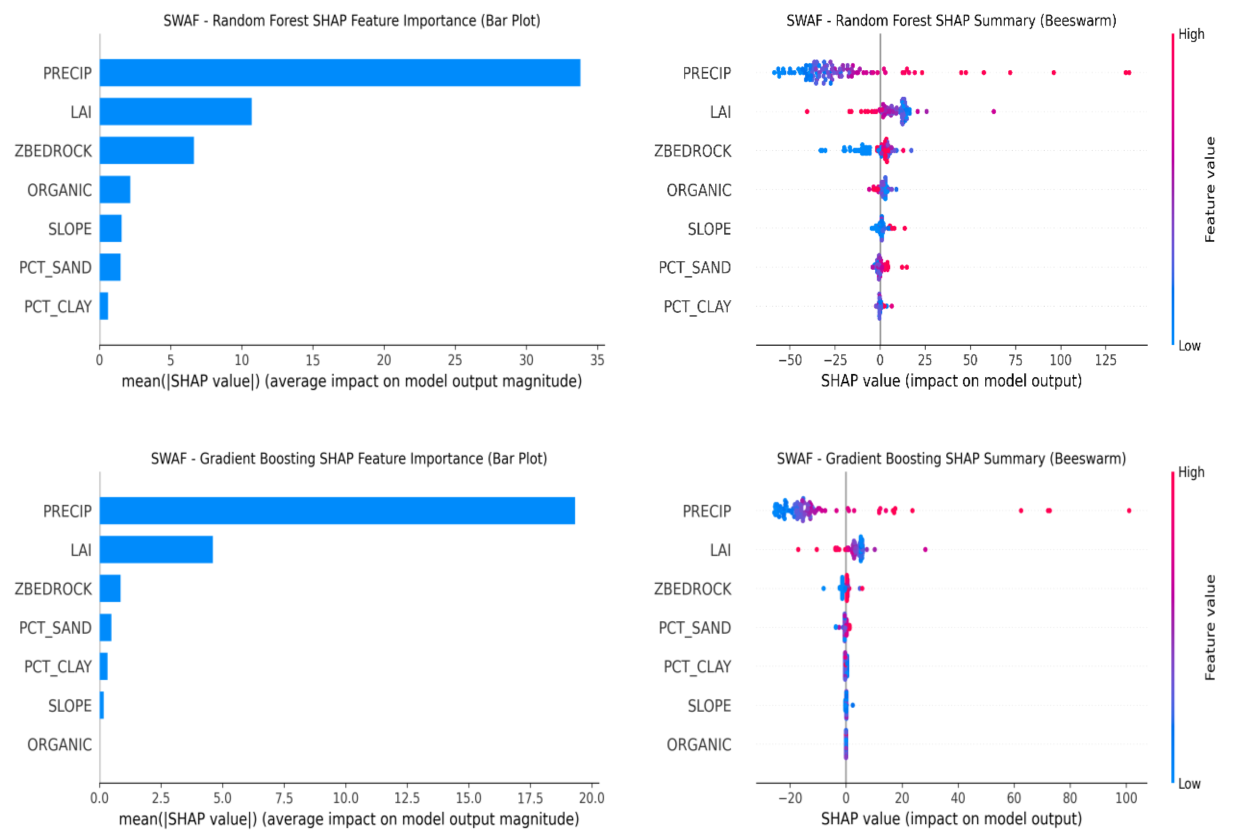

Figure 3.18. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in Southwest Africa, using RF (top) and GB (bottom) models.

❖ Southeast Africa Region

In Southeast Africa, the SHAP results again underscore the dominance of precipitation as the primary driver of recharge variability, with mean SHAP values of approximately 70 mm/year under the Random Forest model and 40 mm/year in Gradient Boosting (Figure 3.19). Extreme rainfall events produce exceptionally high positive SHAP values (over 500 mm/year in RF), confirming the strong hydrological responsiveness of this region. Among land surface factors, soil depth (ZBEDROCK) exerts the largest average influence (approximately 20 mm/year in RF;

approximately 6 mm/year in GB). Deep soils increase recharge predictions, whereas shallow soils can reduce them by more than 50 mm/year. This aligns with a runoff-limited recharge setting, where infiltration opportunity is critical for effective recharge. Vegetation (LAI) and organic matter exhibit negative but relatively modest SHAP values (less than 8 mm/year on average), suggesting that, while evapotranspiration losses are present, they are less significant than soil depth controls. Slope and soil texture have less impact.

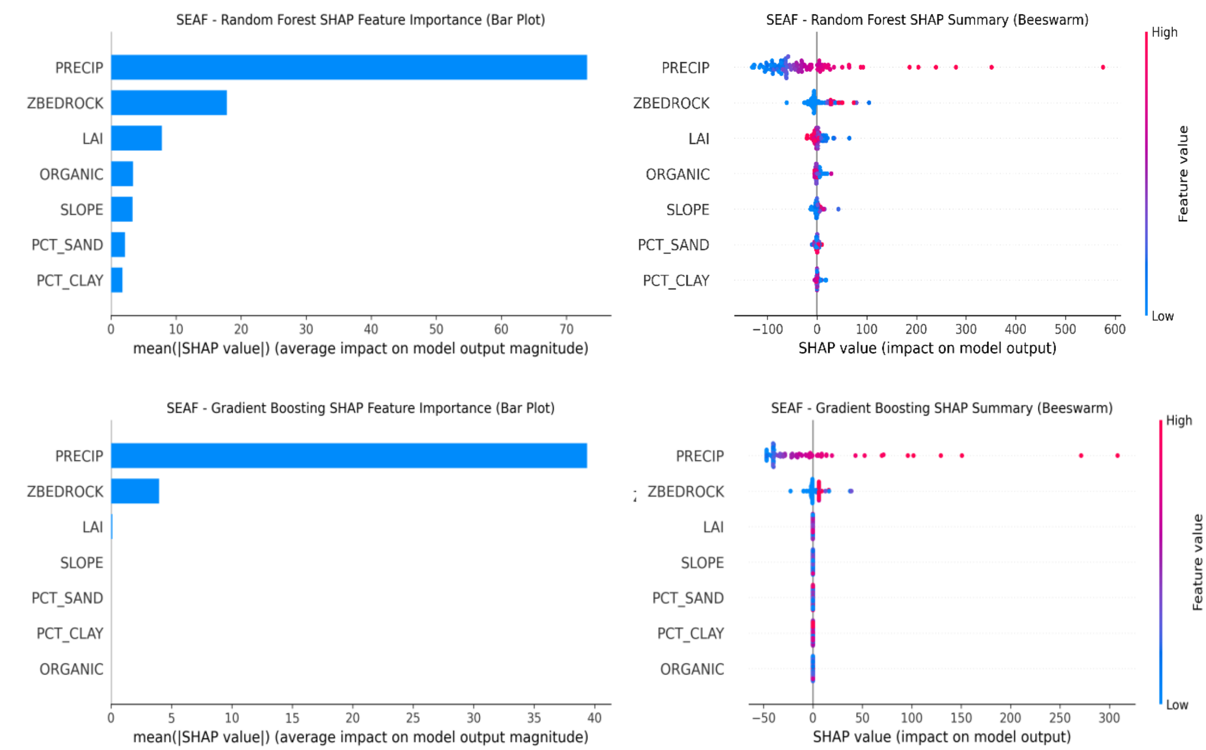

Figure 3.19. SHAP-based feature importance of precipitation and land surface factors controlling GWR variability in Southeast Africa, using RF (top) and GB (bottom) models.

3.5. Land-Surface Modulators of Recharge

The combination of correlation, principal component analysis (PCA), and machine learning interpretation provides a consistent picture of how land-surface variables interact with precipitation to influence recharge partitioning across Africa. From these statistical analyses, the regional contrasts highlight consistent mechanisms modulating recharge efficiency.

- ❖ **Precipitation as the primary driver**

Across all regions, recharge is predominantly controlled by precipitation, as evidenced by its overwhelming SHAP contributions and the strong model dependence on it. However, the amount of precipitation that contributes to recharge is strongly modulated by other variables.

- ❖ **ZBEDROCK: Soil depth as a key modulator**

Shallow soil consistently corresponds with lower recharge predictions, particularly in West Africa, Central Africa, and the Mediterranean. All analysis layers show this evidence:

- Strong negative correlations between soil depth and runoff ratio (RO_R: -0.72 to -0.77) in most of the regions.
- Soil depth contributes most in PC2, Infiltration versus Drainage Gradient, across all regions. This emphasizes its important role.
- In SHAP analysis, low soil depth shifts recharge predictions downward.

These results suggest that shallow soils reduce infiltration and promote surface runoff.

- ❖ **Slope: Influencing runoff and recharge partitioning**

Slope emerges as a secondary but consistent gradient modulator:

- Correlations show positive association with RO_R (up to 0.61 in SWAF), and negative with ET_R (-0.46 in SWAF), indicating its dual role in promoting runoff and limiting retention time for infiltration.
- Slope contributes more to PC2, Infiltration versus Drainage Gradient. This reinforces its role in limiting recharge.
- SHAP analysis also shows that high slope generally indicates downward prediction shifts, particularly in regions like SEAF and MED.

These results suggest that steeper terrain increases runoff generation. Then it decreases the recharge.

- ❖ **Vegetation (LAI): Interaction with atmospheric demand**

The role of vegetation varies by climate region:

- SHAP analysis shows that vegetation consistently reduces predicted recharge in most of the region. This suggests that dense vegetation areas decrease recharge via evapotranspiration loss.
- In drylands (e.g., SAH, MED), LAI has less influence due to a lack of vegetation.

These results suggest that recharge is limited in vegetated regions by evapotranspiration.

❖ **Organic matter and Soil texture (clay and sand): secondary influence**

Organic matter and soil texture variables have less effect across regions:

- Organic matter and clay content are positively correlated with runoff ratio in most of the regions. The PC1, Soil Texture Gradient, shows a strong association between both organic and clay content. Also, in the SHAP analysis, they lead to a downward shift in recharge predictions.
- In contrast, the relationship between sand and ET_R is mixed, which may indicate a trade-off between rapid infiltration and poor water retention. Furthermore, sand contributes to PC1 and manifests a secondary effect in SHAP analysis.

3.6. Implications for water-resource and renewable energy planning

Understanding how land-surface characteristics mediate the conversion of precipitation into groundwater recharge provides valuable insight not only for hydrological management but also for the planning and sustainability of renewable energy systems that depend on water, such as green-hydrogen production.

3.6.1. Regional suitability for recharge-dependent renewable energy deployment

Based on the integrated recharge mechanisms identified in Section 3.5, three regional response types can be identified:

Runoff-limited recharge regimes, characterized by shallow soils and steep slopes, which restrict infiltration despite relatively high precipitation. In such regions, investment in hydropower may be better supported by surface-water management and soil-depth improvement measures (e.g., terracing, infiltration trenches) that enhance recharge and regulate flows.

ET-limited recharge regimes, characterized by moderate to high rainfall but have strong evapotranspirative losses due to high vegetation and organic soil content.

Low-recharge dryland regimes (e.g., SAH, MED) are characterized by a strong dependence on rainfall. These regions are not favourable for recharge-reliant renewable energy development but may be suitable for solar or wind projects with low water demand.

3.6.2. Towards a Water-Energy Nexus perspective

The results highlight the importance of taking a Water-Energy Nexus approach in Africa. The development of renewable energy should not be decoupled from the sustainability of water resources. The variability of groundwater recharge, which is controlled by the interaction between climate and land surface characteristics, creates regionally specific risks and opportunities. Renewable systems supported by groundwater (e.g. hydrogen production) should prioritise regions with high recharge rates (e.g. WAF and CEAF) and avoid areas where steep slopes and shallow soils suppress recharge. Investment in land surface management can be considered a low-cost resilience strategy, enhancing recharge and thereby supporting the reliability of energy systems.

3.6.3. Comparison with Previous Studies

The findings of this study are broadly consistent with earlier work on African recharge dynamics. Fu et al. (2019) identified rainfall as the dominant control on recharge. This aligns with the present results, where precipitation has the highest SHAP contributions across all regions. Similarly, Toure et al. (2024) emphasized the role of slope in modulating infiltration opportunity, supporting the strong positive correlation between slope and runoff observed here. Anand et al. (2025) emphasized that the properties of the soil influence processes of infiltration and percolation, confirming the impact of soil texture (sand, clay) and soil depth in modulating recharge prediction. Unlike previous studies that always point out climatic impacts, the present work highlights the added value of considering land surface characteristics. It presents slope, vegetation, organic matter, soil texture, and soil depth as modulators of recharge efficiency. This extension provides a more integrated framework for water-energy planning.

PARTIAL CONCLUSION

This chapter has shown that groundwater recharge (GWR) across Africa is governed by a combination of precipitation as the dominant driver and land-surface properties that modulate how precipitation is partitioned into evapotranspiration and runoff.

GENERAL CONCLUSION AND RECOMMENDATION

GENERAL CONCLUSION AND RECOMMENDATION

In this thesis, a data-driven framework combining regional-scale correlation analysis, principal component analysis (PCA), and machine-learning (Random Forest and Gradient Boosting) modelling was applied to quantify regional differences in recharge and to understand their relationship with land-surface characteristics. The results provide new insights into the dominant mechanisms affecting recharge partitioning across Africa, while highlighting implications for regional energy-resource planning.

The first key conclusion is that precipitation is the primary determinant of spatial variability in recharge across all African regions, with machine-learning SHAP analysis consistently showing that recharge predictions increase most strongly in association with higher precipitation inputs.

Secondly, land surface characteristics influence recharge by impacting the fraction of rainfall lost to evapotranspiration (ET_R) and runoff (RO_R). Correlation analysis revealed that soil depth and slope are the most important factors that influence the runoff ratio. Shallow soils and steep terrain are strongly associated with higher runoff. Conversely, deeper soils and gentle slopes are associated with reduced runoff fractions.

Thirdly, the principal component analysis results confirm that the land-surface caracteristics can be summarized into two dominant gradients. The Soil Texture Gradient (PC1) associated with storage and vegetation properties that may influence evaporative losses. The Infiltration versus Drainage Gradient (PC2) linked to slope and soil thickness governing runoff.

Machine-learning performance metrics indicate that Random Forest models outperform Gradient Boosting models (R^2 up to 0.98 for RF versus 0.95 for GB), reflecting their greater tolerance for strong non-linear interactions. Nonetheless, SHAP-based interpretation of both models indicates a similar control hierarchy: precipitation dominates, while other variables, particularly soil depth, slope and vegetation (LAI), consistently shift modelled recharge predictions by altering losses.

Three broad regional recharge behavior types are identified by integrating evidence across all analyses:

- Runoff-limited regimes where recharge is limited by shallow soils and steep terrain.

- Evapotranspiration-limited regimes where rainfall is largely lost to evapotranspiration due to high vegetation and organic-rich conditions.
- Precipitation-constrained regimes (e.g., SAH, MED) where low rainfall dominates recharge potential regardless of land surface characteristics.

These findings have implications for Africa's renewable energy transition. It is usful for green hydrogen projects, which require reliable and sustainable water inputs. Regions with higher recharge efficiency, such as West and Central East Africa present promising locations for coupling renewable electricity with groundwater-fed hydrogen electrolysis. Conversely, dryland regimes such as the Mediterranean and Sahara exhibit recharge scarcity. In these areas, Hydrogen production may require alternative water sources, such as desalination or wastewater reuse, to avoid unsustainable groundwater depletion.

The following policy recommendations are proposed based on regional contrasts:

- in West and Central Africa (Runoff-limited regimes): promote soil and slope management interventions such as terracing, bunds, and infiltration trenches to enhance infiltration and reduce surface losses.
- In Ethiopia and highland regions (NEAF): prioritize landscape management to reduce runoff and improve recharge through community-led terracing.
- In Morocco and North Africa (MED): expand adoption of dry cooling technologies for concentrated solar power (CSP) to reduce water demand, while integrating recharge variability into energy feasibility studies.
- In Southern regions (SEAF, SWAF): account for evapotranspiration-driven recharge losses. When planning groundwater-based energy systems, vegetation and land use management should be integrated.

Future work should seek to validate recharge estimates with ground observations, address uncertainties linked to CLM simulations, and integrate socio-economic and energy-demand scenarios to evaluate trade-offs between water use and renewable energy expansion. By strengthening the connection between hydrology, data science, and energy policy, this thesis contributes a framework for ensuring that Africa's pursuit of clean energy, particularly green hydrogen, proceeds without undermining the sustainability of its vital groundwater resources.

BIBLIOGRAPHY REFERENCES

1. Adom, R. K., Simatele, M. D., & Reid, M. (2022). Addressing the challenges of water-energy-food nexus programme in the context of sustainable development and climate change in South Africa. *Journal of Water and Climate Change*, *13*(7), 2761–2779. https://doi.org/10.2166/wcc.2022.099

2. Akpoti, K., Velpuri, N. M., Mizukami, N., Kagone, S., Leh, M., Mekonnen, K., Owusu, A., Tinonetsana, P., Phiri, M., Madushanka, L., Perera, T., Prabhath, P. T., Parrish, G. E. L., Senay, G. B., & Seid, A. (2024). Advancing water security in Africa with new high-resolution discharge data. *Scientific Data* , *11*(1), 1195. https://doi.org/10.1038/s41597-024-04034-0

3. Alam, Md J. B. (2017). Evaluation of Plant Root on the Performance of Evaluation of Plant Root on the Performance of Evapotranspiration (ET) Cover System. Doctorate thesis in Civil Engineering, University of Texas at Arlington, 351p. https://mavmatrix.uta.edu/cgi/viewcontent.cgi?article=1435&context=civilengineering_dissertations

4. Anand, V., Rajput, V. D., Minkina, T., Mandzhieva, S., Sharma, A., Kumar, D., & Kumar, S. (2025). Evaluating groundwater potential with the synergistic use of geospatial methods and advanced machine learning approaches. *Discover Cities*, *2*(1), 56. https://doi.org/10.1007/s44327-025-00095-x

5. Anghileri, D., Pastori, M., Marcos-Garcia, P., Umlauf, G., Crestaz, E., Seliger, R., Iervolino, A., Cordano, E., Cattaneo, L., & Carmona-Moreno, C. (2024). Global Water Challenges in Sub-Saharan Africa and how to strengthen science-policy dialogues on transboundary governance and cooperation. *Journal of Environmental Management*, *365*, 121417. https://doi.org/10.1016/j.jenvman.2024.121417

6. Asuero, A. G., Sayago, A., & González, A. G. (2006). The correlation coefficient: An overview. In *Critical Reviews in Analytical Chemistry* 36(1), 41–59. https://doi.org/10.1080/10408340500526766

7. Atawneh, D. Al, Cartwright, N., & Bertone, E. (2021). Climate change and its impact on the projected values of groundwater recharge: A review. In *Journal of Hydrology* 601, 126602. Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2021.126602

8. Bardsley, W. E. (1995). Using groundwater for hydro electric power generation. *Journal of Hydrology (New Zealand)*, *34*(1), 1–14. http://www.jstor.org/stable/43944742

9. Basset, C., Abou Najm, M., Ghezzehei, T., Hao, X., & Daccache, A. (2023). How does soil structure affect water infiltration? A meta-data systematic review. In *Soil and Tillage Research* 226, 105777. Elsevier B.V. https://doi.org/10.1016/j.still.2022.105577

10. Bayat, B., Oloruntoba, B., Montzka, C., Vereecken, H., & Hendricks Franssen, H. J. (2023). Implications for sustainable water consumption in Africa by simulating five decades (1965–2014) of groundwater recharge. *Journal of Hydrology*, *626*, 130288. https://doi.org/10.1016/j.jhydrol.2023.130288

11. Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1–2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

12. Biazar, S. M., Golmohammadi, G., Nedhunuri, R. R., Shaghaghi, S., & Mohammadi, K. (2025). Artificial Intelligence in Hydrology: Advancements in Soil, Water Resource Management, and Sustainable Development. In *Sustainability (Switzerland),* 17(5), 2250. Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/su17052250

13. Chigozie Ani, E., Kayode Olajiga, O., Queen Sikhakane, Z., & Michael Olatunde, T. (2024). Renewable Energy Integration For Water Supply: A Comparative Review Of African And U.S. Initiatives. *Engineering Science & Technology Journal*, *5*(3), 1086–1096. https://doi.org/10.51594/estj/v5i3.972

14. CTSM. (2020). *CLM5.0 Technical Note: Hydrology. CTSM (Community Terrestrial Systems Model) release-CLM5.0 Documentation.* https://escomp.github.io/CTSM//tech_note/index.html , viewed the 30/09/2025

15. Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

16. Donkor, S. M. K., & Wolde, Y. E. (2022). *Intergrated Water Resource Management in Africa: Issues and Options*. Economic Commission for Africa, 20p. https://communities.adaptationportal.gca.org/knowledge-base/urban-resilience/integrated-water-resources-management-in-africa-issues-and-options

17. Du, Y., & Pechlivanidis, I. G. (2025). Hybrid approaches enhance hydrological model usability for local streamflow prediction. *Communications Earth and Environment*, *6*(1), 334. https://doi.org/10.1038/s43247-025-02324-y

18. Elakkiya, N., & Sankarganesh, E. (2023). Quantum Geographic Information System (QGIS) for Mapping of Study Areas. *Biotica Research Today*, *51*(1), 70–75. https://www.researchgate.net/publication/368449796

19. Fares, M. S. Ben, & Abderafi, S. (2018). Water consumption analysis of Moroccan concentrating solar power station. *Solar Energy*, *172*, 146–151. https://doi.org/10.1016/j.solener.2018.06.003

20. Franzluebbers, A. J. (2002). Water infiltration and soil structure related to organic matter and its stratification with depth. *Soil & Tillage Research*, *66*, 197–205. https://doi.org/10.1016/S0167-1987(02)00027-2

21. Fu, G., Crosbie, R. S., Barron, O., Charles, S. P., Dawes, W., Shi, X., Van Niel, T., & Li, C. (2019). Attributing variations of temporal and spatial groundwater recharge: A statistical

analysis of climatic and non-climatic factors. *Journal of Hydrology*, *568*, 816–834. https://doi.org/10.1016/j.jhydrol.2018.11.022

22. Gebreslassie, H., Berhane, G., Gebreyohannes, T., Hagos, M., Hussien, A., & Walraevens, K. (2025). Water Harvesting and Groundwater Recharge: A Comprehensive Review and Synthesis of Current Practices. In *Water (Switzerland),* 17(7), 976. Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/w17070976

23. Gebru, A. B., Gebreyohannes, T., Kahsay, G. H., & Grum, B. (2024). The dilemma of choosing appropriate groundwater recharge estimation methods in Ethiopia: A systematic review of the existing methods. In *Groundwater for Sustainable Development* 27, 101358. Elsevier B.V. https://doi.org/10.1016/j.gsd.2024.101358

24. Genuer, R., Poggi, J. M., Tuleau-Malot, C., & Villa-Vialaneix, N. (2017). Random forests for big data. *Big Data Research*, *9*, 28-46. https://doi.org/10.1016/j.bdr.2017.07.003

25. Gómez-Escalonilla, V., Martínez-Santos, P., & Martín-Loeches, M. (2022). Preprocessing approaches in machine-learning-based groundwater potential mapping: An application to the Koulikoro and Bamako regions, Mali. *Hydrology and Earth System Sciences*, *26*(2), 221–243. https://doi.org/10.5194/hess-26-221-2022

26. Hasan, B. M. S., & Abdulazeez, A. M. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *2*(1), 20–30. https://doi.org/10.30880/jscdm.2021.02.01.003

27. IRENA (2021). *Making the breakthrough: Green hydrogen policies and technology costs. In* International Renewable Energy Agency, Abu Dhabi, 68p. https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Nov/IRENA_Green_Hydrogen_breakthrough_2021.pdf?la=en&hash=40FA5B8AD7AB1666EECBDE30EF458C45EE5A0AA6

28. Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065), 20150202. Royal Society of London. https://doi.org/10.1098/rsta.2015.0202

29. Jung, H., Saynisch-Wagner, J., & Schulz, S. (2024). Can eXplainable AI Offer a New Perspective for Groundwater Recharge Estimation?—Global-Scale Modeling Using Neural Network. *Water Resources Research*, *60*(4), e2023WR036360. https://doi.org/10.1029/2023WR036360

30. Kaspar, F., Schulzweida, U., & Müller, R. (2010). Climate data operators" as a user-friendly processing tool for CM SAF's satellite-derived climate monitoring products. *Conference: EUMETSAT Meteorological Satellite Conference 2010*, 20–24. https://doi.org/10.13140/RG.2.2.20422.68165 , viewed the 30/09/2025

31. Kim, N. H., & Office, N. D. (2017). *Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1) [Data set]. Data Integration and Analysis System (DIAS)*. *Development of a Phase Model*, Institute of Industrial Science, University of Tokyo, 4p. https://doi.org/10.20783/DIAS.501

32. Lawrence, D. M., Fisher, R. A., Koven, C. D., Swenson, S. C., Bonan, G., van Kampenhout, L., Collier, N., Oleson, K. W., Ghimire, B., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, H., Lombardozzi, D., Riley, W. J., Vertenstein, M., Badger, A. M., Wieder, W. R., Li, F., Sacks, W. J., Shi, M., Xu, C., Ali, A. A., Bisht, G., Burns, S. P., Buzan, J., Fisher, J. B., Gentine, P., Kumar, S., Hoffman, F., Lenaerts, J., Drewniak, B., Pandey, A., Pelletier, J. D., Clark, M., van den Broeke, M., Brunke, M. A., Craig, A., Dahlin, K., Flanner, M., Fox, A. M., Keppel-Aleks, G., Leung, L. R., Knox, R., Lipscomb, W. H., Perket, J., Ricciuto, D. M., Sanderson, B. M., Tang, J., Thomas, R. Q., Lu, Y., Randerson, J. T., Slater, A., Val Martin, M., Subin, Z. M., & Zeng, X. (2019). The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. *Journal of Advances in Modeling Earth Systems, 11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

33. Leigland, J., Trémolet, S., & Ikeda, J. (2016). *Achieving universal access to water and sanitation by 2030: The role of blended finance,* 20. Washington, DC: World Bank. 20p. https://documents1.worldbank.org/curated/en/978521472029369304/pdf/Achieving-universal-access-to-water-and-sanitation-by-2030-the-role-of-blended-finance.pdf https://openknowledge.worldbank.org/server/api/core/bitstreams/86a91dae-20a9-5d73-9a18-36e0f05bca7d/content

34. Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*, 1–10. http://arxiv.org/abs/1705.07874

35. MacDonald, A. M., Bonsor, H. C., Dochartaigh, B. É. Ó., & Taylor, R. G. (2012). Quantitative maps of groundwater resources in Africa. In *Environmental Research Letters*, 7(2), 024009. Institute of Physics Publishing. https://doi.org/10.1088/1748-9326/7/2/024009

36. Maity, R., Srivastava, A., Sarkar, S., & Khan, M. I. (2024). Revolutionizing the future of hydrological science: Impact of machine learning and deep learning amidst emerging explainable AI and transfer learning. *Applied Computing and Geosciences*, *24*, 100206. https://doi.org/10.1016/j.acags.2024.100206

37. Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536. **https://doi.org/10.3390/w10111536**

38. Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. V. (2021). What Role Does Hydrological Science Play in the Age of Machine Learning? In *Water Resources Research*, 57(3), 1–15. Blackwell Publishing Ltd. https://doi.org/10.1029/2020WR028091

39. Nhamo, L., Ndlela, B., Nhemachena, C., Mabhaudhi, T., Mpandeli, S., & Matchaya, G. (2018). The water-energy-food nexus: Climate risks and opportunities in Southern Africa. In *Water (Switzerland)*, 10(5), 567. MDPI AG. https://doi.org/10.3390/w10050567

40. Oloruntoba, B., Kollet, S., Montzka, C., Vereecken, H., & Hendricks Franssen, H.-J. (2025). High-resolution land surface modelling over Africa: the role of uncertain soil properties in combination with forcing temporal resolution. *Hydrology and Earth System Sciences*, *29*(6), 1659–1683. https://doi.org/10.5194/hess-29-1659-2025

41. Pazola, A., Shamsudduha, M., French, J., MacDonald, A. M., Abiye, T., Goni, I. B., & Taylor, R. G. (2024). High-resolution long-term average groundwater recharge in Africa estimated using random forest regression an residual interpolation. *Hydrology and Earth System Sciences*, *28*(13), 2949–2967. https://doi.org/10.5194/hess-28-2949-2024

42. Ersoy, S.R., Terrapon-Pfaff, J. and Agouzoul H. (2022). Sustainable Transformation Of Morocco's Energy System : Development of a Phase Model, Climate Change, Energy and Environment, FRIEDRICH-EBERT-STIFTUNG, 43p. https://epub.wupperinst.org/frontdoor/deliver/index/docId/8487/file/8487_Morocco.pdf

43. Scanlon, B. R. (2010). Chemical tracer methods. In *Estimating Groundwater Recharge*, 136–165, Chapter 7, Healy, R. W. (eds), Cambridge University Press, Cambridge, 257p. https://doi.org/10.1017/cbo9780511780745.008 http://ndl.ethernet.edu.et/bitstream/123456789/57745/1/611.pd

44. Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., & McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9320–9325. https://doi.org/10.1073/pnas.1200311109

45. Scanlon, B. R., Healy, R. W., & Cook, P. G. (2002). Choosing appropriate techniques for quantifying groundwater recharge. *Hydrogeology Journal*, *10*, 18–39. https://doi.org/10.1007/s10040-0010176-2

46. Scholz, M. (2024). New methodology for identifying sustainable freshwater resources for the production of green hydrogen. *International Journal of Sustainable Engineering*, *17*(1), 1–7. https://doi.org/10.1080/19397038.2024.2321612

47. Siabi, E. K., Dile, Y. T., Kabo-Bah, A. T., Amo-Boateng, M., Anornu, G. K., Akpoti, K., Vuu, C., Donkor, P., Mensah, S. K., Incoom, A. B. M., Opoku, E. K., & Atta-Darkwa, T. (2022). Machine learning based groundwater prediction in data scare Volta basin of Ghana. *Applied Artificial Intelligence*, *36*(1), 2138130. https://doi.org/10.1080/08839514.2022.2138130

48. Sishodia, R. P., Shukla, S., Wani, S. P., Graham, W. D., & Jones, J. W. (2018). Future irrigation expansion outweigh groundwater recharge gains from climate change in semi-arid India. *Science of the Total Environment*, *635*, 725–740. https://doi.org/10.1016/j.scitotenv.2018.04.130

49. Tian, W., Li, X., Wang, X.-S., & Hu, B. X. (2012). Coupling a groundwater model with a land surface model to improve water and energy cycle simulation. *Hydrology and Earth System Sciences Discussions, 9*(2), 1163–1205. https://doi.org/10.5194/hessd-9-1163-2012

50. Tonkonogy, B., Brown, J., Micale, V., Wang, X., & Clark, A. (2018). *Blended Finance in Clean Energy: Experiences and Opportunities*, A Report for the Business & Sustainable Development Commission and the Blended Finance Taskforce, 38p. https://climatepolicyinitiative.org/wp-content/uploads/2018/01/Blended-Finance-in-Clean-Energy-Experiences-and-Opportunities.pdf

51. Toure, H., Boateng, C., Gidigasu, S., Wemegah, D., Mensah, V., Aryee, J., Osei, M., Gilbert, J., & Afful, S. (2024). Review of machine learning algorithms used in groundwater availability

studies in Africa: analysis of geological and climate input variables. *Discover Water*, *4*(1), 109. https://doi.org/10.1007/s43832-024-00109-6

52. UN-Water/Africa (2000). *The Africa water vision for 2025: equitable and sustainable use of water for socioeconomic development*. Economic Commission for Africa, African Union, African Development Bank, Addis Ababa, Ethiopia. 34p. https://www.afdb.org/fileadmin/uploads/afdb/Documents/Generic-Documents/african%20water%20vision%202025%20to%20be%20sent%20to%20wwf5.pdf

53. WWAP (2018). The United Nations world water development report 2018: Nature-based solutions for water, United Nations World Water Assessment Programme/ UN-Water, United Nations Educational Scientific and Cultural Organization (UNESCO). Paris, France,139p. https://unesdoc.unesco.org/ark:/48223/pf000026142

54. Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026

55. Wang, L., Dochartaigh, B. Ó., & Macdonald, D. (2010). A literature review of recharge estimation and groundwater resource assessment in Africa, British Geological Survey, Groundwater Resources Programme, Internal Report IR/10/051, 35p. https://nora.nerc.ac.uk/id/eprint/14145/1/BGS_Report-A_literature_review_of_recharge_estimation.pdf

56. Winkler, C., Heinrichs, H., Peña Sanchez, E. U., Ishmam, S., Bayat, B., Lahnaoui, A., Agbo, S., Franzmann, D., Oijeabou, N., Koerner, C., Michael, Y., Oloruntoba, B., Brauner, S., Kuckshinrichs, W., Montzka, C., Vereecken, H., Hendricks Franssen, H., Brendt, J., Venghaus, S., Koné, D., Korgo, B., Ogunjobi, K., Olwoch, J., Chiteculo, V., Getenga, Z., Linßen, J., & Stolten, D. (2025). Participatory mapping of local green hydrogen cost-potentials in Sub-Saharan Africa. *International Journal of Hydrogen Energy, 112*, 289–321. https://doi.org/10.1016/j.ijhydene.2025.02.015

57. Xiang, X., Zhang, H., & Xia, S. T. (2020, August). Label aggregation of gradient boosting decision trees. In *Proceedings of the 2020 2nd International Conference on Image Processing and Machine Vision,* 140-145. https://doi.org/10.1145/3421558.3421581

58. Xing, J., Wang, H., Luo, K., Wang, S., Bai, Y., & Fan, J. (2019). Predictive single-step kinetic model of biomass devolatilization for CFD applications: A comparison study of empirical correlations (EC), artificial neural networks (ANN) and random forest (RF). *Renewable Energy*, *136*, 104–114. https://doi.org/10.1016/j.renene.2018.12.088

59. Zender, C. S. (2008). Analysis of self-describing gridded geoscience data with netCDF Operators (NCO). *Environmental Modelling and Software*, *23*(10–11), 1338–1342. https://doi.org/10.1016/j.envsoft.2008.03.004

60. Zhang, Y., & Haghani, A. (2015). A gradient boosting method to improve travel time prediction. *Transportation Research Part C: Emerging Technologies*, *58*, 308–324. https://doi.org/10.1016/j.trc.2015.02.019