

Addis Ababa University, College of Technology and Built Environment

School of Civil and Environmental Engineering

Machine Learning and Multi-Source Data Integration for Improved Hydrological Forecasting: From Physics-Informed Hybrid Modeling to Transfer Learning Techniques in Data-Scarce Basins

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering (Specialization in Hydraulic Engineering)

By: Juhar Mohammed

June, 2026

Addis Ababa, Ethiopia

Doctoral Dissertation Approval Sheet

Addis Ababa University (AAU)

College of Technology and Built Environment (CTBE)

School of Civil and Environmental Engineering

This is to certify that the dissertation prepared by Juhar Mohammed, entitled, **Machine Learning and Multi-Source Data Integration for Improved Hydrological Forecasting: From Physics-Informed Hybrid Modeling to Transfer Learning Techniques in Data-Scarce Basins**, and submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Civil Engineering (Hydraulic Engineering) complies with the regulation of the university and meets the accepted standards with respect to quality and originality.

Dr. Yenesew Mengiste		13 Aug, 2026
Supervisor	Signature	date
Prof. Mekonnen Gebremichael	Mekonnen Gebremichael	June 15, 2026
Co- Supervisor	Signature	date

APPROVED BY THE BOARD OF EXAMINERS:

Prof. Tsegaye Tadesse	TSEGAYE TADESSE	June 10, 2026
External Examiner	Signature	date
Dr. Agizew Nigusse		12 Aug, 2026
Internal Examiner	Signature	date
Dr. Edom Adane		Aug 12, 2026
Chairperson	Signature	date
Dr. Yenesew Mengiste		13 Aug, 2026
(Dean, SCEE)	Signature	date

Vice Executive Dean for Academic Affairs	Signature	Date

Vice Executive Dean Office for Academic Affairs
Addis Ababa University College of Technology and Built Environment

CamScanner

Abstract

Accurate and robust hydrological forecasting is essential for improving water resource management, enhancing climate change adaptation, and reducing disaster risks. However, forecasting remains challenging, particularly in complex, data-scarce river basins. The lack of sufficient in situ observations, limited skill and inherent uncertainties in forecasting models, complex interactions of hydroclimatic systems, and nonlinear behavior of hydrological processes pose major challenges. To address these challenges, this dissertation employs the Upper Blue Nile Basin as a case study and develops an integrated hydrological forecasting framework that combines multi-source data, process-based modeling, and advanced machine learning techniques, organized around four key objectives. The first objective focuses on developing an advanced machine learning-based multi-source data fusion approach to improve spatiotemporal precipitation estimates in data-scarce river basins. This study employed a Random Forest-based RFmep data fusion method to merge gauge data with satellite (IMERG, CHIRPS, PERSIANN-CCS) and reanalysis (ERA5) products. The merged dataset significantly improved performance compared to individual products, achieving higher correlation (CC > 0.7), improved efficiency (NSE > 0.6), and reduced RMSE and MAE. In addition, categorical metrics such as POD and CSI improved, while FAR decreased. The second objective proposed a novel precipitation forecasting framework that integrates multiple data sources with machine learning models to improve monthly precipitation forecasting. The framework utilized predictors including local meteorological variables, large-scale climate indices, soil moisture, and lagged precipitation. The study employed five machine learning models: ANN, SVM, KNN, RF, and XGBoost. The framework displayed excellent forecasting performance, with R^2 values between 0.64 and 0.87, RMSE from 0.62 to 2.95, MAE between 0.18 and 2.42 respectively. Among the models, ANN, SVM, and KNN outperformed global forecasting systems such as NCEP and ECMWF. The framework performed well during normal and dry months, but performance dropped during wet months due to challenges in capturing extreme conditions. The third objective presents a physics-informed hybrid forecasting framework that integrates multi-source data and process-based hydrological models with machine learning techniques to improve monthly streamflow forecasting. First, the process-based SWAT model generates physics-informed

hydrological variables (e.g., surface runoff, baseflow, evapotranspiration). These, along with local meteorological data and large-scale climate indices, serve as inputs to an LSTM model. The hybrid model achieved excellent performance, with NSE = 0.90, KGE = 0.89, and PBIAS = 2.1%, significantly outperforming standalone machine learning models (standalone LSTM: 0.76, SVR: 0.69, RF: 0.68). This integration improved representation of hydrological processes, reduced systematic errors, and enhanced the model's ability to capture seasonal and interannual variability. The fourth objective addressed the challenge of prediction in ungauged sub-basins using a physics-informed transfer learning framework. A pre-trained LSTM model from a data-rich basin (Guder) was transferred and fine-tuned for a data-scarce basin (Muger). Despite using only 26 local observations, the model achieved strong performance (NSE = 0.77, KGE = 0.702, R^2 = 0.82), effectively capturing both low and high-flow conditions. This demonstrates the potential of transfer learning as a scalable and efficient solution for hydrological prediction in data-limited regions. In general, the integrated hydrological forecasting framework introduced in this study substantially advances the accuracy, reliability, and applicability of precipitation and streamflow predictions, offering a robust and scalable solution for data-scarce basins and supporting informed decision making under rising climate variability.

Keywords: *multi-source data, precipitation data merging, precipitation forecasting, streamflow forecasting, hybrid modeling framework, machine learning, transfer learning*

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my God Allah for granting me strength, wisdom, perseverance and his willingness throughout this journey.

First, I sincerely thank my dedicated principal advisor, **Dr. Yenesew Mengiste**, for her support, guidance, wisdom, and patience throughout my studies. She consistently sets high standards for my work and generously shares her enthusiasm and deep knowledge to help me achieve my goals during my research and thesis writing. I am truly grateful for her unwavering support and excellent supervision. I also want to express my special appreciation to my co-supervisor, **Prof. Mekonnen Gebremichael**, for his incredible support and encouragement, for paying close attention to my work, and for making himself always available whenever I needed help. Lastly, I extend my heartfelt gratitude to **Professor Vijay Singh**, a globally renowned expert in water resource engineering, for his generosity in allowing me to work with him.

My heartfelt thanks go to my family for their unwavering love and support. To my **Mom** and **Dad**, your encouragement and sacrifices have been the foundation of my success. I am eternally grateful for your belief in me and your constant support, prayers, and unconditional love throughout this journey. I extend my sincere appreciation to my sisters **Almenesh Haile** and **Almenesh Tadese** for their invaluable support and cooperation.

I also acknowledge with deep gratitude my **Brother** and **Aunt**, who passed away during this study period. My brother was a cornerstone of my success, a supporter, encourager, and helper in every way. Though he passed away in the final moments before this dissertation dream came true, this success is dedicated to him. I am also grateful to my aunt for her unwavering support during my early education and throughout my journey.

My very special thanks and gratitude go to my friend **Mr. Jiang** for his continuous encouragement and moral support. His kindness and good friendship can never be forgotten.

Lastly, I want to thank all my friends and family, particularly the **Sheikh Oumer, Sofia, Wend, Bereket,** and **Dr. Admasu**, for their companionship and support throughout this journey. Your friendship has made this experience even more meaningful.

List of Published Papers:

1. Mohammed, J., Mengiste, Y., & P. Singh, V. Improving spatio-temporal precipitation estimates in data-scarce river basins: an application of machine learning-based multi-source data merging. *Stoch Environ Res Risk Assess* **37**, 1353–1369 (2023). https://doi.org/10.1007/s00477-022-02346-4
2. Mohammed, J., Mengiste, Y. & Gebremichael, M. Enhancing monthly precipitation forecasting by integrating multi-source data with machine learning models: a study in the Upper Blue Nile Basin. *Model. Earth Syst. Environ.* **11**, 35 (2025). https://doi.org/10.1007/s40808-024-02223-9

Contents

Abstract ii

List of Published Papers: v

List of Tables x

List of Figures xi

Abbreviations xiii

1. Introduction 1

1.1. Background of the Study 1

1.2. Statement of the Problem 4

1.3. Objectives 6

Specific Objectives 6

1.4. Research Questions 6

1.5. Significance of the study 6

1.6. Dissertation Organization 8

2. Description of the Study Area and Data 9

2.1. Location and Geography 9

2.2. Climate 10

2.3. Hydrology 10

2.4. Land Use 11

2.5. Human Influence and Hydraulic Infrastructure Development 12

2.6. Monitoring Stations and Hydro-Meteorological Data Challenges 13

2.7. Data 14

2.8. Meteorological Data Quality Assessment and Station Selection 18

3. Improving Spatio-Temporal Precipitation Estimates in Data-Scarce River Basins: An Application of Machine Learning-Based Multi-Source Data Merging 21

3.1. Introduction 22
3.2. Methodology 25
3.3. Results and Discussion 29
3.3.1. Results 29
3.3.1.1. Performance of unmerged precipitation products 29
3.3.1.2. Performance of merged product 34
3.3.1.3. Qualitative metrics results 39
3.3.2. Discussion 41
3.4. Conclusion 43
4. Enhancing monthly precipitation forecasting by integrating multi-source data with machine learning models 45
4.1. Introduction 46
4.2. Data and Methods 48
4.2.1. Data 48
4.2.2. Selection of Predictors 49
4.3. Forecasting Framework 49
4.4. Machine Learning Models 50
4.4.1. K-Nearest neighbors (KNN) 50
4.4.2. Random Forest 51
4.4.3. Artificial Neural Network 51
4.4.4. Support Vector Machine 52
4.4.5. Extreme Gradient Boosting (XGBoost) 52
4.5. Classification of Normal, Dry, and Wet Months 53
4.6. Forecast Evaluation 53
4.7. Results and Discussion 54

4.8. Conclusion 67

5. Physics- Informed Hybrid Forecasting Framework for Improving Monthly Streamflow Forecasting: Integration of Process-Based Hydrological Models, Multi-Source Data, and Machine Learning Techniques 69

5.1. Introduction 70

5.2. Description of the Study Area 73

5.2.1. Climate and Hydrology 73

5.2.2. Land Use and Soil 74

5.3. Datasets 75

5.3.1. DEM 75

5.3.2. Soil Data 75

5.3.3. LULC data 75

5.3.4. Hydrometeorology data 76

5.4. Methodology 76

5.4.1. Description of the SWAT model 76

5.4.2. Model Setup 77

5.4.3. Calibration and Validation 77

5.4.4. Sensitivity Analysis 78

5.5. Long Short-Term Memory (LSTM) Networks 78

5.6. SWAT-LSTM Hybrid Model 80

5.7. Result 82

5.7.1. Performance of the SWAT Model 82

5.7.2. Quantitative performance evaluation of the proposed hybrid framework 86

5.7.3. Visual Diagnostic Analysis 90

5.7.4. Residual Error Distribution 91

5.8. Discussion 92

5.9. Conclusion 95

6. Leveraging the Power of Transfer Learning to Improve Streamflow Forecasting in a Data-Scarce River Sub-Basin 97

6.1. Introduction 98

6.2. Transfer learning framework 100

6.2.1. Similarity Index 102

6.2.2. K-Means Algorithm 103

6.3. Transfer Learning approach 107

6.4. Result and Discussion 109

6.4.1. Quantitative Performance Metrics 110

6.4.2. Predictive Accuracy and Visual Diagnostics 111

6.5. Conclusion 113

7. Conclusions and Recommendations 115

7.1. Discussion 115

7.2. Conclusions 116

8. Recommendations 119

Reference 122

List of Tables

Table 3.1 Validation metrics and equations 29
Table3.2. Quantitative metrics results based on daily scale measurements 31
Table 4.1: Evaluation metrics and equation 54
Table 4.2. Performance of the forecasting framework 55
Table 5.2. Summary of Performance Metrics 87
Table 6.1. Upper Blue Nile Basin with Sub-Basin Cluster 106
Table 6.2. Comprehensive Performance Metrics for Training and Validation Phases ... 110

List of Figures

Fig. 2.1: Geographic locations and terrain elevation map of the study area 9
Fig.2.2 Map of River Gauge and Meteorological Stations 19
Fig. 3.1. A schematic presentation of the merging procedure 27
Fig 3.2. (a) Pearson coefficient of correlation for respective precipitation gauge station based on daily scale measurements (PGS), (b) Index of Agreement for respective precipitation gauge station (PGS) based on daily scale measurements 32
Fig 3.3. Boxplots representing the performance of individual and merged precipitation datasets based on Kling–Gupta efficiency with respect to ground precipitation gauge stations (PGS) 36
Fig.3.4 (a-f) Mean daily precipitation for the period 2001-2017..... 39
Fig. 3.5 Categorical validation metrics: presentation-based on Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) with respect to ground measurements at precipitation gauge station (PGS). 40
Fig. 4.1. A schematic presentation of the forecasting framework 50
Fig. 4-2. (a- e) Time series graphs observed vs forecasted rainfall 59
Fig. 4.3. (a-e) Observed vs Forecasted monthly Precipitation in mm 62
Fig. 4.4. Comparison of forecast skill for the proposed framework against global precipitation forecasts 63
Fig. 4.5. Performance evaluation of the forecasting framework under different climatic conditions. 66
Fig. 5.1. Study Area Map 73
Fig.5.2. The hybrid streamflow forecasting framework 82
flow for the calibration period flow for the calibration period 85
Fig. 5.3. Scatter and time series plot of observed vs simulated flow for calibration period 85
Fig. 5.4 scatter and time series plot of observed vs simulated flow for the validation period 86
Fig. 5.5 Scatter plot of all models' forecasts vs observed streamflow 88
Fig. 5.6. Kling-Gupta and Nash Sutcliffe Efficiency metric-based comparison 89
Fig. 5.7. Time series plot of observed vs forecasted for all models 90

..... 91
Fig. 5.8. Residual Distribution Plot 91
Fig. 5.9. Taylor Diagram..... 92
Fig 6.1. Map of Source and Target Sub-Basins 105
Fig. 6.2. Performance Metrics Summary: Training vs. Validation..... 111
Fig. 6.3. Scatter Plots of Predicted vs. Actual Streamflow..... 112
Fig. 6.4. Time Series of Actual vs. Predicted Streamflow..... 113

Abbreviations

ANN	Artificial Neural Network
CC	Correlation Coefficient
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station Data
CSI	Critical Success Index
DEM	Digital Elevation Model
DL	Deep Learning
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	ECMWF Reanalysis v5
FAR	False Alarm Ratio
GPM	Global Precipitation Measurement
IAHS	International Association of Hydrological Sciences
IMERG	Integrated Multi-satellitE Retrievals for GPM
KGE	Kling-Gupta Efficiency
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MOWRE	Ministry of Water Resources and Energy
NCEP	National Centers for Environmental Prediction
NMSA	National Meteorological Services Agency
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percent Bias
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System

ANN	Artificial Neural Network
POD	Probability of Detection
PUB	Prediction in Ungauged Basins
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SPI	Standardized Precipitation Index
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
SWAT	Soil and Water Assessment Tool
TL	Transfer Learning
UBNB	Upper Blue Nile Basin
USGS	United States Geological Survey
WMO	World Meteorological Organization
XGBOOST	eXtreme Gradient Boosting

1. Introduction

1.1. Background of the Study

Hydrological forecasting is a critical component of effective water resource management, providing vital information on the availability of surface and groundwater resources across spatial and temporal scales (Haddad et al., 2025; Chang and Guo, 2020; Vişescu et al., 2017). Accurate and reliable forecasts of hydrological components, including precipitation, streamflow, and extreme events, are indispensable for informed decision-making regarding irrigation, reservoir operations, and disaster preparedness, particularly within the context of climate change, rapid population growth, and shifting land-use patterns (Ramachandran et al., 2024; Luo et al., 2023). Furthermore, they provide crucial insights to communities and authorities to support early warning systems that anticipate future anomalous conditions and mitigate the detrimental effects of extreme hydrological events, including floods and droughts (Hahn et al., 2024; Liu et al., 2022).

Despite its paramount importance, hydrological forecasting faces significant challenges that undermine its accuracy and reliability. A primary difficulty stems from the inherent complexity of hydrological systems, which are governed by the intricate interplay of atmospheric, terrestrial, and oceanic processes operating across diverse scales (Zhang et al., 2023; Slater et al., 2023). This systemic complexity is further amplified by local physiographic and anthropogenic factors, such as topography, land use, soil moisture, and vegetation, which affect the movement of water within a specific basin or intra-basin (Slater et al., 2023; Cao et al., 2022; Torre et al., 2019). Consequently, traditional physical and statistical models often struggle to capture the full nonlinearity and dynamics of these systems, particularly under the nonstationary conditions induced by climate change, leading to substantial forecasting errors (Solanki et al., 2025; Lauro et al., 2024; Duethmann et al., 2020).

Moreover, data scarcity frequently exacerbates these challenges, especially in many developing regions where monitoring infrastructure is inadequate or entirely absent, resulting in insufficient and unreliable historical records for model calibration and validation (Haddad et al., 2025; Nearing et al., 2024). This scarcity is further compounded by the uncertainties inherent in data collection methods, including measurement errors and

sensor malfunctions, which often compromise forecast accuracy and reliability (Xu et al., 2024; Nurhamidah et al., 2021). Additionally, uncertainties in model structure, poor parameterization, initial conditions, and the equifinality problem in the calibration process present substantial obstacles to forecast efficacy (Leta et al., 2022; Hallaouin et al., 2020; Foulon & Rousseau, 2018). Furthermore, significant computational demands often constrain the operational deployment of advanced hydrological forecasting techniques. Consequently, running high-resolution, physically based models coupled with ensemble forecasting and sophisticated deep learning procedures requires extensive computational resources, thereby limiting their practical applicability in resource-constrained regions (Slater et al., 2023; Leta et al., 2022).

Nevertheless, recent advancements in computational methodologies, data acquisition technologies, sophisticated modelling theories, and machine learning algorithms present a promising pathway for addressing these persistent challenges. Consequently, a growing body of research has identified several key innovations with significant potential to advance hydrological forecasting skills, including sophisticated artificial intelligence models, model hybridization techniques, quality datasets, ensemble modelling, and advanced optimization methods (Biazar et al., 2025; Hasan et al., 2024; Kumar et al., 2023; Atika et al., 2020). For instance, the quality of hydrometeorological data can be substantially improved by strategically using diverse observational sources, including high-resolution satellite imagery, reanalysis products, and in situ ground-based measurements (Dai et al., 2024; Hasan et al., 2024). Furthermore, the use of data fusion techniques enables the combination of these diverse sources, enhancing their spatial and temporal resolution and yielding more robust datasets for model calibration and validation (Chen et al., 2023; Xiong et al., 2020).

Likewise, hybrid models that synergistically combine process-based hydrological models with data-driven machine learning techniques hold considerable promise for improving hydrological forecasts (Ougahi and Rowan, 2025; Adera et al., 2024; Dai et al., 2024). These approaches are designed to capitalize on their complementary strengths. For example, process-based models help us understand and interpret the physical processes that govern hydrological systems, but struggle to capture complex nonlinear relationships (Bhasme et al., 2022; Yang et al., 2020). Conversely, machine learning (ML) algorithms

excel in identifying complex, nonlinear patterns and quantifying predictive uncertainties (Solanki et al., 2025; Adera et al., 2024). However, pure ML models also suffer from the limitation of lacking physical interpretability (Hasan et al., 2024; Yang et al., 2020). Therefore, by integrating these methodologies, hybrid frameworks effectively address their individual limitations to correct errors, reduce uncertainties, and generate more accurate forecasts.

Additionally, transfer learning, a specialized subset of machine learning, also brings transformative opportunities for predictions in data-scarce regions (Liu et al., 2024; Oruche et al., 2021). This technique involves training a model on a data-rich "source" basin and then transferring the acquired knowledge to a data-scarce "target" basin. This method dramatically reduces dependence on scarce local historical records, enabling faster model deployment and ` improving forecast accuracy in under-gauged basins, offering critical support for water management and disaster preparedness in resource-constrained settings (Qi et al., 2023; Oruche et al., 2021; Ma et al., 2020).

Despite there have been significant advancements and continuous development in models and forecasting techniques, yielding a plethora of methods, no single forecasting approach delivers optimal results under all circumstances and regions (Haddad 2025; Rong et al., 2025; Luo et al.2023; Chong et al., 2022; Zanial et al., 2018; Pagano et al., 2014). Each method exhibits distinct strengths and weaknesses, contingent on specific river basin characteristics or behavioral phases over time. Consequently, there is a pressing need to develop more robust and reliable modelling methods tailored to specific river basins (Larnier & Monnier, 2020; Wang et al, 2018; Razavi & Coulibaly, 2012). Although hybrid modelling and transfer learning techniques have improved forecast accuracy, there is still considerable room for further enhancement, especially in data-scarce or ungauged regions (Solanki et al., 2025; Dai et al., 2024; Oruche et al., 2021). Similarly, integrating process-based models with multi-source data and advanced data-driven methods, such as machine and deep learning, shows promise for improving accuracy; however, these approaches require further research to ensure robustness and generalizability across diverse hydrological contexts (Xiao et al., 2025; Willard et al., 2023).

Therefore, this study aims to address these critical research gaps by developing adaptive, accurate, and scalable forecasting frameworks through the integration of novel approaches, including hybrid modelling, multi-source data fusion, and transfer learning. This dissertation employs the Upper Blue Nile Basin (UBNB) as a case study, where accurate hydrological forecasting is essential for sustainable water management, including transboundary allocation, flood mitigation, dam operations (e.g., the Grand Ethiopian Renaissance Dam), and drought detection to reduce downstream tensions. Despite its strategic importance, hydrological forecasting in the basin remains highly challenging due to its complex terrain, steep elevation gradients, and pronounced spatial heterogeneity, which create strong microclimatic contrasts and highly variable runoff responses (Lakew et al. 2020; Polanco et al., 2017). These complexities are further amplified by diverse seasonal climate regimes, changing land-use patterns, and the increasing influence of climate change. In addition, the basin suffers from high data scarcity, including sparse hydrometeorological monitoring networks and satellite precipitation errors in capturing orographic rainfall in mountainous areas (Mengistu et al., 2020; Nigussie et al., 2019; Abera et al., 2017). These factors contribute substantially to input, parameter, and structural uncertainties in hydrological forecasting models (Takele et al.,2021; Lakew et al. 202). To overcome these challenges, by integrating multi-source data with modular modelling techniques, this study develops a comprehensive framework that enhances predictive capabilities while addressing data scarcity and complex environmental dynamics, ultimately offering scalable, innovative solutions applicable to similar data-scarce regions globally.

1.2. Statement of the Problem

Even though the use of hydrological forecasting is ubiquitous, our ability to generate reliable forecast conditions remains limited due to insufficient ground-based data, particularly in developing regions, and the limited skill of climate and hydrological models in predicting future conditions (Hasan et al., 2024; Ali et al., 2023; Srivastava et al., 2019; Sheffield et al., 2018; Tang et al., 2016; Pagano et al.,2014). The scarcity of data for the calibration, validation, and subsequent use of hydrological forecasting models affects the quality of forecasts and their uptake for decision-making in water resource management.

However, emerging new data sources, such as from low-cost environmental sensors, satellite remote sensing, and reanalysis products, present a significant opportunity to improve the accuracy of hydrological forecasting (Biazar et al., 2025; Ze et al., 2024; Kamyab et al., 2023; Lanza et al., 2022; Sheffield et al., 2018). In addition to that, the skill of forecasting models is improving incrementally as we better understand climate and hydrological systems and as new methods, including machine learning techniques, are developed for collecting and integrating data and models (Ougahi & Rowan, 2025; Slater et al., 2023; Kurtz et al., 2017; Herrnanto et al., 2015).

However, there is no single forecasting model or modelling framework that can be employed for all types of basins around the globe, and each basin requires a unique approach (Ougahi & Rowan, 2025; Slater et al., 2023; Zaher et al., 2020; Sivapragasam et al., 2011). Consequently, there is a persistent need to develop reliable and robust hydrological forecasting modelling frameworks.

In addition, it is not always possible to determine which data is the most accurate across the range of sources, because estimates from each source vary in magnitude, temporal and spatial variability, and reliance on any single source may introduce substantial uncertainty (Hasan et al., 2024; Hartke et al., 2022; Tang et al., 2016; Van der Kwast et al., 2013). Recently, merging data from different sources has emerged as a promising approach to enhance hydrological predictions by leveraging the complementary strengths of individual sources, thereby improving the precision and robustness of the predictions. Furthermore, coupling process-based hydrological models with machine learning techniques through a hybrid framework and transfer learning methods enables the exploitation of the strengths of both approaches and improves forecasting skills, particularly in data-scarce river basins (Solanki et al., 2025; Mohanty et al., 2024; Ali et al., 2023; Ma et al., 2020; Clark et al., 2015; Mendoza et al., 2014). Accordingly, this dissertation aims to advance the accuracy and quality of hydrological forecasting by integrating improved data and models. The developed forecasting frameworks were applied and tested in the Upper Blue Nile River Basin, where forecasting is crucial for water-source planning, allocation, and management amid competing demands.

1.3. Objectives

The general objective of this study is to improve precipitation and streamflow forecasting by integrating machine learning with multi-source data, encompassing physics-informed hybrid modeling and transfer learning techniques, particularly in data-scarce basins.

Specific Objectives

1. To improve spatio-temporal precipitation estimates in data-scarce basins through a machine learning-based multi-source data merging approach
2. To enhance monthly precipitation forecasting by integrating multi-source data with machine learning models
3. To develop a physics-informed hybrid modeling framework that integrates multi-source data, process-based hydrological models, and machine learning techniques to improve monthly streamflow forecasting.
4. To examine the efficacy of transfer learning in enhancing streamflow forecasting for a data-scarce river sub-basin.

1.4. Research Questions

1. Does a machine learning-based multi-source data merging approach improve the spatiotemporal precipitation estimates?
2. Can integrating multiple sources of data with a machine learning model enhance the monthly precipitation forecasting?
3. How can a physics-informed hybrid modeling framework that integrates multi-source data, process-based hydrological models, and machine learning techniques improve monthly streamflow forecasting?
4. How does transferring learning effectively improve streamflow forecasting in data-scarce river sub-basins?

1.5. Significance of the study

In water management communities, hydrological forecasting plays a crucial role in combating water shortages and maximizing water management efficiency. Thus, this research aimed to contribute a significant operational and scientific importance to

forecasting in data-scarce river basins. Operationally, the enhanced spatial-temporal precipitation estimates and improved monthly forecasts are critical inputs for water managers to plan reservoir operation, irrigation scheduling, hydropower generation, water allocations, and agricultural activities.

Moreover, it will provide fundamental information to both operational preparedness agencies and officials on the extension of lead times for extreme event warnings. The advanced streamflow forecasts, especially those enabled by hybridization and transfer learning, directly translate into more accurate and timely early warning systems that save lives, inform infrastructure planning, and manage river basins at national and transnational levels. Furthermore, high-quality streamflow forecasts and their efficient use in water management can yield considerable economic and social benefits across a range of applications.

Scientifically, this study bridges the gap in hydrological forecasting by addressing the fundamental challenge of insufficient data for prediction tasks in poorly gauged basins such as the Upper Blue Nile. The integration of multi-source data (e.g., satellite products and Reanalysis data) with machine learning creates robust proxies for missing ground observations and overcomes their limitations. Furthermore, the application of transfer learning has been a paradigm shift, demonstrating how knowledge from data-rich basins can be systematically used to bootstrap reliable models in hydrologically similar but ungauged catchments, advancing the science of prediction in ungauged basins (PUB).

The methodologies developed in this study for merging data from diverse sources, hybridizing models to capture both physical processes and complex nonlinear relationships, and adopting transfer learning for stream flow forecasting in data-scarce basins provide a valuable blueprint for the field of hydrological forecasting. These findings will contribute to state-of-the-art forecasting frameworks and benchmark case studies, stimulating further innovation and research in multi-source data, model hybridization, and machine learning applications in hydrological forecasting, while closing essential knowledge gaps in similar river basins.

1.6. Dissertation Organization

This dissertation is organized into eight chapters.

Chapter 1 presents the background of research, a general overview of hydrological forecasting, limitations of existing forecasting methods, opportunities and gaps, and briefly discusses the contribution of this dissertation.

Chapter 2 describes the study area, topography, climate, and hydrology. Furthermore, it also briefly states the data used in this dissertation research and their sources.

Chapter 3 focuses on improving the spatiotemporal precipitation estimates using a machine learning-based multi-source data merging approach.

Chapter 4 presents the framework developed to enhance monthly precipitation forecasting by integrating the improved dataset from Chapter 3 with machine learning models.

Chapter 5 introduces a physics-informed hybrid forecasting framework that integrates machine learning techniques with multi-source data and process-based hydrological models to enhance the accuracy of monthly streamflow forecasts.

Chapter 6 details the methodology for improving streamflow forecasting in data-scarce sub-basins using a transfer learning approach, which uses a physics-informed hybrid model and a transfer learning technique based on hydrological similarities between sub-basins.

Chapter 7 summarizes key findings and conclusions regarding the effectiveness of the developed frameworks and their success in improving forecast accuracy. It also offers foundational recommendations for future research.

2. Description of the Study Area and Data

2.1. Location and Geography

The Upper Blue Nile River basin, also known as the Abay basin, is located in the northwestern part of Ethiopia, approximately between Latitudes 7° 40’ and 12° 51’ N, and Longitudes 34 ° 25’ and 39 ° 50’ E (Takele et al., 2021; Polanco et al., 2017). The upper Blue Nile River basin is the largest tributary of the Nile River and covers a drainage area of 175,000 km^2 and contributes about 60% of the annual total Nile discharge (Abebe et al., 2020; Polanco et al., 2017). It covers 17% of Ethiopia’s surface area, generates 43% of the country’s total average runoff, and is home to about 25% of its total population, 40% of the nation’s agricultural production, most of the hydropower potential, and a significant portion of irrigation potential of the country (Lakew, 2020; Taye et al., 2020). The topography is generally rugged and mountainous, ranging from ~490 m a.s.l. near the Sudanese border to ~4300 m a.s.l. on the northeastern part of the basin (Taye et al., 2020; Lakew, 2020; Koukoula et al., 2020). The location and terrain elevation map of the basin is shown in Fig 2.1.

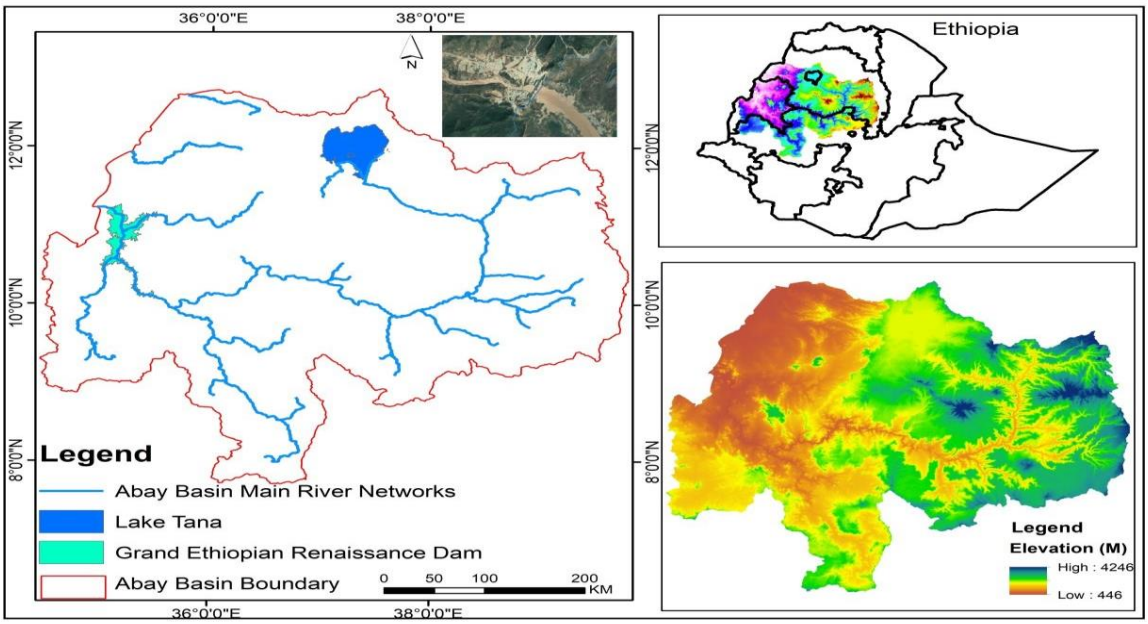


Fig. 2.1: Geographic locations and terrain elevation map of the study area

2.2. Climate

The climate of the Upper Blue Nile River basin varies from humid to semi-arid and highly depends on latitudinal and altitudinal factors (Koukoula et al., 2020). These factors govern a wide range of local climates. For instance, Ayehu et al. (2018) noted that the western lowland areas (near the Ethiopian-Sudanese border) have from hot to arid climates, whereas the highland and eastern mountain summits of the basin are characterized by moderate to even humid-cold climatic patterns. The mean annual temperature varies from 13ºC at the eastern mountain summit to 26°C at the western lowlands of the basin (Lakew et al., 2017). Annual rainfall varies from 400 mm near the Ethiopia–Sudan border to 2,200 mm in the southern areas, such as Didesa and Dabus sub-basins (Roth et al., 2018). According to local climate classification, the basin climate is categorized by three seasons, which are defined as dry season (Bega) from October to the end of February; short rain period (Belg) from March to May; and long rainy period (Kiremt) from June to September. During the major rainy season, which runs from June to September, the basin receives the majority of precipitation. In addition, the basin obtains a significant amount of rain during the short rainy season, which lasts from March to May (Taye et al., 2020; Abebe et al., 2020).

2.3. Hydrology

The Upper Blue Nile Basin's hydrology is shaped by complex topography, a variable monsoon-driven climate, and geological composition. The basin shows strong seasonality, with approximately 80% of its annual stream flow occurring from June to September due to rainfall variability, land use, and reservoir operation (Haile et al. 2017; Teller et al.2013). During the rainy season, the steep slopes and thin, weathered soils generate surface runoff that reaches waterways quickly, producing flashy river discharges with high peak flows and heavy sediment loads (Zargar et al., 2025; Mekonnen et al., 2018). Major tributaries include the Gilgel Abay, Gumara, Ribb, Megech, Jemma, Muger, Guder, Finchaa, Didessa, Dabus, Beles, Anger, and Beshilo rivers.

Groundwater resources are naturally limited due to the dominance of Precambrian crystalline basement rocks of low permeability (Jung et al., 2017). However, localized aquifers in sedimentary basins and fractured zones sustain dry-season base

flow(Mekonnen et al., 2018; Jung et al., 2017). Hydrological connectivity between surface and groundwater varies with seasonal shifts, land cover changes, and human activities such as deforestation, agricultural expansion, and infrastructure development, which alter infiltration, evapotranspiration, and runoff regimes. The basin contributes about 60 % of the Nile's average annual discharge at Aswan (86.5 billion cubic meters) (Yenehun et al., 2021; Mokira et al., 2018). At the Ethiopia–Sudan border, long-term mean discharge is approximately 50 billion cubic meters, historically ranging from 45.6 to 120 billion cubic meters for the whole system (Chakilu et al. 2023; Lemann et al.2018; Jung et al., 2017). Consequently, the basin's hydrology is becoming increasingly surface-runoff-dominated and baseflow-independent(Zargar et al., 2025; Mekonnen et al., 2018).

2.4. Land Use

The land use pattern of the Upper Blue Nile Basin is mainly rain-fed agriculture, which covers more than 50% of the basin area (Zargar et al., 2025; Tewabe et al., 2021; Mekonnen et al., 2018). The remaining part consists of fragmented forests, shrublands, grasslands, wetlands, and major water bodies like Lake Tana and the Blue Nile River system (Jung et al., 2017; Mekonnen et al., 2018). The Upper Blue Nile basin is mostly rural, and the customary land use is subsistence farming and animal husbandry. The major crops grown in the area are teff, maize, barley, wheat, pulses, and the animals raised are cattle, sheep, and goats. The land not suitable for agricultural purposes has been kept as grazing land for sheep and goats. The dominant land-cover class includes cultivated land, while the remaining natural vegetation, like forests, woodlands, and grasslands, have been continuously and adversely degraded by excessive and unsustainable fuelwood collection, rapidly expanding agriculture and population growth, and growth of settlement (Mekonnen et al., 2018; Haregeweyn et al., 2016). Shrubland often serves as a transitional cover type following the degradation of primary forests. While built-up areas have expanded significantly around major urban centers, water bodies have remained largely stable, and bare land is increasingly found in erosion-prone zones and along unstable riverbanks (Leta et al., 2021; Haregeweyn et al., 2016).

Drivers of land-use and land-cover change in the region are primarily linked to population growth and urbanization. In the Lake Tana sub-basin and surrounding watersheds,

agricultural and built-up areas have expanded substantially over recent decades, while natural shrubland, grassland, and bare land categories have declined (Leta et al., 2021). These transitions have significantly increased cultivated land at the expense of natural forests and bushland, a trend that exacerbates land degradation and leads to widespread soil infertility (Haregeweyn et al., 2016). To combat these issues, conservation practices such as terracing, contour plowing, agroforestry, and the construction of stone bunds have been adopted in certain sub-catchments; however, the rate of adoption varies widely depending on land tenure systems, socio-economic factors, and access to modern agricultural inputs (Leta et al., 2021).

2.5. Human Influence and Hydraulic Infrastructure Development

The hydrological system of the Upper Blue Nile Basin has undergone systemic transformation due to human activities. This has led to a shift from natural to engineered hydrological regimes (Mekonnen et al., 2018; Jung et al., 2017). The water budget of the basin has been fundamentally altered by deforestation, agricultural expansion, and urbanization, resulting in a hydrology increasingly dominated by surface runoff (Mekonnen et al., 2018). To address this increased variability, Ethiopia built extensive hydraulic infrastructure to enhance hydropower production, irrigation development, and water supply (Rich et al., 2014; Jung et al., 2017). The most prominent project is the Grand Ethiopian Renaissance Dam, located near the Ethiopia–Sudan border, which is critical for regulating a system where approximately 80% of the annual flow occurs in just four months (Lakew & Moges, 2021; Jung et al., 2017). Other major infrastructures include the Tana-Beles Hydropower Project, Finchaa Reservoir System, Koga Irrigation Scheme, Ribb Dam, and Megech Dam (Mekonnen et al., 2018; Jung et al., 2017). These projects are essential for mitigating historical flow variability, which has ranged significantly from 45.6 to 120 billion cubic meters (Jung et al., 2017). These infrastructures provide substantial socioeconomic benefits, including increased electricity generation and improved water security (Jung et al., 2017; Aich et al., 2014). The UBNB now represents a complex human-water system where hydraulic engineering must mitigate hydrological extremes exacerbated by decades of land degradation (Zargar et al., 2025; Mekonnen et al., 2018). Understanding these anthropogenic influences has become increasingly important for

basin-scale hydrological forecasting and integrated water resources management to ensure long-term water security in a region facing both climatic and developmental pressures (Jung et al., 2017; Aich et al., 2014).

2.6. Monitoring Stations and Hydro-Meteorological Data Challenges

Despite significant development, the Upper Blue Nile Basin remains a functionally data-sparse region (Ali et al., 2023; Wheeler et al., 2020). The core challenges are not merely a lack of measurements but systemic failures in the continuity, spatial representativeness, and physical reliability of existing monitoring infrastructure (Ali et al., 2023; Wedajo et al., 2021).

Temporal Fragmentation and Shallow Historical Records: Systematic hydrological recording in the basin is a relatively recent development, with extensive monitoring initiatives such as the first Abay Master Plan only beginning around 1959 (Gebrehiwot et al., 2011). This leaves a shallow historical benchmark for trend analysis, and while isolated or sporadic records exist from earlier in the 20th century, most historical time series are characterized by severe irregular fragmentation (Julia, 2019; Rientjes et al., 2011). Prolonged data gaps are common due to equipment failure and operational interruptions; for instance, many sub-catchments have continuous records spanning fewer than five years, which is insufficient for traditional model calibration (Dessie et al., 2014; Rientjes et al., 2011). This issue is exacerbated by the basin's extreme seasonality, where approximately 80% of annual discharge occurs during the four-month June–September monsoon period (Amin et al., 2025; Zargar et al., 2025). A sensor failure during this window results in the loss of the year's most critical hydrological signal (Amin et al., 2025).

Spatially Biased Gauge Networks: The distribution of existing rain and stream gauges is systematically biased, with station density significantly higher in the upper highlands than in the lowland regions (Abera et al., 2016). Stations are overwhelmingly concentrated in accessible towns and lower-lying areas along major roads, leaving remote, high-relief sub-catchments with virtually no ground-based coverage (Nigussie et al., 2022; Abera et al., 2016). This spatial imbalance undermines the representation of orographic precipitation, where large topographic variability causes rainfall patterns to vary significantly across the basin (Rientjes et al., 2011). Such bias prevents accurate quantification of runoff

generation and microclimatic variability across elevations that range from 500 to over 4,200 meters (Ayehu et al., 2018; Rientjes et al., 2011).

Infrastructural Deterioration and Measurement Error: Physical monitoring infrastructure is in a state of progressive decline; for example, several automatic stage recorders in the Lake Tana basin ceased operation decades ago (Engida, 2010). During high-flow seasons—the most hydrologically informative periods flashy discharges and heavy sediment loads frequently damage equipment or render manual measurements inaccurate (Dessie et al., 2014; Engida, 2010).

Consequences for Hydrological Analysis: The combined effects of shallow records, fragmented series, and biased gauge placement generate significant uncertainty in flood prediction and drought assessment (Amin et al., 2025; Engida, 2010). These limitations have forced researchers to increasingly integrate sparse ground observations with satellite precipitation products, such as CHIRPS and IMERG, and other remote sensing datasets not merely as an enhancement but as a necessary compensatory strategy for a functionally limited monitoring network (Ali et al., 2023; Nigussie et al., 2022; Wedajo et al., 2021; Ayehu et al., 2018).

2.7. Data

a. Ground-based datasets

Daily meteorological data for the eighty-eight stations were obtained from the Ethiopian National Meteorological Service Agency (NMSA). The data were obtained from 1 January 2010 to 31 December 2017. The spatial distribution of rainfall stations is not uniform, as in the western part, they are very sparse, but fortunately, relatively dense in the topographically complex eastern part of the basin. All stations were made to pass on strict data quality checks for consistency, missing values, data gaps, and outlier tests. After the data quality check, only forty-five stations were qualified for further use in the study. Stations with large data gaps, missing values, and inconsistent records were rejected.

b. Satellite and Reanalysis Precipitation Datasets

In this study, we used three satellite and one reanalysis precipitation products. The four products were selected to capture the full diversity of precipitation estimation

methodologies available for the Upper Blue Nile Basin and represent distinct retrieval methodologies without redundancy, ensuring complementary error structures. IMERG represents advanced multi-satellite microwave-IR fusion; PERSIANN-CCS provides a machine learning-based IR-only alternative; CHIRPS offers long-term satellite-gauge blending calibrated for African conditions; and ERA5 represents the most advanced reanalysis product, selected for its superior spatial resolution (0.25°—the finest among all reanalysis products, including MERRA-2, JRA-55, and ERA-Interim), hourly temporal resolution, demonstrated performance in Ethiopian evaluations, and open data accessibility.

i. CHIRPS

The Climate Hazards Center InfraRed Precipitation with Station data (CHIRPS) is a quasi-global rainfall dataset provided by the Climate Hazards Group at the University of California, Santa Barbara (UCSB) (Funk et al., 2015). This data set is available at a global scale from 1981 to the near present, spanning 50°S-50°N (all longitudes) at 0.05° resolution. The CHIRPS data is estimated for each month based on station data, averaged geostationary satellite thermal IR observations, TRMM 3B42 elevation, latitude, and longitude, and provides a high-resolution imagery system (gridded precipitation datasets) that enables drought monitoring and hydro-climatic system analysis (Funk et al., 2015). The CHIRPS data is characterized by its low latency, high resolution, low bias, and long-term records, which make it suitable for different applications (Belay et al., 2019; Dinku et al., 2018). A detailed description of CHIRPS products can be found in (Funk et al. 2015**).** For this study, the data were accessed from https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_daily/tifs. The datasets were downloaded with the same period as gauge records, i.e., 2010- 2017.

ii. IMERG

The Integrated Multi-satellitE Retrievals for GPM (IMERG) was launched in 2014 jointly by the National Aeronautics and Space Administration (NASA) and the Japan Aeronautics and Exploration Agency (JAXA). It combines data from a group of partner satellites to offer space-based precipitation estimates over the majority of the Earth's surface (Bogerd et al., 2021; Tan et al., 2019). The GPM Core Observatory satellite has an orbit

encompassing 65°S to 65°N, and the algorithm is designed to inter-calibrate, merge, and interpolate all available satellite microwave retrievals, microwave-calibrated IR satellite estimates, and GPCC gauge measurements (Bogerd et al., 2021; Derin et al., 2021). In the latest Version 06 releases of IMERG, the algorithm fuses the early precipitation estimates collected during the era of TRMM satellite missions (2000-2015) with recent estimates during the operation of the GPM satellites (2014 – to date) (Huffman et al., 2019). The IMERG dataset is available at 0.1° spatial resolutions and in three forms, namely, IMERG early, IMERG late, and IMERG final runs with a latency of 6 hours, 18 hours, and 18 months, respectively. Moreover, these products are available at half-hour, daily, and monthly temporal resolutions. A detailed description of GPM IMERG was given in Huffman et al. (2019). In this study, IMERG early data for the period 2010-2017 were accessed from the NASA website:

https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGDF.06.

iii. **PERSIANN-CCS**

The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Cloud Classification System (PERSIANN-CCS) is a high-resolution real-time satellite precipitation product developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI) (Anjum et al., 2022; Gunathilake et al., 2021). PERSIANN-CCS is a cloud-patch-based algorithm where the features are extracted from the cloud coverage under specified temperature thresholds using infrared (10.7 mm) geostationary satellite imagery (Sabir et al., 2016; Hong et al., 2007). A detailed description and algorithm of PERSIANN-CCS were presented in Nguyen et al. (2018) and Karbala et al. (2017). The PERSIANN-CCS data set is freely available (2003 to date) at a (0.04° × 0.04°) resolution. In this study, the PERSIANN-CCS gridded dataset was accessed from the CHRS Data Portal http://chrsdata.eng.uci.edu at a daily resolution.

iv. ERA5

The ERA5 product is the most recent version of the European Centre for Medium-Range Weather Forecasts (ECMWF) dataset, that offers hourly precipitation data with a 0.25° * 0.25° resolution from 1979 to the present (Tarek et al., 2020; Crossett et al., 2020). ERA5

products were developed through advanced data assimilation and modeling techniques in combination outputs from a Numerical Weather Prediction (NWP) model and key atmospheric variable measurements (Jiao et al., 2021; Hersbach et al., 2020). Furthermore, ERA5 incorporates more historical observations, particularly satellite data, estimate accurate atmospheric conditions. Details about ERA5 products can be found in Hersbach et al. (2020) and Bell et al. (2019). For this study, ERA5 precipitation data were downloaded for the period 2010- 2017 from the Copernicus data portal https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset&text=ERA.

c. Ancillary data

DEM

The NASA Shuttle Radar Topographic Mission (SRTM) version 4 digital elevation data model (DEM) at 90 m resolution was downloaded from the USGS National Map Seamless Data Distribution System.

d. Large-Scale Climate Indices

To develop a robust hydrological forecasting framework for the Upper Blue Nile Basin, a set of large-scale climate indices was selected as potential predictors. The selection was guided by their documented teleconnection to regional hydroclimatic variability and hydrological processes, as reported in previous studies (e.g., Habtegebreal et al., 2021; Mohamed et al., 2022; Zaroug et al., 2014). These include ENSO-related indices (Niño 1+2, Niño 3.4, ENSO Precipitation Index, SOI), and QBO (which contributes significantly to the summer rainfall from June to August), both of which were shown to modulate rainfall and streamflow in the basin (Mohamed et al., 2021; Le et al., 2020). Further indicators characteristic of longer time oscillations (PDO, AMO), region-specific indices (including the Sahel Precipitation Index), atmospheric pressure patterns (NAO, AAO), thermal gradients in the Atlantic (TSA, AMM, TNA), a warming index in the Indian Ocean (IOBW), and the associated Dipole Mode Index (DMI) – which controls monsoon rainfall and precipitation extremes – were added owing to their larger teleconnections to the hydro-climatology of East Africa and the Upper Blue Nile Basin (Shiferaw et al., 2023; Habtegebreal et al., 2021; Zaroug et al., 2014).

2.8. Meteorological Data Quality Assessment and Station Selection

Daily meteorological data were obtained from the National Meteorological Service Agency (NMSA) for 88 rainfall stations located within and around the Upper Blue Nile Basin (UBNB). The available records covered the study period and were subjected to a comprehensive quality-control assessment before their use in hydrological forecasting and analysis in this dissertation. Given the challenges associated with hydrometeorological observations in the Upper Blue Nile Basin, including missing records, uneven spatial coverage, station malfunction, and measurement inconsistencies, a rigorous data-screening procedure was implemented to ensure the reliability and representativeness of the selected stations.

i. Completeness Check

The first stage involved evaluating the completeness of rainfall records for all 88 stations. Missing daily observations were identified and aggregated to monthly and annual scales. The percentage of missing data for each station was calculated as:

$$\text{Missing Data Percentage} = \frac{\text{Number of Missing Records}}{\text{Total Number of Records}} * 100 \qquad \text{Eq. 2.1}$$

Stations with excessive missing observations were excluded from further analysis. Following commonly adopted hydrometeorological data-quality standards, stations having less than 20% missing data during the study period were considered acceptable for subsequent analyses. As a result, 45 stations satisfied the completeness criterion and were retained, while the remaining 43 stations were excluded due to excessive missing observations and poor data continuity. The selected 45 stations provided adequate temporal coverage and represented the major climatic and topographic zones of the Upper Blue Nile Basin.

ii. Consistency Check

Consistency was evaluated using the Double Mass Curve (DMC) technique. Cumulative rainfall of a target station was plotted against the cumulative mean rainfall of neighboring stations. Linear relationships indicate consistent records.

iii. Range and Physical Plausibility Tests

Range checks were conducted to identify physically unrealistic observations. Range checks were conducted to identify physically unrealistic observations. The following criteria were applied: negative rainfall values were considered erroneous and removed; maximum temperature values lower than minimum temperature values were identified and corrected where temperature records were available; extremely high rainfall values inconsistent with local climatological conditions were flagged and verified using neighboring station observations.

iv. Verification of Station Metadata

Station metadata, including geographic coordinates, elevation, and station identifiers, were verified using available station inventory records obtained from the NMSA. The spatial distribution and locations of the river gauges and meteorological stations are presented in Figure 2.2.

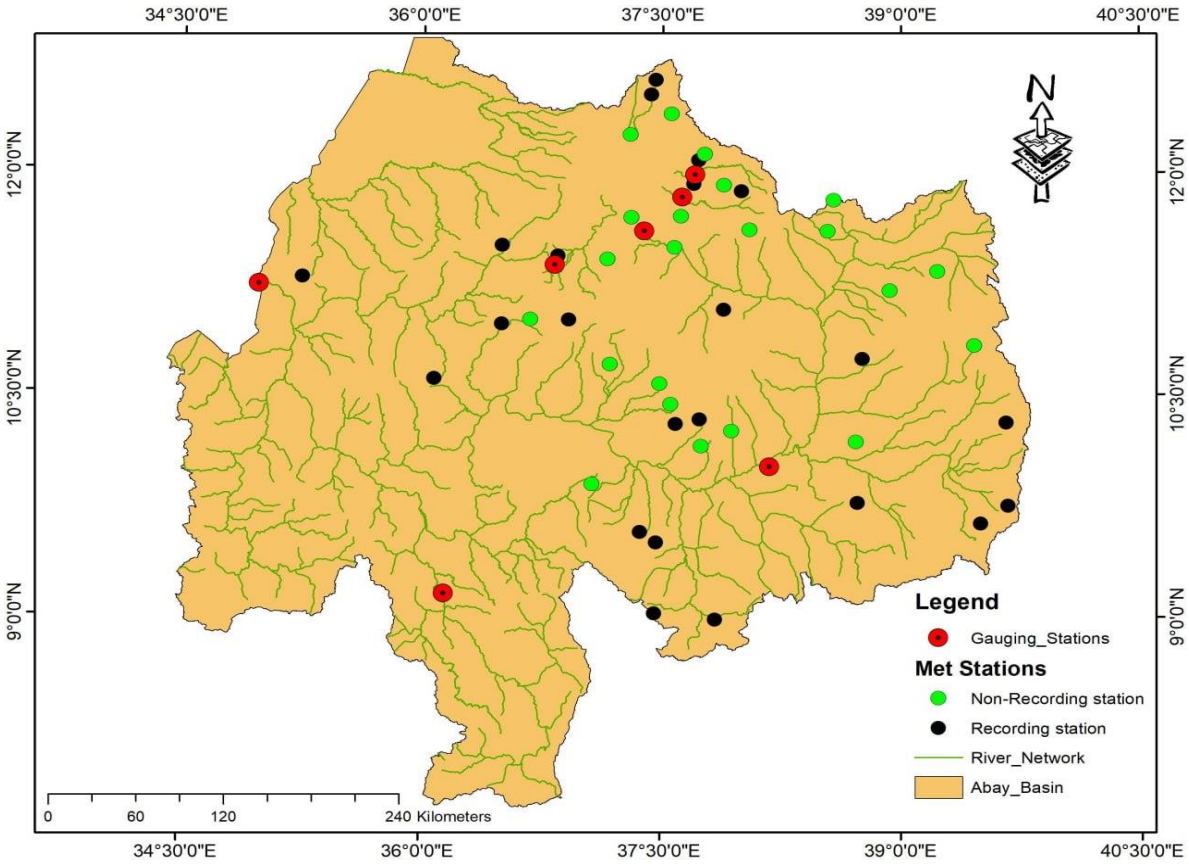


Fig.2.2 Map of River Gauge and Meteorological Stations

v. Missing Data Infilling

Following quality assessment, missing rainfall observations within the selected stations were estimated using the Normal Ratio Method, which is widely recommended when the mean annual rainfall of neighboring stations differs substantially from that of the target station. The Normal Ratio Method estimates missing rainfall as:

$$P_X = \frac{1}{N} = \sum_{i=1}^{n} (\frac{N_X}{N_i}) P_i \qquad \text{Eq.2.2}$$

where:

- P_x = estimated rainfall at the target station,
- P_i = observed rainfall at neighboring station ii,
- N_x = long-term mean annual rainfall at the target station,
- N_i = long-term mean annual rainfall at neighboring station ii,
- n = number of neighboring stations used in the estimation.

vi. Final Station Selection

Following all quality-control procedures, 45 stations were retained. These stations had less than 20% missing data, consistent records, verified metadata, realistic observations, and adequate spatial coverage.

3. Improving Spatio-Temporal Precipitation Estimates in Data-Scarce River Basins: An Application of Machine Learning-Based Multi-Source Data Merging

Abstract

Reliable precipitation estimates are crucial for understanding hydrological processes and engineering applications. Precipitation data can be collected from different sources, such as gauges, satellites, and reanalysis products. However, each source has its own pros and cons, including uncertainties. Merging multi-source data may help improve spatio-temporal precipitation estimates and reduce uncertainties. This study aimed to improve precipitation estimates in the data-scarce river basins, using the upper Blue Nile River basin as a case study, by blending gauge records with IMERG early, CHIRPS, PERSIANN-CCS, and ERA5 products. A machine learning random forest-based merging procedure (RF-MEP) was applied as a merging technique. Daily measured multi-year (2010–2017) precipitation data were used from daily measurements. Several quantitative and qualitative evaluation metrics were applied to assess the performance of the merged product. The performance of individual unmerged rainfall products, namely, CHIRPS, PERSIANN-CCS, IMERG, and ERA-5, was evaluated and compared with the Random Forest-based merged product (gauge, satellite, and reanalysis products) for modeling the spatial–temporal pattern of precipitation in the Upper Blue Nile Basin. Compared with reference data, the merged product exhibited a consistent improvement in terms of Index of Agreement, CC, MAE, NSCE, KGE, and RMSE over the original precipitation products. Moreover, POD, FAR, and CSI also exhibited improvement over the unmerged product. Therefore, the merged precipitation product showed a better spatio-temporal representation and accuracy, which can be successfully used for scientific studies as well as operational purposes in data-scarce river basins.

Keywords: Rain gauge records, Satellite precipitation product, Reanalysis precipitation product, Machine learning

3.1. Introduction

As one of the crucial components of the hydrologic cycle, precipitation is vital for scientific studies as well as engineering applications. Its accurate estimation both in space and time has paramount importance for setting policy and water-based decision-making in different fields (An et al., 2020; Mastrantonas et al., 2019; Kid et al., 2017). For instance, better water resources management and natural disaster prevention plans are dependent on the quality of precipitation estimates, including intensity, magnitude, and spatiotemporal variability (He et al., 2020; Gebremedhin et al., 2020; Hénin et al., 2018). Despite the need for quality precipitation data at reasonable spatial and temporal (e.g., daily) resolutions, it remains a challenging task with a lot of uncertainty (Wu et al., 2020; Gebremedhin et al., 2020; Hu et al., 2019; Hazra et al., 2019). This problem is exacerbated in developing regions where technology and measuring instruments are substantially limited.

Precipitation data, traditionally obtained from gauge stations but in recent decades from weather radar, satellite-based, and reanalysis precipitation products, have been recognized as subordinate sources of precipitation data (De Vera et al., 2021; Kim and Han, 2021; Nguyen et al., 2021; An et al., 2021; Beck et al., 2019). According to Zhang et al. (2021) and Prakash et al. (2019), rain gauge observations are considered the most accurate method to estimate rainfall and are used as the "ground truth" for a wide variety of applications; however, many other studies, such as Gebremedhin et al. (2020), Kidd et al. (2017), and Mohammad-Taghi et al. (2017) have indicated that gauge data by itself have limitations and uncertainties. The obvious limitations of gauges include that they are restricted by point-scale observations (Moraux et al., 2021) but are also subjected to uncertainties and erroneus readings due to wind and evaporation effects (Sun et al., 2018; Mohammad-Taghi et al., 2017). Moreover, uneven and scarce distribution of gauges owing to different factors, such as complex terrain and human geography, limits their ability to capture the spatial and temporal variability of precipitation (Yao et al., 2020; Beck et al., 2019).

Recent satellite precipitation products are considered an attractive option for acquiring rainfall estimates in the absence of or complement to ground measurements and are widely used as forcing data, particularly for regions with data-scarce, ungauged, and complex terrain conditions (Nguyen et al., 2021; Atiah et al., 2020; Beck et al., 2019; Sun et al., 2018). These products are advantageous in terms of temporal and spatial coverage and cost-

effective alternative sources for different applications. Even if satellite precipitation products show ample benefits, they are also affected by substantial errors and uncertainties (Wiwoho et al., 2021; Trinh-Tuan et al., 2019; Sun et al., 2018). They suffer from accuracy issues resulting from the indirect relationship between the remotely sensed signals and rainfall rate, as well as gaps in temporal sampling (e.g., REF). For example, Ma et al. (2018) and Fenta et al. (2018) pointed out the low accuracy of satellite precipitation products at finer temporal scales and variable performance in different climatic and topographic regions (Gohzat et al., 2021; Ullah et al., 2019).

Reanalysis precipitation products are another source of rainfall data, especially for long-term climate studies, and are generated by continuously assimilating data from various satellite meteorological observations into an atmospheric weather model (Hamed et al., 2021; Bhattacharya et al., 2020; Bhuiyan et al., 2019). Although these products are being used widely for different purposes and show enormous benefits, they are affected by irregularly distributed observational stations and orographic effects (Chen et al., 2021; Luo et al., 2019; Essou et al., 2017). Lately, studies have also revealed significant uncertainties in reanalysis datasets throughout warm and cold seasons, in particular over complex terrain, which can limit their use in water resources applications (Lakew, 2020; Luo et al., 2019; Liu et al., 2018).

Even though significant advances have been made in developing more sophisticated instruments, methods, and models, precipitation data remains a major source of uncertainty in various applications (Scanlon et al., 2018; Sun et al., 2018). There is, therefore, a need to improve precipitation data for different purposes (Hu et al., 2019). Recently, merging different data sources in a way that takes the complementary strengths of various precipitation estimates has been considered as one way of improving precipitation estimates at fine spatiotemporal resolutions. The common merging approach is blending gauge observations with an individual satellite precipitation product, but this method does not take advantage of the complementary strengths of various satellite and reanalysis precipitation products (Zhang et al., 2021; Baez-Villanueva et al., 2020; Chen, 2020). Lately, along with the development of machine learning methods, merging multi-source data has emerged as a promising approach to improve precipitation estimates at finer

resolutions (Nguyen et al., 2021; Fan et al., 2021; Zhang et al., 2021; Baez-Villanueva et al., 2020).

The noticeable benefits of machine learning-based merging techniques include great flexibility to incorporate multiple types of explanatory variables, such as topography (Nguyen et al., 2021), their ability to deal with non–linear relationships (Hengl et al., 2018), and the capability of merging multiple products (Zhang et al., 2021; Baez-Villanueva et al., 2020). For instance, many studies (e.g, Yin et al., 2021; Zhang et al., 2021; Wu et al., 2020; Yazdandoost et al., 2020; Bhuiyan et al., 2019) applied machine learning-based merging and obtained a promising quality improvement in the merged products. Furthermore, in comparative studies such as (Baez-Villanueva et al., 2020; Nguyen et al., 2021), it has been shown that machine learning-based approaches significantly outperform other merging techniques. In this study random forest-based merging procedure (RF-MEP) was implemented as a merging technique for multi-source products. RF-MEP has been applied in several watersheds successfully; however, not sufficiently in data-scarce river basins (Nguyen et al., 2021). It is therefore this study intends to explore the applicability and reliability of the machine learning-based merging technique (RF-MEP) in merging multi-source products in data-scarce basins using the upper Blue Nile River Basin as a case study.

The upper Blue Nile River Basin accounts for the largest share of Nile River flow and plays a key role in the hydro-politics of the Nile region. The basin is currently characterized by rapid population growth, hunger, rising energy demand, and persistent conflicts among the Nile River basin's riparian countries (Allam and Eltahir, 2017). Aside from that, to meet the demands of their numerous fast-expanding projects, countries in the region are vying for equitable use of water resources. In the face of these contradictory objectives, reasonable allocation and integrated water resources management are essential to address these conflicts of interest. This can be achieved by understanding the basin's hydrological processes and hydro-climatic factors across diverse spatial and temporal scales (Dile et al., 2018). Meanwhile, the availability of long-term hydro-climatic data in the Upper Blue Nile River Basin is a key challenge for such water studies and technical applications (Khadim et al., 2020; Koukoula et al., 2020; Dile et al., 2018; Abera et al., 2017). In the Upper Blue Nile River basin, most parts of the basin are ungauged, and even the available gauges are of low quality and lack continuous monitoring and maintenance, which makes the

reliability of data highly questionable (Eldardiry and Hossain, 2019; Dile et al., 2018; Polanco et al., 2017).

Furthermore, uneven distribution of existing gauges (cannot often capture the spatial variability of rainfall), missing records (from consecutive days to months and years), and errors due to instrument defects are typical problems that pose a great challenge to examine the basin's natural processes (Bayissa et al., 2021; Abebe et al., 2020; Polanco et al., 2017). In data-scarce river basins, such as the Upper Blue Nile River Basin, where observations are not available or data are of poor quality, a method of improving data is crucial. This study therefore, aimed to improve the spatio-temporal pattern of precipitation through the application of merging multi-source information (rain gauge, satellite, and reanalysis products) using machine learning techniques based on Random Forest (RF model) in the data-scarce Upper Blue Nile Basin.

3.2. Methodology

a. RF-MEP

In this study, a machine learning (Random Forest-based MErging Procedure (RF-MEP)) (Baez-Villanueva et al., 2020) algorithm was applied to improve the spatiotemporal precipitation estimates. This method combines data from ground-based measurements, precipitation products (satellite and reanalysis), and topography-related features. According to Baez-Villanueva et al. (2020), RF-MEP is based on three key assumptions: i) Precipitation measurements from rain gauge stations are accurate at the point scale; ii) satellite and reanalysis precipitation products are generally biased but contain useful information about the spatio-temporal patterns of precipitation; and iii) combination of different precipitation products and rain gauge data can provide a better representation of the spatio-temporal variability of precipitation than any single product.

RF-MEP uses the random forest algorithm to predict the spatial distribution of precipitation by merging data from different sources at a selected temporal scale (e.g., daily, monthly, or annual). Random Forest is known as an ensemble machine learning method that can be used for supervised classification and regression tasks (Zhang et al., 2021). It is based on constructing numerous decision trees, and the final prediction results are taken as the average of individual predictions (Zhang et al., 2021; Hengl et al., 2018; Biau and Scornet,

2016). The individual predictions are generated from a user-defined number of decision tree models based on bootstrap samples using covariates as predictors (Baez-Villanueva et al., 2020; Hengl et al., 2018).

The random forest algorithm is characterized by its capacity to alleviate overfitting problems (Zhang et al., 2021) and gives greater flexibility to work with a small sample size or in the presence of missing data and outliers (Nguyen et al., 2021; Biau and Scornet, 2016). It is also regarded as accurate and easy to implement for merging multiple precipitation products or multiple covariates (Nguyen et al., 2021; Baez-Villanueva et al., 2020). Furthermore, RF can handle high-dimensional data due to its high feature spaces and gives the best performance even for some explanatory variables that do not add information to the predictions (Hengl et al., 2018; Biau and Scornet, 2016). Most notably, compared to other methods, random forest-based merging techniques have exhibited superior efficacy and flexibility (Zhnag et al., 2021; Nguyen et al., 2021; Baez-Villanueva et al., 2020). The random forest model is given by Eq.3.1

$$\Lambda B \quad \theta\ (x) = \frac{1}{B}\sum_{b=1}^{B} t_b^*\,(x) \qquad \text{Eq.3.1}$$

Where ΛB θ is the final prediction; b is the individual bootstrap sample; B is the total number of trees; and t^*_b is the individual decision tree. According to Baez-Villanueva et al. (2020) RF regression models have three parameters to specify: i) the number of regression trees (set at 2000); ii) the number of variables randomly sampled at each decision split (set at one third of the number of covariates); and iii) the node size (i.e., the minimum number of observations per node; set at 5).

b. Merging procedure

The selected satellite precipitation products (SPP), reanalysis precipitation product (RPP), and the digital elevation model (DEM) were downloaded at different spatial resolutions, and they were resampled to a common 0.05° resolution based on the nearest neighbor interpolation approach. The gauge data were randomly divided into training and validation sets (i.e., out of 45 stations that qualified for further use, 20% for validation of the merged product and 80% for the training of the model) in order to capture the spatial variability of

the precipitation patterns. Besides that, the RF-MEP model considers spatial geographical covariates (e.g., coordinates), Euclidean distances to sampling locations, or downslope distances, and process-based covariates (e.g., elevation, rate of elevation change, or aspect) to account for spatial autocorrelations and precipitation distributions in the blending process, since they have a great influence on the quality of final merged products (Nguyen et al., 2021; Baez-Villanueva et al., 2020). Finally, a daily merged high-resolution (0.05°) precipitation data set was produced. For this study, the merging of data was performed using the RFmerge package developed by Zambrano-Bigiarini et al. (2020) in the R environment 3.5.0 (R Core Team, 2018). Details about the RF –MEP and RFmerge can be found in Baez-Villanueva et al. (2020) and Zambrano-Bigiarini et al. (2020), respectively. Fig.3.1 summarizes the steps applied in the merging procedure of this study.

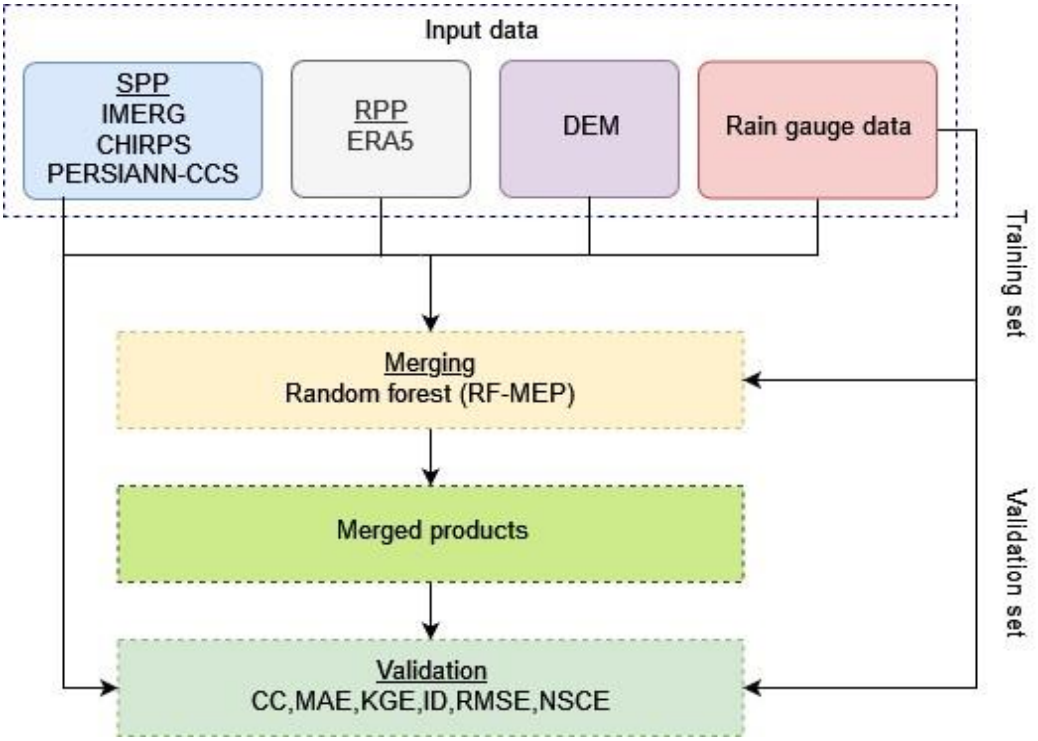


Fig. 3.1. A schematic presentation of the merging procedure

c. Validation metrics

Validation metrics indicate the nearness of the estimated data to an acceptable standard, which in this case is the rainfall data from the gauge stations. Daily precipitation values of the merged, original satellite, and reanalysis precipitation products of given pixels were

extracted using the nearest neighbor method to the point where gauge stations are located. In this analysis, two types of statistical metrics were used for validation purposes, namely, quantitative and qualitative validations (Table 3.1). The quantitative metrics, including Pearson correlation coefficient (CC), Nash-Sutcliffe coefficient of efficiency (NSE), Index of Agreement (d), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Kling-Gupta Efficiency (KGE), were used to measure the accuracy of rain estimates, errors, and biases. On the other hand, the qualitative validation metrics probability of Detection (POD) (measures the ratio of the precipitation product detected rain events to the observed rain events), False Alarm ratio (FAR) (measures the ratio of detected rain events but no observed rain to all detected rain events) were used to reflect the capability of products to detect rainfall events.

Table 3.1 Validation metrics and equations

Validation metrics	Equation	Range of values	Best values
Pearson Correlation Coefficient (CC)	$CC = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\ \sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}}$	[-1,1]	1
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} * \sum_{i=1}^{n} \lvert O_i - P_i \rvert$	[0, + ∞)	0
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(o_{,i} - p_i)^2}{n}}$	[0, + ∞)	0
Kling-Gupta Efficiency (KGE)	$KGE = 1 - \sqrt{(CC-1)^2 + (\frac{cd}{rd} - 1)^2 + (\frac{cm}{rm} - 1)^2}$	(-∞,1]	1
Index of Agreement	$d = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(\lvert P_i - \bar{O} \rvert + \lvert O_i - \bar{O} \rvert)^2}$	[0,1]	1
Nash-Sutcliff Coefficient of Efficiency (NCE)	$NSE = 1 - \frac{\sum_{i=1}^{n}(o_i - p_i)}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$	(-∞,1]	1
Probability of Detection (POD)	$POD = \frac{\sum A}{\sum(A+B)}$	[0,1]	1
False Alarm Ratio (FAR)	$FAR = \frac{\sum C}{\sum(A+C)}$	[0,1]	0
Critical Success Index (CSI)	$CSI = \frac{\sum A}{\sum(A+B+C)}$	[0,1]	1

Notation: where n is the total number of observations, O_i is the observation value (gauge records), and P_i is the modeled value (RPP, SPP, and merged products), and $\bar{o}$ and $\bar{p}$ are the average values of observed and modeled values, respectively. rm is the average of observed values, cm is the average of forecasted values, rd is the standard deviation of observations, and cd is the standard deviation of forecasted values. A is the number of rainfall events detected by both the gauges and the SPPs; B and C are the number of rainfall events detected by the gauges and the number detected by the SPPs. respectively.

3.3. Results and Discussion

3.3.1. Results

3.3.1.1. Performance of unmerged precipitation products

This study provides a comprehensive evaluation of individual satellite and reanalysis precipitation products before merging, using widely used statistical performance metrics.

A total of three different satellite products, CHIRPS, IMERG early, and PERSIANN-CCS, and one reanalysis precipitation product, ERA5, with different spatial resolutions, were examined for their ability to estimate rainfall data at a pixel level, using observations from nine-gauge stations represented as PGS (Precipitation Gauge Stations). These stations were not used in the merging process and were kept for validation purposes. The statistical evaluation results are summarized in the tables and the graphs below.

The performance of both satellite-derived and reanalysis precipitation products exhibits relative biases over the region of interest in the Upper Blue Nile River basin. For instance, the values of daily statistical metrics range from 2.464 to 5.835 mm/d for MAE and from 6.55 to 10.78 mm/d for RMSE. The values of both MAE and RMSE showed high variability across the validation stations. For example, all precipitation (both gauge and satellite) exhibited the highest MAE at gauge station PGS43. In general, CHIRPS consistently showed the highest error, followed by ERA5 and IMERG, while PERSIANNN-CCS showed the lowest error distribution across all stations. These results are comparable to those of An et al. (2020) & Nguyen et al. (2021) in China and Korea, respectively.

Meanwhile, in terms of RMSE, all precipitation products showed variable performance, and IMERG performed best at most stations. ERA5 and CHIRPS had consistently shown approximate values across all stations based on RMSE. The RMSE value of PESIANN-CCS highly fluctuated, as referred to in Table 3.2, and it was observed with the lowest error at most of the stations. According to the NSCE results, both satellite and the reanalysis product data didn’t demonstrate any notable performances. Most results were negative and close to 0, which indicated that the climatic daily average may be an equivalent or better estimator of observed data than the later satellite and reanalysis estimates. This is a weak and unacceptable performance of these products, and the results were consistent with previous studies, such as Gadouali & Messouli (2020). The lower NSE values reported by all satellite products are likely due to NSE's sensitivity to outliers in the data at daily temporal resolution.

Table3.2. Quantitative metrics results based on daily scale measurements

		MERGED	CHIRPS	IMERG	ERA5	PERSIAN-CCS
PGS 8	MAE	2.064	3.213	2.855	3.171	2.568
	NSCE	0.399	-0.464	-0.213	-0.412	-0.032
	RMSE	5.107	7.973	7.258	7.832	6.695
PGS 9	MAE	2.153	4.236	3.731	4.177	3.448
	NSCE	0.63	-0.175	0.031	-0.138	0.167
	RMSE	5.263	9.381	8.518	9.237	7.904
PGS 17	MAE	1.62	3.251	3.304	3.228	2.957
	NSCE	0.743	-0.031	-0.146	-0.001	0.097
	RMSE	3.633	7.272	7.666	7.168	6.807
PGS 18	MAE	2.21	3.432	4.144	3.437	3.857
	NSCE	0.562	-0.125	-0.662	0.113	-0.417
	RMSE	4.602	7.374	8.965	6.552	8.281
PGS 21	MAE	2.5	3.203	2.999	3.195	2.464
	NSCE	0.349	-0.487	-0.786	-0.47	-0.154
	RMSE	5	7.556	8.283	7.518	6.662
PGS 34	MAE	1.668	3.05	2.879	2.625	2.537
	NSCE	0.651	-0.33	-0.47	0.252	-0.076
	RMSE	3.834	7.485	7.868	5.612	6.732
PGS 37	MAE	2.964	3.027	3.38	2.962	3.001
	NSCE	0.246	0.009	-0.295	0.042	-0.007
	RMSE	6.101	6.993	7.991	6.876	7.052
PGS 43	MAE	4.674	5.835	5.348	5.657	5.327
	NSCE	0.331	0.039	0.176	0.097	0.148
	RMSE	8.991	10.778	9.98	10.445	10.153
PGS 46	MAE	2.888	4.053	4.795	3.971	4.386
	NSCE	0.495	-0.032	-0.405	0.143	-0.229
	RMSE	5.961	8.52	9.943	7.764	9.302

The results presented in Fig. 3.2a and b also indicated the performance of gauge and reanalysis precipitation products in terms of Pearson correlation coefficient (CC) and Index of Agreement (d). The Pearson coefficient of correlation reflects the degree of a linear relationship between satellite and reanalysis precipitation products and gauge observations, whereas the index of agreement shows the degree of error in model prediction. In terms of the index of agreement, IMERG outperformed other satellite precipitation products and topped at all stations. The d values for IMERG ranged from 0.485 to 0.741, followed by ERA5, which also showed a marked agreement with gauge measurements, and its values

fluctuated between 0.551 and 0.664 across stations. On average, CHIRPS exhibited the lowest agreement with ground station records. Moreover, CHIRPS exhibited the bottom-most agreement with a value of 0.356 at station PGS8. Most importantly, despite the ability of the index of agreement to detect additive and proportional differences in the observed and simulated means and variances, d is also excessively sensitive to extreme values due to the squared differences. This could be the possible reason for the fluctuation of d values of a single precipitation product across stations. The presence of a few outliers either in precipitation products or ground gauge records notably affected the index of agreement values.

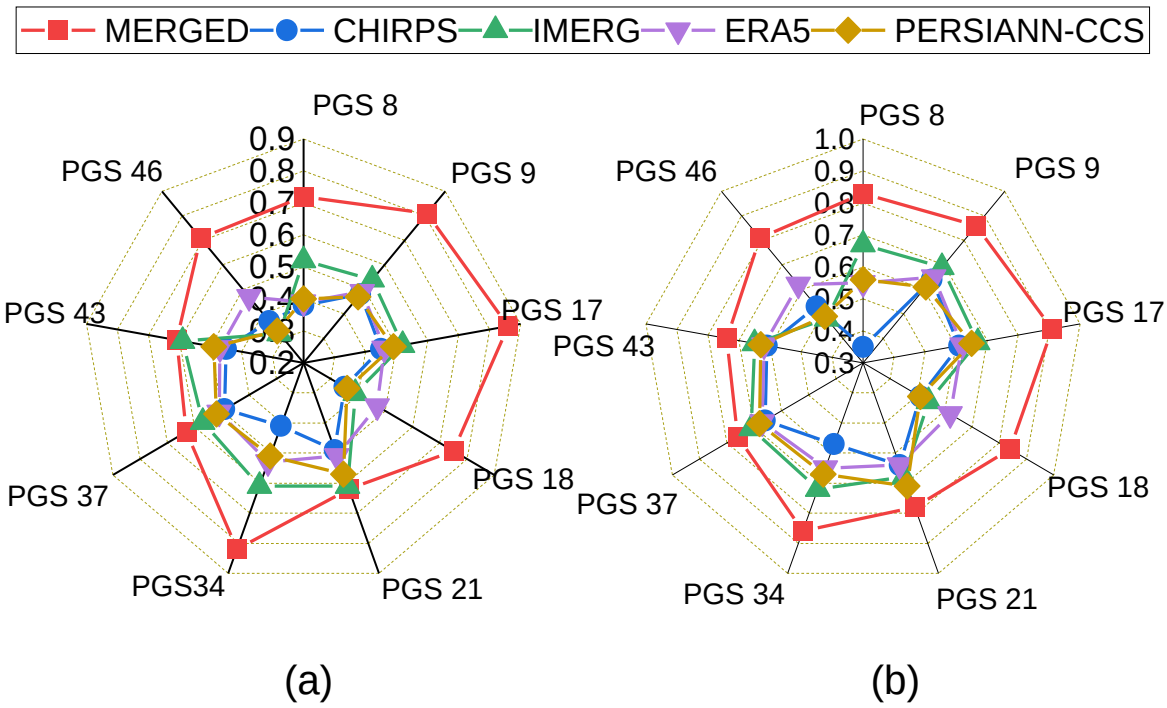


Fig 3.2. (a) Pearson coefficient of correlation for respective precipitation gauge station based on daily scale measurements (PGS), (b) Index of Agreement for respective precipitation gauge station (PGS) based on daily scale measurements

Regarding the Pearson correlation coefficient, again IMERG had the highest Pearson correlation coefficient both in terms of individual stations (0.604) as well as averages of all stations (0.516), followed by ERA5 re-analysis product, PERSIAN CCS, and CHIRPS with average values of Pearson correlation coefficient (0.47, 0.45, 0.43), respectively. Additionally, in most of the stations used for validation, IMERG presented > 0.5 values,

which could be considered as a good linear relationship with ground measurements. However, other precipitation products, both satellite and reanalysis, consistently showed a weak correlation with most values falling between 0.321 to 0.495. The variation of the Pearson correlation coefficient at most stations, even with a single precipitation product, highly depended on the quality of the station datasets.

As a way of summarizing, Fig. 3.3 illustrates the performance of precipitation products based on Kling-Gupta efficiency, which facilitates the analysis of different components (correlation, bias, and variability). The box plot showed the median KGE value for each satellite and reanalysis product. Based on KGE, IMERG showed high fluctuations from bottom to topmost across stations. The median KGE values for the IMERG ranges were from 0.124 to 0.576. Meanwhile, the ERA5 reanalysis product exhibited the highest KGE value on average, whereas the CHIRPS product recorded the lowest performance based on the KGE values. Comparison with existing UBNB studies indicates that CHIRPS performance is highly scale-dependent. Bayissa et al. (2017) and Ayehu et al. (2018) reported good to excellent performance at dekadal and monthly scales; however, Ayehu et al. (2018) documented high false alarms (FAR = 0.31) and 31% overestimation of rainfall frequency—weaknesses consistent with our daily findings. Ahmed et al. (2024) more directly support our results, reporting that at the daily scale, CHIRPS achieves only moderate KGE (0.3–0.6) and underestimates rainfall variability ($\gamma < 0.25$). The primary reason for this discrepancy is temporal scale: Bayissa et al. (2017) deliberately avoided daily evaluation, stating that "a daily comparison was not carried out because of the weak performance reported in previous studies." Thus, our study, along with Ahmed et al. (2024), reveals that CHIRPS skill degrades substantially at daily resolution, particularly in capturing convective rainfall timing, a limitation that metrics such as correlation and bias alone cannot detect. Overall, both satellite and reanalysis showed weak performance except for the index of agreement measurements. Even though IMERG showed > 0.5 according to Pearson correlation, it exhibited weak performance through other performance indicators. Additionally, one product showed a good performance at one station reserved for validation and was weak at the other. This might be the difference in the quality of ground measurements between stations.

3.3.1.2. Performance of merged product

After merging the ground measurements with covariates, including geographical information, satellites, and reanalysis precipitation products, the performance of the merged products was assessed against the station datasets reserved for validation purposes. We used a variety of statistical criteria and found that the merged precipitation estimates from RF-MEP and rainfall observation data were in good agreement. Furthermore, the merging product demonstrated a significant improvement over individual Seattleite and reanalysis products, as can be seen from Table 1. After combining multiple satellite and reanalysis products using the random forest approach, the MAE and RMSE of the merged product decreased to 1.668–4.667 mm/d and 3.833–8.99 mm/d, respectively. The merged product consistently exhibited the lowest mean absolute error and root mean square error across all stations. The lowest and highest errors were recorded at stations PGS 17 and PGS 43. Similarly, original precipitation products (unmerged satellite and reanalysis) also demonstrated the highest error at station PGS 43. This could probably be associated with the quality of data at a given station, since the presence of a few outliers substantially affected the RMSE values.

Meanwhile, the Nash-Sutcliffe coefficient of efficiency values showed a marked improvement of the merged product. Referring to Table 1, at all stations, positive NSCE values ranging from 0.246 to 0.743 were produced. The average NSCE value of the merged product was found to be around 0.482, which was larger than the unmerged individual precipitation products with negative average values. Moreover, according to Moriasi et al. (2007), the NSCE values between zero and one are acceptable; it is therefore, the merged product showed satisfactory to good performances across stations. On the other hand, the Pearson correlation coefficient between observed and merged precipitation showed a remarkable increment from 0.608 to 0.864 values across stations. Referring to Fig. 3a, the Pearson CC values greater than 0.7 were found for six of the nine stations used for comparison, which indicated that there was a high linear relationship between merged and observed records. Again, similar to RMSE and MAE, the lowest and highest CC values were found at stations PGS 43 &17, respectively.

Despite the merged product recording the highest error at station PGS 43, it demonstrated a good correlation value higher than 0.6. Regarding the index of agreement presented in

3b, the merged product exhibited an astonishing performance with values as high as 0.918, which can be considered almost a perfect agreement. At almost all stations used for validation, an index of agreement d value greater than 0.7 was observed. Finally, in Fig. 3.3, the performance of the merged product is presented based on the KGE values. The box plot showed the median value of KGE, which varied from 0.319 at PGS 43 to 0.751 at PGS 17. In general, 77% of the stations assigned for validation exhibited higher than 0.5 Kling-Gupta efficiency values, which were considered as better and more acceptable performance of the merged product.

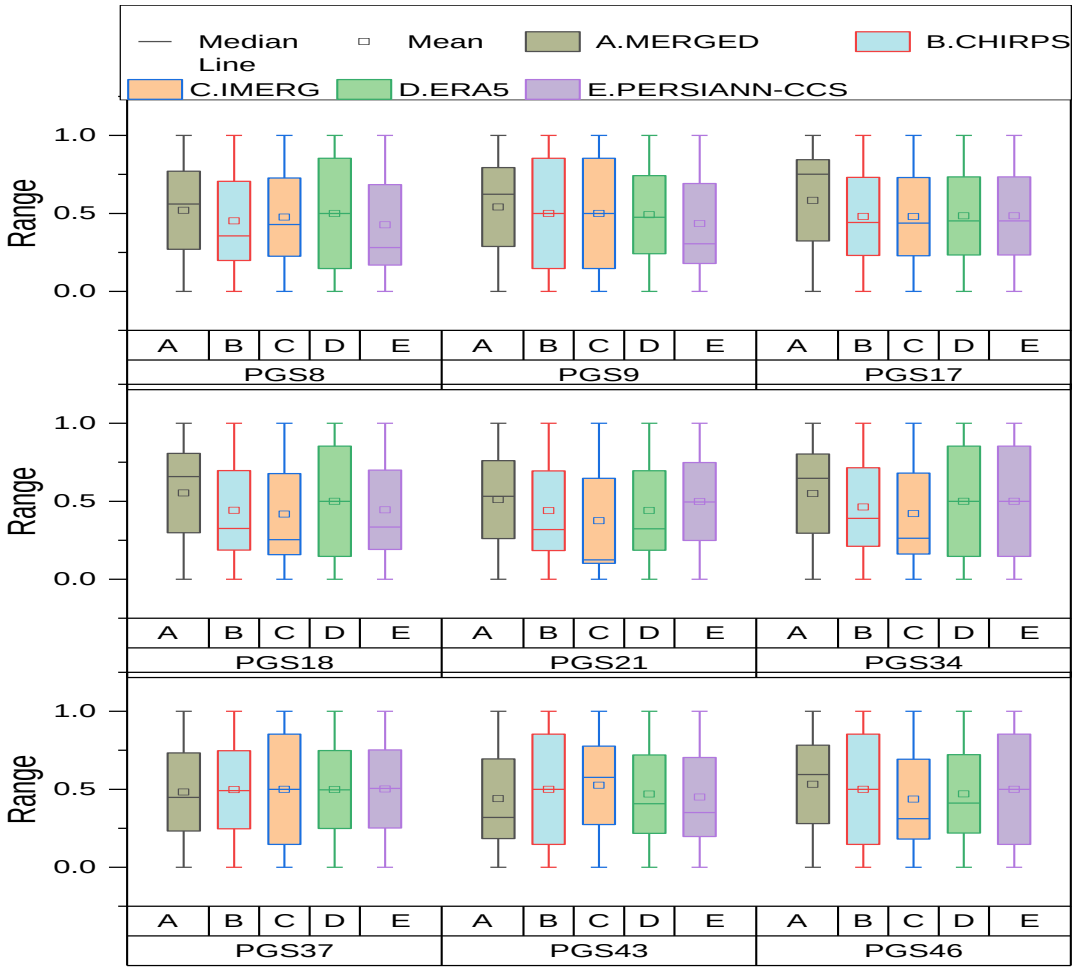
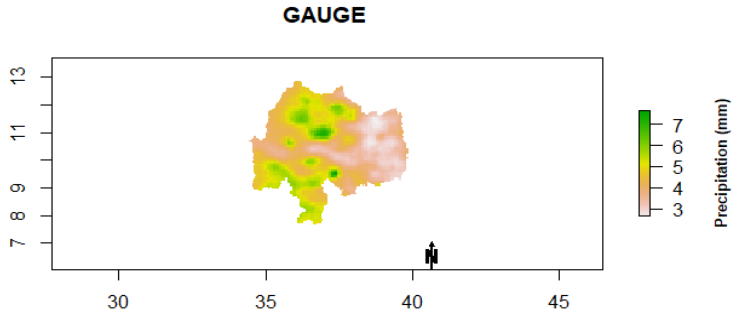


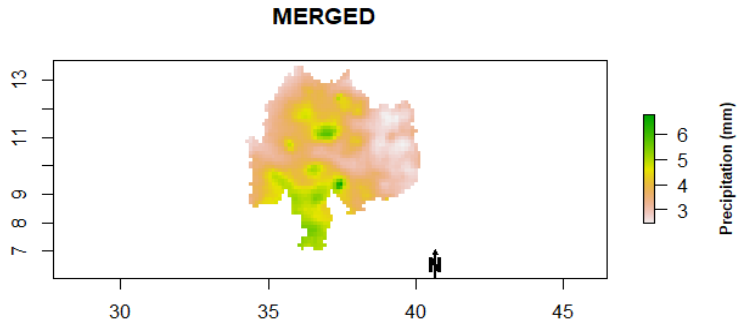
Fig 3.3. Boxplots representing the performance of individual and merged precipitation datasets based on Kling–Gupta efficiency with respect to ground precipitation gauge stations (PGS)

Regarding the spatial performance, the merged product exhibited a marked improvement compared to the individual unmerged products used as covariates. The stations assigned for validation were randomly selected from different parts of the basin to capture the spatial variability of the basin's precipitation. In each station, a substantial improvement was recorded based on all metrics. For example, in seven out of nine stations allotted for validation, a greater than 0.5 KGE value was observed. This highlighted the benefit of merging multiple sources of product using the RF-MEP method. The unmerged products have shown high variability and poor performance in different stations found in different parts of the basin. For instance, they exhibited up to 10.77 RMSE values, which is a marked

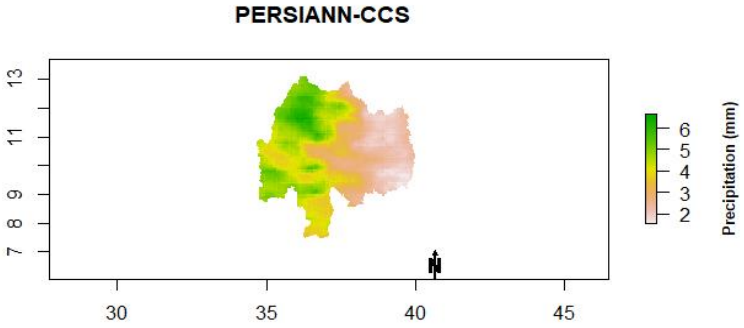
error and a weak correlation of up to 0.2 based on Pearson correlation values. Besides, as it can be observed from the daily mean precipitation plots of Fig. 3 .4(a-f), the merged product (RF MEP) has a better pattern replication of the gauge measurements than the products used as covariates. The merged product showed almost similar distribution and showed reasonable consistency against gauge measurements over the basin, with slight overestimations in the southern and underestimation in the central north part of the basin. On the other hand, both the satellite and reanalysis products exhibited a conspicuous overestimation in the western and southern territories of the basin. However, their pattern of overestimations has improved towards the eastern regions of the basin. The merged product exhibited a slight underestimation of the northern precipitation of the basin. The northern part of the basin is characterized by its complex and mountainous topography. As mentioned by Gebremedhin et al. (2020) and Chen et al. (2020), in such circumstances, the precipitation products underestimate the precipitation compared to the ground measurements. It is, therefore, the underestimation of the merged product in the northern part that possibly arose due to the poor performance of the precipitation products used as covariates for the merging tasks. In general, after merging, the spatial trend has shown remarkable improvements.



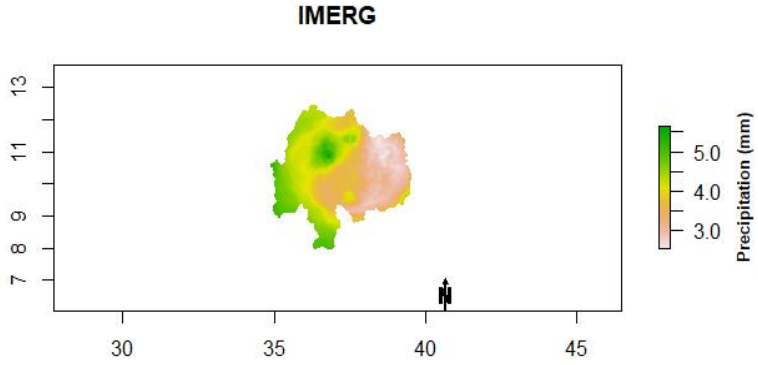
a


MERGED
13
11
9
8
7
30
35
40
45
6
5
4
3
Precipitation (mm)

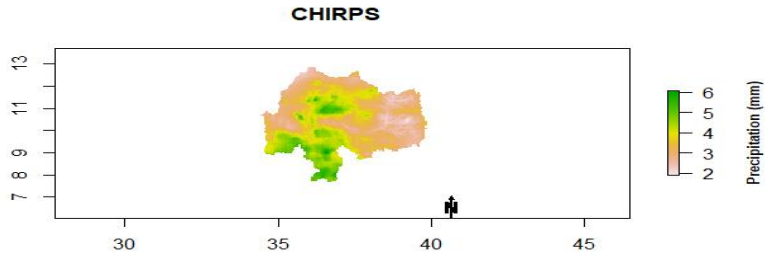
b


PERSIANN-CCS
13
11
9
8
7
30
35
40
45
6
5
4
3
2
Precipitation (mm)

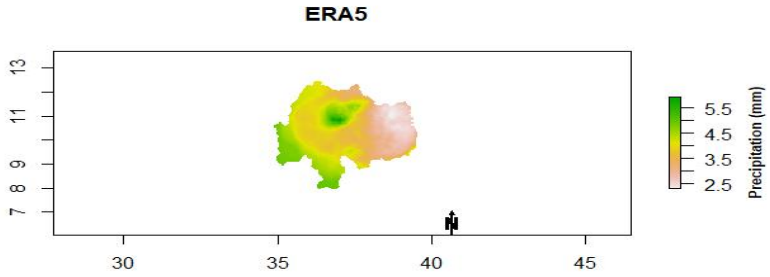
c


IMERG
13
11
9
8
7
30
35
40
45
5.0
4.0
3.0
Precipitation (mm)

d



e



f

Fig.3.4 (a-f) Mean daily precipitation for the period 2001-2017.

3.3.1.3. Qualitative metrics results

Regarding the capability for detecting the occurrence of precipitation at a daily scale, in Fig. 3.5, it is clear that both Satellite and reanalysis precipitation products showed approximate results by all metrics with moderate fluctuations. However, the merged product detected the rain event with more than 90% probability and exhibited remarkable improvements. Despite their poor performance in terms of MAE, RMSE, KGE, and NSCE, IMERG and PERSIANN-CCS exhibited a better probability of detection. On the other hand, CHIRPS recorded the lowest probability of detection across all stations assigned for the evaluation of these products.

The merged product showed the lowest FAR compared to individual satellite and reanalysis precipitation products, with a value of 0.15- 0.5 across stations. Among the original products, IMERG exhibited the highest false alarm ratio, as shown in Fig. 3. In contrast, POD CHIRPS showed the lowest FAR compared to other products. However, the probability of presenting false rain events was substantially reduced in the merged products.

On the other hand, the CSI values for the merged product notably increased up to 0.78. However, the unmerged products presented low and approximate CSI values with slight fluctuations. In general, the averages of all categorical Statistical metrics showed a considerable improvement in the merged product, but values exhibited marked fluctuations for the individual satellite precipitation product, even from station to station. One satellite product outperformed the other by one metric but became the lowest by the other. However, the improvements in the merged product exhibited the same fashion increment across all metrics.

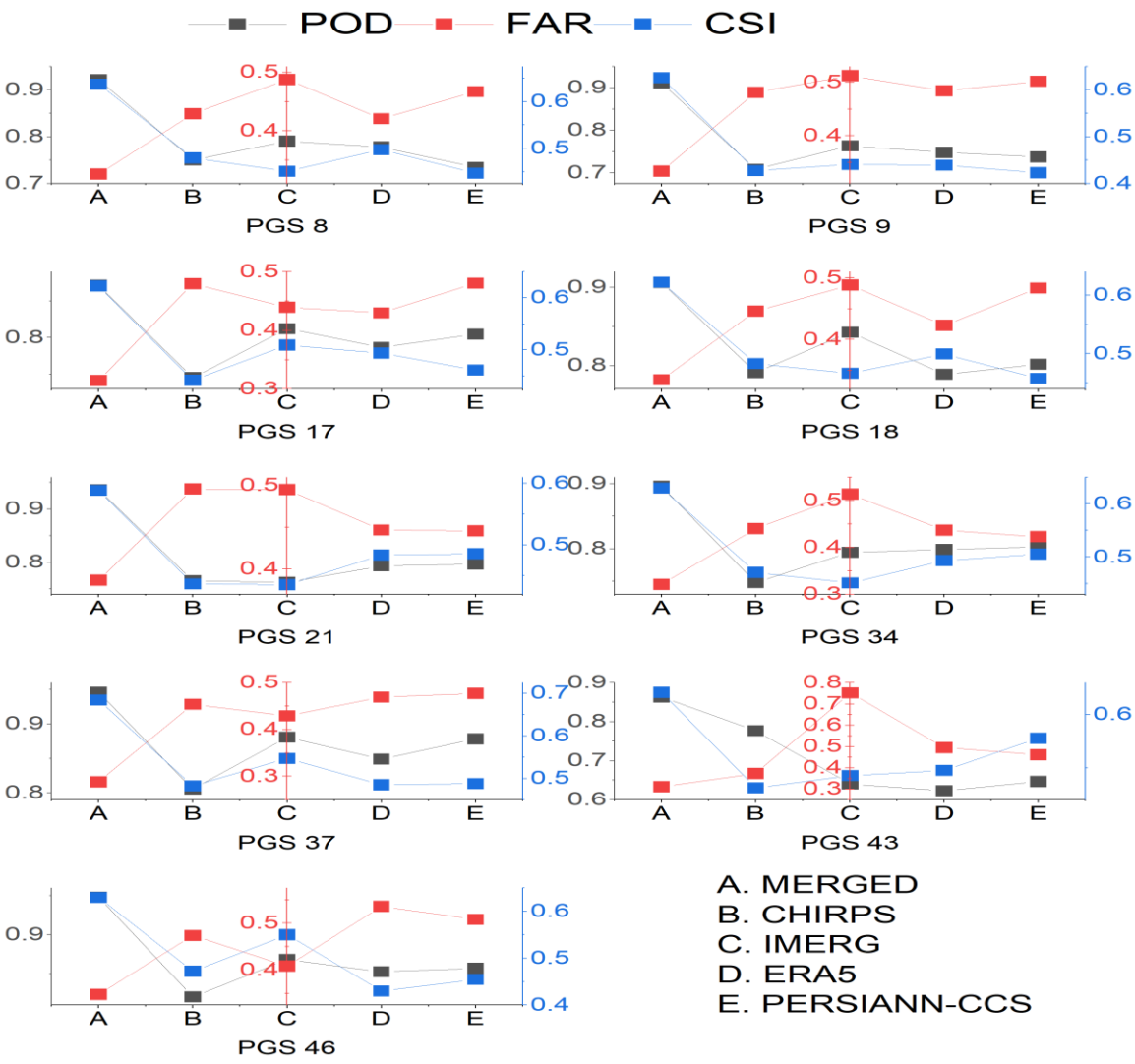


Fig. 3.5 Categorical validation metrics: presentation-based on Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) with respect to ground measurements at precipitation gauge station (PGS).

3.3.2. Discussion

Initially, performance assessments were done for three widely used daily satellites and one reanalysis precipitation product, which had varied spatial resolutions, using ground rainfall data from 2010 to 2017. The results above showed that both the satellite and reanalysis products exhibited a weak performance by all statistical metrics. These poor performances could be attributed to different factors. Among them, the precision of satellite products is determined by several factors, including rainfall retrieval techniques and sensors utilized in these satellites (Hu et al., 2019; Sun et al., 2018). Meanwhile, the reanalysis product heavily depends on the spatial distribution of the assimilated gauge observations and the algorithms applied for modeling (Xu et al., 2020; Bhuiyan et al., 2019). Moreover, the ability of satellite retrievals has limitations in estimating precipitation over complex and heterogeneous topographic terrains. Analogous reasons are also mentioned in other studies, such as (An et al., 2020; Chen et al., 2020; Liu et al., 2019; Satgé et al., 2018; Sun et al., 2018; Zambrano-Bigiarini et al., 2017; Yuan et al., 2017). Likewise, elevation also strongly affected the precision of the products (Navarro et al., 2020).

The Upper Blue Nile River basin, with its complex terrain and wide range of elevation variability (Roth et al., 2018), can be one of the reasons for the poor performance of the precipitation product found in this study. Another notable reason for the weak performance of satellite and reanalysis precipitation products against the ground gauge records could be the scale difference between these two sources. The rain gauges are known for their point-wise observations, but the reanalysis and satellite products are coarse grid-scale measurements, and this creates huge discrepancies. Combined with heterogeneous precipitation distribution, this scale discrepancy can be a huge source of biased evaluation results of these two products. In the same way, Gebremedihin (2020), Lakew (2020), and Rahmawati and Lubczynski (2017) also explained that the weak performance of satellite-derived products at a daily time scale was associated with the fine-scale difference between satellite-derived rainfall products and in situ observations.

On the other hand, these poor performances did not necessarily arise from satellite retrieval or assimilation model errors only, but also probably from the ground measurements, which can be subjected to different uncertainties. The presence of inconsistent records, uneven distribution, and false records is also a huge influence on evaluation statistics. Furthermore,

the existence of limited gauge stations in the study area, which may not represent the actual precipitation distribution within the pixel boundary, can be one of the core reasons associated with the low performance of satellite precipitation products. In their study, Wiwoho (2021), Sharma et al. (2020), and Ullah et al. (2019) also reported similar justification and findings. As described in section one of this study, the Upper Blue Nile River basin is known for its scarce data condition, both in quality and quantity, along with poor distribution of gauge stations that may contribute to the poor performance result of the satellite and reanalysis precipitation products.

After performance evaluation of individual satellite and reanalysis precipitation products, a merged data set was produced, and validation against independent gauges was done. Both the descriptive and categorical statistics results revealed that the merged product notably improved the spatiotemporal representations of precipitation in the study area. These results confirmed other similar studies, such as Zhang et al. (2021), Nguyen et al. (2021), Fan et al. (2020), Wu et al. (2020), and Baez-Villanueva et al. (2020), which reported that merging multisource data improved the precipitation estimate of a given area.

Blending geographic information, gauge, satellite, and reanalysis precipitation data in the upper Blue Nile River basin showed a good improvement over the original precipitation estimates. For instance, the KGE value markedly improved at all stations despite the relative difference among them. These results concurred with the findings of Zhang et al. (2021), Nguyen et al. (2021), and Baez-Villanueva et al. (2020). Likewise, other statistical metrics, such as the Pearson correlation coefficient, NSCE, and index of agreement, also showed a considerable increment and exhibited a better agreement with the ground gauge records than the covariates used for merging. These results are consistent and comparable with other studies, such as Nguyen et al. (2021), who applied the same methodologies. On the other hand, the root means square error and mean absolute error exhibited a marked reduction compared with the covariates, and this agrees with Zhang et al. (2021) and Nguyen et al. (2021).

Similarly, the categorical statistics also revealed a substantial reduction in the false alarm ratio (the number of days that were wrongly classified). Notwithstanding the covariates used for merging, the merged product exhibited a notable improvement in the probability

of detection and critical success index values. These findings are consistent with Zhang et al. (2021), Nguyen et al. (2021), and Baez-Villanueva et al. (2020) studies that reported similar results. The main reason for the increment of POD and reduction of FAR in the merged product is that it takes advantage of multi-satellite precipitation products, which use different and high-quality sensors and methods of detection (Nguyen et al. 2021). It is also worth mentioning that the density of gauges, as well as the error in records, markedly affected the performance of the merged product. Zhang et al. (2021) and Baez-Villanueva et al. (2020) from their experiments confirmed that the use of more gauges for training the model notably improved the quality of the merged products. In the Upper Blue Nile River basin, the available number of gauge stations is limited, and only 36 stations are qualified to train the model as discussed in section 2. Thus, the results of this study are promising from such a limited number of gauge stations and have the potential for improvements if more gauges are used to train the model in the merging process.

3.4. Conclusion

Merging of multisource precipitation data sources (gauge, satellite, and reanalysis) using a machine learning approach was carried out in the upper Blue Nile River basin, which is known for its most complex topography and data-scarce conditions. A random forest-based merging procedure was applied as a blending technique. This method is known for its ability to handle multiple covariates and geographic-related information. Three satellite and one reanalysis products were used in the merging process to improve the spatio-temporal precipitation estimates for the period 2010 -2017.

Before merging, the performance of satellite and reanalysis precipitation products was compared against the independent gauge measurements reserved for validation, and the results showed a weak performance for these products. Based on the statistical metrics, all products showed weak performance except for the relative difference between stations. IMERG exhibited slightly better performances based on the Pearson correlation coefficient > 0.5 values. However, for the rest of the metrics, there was no notable performance difference between products. According to metric results, the CHIRPS product, in general, recorded the worst performance at all stations.

The merged product exhibited a substantial improvement in precipitation estimation based on both descriptive and categorical performance metrics. It showed a good agreement with the independent gauge records, which were excluded from merging operations. These improvements at all stations with different locations indicated that the merged product better represented the spatiotemporal characteristics and distribution of rainfall across the basin territory. The number of days falsely classified notably reduced, the detection of rainfall events remarkably increased, and the overall critical success improved compared to the original data (covariate) used for the merging process. This implies that merging multi-source precipitation data along with geographical covariates will produce improved precipitation data, which can be used for a variety of applications and provide full records in data-scarce conditions. Furthermore, this study demonstrated that the machine learning-based multi-source merging scheme is effective and can enhance the advantages of individual precipitation measurements. The final merged precipitation estimates have a complete spatial coverage, high spatial resolution, and show better accuracy than any single-source input precipitation analysis products.

4. Enhancing monthly precipitation forecasting by integrating multi-source data with machine learning models

Abstract

This study developed a novel framework to improve monthly precipitation forecasts by integrating multi-source data with machine-learning techniques. The framework used a blend of gauge, satellite, and reanalysis data to train the model with high-quality spatial-temporal precipitation information. Predictors such as local meteorological variables, large-scale climate indices, soil moisture data, and lagged precipitation data were used for one-month ahead precipitation forecasting. Five machine learning algorithms, including Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), and Random Forest (RF), were evaluated in the framework for their forecasting performance across five monitoring stations in the upper Blue Nile basin. The results demonstrate that the novel forecasting framework integrating multi-source data with machine learning techniques significantly improved the monthly precipitation forecasts. All models exhibited higher R^2, closer agreement with observed data measured by Willmott's index (d), and lower MAE and RMSE values. During the testing phase, the models recorded 0.64-0.87 R^2, 0.62-2.95 RMSE, 0.18-2.42 MAE, and 0.7-0.85 d values across stations. Among the models evaluated in the framework, ANN, SVM, and KNN consistently showcased better performance both in the training and testing phases, with ANN exhibiting the most promising results overall. However, Extreme Gradient Boosting and Random Forests exhibited comparatively weaker performance across stations. Additionally, the models' performance under different climate categories (dry, normal, and wet months) was also explored. The models excel during normal and dry months, but during wet months their performance diminishes relatively. Furthermore, the proposed framework forecast skills compared with NCEP and ECMWF global precipitation prediction showed a better performance. Overall, the findings highlight the potential of the developed framework to enhance monthly precipitation forecasts, providing valuable insights for decision-makers in various sectors reliant on accurate precipitation information.

Keywords: *Multi-source data integration, Machine learning models, Monthly precipitation forecasting, Global precipitation predictions, Upper Blue Nile Basin*

4.1. Introduction

Forecasting the occurrence and quantity of precipitation several months in advance has been crucial to setting up coping strategies, operational preparedness against potential hazards, and effective utilization of water resources (e.g., Modaresi et al., 2024; Arnone et al., 2020; Poornima et al., 2019; Lee et al., 2018; Kumar et al., 2019). However, its complex interactions with other atmospheric processes and dynamic phenomena make precipitation one of the most difficult tasks to predict accurately (Sarasa-Cabezuelo, 2022; Kumar et al., 2019; Tharun et al., 2018). Thus, characterizing forecast uncertainty and developing methods to improve precipitation forecasting have gained increasing attention in recent years. Research results show that developing a globally generalized precipitation forecasting method remains elusive (Waqas et al., 2023; Danandeh Mehr, 2020; Hossain et al., 2019). Physical models have the advantage of relying on physical processes, but the lack of sufficient data, computational expense, and complexity of atmospheric processes have restricted their uses (Barrera Animas et al., 2022; Nejero et al., 2021; Nourani et al., 2019; Cristiano et al., 2017). Furthermore, their prediction is highly subject to large-scale errors, particularly at regional and watershed scales (Weslati et al., 2023; Yaseen et al., 2019; Aguasca-colomo et al., 2019). On the other hand, stochastic linear models are easy to implement and can be used with limited data conditions (Salih et al., 2020; Unnikrishnan & Jothiprakash, 2018; Nagendra & Ussenaiah, 2018; Zhou et al., 2017). However, accuracy is impeded by rainfall's nonlinear, chaotic nature and its complex interaction with other atmospheric components (Madhu et al., 2021; Narejo et al., 2021; Pena et al., 2020; Esteves et al., 2020).

Machine learning-based forecasting models for hydrological components have gained popularity in recent years due to their exceptional performance in the presence of non-linearity and uncertainties (Tin et al., 2024; Zhou et al., 2021; Lee et al., 2018; Abbot & Marohasy, 2018). These models identify hidden patterns in historical rainfall datasets without the need for knowledge of the underlying physical processes (Weslati et al. 2023; Waqas et al., 2023; Rahman et al., 2022; Ouma et al., 2022; Ridwan et al., 2021). Machine learning models can handle complex situations and irregular datasets without sacrificing forecast accuracy (Weerakody et al., 2021). Furthermore, these techniques have demonstrated high accuracy when dealing with large amounts of non-linear, heterogeneous

data and peak rainfall values (Sarasa-Cabezuelo, 2022; Ridwan et al., 2021; Mehr et al., 2019; Unnikrishnan & Jothiprakash, 2018). As a result, machine-learning methods are increasingly preferred over conventional statistical and physically-based models (Tuysuzoglu et al., 2023; Salaeh et al., 2022; Ghamariadyan& Imteaz, 2021; Nourani et al., 2021; Alizadeh et al., 2021; Bojang et al., 2020; Kang et al., 2020; Chong et al., 2020; Poornima& Pushpalatha, 2019; Jiao et al., 2016; Abbot et al., 2012).

Like other methods, in machine learning methods too, the availability of sufficient data of the right quality and quantity is crucial for producing accurate forecasts (Chengcheng et al., 2022; Hutter et al., 2019; Nourani et al., 2019; Valipour, 2016). However, such data are not readily available at sufficient resolution, coverage, and accuracy, but could be created by combining different datasets with complementary strengths and weaknesses. For instance, Mohammed et al. (2023), Nguyen et al. (2021), Fan et al. (2021), Zhang et al. (2021), and Baez-Villanueva et al. (2020) have verified that merging diverse data sources, including gauge, satellite, and reanalysis data, has led to a substantial enhancement in the spatial-temporal quality and quantity of precipitation data.

The objective of this study is to develop a new methodological framework to forecast monthly precipitation with a one-month lead time that (1) employs machine learning methods, (2) blends multiple precipitation data sources to create reliable precipitation datasets, and (3) utilizes additional predictor data, such as soil moisture, large-scale climate indices, and local meteorological data. The novelty of the proposed framework lies in its use of blended high-quality precipitation data (gauge, satellite, and reanalysis) to train the models. This multi-source data integration, coupled with the use of lagged climate predictors, enhances the accuracy and robustness of the forecasts compared to methods relying solely on single data sources. By merging data from different sources, the proposed framework captures a more comprehensive and detailed representation of precipitation patterns, leading to improved forecasts. The Upper Blue Nile basin is taken as a case study region where rainfall forecasting is highly crucial for hydropower, agriculture, resource planning, and management. The proposed methodological framework will offer a more comprehensive approach to monthly precipitation forecasting, which can help water resource managers and policymakers make informed decisions and develop effective water management strategies.

4.2. Data and Methods

4.2.1. Data

The study utilizes a merged precipitation dataset, synthesized in Chapter 3 from a blend of gauge stations, IMERG Early, CHIRPS, PERSIANN-CCS, and ERA5 data sources. Other ground-based meteorological datasets, including minimum temperature, relative humidity, wind speed, sunshine hours, and maximum temperature, were also collected from the National Meteorological Agency of Ethiopia. For the forecasting task, five stations, namely Debermarkos (DM), Ambo (AM), Gonder (GN), Pawe (PW), and Debretabor (DB) were selected from the Upper Blue Nile Basin based on their relative availability of complete local meteorological datasets and their altitude in such a way that they represent different climate zones of the basin. The global large-scale climate indices data including North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), Tropical Southern Atlantic (TSA), Niño 1+2, Antarctic Oscillation (AAO), Atlantic Meridional Mode (AMM), ENSO Precipitation Index, Sahil Precipitation Index, ENSO precipitation index, Indian Ocean Dipole Mode Index (DMI) were also used as predictors in this study. The monthly values of climate indices were collected from the Climate Prediction Center on the National Oceanic and Atmospheric Administration website. (https://www.esrl.noaa.gov/psd/data/climateindices/list/). Furthermore, soil moisture data were also used, which were accessed from the European Center for Medium-Range Weather Forecasts website: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture.

Additionally, in this study, we used the reforecast data, which consists of forecasts with historical start and projection dates, providing a valuable benchmark for assessing the performance of the proposed framework's predictions against observations (Merz et al., 2020; Manrique-Suñen et al., 2020). The reforecast data used in this study are sourced from the S2S database, which archives reforecast data from global prediction centers at a common 1.5° grid horizontal resolution (Endris et al., 2021). We obtained the Forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP), which were accessed for the study area. Detailed information about the S2S database, ECMWF, and NCEP can be found in Endris et al. (2021) and Zhang et al. (2021).

4.2.2. Selection of Predictors

A cross-correlation analysis was carried out to select the predictor variable and establish a relationship between the lagged climate indices and precipitation of a given month. The local meteorological and large-scale climate indices were found to be highly correlated, including North Atlantic Oscillation (NAO), Pacific Decadal Oscillation (PDO), Tropical Southern Atlantic (TSA), Niño 1+2, Antarctic Oscillation (AAO), Atlantic Meridional Mode (AMM), ENSO Precipitation Index, Sahil Precipitation Index, Enso precipitation index, Indian Ocean Dipole Mode Index (DMI), minimum temperature, maximum temperature, wind speed, relative humidity, and sunshine hour. Additionally, autocorrelation analysis was also done to identify a relationship between the precipitation of a target month and lagged month precipitation phenomena of the past. And, it was found that precipitation in a given month is highly correlated with a 1-month lag and the same month of the previous year's precipitation. Furthermore, from cross-correlation analysis, lagged soil moisture data were also found to be good predictors for the precipitation of a target month.

4.3. Forecasting Framework

The general schematic representation of the forecasting framework, which integrates multi-source data with a machine learning model is shown in Fig.4.1:

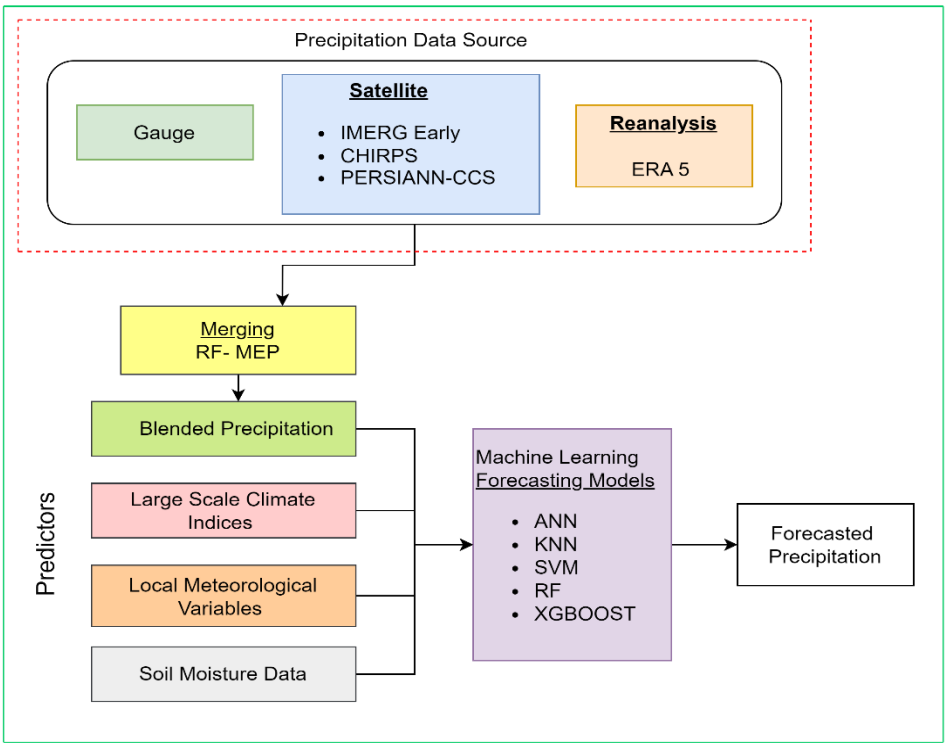


Fig. 4.1. A schematic presentation of the forecasting framework

4.4. Machine Learning Models

This study employed five machine learning models: K-Nearest neighbors (KNN), Random Forest, Artificial Neural Network, Support Vector Machine, and Extreme Gradient Boosting (XGBoost). These models are widely used for rainfall forecasting at different lead times and various regions (e.g., Sarasa-Cabezuelo, 2022; Rahman et al., 2022; Wolfensberger et al., 2021; Anwar et al., 2021; Samantaray et al., 2020; Lee et al., 2018; Nagendra et al., 2018; Huang et al., 2017). The study utilized KNIME, XLSTAT, and R software packages to execute these machine learning models.

4.4.1. K-Nearest neighbors (KNN)

K-Nearest neighbors (KNN) is a popular machine learning algorithm used for classification and regression tasks. The KNN regression works by finding the k nearest neighbors from the input data point based on their Euclidean distances. These neighbors are then used to calculate the average of their output values, which is used as the predicted output value for the input data point (Vinay & Hombaliah, 2021; Huang et al., 2017). The details about the KNN algorithm can be found in Vallam & Qin (2016) and Biruntha et al. (2022).

The Euclidean distance between the new point and each training point is calculated as in Eq. 4.1.

$$D(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad \text{Eq.4.1.}$$

Where (x) is a new point, (y) is an existing point, and i is the i-th feature of nodes and y.

4.4.2. Random Forest

Random forest is a popular machine learning algorithm that is often used for classification and regression tasks. It is an ensemble method that combines multiple decision trees to make more accurate predictions (Pham et al., 2021; Danandeh Mehr et al., 2021). Each decision tree is trained on a randomly selected subset of the data and a random subset of features, which helps to reduce overfitting and improve generalization (Hill & Schumacher, 2021). And the final single prediction will be taken from the average of multiple samples, which are selected through the bootstrap sampling method (Hanoon et al., 2021). Details about the random forest algorithm can be found in Wolfensberger et al. (2021) and Pham et al. (2021).

4.4.3. Artificial Neural Network

An artificial neural network is a type of machine-learning system that is modeled after the way the human brain works. A multi-layer perceptron (MLP) is a type of artificial neural network that is commonly used for supervised learning tasks such as classification and regression. The Multilayer Perceptron (MLP) works by processing input data through interconnected layers of nodes, known as neurons, to generate output predictions (Dongwoo, 2021; Liu et al., 2019). Each neuron applies an activation function to the weighted sum of its inputs, introducing nonlinearity and enabling the network to model complex relationships within the data (Ghamariadyan & Monzu, 2021; Ali et al., 2017). During training, the network adjusts its weights using optimization algorithms, such as backpropagation, to minimize prediction errors (Salaeh et al., 2022; Esteves et al., 2019). A detailed explanation of MLPs can be found in Lee et al. (2018) and is mathematically expressed as in Eq.4.2:

$$y = \varphi(\sum_{i=1}^{n} w_i x_i + b) \qquad \text{Eq.4.2}$$

Where y is the forecasted output value, φ is the activation function for the output neuron, w is the vector of weights, x is the vector of inputs, and b is the bias. The logistic activation function (Sigmoid: σ) is a widely used function that takes whatever data is entered and changes it to a value between 0 and 1 (Esteves et al., 2019; Liu et al., 2018).

$$f(x) = \frac{1}{(1+e^{-x})} \qquad \text{Eq. 4.3}$$

4.4.4. Support Vector Machine

A support vector machine (SVM) is a machine learning algorithm that is used for classification and regression analysis (Montesinos López et al., 2022; Hamidi et al., 2015). It works by finding a hyperplane in the n-dimensional space that best predicts the data (Dawoodi et al., 2021; Danandeh Mehr et al., 2018). This hyperplane is called a decision boundary, and points on either side of it are called support vectors. The SVM algorithm uses a kernel function to transform the input data into a higher-dimensional space where it can be separated by a hyperplane (Samantaray et al., 2021). The kernel function is a mathematical function that maps the input data into a higher-dimensional space.

$$f(x) = w^{Tr}.x + b \qquad \text{Eq.4.4}$$

where x is the input variable, b is the bias, w is the weight, and Tr is the transpose.

4.4.5. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting, often abbreviated as XGBoost, is a powerful supervised machine learning algorithm designed to accomplish complex tasks in the realm of classification and regression (Danandeh Mehr, 2021). Its core lies in an ensemble learning technique called boosting and applying an additive training strategy, where the predictions are made from weak learners that continuously develop over the mistakes of the former learners (Danandeh Mehr, 2021; Ramón et al., 2021; Mai et al., 2021). XGBoost uses a second-order Taylor approximation of the loss function along with the first-order gradient to capture the direction of changes and curvature of the loss function (Wei and Hsu, 2021). The Taylor approximation provides a more accurate approximation, which helps the algorithm converge faster and results in better performance than the conventional gradient boosting technique (Qinghe et al., 2022; Alshboul et al., 2022). More details can be found in Szczepanek (2022) and Hussein et al. (2020).

$$O = \sum_{i=1}^{n}(L(y_I, F(x_i)) + (\sum_{k=1}^{t} R(f_k) + C \qquad \text{Eq. 4.5}$$

where O is the objective function, R(fk) denotes the regularization term at the k iteration time, and C is a constant. The regularization term is expressed as

$$R(f_k) = \alpha H + \frac{1}{2}\eta \sum_{j=1}^{nH}(w_j{}^2, \qquad \text{Eq. 4.6}$$

where α denotes the complexity of leaves, H represents the number of leaves, η denotes the penalty parameter, and w_j is the output of each leaf node.

4.5. Classification of Normal, Dry, and Wet Months

The Standardized Precipitation Index (SPI) is a useful tool for classifying dry, wet, and normal months based on precipitation patterns. The SPI is a statistical index that measures how the precipitation for a specific period compares to the long-term precipitation record for that same period (Moccia et al., 2022; Du et al., 2013). When analyzing the SPI values, we can classify months as dry, wet, or normal based on certain thresholds.

For dry conditions, typically, SPI values below -1 indicate a period of below-average precipitation, indicating a moisture deficit. When the SPI value falls between -1 and +1, it suggests normal precipitation, indicating a balance between water availability and need. Wet months are characterized by SPI values above +1, indicating an excess of precipitation compared to the long-term average (WMO, 2012; McKee et al., 1993).

4.6. Forecast Evaluation

Forecast evaluation metrics indicate the nearness of the forecasted data to observed rainfall data allotted to the training and testing phases. In this study (see Table 4.1), we used statistical metrics including R^2 (correlation coefficient), Willmott's index of Agreement (d), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Table 4.1: Evaluation metrics and equation

Evaluation Metrics	Equation	Value Ranges	Best values
Correlation Coefficient (R^2)	$R^2=\left[\frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}}\right]^2$	[0,1]	1
Mean Absolute Error (MAE)	$MAE=\frac{1}{n}*\sum_{i=1}^{n}\lvert x_i-y_i\rvert$	$[0,+\infty)$	0
Root Mean Square Error (RMSE)	$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(x_{,i}-y_i)^2}{n}}$	$[0,+\infty)$	0
Willmott's Index of Agreement (d)	$d=1-\frac{\sum_{i=1}^{n}(x_i-y)^2}{\sum_{i=1}^{n}(\lvert y_i-\bar{x}\rvert+\lvert x_i-\bar{x}\rvert)^2}$	[0,1]	1

Notation: where n is the total number of observations, x_i is the observation value, and y_i is the forecasted value and $\bar{x}$ and $\bar{y}$ are the average values of observed and forecasted values, respectively.

4.7. Results and Discussion

a. Quantitative metrices and visual diagnostics

This study presents a novel forecasting framework to enhance the accuracy of monthly precipitation forecasts by integrating multi-source data with machine learning models. The framework capitalizes on the advantage of a spatially and temporally enhanced blended dataset comprising gauge, satellite, and reanalysis data as a lagged predictor to forecast the precipitation of the target month. This data set is also employed to train forecasting models. In addition to the precipitation data, a wide array of predictor variables, including local meteorological variables, large-scale climate indices, and soil moisture data, was incorporated into the forecasting model to provide a comprehensive representation of the precipitation driving processes, which contributes to the robustness of the precipitation forecasting framework.

The preliminary analysis involved the utilization of cross-correlation to establish lagged predictors that could significantly influence the monthly rainfall patterns under consideration. This approach ensures that the predictive model accounts for temporal dependencies. The dataset was partitioned into training and testing subsets, with 70% of

the data assigned to model development and the remaining 30% for evaluation. Five machine learning techniques-artificial neural networks, K-Nearest neighbors, random forest, Extreme Gradient Boosting, and support vector machines were each applied as prediction algorithms using records from five monitoring stations in the upper Blue Nile basin. The performance of the developed forecasting framework was evaluated using metrics including the Correlation Coefficient (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Willmott's index of Agreement. Table 4.2 summarizes the performance metrics for both the training and testing phases. Overall, the results demonstrate the potential of this forecasting framework to generate more accurate precipitation forecasts through the integration of multi-source data with machine learning models.

Table 4.2. Performance of the forecasting framework

		DM		**AM**		**GN**		**PW**		**DB**	
	Performance metrics	**TR**	**TEST**	**TR**	**TEST**	**TR**	**TEST**	**TR**	**TEST**	**TR**	**TEST**
ANN	MAE	0.18	0.29	0.30	0.41	0.51	0.64	0.33	0.47	0.75	0.97
	RMSE	0.50	0.62	0.818	1.07	1.35	1.845	1.701	2.04	1.28	1.54
	R^2	0.97	0.78	0.99	0.83	0.99	0.87	0.99	0.82	0.95	0.72
	d	0.93	0.75	0.96	0.81	0.96	0.78	0.95	0.85	0.88	0.81
XGBoost	MAE	1.43	1.88	1.32	1.80	1.69	2.10	1.73	2.42	1.27	2.01
	RMSE	1.60	2.19	1.69	2.10	1.95	2.54	2.16	2.95	1.65	2.28
	R^2	0.74	0.71	0.73	0.70	0.72	0.69	0.70	0.67	0.69	0.66
	d	0.76	0.73	0.74	0.73	0.74	0.71	0.73	0.73	0.72	0.70
RF	MAE	1.31	1.53	1.39	1.59	1.32	2.01	1.64	1.83	1.25	1.70
	RMSE	1.86	2.01	1.75	2.01	1.78	2.19	1.92	2.36	1.94	2.27
	R^2	0.68	0.64	0.74	0.66	0.73	0.66	0.79	0.74	0.77	0.73
	d	0.76	0.75	0.83	0.76	0.84	0.77	0.82	0.77	0.74	0.71
KNN	MAE	1.18	1.96	1.08	1.97	1.08	1.53	1.34	2.04	1.63	1.46
	RMSE	1.75	2.14	1.33	2.72	1.22	1.66	1.86	2.57	1.48	1.99
	R^2	0.73	0.71	0.79	0.71	0.80	0.76	0.83	0.80	0.77	0.75
	d	0.78	0.77	0.83	0.83	0.82	0.81	0.82	0.83	0.86	0.78
SVM	MAE	1.08	1.72	0.92	1.27	1.00	1.55	1.12	1.70	1.24	1.99
	RMSE	1.03	1.74	1.08	1.80	1.58	2.19	1.55	2.16	1.82	2.05
	R^2	0.81	0.78	0.91	0.81	0.87	0.82	0.82	0.80	0.79	0.71
	d	0.84	0.82	0.85	0.84	0.86	0.77	0.84	0.81	0.79	0.76

Referring to Table 4.2 above, during the testing phase, all models produced better results and demonstrated a consistent level of performance, albeit slightly lower than what was achieved during the training phase. For instance, Artificial Neural Networks (ANN) showed up as the best performers. Displaying impressive results across all metrics during both training and testing phases. During the training phase, R-squared values ranged from 0.72 to 0.87, indicating that they effectively explained between 72% and 87% of the variance in observed precipitation data. Additionally, Wilmot's index, a measure of model prediction accuracy, showcased accurate forecasts in the range of 0.75 to 0.85, making it even more reliable with forecasting capabilities.

In contrast, Extreme Gradient Boosting (XGBoost) exhibited comparatively lower performance than the other models during both the training and testing phases. The R-squared values ranged from 0.69 to 0.74 during training and from 0.66 to 0.71 during testing phases. Furthermore, compared to the other forecasting models applied in this study, XGBoost produced a relatively lower Willmott's index of agreement, indicating weaker performance in terms of forecasting accuracy. Conversely, the Random Forest (RF) model showed inconsistent performance across the monitoring stations, exhibiting high variability according to all metrics. However, Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) displayed intermediate performance, surpassing Random Forest (RF) and XGBoost, yet falling short of the precision of Artificial Neural Networks (ANN). Their R-squared values ranged from 0.71 to 0.82 for SVM and from 0.71 to 0.80 for KNN during the testing phases.

Regarding accuracy metrics, all models within the forecasting framework demonstrated reasonably low RMSE and MAE values, with minor differences across stations. ANN and SVM exhibited the lowest MAE values, achieving 0.18 and 0.92 during the training phase and 0.29 and 1.27 during the testing phase, respectively. Similarly, the lowest RMSE values were recorded for ANN and SVM during training (0.50 and 1.03) and testing phases (0.62 and 1.74). These findings underscore the consistency of the forecasting framework in producing forecasts with small errors, enhancing the reliability of its predictions.

Conversely, RF and XGBoost recorded slightly lower performance in terms of RMSE and MAE, suggesting a potential sensitivity to certain station-specific conditions. KNN, on the

other hand, displayed consistent performance across various metrics, indicating an intermediate predictive capability. This superiority of ANN can be attributed to its ability to model smooth nonlinear relationships, handle heterogeneous predictors, and interpolate effectively in high-dimensional spaces with limited data. Unlike tree-based methods, ANN learns smooth, continuous mappings, automatically down-weights irrelevant predictors, and interpolates effectively in high-dimensional spaces. While XGBoost underperformed because its error independence assumption was violated by temporal autocorrelation, the dataset was too small (~59 training samples), and it could not directly capture sequential dependencies.

To evaluate the effectiveness of the forecasting models, we conducted a visual analysis of the time series data. The time series graphs presented below (see Fig. 4 .2a-e) demonstrate that all machine learning models within the forecasting framework successfully captured the essential characteristics of the rainfall data. This is particularly evident in their ability to maintain general patterns and fluctuations over time. Notably, Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) exhibited superior performance in accurately capturing the observed monthly rainfall trends. The framework's ability to replicate actual rainfall behavior further reinforces its forecasting accuracy. However, it is important to acknowledge that there were minor discrepancies between the forecasted and observed values, particularly during peak rainfall months. This suggests potential limitations in the forecasting framework's capacity to predict extreme events or the most intense rainfall periods accurately.

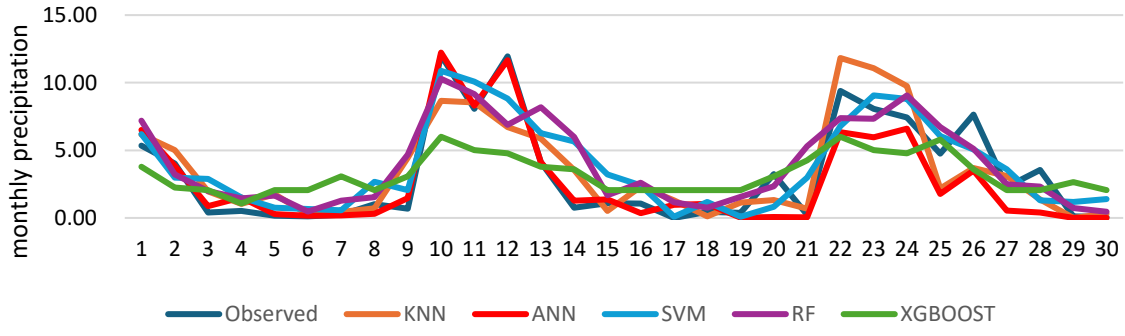


Fig. 4.2 (a) Time Series Graph of Observed vs Forecasted Precipitation @DM Station

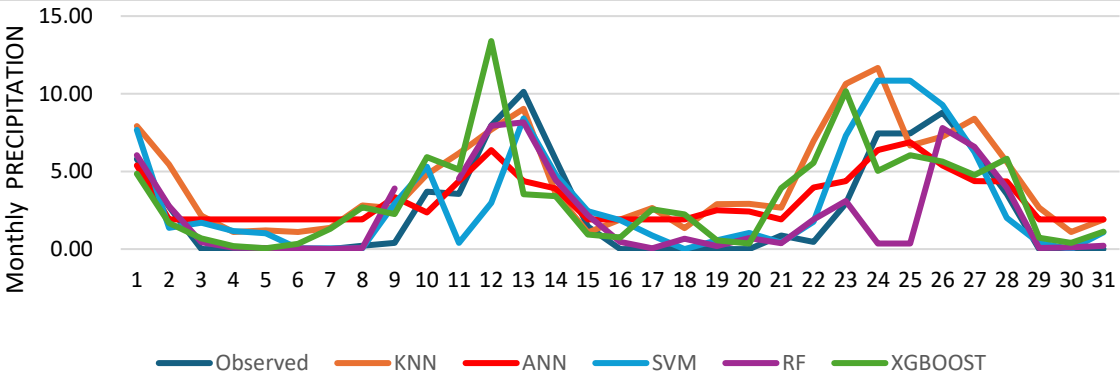


Fig. 4.2 (b) Time Series Graph of Observed vs Forecasted Precipitation @AM Station

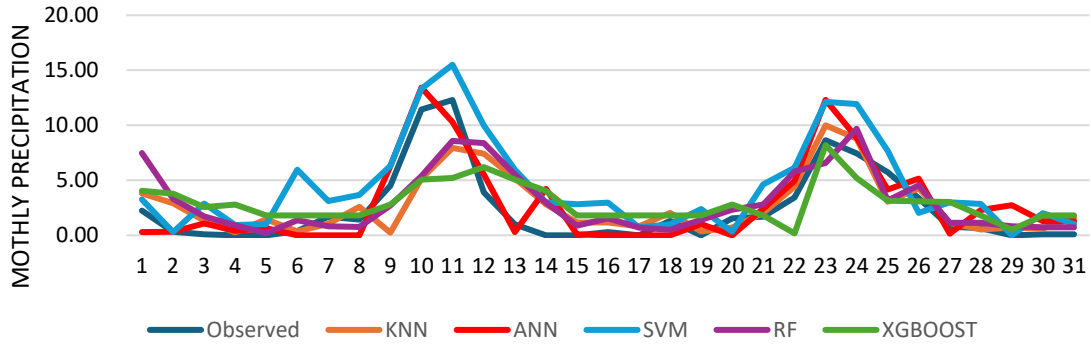


Fig.4.2 (c) Time Series Graph of Observed vs Forecasted Preicpitation @ GN Station

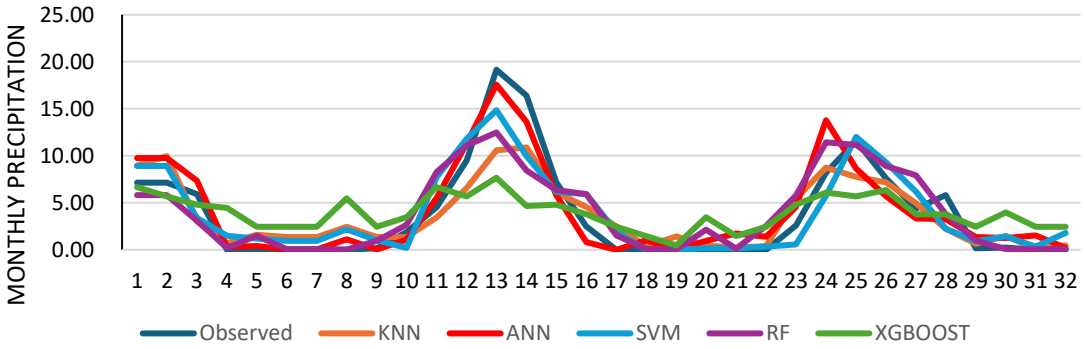


Fig.4.2 (d) Time Series Graph of Observed vs Forecasted Precipitation @ PW Station

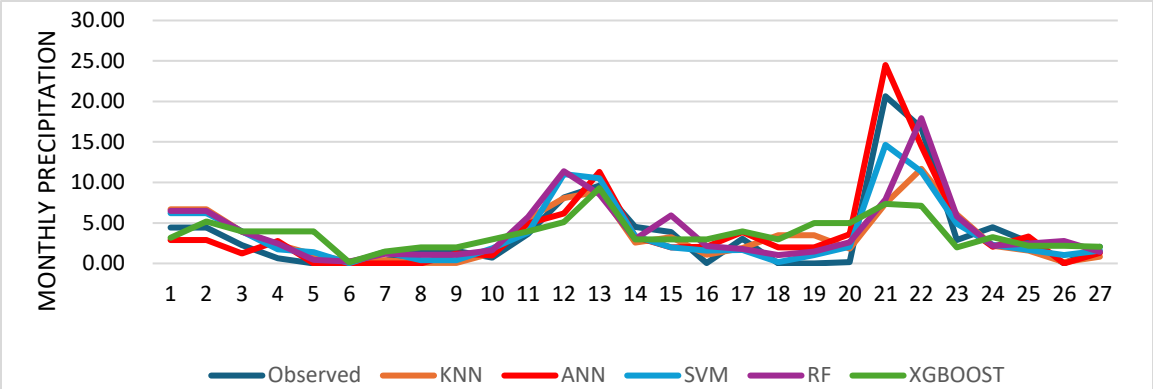


Fig. 4.2 (e) Time Series Graph of Observed vs Forecasted Precipitation @ DB Station

Fig. 4-2. (a- e) Time series graphs observed vs forecasted rainfall

To further evaluate the alignment between predicted and observed rainfall values, we created scatter plots, as illustrated in Figure 4.3 (a-e). These plots reveal a strong correlation between the forecasted values produced by all models within the forecasting framework and the actual rainfall data. While the overall agreement is commendable, it is important to note that some specific peak values do exhibit minor deviations from the observed data. Nevertheless, the machine learning models demonstrate impressive performance in capturing the time series characteristics and aligning with the observed precipitation patterns. These findings instill confidence in the potential of these models as valuable tools for enhancing monthly precipitation forecasts.

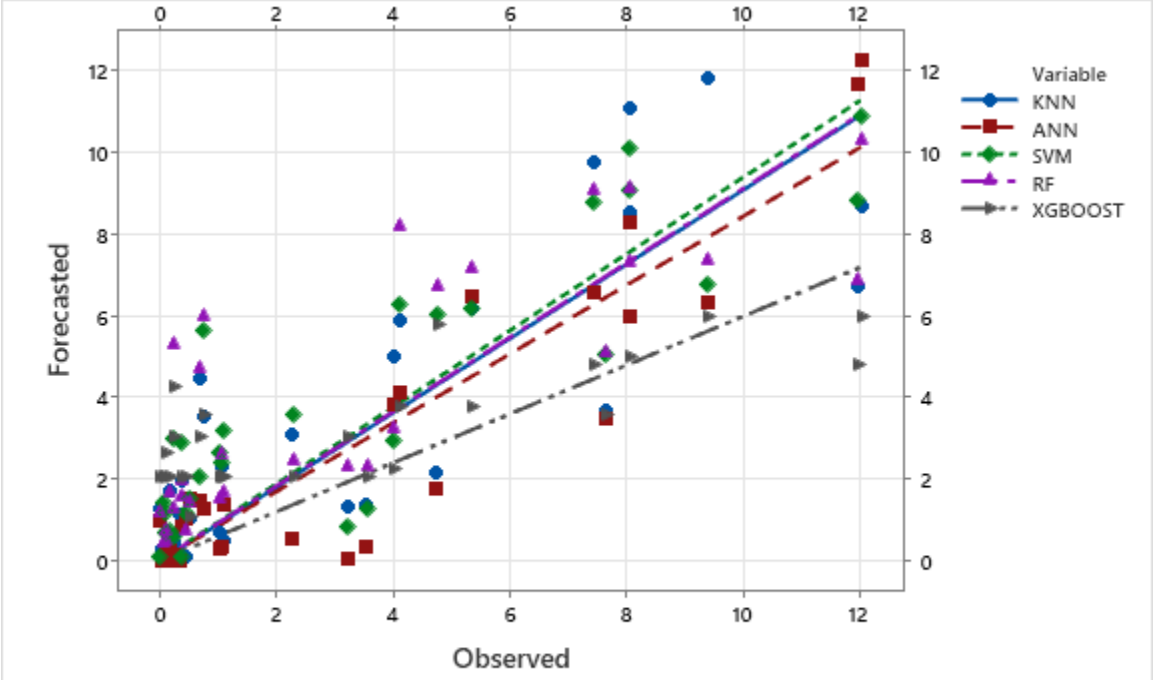


Fig. 4.3 (a) KNN, ANN, SVM, RF, XGBOOST Forecasted vs Observed Precipitation @ DM Station

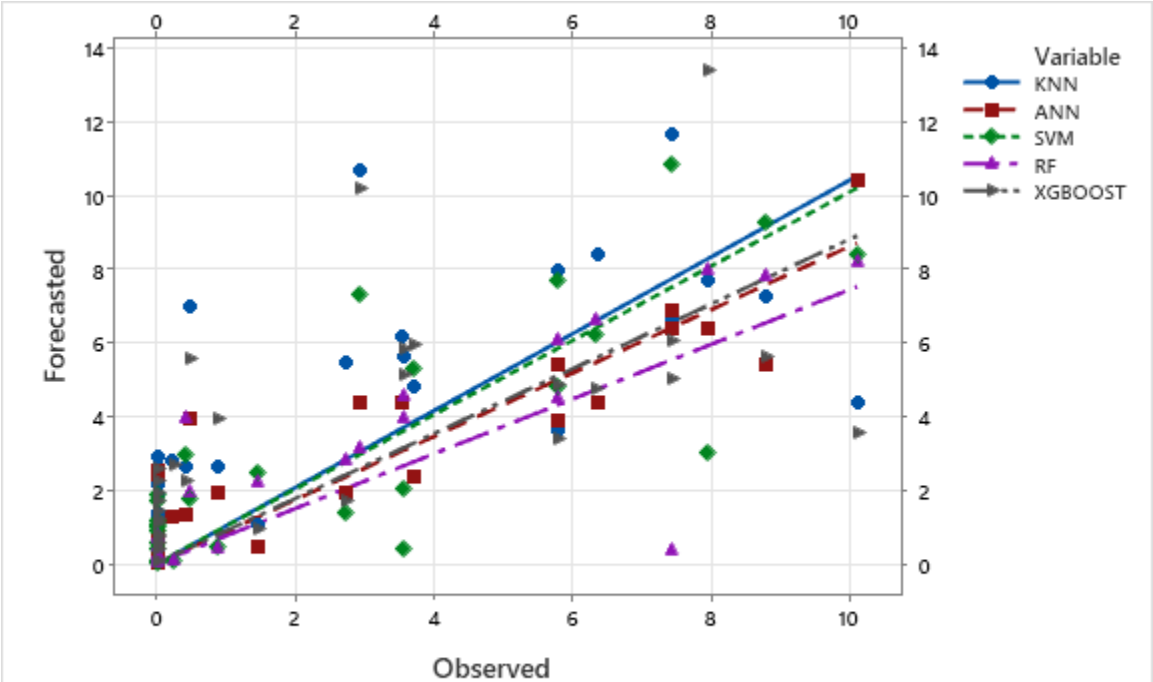


Fig. 4.3 (b) KNN, ANN, SVM, RF, XGBOOST Forecasted vs Observed precipitation @ AM Station

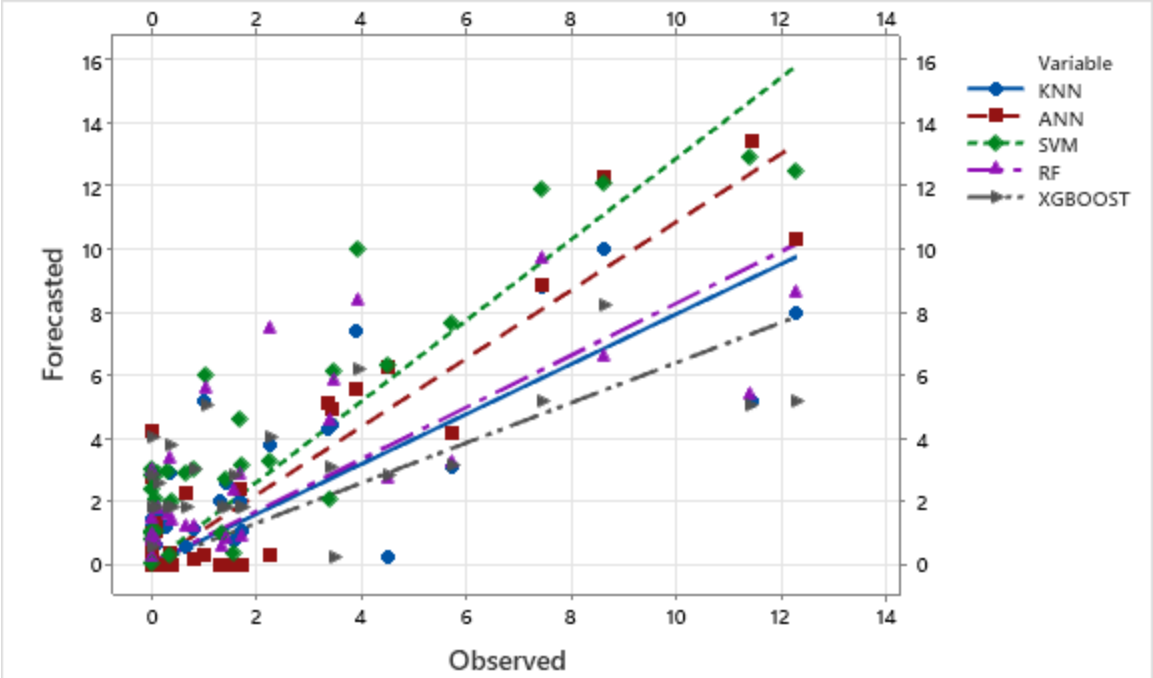


Fig.4.3 (c) KNN, ANN, SVM, RF, XGBOOST Forecasted vs Observed Precipitation @ GN Station

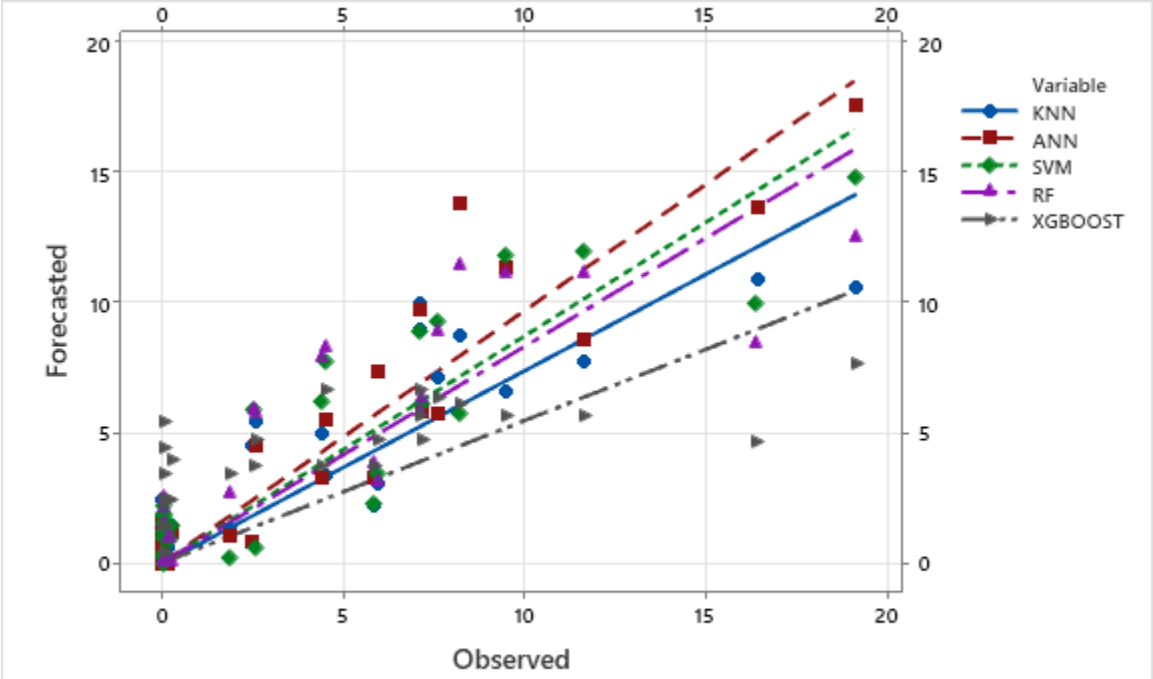


Fig. 4.3 (d) KNN, ANN, SVM, RF, XGBOOST Forecasted vs Observed Precipitation @ PW Station

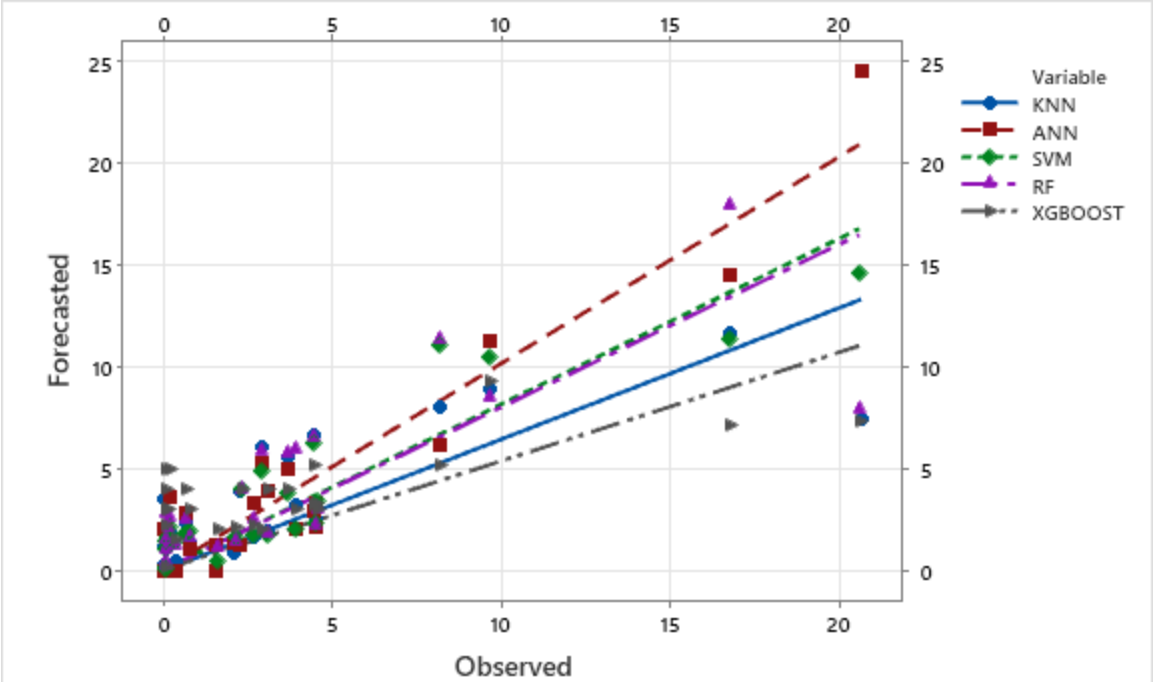


Fig. 4.3 (e) KNN, ANN, SVM, RF, XGBOOST Forecasted vs Observed Precipitation @ DB Station

Fig. 4.3. (a-e) Observed vs Forecasted monthly Precipitation in mm

In this study, we evaluate the performance of a proposed forecasting framework by comparing its predictions with global precipitation reforecast data. The results reveal that the framework consistently demonstrates superior accuracy (see Fig. 4.4) compared to both the NCEP and ECMWF forecasts across all component models. Notably, all models within the framework exhibit higher R^2 values than those predicted by ECMWF and NCEP. These consistently elevated R^2 values indicate a strong correlation between the observed and forecasted precipitation data, suggesting that the framework is more effective in replicating real-world precipitation patterns.

Moreover, our comparative analysis shows that the RMSE values for all models within the framework are significantly lower than those of the global forecasts from ECMWF and NCEP. The reduced RMSE values illustrated in Fig. 4.4 reflect a closer alignment between the framework's forecasts and the actual observed precipitation, underscoring a higher degree of accuracy in the framework's forecasting capabilities.

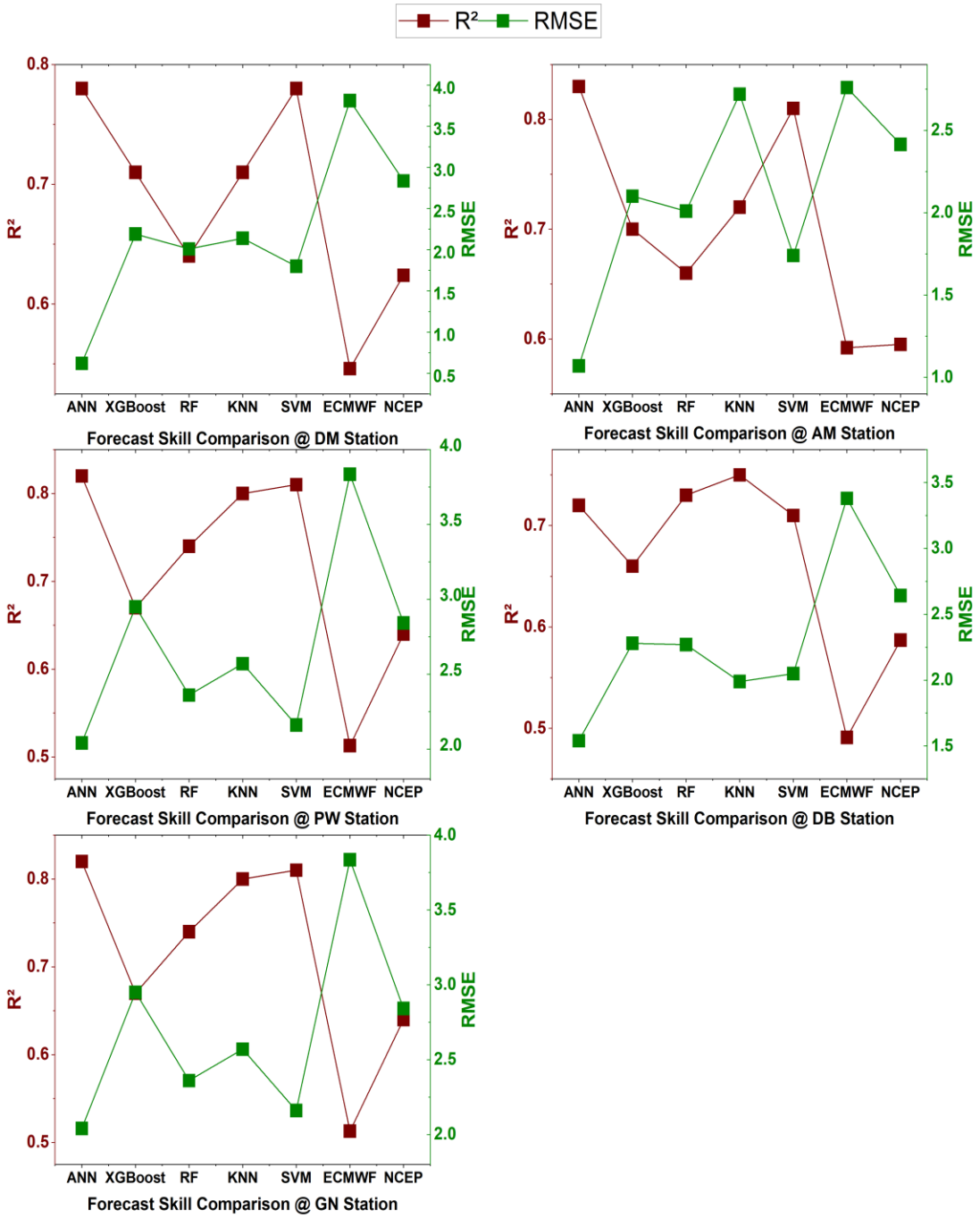


Fig. 4.4. Comparison of forecast skill for the proposed framework against global precipitation forecasts

b. Performance evaluation under different climatic conditions

Normal Months: As displayed in the graphs in Fig.4.5 below, the forecasting framework exhibited remarkable performance during normal months. All models in the forecasting framework demonstrated notable forecasting skill. For instance, Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-Nearest Neighbors (KNN) consistently achieved R^2 values exceeding 0.8 for almost all monitoring stations. Meanwhile, Extreme Gradient Boosting (XGBoost) and Random Forest (RF) also demonstrated reasonably good performance with R^2 values above 0.7, although their performance was suboptimal compared to the other three models. Moreover, a similar performance trend was observed based on accuracy metrics. All models in the forecasting framework recorded relatively lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values. ANN outperformed the other models with MAE and RMSE values of 0.588 and 0.814, respectively. However, as shown in the graph, XGBoost and RF exhibited slightly higher MAE and RMSE values.

Dry Months: Compared to normal months, the performance of the forecasting framework showed a slight drop in forecasting accuracy during dry months. Despite this reduction, the models maintained a respectable level of performance. Specifically, the R^2 values during dry months ranged from 0.701 to 0.866 for XGBoost and ANN, respectively. As presented in Fig. 4.5, SVM and KNN demonstrated competitive performance, with R^2 values around 0.75. We also found similar evidence from the accuracy metrics RMSE and MAE. In terms of MAE, all models in the forecasting framework exhibited values below 2. Furthermore, it is interesting to note that XGBoost and Random Forest displayed relatively lower performance during dry months compared to the other models.

Wet Months: We observed that the forecasting performance during wet months presented unique challenges, with all models in the framework struggling to predict accurately under wet conditions, as indicated by comparatively lower performance metrics. For example, Artificial Neural Networks (ANN) continued to demonstrate the best performance even during wet months, although with diminished accuracy compared to normal and dry periods. Meanwhile, SVM and KNN also exhibited intermediate performance in predicting wet months, suggesting that their modeling approaches managed to capture some of the

complexities associated with high-precipitation events. However, Extreme Gradient Boosting (XGBoost) and Random Forest (RF) consistently displayed the lowest performance across all evaluation criteria during wet months**.** The observed decline in model performance shown above during wet months may be attributed to several interconnected factors. The available record spans only eight years (2010–2017), during which wet months constitute approximately 38% of the data, with months exceeding the 95th percentile represented by only a handful of samples, rendering robust learning of extreme patterns inherently difficult. As a result, the models show a tendency towards average wet conditions instead of capturing the upper tail of the distribution. The situation was further intensified by the Mean Squared Error loss function that gives equal weight to every month, making rare extreme events insignificant contributors to training error. In addition, the rainfall in UNB during the wet season is dominated by convective processes that are chaotically unpredictable, while dry season stratiform rainfall is more predictable; also, regression-toward-the-mean effects systematically compress extreme predictions. Given these limitations, it has a notable impact on flood management, increasing the risk of false negatives during critical peak flow periods and potentially inadequate support for reservoir operations. Future studies could mitigate these limitations through weighted loss functions, assigning higher penalties to wet months, and extension of the temporal record to capture more extreme events.



Fig. 4.5. Performance evaluation of the forecasting framework under different climatic conditions.

In general, the analysis of model performance across dry, normal, and wet months revealed variations in predictive accuracy. We found that Artificial Neural Networks (ANNs), owing to their flexible architecture, consistently maintained their lead in accuracy across different climatic conditions. On the other hand, Support Vector Machines (SVM) and K-

Nearest Neighbors (KNN) followed closely, showcasing their adaptability to diverse precipitation patterns. However, models like Extreme Gradient Boosting (XGBoost) and Random Forest (RF) exhibited more variability in performance, particularly under extremely wet conditions. These findings contribute valuable insights into the framework's suitability for different precipitation scenarios and can guide future refinement efforts.

4.8. Conclusion

In this study, we introduced a novel forecasting framework that integrates multi-source data with machine learning models to enhance the accuracy of monthly precipitation forecasts. The results highlighted improvements in one-month-ahead precipitation forecasts. The framework's models, including ANN, SVM, and KNN, performed best, although Random Forests (RF) and Extreme Gradient Boosting (XGBoost) showed a slight drop in performance. Despite a minor performance drop during the testing phase, all models within the forecasting framework demonstrated generalizability to unseen data, reaffirming their efficacy for real-world precipitation forecasting.

The analysis further delved into the forecasting framework's performance under different meteorological conditions, revealing that ANN, SVM, and KNN excel during normal months, to provide highly accurate forecasts. In contrast, Extreme Gradient Boosting and Random Forest exhibited a slight reduction in forecast accuracy. While experiencing a slight performance reduction during dry months, all models within the forecasting framework remained competitive. However, the models faced challenges in predicting precipitation during wet months, with ANN demonstrating better predictive capabilities in this scenario.

Overall, this research contributes valuable insights for improving precipitation forecasts using multi-source data and machine learning models. However, the performance of the models may vary across different regions and climatic conditions, and the results are specific to the Nile basin. Furthermore, the study is limited by the number of satellite and reanalysis data sources, as well as the machine learning models evaluated. Therefore, we recommend conducting a comprehensive assessment of the framework's efficacy using a broader array of data sources and machine learning approaches, including hybrid and deep learning models, as well as optimization techniques across diverse geographical and

climatic conditions. Future research may also explore various data processing methods, such as data decomposition and sensitivity analysis, to further enhance the framework's capabilities.

5. Physics- Informed Hybrid Forecasting Framework for Improving Monthly Streamflow Forecasting: Integration of Process-Based Hydrological Models, Multi-Source Data, and Machine Learning Techniques

Abstract

Despite significant advances, accurate streamflow forecasting remains a critical challenge in hydrological science due to the lack of sufficient data, inherent complexity, nonlinearity, and spatiotemporal variability of hydrological processes. To enhance the accuracy of streamflow forecasts, this study develops and evaluates a physics-informed hybrid modeling framework that integrates multi-source data with process-based models and machine learning techniques. In the framework, we first employed the calibrated Soil and Water Assessment Tool (SWAT) as a physics-informed feature extractor, generating key intermediate hydrological variables (e.g., surface runoff, baseflow, evapotranspiration) in the Guder sub-basin of the Upper Blue Nile Basin, Ethiopia. These outputs of the SWAT model, combined with local meteorological variables and large-scale climate indices, form a set of predictor pools for a Long Short-Term Memory (LSTM) network to make one-month-ahead streamflow forecasting. The multi-source data integrated SWAT-LSTM hybrid model was trained and tested against observed monthly streamflow, and its performance was benchmarked against three base models, namely standalone LSTM, Random Forest (RF), and Support Vector Regression (SVR) models. Results demonstrate the hybrid model's exceptional performance, achieving a Nash-Sutcliffe Efficiency (NSE) of 0.90, a Kling-Gupta Efficiency (KGE) of 0.89, and a low Percent Bias (PBIAS) of 2.1%, significantly outperforming all benchmark models across all statistical metrics (e.g., NSE values for standalone LSTM, SVR, and RF are 0.76, 0.69, and 0.68, respectively). The integration of multi-source data, especially large-scale climate indices, played a critical role in diminishing systematic bias and improving the representation of climatic drivers. The proposed hybrid framework effectively connects process-based understanding with data-driven prediction, offering a robust, accurate, and physically-consistent operational tool suitable for streamflow forecasting and water resource decision-making in complex watersheds.

Keywords: Multi-Source Data, Hybrid Modeling, LSTM, SWAT, Climate Indices

5.1. Introduction

Accurate streamflow forecasting is a crucial component of effective water resource planning and management, providing vital information for flood control, drought mitigation, irrigation scheduling, agricultural planning, and hydropower operations. Nonetheless, the reliability of streamflow forecasts continues to pose a significant challenge owing to the complexity, nonlinearity, and variability of the hydro-meteorological processes that occur across diverse spatio-temporal scales (Solanki et al., 2025; Wan et al., 2024). Watersheds respond dynamically to a multitude of interacting factors, including precipitation, temperature, evapotranspiration, topography, land use, and soil characteristics. However, these interactions are not only spatially and temporally variable but also subject to uncertainties arising from data limitations, changing climates, and human interventions (Moreno-Pájaro et al., 2025; Tsvetkova & Randhir, 2019).

Streamflow forecasting has relied on two dominant modelling paradigms: process-based hydrological models and data-driven machine learning models (Adera et al., 2024; Pechlivanidis & Du, 2024; Alizadeh et al., 2021). Process-based models, such as the Soil and Water Assessment Tool (SWAT), TOPMODEL, and Variable Infiltration Capacity (VIC), simulate hydrological processes through the distributed parameterization of watershed structures, land surface processes, and meteorological drivers (Jaiswal et al., 2020; Wang et al., 2018). These models are widely used for watershed-scale modelling dynamics because they can represent processes such as surface runoff, baseflow, evapotranspiration, and soil moisture dynamics (Ougahi & Rowan, 2025; Upadhyay et al., 2019) and provide insight into the principles that govern streamflow responses under different climatic and land management scenarios.

However, a notable challenge in calibrating process-based models is equifinality, where different parameter combinations yield similar hydrological simulation results (Ougahi & Rowan, 2025; Duethmann et al., 2020). Moreover, their application encounters uncertainties in the model structure and the accuracy of the input data. These limitations are compounded by the model's inability to capture highly nonlinear input-output relationships, particularly under non-stationary or data-scarce conditions (Adera et al. 2024; Nerantzaki et al., 2020), and their insufficient accuracy in simulating low and high flow (Li et al., 2023; Leta et al., 2022; Duethmann et al., 2020).

Concurrently, machine learning (ML) models such as the Artificial Neural Network, Support Vector Machine, XGBoost, Random Forest, and Long Short-Term Memory have gained significant attention in hydrological forecasting for their ability to model complex, nonlinear relationships directly from data without requiring the knowledge of the underlying physical processes (Hasan et al., 2024; Waqas & Humphries, 2024; Zakaria et al., 2023). Furthermore, their capacity to assimilate heterogeneous inputs and capture long-term temporal dynamics renders them particularly suitable for streamflow prediction (Solanki et al. 2025; Waqas & Humphries, 2024).

However, purely correlative machine learning approaches suffer from their reliance on historical data alone, which often leads to poor generalization in unseen conditions (Wan et al., 2024; Mie et al, 2024), exhibit limited extrapolation capacity beyond training regimes (Gauch et al., 2020; Konapala et al., 2020), and lack a connection to the physical behavior of the watershed (Asadi et al., 2025; Hasan et al., 2024). In addition, it is challenging to answer specific scientific questions using machine learning models, such as the relationship between variable soil moisture and streamflow (Huang et al.,2024; Willard et al., 2023), as the internal relationships of machine learning models may not be straightforwardly interpretable by humans (Asadi et al., 2025; Yokoo et al., 2021; Schmidit et al., 2020). This "black-box" nature of many high-performing ML models, including ensemble tree-based algorithms and neural networks, restricts their transparency, credibility, and acceptance in crucial fields such as streamflow forecasting, groundwater management, and climate adaptation planning (Habibi & Hassanpour, 2025).

This methodological dilemma presents a considerable challenge within contemporary hydrology: the selection between process-based models, which provide a high degree of physical realism and robust explanatory power but require extensive data and substantial computational resources, and data-driven approaches, which offer predictive efficiency and agility but often lack interpretability. Addressing this trade-off is crucial for the advancement of more accurate and reliable forecasting frameworks in hydrological studies (Habibi & Hassanpour, 2025; Chen et al., 2023).

To address these limitations, recent research has focused on hybrid frameworks that integrate process-based hydrological models with machine-learning techniques. These

approaches use streamflow simulations from conceptual models as inputs to enhance the predictive accuracy and robustness of the machine learning models (Solanki et al., 2025; Pechlivanidis & Du, 2024). By leveraging the complementary strengths of both paradigms, hybrid models simultaneously mitigate the key weaknesses of purely data-driven models (e.g., generalization challenges and interpretability issues) and physically based models (e.g., computational intensity, parameter equifinality, and input data uncertainty) (Yuhao et al., 2025; Huang et al., 2024; Mei et al., 2024; Chen et al., 2023).

Accordingly, this study proposes a hybrid modelling framework that integrates multi-source data, the physically based SWAT model, and the deep learning capabilities of LSTM networks. In this framework, SWAT was used not simply as a standalone prediction tool but as a physics-informed generator of intermediate hydrological variables such as surface runoff, and evapotranspiration, which served as informative inputs to the LSTM model. This integration allows the LSTM to be guided by process-based insights, helping it learn more realistic temporal patterns and relationships relevant to streamflow generation (Ahmed et al., 2025; Chen et al.,2023; Oruche et al., 2021; Feng et al., 2020). Unlike previous studies, such as Tao et al. (2024), Huang et al. (2024), and Chen et al. (2023), which used the SWAT output and some meteorological variables as input in the hybridization process, a distinguishing feature of this research is the incorporation of multi-source data, including large-scale climate indices, to enrich the modeling process. These heterogeneous data sources provide complementary information that can address limitations in the quality of input datasets and enhance both physical modeling with SWAT and the learning capacity of the Long Short-Term Model (LSTM) (Asadi et al., 2025; Huang et al., 2024).

In light of these findings, the proposed physics-informed hybrid forecasting framework enhances forecasting accuracy while bridging the critical gap between process-based understanding and data-driven prediction skills. This synergy facilitates the development of computationally efficient models. Furthermore, by leveraging multi-source data, the framework increases model robustness against data gaps and uncertainties, making it a highly suitable solution for operational forecasting in diverse water-resource management applications.

5.2. Description of the Study Area

The Guder sub-basin is located in the central Ethiopian highlands and forms a key tributary of the Upper Blue Nile Basin (Abay Basin) in Ethiopia. Geographically, (see Fig. 5.1) it is located between 9°50'–8°40'N latitude and 37°20'–38°10 'E longitude in the southeastern part of the Blue Nile Basin (Teshome et al., 2023; Duguma, 2022). The topography of the Guder sub-basin is characterized by highly dissected escarpments, valleys, and steeply undulating terrain, with altitudes ranging from 909 to 3,378 m above sea level. The complex topography of the sub-basin significantly affects the local microclimate patterns and hydrology of the basin as a whole.

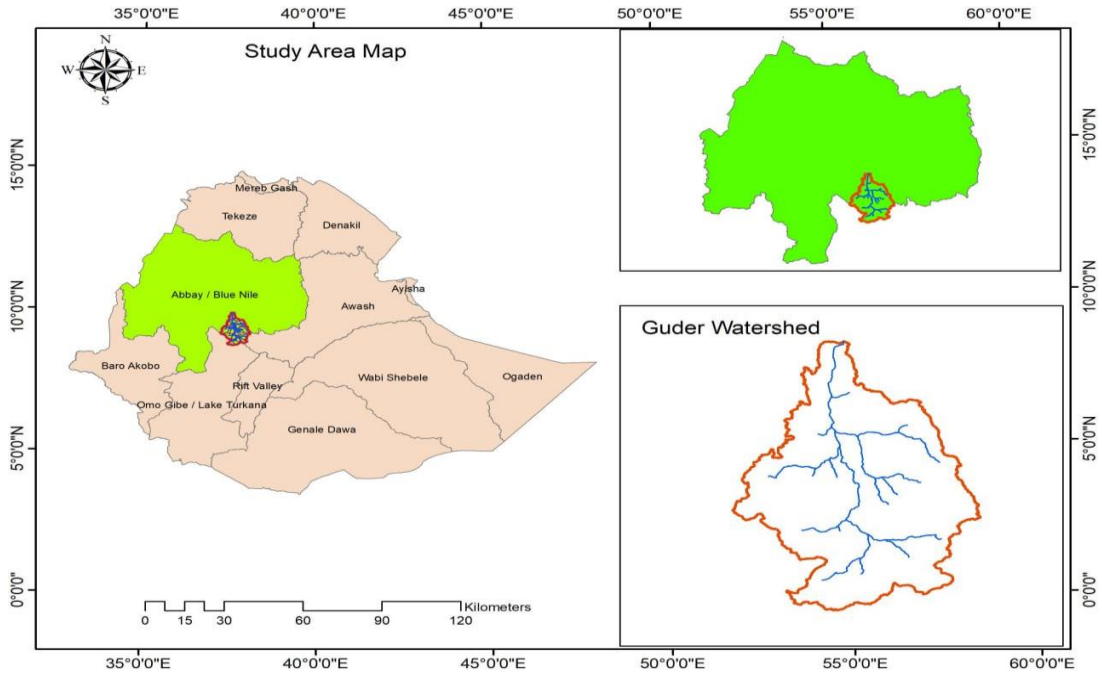


Fig. 5.1. Study Area Map

The Guder sub-basin is bordered by the Muger sub-basin to the east, the Awash Basin to the south, and the Fincha sub-basin to the west. The river flows from south to north and eventually joins the Abbay (Upper Blue Nile) River. The sub-basin is of great importance because of its role in supporting agriculture, providing water resources for local communities, and contributing significant flow and sediment to the Blue Nile River.

5.2.1. Climate and Hydrology

The Guder sub-basin experiences a subtropical highland climate moderated by its elevation, with distinct seasonal precipitation patterns where the rainy season may be divided into

major (Kiremt) and minor rainy seasons (Belg and Tsedey). The mean annual rainfall of the sub-basin averages 1,228–1,445 mm, characterized by a unimodal distribution where 85–90% occurs during the June–September monsoon season (Demessie et al., 2023; Ebabu et al.,2022). During the Belg season, which spans March to May, the basin receives a small amount of rainfall. In the Guder sub-basin, the dry season (Bega) extends from December to February, and Tsedey from September to November, which creates a pronounced wet-dry seasonality in the basin's climatology.

The temperature regime of the basin exhibits moderate annual variation owing to its equatorial proximity but significant diurnal fluctuations. For instance, mean temperatures range between 9.3°C (minimum) and 23.7°C (maximum), with the hottest months (March–April) preceding the monsoon onset and the coolest period (July–August) coinciding with the peak rainfall (Tofu, 2025; Haile & Abebe, 2022). Recent climate model studies by Gemechu et al. (2021) indicate that the region is likely to experience a decrease in precipitation and a significant rise in temperature under future climate scenarios, with projections of up to a 14.4% reduction in rainfall and a 4.4°C temperature increase by the 2080s under RCP 8.5. These changes are expected to reduce water yield and increase evapotranspiration, thereby affecting water availability for agriculture and domestic use in the sub-basin.

Hydrologically, the Guder sub-basin is characterized by a dense drainage network with numerous tributaries feeding into the main Guder River and strong discharge seasonality driven by regional rainfall distribution (Gemechu et al., 2021). The Guder River exhibits a variable flow pattern, with a high-flow period from June to September, which contributes 80–90% of the annual discharge, and a low-flow period from November to April that is baseflow-independent, with frequent stream desiccation (Duguma, 2022).

5.2.2. Land Use and Soil

Land use within the Guder sub-basin is dominated by agricultural activities, with croplands constituting the most extensive land cover type (Demessie et al.2023; Douguma et al. 2021; 2022). The remaining landscape comprised complementary cover classes, including forests, grasslands, bushlands, bare land, urban settlements, and water bodies. In recent decades,

significant land cover transitions have occurred, characterized by the expansion of agricultural and built-up areas and a concurrent reduction in forest and grassland extent (Tofu, 2025). This shift in land use has precipitated adverse environmental consequences, notably exacerbating soil erosion and modifying the hydrological processes. These modifications are manifested through increased streamflow during the wet seasons and diminished baseflow during the dry periods. Furthermore, the soils of the Guder sub-basin are diverse, with major types including Chromic Vertisols, Lithic Leptosols, Cambisols, Haplic Acrisols, and Chromic Luvisols (Ebabu et al., 2022; Duguma, 2022).

5.3. Datasets

5.3.1. DEM

A Digital Elevation Model (DEM) is a critical input for watershed delineation and stream network construction in hydrological studies. In this study, the Shuttle Radar Topography Mission (SRTM) DEM with a 30 m resolution was employed, ensuring adequate representation of terrain features that influence the hydrology of the study area. The DEM was preprocessed to fill sinks, define flow directions, and accumulate flow. Furthermore, the DEM was utilized to generate the slope and to overlay the land use, land cover, and soil data needed to define the hydrologic response unit (HRU) in the SWAT model.

5.3.2. Soil Data

The SWAT model requires soil data to determine the hydrological characteristics of each soil type within each sub-watershed and HRU. Soil data with 1-km resolution were obtained from the Ministry of Water and Energy (MoWE) of Ethiopia. The associated soil properties were obtained from the Harmonized World Soil Database (FAO downloaded from http://www.fao.org/data/en/.

5.3.3. LULC data

The LANDSAT imagery data were downloaded from the US Geological Survey https://earthexplorer.usgs.gov/. ERDAS Imagine 2014 was used to prepare the LULC map from the LANDSAT imagery. During the land-use image classification procedure, the maximum likelihood algorithm was used for supervised image classification. For further hydrological simulation, the entire spatial dataset was resampled to a 12.5 m resolution and projected to WGS 1984 UTM Zone 37 using QGIS.

5.3.4. Hydrometeorology data

This study utilized precipitation data, created from a blend of ground measurements, satellites, and reanalysis precipitation products using the RFmep merging procedure illustrated in Chapter Three of this dissertation. Furthermore, daily minimum and maximum air temperature, relative humidity, sunshine hours, and wind speed data were obtained from the Ethiopian National Meteorological Service Agency (NMSA). Daily streamflow data from 2010 to 2017 were obtained from the Ethiopian MoWE. Large climate index data, including NINO3.4, Southern Oscillation Index (SOI), Atlantic Multidecadal Oscillation (AMO), Dipole Mode Index (DMI), Quasi-Biennial Oscillation (QBO), Tropical North Atlantic Index (TNA), and Indian Ocean Basin-Wide warming (IOBW), were downloaded from https://www.esrl.noaa.gov/psd/data/climateindices/list/. These hydrologic and climatic data were used to create the hydrological and climate balance components and to calibrate and test the hydrological model.

5.4. Methodology

5.4.1. Description of the SWAT model

The Soil and Water Assessment Tool (SWAT) is a semi-distributed process-based hydrological model that is widely applied for simulating water balance components on continuous time scales in large and complex watersheds (Chen et al.,2023; Leta et al., 2022). It can operate on a daily or hourly timescale and over long periods to model streamflow, sediment yield, and nutrient dynamics at catchment scales. SWAT divides a watershed into sub-basins linked by a stream network and further discretizes these sub-basins into Hydrologic Response Units (HRUs), each characterized by unique combinations of soil, slope, and land use (Ren et al., 2025; Huang et al., 2024; Gemechu et al., 2021).

This spatial discretization allows the model to capture the heterogeneity and simulate hydrological processes, such as surface runoff, infiltration, lateral subsurface flow from the soil profile, evapotranspiration, groundwater recharge, and streamflow routing from each hydrological response unit. Surface runoff can be computed using the Soil Conservation Service curve number method (SCS-CN, 1972) or the Green-Ampt infiltration method Zhu et al., 2020). The potential evapotranspiration (PET) can be estimated by the Penman-

Monteith, the Priestly-Taylor, and the Hargreaves methods (Raja et al. 2024). In this study, the Hargreaves method was used to calculate PET, whereas the SCS-CN method was used for surface runoff.

The watershed concentration time was estimated using Manning's formula for both overland and channel flows (Wahyullah et al. 2023). Soil profiles are divided into multiple layers, which influence soil water processes such as plant water uptake, lateral flow, percolation to lower layers, infiltration, and evaporation. The hydrologic cycle simulated by SWAT is based on the water balance equation (5.1).

$$SW_t = SW_o + \sum_{i=1}^{t}(R_{day} - Q_{surf} - E_a - W_{sweep} - Q_{gw}) \qquad \text{Eq. 5.1}$$

Where SW_t and SW_o are the final and initial soil water content respectively in (mm), t is the time in days, R_{day} is the amount of precipitation on day i in (mm), Q_{surf} is the amount of surface runoff on day i (mm), Ea is the amount of evapotranspiration on day i (mm), W$seep$ is the amount of water entering the vadose zone from the soil profile on day i (mm) and Qgw is the amount of return flow on day i (mm).

5.4.2. Model Setup

In this study, the ArcSWAT model is employed to simulate streamflow dynamics in the selected watershed. ArcSWAT is the GIS-based interface of SWAT, providing a user-friendly environment for watershed delineation, Hydrologic Response Unit (HRU) creation, model execution, and evaluating hydrological processes (Premanand et al., 2018). The outputs of ArcSWAT are the basis for subsequent calibration, validation, and sensitivity analyses using SWAT-CUP, with the SUFI-2 algorithm to optimize parameter estimation and quantify uncertainties.

5.4.3. Calibration and Validation

The model was set to run on a daily time step, which ensured an accurate representation of rainfall-runoff processes for the period from 2010 to 2017. The QSWAT model was calibrated using SWAT-CUP with the Sequential Uncertainty Fitting version 2 (SUFI-2) algorithm to match the simulated streamflow to the observed records at the watershed outlet and compared with the observed flows at the gauging stations. This algorithm estimates and optimizes the parameters and accommodates uncertainties. The calibration period

between 2010 and 2014 was chosen based on the available hydrological data and focused on adjusting parameters that have the most impact on hydrological responses, such as surface runoff, evapotranspiration, groundwater, soil water, and channel routing. The performance of the model was evaluated using statistical metrics such as the Nash–Sutcliffe Efficiency (NSE) and percent bias (PBIAS). After calibration, the model was validated using an independent time period (2015-20177) without any further parameter adjustments to test the model's robustness and predictability. The validation results were evaluated using the same statistical criteria, and the consistency of performance across both periods gave us confidence that the QSWAT model can reproduce streamflow, evapotranspiration, and groundwater flow dynamics in the study area.

5.4.4. Sensitivity Analysis

A global sensitivity analysis was conducted to identify the most influential parameters controlling streamflow simulation in the study watershed. This analysis was performed within the SWAT-CUP framework using SUFI-2, which quantifies the relative impact of each parameter based on its t-statistic and corresponding p-value. Parameters with larger absolute t-values were considered highly sensitive, whereas those with smaller p-values indicated greater statistical significance. The sensitivity analysis provided critical insights into the dominant hydrological processes within the basin and guided the calibration procedure by highlighting the parameters that required the most careful adjustment to improve the model performance.

5.5. Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks are a specialized type of recurrent neural network architecture with feedback connections that allow them to exploit temporal dependencies across sequences of data (Oruche et al., 2021; Alizadeh et al., 2021). They are designed to address the vanishing or exploding gradient problem, which is common when training traditional RNNs on data sequences (Asadi et al., 2025; Mei et al., 2024). This feature is especially useful in hydrology, where processes such as streamflow and rainfall-runoff transformation depend heavily on long-term temporal patterns that can extend over days or months (Alizadeh et al., 2021; Feng et al., 2020; Xu et al., 2020). Because LSTMs include memory cells and a system of three gates that regulate information flow, they can selectively remember or forget information over extended periods, making

them effective at capturing historical patterns (Mei et al., 2024; Xu et al., 2020). These gates enable the network to maintain a balance between retaining long-term dependencies and adapting to new input data. This structure is key to resolving the gradient vanishing problem inherent in simpler RNN architectures (Anshuka et al., 2022; Oruche et al., 2021).

The operation of an LSTM cell at each time step, t, is mathematically defined by the following equations:

Forget Gate (f_t): The forget gate determines which information to retain or discard from the memory cell. It takes the current input and the previous hidden cell state C_{t-1} as inputs and, using a sigmoid activation function, produces output values between 0 and 1 for each element of the memory cell. A value of 0 indicates that the information is disregarded, whereas a value of 1 indicates that it is retained.

$$f_t = \sigma(W_f * [h_{t-1,} x_t] + b_f) \qquad \text{Eq. 5.2}$$

Input Gate (i_t) and Candidate Cell State (): This gate determines which new information from the current input x_t should be stored in the cell state. The input gate (i_t) uses a sigmoid function to filter the new inputs to be remembered, similar to the forget gate that uses inputs h_{t-1} and x_t. The candidate cell state, , generates a vector of potential information using a hyperbolic tangent (tan_h) function that gives an output from -1 to +1, which contains all the possible values from h_{t-1} and x_t. Finally, the values of the vector and the regulated values are multiplied to obtain useful information. The equation for the input gate is as follows:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b) \qquad \text{Eq. 5.3}$$

$$\widehat{C_t} = tan_h(W_c * [h_{t-1}, x_t] + b_c] \qquad \text{Eq. 5.4}$$

Cell State Update (Ct): The new cell state is a weighted combination of the old cell state and the new candidate information. This is calculated by an element-wise multiplication of the previous cell state by the forget gate's output, which is then added to the element-wise product of the new candidate state and the input gate's output.

$$C_t = f_t \cdot \odot C_{t-1} + i_t \odot \hat{C}_t \qquad \text{Eq.5.4}$$

where

- $\odot$ denotes element-wise multiplication; tanh is activation function

Output Gate (O_t) and Hidden State (ht): The output gate determines which parts of the updated cell state are passed to the next layer (time step). First, the gate uses a sigmoid function to determine which information from the current cell state will be output. This is done using the previous hidden state, $h_{t-1,}$ and the current input $x_{t:}$

$$O_t = \sigma(W_O * [h_{t-1,}x_t] + b_O) \qquad \text{Eq. 5.5}$$

Subsequently, the current cell state Ct is passed through a $\tan_h$ activation to constrain its values between −1 and +1. Finally, this transformed cell state is then subjected to element-wise multiplication with the output tensor (O_t) to yield the hidden state h_t for the current time step.

$$h_t = O_t \odot \tanh(C_t) \qquad \text{Eq. 5.6}$$

Where: O_t is the output gate activation.; Ct is the current cell state; $\odot$ represents element-wise multiplication; σ is the sigmoid activation function.

5.6. SWAT-LSTM Hybrid Model

In this study, we developed a multi-source data-integrated hybrid SWAT–LSTM framework shown in Fig. 5 .2 to leverage the combined strengths of process-based hydrological modeling with the learning capabilities of deep neural networks for monthly streamflow forecasting in the Guder sub-basin of the Upper Blue Nile Basin. First, the calibrated Soil and Water Assessment Tool (SWAT) was applied during the study period of 2010-2017 to simulate key hydrological variables, including baseflow, surface runoff, and evapotranspiration, which are commonly employed as predictors for streamflow forecasting in the literature (Ahmad et al., 2025; Farmani et al., 2025; Huang & Zhang,2024; Yang et al., 2023; Li et al., 2023; Chen et al., 2023; Tongal & Booij, 2018). These outputs capture the internal catchment processes that govern the water balance and provide a physically consistent memory of watershed dynamics. Unlike raw meteorological inputs alone to the LSTM model, SWAT-generated outputs incorporate the nonlinear transformation of precipitation and temperature through soil, land use, and topography, thereby offering meaningful hydrological signals for stream flow forecasting, which improves forecasting accuracy.

To enhance the predictor pool, SWAT outputs were complemented with local meteorological variables (precipitation, minimum and maximum temperature, humidity, wind speed, and sunshine hours) and large-scale climate indices, such as the El Niño–Southern Oscillation and Indian Ocean Dipole, which are known to influence hydrological variability in the region. We selected predictors using cross-correlation analysis to identify statistically significant contemporaneous and lagged relationships with the observed streamflow values. This is compounded by autocorrelation checks and multicollinearity diagnostics to prevent overfitting, ensuring that the final set of predictors is both physically understandable and statistically relevant. The selected predictors were then used as inputs to a Long Short-Term Memory (LSTM) neural network to form a hybrid setup, where SWAT functions as a physics-based feature extractor, while LSTM serves as a nonlinear mapping tool that links predictors to observed streamflow. We trained the LSTM model to predict monthly streamflow from 't + 1' using the selected predictors as input sequences, where each input window incorporated both concurrent and lagged values to account for memory effects. The network architecture for this forecasting task was optimized through hyperparameter tuning, including the number of hidden layers, memory cells, dropout rates, and learning rate, and the performance was validated via k-fold cross-validation.

Additionally, to enforce physical consistency, the LSTM training loss function (Eq. 5.7) was augmented with two regularization terms: (1) a non-negativity penalty to prevent negative streamflow predictions, and (2) a mass balance violation penalty that discourages predicted streamflow from exceeding the available water (precipitation minus evapotranspiration plus SWAT-simulated baseflow). Training was conducted on a portion of the observed streamflow record from 2010-2014, while a separate dataset from 2015-2017 was reserved for independent testing to assess the predictive skill of the multi-source-informed hybrid forecasting framework.

$$L = MSE + \lambda_1 \cdot ReLU(-Q_pred) + \lambda_2 \cdot ReLU(Q_pred - (P - ET + Q_baseflow)) \quad \text{Eq. 5.7}$$

Where Q_pred is the predicted streamflow, P is the precipitation, ET is the evapotranspiration, and Q_baseflow is the SWAT-simulated baseflow.

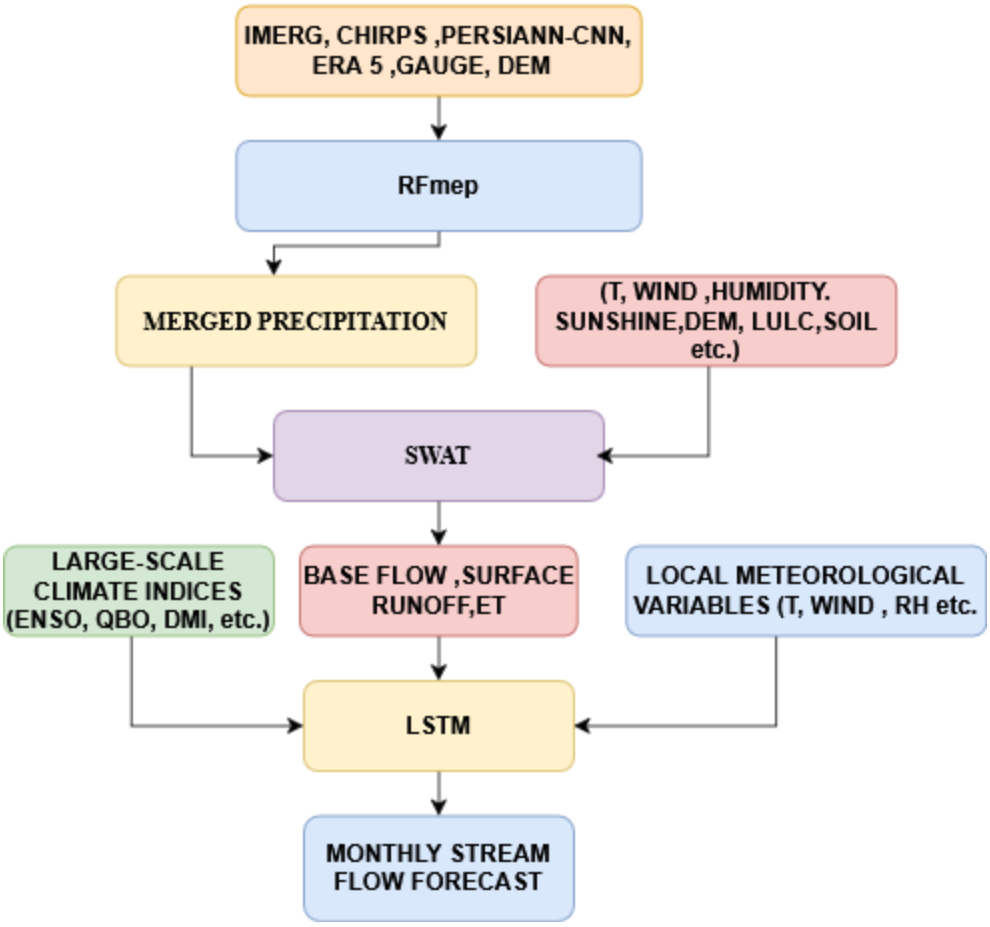


Fig.5.2. The hybrid streamflow forecasting framework

5.7. Result

5.7.1. Performance of the SWAT Model

I. Sensitivity Analysis

To guide calibration and limit model uncertainty, we initially conducted a global sensitivity analysis of 20 key parameters representing major hydrological processes in the Guder sub-basin. These parameters include surface runoff, evaporation, infiltration, soil moisture-holding capacity, groundwater movement, and surface flow routing. They were selected for their substantial influence on runoff generation, baseflow, evapotranspiration, and water transport within the watershed. A total of 500 simulation runs were conducted, and we used t- statistics and p- values to rank parameters by impact. Variables with large t- stat values and small p- values were considered highly sensitive and prioritized in calibration. Accordingly, the analysis identified and ranked the twelve most sensitive parameters, as

shown in Table 5.1. The results indicated that surface runoff processes are the dominant control on the watershed' s hydrological response. For example, the curve number parameter r _ _ CN 2. mgt was the most sensitive, as expected because it directly partitions precipitation into surface runoff and infiltration. Likewise, groundwater parameters exhibited considerable sensitivity, notably the threshold water depth in the shallow aquifer for return flow (v _ _ GWQMN. gw), which ranked second in sensitivity. This highlights the critical role of baseflow in streamflow in the Guder sub- basin, particularly during dry periods. Soil parameters, including the available water capacity (R _ _ SOL _ AWC. sol) and saturated hydraulic conductivity (r _ _ SOL _ K. sol), ranked fifth and seventh, respectively, indicating their moderate influence on soil moisture dynamics and subsurface flow. Parameters related to baseflow recession (v _ _ ALPHA _ BF. gw), channel Manning's coefficient (V _ CH _ N 2. rte), surface runoff lag (V _ _ SURLAG. bsn), and soil evaporation compensation (V _ ESCO. hru) ranked lowest among the twelve parameters. They exhibited p- values exceeding 0. 05, indicating they are less sensitive. When a parameter's p- value is greater than the significance level (here, 0. 0.05), it is usually deemed not significant by the model. Nevertheless, they were retained during calibration to ensure comprehensive representation.

Table 5.1. Ranking of Sensitive Parameters Based on t-Stat and p-Value

S. No	Parameter Name	t-Stat	p-Value	Rank
1	r__CN2.mgt	-12.5	0.00	1
2	v__GWQMN.gw	10.8	0.00	2
3	V__CH_K2.rte	-9.7	0.00	3
4	V__ALPHA_BNK.rte	8.9	0.00	4
5	R__SOL_AWC(..).sol	7.4	0.001	5
6	v__GW_DELAY.gw	6.8	0.002	6
7	r__SOL_K.sol	-5.9	0.003	7
8	V__REVAPMN.gw	5.1	0.04	8
9	v__ALPHA_BF.gw	4.6	0.21	9
10	V__CH_N2.rte	-3.9	0.31	10
11	V__SURLAG.bsn	3.1	0.71	11
12	V__ESCO.hru	2.4	0.74	12

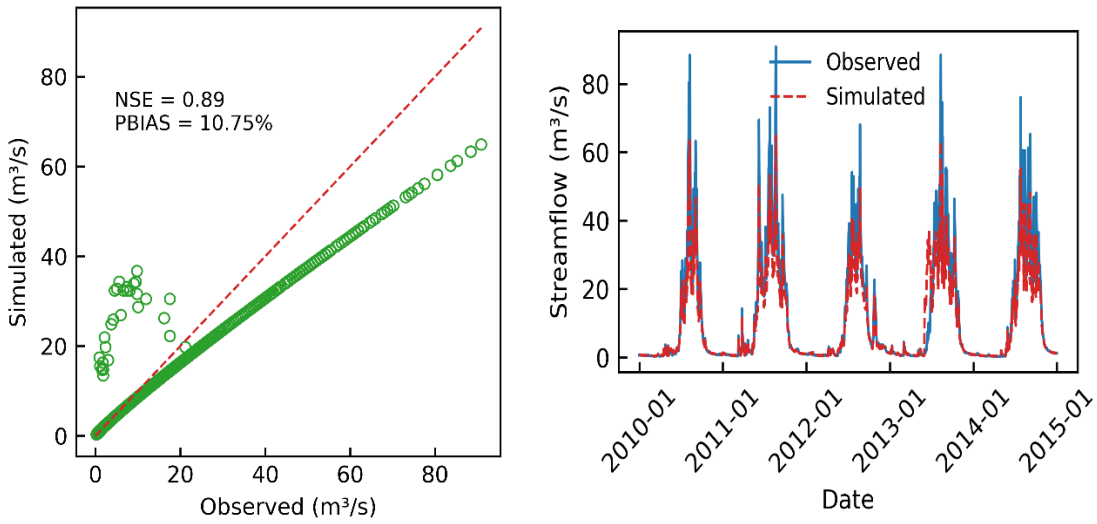
II. Performance of the SWAT model for the calibration and validation period

The performance of the SWAT model in simulating streamflow and key hydrological variables was evaluated through a two-step process of calibration and validation. The

model's accuracy was assessed graphically using time-series plots and scatter plots, and quantitatively using statistical metrics, the Nash-Sutcliffe Efficiency (NSE) and Percent Bias (PBIAS). The results for both periods are presented and discussed below.

a) Calibration Period

The scatter plot for the calibration period is shown in Fig. 5.3a, further confirming this stellar performance. The data points are tightly clustered around the 1:1 line, indicating little bias in the model predictions over the flow conditions. The statistical metrics strongly support this visual assessment. The calibration phase of the model exhibited an excellent performance with the NSE value of 0.89, which is considerably higher than the generally satisfactory threshold value of > 0.5. This indicates a remarkable efficiency and predictive power where the model simulations explain 89% of the variance in the observed stream flow data, signifying excellent model performance for the streamflow simulation model. In contrast, positive PBIAS 10.75% values suggested that the model tends to overestimate stream flow, relative to observed flow. The overestimation rate of 10.75% is regarded as a very good performance since its value is in the range of < ±15%, which is considered a very good model (Cinkus et al. 2023; Moriasi et al., 2007). Furthermore, the graphical presentation of the calibration results demonstrates a very good agreement between the observed and simulated daily stream flows. As shown in Fig. 5.3(b), the simulated hydrograph effectively captures the overall timing and magnitude of the peak flow events throughout the calibration period. The model successfully replicates the recession limbs and the baseflow conditions between storm events, indicating a robust parameterization of the watershed's hydrological processes.



(a) scatter plot of observed vs simulated flow for the calibration period

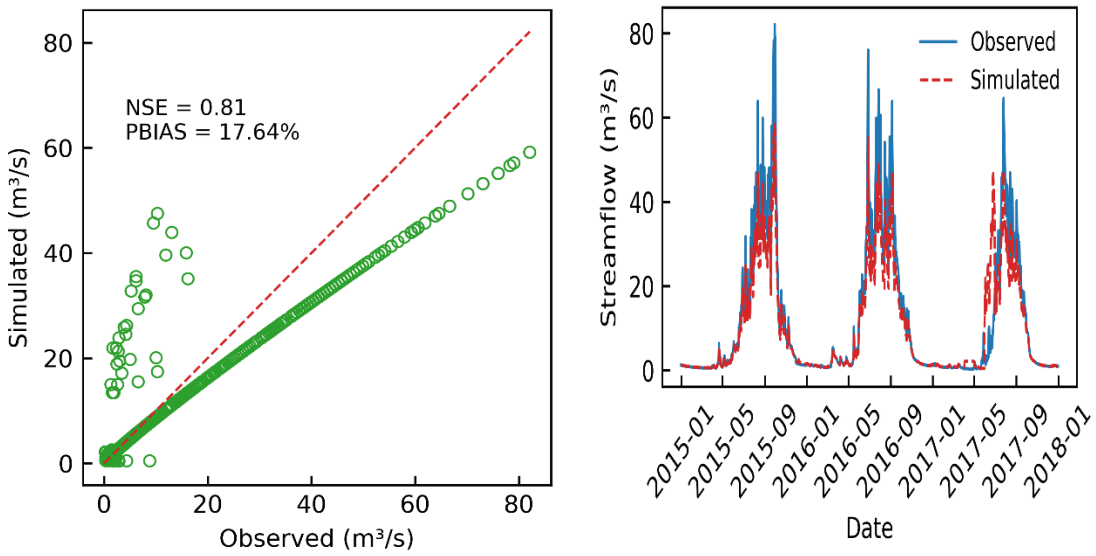
(b) time series plot of observed vs simulated flow for the calibration period

Fig. 5.3. Scatter and time series plot of observed vs simulated flow for calibration period

b) Validation Period

The model's ability to predict streamflow for an independent time period (validation) remained strong, though with a slight decrease in performance compared to the calibration period, which is typical for hydrological modeling (Arnold et al., 2012; Moriasi et al., 2007). Despite this reduction, the 0.81 NSE value remains very good, indicating that the model has retained over 80% of its predictive skill during validation. This confirms that the model has not been overfit to the calibration data and retains its generalizability. The BIAS has increased to 17.64%. This positive value indicates that the model tends to overestimate streamflow during the validation phase. Although it is still deemed within the "satisfactory" rating (if using a threshold of ±15-25%), this represents a degradation from the "very good" performance seen during calibration. This suggests that hydrological conditions or watershed responses during the validation period may have differed slightly

from those during calibration, or that the model has limitations in simulating the highest flow events under the specific conditions of the validation years.



(a) scatter plot of observed vs simulated. flow for the validation period

(b) time series plot of observed vs simulated flow for the validation period

Fig. 5.4 scatter and time series plot of observed vs simulated flow for the validation period

The scatter plot for the validation period (Fig. 5.4a) offers a clear visual assessment of the SWAT model's performance. The data points generally follow the 1:1 line; however, the simulated flow data exhibit noticeably greater scatter than during calibration, particularly for higher flow values (> 40 m³/s). This increased dispersion is also reflected in the percent bias metric. The time-series plot for the validation period (Fig. 5.4b) shows that the model continues to replicate the observed hydrograph reasonably well. The timing of peak flows is generally captured accurately. Nevertheless, some discrepancies are visually evident, especially regarding the magnitude of several high-flow events, where the model tends to overpredict, and a few moderate-flow events, where the fit is less precise than during the calibration phase.

5.7.2. Quantitative performance evaluation of the proposed hybrid framework

To assess the performance of the proposed physics-informed hybrid forecasting framework, which integrates multi-source data, a process-based hydrological model, and a machine

learning technique, we carried out an empirical evaluation. The performance of the proposed Hybrid SWAT- LSTM framework was compared with observed streamflow data and three independent baseline models: standalone LSTM, Random Forest (RF), and Support Vector Regression (SVR), using multiple forecasting skill metrics. The results in Table 5. 5.2 demonstrated a clear and consistent superiority of the hybrid framework across all performance measures. This comprehensive evaluation, based on standard hydrological metrics, provided strong evidence that integrating multi- source data with a process- based hydrological model and machine learning techniques is highly effective.

Table 5.2. Summary of Performance Metrics

Model	R2	NSE	KGE	PBIAS	RMSE	MAE	HSI
LSTM	0.78	0.76	0.75	8.5	15.2	12.3	0.82
RF	0.69	0.68	0.78	12.3	18.7	15.6	0.75
SVR	0.67	0.69	0.58	15.7	21.4	18.2	0.72
Hybrid	0.92	0.90	0.89	2.1	8.5	6.8	0.94

The physics- informed hybrid model (SWAT- LSTM) achieved an exceptional coefficient of determination (R^2) of 0. 92, indicating that the model forecasts explained 92% of the observed streamflow variance. This represented a remarkable improvement over the baseline models, with standalone LSTM (R^2 = 0. 78), RF (R^2 = 0. 69), and SVR (R^2 = 0. 67) all exhibiting substantially lower explanatory power. The improvement over the next-best model (LSTM) was approximately 18%, underscoring the transformative impact of multi- source data integration and model hybridization on forecasting accuracy.

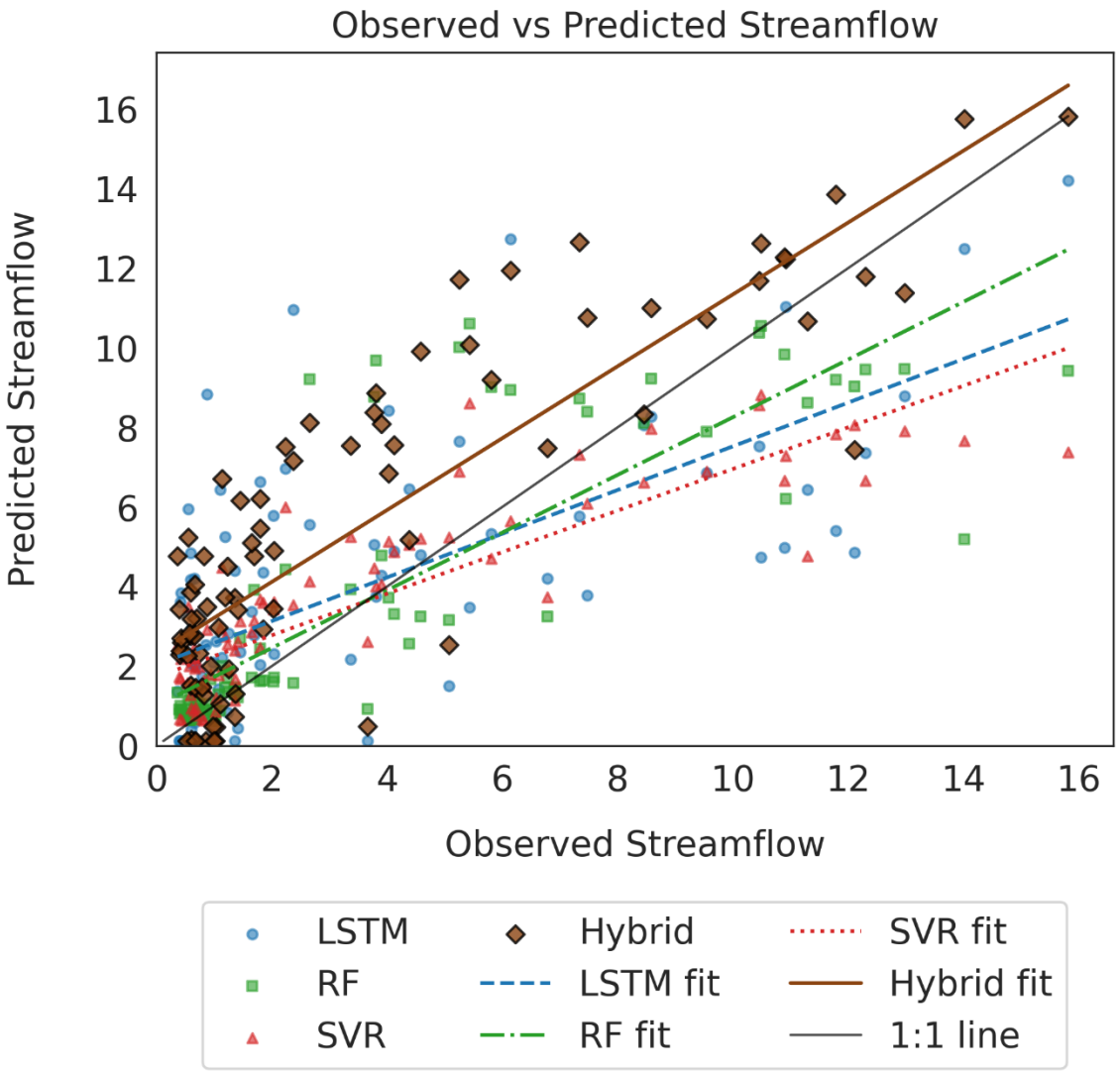


Fig. 5.5 Scatter plot of all models' forecasts vs observed streamflow

As shown in Fig. 5.5 above, the scatterplot effectively reaffirmed these quantitative results, illustrating that predictions made using the hybrid model clustered very tightly along the 1:1 line across all flow regimes. The Nash-Sutcliffe Efficiency (NSE) metric, which is commonly used in hydrological model evaluation, also confirmed the supremacy of the hybrid model in terms of forecasting performance. The hybrid framework had an NSE value of 0.92, indicating near-perfect efficiency (where 1.0 represents a perfect match between observed and simulated hydrographs). This value considerably surpassed the performance of the base models' standalone LSTM (NSE = 0.76), RF (NSE = 0.6897), and

SVR (NSE = 0.6779). Furthermore, the Kling-Gupta Efficiency (KGE), which decomposes model performance into correlation, bias, and variability components, confirmed the hybrid model's performance and balanced capabilities. The model achieved a KGE of 0.89 (shown in Fig. 5.6), whereby it demonstrated excellence in all three components, indicating a high correlation with observations, low systematic bias, and appropriate reproduction of flow variability.

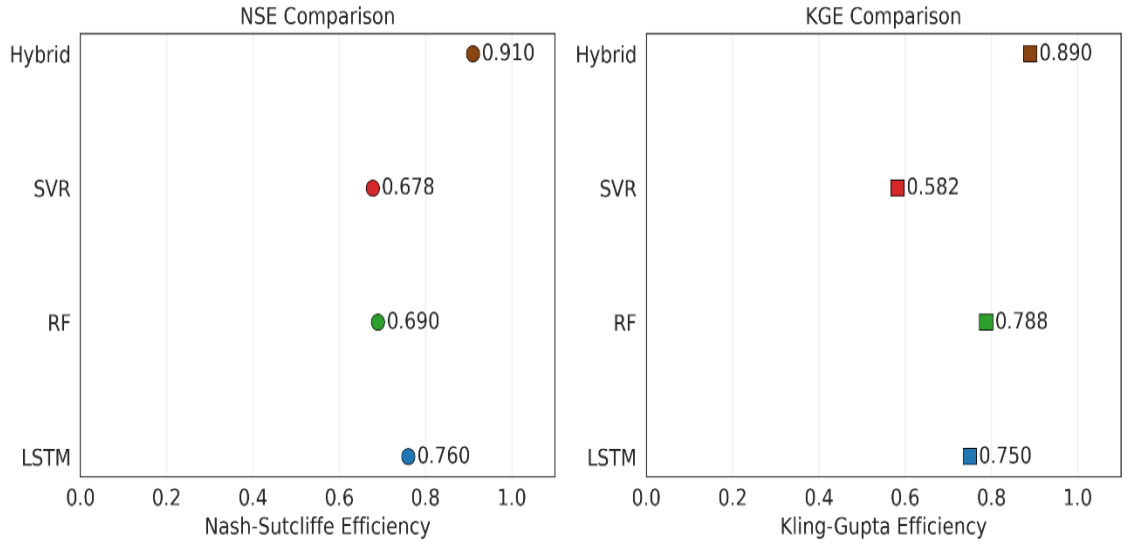


Fig. 5.6. Kling-Gupta and Nash Sutcliffe Efficiency metric-based comparison

The Error-based metrics analysis presented in Table 5.2 also provides notable evidence of how the integration multi- source data into the hybrid model improves the skill of the streamflow forecasts. The hybrid SWAT-LSTM model's Root Mean Square Error (RMSE) of 8.5 represents a 44% reduction compared to the standalone LSTM (RMSE = 15.2), a 55% reduction compared to RF (RMSE = 18.7), and a 60% reduction compared to SVR (RMSE = 21.4). Similarly, the Mean Absolute Error (MAE) of 6.8 of the multi-source data integrated hybrid model shows reductions of 45%, 56%, and 63% relative to LSTM, RF, and SVR models, respectively. These substantial error reductions, a 44% in RMSE and 45% MAE, have direct operational implications for water resource management applications, including water allocation, increasing warning lead times for drought management, and flood hazard prevention.

The hybrid forecasting framework recorded exceptional skill in minimizing systematic bias, as evidenced by its Percent Bias (PBIAS) of 2.1%. This near-perfect forecast presents (Table 5 .2) explicit contrast to the substantial positive biases exhibited by competing models: LSTM (8.5%), RF (12.3%), and SVR (15.7%). The minimal positive bias of the hybrid model indicates systematic over-prediction, reflecting its successful integration of physically-based constraints that enforce fundamental conservation principles, preventing the accumulation of systematic errors that plague purely data-driven approaches. The inclusion of physics-based regularization terms contributed to the hybrid model's low PBIAS (2.1%) by preventing unphysical extrapolations that plagued the standalone LSTM (PBIAS 8.5%).

5.7.3. Visual Diagnostic Analysis

The hybrid model's dominant performance based on quantitative metrics was powerfully reinforced through multiple visual diagnostics that illustrate its enhanced capacity to reproduce both the statistical properties and temporal dynamics of observed streamflow.

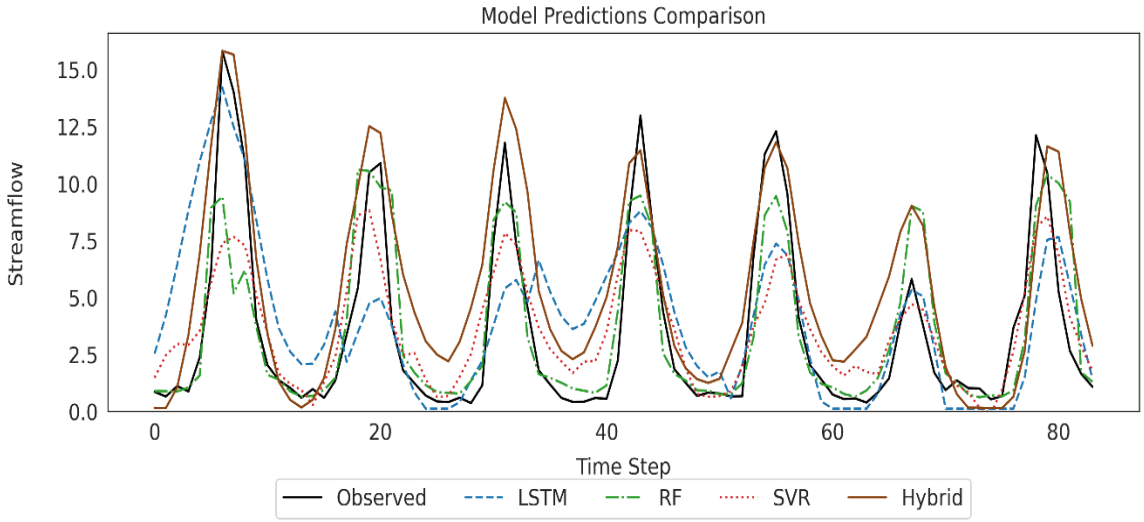


Fig. 5.7. Time series plot of observed vs forecasted for all models

The time series comparison (Fig. 5.7) provided strong evidence that the hybrid model was able to effectively replicate the observed data's temporal patterns. During high-flow events (characterized by sharp peaks and rising limbs in the hydrograph), the hybrid model accurately reproduced both the timing and magnitude of flow variations. On the other hand, for low-flow periods, the shapes of the recession curves mirrored the observations

closely, indicating how the watershed responded dynamically to minor stormwater events. Meanwhile, in both low- and high-flow periods, we found that the standalone LSTM generally followed the overall pattern but exhibited noticeable smoothing, with peak heights diminished and recession phases more gradual than the observed flow. Furthermore, the decline in performance was further exacerbated in the RF and SVR, as shown in Figure 5.7. Their hydrographs were heavily dampened, with peak events flattened into broad, low-amplitude rises that hardly resembled the sharp, distinct peaks seen in the observed streamflow data.

5.7.4. Residual Error Distribution

The violin plots of residuals in Figure 5.8 presented differences in error distributions: the hybrid forecasting framework exhibited a narrow, symmetric distribution centered near zero, with tight concentration. The standalone LSTM shows a wider distribution with positive skew (more frequent over-prediction), consistent with its positive PBIAS. RF and SVR exhibit even broader distributions with complex multi-modal characteristics, reflecting inconsistent performance across different flow conditions.

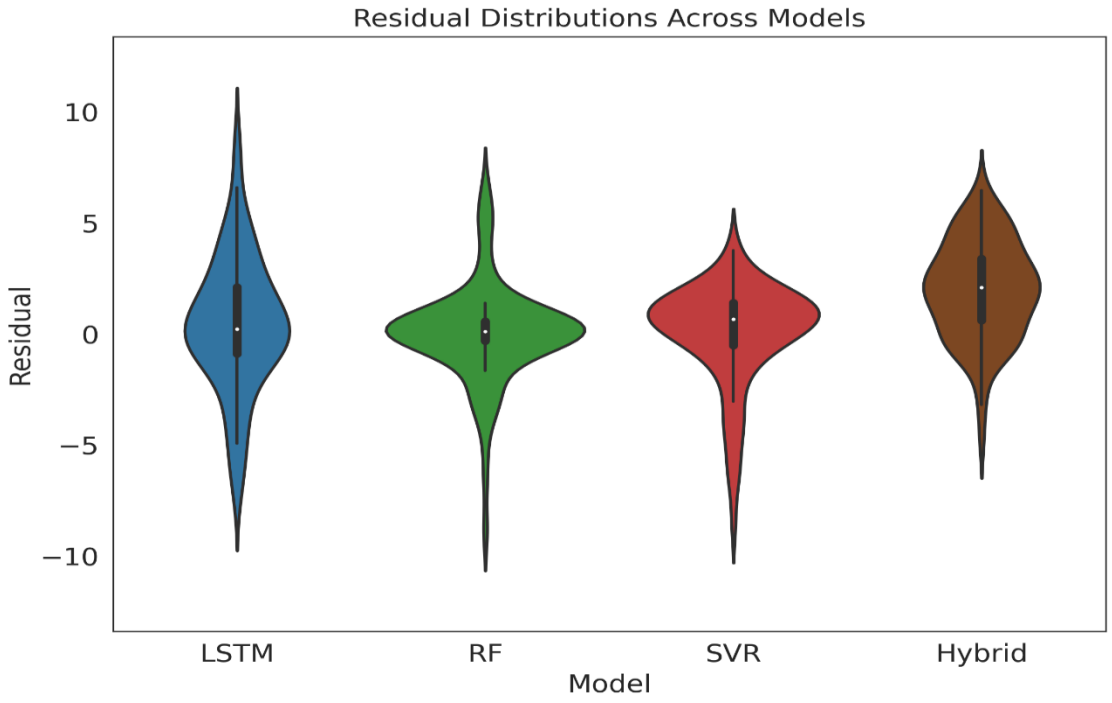


Fig. 5.8. Residual Distribution Plot

The Taylor diagram synthesis depicted in Fig. 5.9 provides a concise visual summary of model performance across three key dimensions: correlation coefficient, standard deviation ratio, and centered root-mean-square error. The hybrid model occupies the optimal position: closest to the reference "observed" point on the x-axis (standard deviation ratio ≈ 1.0), on the highest-correlation arc (r > 0.95), and with minimal distance from the reference point (indicating low error). The standalone LSTM shows good correlation but excessive variability. Its standard deviation ratio is >1.0, indicating a tendency to amplify fluctuations. Meanwhile, the random forest (RF) shows reasonable variability but reduced correlation, while SVR performs poorly across all dimensions, positioned far from the optimal region of the diagram.

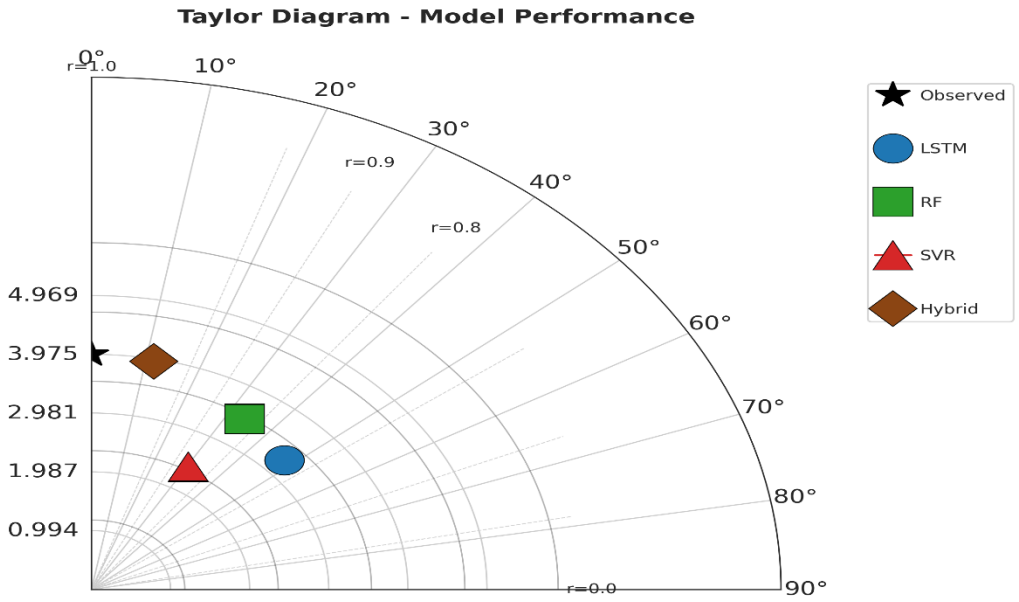


Fig. 5.9. Taylor Diagram

5.8. Discussion

The results indicate that combining multi-source data and physical knowledge of the watershed through SWAT with the temporal learning ability of LSTM significantly enhances streamflow forecasting. The improved performance of the multi-source data-integrated Hybrid SWAT-LSTM foresting framework across all correlation metrics (R^2, NSE), error metrics (RMSE, MAE), and bias indicators (PBIAS) is noteworthy and reflects the successful integration of the process-based and machine learning models. Furthermore, it provides crucial insights into how diverse data are strategically integrated to advance

hydrological forecasting skills. The integration of multi-source data benefited the hybrid forecasting framework by avoiding limitations and uncertainties associated with a single source to enhance the accuracy of forecasts. Analogous findings are reported in similar studies, such as Li et al. (2025), Wang et al. (2024), Idowu et al. (2024), and Chen et al. (2022), which claim that integrating multi-source data improves forecast skill.

The hybrid framework systemically addresses the limitations inherent in a single-paradigm approach (process-based vs. machine learning models). The process-based SWAT component served as a predictor generator or physics-informed feature, where it translated the raw meteorological input to hydrological variables that represented the watershed's hydrological conditions. Mechanistically, SWAT enforces water balance, land surface process representation, and routing constraints, which improve the model's capability to accurately reproduce flow in terms of magnitude, space, and time. These physically simulated hydrological variables can be used as a set of predictors for the LSTM models. This provides a crucial advantage for the LSTM model, which has the opportunity to learn from physically processed physical relationships instead of raw meteorological time series.

Furthermore, physics limits unrealistic behavior, and the LSTM captures non-modeled processes and nonlinearity. As the results demonstrated that integration of the SWAT output into the LSTM model enhanced the skill of stream flow forecasts, and similar improvements are reported by the studies, such as Asadi et al. (2025), Huang et al. (2024), Mie et al. (2024), Chen et al. (2023), and Yang et al. (2023). The hybrid forecasting framework showed exceptional skill in capturing high and low flow, both in magnitude, space, and time, compared to the base models (i.e., the standalone LSTM, Random Forest, and SVR), which lack explicit physical representation of watershed processes and routing physics; thus, tending to produce central tendencies well but struggle with extremes and flow frequency accuracies. For example, RF and SVR, which perform well under normal conditions, tend to smooth out temporal dynamics and underestimate rapid increases (sharp hydrograph features) and recessions. Higher KGE values and lower PBIAS values in the hybrid forecasting framework demonstrated improved correlation and reduced systematic bias.

In drier months, pure data-driven models such as LSTM, RF, and SVR occasionally produce comparable low-flow estimates because these conditions are dominated by stable baseflow processes that are easier to predict from recent data. Taylor diagram analysis reveals higher correlation coefficients and closer standard deviation ratios for the hybrid model compared to standalone machine learning models, confirming that the hybrid approach not only aligns more closely with observations but also better captures inter-annual and intra-annual variability, including extreme events. And, these are similar results to the studies such as Ren et al. (2025), Mie et al. (2024), and Yang et al. (2023). Furthermore, the hybrid framework benefited from integrating large-scale climate indices.

The integration of large-scale climate indices represents one of the main contributions of the hybrid forecasting framework, addressing a critical limitation in conventional hydrological forecasting: the disconnect between local-scale predictions and global-scale climate drivers. Climate teleconnections, statistical relationships between phenomena in geographically separated regions, exert a profound influence on hydrological variability at seasonal and interannual timescales. For example, the El Niño-Southern Oscillation (ENSO) affects precipitation patterns across many regions worldwide, whereas the Pacific Decadal Oscillation (PDO) modulates streamflow regimes over multi-year periods. Therefore, integrating these large-scale climate indices into the proposed hybrid model improved streamflow forecasting. This is consistent with the findings of Li et al. (2025), Aghelpour et al. (2022), Ahmed (2021), Sohrabi et al. (2021), and Kim et al. (2019), who discovered substantial improvement in the streamflow forecast from integrating large-scale climate indices as a predictor at monthly and seasonal scales.

By explicitly incorporating these indices as model inputs, the hybrid framework enables predictions that account for large-scale climate forcing, which would otherwise appear as unexplained variability or require extended local data records to infer indirectly. This capability is particularly valuable for monthly and seasonal forecasting and water resource planning, where the anticipation of wet or dry periods several months in advance supports more proactive management decisions.

5.9. Conclusion

In this study, we presented the proposed physics-informed hybrid modeling framework, which notably improves streamflow forecasting skills by coupling multi-source data with process-based hydrological modeling and machine learning techniques. This comprehensive approach uses the leverage of a physics-based model to simulate key hydrological variables that represent the physical processes of the watershed water balance components and their interactions. Meanwhile, the machine learning model employed to capture complex, nonstationary, and nonlinear temporal dependencies inherent in streamflow data.

The hybrid model uses the SWAT model to generate key hydrological variables such as surface runoff, baseflow, and evapotranspiration, which are added as physics-informed features to the LSTM model input in conjunction with local meteorological variables and large-scale climate indices rather than solely depending on raw meteorological variable inputs. This benefited the hybrid model to overcome the limitations from the independent use of process-based models and purely data-driven models, such as parameter equifinality and input data uncertainty, as well as limited generalization capabilities. The calibration and validation procedures, supported by sensitivity analysis, ensure that the model accurately represents the hydrological behavior of the sub-basin, and enable to incorporation of the most sensitive variables in the predictor pool enhances the prediction accuracy.

The success of the framework is attributed to three key innovations. The framework is successful mainly because of three innovations. The incorporation of multiple data sources into the model, such as remote sensing and high-resolution precipitation data combined with ground measurements, increases the model's ability to represent spatial heterogeneity and temporal variability, making it more reliable, even in data-scarce areas.

Second, using large-scale climate teleconnections as predictors fills an important gap in conventional hydrological forecasting by linking local watershed responses to global climate drivers. As a result of this integration, the model can anticipate seasonal and interannual hydrological variability caused by the El Niño-Southern Oscillation and Indian

Ocean Dipole. Forecast lead times and operational relevance are consequently improved, which is particularly helpful for long-term water resource planning.

Third, the incorporation of physics-informed hydrological knowledge through SWAT added vital constraints such as water balance enforcement and routing processes, which enhanced the realism and interpretability of streamflow predictions. The approach was more credible and transparent because it was physically grounded. It reduced the limitations of "black-box" models of pure ML. At the same time, LSTM captures complex nonlinear relations and dependencies over time, which are difficult to represent in a process-based simulation like SWAT.

This novel hybrid SWAT-LSTM modelling framework is a complementary methodology combining two modelling approaches, namely, process-based models and data-driven models. It also demonstrates the potential of exploiting various data sources to provide an efficient, physically consistent, and statistically powerful tool. Furthermore, this approach enabled us to develop the LSTM component seamlessly with spatio-temporal data simulated by SWAT under different project scenarios. Its demonstrated performance and adaptability make it a promising solution for advancing hydrological forecasting and supporting sustainable water resource management in complex and variable environments.

6. Leveraging the Power of Transfer Learning to Improve Streamflow Forecasting in a Data-Scarce River Sub-Basin

Abstract

Accurate and reliable streamflow forecasting in data-scarce sub-basins remains a persistent challenge due to a lack of sufficient historical records. This problem is especially pronounced in developing regions, particularly in complex, large river basins like the Upper Blue Nile, where some sub-basins have inadequate hydro-meteorological monitoring stations. To address these limitations and improve the skill of forecasting, this study investigated the efficacy of transfer learning, which leverages knowledge from data-rich sub-basins (sources) to enhance predictions in data-scarce sub-basins (targets). Despite the success of transfer learning across basins, including transboundary basins in many studies, its effectiveness at the sub-basin level has not yet been investigated. Thus, in this study, we applied transfer learning in two stages. First, hydrologically analogous sub-basins (Guder and Muger) were identified using K-means clustering based on physical and climatic indices (e.g., aridity, flashiness, and seasonality). Second, process-based knowledge, trained on the data-rich source domain (Guder sub-basin), was transferred via fine-tuning to the data-scarce target domain (Muger sub-basin) using a Long Short-Term Memory (LSTM) network architecture to perform the forecasting task. The results demonstrate high predictive accuracy, with the transfer-learned model achieving a Nash-Sutcliffe Efficiency (NSE) of 0.77, a Kling-Gupta Efficiency (KGE) of 0.702, and an R^2 of 0.82. The model exhibited robust generalization, effectively capturing both baseflow and peak events while successfully integrating physics-based model outputs and large-scale climate indices such as ENSO and AMO. These findings suggest that transfer learning provides a viable and efficient "regionalization" mechanism, enabling reliable streamflow predictions in data-scarce sub-basins to support water resource management and disaster mitigation.

Key words: Transfer Learning, Data-Scarce Sub-basins, Guder Sub-basin, MUGER Sub-basin

6.1. Introduction

Streamflow, the lifeblood of agricultural, industrial, and ecological systems, is of paramount importance for the socioeconomic well-being of nations, particularly in regions heavily reliant on water resources (Thieu et al., 2024; Arsenault et al., 2023; Ullah et al., 2023). However, one of the most persistent and challenging problems in hydrology is the accurate prediction of streamflow in data-scarce basins with limited or no ground-based hydrological observation data. These data-scarce basins, particularly in developing regions, suffer from inadequate hydrometeorological monitoring networks because of logistical, financial, and geopolitical constraints (Feng et al., 2021; Liu et al., 2018). Consequently, traditional hydrological forecasting models, which require substantial data for calibration and validation, often perform poorly or are inapplicable in ungauged contexts.

The concept of "Prediction in Ungauged Basins (PUB)," as promoted by the International Association of Hydrological Sciences (IAHS), has driven substantial research efforts aimed at overcoming the data limitations in such areas (Kuana et al., 2024; Willard et al., 2023; Kratzert et al., 2019). Conventional approaches, which have been widely applied, include regionalization techniques, in which model parameters from gauged basins are transferred to ungauged basins based on hydrological similarities. However, these methods often suffer from significant uncertainties, particularly when donor and target basins differ in climatic or physiographic characteristics (Meng et al., 2020). Additionally, Loukika et al. (2022) and Panthi et al. (2021) noted that the spatial heterogeneity of rainfall, land use, soil types, and human activities can drastically alter streamflow regimes, rendering simplistic regionalization approaches insufficient for accurate predictions.

Recently, deep and machine learning models have gained attention in hydrology because of their ability to model nonlinear relationships between inputs (e.g., precipitation and temperature) and outputs (streamflow) without relying on explicit physical process representations. However, the efficacy of these models depends on the availability of large, high-quality training datasets, including meteorological records and streamflow measurements (Saravanan et al., 2023; Vogeti et al., 2022; Konapala et al., 2020). This requirement poses a significant challenge in data-scarce river basins (Feng et al., 2021; Jia et al., 2020) and necessitates the exploration of alternative methodologies that can leverage knowledge from data-rich regions or analogous hydrological systems to improve

predictions in data-scarce environments (Oruche et al., 2021; Pool and Seibert, 2021; Panthi et al., 2021).

Transfer learning has emerged as a potent paradigm to address data paucity by adapting pre-trained models from data-abundant source domains to target basins with limited data, improving predictive performance without extensive local datasets (Han et al., 2021; Oruche et al., 2021; Ma et al., 2020). This approach involves training a base model using data-rich catchments, then fine-tuning or applying domain adaptation to perform well in data-poor regions, leveraging structural similarities in physiographic, climatic, and hydrological characteristics among basins (Ma et al., 2024; Xu et al., 2023). Studies demonstrate that transfer learning-based models outperform regionalization and standalone ML models in ungauged settings, particularly when combined with spatially distributed inputs such as satellite-derived precipitation, land cover, topography, and soil moisture (Zhu & Zhou, 2024; Arsenault et al., 2023; Oruche et al., 2021; Ma et al., 2020; Kratzert et al., 2019a). Additionally, TL techniques—including domain adaptation, feature representation learning, and few-shot learning—offer nuanced ways to customize models to target basins with minimal additional data.

Despite the growing recognition of the promise of transfer learning (TL) in hydrological forecasting, critical research gaps remain. For instance, transfer learning has yet to be systematically benchmarked across diverse physiographic regions and hydrological regimes, particularly in data-scarce sub-basin scales (Yao et al., 2023; Min et al., 2023; Ma et al., 2020). This spatial data disparity poses a significant challenge for streamflow forecasting and water management, underscoring the absence of robust methods that can leverage data-rich areas to improve predictions by sharing data with neighboring poorly gauged or ungauged regions (Zhang et al., 2025; Nearing et al., 2024; Zhang et al., 2022). Furthermore, although hybrid forecasting models have shown promising capabilities for streamflow forecasting, their efficacy is constrained by the intensive data requirements to train machine learning components from scratch *(Ougahi* and Rowan, 2024; Pathak et al., 2018). However, transfer learning directly addresses this limitation by leveraging knowledge from pretrained models, thereby avoiding the need for extensive, costly, and computationally prohibitive data collection for each sub-basin.

Prior transfer learning studies (Kratzert et al., 2019; Ma et al., 2020; Arsenault et al., 2023; Oruche et al., 2021) operate at basin or continental scales, employ purely data-driven approaches without physics-informed constraints, select source basins arbitrarily or by proximity, and exclude climate teleconnection indices. In contrast, this study provides the first sub-basin scale TL investigation, integrates SWAT outputs for physics-informed learning, implements systematic K-means clustering using six validated similarity indices (BFI, Flow CV, Aridity, Seasonality, Flashiness, PET), and becomes the first to include ENSO, AMO, and TNA indices as dynamic predictors. No existing study combines all four innovations.

Finally, existing studies rarely integrate physics-informed constraints within the TL framework, which could help improve the robustness and physical plausibility of forecasts. Therefore, this study aimed to investigate the efficacy of transfer learning for streamflow forecasting in the Upper Blue Nile Basin, where some sub-basins possess long-term hydrometeorological records (Samy et al., 2019; Abera et al., 2017). In contrast, other sub-basins are highly data-scarce because of historical underinvestment, logistical challenges, and high operational costs associated with streamflow forecasting and maintaining monitoring stations (Khadim et al., 2020; Lakew et al., 2020). Subsequently, this study proposes a novel multi-data source and a physics-informed transfer learning approach for streamflow forecasting, specifically tailored for the Upper Blue Nile Basin.

We aimed to establish the Guder sub-basin as a comparatively data-rich source from which hydrological knowledge (model parameters) derived from both multi-source data inputs and an integrated hybrid modelling approach can be transferred to a Muger sub-basin with similar physiographic and hydrologic characteristics. The main contribution of this study is the systematic application of transfer learning to adapt pre-trained, multi-source data-informed, and hybrid-architected models for effective streamflow forecasting in data-scarce yet hydrologically similar sub-basins within the Upper Blue Nile Basin, which can aid in making informed decisions and preparing for natural disasters.

6.2. Transfer learning framework

The main goal of this research was to implement and evaluate the efficacy of transfer learning for streamflow forecasting in data-scarce river sub-basins. It leverages generalized

hydrological knowledge from sub-basins of the Upper Blue Nile Basin that are relatively data-rich in hydro-climatic information (source task) to improve predictive accuracy and model robustness for a specific data-scarce sub-basin (target task) through a transfer learning approach.

The source task dataset was a Guder sub-basin of the Upper Blue Nile Basin. This basin was chosen because it has relatively high-quality long-term hydrological and meteorological data records, providing a rich source of generalized hydrological patterns and some key hydrological variables generated by the SWAT model from the previous chapter. Furthermore, it enables transfer learning within a physics-informed framework, adding the physical knowledge of the underlying hydrological process.

The target data-scarce sub-basin has been selected from the sub-basins of the Upper Blue Nile Basin, which are geomorphologically and hydrologically similar to the Guder sub-basin. For the selection of the target sub-basin, several similarity indices were employed, which measure hydroclimatic, physiographic, land use, and hydrological response characteristics of the sub-basins. These include:

- **Hydrological response metrics such as** Baseflow Index (BFI), Flow Coefficient of Variation (CV), Seasonality Index, and Flashiness Index (Richards–Baker Index).
- **Hydro-climatic parameters including:** Mean Annual Precipitation (MAP), Potential Evapotranspiration (PET), and Aridity Index (MAP/PET).
- **Topographic and physiographic characteristics such as** elevation, catchment area, and slope-related measurements.

The sub-basins were then grouped into similar clusters using the K-means algorithm. This method has been widely applied and has proven effective in hydrological studies, and it was selected for its computational speed and ability to handle multidimensional data (Tigkas et al. 2015). The clustering process identifies similar sub-basins to provide an organized basis for hydrological modeling and makes it easier to transfer knowledge from data-rich to data-scarce sub-basins. For the clustering process, we gathered all relevant datasets, including streamflow, climate, land use and land cover, physiographic basin

features, rainfall, and soil data from the Abay Basin Authority, the Ministry of Water and Energy (MoWE), the National Meteorological Service Agency (NMSA), and global remote sensing datasets. The next section provides a detailed description of the similarity index used for clustering.

6.2.1. Similarity Index

(1) Median Baseflow Index (BFI)

The Baseflow Index measures the contribution of groundwater (baseflow) to the total streamflow, indicating the catchment storage and regulation capacity. It is calculated as:

$$\mathrm{BFI} = \frac{\mathrm{Q_{base}}}{\mathrm{Q_{total}}} \qquad \text{Eq. 6.1}$$

Where Q_{base} is the baseflow volume derived from hydrograph separation, and Q_{total} is the total streamflow volume over the same period. The median BFI across years was used to minimize the interannual variability.

(2) Flow Coefficient of Variation (Flow CV)

This index captures flow variability and hydrological stability.

$$\mathrm{FlowCV} = \frac{\sigma \mathrm{Q}}{\mathrm{mQ}} \qquad \text{Eq.6.2}$$

In this context, σQ represents the standard deviation of daily discharge, while mQ denotes the mean discharge. Higher values indicate greater variability and more flashy flow regimes.

(3) Aridity Index (AI)

The aridity index describes the balance between water supply and atmospheric demand, defined as:

$$\mathrm{AI} = \frac{\mathrm{P}}{\mathrm{PET}} \qquad \text{Eq. 6.3}$$

where P is the mean annual precipitation, and PET is potential evapotranspiration. Values < 1 indicate arid to semi-arid conditions, while values > 1 indicate humid climates.

(4) Seasonality Index (SI)

The seasonality index reflects the degree of uneven distribution of precipitation across months (Walsh & Lawler, 1981):

$$SI = \left(\frac{1}{P}\sum i = 1 \text{ to } 12\right)\left|P_i - \left(\frac{P}{12}\right)\right| \qquad \text{Eq. 6.4}$$

Pi is the monthly precipitation in month i, and P is the mean annual precipitation. Values close to 0 indicate uniform rainfall distribution, while higher values reflect strong seasonality.

(5) Flashiness Index (FI)

The Richards-Baker Flashiness Index quantifies rapid fluctuations in daily flow:

$$FI = \frac{\left(\sum(t=2\ ton)\ |Q_t - Q_{t-1}|\right)}{\left(\sum(t=1\ ton)\ Q_t\right)} \qquad \text{Eq. 6.5}$$

Where Qt is the daily streamflow on day t. Higher values denote highly variable and responsive systems.

(6) Potential Evapotranspiration (PET)

Potential evapotranspiration was estimated using the Hargreaves–Samani equation (Hargreaves & Samani, 1985).

$$PET = 0.0023 * (T_{mean} + 17.8) * (T_{max} - T_{min})^{0.5} * R_a \qquad \text{Eq. 6.6}$$

Where T_{mean} is mean daily air temperature (°C), T_{max} and T_{min} are daily maximum and minimum temperatures (°C), and Ra is extraterrestrial radiation (MJ m^-2 day^-1).

6.2.2. K-Means Algorithm

K-Means clustering partitions n sub-basins into k groups by minimizing within-cluster variance. The optimization problem is expressed as:

$$J = \sum(i = 1\ \ to\ k \sum(x_j \in C_i)\ \| x_i - \mu_i \|^2 \qquad \text{Eq. 6.7}$$

Where, Ci is the i-th cluster, x_j is the feature vector representing sub-basin j, and μ_i is the centroid of cluster i.

Since the hydrologically meaningful number of sub-basin groups is not known a priori, candidate values of k = 2–6 were tested. Cluster validity was assessed using the silhouette

score, which quantifies cohesion and separation. For a sub-basin i, the silhouette score is defined as:

$$S(i) = \frac{(b(i) - a(i))}{max\{a(i), b(i)\}} \qquad \text{Eq. 6.8}$$

where a(i) represents the mean intra-cluster distance of i, and b(i) denotes the mean nearest-cluster distance. A silhouette scores closer to 1 indicates well-separated and compact clusters, while values near 0 suggest overlapping clusters. The number of clusters yielding the highest average silhouette score was selected as the optimal k. Final clustering was performed using the K-Means algorithm with k equal to the silhouette-optimal solution (for this case, k = 4). The cluster groups are:

- **Cluster 0 (Humid High Rainfall Runoff-Dominated Basins):** represents large western and southwestern tributary basins including Didessa, Dabus, Beles, and Anger, characterized by high rainfall (MAP ≈ 1,600–1,950 mm), relatively low to moderate baseflow contribution (BFI < ~0.50), and moderate to high variability, indicating dominant surface runoff processes, high flow volumes, and strong hydrological response to rainfall.
- **Cluster 1 (Humid Baseflow-Dominated Highland Basins):** represents highland and Lake Tana–influenced basins, including Finchaa, Gilgel Abay, Gumara, Rib, Megech, and Koga, characterized by high rainfall (MAP ≈ 1,500–1,750 mm), high baseflow contribution (BFI > ~0.50), and low to moderate flashiness (RBI < ~0.30), indicating strong groundwater influence, sustained dry-season flow, and relatively stable hydrological regimes.
- **Cluster 2** (Moderate Rainfall Flashy Transitional Basins): represents central and eastern agricultural basins, including Guder, Muger, Jemma, Jedeb, Beshilo, and Temcha, characterized by moderate rainfall (MAP ≈ 1,200–1,400 mm), low to moderate baseflow contribution (BFI ≈ 0.35–0.45), and high flashiness (RBI > ~0.33), reflecting mixed but runoff-dominated processes, rapid response to rainfall, and high flow variability.
- **Cluster 3 (Dry Highly Variable Basin):** represents water-limited eastern basin Wabi, characterized by low rainfall (MAP < ~1,100 mm), low baseflow contribution (BFI < ~0.40), and high flashiness (RBI > ~0.35), indicating weak

groundwater influence, intermittent flow conditions, and highly variable hydrological behavior.

Using the similarity indices, the sub-basins of the Upper Blue Nile Basin were grouped into clusters, as shown in Table 6.1. Based on this clustering, the Muger sub-basin was selected for further analysis because it lies in the same cluster as the Guder sub-basin (the source sub-basin) and belongs to the "Moderate Rainfall Flashy Transitional Basins" cluster. The similarity in hydrological, physiographic, and climatological characteristics between these sub-basins makes them suitable for knowledge transfer from the data-rich to the data-scarce sub-basin and they are located as shown in Fig. 6.1.

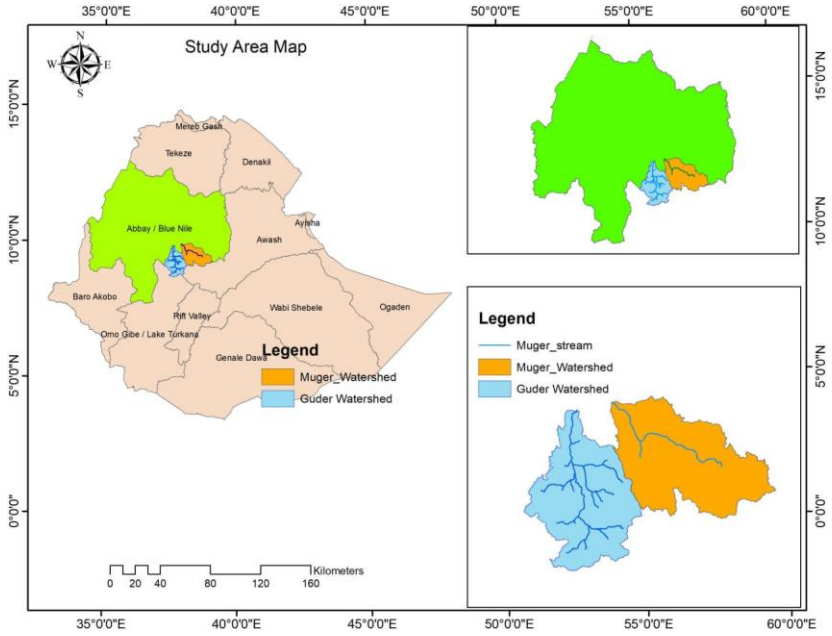


Fig 6.1. Map of Source and Target Sub-Basins

Table 6.1. Upper Blue Nile Basin with Sub-Basin Cluster

Sub-basin	Area (km2)	MAP (mm)	Mean Flow (m3/s)	Elevation (m)	PET (mm)	Median BFI	Flow CV	Seasonality Index	Flashiness RBI	Aridity Index	Cluster
Didessa	19630	1950	229	1750	1400	0.52	1.1	1.1	0.2	1.3929	0
Finchaa	2963	1700	123	2000	1350	0.48	1.25	1.2	0.26	1.2593	1
Dabus	21036	1950	420	1550	1450	0.5	1.15	1.15	0.22	1.3448	0
Beles	14200	1600	186	1600	1500	0.42	1.4	1.35	0.32	1.0667	0
Guder	7011	1300	102	2200	1350	0.4	1.5	1.4	0.34	0.963	2
Muger	7948	1250	89	2400	1300	0.38	1.55	1.45	0.36	0.9615	2
Jemma	4890	1350	54	2150	1350	0.44	1.45	1.35	0.3	1	2
Jedeb	2000	1350	31	2350	1300	0.36	1.6	1.45	0.38	1.0385	2
Koga	297	1400	21	2200	1250	0.48	1.35	1.3	0.28	1.12	1
Gilgel Abay	5088	1650	151	2150	1200	0.58	1.25	1.2	0.22	1.375	1
Ribb	1509	1500	84	2250	1250	0.45	1.35	1.3	0.28	1.2	1
Gumara	1838	1500	80	2200	1250	0.55	1.3	1.25	0.24	1.2	1
Megech	520	1400	41	2050	1300	0.4	1.4	1.35	0.3	1.0769	1
Beshilo	13370	1250	113	2250	1400	0.35	1.7	1.5	0.4	0.8929	2
Wabi	1400	950	27	950	1615	0.32	1.55	1.55	0.4	0.5882	3
Temcha	1450	1300	27	2200	1300	0.38	1.55	1.45	0.34	1	2
Anger	7254	1850	160	1750	1400	0.5	1.2	1.15	0.24	1.3214	0

6.3. Transfer Learning approach

In this study, for transfer learning, we choose a deep learning Long Short-Term Memory (LSTM) architecture, which is known for its ability to handle time-series data. The description of the LSTM model was explained in the previous section.

Source Domain (D_S):

$$D_S = \{X_S, Y_S\}, \quad X_S \in \mathbb{R}^{n_s \times d}, \quad Y_S \in= \mathbb{R}^{n_s} \qquad \text{Eq. 6.9}$$

Target Domain (D_T):

$$D_T = \{X_T, Y_T\}, X_T \in \mathbb{R}^{n_t \times d}, \quad Y_T \in= \mathbb{R}^{n_t} \qquad \text{Eq. 6.10}$$

Where X_S and X_T represent hydrological predictors (precipitation, minimum and maximum temperature, large-scale climate indices, SWAT outputs, etc.), Y_S and Y_T are observed streamflow values, and typically ns and nt; nt < ns, i.e., fewer labelled samples exist in the target basin.

I. The LSTM model is first trained on the source domain:

$$\theta_S = arg\min_{\theta} \mathcal{L}(f(X_S; \theta), Y_S) \qquad \text{Eq. 6.11}$$

Where θ are the trainable parameters (weights, biases), and $\mathcal{L}$ is the loss function (e.g., Mean Squared Error).

Loss Function ($\mathcal{L}$): The model is trained to minimize the difference between its predicted streamflow ($\hat{Q}$_source, t) and the actual observed streamflow (Q_source, t). We use the Mean Squared Error (MSE) as our loss function, which is defined as:

$$\mathcal{L}_{MSE} = \frac{1}{N_{source}} \sum (\hat{Q}_{source,t} - Q_{source,t})^2 \qquad \text{Eq.6.12}$$

Where, N_{source} is the number of data points in the source dataset.

In transfer learning, part of the learnt parameters θ_S are transferred to initialize the target model. Then, fine-tuning is done using the limited target dataset:

$$\theta_T = arg\min_{\theta'} \mathcal{L}(f(X_T; \theta'), Y_T), \theta' \subseteq \theta_S \qquad \text{Eq. 6.13}$$

The loss function for this stage is also MSE, but it's calculated on the target data:

$$\mathcal{L}_{MSE\ fine-tune} = \frac{1}{N_{target}} \sum (\hat{Q}_{target,\ t} - Q_{target,t})^2 \qquad \text{Eq. 6.14}$$

Where Ntarget is the number of data points in the target dataset. This process adjusts the model's parameters to minimize the error on the target-specific data while preserving the knowledge gained during pre-training.

This process ensures that knowledge (temporal dynamics, rainfall-runoff relationships) from the source basin benefits the target basin prediction.

II. Transfer Learning Loss Function

To regularize transfer learning, a joint loss can be defined:

$$\mathcal{L}_{TL} = \mathcal{L}(f(X_T; \theta_T), Y_T) + \lambda \parallel \theta_T - \theta_S \parallel^2 + \lambda_1 \cdot ReLU(-Q_{pre}d) + \lambda_2 \cdot ReLU(Q_{pre}d - (P - ET)) \qquad \text{Eq. 6.15}$$

- ✓ $\mathcal{L}_{TL}$ is the total physics-informed transfer learning loss to be minimized.
- ✓ The first term $\mathcal{L}$ (f (X_t; θ_t), Y_t) is the mean squared error (MSE) between predicted and observed streamflow on the target basin. X_t are input features (precipitation, temperature, climate indices, SWAT outputs, etc.), Y_t is the observed streamflow, and θ_t are the target model parameters.
- ✓ The second term $\lambda||\theta_t - \theta_s||^2$ is the transfer regularization penalty. It penalizes large differences between the target parameters θ_t and the source parameters θ_s (pre-trained on the Guder sub-basin). λ is a coefficient (e.g., 0.001) determined by hyperparameter tuning.
- ✓ The third term $\lambda_1 \cdot$ReLU(-Q_{pre}d) is a non-negativity penalty. It applies a small cost whenever the predicted streamflow Q_{pre}d is negative (physically impossible). ReLU(x) = max (0, x). $\lambda_1 = 0.01$.
- ✓ The fourth term $\lambda_2 \cdot$ReLU(Q_{pre}d - (P -ET)) is an upper-bound penalty. It discourages Q_{pre}d from exceeding the daily water surplus (precipitation P minus actual evapotranspiration ET). This is a necessary condition for mass conservation when storage change is neglected. $\lambda_2 = 0.001$.

- ✓ All penalties are applied during both pre-training (source basin) and fine-tuning (target basin). The ReLU function is differentiable almost everywhere and works with gradient-based optimization.
- ✓ λ is a regularization parameter controlling how strongly the target parameters deviate from the source parameters

To embed physical knowledge into the transfer learning framework, we augment the loss function with two inequality penalty terms. The first term, λ1*ReLU (−Qpred), penalizes negative streamflow predictions, ensuring physical non-negativity. The second term, λ2*ReLU (Qpred−(P−ET)), discourages the model from predicting streamflow that exceeds the daily water surplus (precipitation minus actual evapotranspiration) – a necessary condition for mass conservation when storage change is neglected. The coefficients λ1=0.01λ1=0.01 and λ2=0.001λ2=0.001 were selected via a grid search over the validation set of the target basin, balancing physical constraint satisfaction against predictive accuracy. These penalties are applied during both pre-training on the source basin and fine-tuning on the target basin."

III. Evaluation and Forecasting

Once the model is fine-tuned, its performance is rigorously evaluated on a held-out test set from the target catchment. Statistical metrics: Mean Absolute Error, Root Mean Square Error, Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), and R^2 are used to measure the performance of the final fine-tuned model. Details about these metrics are presented in chapter three table 3.1.

6.4. Result and Discussion

To examine the effectiveness of the physics-informed transfer learning approach in improving streamflow forecasting in data-scarce river basins, we developed a long short-term memory (LSTM) network-based transfer learning model in the Upper Blue Nile sub-basins. Initially, our model was applied and trained on the source basin (Guder sub-basin), which is relatively data-rich, to learn the pattern of interactions and responses between hydroclimatic factors and streamflow process before being applied to the target basin (Muger sub-basin). The initial-layer weights from the LSTM model trained on the Guder sub-basin, which represent the fundamental temporal dynamics of the streamflow process,

were transferred to a new LSTM model. The first layer was then frozen to preserve the transferred knowledge, and thereafter, the model was fine-tuned on limited observed data from the Muger sub-Basin (i.e., target basin). The performance of the transfer learning model was subsequently assessed both quantitatively and through visualization tools, as reported below.

6.4.1. Quantitative Performance Metrics

The study conducted a thorough performance assessment of the transfer learning (TL) approach using standard statistical skill metrics. We compared the model outputs against observed streamflow data from the target basin, splitting the data into training (70%) and validation (30%) sets.

Table 6.2. Comprehensive Performance Metrics for Training and Validation Phases

Dataset	*MAE*	*RMSE*	*R^2*	*NSE*	*KGE*
Training	6.23	8.13	0.87	0.82	0.81
Validation	9.40	10.48	0.82	0.77	0.702

The study conducted a thorough performance assessment of the transfer learning (TL) approach using standard statistical skill metrics. We compared the model outputs against observed streamflow data from the target basin, splitting the data into training (70%) and validation (30%) sets. The model demonstrated strong skill scores, as shown in Table 6.2 and the performance comparison (Fig. 6.2). The validation Nash-Sutcliffe Efficiency (NSE) of 0.77 and Kling-Gupta Efficiency (KGE) of 0.702 classify the model performance as "good" according to established hydrological benchmarks (Ma et al., 2014; Moriasi et al., 2007). Importantly, the validation R^2 of 0.82 indicates that the model explains 82% of the variance in the observed streamflow, a notable achievement given the scarcity of local data. A slight decrease in all metrics from the training to the validation phase (for instance, KGE decreased from 0.81 to 0.702) is expected and minimal, indicating that the model generalizes well without severe overfitting, a common problem when training complex models on small datasets. Similar results are reported in Ouyang et al. (2025); Qin et al. (2024), and Shen et al. (2024).

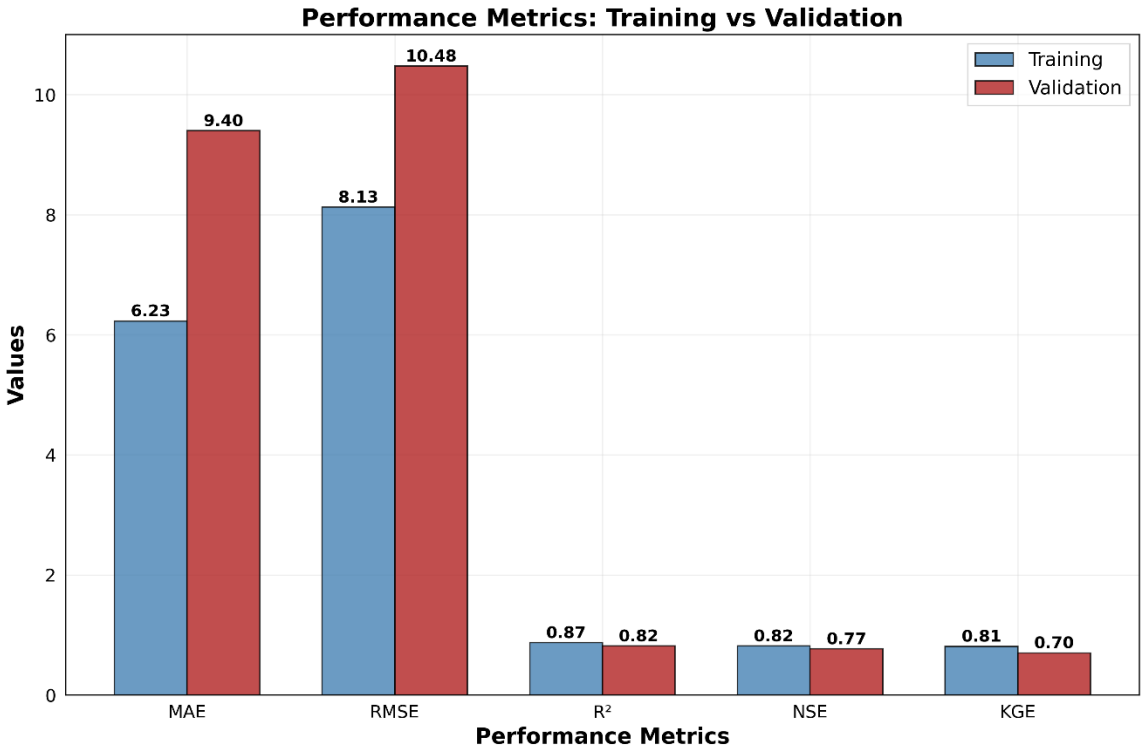


Fig. 6.2. Performance Metrics Summary: Training vs. Validation

The bar chart further compares error metrics (MAE and RMSE) in the training and validation phases. In both phases, the transfer learning model performs well, with a controlled degree of generalization error despite a modest increase in error during the validation phase. The results are consistent with Funato & Sawada (2025); Raza et al. (2025), Kratzert et al. (2019), and Zhang et al. (2022). This minimal difference between the training and validation errors shows that the model captures the underlying patterns of the data, performs well on unseen data, and is not affected by overfitting or noise in the data.

6.4.2. Predictive Accuracy and Visual Diagnostics

The visual presentation tools further explain the forecast skill of the transfer learning model. The scatter plots (Fig.6.3) of forecasted versus observed streamflow display a strong linear agreement along the 1:1 line for both training and validation datasets. The validation plot shows a similar positive relationship to the training plot, though the points are slightly more dispersed than in the training plot. This graphic verifies the high R^2 values (0.82 and 0.87

during the training and validation periods, respectively), which confirm the model's ability to capture both low and high flow events in the target basin.

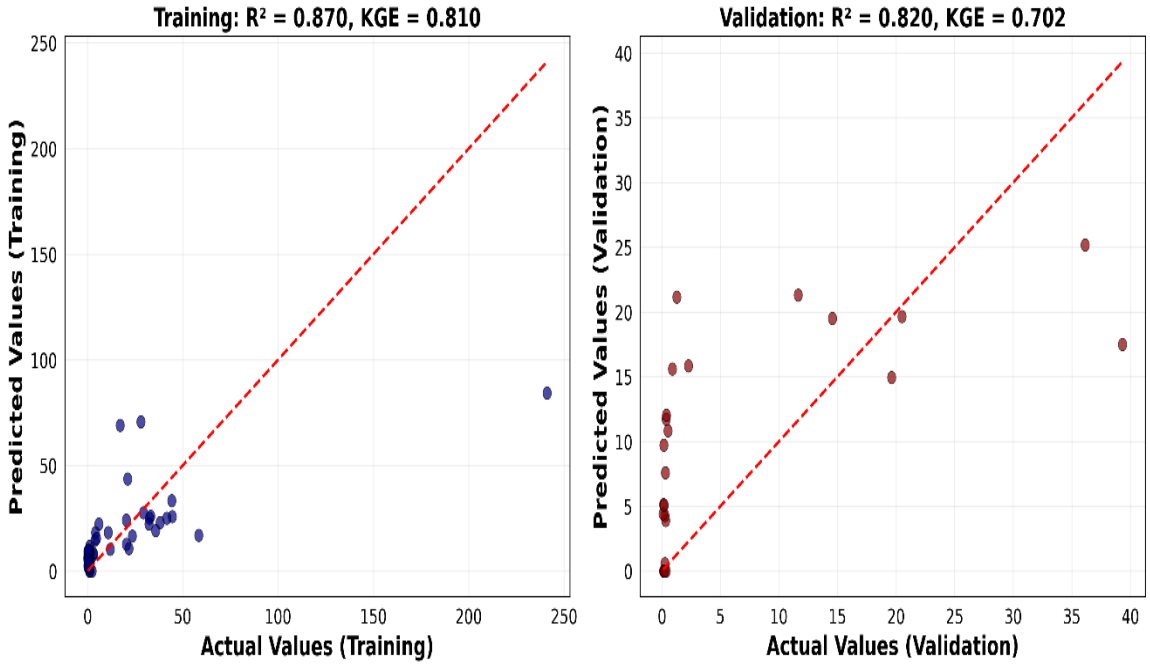


Fig. 6.3. Scatter Plots of Predicted vs. Actual Streamflow

Similarly, Fig. 6.4 shows the time series plot from the transfer learning model. The model showed proficient skill in capturing the rising and falling peaks of the observed data hydrograph properly during both training and validation phases. During the validation period, the model successfully captured the important peak flows (even though the largest peak is under-predicted). This underestimation is a common issue in hydrological modelling, yet the model exhibits no systematic bias or temporal drift. This is also consistent with studies such as Xu et al. (2025), Ueda et al. (2024), Oddo et al. (2024), and Vogeti et al. (2022).

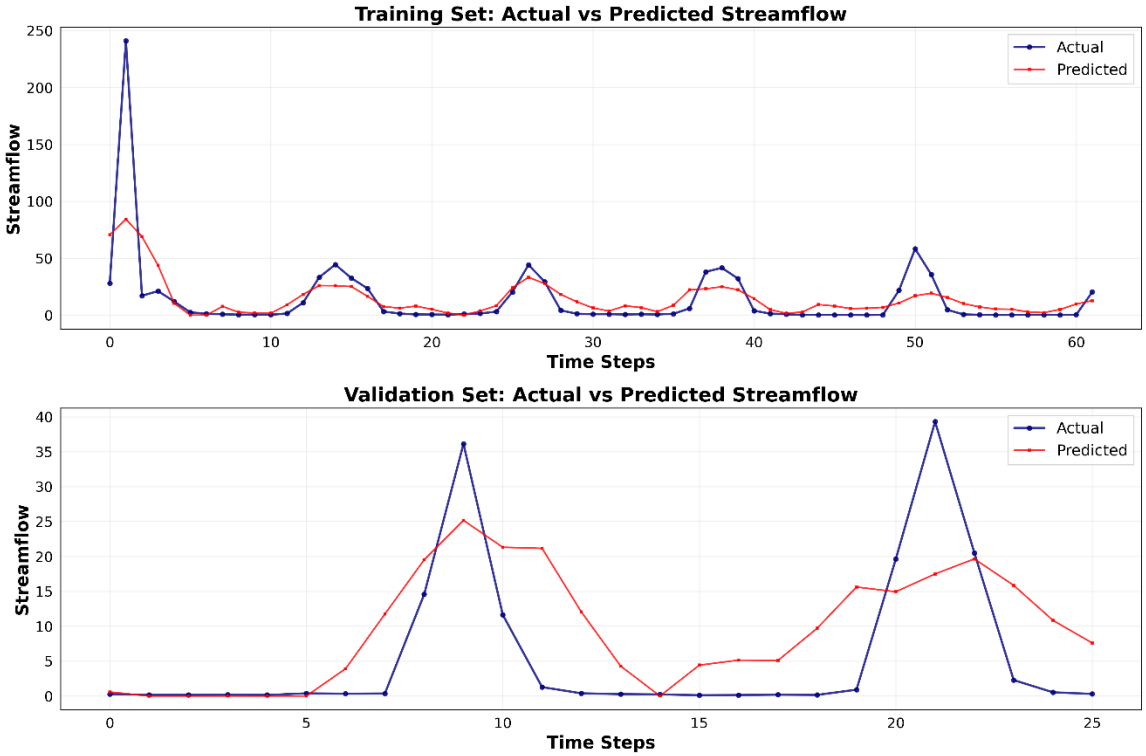


Fig. 6.4. Time Series of Actual vs. Predicted Streamflow.

6.5. Conclusion

The findings of this study showed that physics-informed transfer learning (TL) has great potential to improve streamflow forecasting in data-scarce river basins by effectively leveraging hydrological knowledge from data-rich and hydrologically similar sub-basins. The application of a TL framework based on Long Short-Term Memory (LSTM) for the Upper Blue Nile Basin, utilizing hydrological knowledge transfer from the Guder sub-basin to the Muger sub-basin, proves robust predictive performance despite limited local data availability. This approach overcomes the challenges arising from the need for large amounts of data to train machine learning models and addresses overfitting problems in ungauged or poorly gauged sub-basins by transferring process-based temporal dynamics and rainfall-runoff relationships. Furthermore, the study verified the importance of clustering sub-basins based on physiographic, hydroclimatic, and hydrological response characteristics to provide a systematic basis for selecting appropriate source-target pairs for effective transfer learning.

Unlike conventional regionalization techniques, the success of the transfer learning approaches hinges not on transferring patterns from a source dataset but rather on transferring generalized hydrological knowledge. The pre-trained LSTM model effectively learned how local and large-scale climatic drivers, including major large-scale climate indices like ENSO, AMO, and TNA, interacted with basin characteristics to generate streamflow from the source basins. This learned representation in the Guder sub-basin (source), encoded within the network's weights, proved to be substantially transferable to a hydrologically similar target basin. Furthermore, the fine-tuning process served as a tool to adapt these generalized dynamics for the nuances of the Muger catchment as a form of data-driven regionalization techniques. Meanwhile, the slight gap between training and validation errors revealed that the model generalizes well and is not exposed to overfitting on the limited target domain data. This proves the effectiveness and robustness of the transfer learning framework for reliable forecasting in poorly gauged basins.

Moreover, the integration of a physics-informed, multi-source data framework in the TL architecture substantially enhances model robustness and physical plausibility, countering the limitations of pure data-driven approaches. This study highlights the potential of TL as an effective and scalable approach to the prevalent “Prediction in Ungauged Basins” (PUB) challenge. It provides a computationally efficient solution for regional hydrological prediction with significant relevance for water resource management and disaster preparedness in data-scarce regions.

7. Conclusions and Recommendations

7.1. Discussion

This study pursued four integrated objectives to address hydrological prediction challenges in the data-scarce Upper Blue Nile basin.

Improving precipitation estimates: Individual satellite and reanalysis products performed poorly due to complex terrain, scale mismatches with gauge data, and limited station coverage. A machine learning merging approach substantially improved all statistical metrics (e.g., KGE, POD, FAR). This confirms that blending multiple data sources enhances precipitation representation, even with only 36 training stations. Performance could improve further with denser gauge networks.

Enhancing monthly precipitation forecasting: This study proposes a framework to forecast monthly precipitation using an integrative approach that combines multi-source data with machine learning models, namely, ANN, SVM, CNN, RF, and XGBoost. The model was able to achieve good accuracy for a one-month-ahead forecast. Overall, ANN, SVM, and KNN performed best, though all models had difficulty in wet months, showing challenges of extreme prediction. The framework generalized well to unseen data and can be valuable for practical forecasting applications, though results are basin-specific and require broader testing.

Improving monthly streamflow forecasting: A physics-informed hybrid SWAT-LSTM framework outperformed standalone models (LSTM, RF, SVR) for monthly stream flow forecasting by leveraging complementary strengths. The SWAT model served as a generator of physics-informed features. It enforced both the water balance constraint and the routing constraint, while LSTM captured residual nonlinearities and unmodeled processes and made the prediction. Integrating large-scale climate indices (ENSO, PDO, AMO, TNA) further enhanced skill by linking local flows to global drivers. The hybrid approach excelled at capturing extremes and reducing systematic bias—a clear advantage over pure data-driven or pure process-based models alone.

Examining transfer learning efficacy: a physics-informed LSTM-based transfer learning framework effectively transfers hydrological knowledge from a data-rich source sub-basin (Guder) to a data-scarce target (Muger). Pre-training on source data and fine-tuning on

limited local data led to robust stream flow predictions without overfitting. This indicates that generalized rainfall-runoff relationships are transferable across hydrologically similar basins. Thus, it offers a potential scalable solution for the long-standing Prediction in Ungauged Basins (PUB) problem.

In General, the four objectives form a coherent, progressive strategy for data-sparse regions: merge disparate precipitation products, integrate multi-source data with machine learning for monthly forecasts, embed physical laws via a hybrid SWAT-LSTM framework with climate indices, and transfer hydrological knowledge from data-rich to data-poor sub-basins. This sequence offers a practical and scalable roadmap for improving hydrological forecasting in the Upper Blue Nile and beyond.

7.2. Conclusions

This PhD dissertation research attempted to present a comprehensive development of an integrated hydrological forecasting system, tailored to the Upper Blue Nile Basin, addressing the critical challenges posed by data scarcity, dynamic and complex hydroclimatic variability, nonlinearity, models' limited predictive capability, generalizability, and interpretability. The study achieved its overarching objectives through the utilization of multi-source data fusion, advanced machine learning techniques, and process-based hydrological modeling to substantially advance the state of hydrological forecasting in these data-scarce and diverse hydroclimatic regions. The finding shows that the combination of heterogeneous data from different sources and the coupling of models provides substantial enhancement in forecasting accuracy and robustness of the hydrological forecasting systems, thereby offering both methodological improvements and an operational solution for water resources planning and management.

A random forest-based data fusion method, which blends precipitation datasets from gauges, satellite observations, and reanalysis products considerably improved the spatio-temporal resolution and accuracy of rain estimates. This advancement provides a valuable solution for basins, such as the Upper Blue Nile Basin, characterized by complex topography and inadequate or scarce conventional gauge stations. The merged precipitation product exhibited better correlation with independent gauge data, less misclassification of rainfall events, and more realistically captured the spatial variability of

precipitation across the basin. It provides reliable and robust input for subsequent hydrological modeling and forecasting efforts. *Nevertheless, it should be acknowledged that this component of the study was limited to the year 2017 only, employed a relatively small number of satellite precipitation and reanalysis products, and utilized a digital elevation model with a coarse 90-meter resolution.*

The study also develops a novel precipitation forecasting framework, which integrates these improved precipitation inputs, local meteorological variables, and large-scale climate drivers with multiple machine learning algorithms such as ANN, SVM, KNN, Random Forest, and XGBoost. The framework substantially improved the accuracy of monthly precipitation forecasts, especially at one-month lead times in the Upper Blue Nile basin, and can be adaptable to different climatic regimes. The forecasting framework exhibited good predictive skill for normal and moderately dry conditions while generalizing reasonably to unseen data. Even though challenges remained in accurately capturing extreme wet conditions, the results confirm that multi-source data-driven forecasting approaches remarkably enhance medium-term precipitation predictability, which is useful for climate-sensitive planning and decision-making. Furthermore, in the study, we discovered performance variations between models under different climate conditions from the same input, which highlights the importance of selecting the better models for specific forecasting scenarios. The findings also highlight opportunities for expanding the framework by incorporating hybrid and deep learning models, as well as optimization strategies, to further strengthen forecast robustness and applicability across diverse hydrological settings. However, it must be noted that the precipitation forecasting framework was developed using only five machine learning algorithms, implemented on only five monitoring stations in the Upper Blue Nile Basin, and restricted to a monthly temporal scale. Additionally, component-wise performance and sensitivity analysis of the multi-source inputs was not carried out.

Another contribution of this dissertation work is the physics-informed hybrid modeling framework, which integrates the output from the process-based hydrological model Soil and Water Assessment Tool (SWAT) with Long Short-Term Memory (LSTM) networks. This integration effectively captures the physical watershed processes alongside complex nonlinear temporal dependencies in streamflow data, thereby address the limitations of the

individual approaches. Moreover, the novelty of the framework lies in the inclusion of large-scale climate teleconnection indices such as ENSO and the Indian Ocean Dipole alongside local meteorological and hydrological variables which influence monthly and seasonal hydrological variabilities. The use of this multi-source predictor set reduces systematic biases and errors inherent in the modeling process. Streamflow can then be forecasted with improved physical realism and operational relevance, benefiting upstream short-term management and long-term water resource planning in the basin. Nonetheless, within this hybrid framework, uncertainty analysis was not performed on the SWAT component to quantify how errors propagate through the coupled LSTM model. Furthermore, component-wise interpretability analysis (e.g., SHAP) was not conducted, and the framework was not tested on temporal scales other than monthly.

The study's findings indicate that an effective transfer of hydrological knowledge embedded in an LSTM model, which is trained on a data-rich source basin, provides accurate predictions at a hydrologically similar but data-limited target basin. The systematic clustering of sub-basins based on physiographic and hydroclimatic similarities ensures the relevance of transferred knowledge, while fine-tuning adapts the model to local conditions. This method improves forecast accuracy in limited data situations and provides an efficient scalable solution for regional hydrological prediction, aiding in better disaster risk management and water resources management. However, the transfer learning framework was implemented using only a single donor–target basin pair and was tested exclusively on hydrologically similar basins; its performance on less similar basins was not examined.

In general, the integrated approach developed in this study combines multi-source data merging, machine-learning-based forecasting, physics-informed hybrid modeling, and transfer learning, which enhance forecast skill while mitigating limitations arising from data scarcity and the complex terrain of the Upper Blue Nile Basin. This system generates improved precipitation and streamflow forecasts to ensure greater reliability and spatial completeness of hydrological information for more sustainable water resource management and risk mitigation decisions. The methodologies and findings of this dissertation provide a valuable framework that can be adapted and extended to other

similarly challenging basins worldwide, promoting resilience in water management under changing climatic and environmental conditions.

8. Recommendations

This study demonstrates a significant improvement in hydrological forecasting performance, an enhancement that benefited from the integration of multi-source data, process-based hydrological models, and machine learning techniques. Nevertheless, several limitations warrant acknowledgment, leading to the following recommendations and future research directions:

I. Short-Term Methodological Improvement Recommendations

- The framework successfully demonstrated significant forecast skill improvement for the study period from 2010 to 2017. However, future studies should extend the study period to fully capture climatic variability, including important hydrological extreme events.
- This study produces improved spatiotemporal precipitation estimates by merging gauge measurements, three satellite precipitation products, and one reanalysis dataset. Future work should aim to further enhance accuracy by incorporating additional satellite products, emerging remote sensing technologies, and a broader range of reanalysis datasets.
- Before deploying the monthly precipitation forecasting for operational purposes, it has to be validated with longer data sets, in particular, to deal with performance problems during wet seasons.
- The precipitation forecasting from work should be validated with additional station data.
- The Digital Elevation Model used in the precipitation merging framework is coarser, and the performance has to be tested using a finer resolution DEM.
- The scope of transfer learning in this study is limited to two sub-basins (Guder and MUGER). Even though the framework demonstrated substantial promise in terms of effectiveness, future research is required to test it across broader sub-basins with

varying physiographic characteristics to further validate the physics-informed transfer learning framework for ungauged sub-basins.

- The current framework focuses on monthly precipitation and streamflow forecasting. To further exploit the full potential of the framework, subsequent research should extend the method to other lead times, such as daily and seasonal scales.
- Future studies ought to assess the transferability and adaptability of the hybrid SWAT-LSTM framework to other watersheds with varying climatic, topographic, and land use attributes, including larger river basins.
- To improve model inputs and accuracy forecasting skills, integration of emerging data sources such as high-resolution remote sensing products, citizen science data, soil moisture observations, and groundwater monitoring should be assessed.

II. Long-Term Methodological Improvement Recommendations

- The current streamflow forecasting framework solely utilizes the SWAT and LSTM models. Future studies should investigate the performance of the proposed framework in terms of accuracy, adaptability, and generalization by employing different process-based and machine learning models across heterogeneous hydroclimatic river basins.
- Integrating explainable artificial intelligence (XAI) methods in the LSTM component of the framework could solve the interpretability (black box) problem to enhance transparency and understanding of the internal decision-making processes of the model, and should apply Explainable AI (SHAP) analysis to quantify each feature's marginal contribution to every prediction.
- The hybrid framework relies on the outputs of the process-based SWAT model. It assumes that the SWAT model is based on accurate calibration and parameterization, which may be difficult in a data-scarce region. However, errors and uncertainties induced by the SWAT simulations may propagate through the hybrid modelling framework. Future research should incorporate proper

uncertainty assessment in the hybrid model to characterize forecast confidence and to inform decision-making.

- The multisource data integration enhances the accuracy and robustness of forecasts. However, the datasets come from multiple sources and notably impact the framework's performance due to differences in quality and spatio-temporal resolution, which may be limited and inconsistent in some regions. Therefore, it is essential to resample and validate the input data before using it in the modeling framework.
- The monthly multi-source data integrated precipitation forecasting has been tested using a few machine learning models, namely, ANN, SVM, KNN, Random Forest, and XGBoost. Future studies should further explore the framework using emerging hybrid models, deep learning architectures, and ensemble methods that may offer improved predictive performance.
- This study focuses on knowledge transfer between two hydrologically similar basins. Future work should extend this by exploring multi-source transfer learning, which integrates knowledge from multiple data-rich basins to enhance robustness.

Reference

Abbot, J., & Marohasy, J. (2012). Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences*, *29*(4), 717–730. https://doi.org/10.1007/s00376-012-1259-9

Abbot, J., & Marohasy, J. (2018). Forecasting of medium-term rainfall using artificial neural networks: Case studies from Eastern Australia. In *Engineering mathematics and topical issues in rainfall*. https://doi.org/10.5772/intechopen.72619

Abebe, S., Qin, T., Yan, D., Gelaw, E. B., Workneh, H., Kun, W., Liu, S., & Dong, B. (2020). Spatial and temporal evaluation of the latest high-resolution precipitation products over the Upper Blue Nile River Basin. *Water*, *12*(11), 3072. https://doi.org/10.3390/w12113072

Abera, W., Brocca, L., & Rigon, R. (2016). Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. Institutional Research Information System (Università Degli Studi Di Trento), 471–483. https://doi.org/10.1016/j.atmosres.2016.04.017

Abera, W., Formetta, G., Brocca, L., & Rigon, R. (2017). Modeling the water budget of the Upper Blue Nile basin using the JGrass-NewAge model system and satellite data. *Hydrology and Earth System Sciences*, *21*(6), 3145–3165. https://doi.org/10.5194/hess-21-3145-2017

Aboelnour, M., Gitau, M. W., & Engel, B. A. (2020). A comparison of streamflow and baseflow responses to land-use change and the variation in climate parameters using SWAT. *Water*, *12*(1), 191. https://doi.org/10.3390/w12010191

Ademe, D., Zaitchik, B. F., Tesfaye, K., Simane, B., Alemayehu, G., & Adgo, E. (2020). Climate trends and variability at adaptation scale: Patterns and perceptions in an agricultural region of the Ethiopian Highlands. *Weather and Climate Extremes*, *29*, 100263. https://doi.org/10.1016/j.wace.2020.100263

Adera, S., Bellugi, D., Dhakal, A. S., & Larsen, L. G. (2024). Streamflow prediction at the intersection of physics and machine learning: A case study of two Mediterranean-

climate watersheds. *Water Resources Research*, *60*(7). https://doi.org/10.1029/2023wr035790

Aghelpour, P., Bahrami-Pichaghchi, H., Varshavian, V., Graf, R., Sefidkouhi, M., & Khoshravesh, M. (2022). Evaluating the impact of large-scale climatic indices as inputs for forecasting monthly river flow in Mazandaran Province, Iran. *Pure and Applied Geophysics*, *179*, 1309–1331. https://doi.org/10.1007/s00024-022-02970-9

Aguasca-Colomo, C.-N., & Méndez, A. (2019). Comparative analysis of rainfall prediction models using machine learning in islands with complex orography: Tenerife Island. *Applied Sciences*, *9*(22), 4931. https://doi.org/10.3390/app9224931

Ahmad, K., Iqbal, M., Tariq, M. A. U. R., Khan, A. U., Nadeem, A., Chen, J., Usanova, K., Almujibah, H., Alyami, H., & Abid, M. (2025). Exploring alternate coupling inputs of a data-driven model for optimum daily streamflow prediction in calibrated SWAT-BiLSTM rainfall-runoff modeling. *Frontiers in Water*, *7*, 1558218. https://doi.org/10.3389/frwa.2025.1558218

Ahmed, A., Deo, R., Feng, Q., Ghahramani, A., Raj, N., Yin, Z., & Yang, L. (2021). Deep learning hybrid model with Boruta-Random forest optimiser algorithm for streamflow forecasting with climate mode indices, rainfall, and periodicity. *Journal of Hydrology*, *599*, 126350. https://doi.org/10.1016/j.jhydrol.2021.126350

Ahmed, J. S., Buizza, R., Dell'Acqua, M., et al. (2024). Evaluation of ERA5 and CHIRPS rainfall estimates against observations across Ethiopia. Meteorology and Atmospheric Physics, 136, Article 17. https://doi.org/10.1007/s00703-024-01008-0

Aich, V., Liersch, S., Vetter, T., Huang, S., Tecklenburg, J., Hoffmann, P., Koch, H., Fournet, S., Krysanova, V., Müller, E. N., & Hattermann, F. F. (2014). Comparing impacts of climate change on streamflow in four large African river basins. Hydrology and Earth System Sciences, 18(4), 1305–1321. https://doi.org/10.5194/hess-18-1305-2014

Ali Mohammed, J., Gashaw, T., Worku Tefera, G., Dile, Y. T., Worqlul, A. W., & Addisu, S. (2022). Changes in observed rainfall and temperature extremes in the Upper Blue Nile Basin of Ethiopia. *Weather and Climate Extremes*, *37*, 100468. https://doi.org/10.1016/j.wace.2022.100468

Ali, A. M., Melsen, L., & Teuling, A. J. (2023). Inferring reservoir filling strategies under limited-data-availability conditions using hydrological modeling and Earth observations: the case of the Grand Ethiopian Renaissance Dam (GERD). Hydrology and Earth System Sciences, 27(21), 4057–4086. https://doi.org/10.5194/hess-27-4057-2023

Ali, M. H., Popescu, I., Jonoski, A., & Solomatine, D. (2023). Remote sensed and/or global datasets for distributed hydrological modelling: A review. *Remote Sensing*, *15*(6), 1642. https://doi.org/10.3390/rs15061642

Ali, Z., Hussain, I., Faisal, M., et al. (2017). Forecasting drought using multilayer perceptron artificial neural network model. *Advances in Meteorology*, Article 5681308. https://doi.org/10.1155/2017/5681308

Alizadeh, B., Bafti, A. G., Kamangir, H., Zhang, Y., Wright, D. B., & Franz, K. J. (2021). A novel attention-based LSTM cell post-processor coupled with Bayesian optimization for streamflow prediction. *Journal of Hydrology*, *601*, 126526. https://doi.org/10.1016/j.jhydrol.2021.126526

Allam, M., & Eltahir, E. A. B. (2017). Optimal allocation of land and water resources to achieve water, energy and food security in the upper Blue Nile basin. In *AGU Fall Meeting Abstracts 2017*.

Alshboul, O., Shehadeh, A., Almasabha, G., & Almuflih, A. S. (2022). Extreme gradient boosting-based machine learning approach for green building cost prediction. *Sustainability*, *14*(11), 6651. https://doi.org/10.3390/su14116651

Amin, D., Keshta, E., Ragab, R., Chakrabortty, R., & Elaksher, A. (2025). Impact of Climate Change on Hydrological Extremes (Floods and Droughts) in the Upper Blue Nile Basin. Earth Systems and Environment. https://doi.org/10.1007/s41748-025-00709-9

An, Z. W., Li, C., & Liu, Y. (2020). Evaluation of six satellite and reanalysis precipitation products using gauge observations over the Yellow River Basin, China. *Atmosphere*, *11*(11), 1223. https://doi.org/10.3390/atmos11111223

Anjum, M. N., Irfan, M., Waseem, M., Leta, M. K., Niazi, U. M., et al. (2022). Assessment of PERSIANN-CCS, PERSIANN-CDR, SM2RAIN-ASCAT, and CHIRPS-2.0 rainfall products over a semi-arid subtropical climatic region. *Water*, *14*(2), 147. https://doi.org/10.3390/w14020147

Ansarifard, S., Ghorbanifard, M., Boustani, F., & Abdolazimi, H. (2024). Hydrological simulation and evaluation of drought conditions in the ungauged watershed Parishan lake Iran, using the SWAT model. *Journal of Water and Climate Change*, *15*(9), 4666. https://doi.org/10.2166/wcc.2024.268

Anshuka, A., Chandra, R., Buzacott, A., Sanderson, D., & van Ogtrop, F. (2022). Spatio temporal hydrological extreme forecasting framework using LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment*, *36*(10), 3467. https://doi.org/10.1007/s00477-022-02204-3

Anwar, M. T., et al. (2021). Rainfall prediction using extreme gradient boosting. *Journal of Physics: Conference Series*, *1869*, 8.

Araza, A., Hein, L., Duku, C., Rawlins, M. A., & Lomboy, R. (2020). Data-driven streamflow modelling in ungauged basins: Regionalizing random forest (RF) models. *bioRxiv*. https://doi.org/10.1101/2020.11.14.382598

Arnold, J., Moriasi, D., Gassman, P., Abbaspour, K., White, M., Raghavan, S., Santhi, C., Harmel, D., van Griensven, A., Kannan, N., & Jha, M. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*, 1491–1508. https://doi.org/10.13031/2013.42256

Arnone, E., Cucchi, M., Gesso, S. D., Petitta, M., & Calmanti, S. (2020). Droughts prediction: A methodology based on climate seasonal forecasts. *Water Resources Management*, *34*(14), 4313–4328. https://doi.org/10.1007/s11269-020-02623-3

Arsenault, R., Martel, J., Brunet, F., Brissette, F., & Mai, J. (2023). Continuous streamflow prediction in ungauged basins: Long short-term memory neural networks clearly outperform traditional hydrological models. *Hydrology and Earth System Sciences*, *27*(1), 139. https://doi.org/10.5194/hess-27-139-2023

Asadi, S., Jimeno-Sáez, P., López-Ballesteros, A., & Senent-Aparicio, J. (2025). Coupling SWAT+ with LSTM for enhanced and interpretable streamflow estimation in arid and semi-arid watersheds, a case study of the Tagus Headwaters River Basin, Spain. *Environmental Modelling & Software*, *186*, 106360. https://doi.org/10.1016/j.envsoft.2025.106360

Atiah, W. A., Amekudzi, L. K., Aryee, J. N., Preko, K., & Danuor, S. K. (2020). Validation of satellite and merged rainfall data over Ghana, West Africa. *Atmosphere*, *11*(8), 859. https://doi.org/10.3390/atmos11080859

Ayehu, G. T., Tadesse, T., Gessesse, B., & Dinku, T. (2018). Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. *Atmospheric Measurement Techniques*, *11*, 1921–1936. https://doi.org/10.5194/amt-11-1921-2018

Ayehu, G. T., Tadesse, T., Gessesse, B., & Dinku, T. (2018). Validation of new satellite rainfall products over the Upper Blue Nile Basin, Ethiopia. Atmospheric Measurement Techniques, 11(4), 1921–1936. https://doi.org/10.5194/amt-11-1921-2018

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L., Nauditt, A., Birkel, C., Verbist, K., Giraldo-Osorio, J. D., & Xuan, T. N. (2020). RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, *239*, 111606. https://doi.org/10.1016/j.rse.2019.111606

Bao, T., Johnson, T. T., & Jia, X. (2024). Transfer learning using inaccurate physics rule for streamflow prediction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*. https://doi.org/10.24963/ijcai.2024/793

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, *7*, 100204. https://doi.org/10.1016/j.mlwa.2021.100204

Bayissa, Y., Moges, S., Melesse, A., Tadesse, T., Abiy, A., & Worqlul, A. (2021). Multi-dimensional drought assessment in Abbay/Upper Blue Nile Basin: The importance of shared management and regional coordination efforts for mitigation. *Remote Sensing*, *13*(9), 1835. https://doi.org/10.3390/rs13091835

Bayissa, Y., Tadesse, T., Demisse, G., & Shiferaw, A. (2017). Evaluation of Satellite-Based Rainfall Estimates and Application to Monitor Meteorological Drought for the Upper Blue Nile Basin, Ethiopia. *Remote Sensing*, *9*(7), 669. https://doi.org/10.3390/rs9070669

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., et al. (2019). Daily evaluation of 26 precipitation datasets using stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*, 207–224. https://doi.org/10.5194/hess-23-207-2019

Belay, A. S., Fenta, A. A., Yenehun, A., Nigate, F., Tilahun, S. A., Moges, M. M., Dessie, M., Adgo, E., Nyssen, J., Chen, M., Griensven, A. V., & Walraevens, K. (2019). Evaluation and application of multi-source satellite rainfall product CHIRPS to assess spatio-temporal rainfall variability on data-scarce western margins of Ethiopian highlands. *Remote Sensing*, *11*(22), 2688. https://doi.org/10.3390/rs11222688

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., et al. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4186–4227. https://doi.org/10.1002/qj.4174

Bhasme, P., Vagadiya, J., & Bhatia, U. (2022). Enhancing predictive skills in physically-consistent way: Physics informed machine learning for hydrological

processes. *Journal of Hydrology*, *615*, 128618. https://doi.org/10.1016/j.jhydrol.2022.128618

Bhattacharya, T., Khare, D., & Arora, M. (2020). Evaluation of reanalysis and global meteorological products in Beas river basin of North-Western Himalaya. *Environmental Systems Research*. https://doi.org/10.1186/s40068-020-00186-1

Bhuiyan, E. M., Nikolopoulos, E. I., & Anagnostou, E. N. (2019). Machine learning-based blending of satellite and reanalysis precipitation datasets: A multiregional tropical complex terrain evaluation. *Journal of Hydrometeorology*, *20*(11), 2147–2161. https://doi.org/10.1175/JHM-D-19-0073.1

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*, 197–227. https://doi.org/10.1007/s11749-016-0481-7

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. *Sustainability*, *17*(5), 2250. https://doi.org/10.3390/su17052250

Biruntha, S., Sowmiya, B. S., Subashri, R., & Vasanth, M. (2022). Rainfall prediction using kNN and decision tree. In *2022 International Conference on Electronics and Renewable Systems (ICEARS)*. IEEE. https://doi.org/10.1109/icears53579.2022.9752220

Bogerd, L., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2021). A comprehensive five-year evaluation of IMERG late run precipitation estimates over the Netherlands. *Journal of Hydrometeorology*, *22*(7), 1855–1868. https://doi.org/10.1175/JHM-D-21-0002.1

Bojang, P. O., Yang, T. C., Pham, Q. B., & Yu, P. S. (2020). Linking singular spectrum analysis and machine learning for monthly rainfall forecasting. *Applied Sciences*, *10*(9), 3224. https://doi.org/10.3390/app10093224

Cao, Z., Wang, S., Luo, P., Xie, D., & Zhu, W. (2022). Watershed ecohydrological processes in a changing environment: Opportunities and challenges. *Water*, *14*(9), 1502. https://doi.org/10.3390/w14091502

Chakilu, G., Sándor, S., & Zoltán, T. (2023). The Dynamics of Hydrological Extremes under the Highest Emission Climate Change Scenario in the Headwater Catchments of the Upper Blue Nile Basin, Ethi Lemann, T., Roth, V., Zeleke, G., Subhatu, A. T., Kassawmar, T., & Hurni,, H. (2018). Spatial and Temporal Variability in Hydrological Responses of the Upper Blue Nile basin, Ethiopia. Water. https://doi.org/10.3390/w11010021 opia. Water. https://doi.org/10.3390/w15020358

Chen, C., Zhang, Q., Kashani, M. H., et al. (2022). Forecast of rainfall distribution based on fixed sliding window long short-term memory. *Engineering Applications of Computational Fluid Mechanics*, *16*(1), 248–261. https://doi.org/10.1080/19942060.2021.2009374

Chen, H., Huang, J., & Huang, J.-C. (2023). Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach. *Journal of Hydrology*, *622*(Part A), Article 129734. https://doi.org/10.1016/j.jhydrol.2023.129734

Chen, H., Xu, Y., Teegavarapu, R., Guo, Y., & Xie, J. (2021). Assessing different roles of baseflow and surface runoff for long-term streamflow forecasting in southeastern China. *Hydrological Sciences Journal*, *66*, 2312–2329. https://doi.org/10.1080/02626667.2021.1988612

Chen, J., Zhang, X., Xiong, L., & Chen, H. (2022). A framework of integrating heterogeneous data sources for monthly streamflow prediction using a state-of-the-art deep learning model. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2022.128599

Chen, L., Yao, X., Liu, K., Tan, C., & Moon, S. K. (2023). Multisensor fusion-based digital twin in additive manufacturing for in-situ quality monitoring and defect correction. *arXiv*. https://doi.org/10.48550/arXiv.2304.05685

Chen, S., Xiong, L., Ma, Q., Kim, J. S., Chen, J., & Xu, C. Y. (2020). Improving daily spatial precipitation estimates by merging gauge observation with multiple satellite-based precipitation products based on the geographically weighted ridge regression method. *Journal of Hydrology*, *589*, 125156. https://doi.org/10.1016/j.jhydrol.2020.125156

Chen, Y., Sharma, S., Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., & K, N. (2021). Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. *Atmospheric Research*, *250*, 105365. https://doi.org/10.1016/j.atmosres.2020.105365

Chong, K. L., Huang, Y. F., Koo, C. H., Sherif, M., Ahmed, A. N., & El-Shafie, A. (2022). Investigation of cross-entropy-based streamflow forecasting through an efficient interpretable automated search process. *Applied Water Science*, *13*(1). https://doi.org/10.1007/s13201-022-01790-5

Chong, K. L., Lai, S. H., Yao, Y., Ahmed, A. N., Jaafar, W. Z. W., & El-Shafie, A. (2020). Performance enhancement model for rainfall forecasting utilizing integrated wavelet-convolutional neural network. *Water Resources Management*, *34*(8), 2371–2387. https://doi.org/10.1007/s11269-020-02554-z

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., & Goldscheider, N. (2023). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrology and Earth System Sciences*, *27*, 2397. https://doi.org/10.5194/hess-27-2397-2023

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., & Wood, A. W. (2015). A unified approach for process-based hydrologic modeling: Model implementation and case studies. *Water Resources Research*, *51*, 2515.

Cristiano, E., Veldhuis, T., & Van De Giesen, M. (2017). Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas: A review. *Hydrology and Earth System Sciences*, *21*(7), 3859–3878.

Crossett, C. C., Betts, A. K., Dupigny-Giroux, L. A., & Bomblies, A. (2020). Evaluation of daily precipitation from the ERA5 global reanalysis against GHCN observations in the Northeastern United States. *Climate*, *8*(12), 148. https://doi.org/10.3390/cli8120148

Dai, T., Maher, K., & Perzan, Z. (2024). Machine learning surrogates for efficient hydrologic modeling: Insights from stochastic simulations of managed aquifer recharge. *arXiv*. https://doi.org/10.48550/arxiv.2407.20902

Danandeh Mehr, A. (2020). An ensemble genetic programming model for seasonal precipitation forecasting. *International Journal of Environmental Science and Technology*, *16*(1), 335–346. https://doi.org/10.1007/s13762-018-1674-2

Danandeh Mehr, A. (2021). Seasonal rainfall hindcasting using ensemble multi-stage genetic programming. *Theoretical and Applied Climatology*, *143*(1), 461–472.

Danandeh Mehr, A., Nourani, V., Karimi Khosrowshahi, V., & Ghorbani, M. A. (2019). A hybrid support vector regression–firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology*.

Danandeh Mehr, A., Torabi Haghighi, A., Jabarnejad, M., Safari, M. J. S., & Nourani, V. (2022). A new evolutionary hybrid random forest model for SPEI forecasting. *Water*, *14*(5), 755. https://doi.org/10.3390/w14050755

Dawoodi, H. H. (2021). Rainfall prediction in north Maharashtra region using support vector machine. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2020.1734813

De Vera, A., Alfaro, P., & Terra, R. (2021). Operational implementation of satellite-rain gauge data merging for hydrological modeling. *Water*, *13*, 533. https://doi.org/10.3390/w13040533

Demessie, S., Dile, Y., Bedadi, B., Gashaw, T., & Tefera, G. (2023). Evaluations of regional climate models for simulating precipitation and temperature over the Guder sub-basin of Upper Blue Nile Basin Ethiopia. *Modeling Earth Systems and Environment*, *9*, 4455–4476. https://doi.org/10.1007/s40808-023-01751-0

Derin, Y., Kirstetter, P., & Gourley, J. (2021). Evaluation of IMERG satellite precipitation over the land–coast–ocean continuum Part I: Detection. *Journal of Hydrometeorology*, *22*(11), 2843–2859. https://doi.org/10.1175/JHM-D-21-0234.1

Dessie, M., Verhoest, N. E. C., Pauwels, V., Admasu, T., Poesen, J., Adgo, E., Deckers, J., & Nyssen, J. (2014). Analyzing runoff processes through conceptual hydrological modeling in the Upper Blue Nile Basin, Ethiopia. Hydrology and Earth System Sciences, 18(12), 5149–5167. https://doi.org/10.5194/hess-18-5149-2014

Dile, Y. T., Tekleab, S., Ayana, E. K., Gebrehiwot, S. G., Worqlul, A. W., Bayabil, H. K., Yimam, Y. T., Tilahun, S. A., Daggupati, P., Karlberg, L., & Srinivasan, R. (2018). Advances in water resources research in the Upper Blue Nile basin and the way forward: A review. *Journal of Hydrology*, *560*, 407–423. https://doi.org/10.1016/j.jhydrol.2018.03.042

Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., & Ceccato, P. (2018). Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quarterly Journal of the Royal Meteorological Society*, *144*, 292–312. https://doi.org/10.1002/qj.3244

Du, J., Fang, J., Xu, W., & Shi, P. (2013). Analysis of dry/wet conditions using the standardized precipitation index and its potential usefulness for drought/flood monitoring in Hunan Province, China. *Stochastic Environmental Research and Risk Assessment*, *27*(2), 377–387. https://doi.org/10.1007/s00477-012-0589-6

Duethmann, D., Blöschl, G., & Párajka, J. (2020). Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? *Hydrology and Earth System Sciences*, *24*(7), 3493. https://doi.org/10.5194/hess-24-3493-2020

Duguma, T. A. (2022). Soil erosion risk assessment and treatment priority classification: A case study on Guder watersheds, Abay river basin, Oromia, Ethiopia. *Heliyon*, *8*(8). https://doi.org/10.1016/j.heliyon.2022.e10183

Ebabu, K., Taye, G., Tsunekawa, A., Haregeweyn, N., Adgo, E., Tsubo, M., Fenta, A., Meshesha, D., Sultan, D., Aklog, D., Admasu, T., Van Wesemael, B., & Poesen, J. (2022). Land use management and climate effects on runoff and soil loss responses in the highlands of Ethiopia. *Journal of Environmental Management*, *326*(Pt A), 116707. https://doi.org/10.1016/j.jenvman.2022.116707

Eldardiry, H., & Hossain, F. (2019). Understanding reservoir operating rules in the transboundary Nile River Basin using macroscale hydrologic modeling with satellite measurements. *Journal of Hydrometeorology*, *20*(11), 2253–2269. https://doi.org/10.1175/JHM-D-19-0058.1

Engida, A. N. (2010). Hydrological and Suspended Sediment Modeling in the Lake Tana Basin, Ethiopia. HAL (Le Centre Pour La Communication Scientifique Directe). https://theses.hal.science/tel-00539152

Essou, G. R. C., Brissette, F., & Lucas-Picher, P. (2017). Impacts of combining reanalysis and weather station data on the accuracy of discharge modelling. *Journal of Hydrology*, *545*, 120–131. https://doi.org/10.1016/j.jhydrol.2016.12.021

Esteves, J. T., de Souza, R. G., & Ferraudo, A. S. (2019). Rainfall prediction methodology with binary multilayer perceptron neural networks. *Climate Dynamics*, *52*(3–4), 2319–2331. https://doi.org/10.1007/s00382-018-4252-x

Estévez, J., Bellido-Jiménez, J. A., Liu, X., & García-Marín, A. P. (2020). Monthly precipitation forecasts using wavelet neural networks models in a semiarid environment. *Water*, *12*(7), 1909. https://doi.org/10.3390/w12071909

Fan, Z., Li, W., Jiang, Q., Sun, W., Wen, J., & Gao, J. (2021). A comparative study of four merging approaches for regional precipitation estimation. *IEEE Access*, *9*, 33625–33637. https://doi.org/10.1109/ACCESS.2021.3057057

Farmani, M., Tavakoly, A., Behrangi, A., Qiu, Y., Gupta, A., Jawad, M., Sohi, H., Zhang, X., Geheran, M., & Niu, G. (2025). Improving streamflow predictions in the arid southwestern United States through understanding of baseflow generation mechanisms. *Water Resources Research*, *61*. https://doi.org/10.1029/2024wr039479

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, *56*(9). https://doi.org/10.1029/2019wr026793

Feng, D., Lawson, K., & Shen, C. (2021). Mitigating prediction error of deep learning streamflow models in large data-scarce regions with ensemble modeling and soft data. *Geophysical Research Letters*, *48*(14). https://doi.org/10.1029/2021gl092999

Fenta, A. A., Yasuda, H., Shimizu, K., Haregeweyn, N., Kawai, T., Sultan, D., & Belay, A. S. (2017). Spatial distribution and temporal trends of rainfall and erosivity in the eastern Africa region. *Hydrological Processes*, *31*, 4555–4567. https://doi.org/10.1002/hyp.11378

Fenta, M. D., & Disse, M. (2018). Analyzing the future climate change of the Upper Blue Nile River basin using statistical downscaling techniques. *Hydrology and Earth System Sciences*, *22*, 2391–2408. https://doi.org/10.5194/hess-22-2391-2018

Funk, C., Peterson, P., Landsfeld, M., et al. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. https://doi.org/10.1038/sdata.2015.66

Gadouali, F., & Messouli, M. (2020). Evaluation of multiple satellite-derived rainfall products over Morocco. *International Journal of Hydrological Sciences and Technology*, *10*(1), 72. https://doi.org/10.1504/ijhst.2020.104988

Gauch, M., Mai, J., & Lin, J. (2020). The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. *Environmental Modelling & Software*, *135*, 104926. https://doi.org/10.1016/j.envsoft.2020.104926

Gebrehiwot, S. G., Ilstedt, U., Gärdenäs, A. I., & Bishop, K. (2011). Hydrological characterization of watersheds in the Blue Nile Basin, Ethiopia. Hydrology and Earth System Sciences, 15(1), 11–20. https://doi.org/10.5194/hess-15-11-2011

Gebremedhin, M. A., Lubczynski, M. W., Maathuis, B. H., & Teka, D. (2020). Novel approach to integrate daily satellite rainfall with in-situ rainfall, Upper Tekeze Basin, Ethiopia. *Atmospheric Research*, *248*, 105135. https://doi.org/10.1016/j.atmosres.2020.105135

Gemechu, T. M., Zhao, H., Bao, S.-S., Cidan, Y., Liu, Y., Li, F., & Li, H. (2021). Estimation of hydrological components under current and future climate scenarios in Guder catchment, Upper Abbay Basin, Ethiopia, using the SWAT. *Sustainability*, *13*(17), 9689. https://doi.org/10.3390/su13179689

Ghamariadyan, M., & Imteaz, M. A. (2021). Monthly rainfall forecasting using temperature and climate indices through a hybrid method in Queensland, Australia. *Journal of Hydrometeorology*. https://doi.org/10.1175/jhm-d-20-0169.1

Ghozat, A., Sharafati, A., & Hosseini, S. A. (2022). Satellite-based monitoring of meteorological drought over different regions of Iran: Application of the CHIRPS precipitation product. *Environmental Science and Pollution Research International*, *29*, 36115–36132. https://doi.org/10.1007/s11356-022-18773-3

Gunathilake, M. B., Senerath, T., & Rathnayake, U. (2021). Artificial neural network based PERSIANN datasets in evaluation of hydrologic utility of precipitation estimations in a tropical watershed of Sri Lanka. *AIMS Geosciences*, *7*(3), 478–489. https://doi.org/10.3934/geosci.2021027

Habibi, S., & Hassanpour, S. T. (2025). An explainable machine learning framework for forecasting lake water equivalent using satellite data: A 20-year analysis of the Urmia Lake Basin. *Water*, *17*(10), 1431. https://doi.org/10.3390/w17101431

Habtegebreal, A. K., Alemu, A. B., & Raju, U. J. P. (2021). Examining the Role of Quasi-biennial Oscillation on Rainfall patterns over Upper Blue Nile Basin of Ethiopia. SHILAP Revista de Lepidopterología, 8(3), 190. https://doi.org/10.3934/environsci.2021013

Haddad, K. (2025). A comprehensive review and application of Bayesian methods in hydrological modelling: Past, present, and future directions. *Water*, *17*, 1095.

Haile, A. T., Akawka, A. L., Berhanu, B., & Rientjes, T. (2017). Changes in water availability in the Upper Blue Nile basin under the representative concentration pathways scenario. *Hydrological Sciences Journal*, *62*(13), 2139–2149. https://doi.org/10.1080/02626667.2017.1365149

Haile, A., Akawka, A. L., Berhanu, B., & Rientjes, T. (2017). Changes in water availability in the Upper Blue Nile basin under the representative concentration pathways scenario. Hydrological Sciences Journal, 62, 2139 - 2149. https://doi.org/10.1080/02626667.2017.1365149

Haile, M., & Abebe, A. (2022). GIS and fuzzy logic integration in land suitability assessment for surface irrigation: The case of Guder watershed Upper Blue Nile Basin Ethiopia. *Applied Water Science*, *12*. https://doi.org/10.1007/s13201-022-01761-w

Hallouin, T., Bruen, M., & O'Loughlin, F. (2020). Calibration of hydrological models for ecologically relevant streamflow predictions: A trade-off between fitting well to data and estimating consistent parameter sets? *Hydrology and Earth System Sciences*, *24*(3), 1031. https://doi.org/10.5194/hess-24-1031-2020

Hamed, M. M., Nashwan, M. S., & Shahid, S. (2021). Performance evaluation of reanalysis precipitation products in Egypt using fuzzy entropy time series similarity analysis. *International Journal of Climatology*, *41*(11), 5431–5446. https://doi.org/10.1002/joc.7286

Hamidi, O., Poorolajal, J., Sadeghifar, M., et al. (2015). A comparative study of support vector machines and artificial neural networks for predicting precipitation in Iran. *Theoretical and Applied Climatology*, *119*(3–4), 723–731. https://doi.org/10.1007/s00704-014-1141-z

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Zhang, L., Han, W., Huang, M., Jin, Q., Lan, Y., Liu, Y., Liu, Z., Lu, Z., Qiu, X., Song, R., Tang, J., Wen, J.-R., … Zhu, J. (2021). Pre-trained models: Past, present and future. *arXiv*. https://doi.org/10.48550/arXiv.2106.07139

Hanoon, M. S., Ahmed, A. N., Zaini, N., et al. (2021). Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia. *Scientific Reports*, *11*(1), 18935. https://doi.org/10.1038/s41598-021-96872-w

Haregeweyn, N., Tsunekawa, A., Poesen, J., Tsubo, M., Meshesha, D., Fenta, A., Nyssen, J., & Adgo, E. (2016). Comprehensive assessment of soil erosion risk for better land use planning in river basins: Case study of the Upper Blue Nile River.. The Science of the total environment, 574, 95-108 . https://doi.org/10.1016/j.scitotenv.2016.09.019

Hartanto, I., Almeida, C., & Alexandridis, T. (2015). Merging earth observation data, weather predictions, in-situ measurements and hydrological models for water information services. *Journal of Environmental Engineering and Management*, *14*, 2031–2042.

Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D., & Khan, S. (2022). Ensemble representation of satellite precipitation uncertainty using a nonstationary, anisotropic autocorrelation model. *Water Resources Research*, *58*(8). https://doi.org/10.1029/2021wr031650

Hasan, F. M., Medley, P., Drake, J., & Chen, G. (2024). Advancing hydrology through machine learning: Insights, challenges, and future directions using the CAMELS, Caravan, GRDC, CHIRPS, PERSIANN, NLDAS, GLDAS, and GRACE datasets. *Water*, *16*(13), 1904. https://doi.org/10.3390/w16131904

Hazra, A., Maggioni, V., Houser, P., Antil, H., & Noonan, M. (2019). A Monte Carlo-based multi-objective optimization approach to merge different precipitation estimates for land surface modeling. *Journal of Hydrology*, *570*, 454–462. https://doi.org/10.1016/j.jhydrol.2018.12.039

He, X., Pan, M., Wei, Z., Wood, E. F., & Sheffield, J. (2020). A global drought and flood catalogue from 1950 to 2016. *Bulletin of the American Meteorological Society*, *5*, E508–E535. https://doi.org/10.1175/BAMS-D-18-0269.1

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*, e5518. https://doi.org/10.7717/peerj.5518

Hénin, R., Liberato, M., Ramos, A., & Gouveia, C. (2018). Assessing the use of satellite-based estimates and high-resolution precipitation datasets for the study of extreme precipitation events over the Iberian Peninsula. *Water*, *10*(11), 1688. https://doi.org/10.3390/w10111688

Hersbach, H., Bell, B., Berrisford, P., et al. (2023). ERA5 hourly data on single levels from 1940 to present [Data set]. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Hill, A. J., & Schumacher, R. S. (2021). Forecasting excessive rainfall with random forests and a deterministic convection-allowing model. *Weather and Forecasting*, *36*(5), 1693–1711. https://doi.org/10.1175/waf-d-21-0026.1

Hofmeister, M., Mosbach, S., Hammacher, J., Blum, M., Röhrig, G., Dörr, C., Flegel, V., Bhave, A., & Kraft, M. (2021). Resource-optimised generation dispatch strategy for district heating systems using dynamic hierarchical optimisation. *Applied Energy*, *305*, 117877. https://doi.org/10.1016/j.apenergy.2021.117877

Hong, Y., Gochis, D., Cheng, J., Hsu, K., & Sorooshian, S. (2007). Evaluation of PERSIANN-CCS rainfall measurement using the NAME event rain gauge network. *Journal of Hydrometeorology*, *8*(3), 469–482. https://doi.org/10.1175/JHM574.1

Hong, Y., Hsu, K. L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*, 1834–1853. https://doi.org/10.1175/JAM2173.1

Hossain, I., Rasel, H. M., Imteaz, M. A., & Mekanik, F. (2020). Long-term seasonal rainfall forecasting using linear and non-linear modelling approaches: A case study for Western Australia. *Meteorology and Atmospheric Physics*, *132*(1), 131–141. https://doi.org/10.1007/s00703-019-00679-4

Hossain, M. M., Anwar, A. H. M. F., Garg, N., Prakash, M., & Bari, M. (2022). Monthly rainfall prediction at catchment level with the Facebook Prophet model using observed and CMIP5 decadal data. *Hydrology*, *9*(6), 111. https://doi.org/10.3390/hydrology9060111

Hu, Q., Li, Z., Wang, L., Huang, Y., Wang, Y., & Li, L. (2019). Rainfall spatial estimations: A review from spatial interpolation to multi-source data merging. *Water*, *11*(3), 579. https://doi.org/10.3390/w11030579

Huang, C., Zhang, Y., & Hou, J. (2024). Soil and Water Assessment Tool (SWAT)-informed deep learning for streamflow forecasting with remote sensing and in situ precipitation and discharge observations. *Remote Sensing*, *16*, 3999. https://doi.org/10.3390/rs16213999

Huang, F., & Zhang, X. (2024). A new interpretable streamflow prediction approach based on SWAT-BiLSTM and SHAP. *Environmental Science and Pollution Research*, *31*, 23896–23908. https://doi.org/10.1007/s11356-024-32725-z

Huang, M., Runsheng, S., & Xing, T. (2017). A novel approach for precipitation forecast via improved K-nearest neighbor algorithm. *Advanced Engineering Informatics*. https://doi.org/10.1016/j.aei.2017.05.003

Huffman, G. J., Bolvin, D. T., Braithwaite, D., et al. (2020). Integrated multi-satellite retrievals for the Global Precipitation Measurement (GPM) mission (IMERG). In *Advances in global change research* (pp. 343–353). Springer. https://doi.org/10.1007/978-3-030-24568-9_19

Hutter, F. (2019). *Automated machine learning: Methods, systems, challenges*. Springer.

Huynh, N. N. T., Garambois, P., Colléoni, F., Renard, B., Roux, H., Demargne, J., & Javelle, P. (2023). Learning regionalization within a differentiable high-resolution

hydrological model using accurate spatial cost gradients. *Zenodo*. https://doi.org/10.5281/zenodo.xxx [Note: DOI missing; kept as original]

Idowu, M., Kulls, C., & Kisi, O. (2024). A new method for monthly streamflow prediction using multi-source data: Range-dependent multivariate adaptive regression splines–genetic algorithm. *Hydrological Sciences Journal*, *69*, 1860–1880. https://doi.org/10.1080/02626667.2024.2394639

Jaiswal, R. K., Ali, S., & Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, *10*(1). https://doi.org/10.1007/s13201-019-1122-6

Jang, D. (2021). An application of ANN ensemble for estimating of precipitation using regional climate models. *Advances in Civil Engineering*, *2021*, 1–16. https://doi.org/10.1155/2021/7363471

Jia, X., Zwart, J. A., Sadler, J. M., Appling, A., Oliver, S. K., Markstrom, S. L., Willard, J., Xu, S., Steinbach, M., Read, J. S., & Kumar, V. (2020). Physics-guided recurrent graph networks for predicting flow and temperature in river networks. *arXiv*. https://doi.org/10.48550/arXiv.2009.12575

Jiao, D., Xu, N., Yang, F., & Xu, K. (2021). Evaluation of spatial-temporal variation performance of ERA5 precipitation data in China. *Scientific Reports*, *11*(1), 17956. https://doi.org/10.1038/s41598-021-97432-y

Jiao, G., Guo, T., & Ding, Y. (2016). A new hybrid forecasting approach applied to hydrological data: A case study on precipitation in northwestern China. *Water*, *8*(9), 367. https://doi.org/10.3390/w8090367

Julia, T. (2019). Environmental Change and African Societies. https://doi.org/10.1163/9789004410848

Jung, H. C., Getirana, A., Policelli, F., McNally, A., Arsenault, K. R., Kumar, S. V., Tadesse, T., & Peters-Lidard, C. D. (2017). Upper Blue Nile basin water budget

from a multi-model perspective. Journal of Hydrology, 555, 535 – 546. https://doi.org/10.1016/j.jhydrol.2017.10.040

Jung, H. C., Getirana, A., Policelli, F., McNally, A., Arsenault, K. R., Kumar, S. V., Tadesse, T., & Peters-Lidard, C. D. (2017). Upper Blue Nile basin water budget from a multi-model perspective. Journal of Hydrology, 555, 535 – 546. https://doi.org/10.1016/j.jhydrol.2017.10.040

Jung, H. C., Getirana, A., Policelli, F., McNally, A., Arsenault, K. R., Kumar, S. V., Tadesse, T., & Peters-Lidard, C. D. (2017). Upper Blue Nile basin water budget from a multi-model perspective. Journal of Hydrology, 555, 535 – 546. https://doi.org/10.1016/j.jhydrol.2017.10.040

Kalteh, A. M. (2017). Enhanced monthly precipitation forecasting using artificial neural network and singular spectrum analysis conjunction models. *INAE Letters*, *2*(3), 73–81. https://doi.org/10.1007/s41403-017-0025-9

Kamyab, H., Khademi, T., Chelliapan, S., SaberiKamarposhti, M., Rezania, S., Yusuf, M., Farajnezhad, M., Abbas, M., Jeon, B., & Ahn, Y. (2023). The latest innovative avenues for the utilization of artificial intelligence and big data analytics in water resource management. *Results in Engineering*, *20*, 101566. https://doi.org/10.1016/j.rineng.2023.101566

Kang, J., Wang, H., Yuan, F., Wang, Z., Huang, J., & Qiu, T. (2020). Prediction of precipitation based on recurrent neural networks in Jingdezhen, Jiangxi province, China. *Atmosphere*, *11*(3), 246. https://doi.org/10.3390/atmos11030246

Kellengere Shankarnarayan, V., & Ramakrishna, H. (2021). Comparative study of three stochastic future weather forecast approaches: A case study. *Data Science and Management*, *3*, 3–12. https://doi.org/10.1016/j.dsm.2021.07.002

Keyantash, J. (2023). Standardized Precipitation Index (SPI). In *The Climate Data Guide*. National Center for Atmospheric Research. Retrieved November 20, 2023, from https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi

Khadim, F., Dokou, Z., Lazin, R., Moges, S., Bagtzoglou, A., & Anagnostou, E. (2020). Groundwater modeling in data scarce aquifers: The case of Gilgel-Abay, Upper Blue Nile, Ethiopia. *Journal of Hydrology*, *590*, 125214. https://doi.org/10.1016/j.jhydrol.2020.125214

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*, 69–78. https://doi.org/10.1175/bams-d-14-00283.1

Kim, J., & Han, H. (2021). Evaluation of the CMORPH high-resolution precipitation product for hydrological applications over South Korea. *Atmospheric Research*, *258*, 105650. https://doi.org/10.1016/j.atmosres.2021.105650

Kim, T., Shin, J., Kim, H., Kim, S., & Heo, J. (2019). The use of large-scale climate indices in monthly reservoir inflow forecasting and its application on time series and artificial intelligence models. *Water*. https://doi.org/10.3390/w11020374

Konapala, G., Kao, S., Painter, S., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344. https://doi.org/10.1029/2019wr026065

Kreibich, H., Pantillon, F., Parolai, S., Pinto, J. G., Punge, H. J., Rivalta, E., Schröter, K., Strehlow, K., Weisse, R., & Wurpts, A. (2020). Impact forecasting to support emergency management of natural hazards. *Reviews of Geophysics*, *58*(1), 1–52. https://doi.org/10.1029/2020RG000704

Kuana, L. A., Almeida, A. S., Mercuri, E. G. F., & Noe, S. M. (2024). Regionalization of GR4J model parameters for river flow prediction in Paraná, Brazil. *Hydrology and*

Earth System Sciences, *28*(14), 3367. https://doi.org/10.5194/hess-28-3367-2024

Kumar, D., Singh, A., Samui, P., & Jha, R. K. (2019). Forecasting monthly precipitation using sequential modeling. *Hydrological Sciences Journal*, *6*, 690–700.

Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D., & Tota-Maharaj, K. (2023). The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. *Sustainability*, *15*, 10543.

Kurtz, W., Lapin, A., & Schilling, O. (2017). Integrating hydrological modeling, data assimilation and cloud computing for real-time management of water resources. *Environmental Modelling & Software*, *3*, 418–435.

Lakew, H. B. (2020). Investigating the effectiveness of bias correction and merging MSWEP with gauged rainfall for the hydrological simulation of the upper Blue Nile basin. *Journal of Hydrology: Regional Studies*, *32*, 100741. https://doi.org/10.1016/j.ejrh.2020.100741

Lakew, H. B., & Moges, S. A. (2021). Dynamical bias correction procedure to improve global gridded daily streamflow data for local application in the Upper Blue Nile basin. Journal of Hydrology and Hydromechanics, 69(1), 41–48. https://doi.org/10.2478/johh-2020-0040

Lakew, H. B., & Moges, S. A. (2021). Dynamical bias correction procedure to improve global gridded daily streamflow data for local application in the Upper Blue Nile basin. Journal of Hydrology and Hydromechanics, 69(1), 41–48. https://doi.org/10.2478/johh-2020-0040

Lakew, H. B., Moges, S. A., & Asfaw, D. H. (2017). Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile Basin: A case study of Gilgel Abbay. *Hydrology*, *4*(3), 39. https://doi.org/10.3390/hydrology4030039

Lanza, L. G., & Cauteruccio, A. (2022). Precipitation measurement instruments: Calibration, accuracy and performance. *Water*, *14*(5), 811. https://doi.org/10.3390/w14050811

Larnier, K., & Monnier, J. (2020). Hybrid neural network – variational data assimilation algorithm to infer river discharges from SWOT-like data. *Nonlinear Processes in Geophysics Discussions*. https://doi.org/10.5194/npg-2020-32

Lauro, C. di, Vich, A. I. J., Rivera, J. A., & Moreiras, S. M. (2024). Non-stationary models for hydrological extremes in the mountain rivers of the Argentinean Central Andes. *Hydrological Sciences Journal*, *69*(10), 1301. https://doi.org/10.1080/02626667.2024.2369627

Lee, J., Kim, C. G., Lee, J., Kim, N., & Kim, H. (2018). Application of artificial neural networks to rainfall forecasting in the Geum River Basin, Korea. *Water*, *10*(10), 1448. https://doi.org/10.3390/w10101448

Leta, M. K., Ebsa, D. G., & Regasa, M. S. (2022). Parameter uncertainty analysis for streamflow simulation using SWAT model in Nashe watershed, Blue Nile River Basin, Ethiopia. *Applied and Environmental Soil Science*, *2022*, 1. https://doi.org/10.1155/2022/1826942

Leta, M., Demissie, T. A., & Tränckner, J. (2021). Modeling and Prediction of Land Use Land Cover Change Dynamics Based on Land Change Modeler (LCM) in Nashe Watershed, Upper Blue Nile Basin, Ethiopia. Sustainability. https://doi.org/10.3390/su13073740

Li, J., Ai, P., Xiong, C., & Song, Y. (2025). Leveraging multi-source data and teleconnection indices for enhanced runoff prediction using coupled deep learning models. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-00115-1

Li, J., Yuan, X., & Ji, P. (2023). Long-lead daily streamflow forecasting using Long Short-Term Memory model with different predictors. *Journal of Hydrology: Regional Studies*. https://doi.org/10.1016/j.ejrh.2023.101471

Liu, J., Koch, J., Stisen, S., Troldborg, L., & Schneider, R. (2024). A national-scale hybrid model for enhanced streamflow estimation – consolidating a physically based hydrological model with long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *28*(13), 2871. https://doi.org/10.5194/hess-28-2871-2024

Liu, J., Xia, J., She, D., Li, L., Wang, Q., & Zou, L. (2019). Evaluation of six satellite-based precipitation products and their ability for capturing characteristics of extreme precipitation events over a climate transition area in China. *Remote Sensing*, *11*(12), 1477. https://doi.org/10.3390/rs11121477

Liu, Q., Zou, Y., Liu, X., & Linge, N. (2019). A survey on rainfall forecasting using artificial neural network. *International Journal of Embedded Systems*, *11*(2), 240. https://doi.org/10.1504/ijes.2019.098300

Liu, S., Li, Y., Pauwels, V., & Walker, J. P. (2018). Impact of rain gauge quality control and interpolation on streamflow simulation: An application to the Warwick catchment, Australia. *Frontiers in Earth Science*, *5*. https://doi.org/10.3389/feart.2017.00114

Liu, Z. J., Liu, Y. S., Wang, S. S., Yang, X. J., Wang, L. C., Baig, M. H. A., Chi, W. F., & Wang, Z. S. (2018). Evaluation of spatial and temporal performances of ERA-Interim precipitation and temperature in Mainland China. *Journal of Climate*, *31*(11), 4347–4365. https://doi.org/10.1175/JCLI-D-17-0212.1

Loukika, K. N., Reddy, K. V., Buri, E. S., & Sridhar, V. (2022). Predicting the effects of land use land cover and climate change on Munneru River Basin using CA-Markov and Soil and Water Assessment Tool. *Sustainability*, *14*(9), 5000. https://doi.org/10.3390/su14095000

Luo, C., Ding, W., Zhang, C., & Yang, X. (2023). Exploiting multiple hydrologic forecasts to inform real-time reservoir operation for drought mitigation. *Journal of Hydrology*, *618*, 129232. https://doi.org/10.1016/j.jhydrol.2023.129232

Luo, H., Ge, F., Yang, K., Zhu, S., Peng, T., Cai, W., Liu, X., & Tang, W. (2019). Assessment of ECMWF reanalysis data in complex terrain: Can the CERA-20C

and ERA-Interim datasets replicate the variation in surface air temperatures over Sichuan, China? *International Journal of Climatology*, *39*, 5619–5634. https://doi.org/10.1002/joc.6175

Ma, K., Feng, D., Lawson, K., Tsai, W., Liang, C., Huang, X., Sharma, A., & Shen, C. (2020). Transferring hydrologic data across continents – leveraging US data to improve hydrologic prediction in other countries. *ESS Open Archive*. https://doi.org/10.1002/essoar.10504132.1

Ma, K., Shen, C., Xu, Z., & He, D. (2024). Transfer learning framework for streamflow prediction in large-scale transboundary catchments: Sensitivity analysis and applicability in data-scarce basins. *Journal of Geographical Sciences*, *34*(5), 963. https://doi.org/10.1007/s11442-024-2235-x

Ma, Y., Sun, X., Chen, H., Hong, Y., & Zhang, Y. (2021). A two-stage blending approach for merging multiple satellite precipitation estimates and rain gauge observations: An experiment in the northeastern Tibetan Plateau. *Hydrology and Earth System Sciences*, *25*, 359–374. https://doi.org/10.5194/hess-25-359-2021

Madhu, B., Rahman, M. A., Mukherjee, A., Islam, M. Z., Roy, R., & Ali, L. E. (2021). A comparative study of support vector machine and artificial neural network for option price prediction. *Journal of Computer and Communications*, *09*(05), 78–91. https://doi.org/10.4236/jcc.2021.95006

Mai, Y., Sheng, Z., Shi, H., & Liao, Q. (2021). Using improved XGBoost algorithm to obtain modified atmospheric refractive index. *International Journal of Antennas and Propagation*, *2021*, 1–11. https://doi.org/10.1155/2021/5506599

Malede, D. A., Agumassie, T. A., Kosgei, J. R., Andualem, T. G., & Diallo, I. (2022). Recent approaches to climate change impacts on hydrological extremes in the upper Blue Nile basin, Ethiopia. *Earth Systems and Environment*, *6*(3), 669–679. https://doi.org/10.1007/s41748-021-00287-6

Manrique-Suñen, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N., & Doblas-Reyes, F. J. (2020). Choices in the verification of S2S forecasts and their implications for

climate services. *Monthly Weather Review*, *148*, 3995–4008. https://doi.org/10.1175/MWR-D-20-0067.1

Mastrantonas, N., Bhattacharya, B., Shibuo, Y., Rasmy, M., Espinoza-Dávalos, G., & Solomatine, D. (2019). Evaluating the benefits of merging near-real-time satellite precipitation products: A case study in the Kinu Basin Region, Japan. *Journal of Hydrometeorology*, *20*(6), 1213–1233. https://doi.org/10.1175/JHM-D-18-0190.1

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology* (pp. 179–184). American Meteorological Society.

Mehr, A. D., Nourani, V., Khosrowshahi, V. K., & Ghorbani, M. A. (2019). A hybrid support vector regression-firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology*, *16*(1), 335–346.

Mei, Z., Peng, T., Chen, L., Singh, V., Yi, B., Leng, Z., Gan, X., & Xie, T. (2024). Coupling SWAT and LSTM for improving daily streamflow simulation in a humid and semi-humid river basin. *Water Resources Management*, *39*, 397–418. https://doi.org/10.1007/s11269-024-03975-w

Mekonnen, D. F., Duan, Z., Rientjes, T. H. M., & Disse, M. (2018). Analysis of combined and isolated effects of land-use and land-cover changes and climate change on the upper Blue Nile River basin's streamflow. Hydrology and Earth System Sciences, 22(12), 6187–6207. https://doi.org/10.5194/hess-22-6187-2018

Mekonnen, D. F., Duan, Z., Rientjes, T. H. M., & Disse, M. (2018). Analysis of combined and isolated effects of land-use and land-cover changes and climate change on the upper Blue Nile River basin's streamflow. Hydrology and Earth System Sciences, 22(12), 6187–6207. https://doi.org/10.5194/hess-22-6187-2018

Mekonnen, D. F., Duan, Z., Rientjes, T. H. M., & Disse, M. (2018). Analysis of combined and isolated effects of land-use and land-cover changes and climate change on the

upper Blue Nile River basin's streamflow. Hydrology and Earth System Sciences, 22(12), 6187–6207. https://doi.org/10.5194/hess-22-6187-2018

Mendoza, P., Rajagopalan, B., Clark, M., Cortés, G., & McPhee, A. (2014). A robust multi-model framework for ensemble seasonal hydro-climatic forecasts. *Water Resources Research*, *50*, 6030.

Meng, F., Sa, C., Liu, T., Luo, M., Liu, J., & Tian, L. (2020). Improved model parameter transferability method for hydrological simulation with SWAT in ungauged mountainous catchments. *Sustainability*, *12*(9), 3551. https://doi.org/10.3390/su12093551

Mengistu, D. A., Bewket, W., Dosio, A., & Panitz, H.-J. (2020). Climate change impacts on water resources in the Upper Blue Nile (Abay) River Basin, Ethiopia. *Journal of Hydrology*, *592*, 125614. https://doi.org/10.1016/j.jhydrol.2020.125614

Min, X., Hao, B., Sheng, Y., Huang, Y., & Qin, J. (2023). Transfer performance of gated recurrent unit model for runoff prediction based on the comprehensive spatiotemporal similarity of catchments. *Journal of Environmental Management*, *330*, 117182. https://doi.org/10.1016/j.jenvman.2022.117182

Moccia, B., Mineo, C., Ridolfi, E., Russo, F., & Napolitano, F. (2022). SPI-based drought classification in Italy: Influence of different probability distribution functions. *Water*, *14*, 3668. https://doi.org/10.3390/w14223668

Modaresi, F., Danandeh Mehr, A., & Kazemi Choolanak, A. (2024). SMRF: A new class-based probabilistic approach for season-ahead monthly rainfall forecasting. *Acta Geophysica*, *5*, 1–15.

Mohamed, M. A., El Afandi, G. S., & El-Mahdy, M. E. S. (2022). Impact of climate change on rainfall variability in the Blue Nile basin. *Alexandria Engineering Journal*, *61*(4), 3265–3275. https://doi.org/10.1016/j.aej.2021.08.056

Mohammad-Taghi, S., Ali, R. J., & Andrew, K. (2017). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, *48*(4), 1032–1044. https://doi.org/10.2166/nh.2016.364

Mohammed, J., Mengiste, Y., & Singh, V. (2023). Improving spatio-temporal precipitation estimates in data scarce river basins: An application of machine learning-based multi-source data merging. *Stochastic Environmental Research and Risk Assessment*, *37*, 1353–1369. https://doi.org/10.1007/s00477-022-02346-4

Mohanty, A., Sahoo, B., & Kale, R. V. (2024). A hybrid model enhancing streamflow forecasts in paddy land use-dominated catchments with numerical weather prediction model-based meteorological forcings. *Journal of Hydrology*, *635*, 131225. https://doi.org/10.1016/j.jhydrol.2024.131225

Mokria, M., Gebrekirstos, A., Abiyu, A., & Bräuning, A. (2018). Upper Nile River flow reconstructed to A.D. 1784 from tree-rings for a long-term perspective on hydrologic-extremes and effective water resource management. Quaternary Science Reviews. https://doi.org/10.1016/j.quascirev.2018.09.011

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). Support vector machines and support vector regression. In *Multivariate statistical machine learning methods for genomic prediction* (pp. 337–378). Springer. https://doi.org/10.1007/978-3-030-89010-0_9

Moraux, A., Dewitte, S., Cornelis, B., & Munteanu, A. (2021). A deep learning multimodal method for precipitation estimation. *Remote Sensing*, *13*(16), 3278. https://doi.org/10.3390/rs13163278

Moreno-Pájaro, Á. M., Osorio-Gastelbondo, A., Moreno-Egel, D. A., Coronado-Hernández, Ó. E., Narváez-Cuadro, M. A., Saba, M., & Arrieta-Pastrana, A. (2025). Hydrological modelling and remote sensing for assessing the impact of vegetation cover changes. *Hydrology*, *12*(5), 107. https://doi.org/10.3390/hydrology12050107

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*, 885–900. https://doi.org/10.13031/2013.23153

Muhammad, A. U., & Abba, S. (2023). Transfer learning for streamflow forecasting using ungauged MOPEX basins data set. *Earth Science Informatics*, 1–24.

Nagendra, K., & Ussenaiah, M. (2018). A comparative study on the performance of support vector machines with artificial neural networks in rainfall forecasting. *JCRT*, *6*, 8.

Narejo, S., Jawaid, M. M., Talpur, S., Baloch, R., & Pasero, E. G. A. (2021). Multi-step rainfall forecasting using deep learning approach. *PeerJ Computer Science*, *7*, e514. https://doi.org/10.7717/peerj-cs.514

Navarro, A., García-Ortega, E., Merino, A., & Sánchez, J. L. (2020). Extreme events of precipitation over complex terrain derived from satellite data for climate applications: An evaluation of the southern slopes of the Pyrenees. *Remote Sensing*, *12*(13), 2171. https://doi.org/10.3390/rs12132171

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T., Weitzner, D., & Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559. https://doi.org/10.1038/s41586-024-07145-1

Nerantzaki, S. D., Hristopulos, D. T., & Nikolaidis, N. P. (2020). Estimation of the uncertainty of hydrologic predictions in a karstic Mediterranean watershed. *Science of the Total Environment*, *717*, 137131. https://doi.org/10.1016/j.scitotenv.2020.137131

Nguyen, G. V., Le, X. H., Van, L. N., Jung, S., Yeon, M., & Lee, G. (2021). Application of random forest algorithm for merging multiple satellite precipitation products across South Korea. *Remote Sensing*, *13*(20), 4033. https://doi.org/10.3390/rs13204033

Nguyen, P., Ombadi, M., Sorooshian, S., Hsu, K., AghaKouchak, A., Braithwaite, D. A., Shouri, H., & Thorstensen, A. R. (2018). The PERSIANN family of global satellite precipitation data: A review and evaluation of products. *Hydrology and Earth System Sciences*, *22*, 5801–5816. https://doi.org/10.5194/hess-22-5801-2018

Nigussie, A. B., Tenfie, H. W., Zimale, F. A., Endalew, A., & Wudiye, G. (2022). Evaluation of multiple bias correction methods with different satellite rainfall products in the Main Beles Watershed, Upper Blue Nile (Abbay) Basin, Ethiopia. Journal of Water and Climate Change, 14(1), 156–174. https://doi.org/10.2166/wcc.2022.244

Nigussie, G., Moges, M. A., Moges, M. M., & Steenhuis, T. S. (2019). Assessment of suitable land for surface irrigation in ungauged catchments: Blue Nile Basin, Ethiopia. *Water*, *11*(7), 1465. https://doi.org/10.3390/w11071465

Nourani, V., Uzelaltinbulat, S., Sadikoglu, F., & Behfar, N. (2019). Artificial intelligence-based ensemble modeling for multi-station prediction of precipitation. *Atmosphere*. https://doi.org/10.3390/atmos10020080

Nurhamidah, N., Hanwar, S., Junaidi, A. B., Februarman, F., & Sunaryo, S. (2021). Effect of missing rainfall data on the uncertainty of design floods. In *IOP Conference Series: Materials Science and Engineering* (Vol. 12007). IOP Publishing.

Oruche, R., Egede, L., Baker, T. L., & O'Donncha, F. (2021). Transfer learning to improve streamflow forecasts in data scarce regions. *arXiv*. https://doi.org/10.48550/arxiv.2112.03088

Ougahi, J. H., & Rowan, J. S. (2025). Enhanced streamflow forecasting using hybrid modelling integrating glacio-hydrological outputs, deep learning and wavelet transformation. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-87187-1

Ouma, Y. O., Cheruyot, R., & Wachera, A. N. (2022). Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: Case study of Nzoia hydrologic basin. *Complex & Intelligent Systems*, *8*(1), 213–236. https://doi.org/10.1007/s40747-021-00365-2

Özdoğan-Sarıkoç, G., & Dadaşer-Çelik, F. (2024). Physically based vs. data-driven models for streamflow and reservoir volume prediction at a data-scarce semi-arid

basin. *Environmental Science and Pollution Research*, *31*(27), 39098. https://doi.org/10.1007/s11356-024-33732-w

Pagano, T. C., Wood, A. W., Ramos, M., Cloke, H., Pappenberger, F., Clark, M., Cranston, M., Kavetski, D., Mathevet, T., Sorooshian, S., & Verkade, J. (2014). Challenges of operational river forecasting. *Journal of Hydrometeorology*, *15*(4), 1692. https://doi.org/10.1175/jhm-d-13-0188.1

Panthi, J., Talchabhadel, R., Ghimire, G. R., Sharma, S., Dahal, P., Baniya, R., Boving, T. B., Pradhanang, S. M., & Parajuli, B. (2021). Hydrologic regionalization under data scarcity: Implications for streamflow prediction. *Journal of Hydrologic Engineering*, *26*(9). https://doi.org/10.1061/(asce)he.1943-5584.0002121

Pathak, J., Wikner, A., Fussell, R. K., Chandra, S., Hunt, B. R., Girvan, M., & Ott, E. (2018). Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(4). https://doi.org/10.1063/1.5028373

Pechlivanidis, I., & Du, Y. (2024). Hybrid approaches enhance hydrological model usability for local streamflow prediction. *Research Square*. https://doi.org/10.21203/rs.3.rs-5142634/v1

Peña, M., Vázquez-Patiño, A., Zhiña, D., Montenegro, M., & Avilés, A. (2020). Improved rainfall prediction through nonlinear autoregressive network with exogenous variables: A case study in Andes high mountain region. *Advances in Meteorology*, *2020*, 1–17. https://doi.org/10.1155/2020/1828319

Pham, L. T., Luo, L., & Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. *Hydrology and Earth System Sciences*, *25*(6), 2997–3015. https://doi.org/10.5194/hess-25-2997-2021

Polanco, E. I., Fleifle, A., Ludwig, R., & Disse, M. (2017). Improving SWAT model performance in the upper Blue Nile Basin using meteorological data integration and sub-catchment discretization. *Hydrology and Earth System Sciences*, *21*, 4907–4926. https://doi.org/10.5194/hess-21-4907-2017

Pool, S., & Seibert, J. (2021). Gauging ungauged catchments – Active learning for the timing of point discharge observations in combination with continuous water level measurements. *Journal of Hydrology*, *598*, 126448. https://doi.org/10.1016/j.jhydrol.2021.126448

Poornima, S., & Pushpalatha, M. (2019). Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. *Atmosphere*, *10*(11), 668. https://doi.org/10.3390/atmos10110668

Prakash, S., Seshadri, A., Srinivasan, J., & Pai, D. S. (2019). A new parameter to assess impact of rain gauge density on uncertainty in the estimate of monthly rainfall over India. *Journal of Hydrometeorology*, *20*(5), 821–832. https://doi.org/10.1175/JHM-D-18-0161.1

Qi, S., He, M., Hoang, R., Zhou, Y., Namadi, P., Tom, B., Sandhu, P., Bai, Z., Chung, F. I., Ding, Z., Anderson, J., Roh, D. M., & Huynh, V. (2023). Salinity modeling using deep learning with data augmentation and transfer learning. *Water*, *15*(13), 2482. https://doi.org/10.3390/w15132482

Qinghe, Z., Wen, X., Boyan, H., Jong, W., & Junlong, F. (2022). Optimised extreme gradient boosting model for short-term electric load demand forecasting of regional grid system. *Scientific Reports*. https://doi.org/10.1038/s41598-022-22024-3

Rahman, A. U., Abbas, S., Gollapalli, M., et al. (2022). Rainfall prediction system using machine learning fusion for smart cities. *Sensors*, *22*(9), 3504. https://doi.org/10.3390/s22093504

Rahmawati, N., & Lubczynski, M. W. (2017). Validation of satellite daily rainfall estimates in complex terrain of Bali Island, Indonesia. *Theoretical and Applied Climatology*, *134*, 513–532. https://doi.org/10.1007/s00704-017-2290-7

Raja, O., Parsinejad, M., & Tajrishy, M. (2024). Evaluation of coupled SWAT-MODFLOW-NWT model for conjunctive use of surface water and groundwater resources in the Mahabad plain of Iran. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-83892-5

Ramachandran, A., Mousa, H. Y. S., Maier, A., & Bayer, S. (2024). Week-ahead water demand forecasting using convolutional neural network on multi-channel wavelet scalogram. *Engineering Proceedings*, *51*, 179. https://doi.org/10.3390/engproc2024069179

Ramón, A., Torres, A. M., Milara, J., et al. (2022). Extreme gradient boosting-based method to classify patients with COVID-19. *Journal of Investigative Medicine*, *70*, 1472–1480.

Razavi, T., & Coulibaly, P. (2012). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering*, *18*(8), 958. https://doi.org/10.1061/(asce)he.1943-5584.0000690

Ren, C., Chang, J., Wang, X., Niu, C., Wang, L., & Zhang, J. (2025). Coupled SWAT, stationary wavelet transform, and interpretable machine learning to improve watershed streamflow simulation. *Water Resources Management*, *39*, 3483–3498. https://doi.org/10.1007/s11269-025-04117-6

Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1651–1663. https://doi.org/10.1016/j.asej.2020.09.011

Rientjes, T. H. M., Perera, B. U. J., Haile, A. T., Reggiani, P., & Muthuwatta, L. (2011). Regionalisation for lake level simulation – the case of Lake Tana in the Upper Blue Nile, Ethiopia. Hydrology and Earth System Sciences, 15(4), 1167–1183. https://doi.org/10.5194/hess-15-1167-2011

Rong, Z., Sun, W., Xie, Y., Huang, Z., & Chen, X. (2025). Mixture of experts leveraging informer and LSTM variants for enhanced daily streamflow forecasting. *Journal of Hydrology*, 132737. https://doi.org/10.1016/j.jhydrol.2025.132737

Roth, V., Lemann, T., Zeleke, G., Subhatu, A. T., Nigussie, T., & Hurni, H. (2018). Effects of climate change on water resources in the upper Blue Nile Basin of Ethiopia. *Heliyon*, *4*(9), e00771. https://doi.org/10.1016/j.heliyon.2018.e00771

Saber, M., Saeed, G., Yang, H., Chen, S., & Reza, K. M. (2016). Comprehensive evaluation of four high-resolution satellite precipitation products under diverse climate conditions in Iran. *Hydrological Sciences Journal*, *61*(2), 420–440. https://doi.org/10.1080/02626667.2014.987675

Salaeh, N., Ditthakit, P., Pinthong, S., et al. (2022). Long-short term memory technique for monthly rainfall prediction in Thale Sap Songkhla river basin, Thailand. *Symmetry*, *14*(8), 1599. https://doi.org/10.3390/sym14081599

Samantaray, S., Tripathy, O., Sahoo, A., & Ghose, D. K. (2020). Rainfall forecasting through ANN and SVM in Bolangir watershed, India. In *Smart intelligent computing and applications* (Smart Innovation, Systems and Technologies, pp. 767–774). Springer. https://doi.org/10.1007/978-981-13-9282-5_74

Samy, A., Ibrahim, M., Mahmod, W., Fujii, M., Eltawil, A., & Daoud, W. (2019). Statistical assessment of rainfall characteristics in Upper Blue Nile Basin over the period from 1953 to 2014. *Water*. https://doi.org/10.3390/W11030468

Sarasa-Cabezuelo, A. (2022). Prediction of rainfall in Australia using machine learning. *Information*, *13*(4), 163. https://doi.org/10.3390/info13040163

Saravanan, S., Reddy, N. M., Pham, Q. B., Alodah, A., Abdo, H. G., Almohamad, H., & Dughairi, A. A. A. (2023). Machine learning approaches for streamflow modeling in the Godavari Basin with CMIP6 dataset. *Sustainability*, *15*(16), 12295. https://doi.org/10.3390/su151612295

Satgé, F., Hussain, Y., Bonnet, M. P., Hussain, B., Martinez-Carvajal, H., Akhter, G., & Uagoda, R. (2018). Benefits of the successive GPM based satellite precipitation estimates IMERG–V03, –V04, –V05 and GSMaP–V06, –V07 over diverse geomorphic and meteorological regions of Pakistan. *Remote Sensing*, *10*(9), 1373. https://doi.org/10.3390/rs10091373

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., van Beek, L. P., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National*

Academy of Sciences USA, *115*, E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Schmidt, L., Heße, F., Attinger, S., & Kumar, R. (2020). Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research*, *56*(5). https://doi.org/10.1029/2019wr025924

Sharma, S., Khadka, N., Hamal, K., Shrestha, D., Talchabhadel, R., & Chen, Y. (2020). How accurately can satellite products (TMPA and IMERG) detect precipitation patterns, extremities, and drought across the Nepalese Himalaya? *Journal of Earth and Space Science*, *7*, e2020EA001315. https://doi.org/10.1029/2020EA001315

Sheffield, J., Wood, F., Pan, M., Beck, H., Serrat-Capdevila, A., & Verbist, K. (2018). Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions. *Water Resources Research*, *54*, 9724–9758.

Shiferaw, H., Girma, A., Hadush, K., Mariam, H. G., Yazew, E., & Zenebe, A. (2023). Long-term hydroclimatic variability over the semi-arid Ethiopian highlands in relation to ENSO and IOD teleconnection signals. *Theoretical and Applied Climatology*, *153*, 193. https://doi.org/10.1007/s00704-023-04450-z

Sivapragasam, C., Arun, V. M., & Muttil, N. (2011). Long term flow forecasting for water resources planning in a river basin. In *19th International Congress on Modelling and Simulation*, Perth, Australia, 12–16 December.

Slater, L., Arnal, L., Boucher, M., Chang, A. Y.-Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A. W., & Zappa, M. (2023). Hybrid forecasting: Blending climate predictions with AI models. *Hydrology and Earth System Sciences*, *27*(9), 1865. https://doi.org/10.5194/hess-27-1865-2023

Sohrabi, S., Brissette, F., & Arsenault, R. (2021). Coupling large-scale climate indices with a stochastic weather generator to improve long-term streamflow forecasts in a

Canadian watershed. *Journal of Hydrology*, *594*, 125925. https://doi.org/10.1016/j.jhydrol.2020.125925

Solanki, H., Vegad, U., Kushwaha, A. P., & Mishra, V. (2025). Improving streamflow prediction using multiple hydrological models and machine learning methods. *Water Resources Research*, *61*(1). https://doi.org/10.1029/2024wr038192

Srivastava, A., Deb, P., & Kumari, N. (2019). Multi-model approach to assess the dynamics of hydrologic components in a tropical ecosystem. *Water Resources Management*, *34*(1), 327. https://doi.org/10.1007/s11269-019-02452-z

Sun, E., Cui, Y., Liu, P., & Yan, J. (2025). A decade of deep learning for remote sensing spatiotemporal fusion: Advances, challenges, and opportunities. *arXiv*. https://doi.org/10.48550/ARXIV.2504.00901

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. (2018). A review of global precipitation datasets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, *56*, 79–107. https://doi.org/10.1002/2017RG000574

Szczepanek, R. (2022). Daily streamflow forecasting in the mountainous catchment using XGBoost, LightGBM, and CatBoost. *Hydrology*, *9*, 12. https://doi.org/10.3390/hydrology9120226

Takele, G. S., Gebre, G. S., Gebremariam, A. G., & Engida, A. N. (2021). Hydrological modeling in the Upper Blue Nile basin using soil and water analysis tool (SWAT). *Modeling Earth Systems and Environment*, *8*, 277–292. https://doi.org/10.1007/s40808-021-01085-9

Takele, G. S., Gebrie, G. S., Gebremariam, A. G., & Engida, A. N. (2022). Future climate change and impacts on water resources in the Upper Blue Nile basin. *Journal of Water and Climate Change*, *13*(2), 908–925.

Tan, J., Huffman, G. J., Bolvin, D. T., & Nelkin, E. J. (2019). IMERG V06: Changes to the morphing algorithm. *Journal of Atmospheric and Oceanic Technology*, *36*(12), 2471–2482.

Tang, Q., Zhang, X., Duan, Q., Huang, S., Yuan, X., Cui, H., Li, Z., & Liu, X. (2016). Hydrological monitoring and seasonal forecasting: Progress and perspectives. *Journal of Geographical Sciences*, *26*(7), 904. https://doi.org/10.1007/s11442-016-1306-z

Tao, J., Gu, Y., Yin, X., Chen, J., Ao, T., & Zhang, J. (2024). Coupling SWAT and transformer models for enhanced monthly streamflow prediction. *Sustainability*, *16*(19), 8699.

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, *24*, 2527–2544. https://doi.org/10.5194/hess-24-2527-2020

Taye, M., Sahlu, D., Zaitchik, B. F., & Neka, M. (2020). Evaluation of satellite rainfall estimates for meteorological drought analysis over the Upper Blue Nile Basin, Ethiopia. *Geosciences*, *10*(9), 352. https://doi.org/10.3390/geosciences10090352

Tekleab, S., Mohamed, Y., & Uhlenbrook, S. (2013). Hydro-climatic trends in the Abay/Upper Blue Nile basin, Ethiopia. Physics and Chemistry of The Earth, 61, 32-42. https://doi.org/10.1016/j.pce.2013.04.017

Teshome, D. S., Leta, M. K., Taddese, H., Moshe, A., Tolessa, T., Ayele, G. T., & You, S. (2023). Watershed hydrological responses to land cover changes at Muger watershed, Upper Blue Nile River Basin, Ethiopia. *Water*, *15*(14), 2533. https://doi.org/10.3390/w15142533

Tewabe, D., Dessie, M., Asmamaw, D. K., Tamiru, E., Adgo, E., Nyssen, J., Walraevens, K., & Cornelis, W. (2021). Comparative analysis of groundwater conditions on rain-fed and irrigated agriculture in the upper Blue Nile basin, Ethiopia. Journal of Hydrology: Regional Studies. https://doi.org/10.1016/j.ejrh.2021.100916

Tharun, V. P., Prakash, R., & Devi, S. R. (2018). Prediction of rainfall using data mining techniques. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. IEEE. https://doi.org/10.1109/icicct.2018.8473177

Thieu, N. V., Nguyen, N. H., Sherif, M., El-Shafie, A., & Ahmed, A. N. (2024). Integrated metaheuristic algorithms with extreme learning machine models for river streamflow prediction. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-63908-w

Tin, T. T., Sheng, E. H. C., Xian, L. S., Yee, L. P., & Kit, Y. S. (2024). Machine learning classification of rainfall forecasts using Austin weather data. *International Journal of Innovative Research in Science Studies*, *7*(2), 727–741. https://doi.org/10.53894/ijirss.v7i2.2881

Tofu, D. (2025). Prediction of future climate variability and its implications for rainfed agriculture in the Guder watershed Ethiopia. *Scientific African*. https://doi.org/10.1016/j.sciaf.2025.e02531

Tongal, H., & Booij, M. (2018). Simulation and forecasting of stream flows using machine learning models coupled with base flow separation. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2018.07.004

Torre, A. M. la, Blyth, E., & Robinson, E. L. (2019). Evaluation of drydown processes in global land surface and hydrological models using flux tower evapotranspiration. *Water*, *11*(2), 356. https://doi.org/10.3390/w11020356

Trinh-Tuan, L., Matsumoto, J., Ngo-Duc, T., et al. (2019). Evaluation of satellite precipitation products over Central Vietnam. *Progress in Earth and Planetary Science*, *6*, 54. https://doi.org/10.1186/s40645-019-0297-7

Tsvetkova, O., & Randhir, T. O. (2019). Spatial and temporal uncertainty in climatic impacts on watershed systems. *Science of the Total Environment*, *687*, 618. https://doi.org/10.1016/j.scitotenv.2019.06.141

Tuysuzoglu, G., Birant, K. U., & Birant, D. (2023). Rainfall prediction using an ensemble machine learning model based on K-stars. *Sustainability*, *15*(7), 5889. https://doi.org/10.3390/su15075889

Ullah, B., Fawad, M., Khan, A. U., Mohamand, S. K., Khan, M., Iqbal, M. J., & Khan, J. (2023). Futuristic streamflow prediction based on CMIP6 scenarios using machine learning models. *Water Resources Management*, *37*(15), 6089. https://doi.org/10.1007/s11269-023-03645-3

Ullah, W., Wang, G., Ali, G., Tawia Hagan, D., Bhatti, A., & Lou, D. (2019). Comparing multiple precipitation products against in-situ observations over different climate regions of Pakistan. *Remote Sensing*, *11*(6), 628. https://doi.org/10.3390/rs11060628

Unnikrishnan, P., & Jothiprakash, V. (2018). Daily rainfall forecasting for one year in a single run using singular spectrum analysis. *Journal of Hydrology*, *561*, 609–621. https://doi.org/10.1016/j.jhydrol.2018.04.032

Upadhyay, H., Mittal, H., Singh, M., & Hirapara, J. G. (2019). Application of SWAT model for estimation of runoff in Pindwara watershed and assessment of its feasibility. *International Journal of Current Microbiology and Applied Sciences*, *8*(6), 3056. https://doi.org/10.20546/ijcmas.2019.806.364

Valipour, M. (2016). How much meteorological information is necessary to achieve reliable accuracy for rainfall estimations? *Agriculture*, *6*(4), 53. https://doi.org/10.3390/agriculture6040053

Vallam, P., & Qin, X. S. (2016). Multi-site rainfall simulation at tropical regions: A comparison of three types of generators. *Meteorological Applications*, *23*(3), 425–437. https://doi.org/10.1002/met.1567

Van Der Kwast, H., Hartanto, I., & Van Andel, S. J. (2013). *Integrating in-situ measurements and remote sensing data in process-based hydrological modeling*. VUB, Brussels, Belgium.

Vişescu, M., Beilicci, E., & Beilicci, R. (2017). Integrated hydrographical basin management. Study case – Crasna River Basin. *IOP Conference Series: Materials Science and Engineering*, *245*, 32038. https://doi.org/10.1088/1757-899x/245/3/032038

Vogeti, R. K., Mishra, B. R., & Raju, K. S. (2022). Machine learning algorithms for streamflow forecasting of Lower Godavari Basin. *H2Open Journal*, *5*(4), 670. https://doi.org/10.2166/h2oj.2022.240

Wahyullah, W., Hendrayanto, H., & Suharnoto, Y. (2023). Simulation of the impact of land use change on surface run-off in Karst Leang Lonrong sub-watershed. *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan*, *13*(2), 313. https://doi.org/10.29244/jpsl.13.2.313-326

Wan, S., Shah, R., Deng, Q., Sabo, J. L., Liu, H., & Selçuk, K. (2024). Spatio-temporal causal learning for streamflow forecasting. *arXiv*. https://doi.org/10.48550/arxiv.2411.17937

Wang, L., Wang, Z., Yu, J., Zhang, Y., & Dang, S. (2018). Hydrological process simulation of inland river watershed: A case study of the Heihe River Basin with multiple hydrological models. *Water*, *10*(4), 421. https://doi.org/10.3390/w10040421

Wang, W., Chen, L., & Xu, C. (2021). Hydrological modeling in water cycle processes. *Water*, *13*(14), 1882. https://doi.org/10.3390/w13141882

Wang, Z., Xu, N., Bao, X., Wu, J., & Cui, X. (2024). Spatio-temporal deep learning model for accurate streamflow prediction with multi-source data fusion. *Environmental Modelling & Software*, *178*, 106091. https://doi.org/10.1016/j.envsoft.2024.106091

Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, *13*, 102946. https://doi.org/10.1016/j.mex.2024.102946

Waqas, M., Humphries, U. W., Wangwongchai, A., Dechpichai, P., & Ahmad, S. (2023). Potential of artificial intelligence-based techniques for rainfall forecasting in

Thailand: A comprehensive review. *Water*, *15*(16), 2979. https://doi.org/10.3390/w15162979

Wedajo, G. K., Muleta, M. K., & Awoke, B. G. (2021). Performance evaluation of multiple satellite rainfall products for Dhidhessa River Basin (DRB), Ethiopia. Atmospheric Measurement Techniques, 14(3), 2299–2316. https://doi.org/10.5194/amt-14-2299-2021

Weerakody, P. B., Wong, K. W., Wang, G., & Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, *441*, 161–178. https://doi.org/10.1016/j.neucom.2021.02.046

Wei, C. C., & Hsu, C. C. (2020). Extreme gradient boosting model for rain retrieval using radar reflectivity from various elevation angles. *Remote Sensing*, *12*(14), 2203. https://doi.org/10.3390/rs12142203

Weslati, O., Bouaziz, M., & Serbaji, M. M. (2023). Precipitation forecasting and monitoring in degraded land: A study case in Zaghouan. *Land*, *12*(4), 738. https://doi.org/10.3390/land12040738

Wheeler, K., Jeuland, M., Hall, J. W., Zagona, E., & Whittington, D. (2020). Understanding and managing new risks on the Nile with the Grand Ethiopian Renaissance Dam. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-19089-x

Willard, J., Varadharajan, C., Jia, X., & Kumar, V. (2023). Time series predictions in unmonitored sites: A survey of machine learning techniques in water resources. *arXiv*. https://doi.org/10.48550/arxiv.2308.09766

Wiwoho, B. S., Astuti, I. S., Alfarizi, I. A. G., & Sucahyo, H. R. (2021). Validation of three daily satellite rainfall products in a humid tropic watershed, Brantas, Indonesia: Implications to land characteristics and hydrological modelling. *Hydrology*, *8*(4), 154. https://doi.org/10.3390/hydrology8040154

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., & Berne, A. (2021). RainForest: A random forest algorithm for quantitative precipitation estimation

over Switzerland. *Atmospheric Measurement Techniques*, *14*, 3169–3193. https://doi.org/10.5194/amt-14-3169-2021

World Meteorological Organization. (2012). *Standardized Precipitation Index user guide* (M. Svoboda, M. Hayes, & D. Wood, Eds.; WMO-No. 1090).

Wu, H., Yang, Q., Liu, J., & Wang, G. (2020). A spatiotemporal deep fusion model for merging satellite and gauge precipitation in China. *Journal of Hydrology*, *584*, 124664. https://doi.org/10.1016/j.jhydrol.2020.124664

Xiao, Y., Wang, J., & Zhou, J. (2025). Hydrosphere under the driving of human activity and climate change: Status, evolution, and strategies. *Sustainability*, *17*(7), 3257. https://doi.org/10.3390/su17073257

Xiong, Z., Sang, J., Sun, X., Zhang, B., & Li, J. (2020). Comparisons of performance using data assimilation and data fusion approaches in acquiring precipitable water vapor: A case study of a western United States of America area. *Water*, *12*(10), 2943. https://doi.org/10.3390/w12102943

Xu, L., Chen, N., Moradkhani, H., Zhang, X., & Hu, C. (2020). Improving global monthly and daily precipitation estimation by fusing gauge observations, remote sensing, and reanalysis datasets. *Water Resources Research*, *56*, e2019WR026444. https://doi.org/10.1029/2019WR026444

Xu, Q., Shi, Y., Bamber, J. L., Ouyang, C., & Zhu, X. X. (2024). Large-scale flood modeling and forecasting with FloodCast. *Water Research*, *264*, 122162. https://doi.org/10.1016/j.watres.2024.122162

Xu, W., Jiang, Y., Zhang, X., Li, Y., Zhang, R., & Fu, G. (2020). Using long short-term memory networks for river flow prediction. *Hydrology Research*, *51*(6), 1358. https://doi.org/10.2166/nh.2020.026

Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., & Ran, G. (2023). Deep transfer learning based on transformer for flood forecasting in data-scarce basins. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2023.129956

Yang, S., Tan, M., Song, Q., He, J., Yao, N., Li, X., & Yang, X. (2023). Coupling SWAT and Bi-LSTM for improving daily-scale hydro-climatic simulation and climate change impact assessment in a tropical river basin. *Journal of Environmental Management*, *330*, 117244. https://doi.org/10.1016/j.jenvman.2023.117244

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lü, W., & Zhao, B. (2020). A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. *Journal of Hydrology*, *590*, 125206. https://doi.org/10.1016/j.jhydrol.2020.125206

Yao, J. Q., Chen, Y. N., & Chen, J. (2020). Intensification of extreme precipitation in arid Central Asia. *Journal of Hydrology*, *598*(12), 125760. https://doi.org/10.1016/j.jhydrol.2020.125760

Yao, Y., Zhao, Y., Li, X., Feng, D., Shen, C., Liu, C., Kuang, X., & Zheng, C. (2023). Can transfer learning improve hydrological predictions in the alpine regions? *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2023.130038

Yaseen, Z., Ebtehaj, I., Kim, S., et al. (2019). Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. *Water*, *11*(3), 502. https://doi.org/10.3390/w11030502

Yazdandoost, F., Moradian, S., Izadi, A., & Bavani, A. M. (2020). A framework for developing a spatial high resolution daily precipitation dataset over a data-scarce region. *Heliyon*, *6*(9), e05091. https://doi.org/10.1016/j.heliyon.2020.e05091

Yenehun, A., Dessie, M., Azeze, M., Nigate, F., Belay, A. S., Nyssen, J., Adgo, E., Van Griensven, A., Van Camp, M., & Walraevens, K. (2021). Water Resources Studies in Headwaters of the Blue Nile Basin: A Review with Emphasis on Lake Water Balance and Hydrogeological Characterization. Water, 13, 1469. https://doi.org/10.3390/w13111469

Yihun, D., Sirak, T., Esayas, K. A., Solomon, G., Abeyou, W., Haimanot, K. B., Yohanese, T., Seifu, A., Prasad, D., Louise, K., & Raghavan, S. (2018). Advance in water resource research in the upper Blue Nile river basin and the way forward. *Journal of Hydrology*, *560*, 407–423.

Yin, J., Guo, S., Gu, L., Zeng, Z., Liu, D., Chen, J., et al. (2021). Blending multi-satellite, atmospheric reanalysis and gauge precipitation products to facilitate hydrological modelling. *Journal of Hydrology*, *593*, 125878. https://doi.org/10.1016/j.jhydrol.2020.125878

Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., & Amagasaki, M. (2021). Capabilities of deep learning models on learning physical relationships: Case of rainfall-runoff modeling with LSTM. *Science of the Total Environment*, *802*, 149876. https://doi.org/10.1016/j.scitotenv.2021.149876

Yuan, F., Zhang, L., Win, K., Ren, L., Zhao, C., Zhu, Y., Jiang, S., & Liu, Y. (2017). Assessment of GPM and TRMM multi-satellite precipitation products in streamflow simulations in a data-scarce mountainous watershed in Myanmar. *Remote Sensing*, *9*(3), 302. https://doi.org/10.3390/rs9030302

Yuhao, Z., Pan, J., & Shao, G. (2025). A comparative study of a two-dimensional slope hydrodynamic model (TDSHM), long short-term memory (LSTM), and convolutional neural network (CNN) models for runoff prediction. *Water*, *17*(9), 1380. https://doi.org/10.3390/w17091380

Zaher, M., Hossam, F., & Vadahir, A. (2020). Hybridized extreme learning machine model with slap swarm algorithm: A novel predictive model for hydrological application. *Hindawi Complexity*, *2020*, 80645.

Zakaria, M. N. A., Ahmed, A. N., Malek, M. A., Birima, A. H., Khan, M. M. H., Sherif, M., & El-Shafie, A. (2023). Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia. *Heliyon*, *9*(7). https://doi.org/10.1016/j.heliyon.2023.e17689

Zambrano-Bigiarini, M., Baez-Villanueva, O. M., & Giraldo-Osorio, J. (2020). *RFmerge: Merging of satellite datasets with ground observations using random forests* (R package version 0.1-10). https://CRAN.R-project.org/package=RFmerge

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., & Ribbe, L. (2017). Temporal and spatial evaluation of satellite-based rainfall estimates across the complex

topographical and climatic gradients of Chile. *Hydrology and Earth System Sciences*, *21*, 1295–1320. https://doi.org/10.5194/hess-21-1295-2017

Zanial, W. N. C. W., Malek, M. A., & Reba, M. N. M. (2018). A review on rainfall runoff simulation at ungauged catchment. *International Journal of Engineering & Technology*, *7*, 162.

Zargar, M., Bronstert, A., Francke, T., Zimale, F. A., Worku, K. B., Wiegels, R., Lorenz, C., Hageltom, Y., Sawadogo, W., & Kunstmann, H. (2025). Comparison and hydrological evaluation of different precipitation data for a large tropical region: the Blue Nile Basin in Ethiopia. Frontiers in Water, 7. https://doi.org/10.3389/frwa.2025.1536881

Zargar, M., Bronstert, A., Francke, T., Zimale, F. A., Worku, K. B., Wiegels, R., Lorenz, C., Hageltom, Y., Sawadogo, W., & Kunstmann, H. (2025). Comparison and hydrological evaluation of different precipitation data for a large tropical region: the Blue Nile Basin in Ethiopia. Frontiers in Water, 7. https://doi.org/10.3389/frwa.2025.1536881

Zargar, M., Bronstert, A., Francke, T., Zimale, F. A., Worku, K. B., Wiegels, R., Lorenz, C., Hageltom, Y., Sawadogo, W., & Kunstmann, H. (2025). Comparison and hydrological evaluation of different precipitation data for a large tropical region: the Blue Nile Basin in Ethiopia. Frontiers in Water, 7. https://doi.org/10.3389/frwa.2025.1536881

Zargar, M., Bronstert, A., Francke, T., Zimale, F. A., Worku, K. B., Wiegels, R., Lorenz, C., Hageltom, Y., Sawadogo, W., & Kunstmann, H. (2025). Comparison and hydrological evaluation of different precipitation data for a large tropical region: the Blue Nile Basin in Ethiopia. Frontiers in Water, 7. https://doi.org/10.3389/frwa.2025.1536881

Zaroug, M., Eltahir, E. A. B., & Giorgi, F. (2014). Droughts and floods over the upper catchment of the Blue Nile and their connections to the timing of El Niño and La Niña events. *Hydrology and Earth System Sciences*, *18*(3), 1239. https://doi.org/10.5194/hess-18-1239-2014

Ze, L., Zhou, J., Yang, X., Zhao, Z., & Lv, Y. (2024). Research on water resource modeling based on machine learning technologies. *Water*, *16*(3), 472. https://doi.org/10.3390/w16030472

Zhang, J., Bonas, M., Bolster, D., Fuglstad, G., & Castruccio, S. (2023). High resolution global precipitation downscaling with latent Gaussian models and nonstationary SPDE structure. *arXiv*. https://doi.org/10.48550/arXiv.2302.03148

Zhang, L., Kim, T., Yang, T., Hong, Y., & Zhu, Q. (2021). Evaluation of sub-seasonal-to-seasonal (S2S) precipitation forecast from the North American multi-model ensemble phase II (NMME-2) over the contiguous US. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2021.127058

Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., & Ge, Y. (2021). Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology*, *594*, 125969. https://doi.org/10.1016/j.jhydrol.2021.125969

Zhang, M., Yao, T., Gu, H., Wang, W., L, P., Gu, H., Pei, Y., & Lu, B. (2025). A Hybrid Runoff Forecasting Framework Integrating Hydrological Physics and Data-Driven Models. Sustainability, 17(24), 11120–11120. https://doi.org/10.3390/su172411120

Zhang, S., Zhang, D., Huang, G., Wan, J., Kang, Y., Jiang, D., Xia, B., Zhao, Z. G., & Liu, R. (2025). A novel framework for predicting water level by spatial-temporal deep learning models based on integrated physical models in ungauged basins. https://doi.org/10.2139/ssrn.5132775

Zhang, Y., Ragettli, S., Molnár, P., Fink, O., & Peleg, N. (2022). Generalization of an encoder-decoder LSTM model for flood prediction in ungauged catchments. *Journal of Hydrology*, *614*, 128577. https://doi.org/10.1016/j.jhydrol.2022.128577

Zhou, T., Wang, F., & Yang, Z. (2017). Comparative analysis of ANN and SVM models combined with wavelet preprocess for groundwater depth prediction. *Water*, *9*(10), 781. https://doi.org/10.3390/w9100781

Zhou, Z., Ren, J., He, X., & Liu, S. (2021). A comparative study of extensive machine learning models for predicting long-term monthly rainfall with an ensemble of climatic and meteorological predictors. *Hydrological Processes*, *35*(11). https://doi.org/10.1002/hyp.14424

Zhu, H., & Zhou, Q. (2024). Advancing satellite-derived precipitation downscaling in data-scarce area through deep transfer learning. *IEEE Transactions on Geoscience and Remote Sensing*, *62*, 1–13. https://doi.org/10.1109/TGRS.2024.3367332

Zhu, Y., Ishikawa, T., & Subramanian, S. S. (2020). Simulation of runoff and infiltration using iterative cross-coupled surface and subsurface flows. *Japanese Geotechnical Society Special Publication*, *8*(3), 41. https://doi.org/10.3208/jgssp.v08.j13