

Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin,
Ethiopia



Gutema Genene Haleke

A Thesis submitted to the Department of Water Resources Engineering,
School of Civil Engineering and Architecture

Office of Graduate Studies

Adama Science and Technology University

March, 2023

Adama, Ethiopia

Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin,
Ethiopia

Gutema Genene Haleke

Advisor: Dereje Birhanu (PhD)

A Thesis Submitted to the department of Water Resources Engineering,
School of Civil Engineering and Architecture

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in Water
Resources Engineering (Specialization in Hydrology and Water Resource Management)

Office of Graduate Studies

Adama Science and Technology University

March, 2023

Adama, Ethiopia

DECLARATION

I hereby declare that this Master Thesis entitled “Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin, Ethiopia” is my original work. That is, it has not been submitted for the award of any academic degree, diploma, or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation

Gutema Genene

Name of the student

Signature

Date

RECOMMENDATIONS

I, the advisor of this thesis, hereby certify that I have read the revised version of the thesis entitled “Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin, Ethiopia” prepared under my guidance by Gutema Genene submitted in partial fulfillment of the requirements for the degree of Master’s of Science in Water Resources Engineering (Specialization in Hydrology and Water Resource Management). Therefore, I recommend the submission of the revised version of the thesis to the department following the applicable procedures.

Dereje Birhanu (PhD)

Major Advisor/Supervisor

Signature

Date

APPROVAL PAGE

I, the advisor of the thesis entitled “Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin, Ethiopia” developed by Gutema Genene, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Dereje Birhanu (PhD)

Major Advisor

Signature

Date

We, the undersigned, members of the Board of Examiners of the thesis by Gutema Genene have read and evaluated the thesis entitled “Hybrid Modeling Approach for Streamflow Simulation: The Case of Upper Awash Sub-basin, Ethiopia” and examined the candidate during the open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Water Resources Engineering (Specialization in Hydrology and Water Resource Management).

Mesfin Benti (PhD)

Chairperson

Signature

Date

Abdulkerim B.Serur (PhD)

Internal Examiner

Signature

Date

Kassa Tadele (Dr.Ing)

External Examiner

Signature

Date

Final approval and acceptance of the thesis is contingent upon the submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department Head

Signature

Date

School Dean

Signature

Date

Office of Postgraduate Studies, Dean

Signature

Date

ACKNOWLEDGMENTS

I want to thank the Almighty God first and foremost for his immeasurable mercy shown to me throughout my life and for giving me the capacity to complete this study. Then, I want to convey my profound gratitude to my adviser, Dereje Birhanu (PhD), for his wise advice, ongoing encouragement, and patience during my MSc studies. This thesis' extensive study and writing benefited from his extensiveness of experience and depth of understanding. Being able to work and study under his direction was a huge honor and privilege.

I want to sincerely thank Wollega University (WU) for giving me the chance to pursue a master's degree and for their financial assistance. Next, I want to express my gratitude to the entire staff of the Water Resources Engineering department at Adama Science and Technology University (ASTU) for providing me with a secure work environment and for their ongoing technical assistance throughout my study. I also want to thank the Ministry of Water and Energy (MoWE) and the Ethiopian Meteorology Institute (EMI) for helping to provide the necessary data. Last but not least, I want to thank my family for their love, prayers, concern, and sacrifices made to raise me and prepare me for the future.

TABLE OF CONTENTS

DECLARATION.....	iii
RECOMMENDATIONS.....	iv
APPROVAL PAGE	v
ACKNOWLEDGMENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF TABLES IN THE APPENDIXS	xiv
LIST OF FIGURES IN THE APPENDIXS	xv
LIST OF ACRONYMS AND ABBREVIATIONS	xvi
ABSTRACT	xvii
CHAPTER ONE	1
1. INTRODUCTION.....	1
1.1 Background	1
1.2 Statement of the Problem	3
1.3 Objectives.....	4
1.3.1 General objective.....	4
1.3.2 Specific objectives.....	4
1.4 Research Questions	5
1.5 Significance of the Study	5
1.6 Scope and Limitation of the Study.....	6
1.7 Thesis Organization.....	7
CHAPTER TWO	8
2. LITERATURE REVIEW	8
2.1 General Overview	8
2.2 Hydrological Modelling	8
2.3 Classification of Hydrological Models	9
2.3.1 Data-Driven Models	9
2.3.2 Conceptual Models	10
2.3.3 Physical-Based Models.....	10
2.4 Hydrological Model Selection Criteria	10
2.5 Soil and Water Assessment Tool (SWAT)	11

2.5.1 Components of SWAT Hydrological Model.....	11
2.5.2 Application of SWAT Model	16
2.6 Application of Machine Learning in Water Resources.....	18
2.6.1 Artificial neural network (ANN)	18
2.6.1.1 Multilayer Perceptron (MLP)	19
2.6.1.2 Transfer functions	20
2.7 Hybrid Modeling.....	21
2.7.1 Definition and concept of hybrid modeling.....	21
2.7.2 Classification of hybrid models.....	22
2.7.3 The state of the art in hybrid hydrological modeling today	23
2.7.4 Post-processor framework	25
2.7.5 Previous works in the study area	26
CHAPTER THREE	28
3. MATERIALS AND METHODS	28
3.1 Study Area Description	28
3.1.1 Location	28
3.1.2 Climate.....	28
3.1.3 Topography.....	29
3.2 Data Collection.....	30
3.2.1 Weather Data	30
3.2.2 Hydrological data	30
3.2.3 Digital elevation model data (DEM)	32
3.2.4 Land Use and Land Cover (LULC) data	33
3.2.5 Soil data	35
3.3 Data Analysis	35
3.3.1 Completing missing data	35
3.3.2 Test for data consistency and homogeneity.....	37
3.3.3 Rainfall data.....	37
3.3.3.1 Weighted area rainfall.....	38
3.3.3.2 The correlation between rainfall and streamflow	39
3.3.4 Temperature.....	40
3.4 Materials and Software's.....	41
3.5 Methodology	41

3.5.1 SWAT hydrological model setup	41
3.5.1.1 Watershed delineation.....	41
3.5.1.2 Hydrologic response unit (HRU) determination.....	42
3.5.1.3 Weather data input	43
3.5.1.4 SWAT Simulation.....	43
3.5.1.5 Sensitivity analysis.....	44
3.5.1.6 Calibration of model	45
3.5.2 Artificial Neural Networks Development.....	45
3.5.2.1 Input selection.....	46
3.5.2.2 Data division (train, validation and test data)	47
3.5.2.3 Data Preprocessing.....	48
3.5.2.4 Model architecture selection	49
3.5.2.5 Model network (structure) selection	51
3.5.2.6 Model training.....	52
3.5.2.7 Hyperparameters optimization.....	53
3.5.2.8 Testing the model.....	55
3.5.3 Hybrid model.....	55
3.5.3.1 Scheme 1.....	55
3.5.3.2 Scheme 2.....	56
3.6 Model Performance Evaluation.....	57
3.7 Conceptual Framework of the Study.....	61
CHAPTER FOUR.....	62
4. RESULTS AND DISCUSSIONS	62
4.1 Simulation Streamflow using SWAT Model	62
4.1.1 Streamflow parameter sensitivity analysis	62
4.1.2 Calibration and validation of SWAT model.....	64
4.2 Hybrid model.....	66
4.2.1 Scheme 1.....	66
4.2.2 Scheme 2.....	71
4.2.2.1 Residual analysis (post-processing).....	71
4.2.2.2 Modeling residual error.....	73
4.3. Model Performance Evaluation.....	80
4.3.1 Hybrid model performance at the selected calibration sites.....	80

4.3.2 Hybrid models performance for extreme events	90
CHAPTER FIVE	97
5. CONCLUSIONS AND RECOMMENDATIONS.....	97
5.1 Conclusions	97
5.2 Recommendations	99
REFERENCES.....	100
APPENDICES	111

LIST OF TABLES

Table 3.1. Streamflow gauging stations.....	32
Table 3.2. Land use and land cover categories identified at the study area.....	34
Table 3.3. Correlation coefficients between the observed rainfall and streamflow.....	40
Table 3.4. Flow Parameters for calibration and validation of the SWAT model	44
Table 3.5. Statistical properties of training and testing datasets	48
Table 3.6. The hyperparameters values used for model optimization.	54
Table 3.7. Performance ratings recommended for daily and monthly streamflow simulation.	60
Table 4.1. Shows the best optimized fitted parameter value during calibration.....	63
Table 4.2. Training and testing performance of the screened Hybrid1 and stand-alone ANN models	69
Table 4.3. Results of the hyperparameter tuning for stand-alone ANN and Hybrid1.	70
Table 4.4. Residual error descriptive statistics	72
Table 4.5. Screened best models for residual simulation.....	78
Table 4.6. Results of the hyper parameter tuning for residual and Hybrid2.....	78
Table 4.7. Streamflow simulation performance statistics	88
Table 4.8. Classifications of hydrological flow conditions	93

LIST OF FIGURES

Figure 2.1. Overview of activation functions	21
Figure 3.1. Location of the study area	29
Figure 3.2. Monthly flow and rainfall characteristics of gauge stations under the study area	31
Figure 3.3. Annual flow characteristics of gauge stations under the study area.....	31
Figure 3.4. Digital Elevation and Slope map of the Upper Awash Sub-basin.....	33
Figure 3.5. Land use and land cover and soil map of the Upper Awash Sub-basin	35
Figure 3.6. Long-term average monthly rainfall pattern of the selected stations	38
Figure 3.7. Annual rainfall pattern of the selected stations	38
Figure 3.8. Thiessen polygon maps	39
Figure 3.9. Average annual maximum and minimum temperature of the selected stations	40
Figure 3.10. Watershed delineated sub-basins of the Upper Awash Sub-basin	42
Figure 3.11. Structure of a multilayer perceptron ANN	51
Figure 3.12. Conceptual framework for the study	61
Figure 4.1. Streamflow-sensitive parameters.....	62
Figure 4.2. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Akaki station.	65
Figure 4.3. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Melka Kunture station.....	66
Figure 4.4. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Hombole station	66
Figure 4.5. Correlation between observed streamflow, climate variables, and simulated streamflow.....	67
Figure 4.6. The relationship between residual error and precipitation at three stations	72
Figure 4.7. Residual error frequency distribution histogram.	73
Figure 4.8. Correlation between observed streamflow, climate variables, and residual error.....	74

Figure 4.9. Akaki station Cross-correlation graphs	76
Figure 4.10. Melka Kunture Station Cross-correlation graphs	76
Figure 4.11. Hombole station Cross-correlation graph.....	77
Figure 4.12. Observed and estimated residual errors time series	79
Figure 4.13. Observed and simulated streamflow scatter plots at Akaki station.....	81
Figure 4.14. Hydrographs of Akaki station during the calibration and validation period.....	82
Figure 4.15. Observed and simulated streamflow scatter plots at Melka Kunture station.	84
Figure 4.16. Hydrographs of Melka Kunture station during the calibration & validation period	85
Figure 4.17. Observed and simulated streamflow scatter plots at Hombole station.....	85
Figure 4.18. Hydrographs of Hombole station during the calibration and validation period	87
Figure 4.19. Boxplots of observed versus simulated streamflow for the Akaki station.	91
Figure 4.20. Box-plots of the observed versus simulated for the Melka Kunture station	92
Figure 4.21. Box plots of observed versus simulated streamflow for the Hombole station.....	92
Figure 4.22. Summary of model performance in terms of root mean squared error (RMSE) for daily streamflow simulations	95
Figure 4.23. Validation period flow duration curves.....	95

LIST OF TABLES IN THE APPENDICES

Appendix A.1. Description of meteorological stations.....	111
Appendix A.2: Percentage of missing data for each meteorological station	111
Appendix A.3 Major Soil groups in the study area	112

LIST OF FIGURES IN THE APPENDICES

Appendix A.3. Major Soil groups in the study area	112
Appendix A.4. Double mass curve representative of the Upper Awash Sub-basin meteorological stations	112
Appendix B.1. Streamflow calibration 95PPU plot.....	113
Appendix B.2. Streamflow validation 95PPU plot.....	113
Appendix C 1. Cross-correlation graphs between residual of SWAT with observed streamflow and areal precipitation.....	114
Appendix E1. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Akaki station NB: The optimum hyper parameters in figure c and d are marked in green.....	120
Appendix E2. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Melka Kunture station.....	121
Appendix E3. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Hombole station	122
Appendix F1. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Akaki station	123
Appendix F2. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Melka Kunture station.....	124
Appendix F3. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Hombole station	125

LIST OF ACRONYMS AND ABBREVIATIONS

Adam	Adaptive Moment Estimation
ANN	Artificial Neural Network
CN	Curve Number
DEM	Digital Elevation Model
DMC	Double Mass Curve
FAO	Food and Agricultural Organization
FC	Filed Capacity
FFNN	Feed Forward Neural Network
GIS	Geographic Information System
HBV	Hydrologiska Byrans Vattenbalansavdelning
HEC-HMS	Hydrologic Engineering Center-Hydrologic Modeling System
HRUs	Hydrologic Response Units
Km ²	Square kilometer
LULC	Land Use Land Cover
m.s.l	Mean Sea Level
ML	Machine Learning
MLP	Multilayer Perceptron
MoWE	Ministry of Water and Energy
EMI	Ethiopian Meteorology Institute
NSE	Nash-Sutcliffe Efficiency
PBIAS	Percent Bias
R ²	Coefficient of Determination
ReLU	Rectified Linear Unit Activation Function
RMSE	Root mean square error
SCS	Soil Conservation Services
SDGs	Sustainable Development Goals
SGD	Stochastic Gradient Descent
SRTM	Shuttle Radar Topography Mission
SUFI-2	Sequential Uncertainty Fitting Version 2
SWAT	Soil and Water Assessment Tool
SWAT-CUP	SWAT Calibration and Uncertainty Programs
USGS	United State Geological Survey
UTM	Universal Transverse Mercator

ABSTRACT

Hydrological models are effective tools for simulating the effect of watershed processes and management on soil and water resources. However, incomplete representations of physical processes often lead to structural errors in physical-based hydrologic models. A data-driven model, on the other hand, can reduce streamflow modeling errors without relying on the underlying physical process. But, in dynamic climates and environments, a data-driven model may be unreliable in simulating future hydroclimate variables. Therefore, to overcome the limitations of physical-based and data-driven models, two hybrid models (Hybrid1 and Hybrid2) were built and evaluated by integrating Soil and Water Assessment Tool (SWAT) outputs with the artificial neuron network (ANN) for daily streamflow simulation using multisite calibration at the Upper Awash Sub-basin. The two hybrid models were built with different levels of integration: the SWAT model was used as the first model layer in the Hybrid1 model, whereas the SWAT residuals were simulated in the Hybrid2 model as an error-correcting layer. In order to meet the objectives of the study, first the SWAT model's performance in simulating streamflow was evaluated through sensitivity analysis. Then SWAT was calibrated (from 2003–2011) and validated (from 2012–2015) using the SUF-2 algorithm. Following this, hybrid models were built and trained using the TensorFlow v2 Python-based Keras machine learning platform. The model's performance was evaluated by statistical techniques including the coefficient of determination (R^2), Nash Sutcliffe efficiency (NSE), and percent of bias (PBIAS). Moreover, the performance of the models for extreme flow events was also assessed. The average calibration and validation period model performance at the sub-basin show the value of R^2 varied from 0.53 to 0.7, 0.63 to 0.81, and 0.8 to 0.9 for SWAT, Hybrid1, and Hybrid2 models, respectively. The NSE value varied from 0.51 to 0.68, 0.59 to 0.8, and 0.72 to 0.88 for SWAT, Hybrid1, and Hybrid2 models, respectively. The PBIAS for the validation period also varied from 16.4 to 16.9, 26 to -12.2, and -4.9 to 3.9 for SWAT, Hybrid1, and Hybrid2 models, respectively. The average root mean squared error (RMSE) of the SWAT model in the calibration and validation period was reduced by 2%, 22%, and 20% by Hybrid1 and by 18%, 23%, and 34% by Hybrid2 at the Akaki, Melka Kunture, and Hombole watersheds, respectively. For extreme flow simulation, the Hybrid2 model performed better in all flow conditions, while the Hybrid1 model performed relatively well in low flow. The findings of this research demonstrate that hybrid modeling approaches can well reproduce historical daily streamflow series and outperform SWAT and stand-alone ANN models.

Keywords: ANN, Data-driven model, Hybrid model, Physical-based model, SWAT Model, Streamflow, Upper Awash Sub-basin.

CHAPTER ONE

1. INTRODUCTION

1.1 Background

The amount of freshwater available for the use of humans and other living things in the world is limited. Due to climate change, increasing population, and urbanization, ensuring the quantity and quality of this natural asset has become critical. Furthermore, the public has become more sensitive to water resource problems, while water demand has increased in parallel with the rapidly developing world. This sensitivity led to a systematic review of all activities related to the management of water resources, which required complementary hydrological studies to determine the quality and quantity of resource availability on a spatiotemporal scale (Farfán et al., 2020; Ghaith et al., 2020).

Streamflow is one of the most important components of the hydrological cycle. At a watershed level, streamflow simulation is necessary for many water resource-related activities, such as management, planning, and design of water resource systems (Sharma et al., 2021). A hydrological model is an abstraction of a complex, non-linear, time and space-varying hydrological process that attempts to mimic reality and is mainly used for predicting system behavior and understanding various hydrological processes (Akoko et al., 2021). Since the development of the first rainfall-runoff model in the late nineteenth century, hydrological models have evolved to encompass a broad range of complexities and approaches. At present, many hydrological models have been developed to simulate hydrological processes, and they are important tools for estimating flow values, which can establish relationships between rainfall and runoff. The two common modeling paradigms in hydrologic modeling are physical-based and data-driven, and these two modeling methods have been studied separately by many researchers (Jaiswal et al., 2020).

Physical-based hydrological models capture simplified simulations of the underlying physical processes that drive water flow. These models often use various data sources as the basis for their configuration, such as land cover maps, soil maps, or digital elevation models (Gauch et al., 2019). Among many available physical-based models used for simulating streamflow processes, the Soil

and Water Assessment Tool (SWAT) is the most widely applicable and consistently evolving. Several researchers have utilized the SWAT model in their study and strongly suggest its effectiveness in rainfall-runoff modeling since it allows for long-term simulation of hydrometric output data (Serur and Sarma, 2016; Khalid et al., 2016; Gonfa and Kumar, 2018; Raihan et al., 2020; El Harraki et al., 2021).

On the other hand, machine learning for hydrological simulation has made significant progress, attracting the attention of many scientists and researchers in this field (Darbandi and Pourhosseini, 2018). An artificial neural network (ANN) is one of the results of evaluating complex unknown phenomena by modeling human brain networks. One of the benefits of ANN is that they can perform specific functions to express the relationship between input and output data, and they can also extract the most important information from the available data (Lv et al., 2020). ANN have been used successfully in hydrological research for a variety of purposes, including uncertainty prediction, rainfall-runoff modeling, runoff simulation, storage inflow simulation, river sediment modeling, and the assessment of water-powered energy (Sharma et al., 2015; Choong et al., 2020; Dalkiliç and Hashimi, 2020).

Furthermore, hybrid hydrological modeling, which refers to an integrated modeling framework that includes more than one modeling framework, has gained popularity. The goal of using hybrid models in different disciplines is to combine prior knowledge such as conservation and kinetic laws with data-driven, non-parametric models (von Stosch et al. 2014). In this regard, hydrologists have recently worked hard to integrate data-driven hydrological models with physical-based models and found a promising result in terms of model performance improvement (Mekonnen et al., 2015; Noori and Kalin, 2016; Koycegiz and Buyukyildiz, 2019; Wu et al., 2019).

Often, single hydrological models cannot equally describe the characteristics of hydrological processes for all ranges of model outputs (streamflows), due to the multiple hydrological responses and their values at different magnitudes. The hybrid approach opens up possibilities for handling such difficulties and allows for improving the predictive capability of models.

Konapala et al. (2020), for example, reported that after experiments in 531 catchments representing diverse climate conditions across the Continental United States, the ability of the hybrid of physical-based and machine-learning models to simulate streamflow demonstrated reasonable success in several catchments. Similarly, Ghaith et al. (2020) used a hybrid of conceptual and data-driven models to simulate streamflow in Texas' Guadalupe Basin and reported the RMSEs of the validation period as 34 and 44 m³/s for the hybrid and conceptual models, respectively. Thus, the primary goal of this paper is to evaluate the robustness of hybrid modeling approaches for Upper Awash Sub-basin streamflow simulations using multi-site calibration with a hybrid of two popular models, SWAT and ANN.

1.2 Statement of the Problem

In recent decades, water scarcity, climatic variability, and extreme weather events have posed significant challenges in water resource planning and management. As a result, understanding a river basin's hydrological characteristics is critical for estimating the amount of water available over time. Water is a prerequisite for achieving all of the core Sustainable Development Goals (SDGs). Water resource development and availability are critical components of agricultural and broader economic growth, which means that without significant investments in water, it will be impossible to end poverty (Goal 1) or achieve the majority of the other SDGs.

According to the Food and Agricultural Organization (FAO) (2020) report, the Awash Basin is the most utilized river basin in Ethiopia, hosting most of the industrial activities in the country, a number of small-to-large scale irrigation schemes, and the main population centers of the country. Due to this, the basin faces high water stress during the peak of the irrigation season and frequent flooding in rainy seasons. Therefore, in basins like Awash, where there is a high demand for water and the frequency of extreme events is increasing, having accurate and relevant hydrological knowledge of the volume and variations of streamflow across the time dimension (short-term and long-term) allows managers and policy makers to make the right decisions regarding the design of water facilities, hydropower generation, allocating surface waters in irrigation scheduling, and water balance management at the catchment scale. However, due to the complexity of the hydrological cycle and other factors, accurate estimation of streamflow using hydrological models

remains challenging. This challenge may arise from the selection and implementation of these models for a specific basin (Mengistu et al., 2016).

Furthermore, in developing countries like Ethiopia, there are several bottlenecks that reduce the quality of hydrological model output. To name a few, there is a lack of accurate and relevant data, a lack of computational resources, complex topographic features, and an uneven distribution of metrological stations (Meshesha and Abdi, 2019; Adane et al., 2021).

To overcome those and other types of constraints, researchers have recently worked to improve the accuracy of hydrological outputs by using hybrid modeling and have successfully integrated different models in various parts of the world with significant improvements in model accuracy (Robertson et al., 2013; Humphrey et al., 2016). As a result, besides using the traditional modeling approach, using the hybrid model is essential. However, there have only been a few studies in Ethiopia that have used various modeling approaches and no studies have been conducted in the study area specifically using a hybrid modeling approach, which is a relatively new and promising approach for streamflow modeling. As a result of this research gap, the current study focuses on evaluating the performance of a hybrid model for daily streamflow simulation at the Upper Awash Sub-basin by combining a physical-based SWAT with a data-driven ANN model.

1.3 Objectives

1.3.1 General objective

The main objective of this research is to simulate streamflow in the Upper Awash Sub-basin by using a hybrid modeling approach.

1.3.2 Specific objectives

In order to meet the general objective of the study, the following specific objectives were addressed:

- To develop and calibrate a physical-based SWAT model at multiple sites in the Upper Awash Sub-basin for streamflow simulation.

- To enhance the accuracy of hydrological models through a hybrid of SWAT and a data-driven ANN model.
- To develop an error correction (post-processing) model for SWAT hydrological model using a data-driven ANN model.
- To evaluate and compare the performance of SWAT, ANN, and hybrid models in simulating streamflow extremes at multiple sites in the Upper Awash Sub-basin.

1.4 Research Questions

The main and specific objectives of this study are derived from research questions relevant to the water resources development and management of the study area. Therefore, the current study shall address the following research questions.

1. How to develop and perform a multi-site calibration of the SWAT model for streamflow simulation in the Upper Awash Sub-basin?
2. How does a hybrid model improve the hydrological model's accuracy?
3. How to develop an error correction (post-processing) model for SWAT hydrological model using a data-driven ANN model?
4. How have SWAT, ANN, and hybrid models performed for simulating streamflow extremes at the Upper Awash river basin?

1.5 Significance of the Study

Hydrological models are essential for assessing water resources through the analysis of streamflow quality and quantity. Depending on the objectives, resources, and data available, a variety of hydrological modeling techniques have been used in several studies. For streamflow simulation, hydrological models play an important role, but not all hydrological modeling approaches produce the same results or perform in the same way (Devia et al., 2015).

Accurate simulation of the rainfall-runoff relationship is required for a variety of water-related applications, including flood forecasting (Hu et al., 2018). In general, physical-based hydrological

models are frequently used to simulate nonlinear relationships between rainfall and runoff, and these models are used in a variety of hydrologic applications, including short-term simulations for flood prediction and long-term simulations for drought analysis and water resource management (Wang et al., 2021). In several hydrologic models, multiple parameters are used to calibrate modeling processes and reduce output errors or residuals. Errors in streamflow simulation may result in incorrect hydrological analysis when using hydrologic model results. Decision makers running water-related infrastructures for flood forecasting, irrigation control, drought prediction, and other water management objectives, in particular, require credible hydrologic modeling conclusions that take into account the uncertainties inherent in the results (Shrestha and Solomatine, 2008).

Therefore, in this study, a hybrid modeling framework was adopted for streamflow simulation in the Upper Awash Sub-basin. Since the hybrid modeling framework is a new modeling insight, its performance has been tested in different parts of the world and shows a significant improvement, but in the study area, its performance has not been tested yet. As a result, the findings of this study will make a significant contribution to the Upper Awash Basin's effective water management. Moreover, the study will contribute to water-related project work in the sub-basin as well as have a significant impact on socioeconomic and environmental aspects.

1.6 Scope and Limitation of the Study

The study focuses on streamflow simulation in the Upper Awash Sub-basin, which is located in the upper part of the Awash Basin, using three different modeling approaches: physical-based, data-driven, and hybrid. SWAT and ANN are the two selected models, and besides the standard methodology, the hybrid model of SWAT and ANN was implemented based on two different integration schemes. Then the models were calibrated at three sites in the Upper Awash Sub-basin located between longitudes 37°50'10"E and 39°30'20"E and latitudes 8°18'45"N and 9°20'7"N. The performance of each model was evaluated using a statistical and graphical method; furthermore, the model's performance in simulating extreme flow was also evaluated accordingly. As an input, land use and land cover data, soil data, and topographic data were used for the SWAT physical-based model, and the streamflow data was used to calibrate and validate the modes. Daily

hydro-meteorological data was collected from the respective institutions over a 16-year period, from 2000 to 2015. However, this study investigates the performance of physical-based SWAT and multilayer perceptron neural network (MLPNN) for streamflow simulation and is limited to investigating the hybrid of a hydrological model with other data-driven models. Another limitation of this study is that, despite multisite calibration, the approach's applicability was not tested across a wide range of climate zones and watershed characteristics.

1.7 Thesis Organization

This thesis was organized into five chapters. The first chapter is the introduction, which contains the study's background and provides an overview of the research. Furthermore, the study's objectives, significance, and scope/limitations were discussed. The second chapter (literature review) focuses on providing a thorough explanation of the hybrid modeling approach. It also includes the application of a physical-based and data-driven model for water resources by reviewing the global and national studies. In addition, the concept of residual error modeling or post-processing framework was thoroughly discussed. The third chapter (materials and methods) provides a brief description of the study area, materials and software used, required data preparation, data quality analysis, and step-by-step methodology adopted to accomplish the study's objectives. The fourth chapter explains the results of the study, focusing on the research objectives and discussing previous related works. Finally, under the fifth chapter, conclusions and recommendations are provided.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 General Overview

In this chapter, a detailed and relevant literature review includes the definition and theory of hydrological modeling, particularly physical-based and machine-learning approaches discussed, SWAT, ANN, and hybrid modeling approaches, and the concept of model performance evaluation reviewed from related literature by evaluating the scientific works that are associated with the subject of this study.

2.2 Hydrological Modelling

Models are representations of real-world systems that contain the essential structure. Some models are physical representations of natural systems in miniature. A series of equations is sometimes used to represent the systems, resulting in mathematical models. The number, form, and interconnections of these equations in a model can range from simple to complex. The equations in mathematical models can be derived from fundamental physical laws or from statistical analysis of observed data (empirical equations) (Butcher, 2008).

A runoff model is a series of equations that assist in the estimation of streamflow as a function of various factors used to describe watershed characteristics (Sorooshian et al., 2008). The major purpose of hydrological modeling is to get a better understanding of the performance of a real-world domain, such as a river catchment or reach, an aquifer, a water distribution or drainage network, an estuary, or coastal waters. The capacity to produce performance forecasts is the second goal, which will require reproducing prior performances and analyzing why they occur (Price and Solomatine, 2009). Since its humble beginnings in the 1850s, hydrologic modeling has come a long way. Modeling advances have accelerated, owing mostly to easy access to nearly infinite processing power, sophisticated equipment, and remote sensing and GIS capabilities (Singh, 2018).

2.3 Classification of Hydrological Models

Many hydrological models with unique and common properties are being produced on a daily basis. Engineers, professionals, and researchers can benefit from proper classification by understanding the properties of models before choosing to use them in their work (Jajarmizad et al., 2012). There are several classes of hydrological models, all with the goal of making the models simple to learn and run. It should be emphasized, however, that most hydrological models are not particularly different from one another; the major variation is in their setup and operation technique, not in their ideas (Devia et al., 2015). Hydrological models have been distinguished and classified based on their function and objectives, structure, and spatial disaggregation level. Because there are various ways to classify hydrological models, the more-or-less standard classification used in this thesis is introduced here.

2.3.1 Data-Driven Models

These are observation-oriented models that utilize information from current data without taking into account the characteristics and processes of the hydrological system. They are based on mathematical equations generated from contemporaneous input and output time series rather than physical catchment processes. These models are valid within the boundaries. Data-driven modeling (DDM) is based on an examination of the data that characterizes the system under consideration. A model may then be built based on the linkages between the system state variables (input, internal, and output variables), with only a few assumptions about the system's "physical" behavior. Modern methodologies in hydraulic engineering and hydrology can go much further than traditional empirical modeling methods. They enable the solution of numerical simulation problems, the reconstruction of highly nonlinear functions, the classification and grouping of data, and the construction of rule-based systems (Solomatine and Ostfeld, 2008). This approach is illustrated by the unit hydrograph. Statistical approaches employ regression and correlation models to determine the functional connection between inputs and outputs. In rainfall-runoff modeling, computational intelligence models such as the artificial neural network (ANN), fuzzy inferencing, and neuro-fuzzy systems are increasingly being applied (Price and Solomatine, 2009).

2.3.2 Conceptual Models

This model explains all of the hydrological processes that make up the system. It is made up of a series of interconnected reservoirs that reflect the physical elements in a watershed and are refilled by rainfall, infiltration, and percolation and emptied by evaporation, runoff, and drainage, among other things (Lee and Moon, 2007). This approach employs semi-empirical equations, and model parameters are determined not only from field data but also through calibration. Calibration necessitates a large number of meteorological and hydrological records in order to make use of conceptual representations of the hydrological processes that appear to be relevant in the system.

2.3.3 Physical-Based Models

Physical-based models are mathematical representations of real-world systems. These models require parameters that can be monitored and are time- and space-dependent. They have low requirements for hydrological and meteorological data for calibration, but it also includes estimating various variables that describe the physical properties of a catchment, such as soil moisture content, beginning water depth, slope, topology, and so on (Sorooshian et al., 2008; Devia et al., 2015). In contrast to conceptual models, physical-based models do not consider water flow in a watershed that occurs between a few storage units. Hydrological processes involving water and energy movement, such as Richards' equation for vadose zone flow, Boussinesq's equation for groundwater flow, and Saint Venant equations for overland and channel routing, are evaluated using partial differential equations. Physical-based models produce more complete and exact results than other model types (Abdulkareem et al., 2018).

2.4 Hydrological Model Selection Criteria

There are multiple hydrological models available today that simulate the hydrological process at various geographical and temporal dimensions. Although there are no hard and fast rules for choosing between models, some basic recommendations may be mentioned (Abdella, 2013). However, one of the most important factors to consider when choosing a hydrological model is the availability of input data to model the watershed. As a result, the simplest approach that can offer answers to the questions must be adopted. It is especially important to identify the simplest model that will produce adequate accuracy. There are several factors that may be utilized to select

the best hydrologic model. Because each project has its own unique set of objectives and demands, these criteria are always project-specific. There are four primary, common, and essential ones that must always be considered among the different project-dependent selection criteria (Haileyesus, 2011):

- i. Required model outputs are crucial for the required purpose and must be estimated by the model. (Does the model forecast the variables required by the project, such as peak flow, event volume, hydrograph, and long-term flows?)
- ii. Hydrologic processes must be represented in order to accurately estimate desired outputs (Can the model simulate regulated reservoir operation, a single event, or continuous processes?)
- iii. Availability of input data (Can all of the model's inputs be given within the project's schedule and cost constraints?)
- iv. Cost (does the expenditure appear to be worthwhile in terms of the project's objectives?)

2.5 Soil and Water Assessment Tool (SWAT)

Arnold et al. (1998) developed the Soil and Water Assessment Tool (SWAT), a semi-distributed continuous non-point source model in which the watershed is divided into a large number of sub-basins, which are then subdivided into unique soil or land use classes, which are usually referred to as the hydrologic response units. SWAT is based on a water balance for each of the four storage volumes: snow, soil profile, shallow aquifer, and deep aquifer.

SWAT operates from small watersheds to river basin-scale and is used to simulate the quality and quantity of surface and groundwater and predict the environmental impact of land use, land management practices, and climate change. In addition, it has also been applied in categories such as curve number modifications, wetland applications, and best management practices. As a result of over 30 years of USDA research, the model developed continuously and became one of the most commonly used and effective models (El Harraki et al., 2021).

2.5.1 Components of SWAT Hydrological Model

In the SWAT model, the hydrology of the watershed is simulated in two ways. These are;

- i. The land phase hydrology cycle: It governs the quantity of water, nutrient, sediment, and pesticide loading to the channel in individual sub-watershed.
- ii. Water/routing phase of the hydrological cycle: It governs the movement of water, and sediments within the channel system of the watershed to the outlet. It is based on the principles of the water balance:

$$SW_t = SW_0 + \sum_{i=1}^t (R_{day} - Q_{surf} - E_a - \omega_{seep} - Q_{gw}) \dots \dots \dots (2.1)$$

Where; SW_t Final soil water content (mm)

SW_0 Initial soil water content on the day i (mm)

t time (days)

R_{day} The amount of precipitation on the day i (mm)

Q_{surf} The amount of surface runoff on the day i (mm)

E_a The amount of water evapotranspiration on the day i (mm)

Q_{gw} The amount of water return flow on the day i (mm)

W_{seep} The amount of water entering the vadose zone from the soil profile on the day i(mm)

In SWAT, the principle that surface runoff occurs whenever the rate of water application to the ground surface exceeds the rate of infiltration is applied. Based on this principle, SWAT employs two methods for calculating surface runoff: the Soil Conservation Service Curve Number technique (USDA Soil Conservation Service, 1972) and the Green and Ampt infiltration methods..

The Soil Conservation Service (SCS) curve number approach, also known as the curve-number (CN) method, combines land use and soil properties into a single parameter. The SCS curve number approach is simple, frequently used, and effective for estimating the quantity of runoff following a rainfall event under different land uses and soil types. The SCS curve number equation is (SCS, 1972):

$$Q_{surf} = \frac{(R_{day} - I_a)^2}{(R_{day} - I_a + S)} \dots \dots \dots (2.2)$$

Where;

Q_{surf} Accumulated runoff (mm water)
R_{day} Daily rainfall (mm water)
I_a Initial abstraction (mm)
S Retention parameter (mm water),

The retention parameter varies spatially due to changes in soils, Land use, management and slope and temporally due to changes in soil water content. The retention parameter is defined as:

$$S = 25.4 \left(\frac{100}{CN} - 10 \right) \dots \dots \dots (2.3)$$

Where;

CN daily curve number

The initial abstraction, I_a is commonly approximated as 0.2S and the above equation becomes

$$Q_{surf} = \frac{(R_{day} - 0.2S)^2}{(R_{day} + 0.8S)} \dots \dots \dots (2.4)$$

For R_{day} > I_a (=0.2S)

In the SWAT model the peak runoff rate can be computed by the use of a modified rational method (Equation 2.5);

$$q_{peak} = \frac{atc * Q_{surf} * A}{3.6 * t_{conc}} \dots \dots \dots (2.5)$$

Where;

Q_{peak} Peak runoff rate (m³ /s)
Q_{surf} Surface runoff (mm)
atc Fraction of daily rainfall
t_{conc} Time of concentration for sub-basin (hr)
A Area the sub-basin area (km²)

3.6 A unit conversion factor.

Penman-Monteith, Monteith (1965), Priestley and Taylor (1972), and Hargreaves et al. (1985) methods are provided in the SWAT model to estimate potential evapotranspiration (PET). The PeEMIn-Monteith method computed potential evapotranspiration using equation 2.6.

$$E = \frac{\Delta \cdot (H_{\text{net}} - G) + \rho_{\text{air}} \cdot [e_z^0 - e_z] / r_a}{\Delta + \gamma(1 + \frac{r_c}{r_a})} \dots \dots \dots (2.6)$$

Where;

λE	the latent heat flux density (MJ m ⁻² d ⁻¹)
E	Depth evaporation rate [mm/day]
H_{net}	Net radiation at the crop surface [MJm/day]
G	Soil heat flux density to the ground [MJ/m ² day]
Δ	Slope of the saturated vapor pressure curve, $de/[kpa/^\circ C]$
$\odot$	Psychrometric constant [kpa/°C], ρ_{air} is the air density (kg/m ³)
C_p	Specific heat at constant pressure (MJ/kg °C)
e_z^0	Saturated vapor pressure of air at height z [kpa]
e_z	Water vapor pressure of air at height z [kpa]
r_c	Plant canopy resistance (s/m)
r_a	Diffusion resistance of the air layer (aerodynamic resistance) (s/m).

Percolation occurs into the soil layers when the soil water content is above the field capacity (FC). In SWAT mode, the quantity of water that percolates in the soil layer is computed using the following equation:

$$\text{Swly, excess} = \text{Swly} - \text{Fcly} \dots \dots \dots (2.7)$$

Where; $\text{Swly} > \text{Fcly}$

Swly, excess	Drainage quantity of water in the soil layer (mm)
Swly	Water content of the soil layer on a given day (mm)

Fcly Water content of the soil layer at field capacity (mm)

The simulation of groundwater flow in the SWAT model contributed to total streamflow and can be divided into two (2) aquifer systems, such as a shallow aquifer (2–20 m) and a deep aquifer (> 20 m). The water that enters the shallow aquifer can contribute to the total streamflow inside the watershed, whereas the water that infiltrates the deep aquifer contributes to the streamflow outside the watershed (Arnold and Allen, 1996). The water balance for a shallow aquifer is calculated using equation 2.8 (Neitsch et al., 2011).

$$Aqsh_i = Aqsh_{i-1} + Wrchrg_i - Qgw - Wrevap - Wdeep - Wpump_h \dots \dots (2.8)$$

Where;

Aqsh_i Quantity of water stored in the shallow aquifer on the day i (mm)
Aqsh_{i-1} Volume of water stored in the shallow aquifer on day i-1(mm)
Wrchrg_i Quantity of recharge entering the aquifer on the day i (mm)
Qgw Groundwater flow, or base flow, into the main channel on the day i (mm)
Wrevap Amount of water moving into the soil zone in response to water deficiencies on the day i (mm).
Wdeep Amount of water percolating from the shallow aquifer into the deep aquifer on the day i (mm).
Wpump_h Amount of water removed from the shallow aquifer by pumping on the day i (mm).

To compute the flow in the channel system, two (2) approaches are available. The Muskingum and variable storage techniques are the two available methods. The Muskingum technique, when combined with wedge and prism storage, can represent the quantity of storage in a channel length. The variable storage approach, on the other hand, employs a continuity equation. The variable storage approach developed by Williams (1969) was used in this investigation and is represented by the following equation:

$$\Delta V_{stored} = V_{in} - V_{out} \dots \dots \dots (2.9)$$

Where;

ΔV_{stored}	Change in the quantity of storage during the time step (m^3)
V_{in}	Quantity of inflow during the time step (m^3)
V_{out}	Quantity of outflow during the time step (m^3)

2.5.2 Application of SWAT Model

SWAT is widely regarded as one of the most effective hydrological models for addressing hydrologic and environmental issues around the world. However, while the model's effectiveness and reliability have been demonstrated on other continents, its applicability in Africa is still questionable (Akoko et al., 2021). When we come to Ethiopia, the SWAT model is utilized in many watersheds in the country under different climate and catchment characteristics. The majority of the studies address the issues of applications considering: water resources and streamflow; erosion and sedimentation; land-use management and agricultural-related contexts; climate-change contexts; and model parameterization in various river basins. Some studies are discussed as follows:

The performance of the SWAT model was evaluated in the Weyib river catchment in Ethiopia by Serur and Sarma (2016) for runoff and sediment yield simulations under comparable accuracy in simulations, and the results of calibration and validation of the model at a daily time step were reported as NSE values of 0.92 and 0.95, respectively, and the most sensitive parameter was found to be Curve Number (CN).

Tufa and Sime (2020) applied the SWAT model to simulate flow and evaluate model performance in a basin 1711 km^2 wide of the Toba sub-basin located in southwest Ethiopia. The sensitivity analysis, calibration, and validation of the model were performed using the sequential uncertainty fitting method (SUFI-2) and showed that the groundwater delay and the curve number were the two most sensitive parameters, and the performance of the model for the calibration and validation periods was reported as R^2 , 0.82, and 0.70, and NSE, 0.72, and 0.68, respectively. Likewise, Aliye et al. (2020) compared the performance of two hydrologic models in streamflow simulation of the Katar River basin in Ethiopia, finding that the HEC-HMS model outperformed the SWAT model. The efficacy of the models was measured in terms of R^2 and NSE, with SWAT scoring 0.8 and

0.69 for calibration and 0.78 and 0.67 for validation, respectively, and HEC-HMS scoring 0.88 and 0.75 for calibration and 0.87 and 0.73 for the validation period.

Fanta and Sime (2021) also investigated the performance of SWAT and HEC-HMS models in a daily runoff simulation of the Toba watershed using the same input data set and found that both SWAT and HEC-HMS models performed well in the runoff simulation, with statistical justifications provided in terms of R^2 , NSE, PBIAS, and RMSE. The respective values for HEC-HMS were reported as 0.85, 0.805, 33.7%, and 22.7 m³/s during calibration and 0.87, 0.84, 2.71%, and 20.7 m³/s during validation. For the SWAT model, the respective values were 0.69, 0.61, – 26.1%, 31.85 m³/s, for calibration, and 0.62, 0.54, – 18.3, 35.24 m³/s, for the validation period.

Similarly, several studies have been conducted in different watersheds found in the Awash Basin, the basin where the study area is located. Tessema et al. (2015) examined the applicability of the SWAT in the Upper Awash Sub-basin. The model's performance was assessed using streamflow predictions at four sub-basin outlets and the river basin's main outlet. Model validation revealed that daily streamflow was reasonably predicted, as evidenced by Nash-Sutcliffe values ranging from 0.55 to 0.71.

Tessema and Kebede (2021) analyzed and quantified the impacts of climate change on streamflow in the Kesem sub-basin and the Awash River basin using the SWAT. The statistical performance of the SWAT model in simulating the streamflow was reported with a coefficient of determination (R^2) of 0.84 and a Nash-Sutcliffe Efficiency (NSE) of 0.78 for monthly calibration and an R^2 of 0.78 and an ENS of 0.73 for monthly validation periods. Gonfa and Kumar (2018) also calibrated the SWAT model for the simulation of monthly runoff and sediment load from the Mojo watershed, which has a total area of 2017. 21 km² and they reported that surface runoff and ground water were found to be the most sensitive parameters for runoff. In addition, they found that the model performance was good, with a value of NSE = 0.76 during calibration and NSE = 0.70 during validation for runoff simulation.

In conclusion, with the exception of a few studies, the majority of the studies simulated streamflow on a daily basis. Because daily flow information is just as important as monthly flow information,

especially for flood forecasting and warning, the hydrological model output for daily time steps should be improved.

2.6 Application of Machine Learning in Water Resources

Machine learning (ML) is the collection of methodologies and algorithms that enable computers to automatically improve performance via experience. As such, they show "data-driven" thinking as opposed to "knowledge-driven" reasoning that supports most physical science disciplines (Xu and Liang, 2021). Recent advances in artificial intelligence methodologies and applications have made significant contributions to the modeling and simulation of hydrological processes, climate change, and earth systems.

Deep learning and machine learning approaches, in particular, have been shown to be critical for obtaining improved accuracy, resilience, efficiency, computing cost, and overall model performance (Ardabili et al., 2020; Lange and Sippel, 2020). With the popularity of machine learning and artificial intelligence technology in recent years, mathematical models such as artificial neural networks, fuzzy inference systems, support vector machines, long short-term memory networks, and others have been increasingly used in hydrological simulation research. Simultaneously, society and humans are paying increasing attention to the river basin's water environment (Gui et al., 2021). For this study, the most popular form of machine learning algorithm, ANN, was chosen.

2.6.1 Artificial neural network (ANN)

The artificial neural network (ANN) is one of the most successful algorithms in machine learning (ML) and a data-driven modeling approach that is becoming increasingly popular. This paradigm is inspired by how the human brain processes information, which is done by a biological neural system that accomplishes a specific task that it has been trained to do by handling a number of incoming signals, processing them, and outputting the result (Price and Solomatine, 2009). Even if the rainfall–runoff mechanism is seen as a "black box" system by neurohydrologists, the ANN is capable of dealing with noisy and ambiguous data. On the other hand, ANNs are often less

expensive and easier to implement than their physical-based counterparts, and they are also well suited to dealing with dynamic challenges (Jimeno-Sáez, 2018).

2.6.1.1 Multilayer Perceptron (MLP)

There are several types of ANN, including feed-forward neural networks (FFNN) or multilayer perceptron neural networks (MLPNN), generalized feed-forward neural networks (GFNN), radial basis function neural networks (RBFNN), and adaptive neuro-fuzzy inference systems (ANFIS), but FFNN with back-propagation or multilayer perceptron neural networks (MLPNN) are used to solve complicated problems by modeling complex input-output relationships (Daniel et al., 2011). A MLP is a mathematical model that simulates the functioning of the human brain. MLP neural networks are composed of many linked neurons that may mimic brain synapses. Each neuron-to-neuron connection has a weight that indicates the degree of output influence at the input to the following neuron. Back Propagation (BP) is a common technique for reducing the error of MLP neural networks. BP technique, the error per training pair is estimated and also applied to adjust the weights of the neurons to fit the wanted output (Gunathilake et al., 2021; Choong et al., 2020). According to Behmanesh and Ayashm (2015), the primary benefit of employing a back propagation algorithm is:

- i. Adaptive learning, which means self-correction of the network,
- ii. Self-organization means that the neurons will adapt themselves to the new regulations and respond to the input will change,
- iii. Immediate performance: this means that computations can be conducted in a parallel manner,
- iv. Error tolerance: which means the overall performance of the model remains unscathed if a small part faces an error,
- v. Grouping: this is the ability to categorize the inputs in order to find an appropriate output,
- vi. Generalization, which is the ability of the model to generate a general rule by facing only a few species and expanding the results to new inputs, and
- vii. The sustainability and flexibility of the technique.

2.6.1.2 Transfer functions

In a neural network, the only processing function is the transfer function (also known as an activation function). It is used to limit the amplitude of a neuron's output. The selection of data transfer functions with appropriate weights from input to hidden and from hidden layer to output layer is a critical issue in artificial neural network modeling (Gharde et al., 2016). Sigmoid nonlinearity is a common transfer function for MLP. It is given by the following equation:

$$f(x) = \frac{1 - e^x}{1 + e^x} \dots \dots \dots (2.10)$$

This type of transfer function takes the input data can be in the range of plus infinity to minus infinity, and it converts the output into the range -1 to 1. Another type of popular activation function is Rectified Linear Unit Activation Function (ReLU) this type of activation function is responsible for transforming the weighted input that is summed up from the node to the strict output or proportional sum. These functions are piecewise linear functions that usually output the positive input directly; otherwise it the output is zero. ReLU function is the most widely used function and performs better than other activation functions in most cases.

$$f(x) = x^+ = \max(0, x) \dots \dots \dots (2.11)$$

According to Szandaa's (2021) review, there is no absolute rule for selecting the activation function, and the activation functions can adjust to the process better or worse because they can improve the learning of specific data patterns. Figure 2.1 illustrates an overview of various activation functions, the mathematical formula, the models to use them in, and a graphical representation.

The effective application of artificial neural networks (ANNs) to hydrological modeling involves a number of important decisions. For example, the types of networks and training algorithms are constantly changing, and some of these operations may be automated, but many decisions must still be made through a process of trial and error (Jimeno-Sáez et al. 2018). However, nowadays, researchers are successfully applying ANN to not only predict hydrological variables but also to predict the error of hydrological models (Sikorska-Senoner and Quilty, 2021).

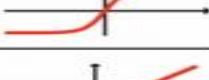
Activation Function	Equation	Example	1D Graph
Linear	$\phi(z) = z$	Adaline, linear regression	
Unit Step (Heaviside Function)	$\phi(z) = \begin{cases} 0 & z < 0 \\ 0.5 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Sign (signum)	$\phi(z) = \begin{cases} -1 & z < 0 \\ 0 & z = 0 \\ 1 & z > 0 \end{cases}$	Perceptron variant	
Piece-wise Linear	$\phi(z) = \begin{cases} 0 & z \leq -\frac{1}{2} \\ z + \frac{1}{2} & -\frac{1}{2} \leq z \leq \frac{1}{2} \\ 1 & z \geq \frac{1}{2} \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multilayer NN	
Hyperbolic Tangent (tanh)	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multilayer NN, RNNs	
ReLU	$\phi(z) = \begin{cases} 0 & z < 0 \\ z & z > 0 \end{cases}$	Multilayer NN, CNNs	

Figure 2.1. Overview of activation functions (Raschka and Mirjalili, 2018)

2.7 Hybrid Modeling

2.7.1 Definition and concept of hybrid modeling

In the last two decades, hybrid modeling has received increasing attention in a wide range of environmental and engineering fields where complex and dynamic systems are prevalent (Nielsen et al., 2021). The objective of using hybrid modeling in many disciplines is to reduce the complexity of the process or to improve prediction accuracy. Various researchers refer to approaches that directly combine a hydrological model with a data-driven model as hybrid models (Konapala et al., 2020), hydrologically-informed machine learning (Herath et al., 2021), and physics-informed hybrid models (Lu et al., 2021). Which are sometimes known as theory-driven or process-driven models (Adombi et al., 2021; Read et al., 2019).

According to Corzo et al. (2009), the term "hybrid modeling" in the context of hydrological modeling refers to an integrated modeling framework that may include several models based on various modeling paradigms. In particular, mechanistic modeling of complex dynamic systems involves the coupling of models based on various existing modeling techniques to varying degrees of complexity, ranging from simple comparison to bridging to total fusion (Parrott, 2011). The

concept of hybrid modeling is not new, but it is very rare in its application to streamflow modeling (Wang et al., 2021).

Hybrid models can overcome issues caused by differences in the temporal and spatial scales of flow processes in the channel, overland plane, and subsurface. The concept of hybrid modeling or model integration is introduced to address some of the issues associated with the fully physics-based representation of all subsystem processes of a watershed while providing a much better and more sophisticated interpretation than an empirically based lumped parameter model (Eslamian, 2014). The small-scale requirements of overland and unsaturated zone flow domains for watershed modeling impose severe constraints on efforts to fully integrate the system. In contrast to the physical-based hydrologic model, which conceptually simulates the rainfall-runoff process of the watershed, the data-driven hydrologic model is built through a process that users do not understand, so prediction errors may occur due to input data with different features. As a result, completely replacing hydrologic modeling with a data-driven model would be difficult. Therefore, a hybrid modeling approach is preferable, in which distributed and lumped parameter models are linked and combined to produce a semi-distributed watershed model (Aral and Gunduz, 2006; Kwak et al., 2022).

2.7.2 Classification of hybrid models

According to the Corzo (2009) classification, hybrid models are divided into three types:

1. Data-driven models with incorporated hydrological knowledge: hybrid models of this type are based on the separation of input space or the identification of processes and output regimes based on hydrological knowledge.
2. Process-based models using data-driven techniques, or with some components replaced by data-driven models to solve complex processes in a physical-based model: This category contains hydrological process models that use data-driven (machine learning) techniques or have some components replaced by DDMs.
3. Data-driven models used in parallel: In this model category, there are two setups: parallel and serial (sequential) model operations. In a parallel scheme, two or more models operate concurrently to solve the same problem, and this architecture is commonly referred to as an

ensemble. In a serial setup, there are also two schemes commonly used: in the first, one model supports the other by feeding information into it (e.g., to update its states); in the second, one model corrects the output of another (error-correcting schemes). In fact, in many operational flow forecasting systems based on hydrological models, the serial setup is implemented in the form of data assimilation to update the model's internal states and/or correct its outputs.

Two sequential hybrid models were tested in this study to improve the performance of streamflow simulation by correcting the SWAT model output with an ANN model.

2.7.3 The state of the art in hybrid hydrological modeling today

There are numerous methods for integrating hydrological models with data-driven models that are addressed in the literature these days. For example, Latifolu et al. (2015) presented a method for forecasting hydrological data for the Kizilirmak River in Kayseri, Turkey, using singular spectrum analysis (SSA), artificial neural networks (ANN), and a hybrid of SSA and ANN. The hybrid SSA–ANN model was a very promising approach for forecasting hydrological data based on statistical parameters. The statistical performance parameters were $MSE = 0.00088$, $MAE = 0.0217$, and $R^2 = 0.986$. When the ANN model and the SSA-ANN model results were compared, the SSA-ANN model predicted 1 month-ahead streamflow data much better than the ANN model.

Ghaith et al. (2020) conducted research in the Guadalupe Basin in Texas using a hybrid of machine learning and HYMOD conceptual hydrological models, and they successfully developed a combined model called HHDD to forecast daily flow. The proposed modeling approach was successful in improving the daily prediction compared with the HYMOD conceptual model, and it was found very helpful in avoiding the complexity of parameters in calibration. The performance of both models was reported using two statistical criteria: the root-mean-square error (RMSE) and the mean absolute error (MAE). For the validation period, the RMSE values of hybrid and conceptual models were reported at 34 and 44 m^3/s , respectively, and the MAE values were 7.5 and 10 m^3/s , respectively.

Kumanlioglu and Fitikoglu (2020) combined an artificial neural network and a genetic algorithm (GA) with a conceptual model, GR4J, to develop a daily rainfall-runoff model in Turkey's Gediz River Basin. By using only future precipitation and evapotranspiration conditions, the developed

hybrid GR4J-ANN-GA model with fewer parameters performed better in predicting future runoff. The main difference between this hybrid model and regular ANN-based models is that runoff observations are not required as model input, and incorporating ANN reduces the number of GR4J model parameters to one.

Farfán et al. (2020) also conducted a study for two rivers in Ecuador's Tropical Andes, where the region's complex climatic variability poses significant challenges in flow forecasting. Two physical models (WEAP and GR2M) and two ANN-based models with meteorological input variables (precipitation and evapotranspiration) were used. Individual performances are analyzed and compared to the results of the hybrid technique using a variety of evaluation criteria, including a supply-demand analysis for the region of interest. The goal of their research was to develop and test a hybrid technique that takes the monthly simulated flow of physical and ANN-based models as input variables and configures a new ANN with the observed flow as the target variable. The proposed technique aims to overcome the individual limitations of physical and data-driven approaches by combining the robustness and hydrological significance of physical models with the black-box predicting abilities of data-driven methods. The results also reveal that the hybrid technique is more consistent than the individual models, presenting similar results in the calibration and validation phases.

The two popular models, a physical-based SWAT and a data-driven ANN, are also successfully integrated and reported in the literature to simulate streamflow.

Mekonnen et al. (2015), for example, used a hybrid of the SWAT and ANN models in their study of the North American grassland region. The findings of their study imply that combining a process-based SWAT model with a data-driven model like ANN can give superior modeling skills for the wet season. The performance of the model for the calibration period was reported in terms of NSE as 0.75, 0.90, and 0.91 for SWAT, ANN, and hybrid models, respectively. For validation, the NSE reported 0.81, 0.89, and 0.94 for SWAT, ANN, and hybrid models, respectively.

Noori and Kalin (2016) used ANNs to forecast daily river flows and incorporated simulated data such as soil moisture, effective rainfall, surface runoff, infiltration depths, and base flow and

stormflow information derived from conceptual models. The researchers tested the ability and efficiency of the ANN model combined with SWAT for daily streamflow prediction in 29 nearby watersheds in and around Atlanta, Georgia, in the Southeastern United States, and discovered that the hybrid SWAT-ANN model performs better than the SWAT model.

Lv et al. (2020) also presented a method for integrating the two models using the suggested improved SWAT-ANN approach for runoff simulation in the Baliqlu Chai Watershed in Iran's Ardabil Province, which showed improved model performance. The models' performance was measured using the coefficient of determination (R^2), relative error (RE), root mean square error (RMSE), and the Nash-Sutcliffe coefficient (NSE), among other metrics. The SWAT model had 0.76, 0.14, 2.19, and 0.74 for calibration and 0.72, 0.35, 2.70, and 0.58 for validation, respectively. For hybrid SWAT-ANN, the performance values were 0.80, 0.11, 1.61, and 0.79 for the calibration period and 0.78, 0.44, 2.45, and 0.64 for the validation period, respectively.

2.7.4 Post-processor framework

Hydrological model errors affect predictions of design floods, risk assessment, and water resource management in general and are caused by a variety of sources of uncertainty, including model structure, input and output variables, parameters, initial circumstances, and so on (Farmer and Vogel, 2016). The difficulty in correctly estimating the geographical and temporal variability in hydrologic modeling inputs across broad modeling domains is a major contributor to errors (Shrestha and Solomatine, 2009). Although hydrological model errors have traditionally been considered to be independent and normally distributed, it is now widely understood that this assumption is no longer valid since hydrological model residuals are significantly auto-correlated (Sikorska et al., 2012; Sikorska-Senoner and Quilty, 2021). As a result, they cannot be represented as a regularly distributed random variable and alternative methods in hydrology that do not rely on these limiting assumptions have attracted extensive attention.

Several studies have used data-driven models to improve the performance of hydrologic models by predicting and reducing model uncertainty. Abebe and Price (2003), for example, presented a complementary framework for analyzing uncertainty in hydrologic models using an ANN model. Their research looked at the relationships between precipitation and model error in the runoff, and

they discovered that using a data-driven model (e.g., an ANN model) for error estimation in hydrologic models can significantly improve model performance.

Kwak et al. (2021) also used a combination of two models to predict runoff in the Hyeongsan River basin in South Korea: the long-short-term memory model (LSTM) and the TANK model by modeling the residual using LSTM. When the performance of each model was compared, they discovered that the hybrid model outperformed the separate models. As a result, they concluded that relying solely on deep-learning model results for applications that directly affect people's lives, such as flood and drought forecasting or water resource management, is insufficient and misleading unless supplemented with other physical-based models.

Wang et al. (2021) have presented a hybrid model that employs the output of the Xinanjiang hydrological model along with wavelet-decomposed sub-series of earlier streamflow observations as input to Random Forests (RF) for simulating streamflow residual error. Sikorska-Senoner and Quilty (2021), in a study conducted in three Swiss catchments, found that the residuals from an ensemble of hydrological model simulations were changed by a data-driven model for each hydrological model simulation, resulting in improved ensemble streamflow simulations. The study demonstrated that utilizing a nonlinear data-driven model to simulate hydrological model residuals and then incorporating the output into hydrological model simulations can result in more accurate streamflow simulations than using a stand-alone hydrological model. Furthermore, Quilty et al. (2022) proved the effectiveness of this approach by creating the stochastic version, which accounts for input data, input variable selection, and model output uncertainty in addition to parameter uncertainty. In this study, the residual error of the physical-based SWAT model was reduced by using the ANN model.

2.7.5 Previous works in the study area

In Ethiopia, specifically in the study area, there is no direct study conducted to apply the hybrid methodology for streamflow simulation improvement, but some studies related to the topic are presented as follows: Belayneh et al. (2016) investigated five drought forecasting models for short lead times in the Awash River Basin using a hybrid technique. Traditional drought forecasting methods such as ARIMA models were compared with machine learning techniques such as ANN

and support vector regression (SVR), as well as hybrid models (WA-ANN and WA-SVR) that preprocess input data using wavelet analysis (WA). The forecast results of this study show that the hybrid wavelet neural network (WA-ANN) models were the best models for forecasting SPI 3 and SPI 6 values in the Awash River Basin over lead times of 1 and 3 months.

Gleason et al. (2018) conducted a study by combining remotely sensed discharge estimates made via at-many-station hydraulic geometry (AMHG) and the PCR-GLOBWB hydrological model to estimate discharge over the Lower Nile. First, they estimated initial discharges from 87 Landsat images and AMHG, and then used these flow estimates to tune the model. The resulting tuned modeled hydrograph shows a large improvement in flow magnitude: validation of the tuned monthly hydrograph against a historical gauge yields an RMSE of 439 m³/s (40.8%). Based on the results obtained, they demonstrated that coupling physical models and remote sensing is a promising proof of concept toward future modeling.

Quilty and Adamowski (2021) also presented the hybrid wavelet decomposition settings method of maximal overlap discrete wavelet packet transform (MODWPT), which enables the integration of multiple wavelet decomposition within the wavelet data-driven forecasting framework for the purpose of forecasting monthly rainfall at six meteorological stations in the Awash River Basin (Ethiopia), and the performance of each individual method was compared with the hybrid one. Finally, they concluded that the integrated MODWPT is a promising new approach for monthly rainfall forecasting, and it is expected to be useful for forecasting other hydrological variables such as streamflow and temperature.

As a final point, the literature supports the hybrid modeling approach's high performance in modeling streamflow in various parts of the world, with the suggestion that integrating different models will improve the accuracy of streamflow simulation. However, according to the review, no studies have been reported to improve the accuracy of streamflow simulation using a hybrid modeling approach in Ethiopia, particularly in the Upper Awash Basin, where there is frequent high water stress and flooding. Therefore, having accurate and relevant streamflow information is critical for decision-makers and water resource managers. As a result, the applicability of a hybrid of SWAT and ANN models in the Upper Awash Sub-basin was evaluated in this study.

CHAPTER THREE

3. MATERIALS AND METHODS

This chapter addresses the study area, materials and software used, needed data, and data quality analysis. Furthermore, a comprehensive methodology of SWAT, ANN, and a hybrid model for streamflow simulation, as well as hydrological residual modeling, were described. Finally, the framework for the overall methodology was designed and presented.

3.1 Study Area Description

3.1.1 Location

The Awash River Basin is a significant development area that is at the forefront of Ethiopia's development practices, particularly in irrigation. The basin begins at an elevation of about 3000 m a.s.l. in the central high plateau near Ginchi town west of Addis Ababa and flows for 1200 km to join Lake Abe at an elevation of about 250 m a.s.l (Tessema, 2011). The Upper Awash Sub-basin is situated in the western part of the Awash River Basin in the Oromia Regional State, and the sub-basin covers a total area of 7656 km² at the Hombole hydrometric station. It is geographically located between longitudes 37°50'10"E and 39°30'20"E, and latitudes 8°18'45"N and 9°20'7"N, and includes the capital city of Addis Ababa (Figure 3.1).

3.1.2 Climate

The climate of the Upper Awash Sub-basin ranges from humid to sub-humid in the highlands and from semi-arid to arid in the Rift Valley and is influenced by the Intertropical Convergence Zone (ITCZ). The convergence of dry tropical winter winds and humid equatorial westerly winds is marked by this low-pressure area. The annual migration of the ITCZ determines the seasonal distribution of rainfall in the sub-basin. In March, the ITCZ begins its trans-basin journey from the south, bringing with it light or spring rains. In June and July, it reaches the northernmost point outside the basin, where it experiences heavy rainfall for the rest of the year. During August, September, and October, the ITCZ migrates southward, restoring a drier easterly air stream that lasts until March, when the ITCZ migrates northward again. Orographic effects tend to change the

distribution of rainfall over highland areas (Fufa, 2016). Annual rainfall in the plains ranges from 850mm to 1000mm, with higher amounts in the mountains of the Upper Awash Sub-basin. In the highlands, the average annual temperature is around 15 °C, while in the lowlands, it is around 21 °C (Dilnesaw et al., 2007).

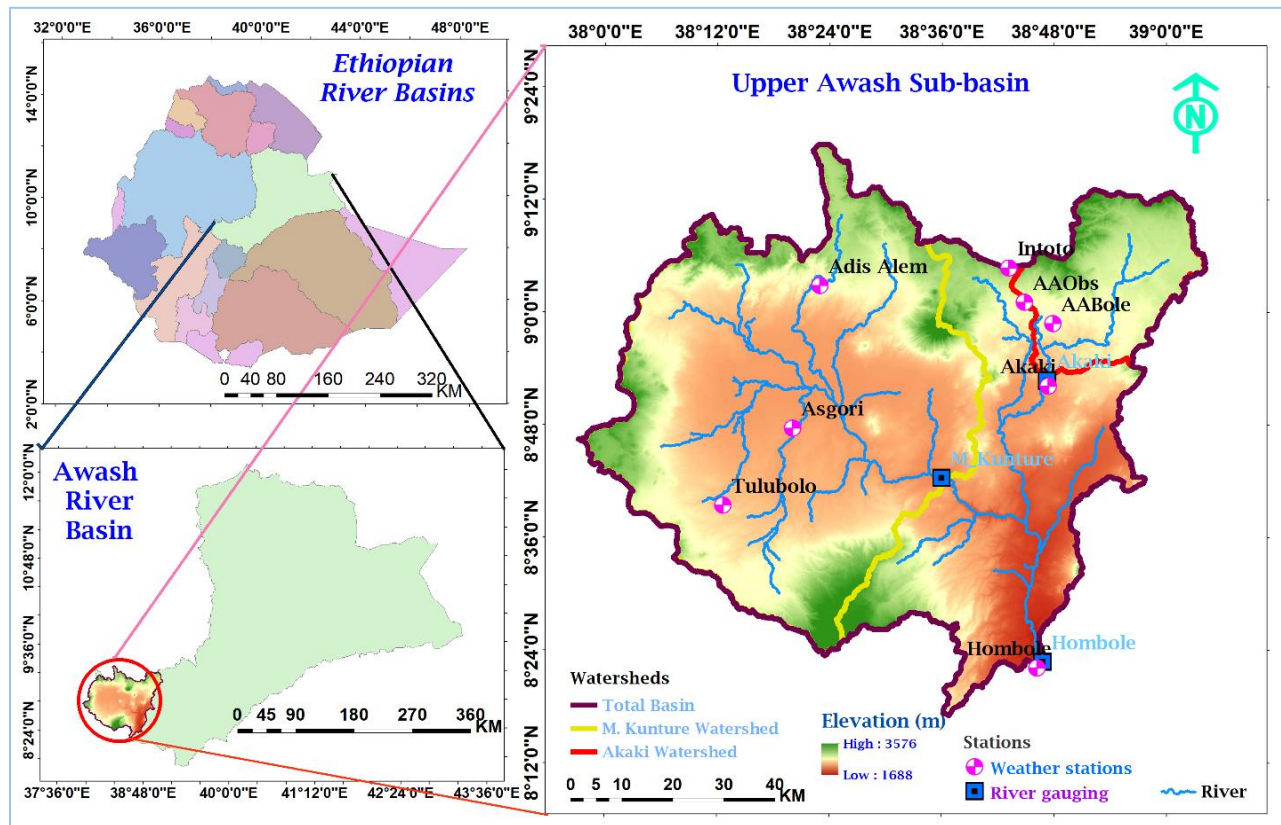


Figure 3.1. Location of the study area

3.1.3 Topography

The average altitude of the Upper Awash Sub-basin is 2632 meters above mean sea level and ranges from 1688 to 3576 meters a.m.s.l. The high plateau is located to the west of Addis Ababa. There are large elevation drops from the western and northern parts of the Sub-basin to its southern side. The total drop in elevation is above 1888 meters from north to south at the Hombole hydrometric station.

3.2 Data Collection

To achieve the goal of the study, various types of data were collected and used , including land use and land cover data, a digital elevation model (DEM), soil data, weather data, and streamflow data. The description of the data and the analysis are provided in the following section.

3.2.1 Weather Data

The weather of the basin represents the input of water and energy that controls the water balance and determines the relative importance of the various components of the hydrological cycle. (Neitsch et al., 2011).

The SWAT model requires daily precipitation, maximum and minimum temperatures, solar radiation, wind speed, and relative humidity as meteorological variables. This data was gathered from the Ethiopian Meteorology Institute (EMI). Weather data from Sub-basin stations (Addis Ababa Bole, Addis Ababa Observatory, Asgori, Hombole, Tulubolo, Akaki, and Intoto) was collected from 2000 to 2019, but the most recent record of streamflow data obtained from the Ministry of Water and Energy (MoWE) was only up to 2015. So climate data from 2000 to 2015 was considered only. The description of weather stations included in this study were all located in the Upper Awash Sub-basin given in appendix A1.

3.2.2 Hydrological data

Streamflow data was used to calibrate the SWAT model and train the ANN model in this study because it is critical for hydrological modeling, particularly model calibration and training. The Awash basin has 76 river gauging stations, 21 of which are located in the Upper Awash Sub-basin, particularly upstream of Koka Dam and seven of which are equipped with automatic water level recorders, including Bello, Melka Kunture, Melka Hombole, Mojo, and Akaki (Mengistu, 2008; Sileshi Gobena, 2010).

The Ministry of Water and Energy (MoWE), which is in charge of collecting and disseminating hydrological data, provided daily flow data from the Upper Awash Sub-basin for 16 years (2000-2015) at three distinct locations representing the outlet points of two watersheds: Akaki, M.

kunture (located within the sub-basin), and Hombole (representing total sub-basin gauge sites). The three gauging stations that were chosen for multi-site calibration in this study represent different watersheds with different characteristics.

Figure 3.2 indicates the peak flow occurs from June to October and minimum flow occurs from November to May at three gauging stations

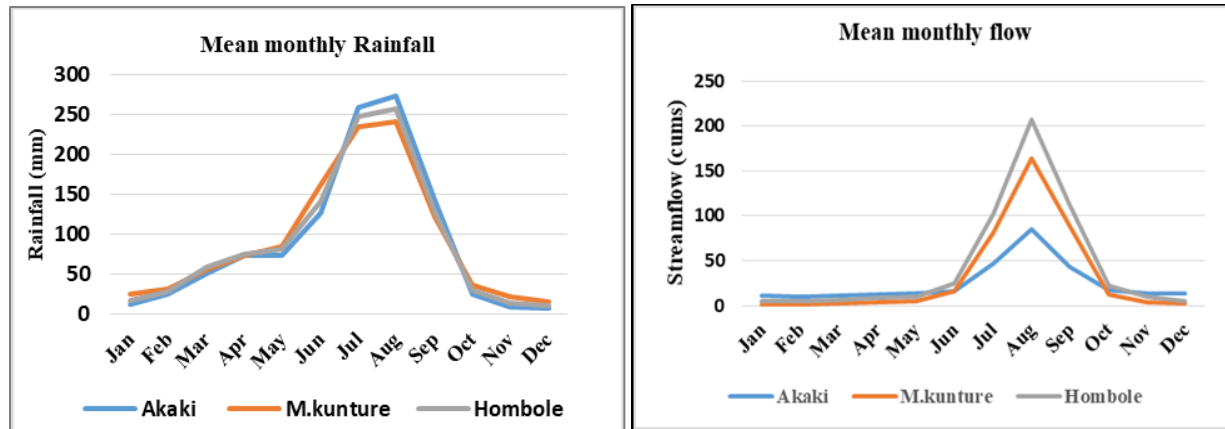


Figure 3.2. Monthly flow and rainfall characteristics of gauge stations under the study area

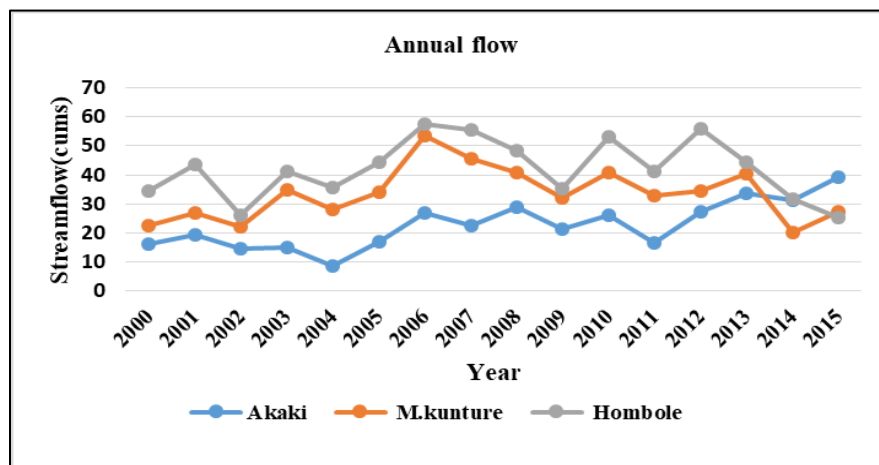


Figure 3.3. Annual flow characteristics of gauge stations under the study area

Table 3.1. Streamflow gauging stations

SN.	Station name	Longitude	Latitude	Area (km ²)	Length of streamflow data
1	Akaki	38.47	8.53	884.4	16 years (2000-2015)
2	Melka Kunture	38.36	8.42	4456	16years (2000-2015)
3	Hombole	38.47	8.23	7656	16 years (2000-2015)

Source: MoWE, 2022

3.2.3 Digital elevation model data (DEM)

A digital elevation model (DEM) is any digital representation of a topographic surface, and it is specifically made available in the form of a raster or regular grid of spot heights. A 30 m × 30 m resolution. For watershed delineation, a DEM of the study area was obtained from the Earth Explorer website (<https://earthexplorer.usgs.gov/>). Furthermore, it is used for the land slope class classification, which was the basic parameter for the Hydrological Response Unit (HRU) definition. The different land slope classes are shown in Figure 3.4.

The slope class between 1–5% and 5–10% is the majority of the land slope class of the sub-basin, which covers 39.6% and 29.75%, respectively. The remaining land slope classes (10–15%, 15–25%, and greater than 25%) cover 12.5%, 10%, and 7.35% of the watershed, respectively. These different land slope classes show that the majority of the Upper Awash Sub-basin area has been dominated by a flatter surface.

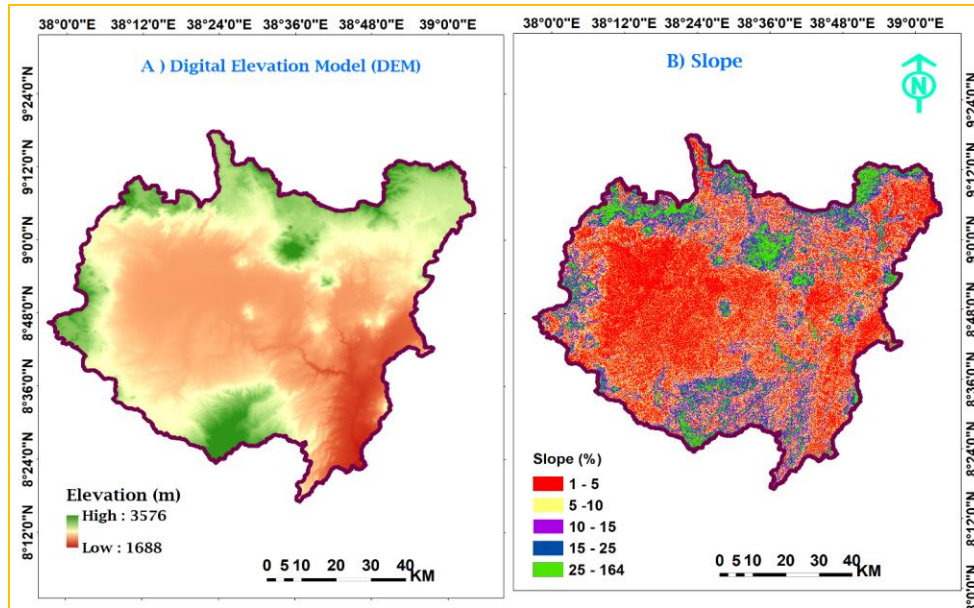


Figure 3.4. Digital Elevation and Slope map of the Upper Awash Sub-basin

3.2.4 Land Use and Land Cover (LULC) data

Land use and land cover change affect the hydrological cycle components such as surface runoff, evapotranspiration, percolation, and infiltration. As a result, LULC is considered one of the significant inputs for the SWAT model to describe the hydrological response units (HRUs) of the watershed along with the slope and soil map.

Image Processing

To make land use and land cover (LULC) data in line with climate and streamflow data, the LULC data for the year 2008 was derived from Landsat 7 Enhanced Thematic Mapper Plus (L7-ETM+) at 30 m resolution and obtained from the United States Geological Survey (USGS) earth explorer website (<http://earthexplorer.usgs.gov/>) in GeoTIFF file format. The selection of the Landsat satellite image data was influenced by the quality of the image, caused by solar, radiometric and atmospheric condition. Therefore, to prevent seasonal variation of vegetation coverage and to avoid error with cloud cover the image was almost selected in a cloud-free season.

Preprocessing satellite imagery before image classification is critical for minimizing instrumental errors and building a more thorough association between the obtained data and ground surface features. Before classification, ERDAS IMAGINE 2015 was used for layer stacking, mosaicking different layers of satellite image into one layer, radiometric correction such as noise reduction, histogram matching, periodic noise removal, and distrip, TM data, and masking the area of interest with the help of a shape file. The pre-processed satellite image was then classified by supervised classification using the maximum likelihood algorithm in the ERDAS IMAGINE 2015 environment. As a result, seven categories based on the nature and similarity of the land use and cover types were identified.

Table 3.2. Land use and land cover categories identified at the study area

Land use land changes types	General description	SWAT code
Agriculture	Areas covered with annual crops followed by harvest and bare soil period	AGRL
Forest land	Areas composed of forest land and transplanted	FRSE
Shrubs land	This category includes low woody plants, generally less than three meters in height, usually with multiple stems, and grow vertically.	RNGB
Grassland	Landscapes that have a ground story in which grasses are the dominant vegetation forms	RNGE
Bare land	Areas with little or no vegetation cover consisting of exposed soil and/or bedrock	BARR
Urban	Where the dominant area covered by building and houses	URBN
Water body	Permanent open water, lakes, reservoirs and streams	WATR

Accuracy assessment

The development of a classification error matrix (also known as a confusion matrix or a contingency table) is one of the most frequent ways of displaying classification accuracy. Error

matrices compare the relationship between known reference data (ground truth) and the related results of automated categorization on a category-by-category basis. This was done based on the ground truth in ERDAS Imagine 2015 and Google Earth Pro 7.3.3 software for the LULC map of 2008. The accuracy of classification was measured using measures generated from the error matrices, such as user, producer, and overall accuracy, as well as the kappa coefficient.

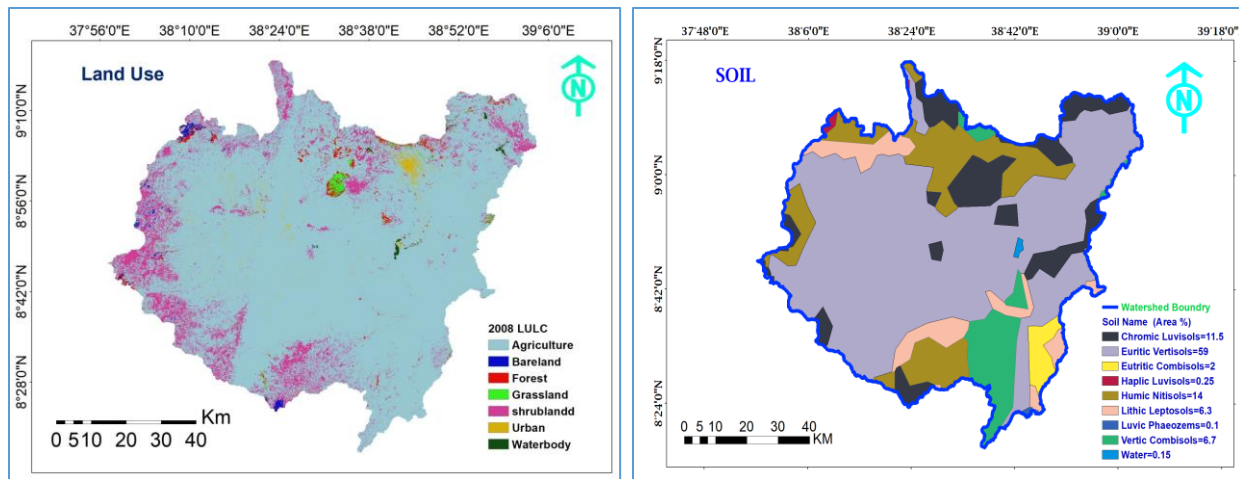


Figure 3.5. Land use and land cover and soil map of the Upper Awash Sub-basin

3.2.5 Soil data

The SWAT model requires soil data from the study area, which includes physical and chemical properties such as soil texture, available water content, hydraulic conductivity, bulk density, and organic carbon content. For this study, the Harmonized World Soil Database (HWSD) was used, and the soil raster image and the database were obtained from HWSD. Accordingly, vertisols and livisols were found to be the two dominant soil groups in the study area. Consequently, a user-defined soil database was prepared for each soil layer and added to the SWAT user soil database.

3.3 Data Analysis

3.3.1 Completing missing data

Missing data can disrupt the continuity of recorded data due to a variety of factors, including damage or a fault in the gauging station during the measuring period. As a result, before beginning

any model simulation, ensure that the data is homogeneous, consistent, sufficient, and complete, with no missing data. Missing climate data, especially precipitation data, can cause a number of issues in hydrologic analysis and design. Because of the high cost of data collection, as well as some natural and man-made factors, keeping complete records of data at each station can be difficult. The above-mentioned conditions can sometimes make it difficult to collect quantitative and qualitative data from the study area. Failure or absence of the observer to make the necessary visit to the gauge, destruction of recording gauges, and instrument failure due to mechanical or electrical malfunctioning can result in missing data for gauges that require periodic observation. All of these causes of instrument failure reduce the precipitation record's length and information content.

There are several techniques available to fill in missing data, including the arithmetic mean method, normal ratio method, and invers distance weighted methods. The appropriate method is chosen based on the percent of missing data and the number of stations. To select the appropriate method for filling in the missing data, first the missing percent of the selected stations was computed using equation (3.1). Although the percent of missing stations is less than ten percent, which allows us to use the arithmetic mean method, the normal ratio method was found to be sounder than the arithmetic mean method. Therefore, the normal ratio method was used to fill in the missing data at all stations.

$$\text{Percent (\%)} \text{ of missing data} = \frac{X}{Y} \times 100\% \dots \dots \dots (3.1)$$

Where, X is count of missed data and Y is whole data set

$$P_x = \frac{1}{n} \sum_{i=1}^{i=n} \frac{N_x}{N_i} P_i \dots \dots \dots (3.2)$$

Where P_x is the missing precipitation for any storm at the interpolation station 'x', P_i is the precipitation for the same period for the same storm at the "ith" station of a group of index stations, N_x is the normal annual precipitation value for the "x" station, and N_i is the normal annual precipitation value for the "ith" station. The percentage of missing data for each meteorological station was presented in Appendix A2.

3.3.2 Test for data consistency and homogeneity

The consistency and homogeneity of recorded rainfall data were checked after missing rainfall data was filled in. The theory that a plot of two cumulative quantities measured for the same time period should be a straight line with their proportionality unchanged, which is represented by slope, was used to analyze the consistency of time series data. As a result, the double-mass curve technique was used to correct the record's inconsistency. This method is based on the idea that when data is collected from the parent population, it is consistent. To account for non-representative factors such as changes in location or rain gauge exposure, the double mass curve technique was used to adjust precipitation records. From the double mass curve analysis, almost all meteorological stations in the study area show a relatively constant result.

Homogeneity analysis was used to separate a change in the statistical properties of the time series data. The causes can be either natural or man-made. These include alterations to land use and the relocation of the observation gauging station. The Rainbow software package was used for checking homogeneity tests. In RAINBOW, the test for homogeneity is based on the adjusted partial sums or cumulative deviations from the mean (Buishand, 1982). The rainfall and streamflow data have been tested for homogeneity at different significance levels and they are found to be generally homogeneous. The double mass curve and homogeneity test results are presented in Appendices A3 and A4, respectively.

3.3.3 Rainfall data

From the rainfall data analysis, bi-modal rainfall patterns were detected, which means In the Upper Awash Sub-basin, there are two rainy seasons, such as heavy rainfall from June to August and low rainfall from March to April. The watersheds receive maximum rainfall from June to September, about 74% of the annual rainfall, and the second rainy period covers the period from February to May. Figures 3.6 and 3.7 indicate the 16-year long-term average monthly and annual rainfall patterns of the selected stations in the study region, respectively. The highest records are observed in July and August, whereas the lowest records are observed in November and December.

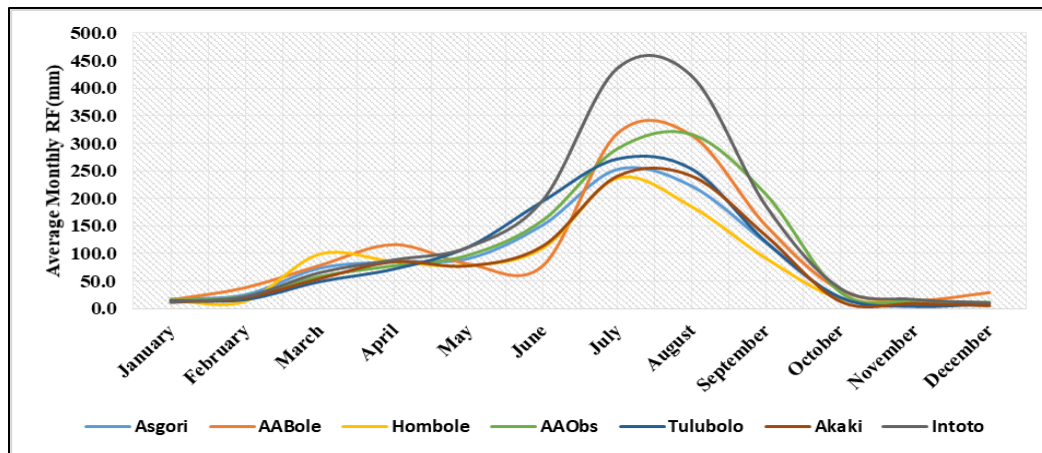


Figure 3.6: Long-term average monthly rainfall pattern of the selected stations

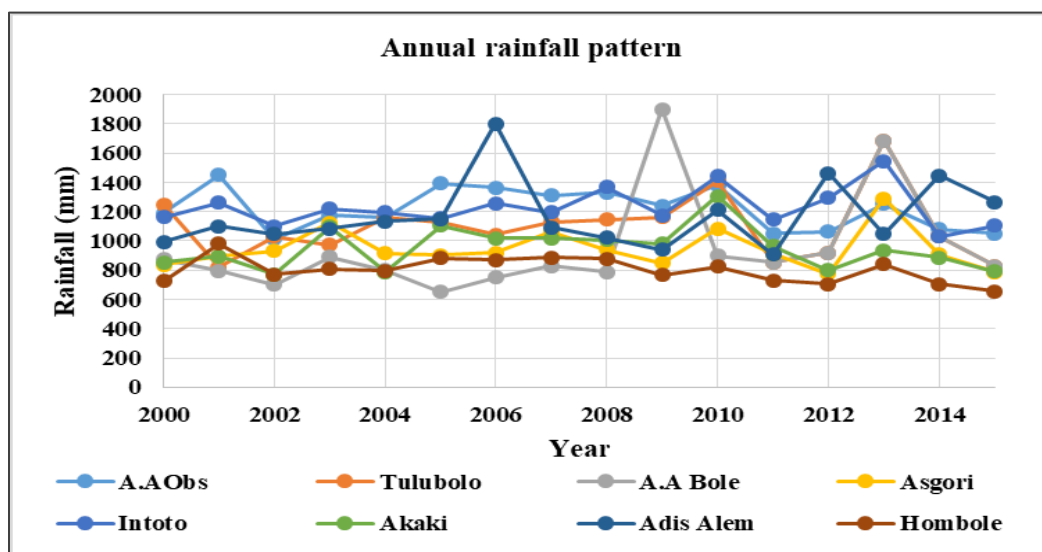


Figure 3.7. Annual rainfall pattern of the selected stations

3.3.3.1 Weighted area rainfall

In certain basins or watershed areas, rain gauge stations distributed in the basin or watershed represent the point measurement. Therefore, rain from the station in the basin may be different from the second station in the same basin area. As a result, we must estimate areal precipitation for the entire study area. Rain gauge measurements can usually be used in three ways to calculate the area of precipitation over a catchment. The arithmetic mean, the Thiessen polygon, and the isohyetal method are three of these methods. Because of its solid theoretical foundation and the

availability of computational tools, the Thiessen polygon method was chosen for this study. The method, however, is reliant on a well-connected network of representative rain gauges and does not allow analysts to account for topographic influences (Brown et al., 2012). The Upper Awash Sub-basin's Thiessen polygon map, as well as the two watersheds, Akaki and Melka Kuntura, were created, and the representative area of each station was calculated using Arc GIS, followed by the areal precipitation. Using equation 3.3.

$$P_{Avg} = \frac{P_1A_1 + P_2A_2 + P_3A_3 + \dots + P_nA_n}{A_1 + A_2 + A_3 + \dots + A_n} \dots \dots \dots (3.3)$$

Where P1, P2 and Pn are rainfall records recorded by the stations 1, 2, and 3 respectively, and A1, A2 and A3 are respective areas of the Thiessen polygons, and n is the number of stations.

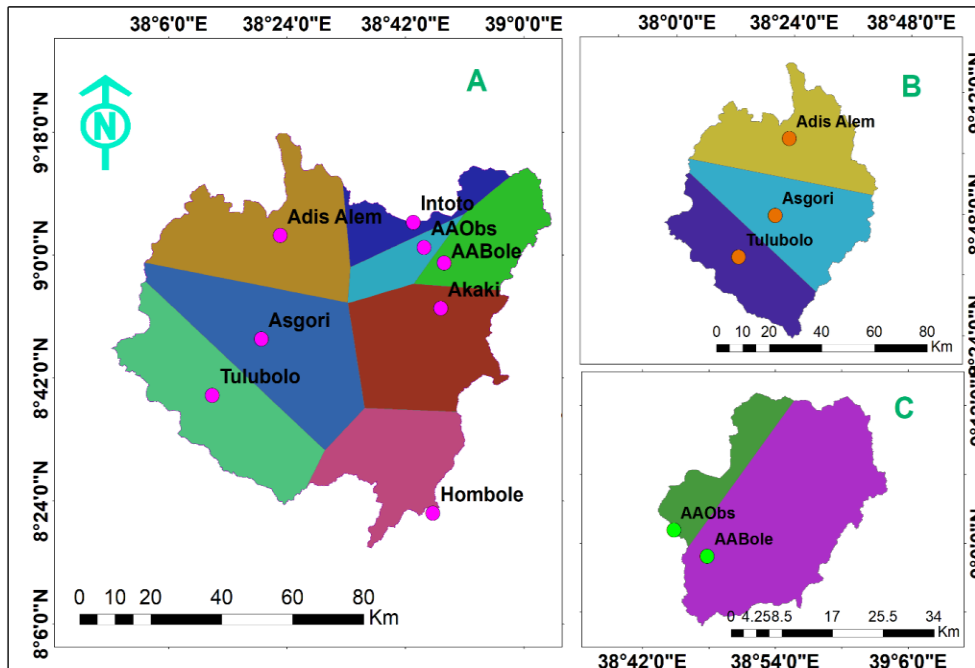


Figure 3.8. Thiessen polygon maps for Hombole (A), Melka Kuntura (B), and Akaki (C) watersheds

3.3.3.2 The correlation between rainfall and streamflow

According to Arnold et al. (1998), using a single gage to represent an entire sub-watershed or even using area-weighted methods for rainfall representation can result in significant errors in runoff

estimation. The correlation coefficient between precipitation data and observed streamflow, on the other hand, may aid in explaining model performance in different sub-watersheds.

Table 3.3. Correlation coefficients between the observed rainfall and streamflow

SN.	Rainfall Station name	Akaki (Q)	M.Kunture(Q)	Hombole(Q)
1	Addis Ababa Observation	0.35	0.37	0.37
2	Addis Ababa Bole	0.3	0.35	0.38
3	Hombole	0.21	0.23	0.22
4	Asgori	0.26	0.31	0.29
5	Tulubolo	0.27	0.33	0.31
6	Akaki	0.31	0.34	0.33
7	Intoto	0.4	0.43	0.44
Areal Rainfall	Akaki area weighted	0.45	0.47	0.48
	M. Kunture area weighted	0.33	0.54	0.39
	Hombole area weighted	0.45	0.5	0.56

3.3.4 Temperature

Temperature varies greatly depending on altitude within the sub-basin. The mean annual maximum temperature of the study area varies from 18.4°C to 28.14°C, while the mean minimum temperature varies from 7.8°C to 11.5°C. The mean annual temperature of the basin, estimated according to the data of the seven stations, varies from 14.2°C to 19.6°C.

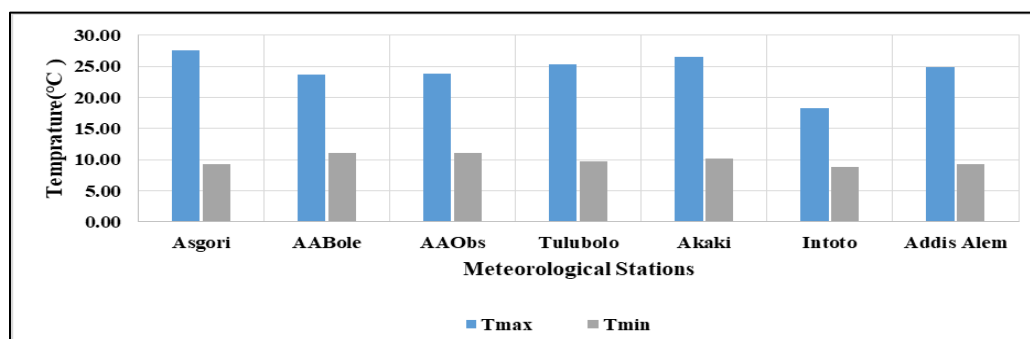


Figure 3. 9. Average annual maximum and minimum temperature of the selected stations

3.4 Materials and Software's

The Geographical Information System (GIS), the ArcGIS-interfaced Soil and Water Assessment Tool (ArcSWAT-2012), and SWAT-CUP, as well as Microsoft Excel and a Python-based tensor flow machine learning platform, were used in this study. The RAINBOW software was used to test the homogeneity of climate data. ArcGIS software was used to prepare spatial data. Googleearth-win-pro-7.3 and ERDAS Imagine 2015 are used for LULC image classifications. The SWAT model was used to simulate streamflow, and the SWAT-CUP software package was used for model sensitivity, calibration, and validation. The Python programming language was used as the main tool for building and training neural networks, like the Keras package that is included in TensorFlow v2.

3.5 Methodology

3.5.1 SWAT hydrological model setup

SWAT is a semi-distributed model that can account for the heterogeneity of a watershed by dividing it into sub-watersheds based on the river network and terrain; sub-watersheds are then divided into hydrologic response units (HRUs), which group land regions with distinct soil, land cover, and slope combinations. SWAT uses water balance to model the hydrologic cycle, which is influenced by climatic variables such as daily precipitation and maximum and minimum air temperatures (Jimeno-Sáez et al., 2018). The SWAT model was used in this study to set up a SWAT model for the Upper Awash Sub-basin by integrating it with a GIS that combines various spatial data (LULC, DEM, slope, and soil) with climate data, which is currently embedded in an ArcGIS interface called ArcSWAT-2012.

3.5.1.1 Watershed delineation

The first step in creating SWAT model input is delineating the watershed from the DEM. The SWAT model organizes the inputs into spatially distinct groups. Before going into a hand with spatial input data, the soil map, LULC map, and DEM were all projected into the same projection called UTM Zone 37N, which is a projection parameter for Ethiopia. A watershed is divided into sub-basins for modeling purposes.

The five major steps in the watershed delineation process are:-DEM setup; stream definition; outlet and inlet definition; watershed outlet selection and definition; and calculation of sub-basin parameters. In the beginning, the downloaded DEM was projected and imported into the SWAT project, and then flow direction and flow accumulation were computed based on the DME. Following the completion of flow direction and accumulation, stream network functional units were also created. The threshold-based stream definition option was used to define the minimum size of the sub-basins for stream definition, and in the outlet and inlet selection steps, the locations of the three selected hydrometric gauging stations (Akaki, Melka Kunture, and Hombole) were manually added during the model setup process. The Upper Awash Sub-basin was separated into 27 small sub-basins with a total drainage area of 7654 km², out of total 27 sub-basin 3 (three) were located inside the Akaki Watershed and 19 (nineteen) were located inside the Melka Kunture Watershed, as presented in figure 3.10.

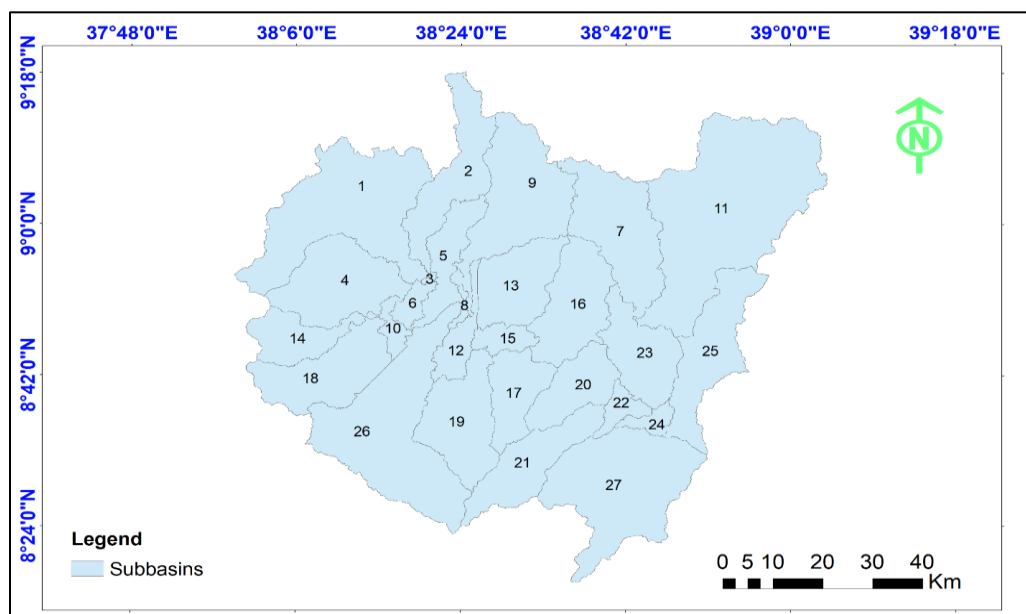


Figure 3.10. Watershed delineated sub-basins of the Upper Awash Sub-basin

3.5.1.2 Hydrologic response unit (HRU) determination

The unique occurrences of soil type, land cover, and slope class are defined as HRUs in SWAT. Any land parcels within a sub-basin with the same combination of these three characteristics will be considered one HRU. SWAT models all processes for each individual HRU in the watershed,

regardless of sub-basin location. For simplicity and time savings during HRU parameterization, multiple HRU slope discretization was selected. To eliminate minor land uses, soils, or slope classes in each sub-basin, a threshold level was used in multiple HRU definitions. Land uses or soils with less than the required coverage are eliminated. Then, the Awash Sub-basin was divided into 265 HRU, with each having unique land use, slope, and soil combinations.

3.5.1.3 Weather data input

Once HRU analysis is complete, the weather data is used to simulate watershed hydrology. The data was prepared and imported into the SWAT project with Six (6) weather variables, including weather generator, rainfall, wind speed, temperature, solar radiation, and relative humidity data, are available in the weather data definitions. The all-weather parameters of the station were prepared in text file format for the SWAT model. The Addis Ababa observatory station contains all the weather parameters; therefore, it was used as the principal station for generating weather parameter values for the remaining stations. Then, the weather generator data file (WGEN_user), rainfall, temperature, solar radiation, humidity, and wind speed were loaded into the model, respectively, for the weather generator database definition.

3.5.1.4 SWAT Simulation

After the completion of loading all the necessary data into the SWAT project, the database files were constructed that held the information required to produce the default input for the SWAT model. Using a modified soil conservation method, the SWAT model was used to simulate the Awash Sub-basin streamflow in a daily time step. Since sub-daily rainfall data for the Green and Ampt infiltration methods are not available, the modified Soil Conservation Service curve number equation (USDA, 1972) was utilized to estimate the catchment's surface runoff.

The Penman-Monteith method for potential evapotranspiration was selected. This is because this method uses all accessible weather data; the more data it takes, the more accurate the result. According to the SWAT user's manual, a warm-up period of a minimum of 3-5 years is recommended. In this study, a three (3) year warm-up period was used in order to reduce the effects of the initial condition of the model. Finally, a model simulation run was conducted on a daily time step for 16 years, from 2000 to 2015.

3.5.1.5 Sensitivity analysis

Sensitivity analysis is a technique for determining the sensitivity of various parameters in a hydrological process simulation. The change in model output per change in parameter input is known as "model sensitivity" (Arnold et al., 2012). The most sensitive parameters for any hydrological model, such as SWAT, have been determined for model calibration and validation purposes. There are two types of sensitivity analysis in the SWAT-CUP algorithm. These are one-at-a-time (OAT) and global sensitivity analyses. One-at-a-time sensitivity analysis works by varying each parameter while holding all others unchanged, whereas a global sensitivity analysis involves varying all parameters at once (Abbaspour, 2015). Generally, eleven (11) most sensitive parameters were selected from previous studies done in the Upper Awash Sub-basin (Gessesse et al., 2014; Shawul et al., 2019).

Table 3.4. Flow Parameters for calibration and validation of the SWAT model

No	Parameter	Description	Max	Min
1	r__CN2	Runoff Curve Number	35	98
2	v__ALPHA_BF	Base flow alpha factor (days)	0	1
3	v__GW_DELAY	Groundwater delay (day)	30	450
4	a__GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur	0	25
5	r__SOL_AWC	Available water capacity of a soil	-0.5	0.5
6	v__REVAPMN	Threshold depth of water (mm)	0	500
7	a__ESCO.hru	Soil evaporation compensation factor	0	0.2
8	r__SLSUBBSN	Average slope length	0	0.2
9	r__HRU_SLP	Average slope steepness	0	0.2
10	v__SURLAG	Surface runoff lag coefficient	0	10
11	r__OV_N	Manning's "n" value for overland flow	-0.2	0

3.5.1.6 Calibration of model

Model calibration is the process of estimating model parameters by comparing model predictions for a given set of assumed conditions with observed data for the same conditions. The calibration of the model has been done based on the assumption that there is a linear relationship between the observed and the simulated ones. That means all of the error variances is contained in the simulated values, and the measured data is free of error. But in reality, the measured data is not free of error (Moriassi et al., 2007). The goal of calibration is to find those sets of parameter values for the model that give a simulated hydrological series that adequately matches the observed series. Therefore, for this study, the calibration was carried out at multiple sites (Akaki, Melka Kunture, and Hombole) using observed streamflow data. The 16 years of measured streamflow data (2000–2015) were used to calibrate and validate the model. The data was separated into warm-up (2000–2002), calibration (2003–2011), and validation (2012–2015) periods. Finally, calibration and validation were done using SWAT-CUP 2019 by using the Sequential Uncertainty Fittings Program (SUFI 2). This is because it combines optimization with uncertainty analysis.

3.5.2 Artificial Neural Networks Development

One of the unsolved questions in utilizing ANNs to model the hydrological process in which architecture to utilize. Aside from hidden layer units and weights, choosing an appropriate input vector (significant independent regressor variables) is a difficult challenge. In contrast to physical-based models, the sets of variables that influence the system output are not known in advance when creating ANN models. As a result, selecting a suitable and sparse input vector that allows the ANN to properly map to the desired output vector is a tough challenge.

To properly use ANN, we should first understand the relative benefits of the different networks available and the best technique to train them. Some decisions should be made when using a neural network to model streamflow. Initially, we have to choose the appropriate neural network type. Following that, we must select the appropriate training algorithm, training period, and network structure. Finally, we need to determine how to pre-and post-process the input and output data.

There are several steps involved in creating a neural network model. The following steps for developing an ANN model were outlined (Maier and colleagues, 2010). Input selection, data division, model architecture, Model network (structure), model training (calibration), and model testing (validation).

3.5.2.1 Input selection

Identifying the model inputs is a critical stage in the construction of an ANN model. If one or more important inputs are not included, the resulting model will be unable to construct the best feasible input-output relationship. In time series applications, where prospective model inputs include not only distinct input variables but also their lagged values, the absence of critical model inputs is more likely (unless recurrent neural network architectures are used). This greatly expands the number of potential model inputs (as opposed to input variables), and lags have been used in many previous studies (Rezaei et al., 2015).

The employed statistical approach in this study was suggested by Sudheer et al. (2002) to identify the appropriate input vector for a data-driven model. The method is based on the heuristic that the possible influencing variables, related to different time lags, can be identified through cross-correlation analysis. As a result, selecting all of the lagged attributes as input attributes would almost certainly degrade the model learning process due to redundant information (Shanmuganathan and Samarasinghe, 2016).

The addition of too many inputs is caused mainly by input redundancy, in which some of the selected inputs provide substantial information but are related to each other and hence supply redundant information. This can lead to a variety of issues. For starters, redundant inputs increase the likelihood of overfitting (overtraining). This is due to the fact that a greater number of inputs increases the size of the network, and hence the number of connection weights (i.e., model parameters) that must be calibrated. Because the number of training samples is often fixed, adding redundant model inputs raises the ratio of connection weights to training samples, increasing the likelihood of overfitting while contributing no more information to the model. Furthermore, including duplicate model inputs adds more local minima to the error surface in weight space. For example, if two model inputs (y_1 and y_2) are highly correlated and hence represent the same input

information, there are likely to be many weight combinations that result in equal model performance (Yenigün et al., 2010). In general, when developing an ANN model, more attention should be paid to the proper selection of input vectors. Therefore, the input you select can have a significant impact on the performance of your model. The response of a watershed to any quantity of rainfall that falls across it is impacted by both precipitation and catchment characteristics. However, it is difficult to establish the relationship between the physical model's residuals and the time series of specified state variables.

Correlation and cross-correlation analysis were performed on various parameters to determine the best input vector for residual (error) and streamflow time series. Therefore, in this study, daily areal precipitation, preceding day streamflow, residual (error), and daily streamflow at prior time steps were analyzed. Based on the correlation analysis, the possible set for training a stand-alone ANN and hybrid model was identified, and the final input combination was chosen through trial and error by examining the models with the list RMSE.

3.5.2.2 Data division (train, validation and test data)

Data division, or "splitting," is an important concern for the ANN model development process. The available data is generally splatted into training, testing, and validation subsets. The training set is used to estimate the unknown connection weights; the testing set is used to decide when to stop training in order to avoid overfitting and/or determine which network structure is optimal, and the validation set is used to assess the generalization ability of the trained model. Consequently, the training, testing, and validation sets should have the same statistical properties in order to develop the best possible model given the available data (Devi, 2015). Different studies used different percentages for data division into training, validation, and testing. The most common data division that was used in machine learning algorithms was 70% for training purposes and 30% for testing and validation (Basin et al., 2017; Singh et al., 2017). Therefore, to keep consistency with the physical-based SWAT model, for the development of hybrid models, 70% of the data was used for training, which covered the period from 2003–2011, and the remaining four years (2012–2015) were used for validation. During data division, the statistical properties of both the training and

test dataset should be checked before the model evaluation. The statistical property of the datasets are given in table 3.5.

Table 3.5. Statistical properties of training and testing datasets

Station name		Akaki		Melka Kunture		Hombole	
	Variables	Mean	SD	Mean	SD	Mean	SD
Training period (2003-2011)	Q	20.30	25.4	38.1	71	45.71	73.81
	P	0.52	1.09	1.90	1.16	1.71	1.23
	T _{max}	26.20	1.65	25.4	1.53	24.0	1.67
	T _{min}	9.79	2.0	10.0	2.40	10.50	1.97
Testing period (2012-2015)	Q	32.80	30.80	32.5	64.3	39.48	70.58
	P	0.40	0.94	1.77	1.01	1.58	1.28
	T _{max}	25.30	1.92	25.80	1.44	25.10	1.79
	T _{min}	10.0	1.80	9.80	2.23	10.10	1.90

Where SD stands for standard deviation, Q stands for streamflow, P stands for precipitation, and T_{max} and T_{min} stand for maximum and minimum air temperature.

3.5.2.3 Data Preprocessing

Machine learning algorithms are designed to discover meaningful correlations between explanatory variables and a target variable of interest. Explanatory variables must be preprocessed to eliminate bias linked to the ranges of input variables. Scaling is recommended because some algorithms assign weight to explanatory variables based on their magnitude. Furthermore, algorithms that use gradient descent as an optimization approach (such as logistic regression and neural networks) require data scaling since range differences result in different step sizes in the gradient descent formula for each feature, causing computation time to increase. Different ranges have an impact on algorithms like K-neighbors and support vector machines, which rely on distances between data points to determine similarity.

Standardization/Normalization

The standardized explanatory variables are obtained by removing the mean and scaling them to unit variance (Casari and Zheng, 2018). According to Wu et al. (2012), standardization and normalization are critical in facilitating the improvement of machine learning model training processes. By computing the relevant statistics on the samples in the training set, centering and scaling happen independently of each explanatory variable. The mean and standard deviation are then saved and used on subsequent data using a back-scale. On the other hand, normalization is the process of casting the data into a specific range, like between 0 and 1, or between -1 and +1 and it is required when there are big differences in the ranges of different features. This scaling method is useful when the data set does not contain outliers. Equations 4.1 and 4.2 were used for standardization and normalization, respectively.

$$x^s = (x - u)/s \dots\dots\dots (4.1)$$

Where x^s is the new value of the sample, x is the old value of the sample, u is the mean of the training samples, and s is the standard deviation of the training samples,

$$x^n = (x - x_{min})/(x_{max}-x_{min}) \dots\dots\dots (4.2)$$

Where x^n is the new value of the sample, x is the old value of the sample, x_{min} is the minimum and x_{max} is the maximum value of the training samples.

3.5.2.4 Model architecture selection

The model (network) architecture determines the overall structure and information flow in ANN models. Consequently, it has a significant impact on the functional form of the relationship between model inputs and outputs. Traditionally, ANN architectures have been divided into feedforward and recurrent networks. In feed-forward networks, information propagation is only in one direction, i.e., from the input layer to the output layer. Multilayer perceptrons (MLPs) are the most common form of the feed-forward model architecture. The MLP is the most popular neural network architecture in use today. In the majority of studies, the MLP is trained using the error-back propagation algorithm. Generally, the MLPNN (FFNN) normally consists of three layers: an

input layer, a hidden layer, and an output layer. Supervised learning is used for adjusting the weights connected to each layer. For adjusting weights in MLP, supervised learning is used, and the error (difference between actual and expected results) is calculated. The idea of the back propagation algorithm was to reduce this error until the ANN learned the training data. The neurons of the input layer receive the input vector and transmit the values to the next layer of elements to be processed across connections; this process continues up to the output layer. The number of hidden layers and neurons is set according to the output accuracy required, and the error is minimized by the trial-and-error method (Dawson and Wilby, 2001). In this study, a multilayer perceptron (MLP) type was used, which is a feedforward neural network architecture with full unidirectional connections between successive layers and is widely used in water resources engineering applications (Tamiru and Dinka, 2021). The structure of the MLP-ANN is given in Figure. 3.11, which includes input, hidden, and output layers of neurons. These three layers of weight were linked by strong connections. Thus, there are two sets of weights: the input-hidden layer weights (w_{ji}) and the hidden-output layer weights (w_{kj}), which provide a high degree of flexibility in adapting to the data. The output of typical MLPNN with a single layer of hidden neurons is given by the following equation:

$$Y_k = fo \left[\sum_{j=1}^m (W_{kj} * fh \left(\sum_{i=1}^n (w_{ji} + x_i) + b_j \right)) + b_k \right] \dots \dots \dots (3.4)$$

Where: Y_k is the output variable
 x_i is the input variable
 W_{ji}, w_{kj} the input-hidden layer weights and the hidden-output layer weights respectively
 n, m is the number of neurons in the input and hidden layer respectively
 b_j, b_k is the bias of the hidden, and output layer respectively
 fh, fo is the activation function of the hidden and output layer respectively

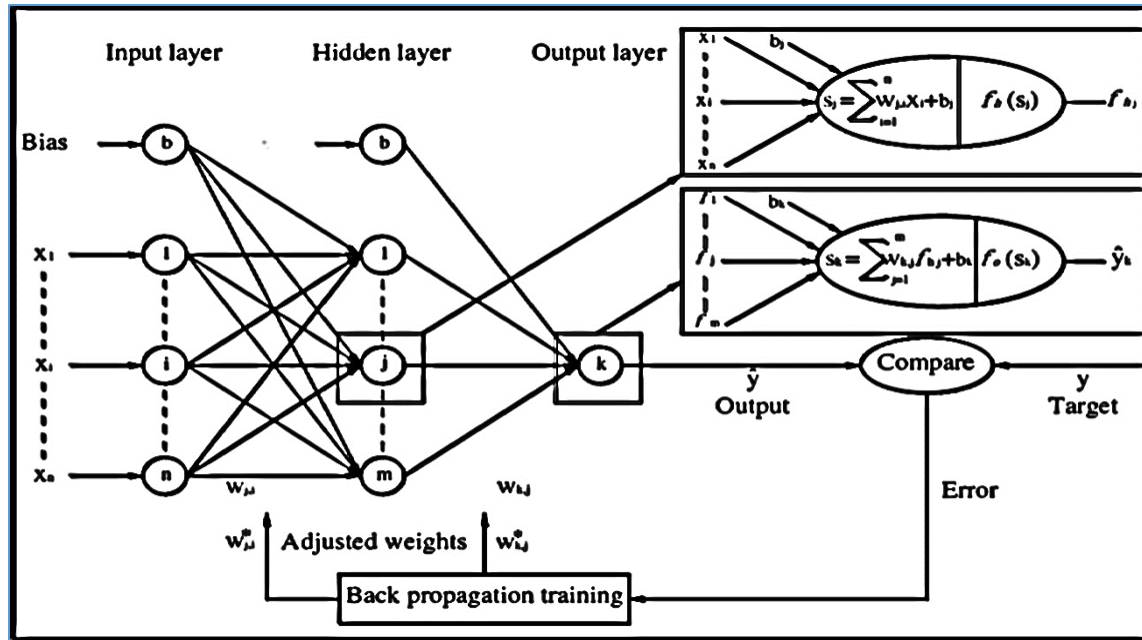


Figure 3.11. Structure of a multilayer perceptron ANN. Source: (Kassem et al., 2020)

3.5.2.5 Model network (structure) selection

The functional form of the relationship between model inputs and outputs is defined by the model (network) structure and the model (network) architecture. The number of hidden nodes, how they are arranged (e.g., number of layers, number of nodes per layer), and how they process incoming signals are all factors in determining an appropriate network structure (e.g., type of transfer function, etc.). In general, the best network structure strikes a balance between generalization ability and network complexity (e.g., network size and the number of free parameters). The network may be unable to capture the desired relationship if network complexity is too low or an inappropriate functional form is chosen. However, if network complexity is too high, generalization and processing speed suffers, and the network becomes more difficult to calibrate and transparent (Shirke and Kawitkar, 2012; Singh et al., 2011; Jimeno-Sáez et al., 2018; Dalkılıç and Hashimi, 2020). According to Wanas et al. (1998), the optimal number of hidden neurons is governed by the number of input and output data points, and the best number of hidden neurons with the highest performance occurs when the number of hidden nodes equals the logarithm of the training data points. As per Mishra and Desail (2006), the optimal number of hidden neurons is $2N+1$, where N is the number of input neurons. The number of nodes is further influenced by the

number of training instances, the degree of noise in the intended output, the complexity of the function or classification to be learned, and the design and training techniques.

3.5.2.6 Model training

Model training is the process of determining a set of model parameters (for example, connection weights) that will allow a model with a specific functional form to best represent the desired input/output relationship. To develop and train the models, the Python programming environment (version 3.7) was chosen, which has significant advantages over corresponding languages in that it is easier to design applications while also being a high-level language, making it more flexible and easily accessible, and understood by human nature. This programming language was used in conjunction with the Jupyter Notebook and Google Colab platforms, which offered a free cloud service with a free GPU.

The TensorFlow machine learning platform was used, which was developed by Google in 2015. Its core is written in C++/CUDA and serves as a platform for developing ML models, particularly neural networks. TensorFlow's execution context is heterogeneous, consisting of a multi-core CPU and several many-core GPU environments. TensorFlow has become one of the most popular frameworks due to the amount of support and documentation produced in recent years, as well as the ease with which calculations can be distributed to a variety of devices and its excellent visualization tool (Iruela et al., 2021).

Before training the network, the input data was normalized in order to change the values of numeric columns on a common scale. The simplest implementation of backpropagation learning updates the network weights and biases in the negative of the gradient, which is the direction in which the performance function decreases the most rapidly. After each input is applied to the network, the gradient is computed, and the weights are updated in incremental mode. Before the weights are updated in batch mode, all of the inputs are applied to the network. Batch training (train) mode was used in this study.

Overfitting is one of the issues that arise during neural network training. There is little or no chance of overfitting if the number of parameters in the network is much smaller than the total number of

points in the training set. If overfitting is a concern, the best generalization ability is achieved by minimizing a suitable error measure between actual and predicted outputs in the test set, assuming that the training and testing data are representative of the modeling domain. When overfitting is not an issue and the training datasets are representative of the modeling domain, an appropriate error measure between actual and predicted training outputs is minimized (Beyen, 2018). The Python libraries that were used and are necessary to build a neural model are:

1. Tensorflow: mathematical library compiled by the Google group and is applicable to machine learning and neural networks
2. Keras: library suitable for neural networks, with excellent results in deep learning, which has been written in Python
3. Pandas: suitable for analyzing and processing data in Python programming language
4. Numpy: mathematical library for the implementation of complex mathematical calculations between tables of different dimensions
5. Matplotlib: graphical depictions library
6. HParams dashboard: TensorBoard provides several tools to help with the hyperparameter optimization and visualization process, which identify the best experiment or most promising sets of hyperparameters.

3.5.2.7 Hyperparameters optimization

Grid search (GS) is the most widely used method for learning hyperparameter configuration space. It can be described as an exhaustive exploration or a brute-force method that tests all hyperparameter combinations given to the grid configuration. GS works by calculating the Cartesian product of a finite set of user-defined values (Zoller and Huber, 2019). GS will not be able to fully exploit the high-performing areas on its own. To be successful, a hyperparameter search should start with a large search space based on previous findings of high-performing hyperparameter environments. However, the main disadvantage of GS is its ineffectiveness in the high-dimensional configuration space; as the number of parameters increases, the possible combinations of parameters also grow exponentially.

Before grid searching was applied to the algorithm, a dataset was divided into training and validation sets. Then the wide range of hyperparameters was selected by referring to the previous work as presented in Table 3.6. To reduce the model's complexity and dimensionality issue, the number of hidden layers was fixed at one, and a grid search strategy was employed to tune the number of neurons in the hidden layer as well as other hyperparameters, including the learning rate and optimizers. Finally, a validation dataset was used to validate the models, and a model with all possible combinations of hyperparameters was tested on the validation set to choose the best hyperparameters.

Table 3.6. The hyperparameters values used for model optimization.

No	Hyperparameters	Range of values
1	No of nodes in the first hidden layer	Log (m) to $2N+1$
2	Optimizers	Adam, SGD
3	Learning rate	0.1, 0.01, 0.001

Where m is the number of training data points and N is the number of input features SGD is Stochastic Gradient Descent and Adam is Adaptive Moment Estimation.

Optimization algorithm

Optimization is one of the core components of machine learning. The essence of most machine learning algorithms is to build an optimization model and learn the parameters of the objective function from the given data. An optimization algorithm finds the values of the parameters (weights) that minimize the error when mapping inputs to outputs. These optimization algorithms, or optimizers, greatly affect the accuracy of the ML model. They also have an impact on the model's speed training. Almost all popular optimizers in ML are based on gradient descent. This means that they repeatedly estimate the slope of a given loss function and move the parameters in the opposite direction. The simplest example for such an optimizer is probably stochastic gradient descent (or SGD) which has been used since the mid-20th century (Robbins et al., 1951). In the 2010s, the use of adaptive gradient methods such as AdaGrad or Adam has become increasingly popular (Duchi et al., 2011). However, recent trends show that parts of the research community

are moving back towards using SGD over adaptive gradient methods. In this study, out of two optimizers, the best optimizers for each model was tuned using a grid search strategy.

3.5.2.8 Testing the model

The performance of a trained network is evaluated and tested after it has been trained using an independent data set and predetermined performance criteria. It is critical that the data collected during the testing process not be used in any way during the training process. The goal of model testing is to validate the network on samples that were not used in the training process, referred to as the testing set. Among the other networks that are trained all the way to convergence on the training samples, the network that can yield the best performance on the validation samples would be the most accurate model (Jimeno-Sáez et al., 2018).

3.5.3 Hybrid model

As discussed in the previous chapter, the literature demonstrated that combining data-driven modeling with physical or process-based modeling is not only feasible but also has the potential to improve model performance. To integrate different models, researchers use various combinations of models and methods. In this study, two popular hydrological models, SWAT and ANN, were combined into two different schemes.

3.5.3.1 Scheme 1

The SWAT model output, as well as the time series of climatic variables and time lag of climate variables, were used to develop the hybrid model in Scheme 1 (hereafter Hybrid1). The rationale for integrating the output of the SWAT model into ANN is to increase the correlation between input variables and output (streamflow) through the more generalized output of the model that encounters the effect of the spatiotemporal variation of the hydrological processes in a distributed manner, which is expected to increase the accuracy of ANN. The output of the SWAT model was combined with other time series, such as area rainfall, temperature, and the preceding time step of those variables was utilized based on correlation analysis. As a result, input data in various combinations was prepared and treated as an individual model, and each model was trained individually. Humphrey et al. (2016); Ghaith et al. (2020); Konapala et al. (2019); and Farfán et

al. (2020) also used this strategy successfully and reported improved model results. At first, the calibrated result of the SWAT model was used as input along with climate variables. To make the training process of the model easier and to prevent the overfitting problem, all ANN models were trained with a single hidden layer, and the number of neurons was chosen by trial and error, which gave the lowest RMSE. Daily streamflow was simulated as a function of:- daily precipitation (P_t), maximum daily temperature ($T_{\max}$), minimum daily temperature ($T_{\min}$), precipitation of the previous n days (P_{t-n}), and cumulative rainfall of the preceding n days (R_n). In this study, the most suitable delays for climate variables were determined using cross-correlation analyses. After considering the possible input combinations, selected ANN models were built and trained following the normal protocol discussed in Section 3.5.2.

3.5.3.2 Scheme 2

One straightforward way to combine physics-based and data-driven models are to feed the error of a physics-based model as an input to a data-driven model. The second scheme of hybrid modeling adopted in this study was based on post-processing methods that can reduce and correct biases in simulated hydrological model (SWAT). The residual error generated from the physical-based semi-distributed SWAT model can be modeled by finding any systematic patterns in the historical residuals of the physical model and linking them to state variables chosen for the development of the ANN model (hereafter Hybrid2). The role of the data-driven ANN model in this scheme was to better describe the nonlinear relationship between the time series of climate variables and the autocorrelation of the residual error itself in order to better characterize the SWAT model's residual error.

The approach is built on the assumption that any patterns discovered between model residuals and state variables are a reflection of structural issues in the data and/or model itself. As a result, residual data is used to construct an ANN model that captures the pattern and aids in correcting the error of the SWAT model simulation. In general, this method has been used successfully in many studies to correct physical and conceptual model errors (Abebe and Price, 2003; Lv et al., 2020; Kwak et al., 2021; Cho and Kim, 2022).

The designed methodology was started by applying the physical model to the watershed. Therefore, the SWAT model was used to simulate the streamflow of the Upper Awash Sub-basin, and the model was calibrated at three gauging stations. All the necessary steps, model input data, and model setup for SWAT development were given in Section 3.5.1. Then the residual, or error, of the SWAT model was calculated by using equation 3.5, given below.

$$Q_{Er} = Q_{obs} - Q_{SWAT} \dots \dots \dots (3.5)$$

Where Q_{Er} is error or residual of SWAT model, Q_{SWAT} is the streamflow found by the SWAT model, and Q_{obs} is the recorded value of streamflow.

Finally, the result of equation 3.5 was used as the target variable in the building of the ANN model that simulates streamflow uncertainty (residuals). The results of the ANN and SWAT models were then combined using a simple ANN model as a total flow (Lv et al., 2020).

The procedure for applying Scheme 2 is shown in the following steps:

1. Apply the SWAT model to historical data from the sub-basin and simulate streamflow by following the standard procedures for developing the SWAT model as discussed in Section 3.5.1.
2. Apply equation 3.5 to compute the historical residual error time series between the predicted and observed discharge.
3. The relationship between the residual error time series and other time series at varying lag times, such as precipitation, temperature, and streamflow, is analyzed using cross-correlation.
4. Select the time series that are most correlated to the historical residual time series based on the results of step 3. Also, choose the lag times that are highly correlated.
5. Set up an ANN model that targets the residual errors using the best-related time series as input (at corresponding lag times). This process also follows the standard procedures for developing ANN as discussed in section 3.5.2.

6. Apply the ANN model to predict the final corrected flow by combining the predicted error with the SWAT simulated streamflow.

3.6 Model Performance Evaluation

As models are only representative of real processes, limitations are observed in terms of performance. Therefore, this step is necessary to evaluate the model outputs against the observed data. The performance of a model is evaluated either subjectively or objectively. Subjective evaluation is based on visual inspection to determine how closely the simulated and actual discharges match, and the model's systematic (underestimation/overestimation) or dynamic (periodic pattern) behavior is observed, whereas objective evaluation is based on a mathematical examination of how well the two discharges fit one another (Waseem et al., 2017). There are several criteria for evaluating hydrological models. However, for this study, performance evaluation criteria such as coefficient of determination (R^2), Nash and Sutcliffe efficiency (NSE), percent bias (PBIAS), and root mean square error (RMSE) are discussed. The equation used to compute them is given as follows:

$$R^2 = \frac{(\sum_{i=1}^n (O_i - O_{av})(P_i - P_{av}))^2}{\sum_{i=1}^n (O_i - O_{av})^2 \sum_{i=1}^n (P_i - P_{av})^2} \dots \dots \dots (3.6)$$

$$NSE = \frac{\sum_{i=1}^n (O_i - O_{av})^2 - \sum_{i=1}^n (P_i - P_{av})^2}{\sum_{i=1}^n (O_i - O_{av})^2} \dots \dots \dots (3.7)$$

$$PBIAS = \frac{\sum_{i=1}^n (O_i - P_i)}{\sum_{i=1}^n (O_i)} \times 100\% \dots \dots \dots (3.8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \dots \dots \dots (3.9)$$

Where;

- O_i is the observed streamflow.
- P_i is the predicted streamflow.
- O_{av} is the average observed streamflow.
- P_{av} is the average predicted streamflow.

- i. The coefficient of determination (R^2) is the value that shows the strength of the association between the observed and simulated values. The coefficient of determination has a value that ranges from zero to one; a value of one indicates better agreement, while a value of zero indicates that there is no correlation between predicted and observed values (Moriasi et al., 2015). The coefficient of determination has some limitations as it evaluates only the linear relationship between variables. It can be explained as if $S_i = (MO_i + N)$, where M is a non-zero value and N is any value. In this case, the coefficient of determination (R^2) of S (simulated) and O (observed) will be equal to one. This reflects the fact that R^2 is very exhaustive for additive and proportional differences between observed and model-simulated values (Waseem et al., 2017). Hence, higher values of the coefficient of determination can be achieved even when observed and model-simulated values vary significantly.
- ii. The Nash and Sufficient Efficiency (NSE) indicates how well the plot of observed versus simulated data fits the 1:1 line and is recommended because it is very commonly used and provides extensive information on reported values. The range lies between 1 and, $-\infty$, with 1 being the perfect fit (Moriasi et al., 2007). The NSE criterion has a major disadvantage in that it leads to overestimation or underestimation depending on the dominant error of the model under consideration. Another disadvantage is that it calculates the variance using the observed mean. This means that for catchments with highly variable discharges, the efficiency is overestimated.
- iii. Percent bias (PBIAS): Percent bias measures the average tendency of simulated data to be larger or smaller than their observed counterparts. PBIAS has an optimal value of 0.0, with low-magnitude values indicating accurate model simulation. Positive values indicate underestimation bias in the model, while negative values indicate overestimation bias in the model.
- iv. Root mean square error (RMSE): The root mean square error measures the average error between the observed and simulated values and provides model performance in the units of the output response of interest. The lower the RMSE value, the better the model's performance. The RMSE can be used to measure the agreement between the observed and simulated values.
- v. P-factor and r-factor: The p- and r-factors are closely related to each other, which indicates that a larger p-factor can be achieved only at the expense of a higher r-factor. The Latin hypercube

sampling method was employed for 95PPU and for obtaining the final cumulative distribution of the model outputs. During the initialization of model parameters, SUFI-2 assumed a large parameter uncertainty and then decreased this uncertainty through the p-factor and r-factor performance statistics. The range of the p-factor varied from 0 to 1, with values close to 1 indicating very high model performance and efficiency, while the r-factor was the average width of the 95PPU band divided by the standard deviation of the measured variable and varied in the range 0–1 (Abbaspour et al., 2007; Yang et al., 2008). After balancing these two factors and determining an acceptable value for the r and p-factors, the calibrated parameter ranges could be generated.

Table 3.7. Performance ratings recommended for daily and monthly streamflow simulation.

(Moriasi et al., 2015)

Performance Rating	Performance evaluation criteria for streamflow simulation for both monthly and daily time step		
	R ²	NSE	PBIAS
Very good	> 0.85	> 0.80	< ± 5
Good	0.75 - 0.85	0.70- 0.80	±5 - ± 10
Satisfactory	0.6-0.75	0.50 -0.70	±10 - ± 15
Unsatisfactory	< 0.60	< 0.50	> ± 15

In this study, the two models, SWAT and ANN, were integrated to improve streamflow simulation of the Upper Awash Sub-basin at multiple sites inside the sub-basin. Figure 3.13 describes the sequential steps of the methodologies and the data required by each model to achieve the objectives of this study. The following framework illustrates the general workflow of the study



CHAPTER FOUR

4. RESULTS AND DISCUSSIONS

This chapter describes the results of a multi-site calibration of the SWAT model, two hybrid SWAT and ANN models (Hybrid1 and Hybrid2), and one stand-alone data-driven model (ANN) at the Upper Awash Sub-basin. In addition, the effective time series data for simulating SWAT model residual error was investigated and analyzed in relation to streamflow simulation performance. Finally, the model's performance in streamflow simulation was compared.

4.1 Simulation Streamflow using SWAT Model

4.1.1 Streamflow parameter sensitivity analysis

The initial stage in developing an effective and reliable physical-based hydrological model is sensitivity analysis, which assists in finding and sorting the parameters that have a substantial influence on specific model outcomes (Saltelli et al., 2000). Table 3.5 shows how the variation ranges of these parameters are affected by the absolute SWAT value. SWAT-CUP ranks the sensitivity of parameters based on their t-stat and p-value after running a series of simulations.

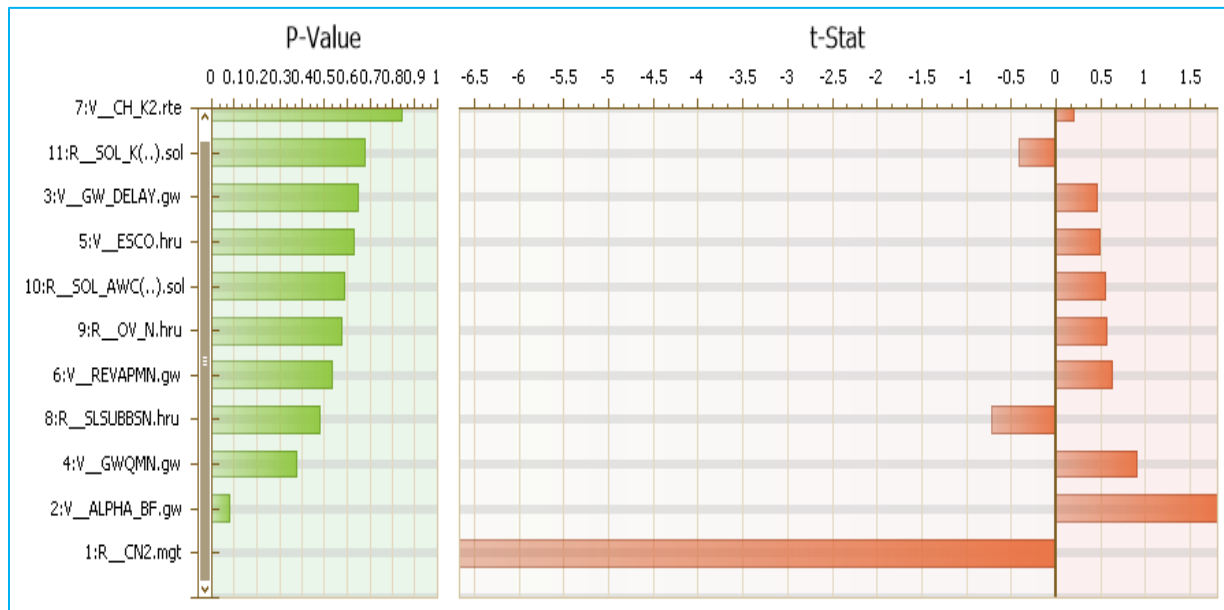


Figure 4.1. Streamflow-sensitive parameters

The highest t-stat value indicates the ratio of the high parameter coefficient to the standard error, and the lowest p-value indicates the rejection of the hypothesis that increasing the parameter value results in a substantial increase in the variable response (Abbaspour et al., 2017). Figure 4.1 shows the most sensitive hydrologic flow parameters during calibration and validation.

The surface runoff response parameter SCS CN II value (CN2) was found to be the most sensitive flow parameter in the sub-basin, followed by the base flow alpha-factor (ALPHA_BF), the threshold depth of water in the shallow aquifer required for return flow to occur (GWQMN), and the average slope length (SLSUBBSN), as shown in Figure 4.1. The threshold depth of water in the shallow aquifer for re-evaporation to occur (REVAPMN), Manning's "n" value for overland flow (OV_N), available water capacity of the soil layer (SOL_AWC), soil evaporation compensation factor (ESCO), and groundwater delay (GW_DELAY) also have the highest influence in controlling the streamflow.

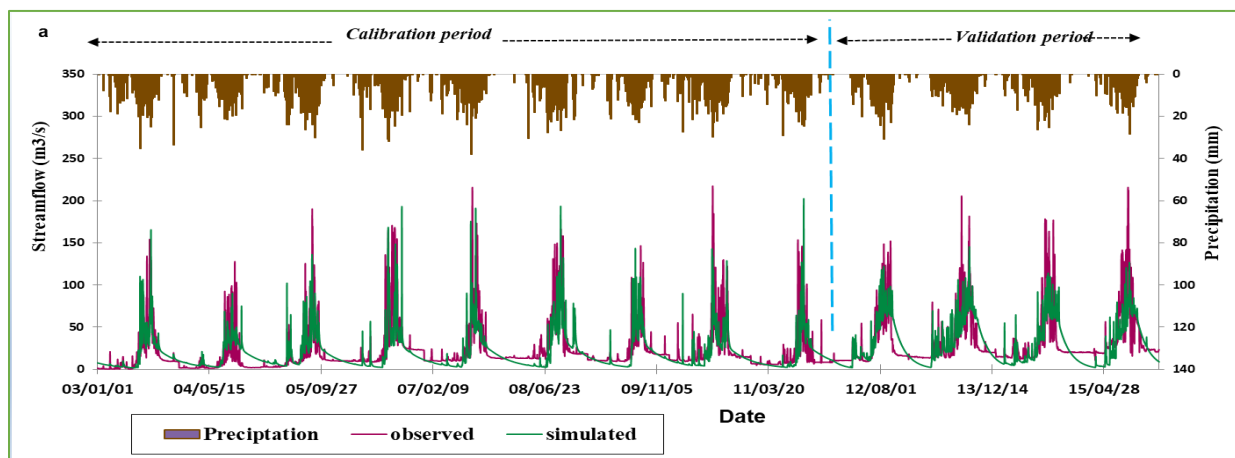
Table 4.1. Shows the best optimized fitted parameter value during calibration.

Parameter_Name	Fitted_Value	Min_Value	Max_Value	Rank
R__CN2.mgt	11.83%	-20%	20%	1
V__ALPHA_BF.gw	0.2	0.2	0.8	2
V__GWQMN.gw	226.23	123.85	1200	3
R__SLSUBBSN.hru	76.2	20	150	4
V__REVAPMN.gw	282.30	20.0	450	5
R__OV_N.hru	-0.082	-0.22	-0.07	6
R__SOL_AWC(..).sol	0.73	0.638	0.99	7
V__ESCO.hru	0.52	0.223	0.824	8
V__GW_DELAY.gw	268.45	62.5	320.25	9
R__SOL_K(..).sol	-0.183%	-0.27%	0.77%	10
V__CH_K2.rte	136.08	56.0	160.0	11

4.1.2 Calibration and validation of SWAT model

In the calibration of SWAT model, the basin parameters are at a global basin scale for all sub-basins, and basin parameters cannot be applied simultaneously in multisite calibration to each sub-basin level (Malik et al., 2022). Therefore, the model was calibrated simultaneously for all three gauging stations using a common set of hydrological parameters since all the selected parameters are varied only with HRU. The model performance indices for both the calibration and validation periods were presented in Table 4.7.

Based on the uncertainty result, the P-factor for the Akaki, Melka Kunture, and Hombole watersheds is 0.53; 0.61; and 0.89, respectively, while the r-factor is 0.99; 0.46; and 1.26, respectively. The P-factor for the Hombole is thus acceptable, but not for Akaki and Melka Kunture. In the case of the r-factor, the value is within acceptable limits except for Hombole, which indicates that the parameter uncertainties are not within the desired ranges for the Akaki and Melka Kunture watersheds. Hence, most of the measured values were not bracketed by the 95PPU, indicating that the SWAT model could not fully simulate the flow processes. This problem might be due to information related to measured rainfall data, measured discharge data, and underlying hydrologic processes. Model structural errors and poor spatial coverage of rainfall data within the watershed could be other reasons.



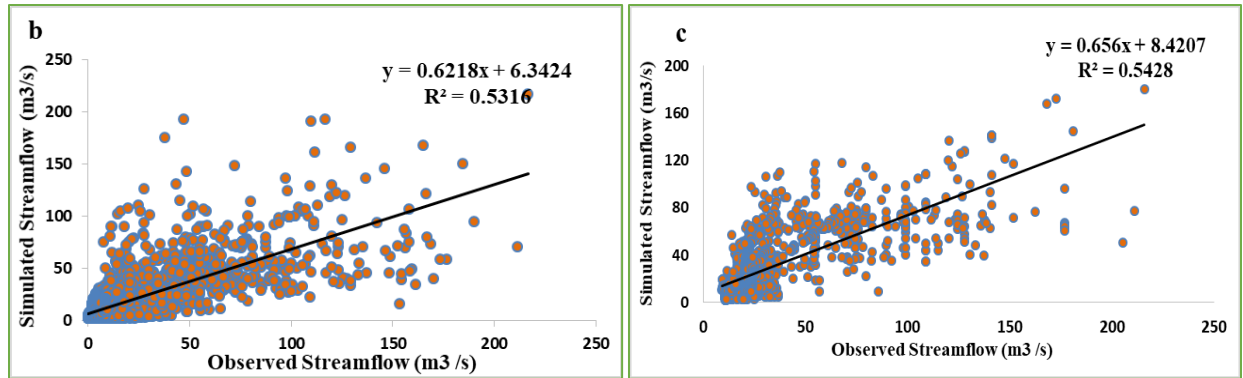


Figure 4.2. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Akaki station: (a) hydrograph during the calibration and validation periods; (b) scatter plot during the calibration period; (c) scatter plot during the validation period.

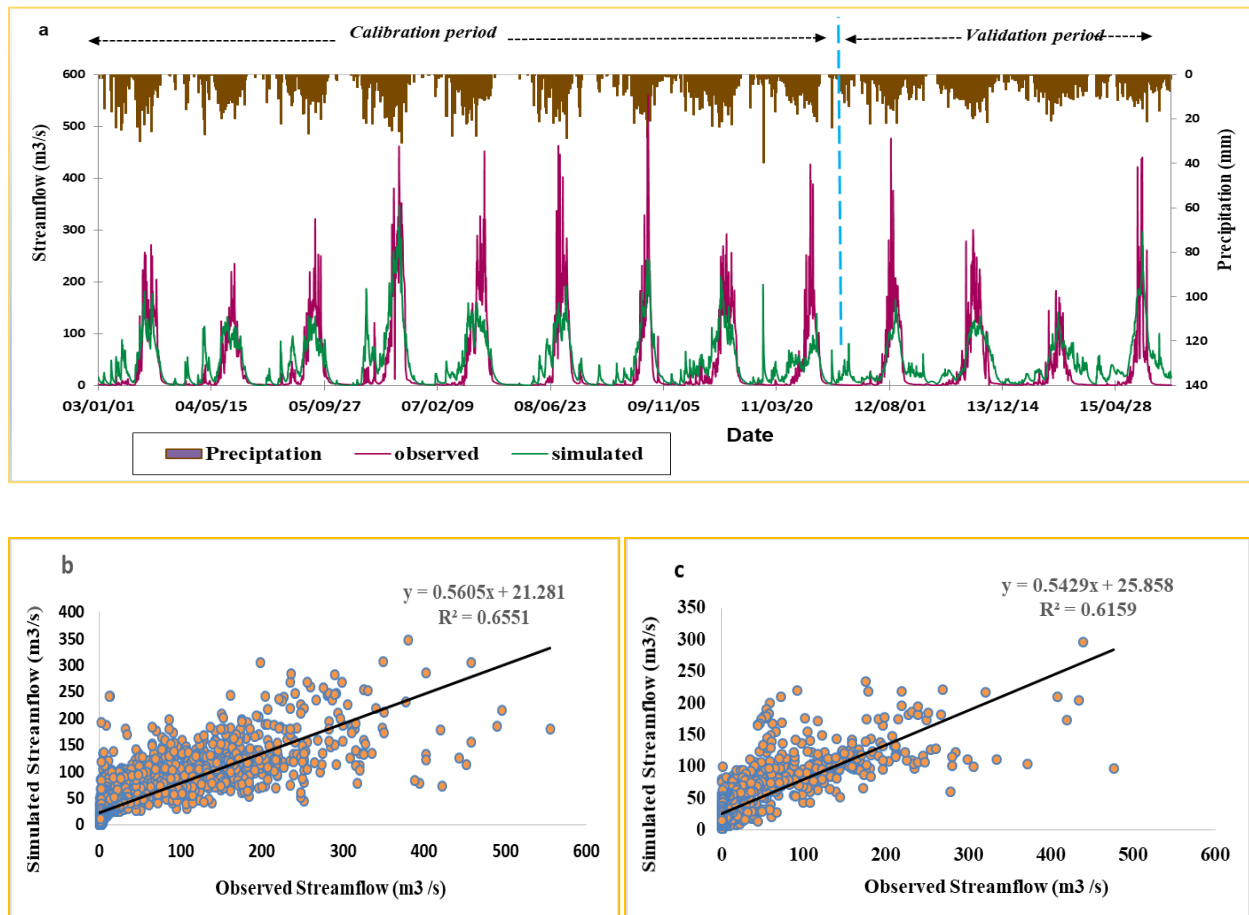


Figure 4.3. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Melka Kunture station: (a) hydrograph during the calibration and validation periods; (b) scatter plot during the calibration period; (c) scatter plot during the validation period.

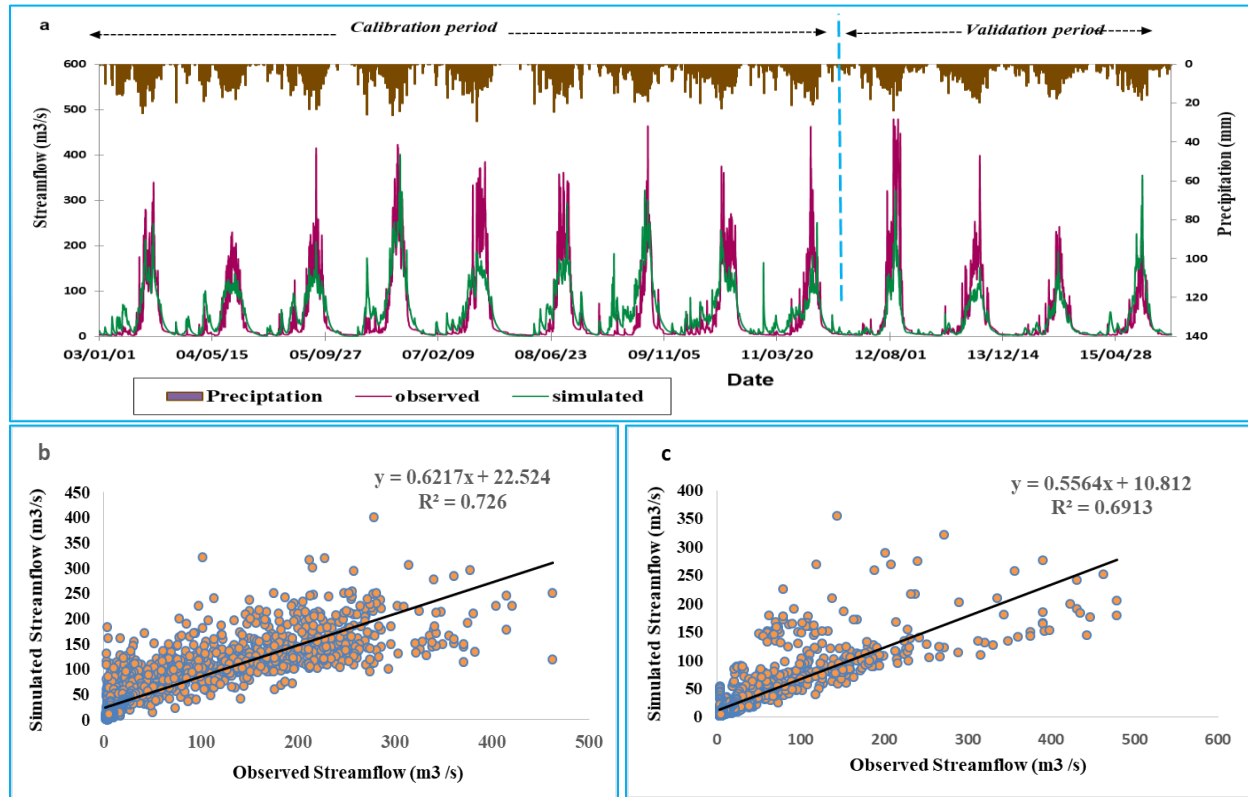


Figure 4.4. Observed and simulated streamflow hydrographs and scatter plots of the SWAT model at Hombole station: (a) hydrograph during the calibration and validation periods; (b) scatter plot during the calibration period; (c) scatter plot during the validation period.

4.2 Hybrid model

4.2.1 Scheme 1

The results of the hybrid1 model, training, and evaluation for streamflow simulation of the Upper Awash Sub-basin at three hydrometric stations using the methods and tools presented in Scheme 1 are discussed in this section. In addition to the hybrid1 model, the training result of a stand-alone ANN was included. Following multiple experiments with all possible combinations of input

features, the six models with the lowest RMSE for the Hybrid1 model and the five models for the stand-alone ANN model were identified and presented in Table 4.2.

For streamflow simulation by a stand-alone ANN model, daily precipitation (P_t), maximum daily temperature ($T_{\max}$), minimum daily temperature ($T_{\min}$), and cumulative rainfall of the preceding n days (R_n) were considered as potential predictors of streamflow. In addition, the lags of those variables were analyzed by cross-correlation analysis and considered as input. The result of the correlation analysis of variables is presented in Appendix C.

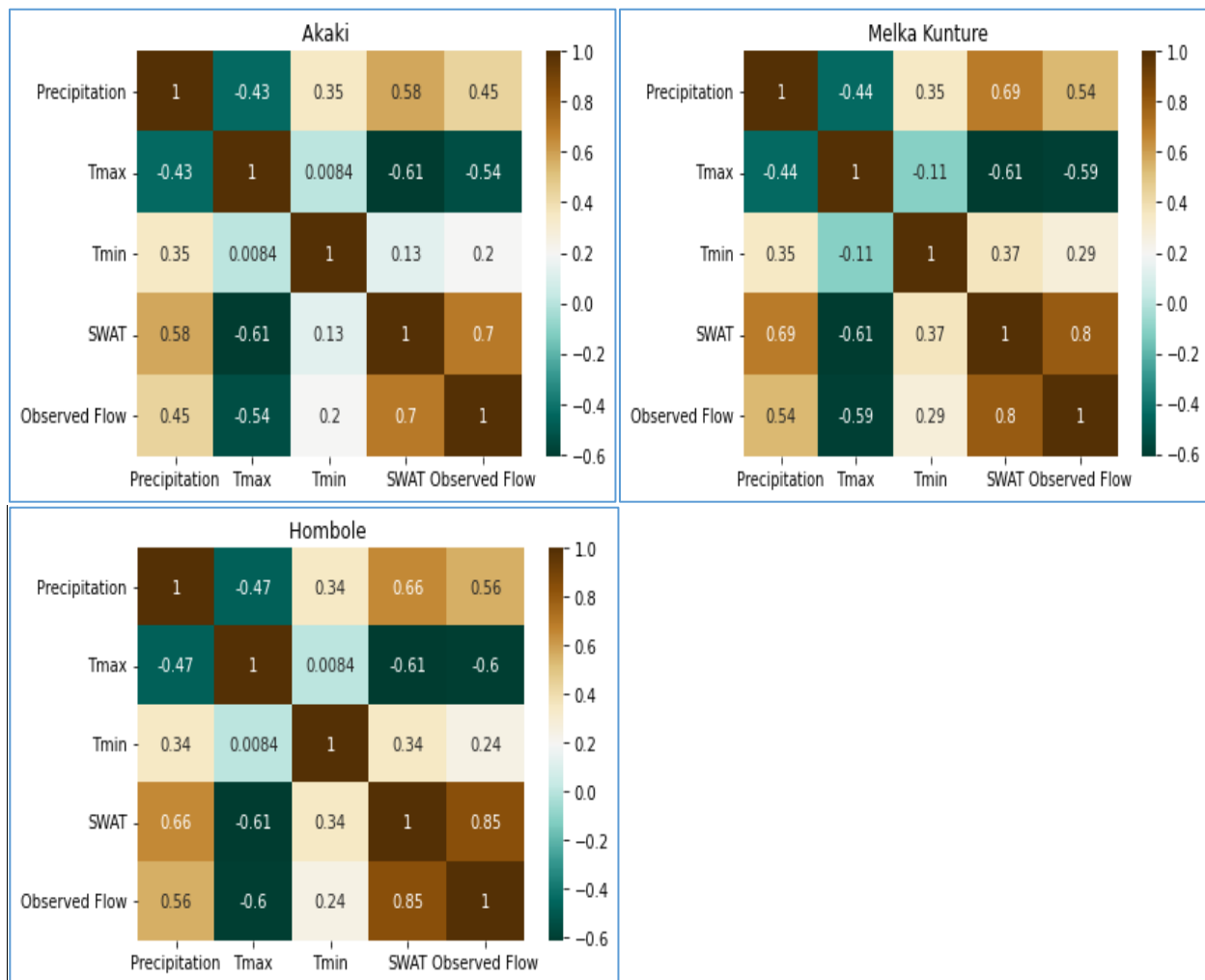


Figure 4.5. Correlation between observed streamflow, climate variables, and simulated streamflow.

The correlation between rainfall and streamflow at Akaki station is weak and positive, as illustrated by the heat maps above. However, the correlation is slightly stronger in Melka Kuntura and Hombole. The maximum temperature has a moderately negative correlation across all stations, whereas the minimum temperature has a weakly positive correlation. The lack of a significant correlation between climate variables and streamflow significantly reduced the performance of the data-driven model. However, because observed and SWAT-simulated streamflow have a strong positive correlation, the Hybrid1 model significantly outperformed the stand-alone ANN model.

Model training and hypermeter optimization

In this study, a feed-forward and back-propagation (FFBP) network with a single hidden layer was trained (along with the usual input and output layers, representing the input variables and target, respectively). In the hidden layer, all ANN models used a ReLU (Rectified Linear Unit) activation function, which is a relatively faster and more effective one. Before training the model, all inputs were standardized.

For optimizers, the number of nodes in the first hidden layer and the learning rate were adjusted to achieve the best results. After several iterations and experiments on the data, the optimum set of hyperparameters was discovered. At Akaki Station hybrid1 model with the 6-1-13-1 structure and stand-alone ANN model with the 5-1-3-1 structure were found to be the best models with a low MRSE. The input features for Akaki station that produced the best representation of streamflow simulation were areal precipitation with a lag of one to two days, 32 days of cumulated rainfall, and maximum air temperature for the ANN stand-alone model. Similarly, this input data set was identified as the best input for the Hybrid1 model, along with the output of the SWAT model simulation, to deliver the best representation for optimized streamflow simulation. The optimized Hybrid1 model was developed using the SGD optimizer, with the hyperparameter set at 13 nodes in the hidden layer and a learning rate of 0.001. In the instance of the stand-alone ANN model, 3 nodes in the hidden layer and a learning rate of 0.001 with the SGD optimizer produced the best results. The result of hyperparameter optimization for the Akaki station is given in Appendix E1.

At Melka Kunture stations, the model with 4-1-12-1 and 7-1-7-1 structures fits the stand-alone ANN and Hybrid1 models best, respectively. The stand-alone ANN model was fitted with four input characteristics, including areal precipitation with a one-day lag, cumulated rainfall over 38 days, and maximum air temperature, whereas Hybrid1 was fitted with two extra inputs, including minimum temperature and two-day lag precipitation. The optimal solution for a stand-alone ANN model was found with the hyperparameter set to be 12 nodes in the hidden layer and a 0.01 learning rate with the Adam optimizer, and for the Hybrid1 model, 7 nodes in the hidden layer and a 0.001 learning rate with the Adam optimizer. The optimum set of parameters for the Melka Kunture station was visualized in Appendix E2.

The model with 6-1-12-1 structures fits the stand-alone ANN best for the Hombole station, and the model with 7-1-13-1 structures fits the Hybrid1. The stand-alone ANN model was fitted with six input features, including areal precipitation with a one- to two-day lag, cumulated rainfall over 40 days, and maximum and minimum air temperature, whereas Hybrid1 was developed with the same input data by only adding the result of SWAT. The Adam optimizer determined the optimal solution for a stand-alone ANN model with 12 nodes in the hidden layer and a 0.01 learning rate, and the Adam optimizer found the optimal solution for the Hybrid1 model with 13 nodes in the hidden layer and a 0.01 learning rate. Appendix E3 illustrates the best set of parameters for the stand-alone ANN and hybrid models, as well as the training history, which shows the loss and validation loss during training and validation.

Table 4.2. Training and testing performance of the screened Hybrid1 and stand-alone ANN models

Stations	No of Model	Model input parameter	Model Structure	Training		Testing	
				R ²	RMSE	R ²	RMSE
Akaki	1	$Q_t = f(Q_{SWAT})$	1-1-3-1 _h	0.56	18.7	0.53	22.4
	2	$Q_t = f(P_t)$	1-1-3-1 _a	0.51	22.5	0.48	28
		$Q_t = f(Q_{SWAT}, P_t)$	2-1-5-1 _h	0.58	17.6	0.55	21.3
	3	$Q_t = f(P_t, R_{32})$	2-1-4-1 _a	0.53	19.2	0.51	25.6
		$Q_t = f(Q_{SWAT}, P_t, R_{32})$	3-1-6-1 _h	0.6	16.2	0.56	21.5
	4	$Q_t = f(P_t, P_{t-1}, P_{t-2}, R_{32})$	4-1-7-1 _a	0.59	16.8	0.58	24.2
		$Q_t = f(Q_{SWAT}, P_t, P_{t-1}, P_{t-2}, R_{32})$	5-1-9-1 _h	0.63	15.9	0.6	22.5
	5	$Q_t = f(P_t, P_{t-1}, P_{t-2}, R_{32}, T_{max})$	5-1-3-1 _a	0.58	16.5	0.62	23.7
		$Q_t = f(Q_{SWAT}, P_t, P_{t-1}, P_{t-2}, R_{32}, T_{max})$	6-1-13-1 _h	0.64	15.3	0.62	21

	6	$Q_t = f(P_t, P_{t-1}, T_{\max}, T_{\min}, R_{32})$	5-1-10-1 _a	0.57	17.8	0.58	23.1
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, T_{\max}, T_{\min}, R_{32})$	6-1-15-1 _h	0.65	15.2	0.57	21.6
Melka Kunture	1	$Q_t = f(Q_{\text{SWAT}})$	1-1-3-1 _h	0.71	40.7	0.69	38.9
	2	$Q_t = f(P_t)$	2-1-2-1 _a	0.61	42.15	0.6	40.2
		$Q_t = f(Q_{\text{SWAT}}, P_t)$	2-1-3-1 _h	0.67	40.15	0.62	38.2
	3	$Q_t = f(P_t, R_{38})$	3-1-5-1 _a	0.64	36.6	0.61	37.9
		$Q_t = f(Q_{\text{SWAT}}, P_t, R_{38})$	3-1-10-1 _h	0.77	33.6	0.71	34.7
	4	$Q_t = f(P_t, P_{t-1}, R_{38}, T_{\max})$	4-1-12-1_a	0.78	34.9	0.68	35.6
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, R_{38}, T_{\max})$	5-1-12-1 _h	0.79	32.3	0.72	33.2
	5	$Q_t = f(P_t, P_{t-1}, R_{38}, T_{\max})$	5-1-10-1 _a	0.7	35.2	0.69	36.5
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, R_{38}, T_{\max})$	6-1-7-1 _h	0.76	33	0.72	34.7
	6	$Q_t = f(P_t, P_{t-1}, ET_t, T_{\max}, T_{\min}, R_{38})$	6-1-9-1 _a	0.74	33.5	0.68	36.2
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, P_{t-2}, T_{\max}, T_{\min}, R_{38})$	7-1-7-1_h	0.81	32.1	0.72	33.3
Hombole	1	$Q_t = f(Q_{\text{SWAT}})$	1-1-3-1 _h	0.69	36.4	0.67	38.4
	2	$Q_t = f(P_t)$	1-1-3-1 _a	0.63	37	0.58	43.5
		$Q_t = f(Q_{\text{SWAT}}, P_t)$	2-1-5-1 _h	0.75	35.9	0.72	36
	3	$Q_t = f(P_t, R_{40})$	2-1-4-1 _a	0.66	36.5	0.64	42.7
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, R_{40})$	4-1-8-1 _h	0.76	35.2	0.77	35.4
	4	$Q_t = f(P_t, P_{t-1}, R_{40}, T_{\max})$	4-1-5-1 _a	0.71	35.1	0.67	40.9
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, R_{40}, T_{\max})$	5-1-5-1 _h	0.78	34.5	0.73	35.8
	5	$Q_t = f(P_t, P_{t-1}, P_{t-2}, R_{40}, T_{\max})$	5-1-11-1 _a	0.72	34	0.67	41.8
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, P_{t-2}, R_{40}, T_{\max})$	6-1-12-1 _h	0.8	33.5	0.76	36.6
	6	$Q_t = f(P_t, P_{t-1}, P_{t-2}, T_{\max}, T_{\min}, R_{40})$	6-1-7-1_a	0.82	31.5	0.68	40.1
		$Q_t = f(Q_{\text{SWAT}}, P_t, P_{t-1}, P_{t-2}, T_{\max}, T_{\min}, R_{40})$	7-1-13-1_h	0.83	32.2	0.78	35.1

Subscript “a” denotes a stand-alone ANN, and “h” denotes a Hybrid1.

The hyperparameters are tuned like knobs to increase the neural network's performance, resulting in an optimal model. Hyperparameters are commonly defined as parameters that are used to optimize other model parameters. The outcome of grid search-based hyperparameter optimization was presented, and the results were visualized using Tensorflow's TensorBoard tool, which aids in the understanding of a neural network via interactive graphs and statistics. Table 4.3 shows the results of the optimum hyperparameter for the best fitted models.

Table 4.3. Results of the hyperparameter tuning for stand-alone ANN and Hybrid1.

Station	Models	Model structure	Hyperparameters			Training time(sec)
			Optimum No neuron in a hidden layer	Optimizer	Learning rate	
Akaki	ANN	5-1-3-1	3	SGD	0.001	56

	Hybrid1	6-1-13-1	13	SGD	0.001	63
Melka Kunture	ANN	4-1-12-1	12	Adam	0.01	35
	Hybrid1	7-1-7-1	7	Adam	0.001	45
Hombole	ANN	6-1-12-1	12	Adam	0.01	28
	Hybrid1	7-1-13-1	13	Adam	0.01	33

4.2.2 Scheme 2

4.2.2.1 Residual analysis (post-processing)

Residual errors are the differences between the measured and simulated values, and throughout this study, the terms residual and error are considered similar and used interchangeably. This means that negative residuals indicate an overestimation of discharge by the model, while positive residuals indicate an underestimation.

Residuals can be analyzed without losing the time reference of the data, which allows for the consideration of seasonal effects. Also, it is possible to examine specific parts of the hydrograph separately, like, for example, the rising limb and the recession curve or high and low flows. In spite of the diverse possibilities offered for model evaluation by residual analysis, it is not very commonly used in hydrological modeling.

In this study, the residual error of the SWAT model at three stations was analyzed, with the nine-year model calibration period comprising a total of 3287 days and the four-year model validation period comprising 1461 days. Out of a total of 3287 days in the calibration period, streamflow is overestimated by the model on 1583, 2460, and 2350 days for Akaki, Melka Kunture, and Hombole watersheds, respectively, as indicated by the negative residuals, and underestimated on 1704, 827, and 937 days, as indicated by the positive residuals. Similarly, during the validation period, Akaki, Melka Kunture, and Hombole each have 558, 1239, and 1052 days with a negative residual, while 903, 222, and 409 days have a positive residual out of a total of 1461 days. Monthly residuals were compared across three stations in the study area, and in all stations, a clear seasonality was observed, which represents a more substantial variability in the Kiremt season (June-September) than in Bega (October-January). The range of the residual usually expands in Kiremt. Figure 4.6 shows seasonality in the residual error.

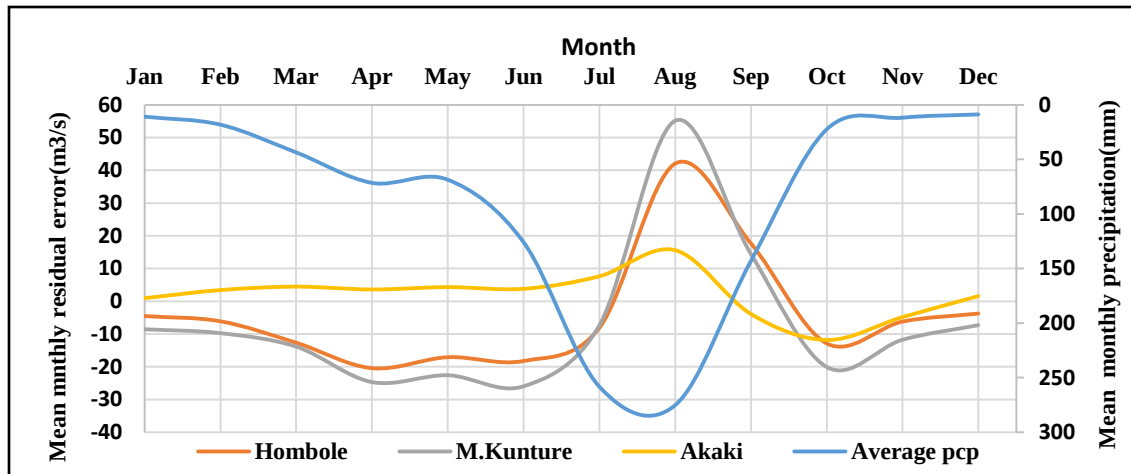


Figure 4.6. The relationship between residual error and precipitation at three stations

As illustrated in figure 4.6, the largest residual error occurs in August, which is the peak rainy season for the Upper Awash Sub-basin. This can be explained by the influence of the Kiremt rainfall, which brings about frequent and heavy precipitation events and consequently high discharges. The residuals indicate that the peaks are underestimated and not simulated very well by the physical-based SWAT model yet. Nevertheless, there is no clear trend toward positive or negative values, which indicates that some model variable causes it to overestimate discharge in certain situations and underestimate it in others.

Table 4.4. Residual error descriptive statistics

Statistics	Akaki	Melka Kunture	Hombole
Mean	2.1	-6.8	-4.1
Median	1.3	-8.8	-5.7
Mode	-0.1	-0.5	-2.4
Standard Deviation	20.2	41.4	39.5
Kurtosis	10.6	15.5	13.2
Skewness	1.1	2.6	2.4
Range	300.6	607.6	563.6
Minimum	-145.8	-228.2	-220.1
Maximum	154.7	379.4	343.5

The residual for the entire period, including calibration and validation, with a total of 4748 data points, was used to plot the histogram. From the statistical analysis, the residual data at all stations did not follow the normal distribution. In particular, the residual data at Melka Kuntura and Hombole stations showed a right-skewed distribution with skewness coefficients of 2.6 and 2.4, respectively, whereas Akaki station has a 1.1 skewness coefficient, which is a relatively small skewness coefficient. According to Özger (2009), the normality assumption is not restrictive and good results can be obtained by using real-world observations directly. As a result, the data was used directly to train a data-driven model without transformation.

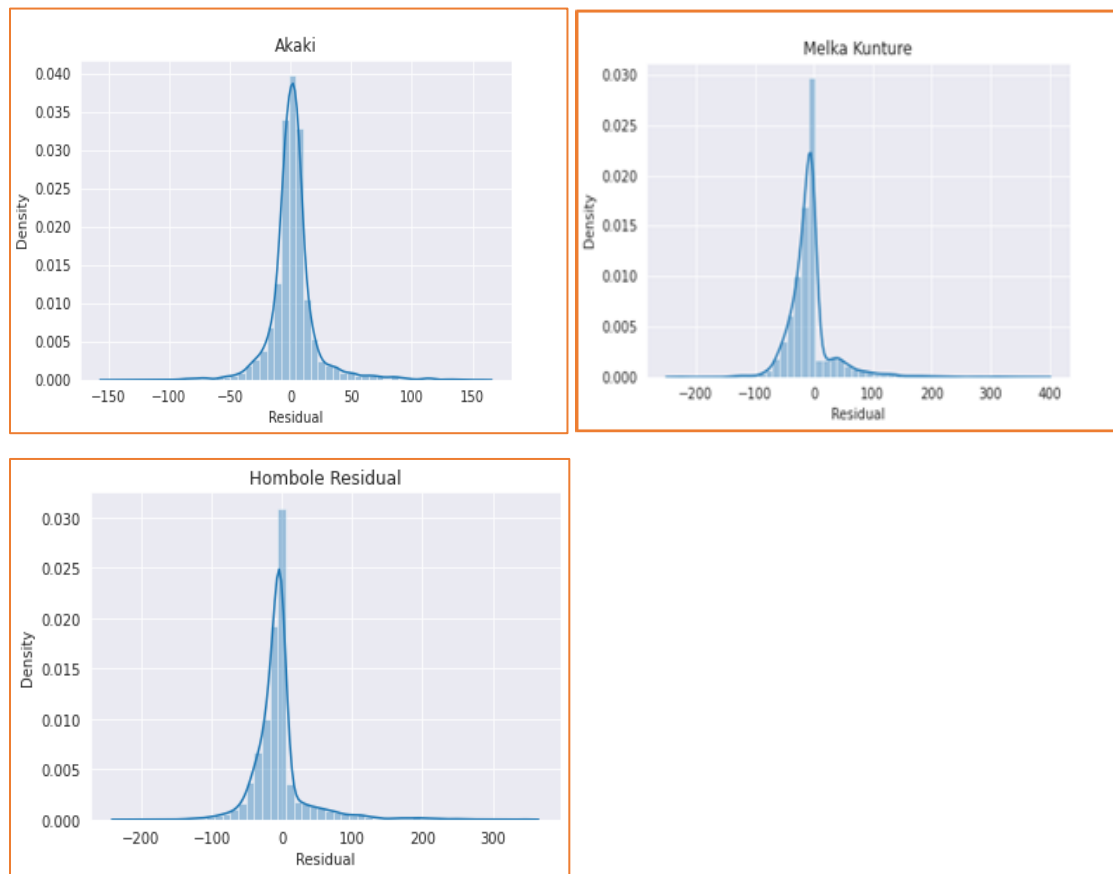


Figure 4.7. Residual error frequency distribution histogram for Akaki, Melka Kuntura, and Hombole stations, respectively.

4.2.2.2 Modeling residual error

According to new hydrological insights tested in many parts of the world by studying many hydrologic scenarios, hydrologic model residual errors frequently follow some patterns that are

highly correlated with the model inputs, model state variables, and output variables, and such patterns can be effectively learned via data-driven models (Frame et al., 2020; Nearing et al., 2020).

Input Selection

The identification of key input variables is a critical stage in the development of data-driven models. Climate variables (precipitation, maximum and minimum air temperature), and observed streamflow time derivatives were used as input variables to the ANN model. The relationship between the residuals of the SWAT model and the time series of selected variables is studied first using a method based on correlation analysis by standardizing the data.

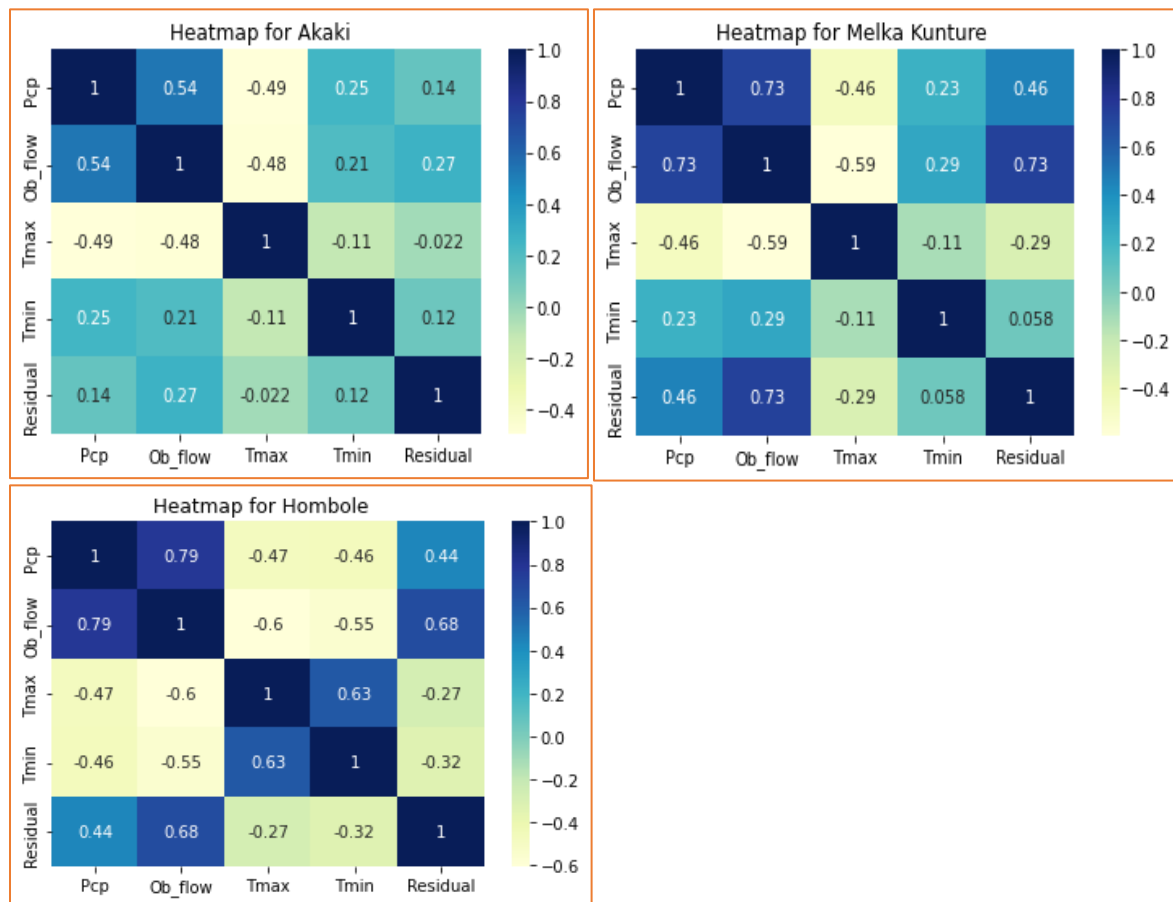


Figure 4.8. Correlation between observed streamflow, climate variables, and residual error.

Based on Correlation and cross-correlation analysis, the following input variables were selected: observed streamflow, precipitation, and air temperature, for the current and/or previous days. As shown in Figures 4.9, 4.10, and 4.11, the correlation between residual error and streamflow shows a strong positive correlation that decreases with increasing time of lag in all three stations. As a result, inputs, including observations from the three days preceding the simulation day ($t-1$ through $t-3$), were considered inputs, where "t" represents the time step. Similarly, precipitation shows a weak positive correlation at Akaki station and a relatively higher correlation at Melka Kunture and Hombole stations. The maximum correlation occurs at one day's lag ($t-1$), and it continues to decrease with time for all stations. The correlation between residual error and maximum temperature was weakly negative, with maximum lags occurring at $t-1$, $t-15$, and $t-20$ for Akaki, Melka Kunture, and Hombole stations. For the case of the minimum air temperature, the correlation coefficient is very weak at all stations. Generally, based on the correlation analysis result and the training performance, different input combinations of observed streamflow, precipitation, and maximum and minimum air temperatures were considered.

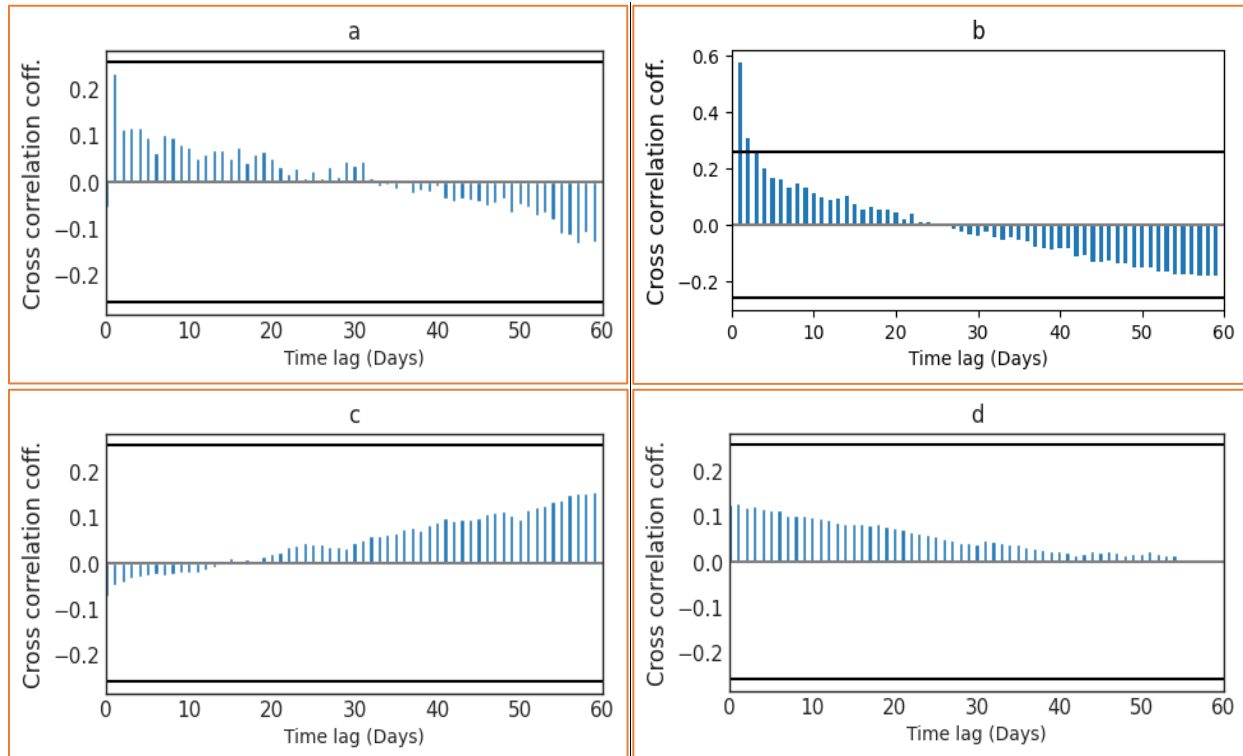


Figure 4.9. Akaki station Cross-correlation graphs: a) residual and precipitation; b) residual and observed streamflow; c) residual and maximum temperature; and d) residual and minimum temperature.

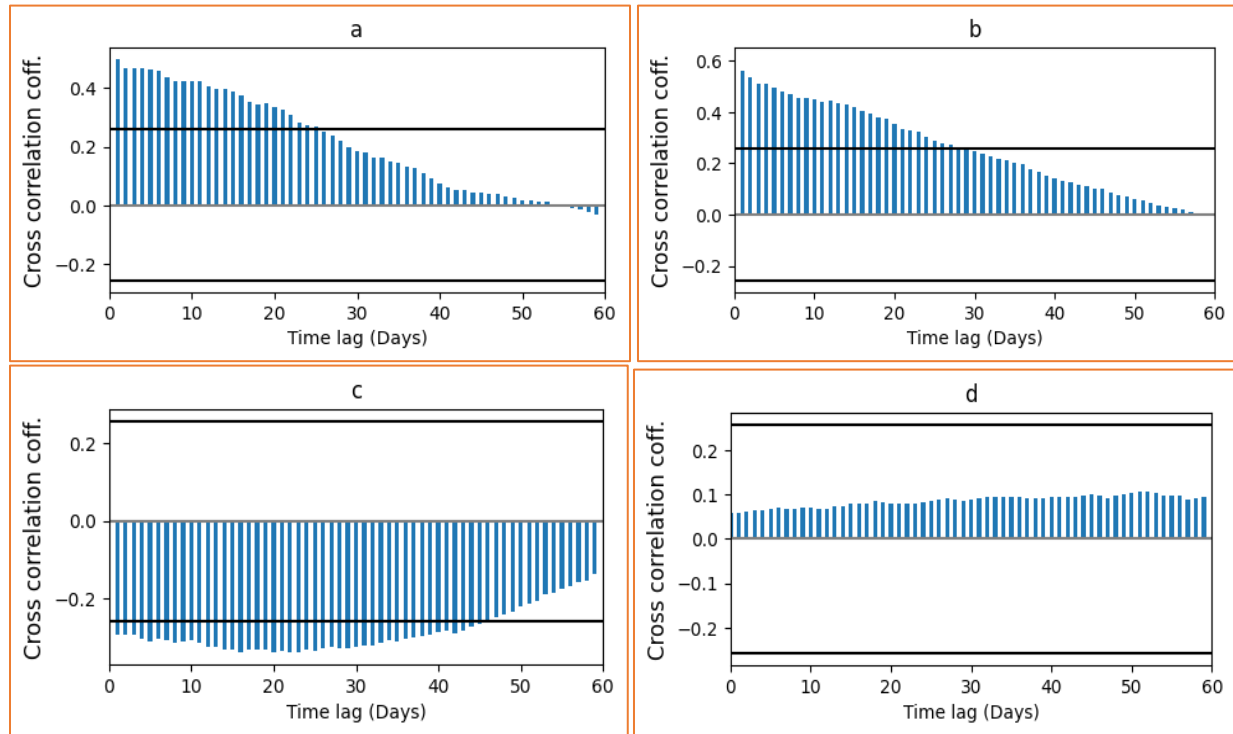


Figure 4.10. Melka Kunture Station Cross-correlation graphs: a) residual and precipitation; b) residual and observed streamflow; c) residual and maximum temperature; and d) residual and minimum temperature.

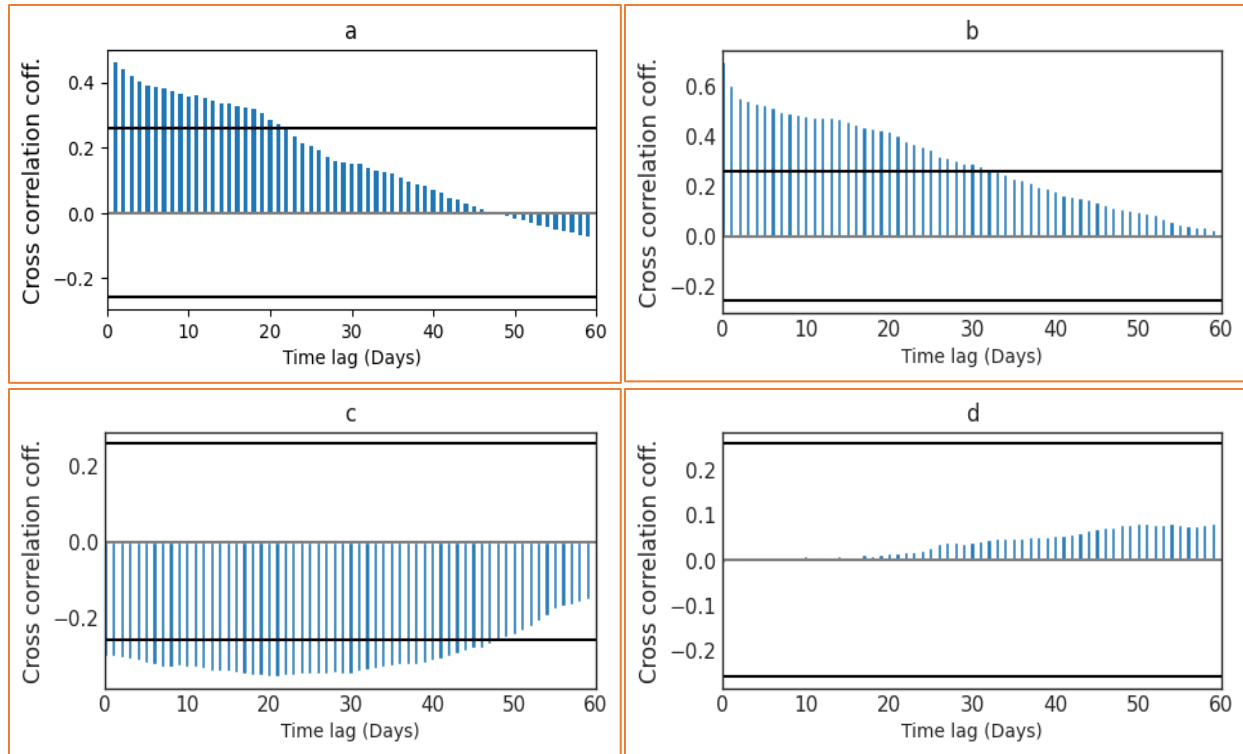


Figure 4.11. Hombole station Cross-correlation graphs: a) residual and precipitation; b) residual and observed streamflow; c) residual and maximum temperature; and d) residual and minimum temperature.

According to the cross-correlation analysis, precipitation at the time of simulation and a lag of two preceding days were chosen as significant inputs for all stations, and streamflow from lag one to lag three was found important. In the case of the maximum temperature, observation at the time of simulation and a lag of fifteen ($t-15$) and twenty ($t-20$) days were used for Melka Kunture and Hombole stations, respectively, and the minimum air temperature at the time of simulation was considered for the Akaki station. At each calibration site, the cross-correlation analysis yields completely different result. This could be due to catchment characteristics such as shape and size, which have a large impact on the time delay of the variables. Table 4.5 displays the top three models that perform well, with model number 3 being chosen for further hyperparameter optimization.

Table 4.5. Screened best models for residual simulation.

Stations	Models	Model input parameter	Model Structure (I-H-O)	Training		Testing	
				R ²	RMSE	R ²	RMSE
Akaki	1	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, T_{min})$	5 1 1	0.48	15.8	0.45	17.3
	2	$Q_{Er t} = f(P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}, T_{min})$	5 1 1	0.48	15.6	0.46	17.1
	3	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}, T_{min})$	6 1 1	0.49	15.2	0.46	16.8
Melka	1	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2})$	4 1 1	0.6	27.3	0.51	27.8
	2	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, T_{max})$	5 1 1	0.63	27.4	0.54	26.4
	3	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}, T_{max t-15})$	6 1 1	0.64	26.5	0.55	25.5
Hombole	1	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, T_{max})$	5 1 1	0.62	25.2	0.56	28.3
	2	$Q_{Er t} = f(P_t, P_{t-1}, P_{t-2}, Q_{t-1}, Q_{t-2}, T_{max t-20})$	6 1 1	0.65	24.8	0.58	27.2
	3	$Q_{Er t} = f(P_t, P_{t-1}, Q_{t-1}, Q_{t-2}, Q_{t-3}, T_{max t-20})$	6 1 1	0.66	24.5	0.58	27.4

Table 4.6. Results of the hyper parameter tuning for residual and Hybrid2.

Station	Models	Model structure	Hyperparameters			Total Training time(sec)	Validation RMSE
			Optimum No of neuron in a hidden layer	Optimizer	Learning rate		
Akaki	Residual	6-1-9-1	9	SGD	0.001	63	15.4
	Hybrid2	2-1-3-1	3	SGD	0.001	38	16.2
Melka Kunture	Residual	6-1-13-1	13	Adam	0.001	32	23.1
	Hybrid2	2-1-4-1	4	Adam	0.01	23	32.3
Hombole	Residual	6-1-3-1	3	Adam	0.01	29	25.7
	Hybrid2	2-1-5-1	5	SGD	0.001	42	27.2

In addition to the performance measure presented in table 4.5, in figure 4.12, the performance of the data-driven ANN model is presented visually and shows a good performance in estimating the residual error for both the calibration and validation periods for all stations.

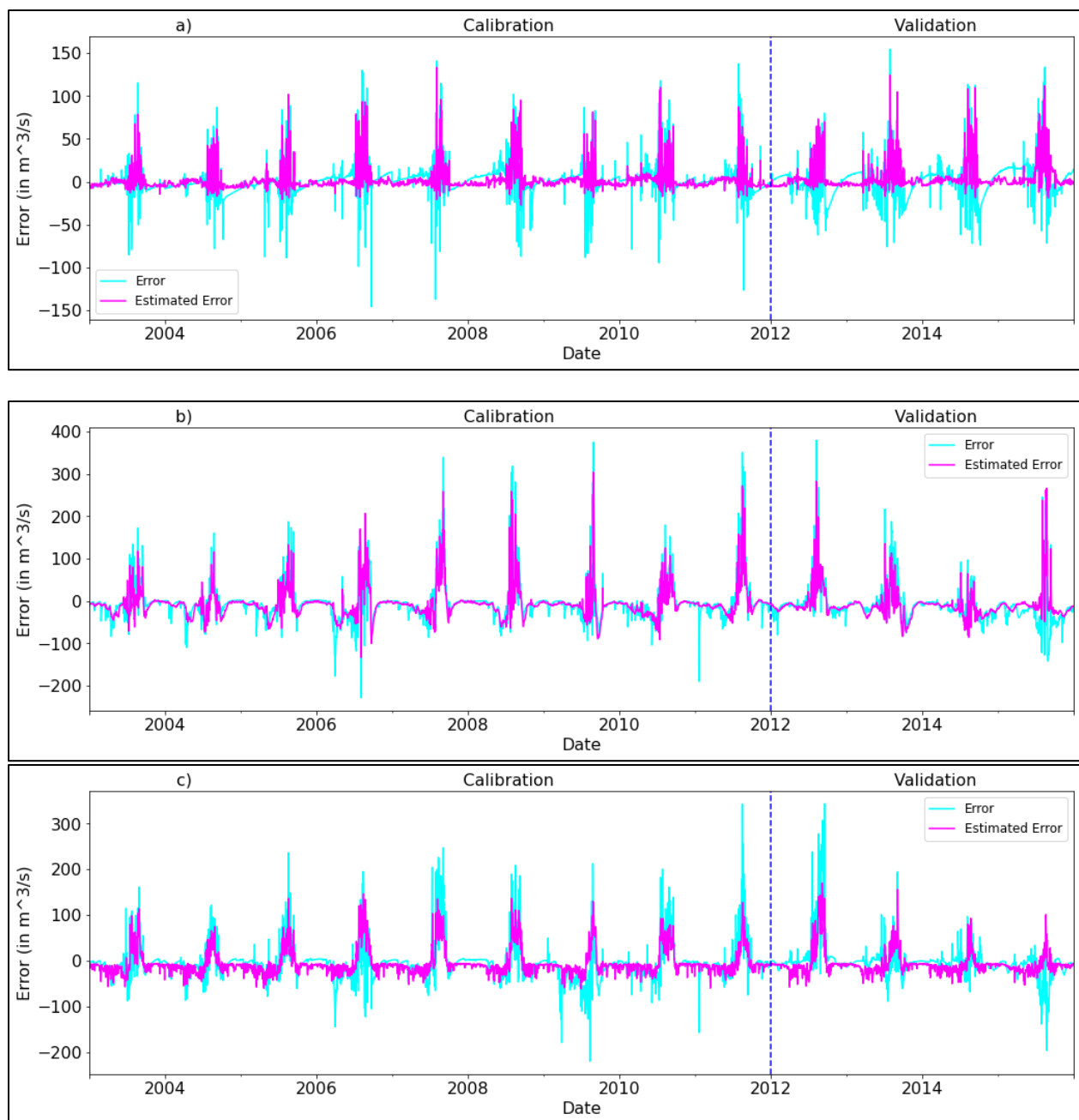


Figure 4.12. Observed and estimated residual errors time series: (a) Akaki, (b) Melka Kunture, and (c) Hombole.

To obtain the final streamflow simulation using Scheme2, the estimated results of the residual model and the SWAT model must be combined. Many studies (Abebe and Price, 2003; Wu et al. 2019, and Konapala et al. 2020) combined estimated residual error and simulated hydrological

models. However, there are some cases when the sum of the estimated residual error and the hydrological model simulation produces some unrealistic zeros and/or negative flows. The study by Lv et al., 2020, on the other hand, successfully combines the two components using the ANN model without this issue. As a result, the Hybrid2 model is the final ANN model used to combine the estimated residual error and the simulated streamflow by SWAT. Because the model inputs (simulated streamflow by SWAT and estimated residual error) are already known, training the ANN model is quite simple, and finally, the hyperparameters for Hybrid2 are optimized, as shown in Table 4.6.

4.3. Model Performance Evaluation

Three modeling approaches were used in this study to simulate the streamflow of the Upper Awash Sub-basin, and the models' performances were evaluated at the Hombole outlet and two interior stations, Akaki and Melka Kunture, located within the sub-basin. The performance evaluation indices R^2 , NSE, PBAIS, RSR, and RMSE were calculated and presented in Table 4.7 to evaluate the model performances. The model performance of all four models, namely SWAT, ANN, Hybrid1, and Hybrid2, was evaluated by comparing the calculated performance indices with the updated model performance evaluation criteria of Marias et al. (2015).

4.3.1 Hybrid model performance at the selected calibration sites

In the Akaki watershed, which covers an area of 844.4 km², the performance of two hybrid models, the stand-alone ANN and SWAT model was evaluated. The evaluation results also revealed that the hybrid model outperformed the physical-based SWAT model in simulating streamflow at the watershed. According to the model performance rating criteria reported in table 3.7, the SWAT model lies in satisfactory range except for the R^2 value, and the model has the performance metrics of R^2 , NSE, and PBAIS of 0.53 to 0.52 and 13.4 respectively for the calibration period and 0.54, 0.50, and 16.4 for the validation period. The performance of the SWAT also showed a similar result to the study of Tessema (2011) for daily flow simulation at the watershed and equivalent performance to many studies reported in the study area.

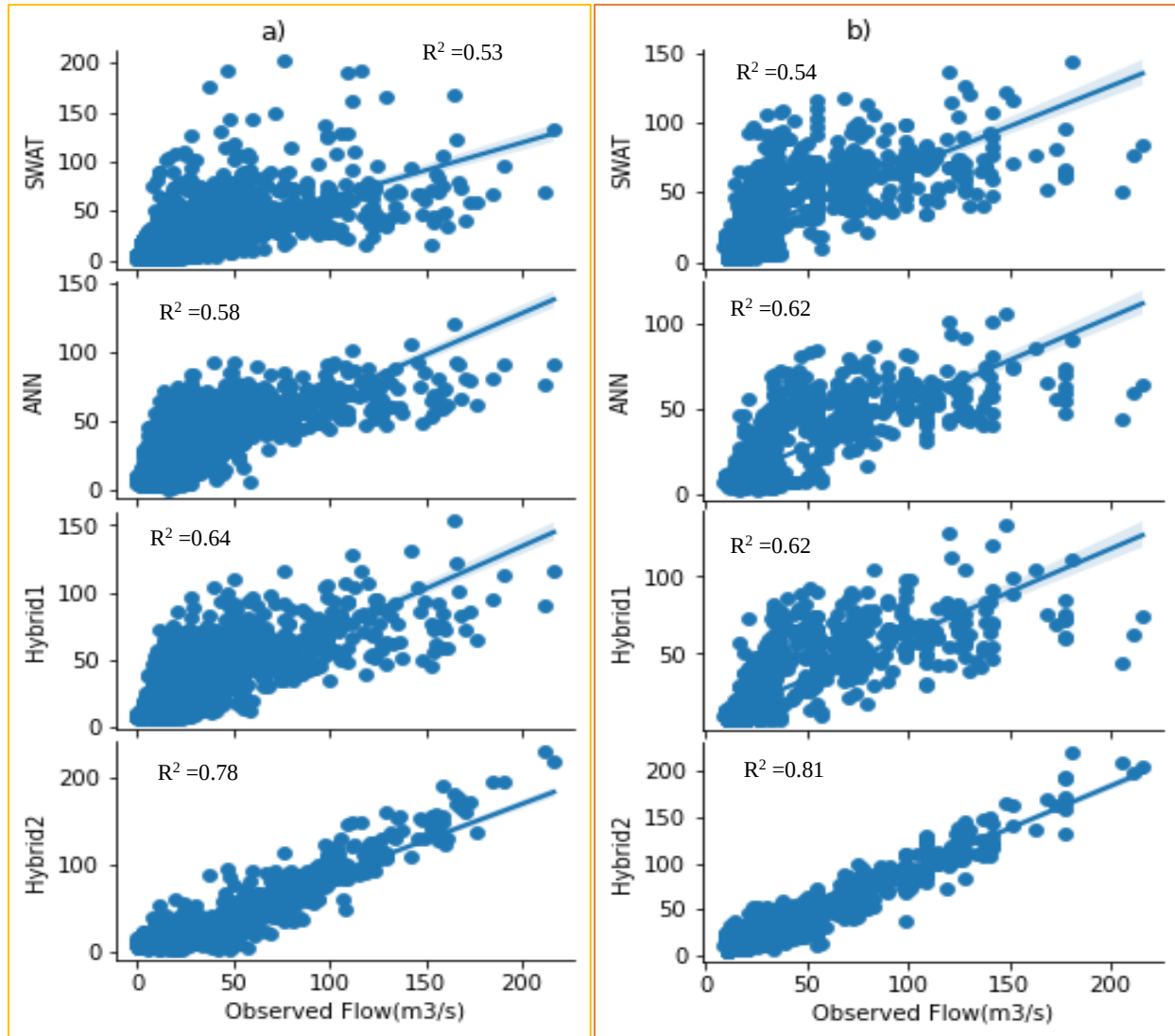


Figure 4.13. Observed and simulated streamflow scatter plots of all models at Akaki station: (a) scatter plot during the calibration period, and (b) scatter plot during the validation period. The y-axis of the graphs represents the simulated flow in (m^3/s) for the indicated models.

Similarly, the performance of the stand-alone ANN model for the watershed is satisfactory, with R^2 , NSE, and PBAIS of 0.58, 0.58, and 2.4, respectively, for the calibration period and 0.62, 0.41, and 41.3 for the validation period. Unlike the SWAT model, the ANN model shows less performance in the validation period. In the case of hybrid models, the performance was satisfactory and good for Hybrid1 and Hybrid2 models, respectively. The performance indices R^2 , NSE, and PBAIS for Hybrid1 were 0.64, 0.64, and -1.39 for the calibration period and 0.62, 0.54

and 26 for the validation period, respectively. Likewise, the Hybrid2 model also showed a tremendous improvement with R^2 , NSE, and PBAIS of 0.78, 0.72, and -0.89, respectively, for the calibration period and 0.81, 0.72, and -4.9, respectively, for the validation period, and the negative value of the percent of bias in the Hybrid2 indicates the underestimation of the total simulated flow. For a better understanding, in figures 4.13 and 4.14, the scatter and hydrograph were plotted against the observed values of streamflow for the calibration/ training, and validation periods.

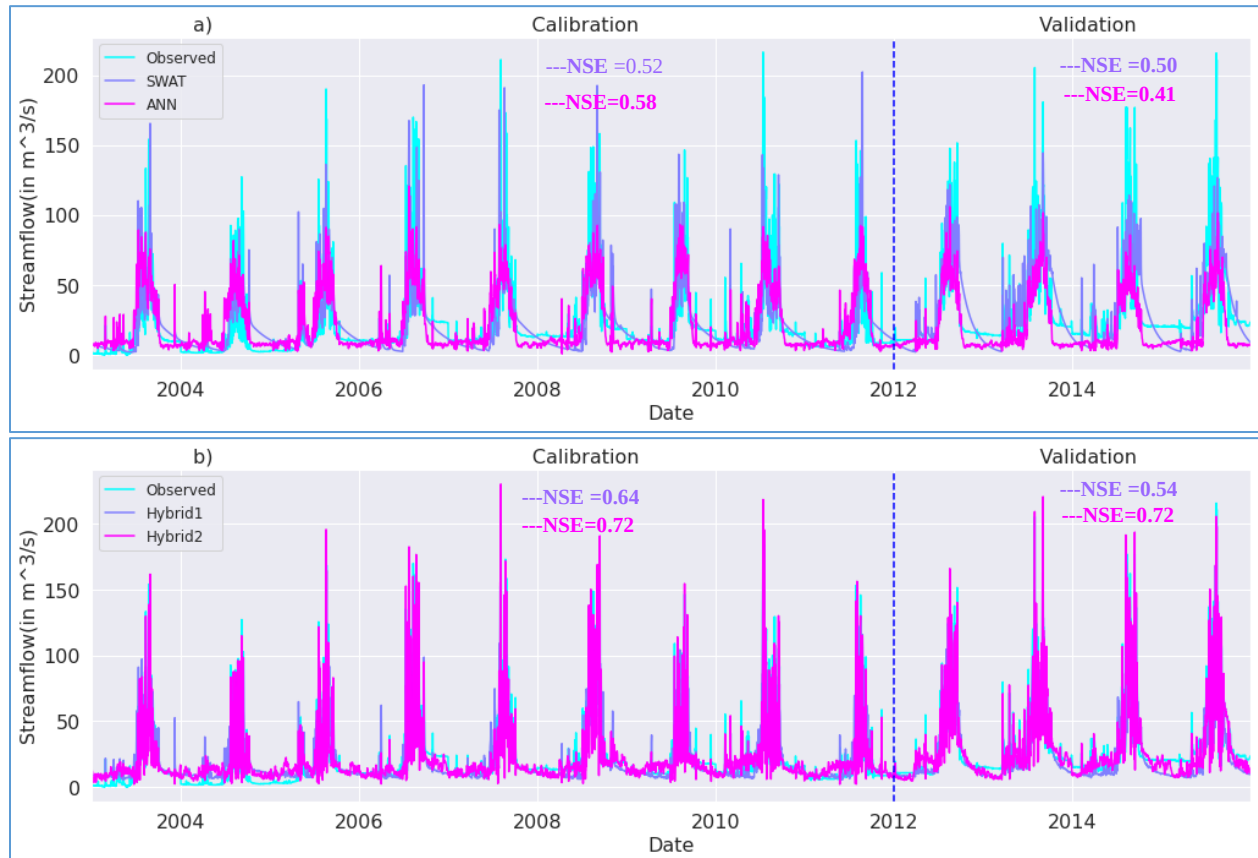


Figure 4.14. Hydrographs of Akaki Station during the calibration and validation period: (a) for observed streamflow, SWAT and ANN models; and (b) for observed streamflow, Hybrid1 and Hybrid2.

The performance of four models was also evaluated at the second calibration site, Melka Kunture, which has a total area of 4456 km². The performance of the physical-based SWAT model at the watershed was satisfactory for both calibration and validation in all performance indices except for the percent of bias in the validation period. The SWAT model has the performance metrics of

R^2 , NSE, and PBAIS of 0.65 to 0.64 and -11.8, respectively, for the calibration period and 0.61, 0.57, and -20.9 for the validation period. The performance of the stand-alone ANN model for the watershed is within the good range for the calibration period except for the value of NSE and in the satisfactory range for the validation period, with R^2 , NSE, and PBAIS of 0.78, 0.69, and 0.77, respectively, for the calibration and 0.67, 0.59, and 7.37, respectively, for the validation period.

In the case of hybrid models, the performance of the model range was good for the calibration period and satisfactory for the validation period. The performance indices R^2 , NSE, and PBAIS for Hybrid1 were 0.81, 0.79, and 1.61, respectively, for the calibration period and 0.72, 0.7, and -14.5, respectively, for the validation period. In the case of the Hybrid2 model, it also showed a significant improvement, with very good results for the calibration period and good results for the validation period, with R^2 , NSE, and PBAIS of 0.87, 0.86, and 0.85, respectively, for the calibration period and 0.72, 0.72, and -1.25, respectively, for the validation period.

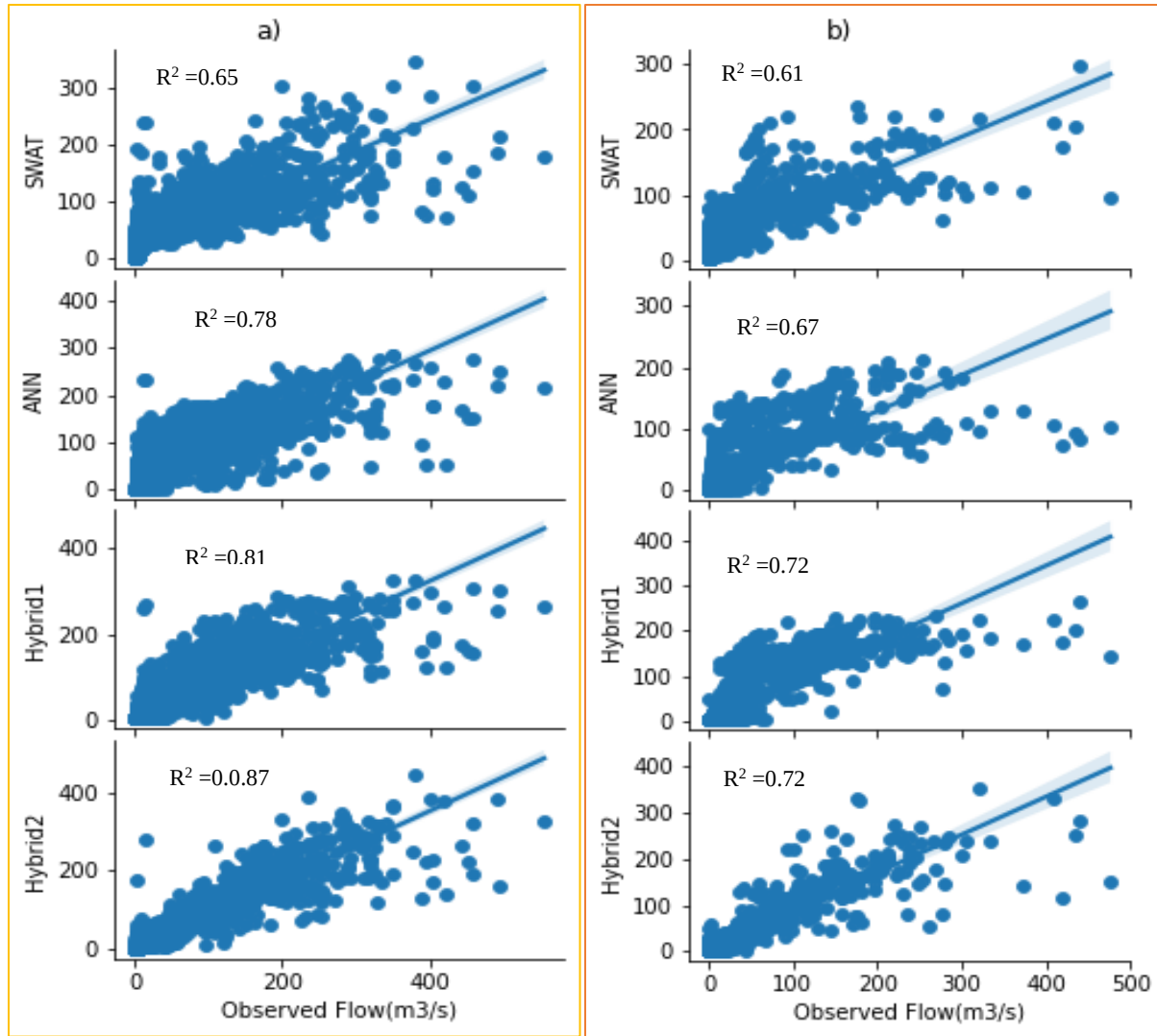


Figure 4.15. Observed and simulated streamflow scatter plots of all models at Melka Kunture station: (a) scatter plot during the calibration period, and (b) scatter plot during the validation period. The y-axis of the graphs represents the simulated flow in (m³/s) for the indicated models.

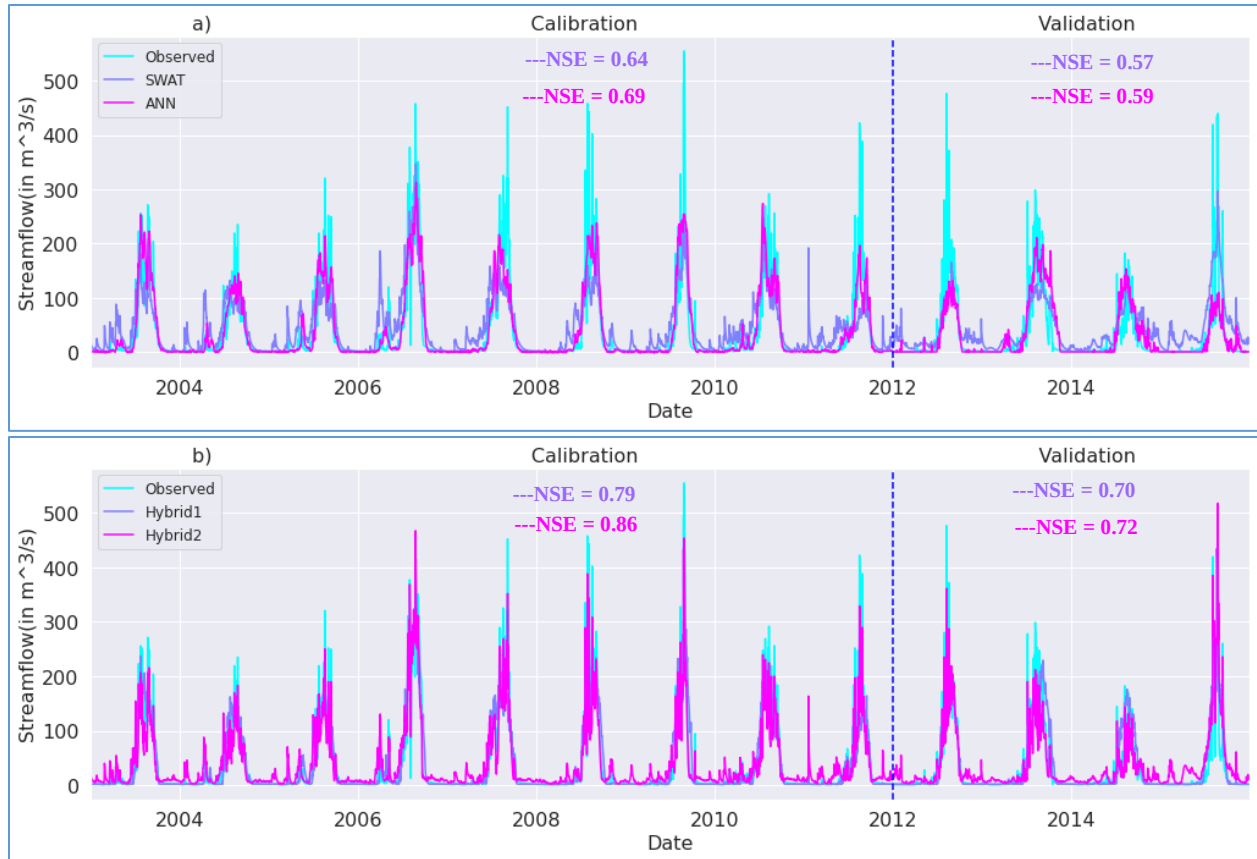


Figure 4.16. Hydrographs of Melka Kunture Station during the calibration and validation period: (a) for observed streamflow, SWAT and ANN models; and (b) for observed streamflow, Hybrid1 and Hybrid2.

Finally, the performance of the models was evaluated at the Hombole outlet, which has a total area of 7656 km^2 . According to the performance evaluation result, the SWAT model has good performance, with the performance indices of R^2 , NSE, and PBAIS of 0.71, 0.69, and -9.4 for the calibration period, and 0.69, 0.65, and 16.9 for the validation period. The performance of the stand-alone ANN model at this outlet was also found within the very good range for the calibration period, except for the value of R^2 , which was found in the range of good. The results of performance were found to be 0.82, 0.74, and 3.8 for R^2 , NSE, and PBIAS, respectively, for the calibration period. However, for the validation period, the performance of the ANN model was substantially decreased to satisfactory levels, with R^2 , NSE, and PBAIS of 0.68, 0.69, and -4.2, respectively. In the case of hybrid models, the performance of the model was very good. The performance indices R^2 , NSE, and PBAIS for Hybrid1 were 0.83, 0.8, and 0.2, respectively, for

the calibration period and 0.79, 0.79, and -12.2, respectively, for the validation period. Likewise, the performance of the Hybrid2 model also showed a remarkable improvement, with very good results for both the calibration and validation periods, with R^2 , NSE, and PBAIS of 0.92, 0.91, and 1.17, respectively, for the calibration period and 0.87, 0.86, and 3.9, respectively, for the validation period. In the case of hybrid models, the performance of the model was found to be very good.

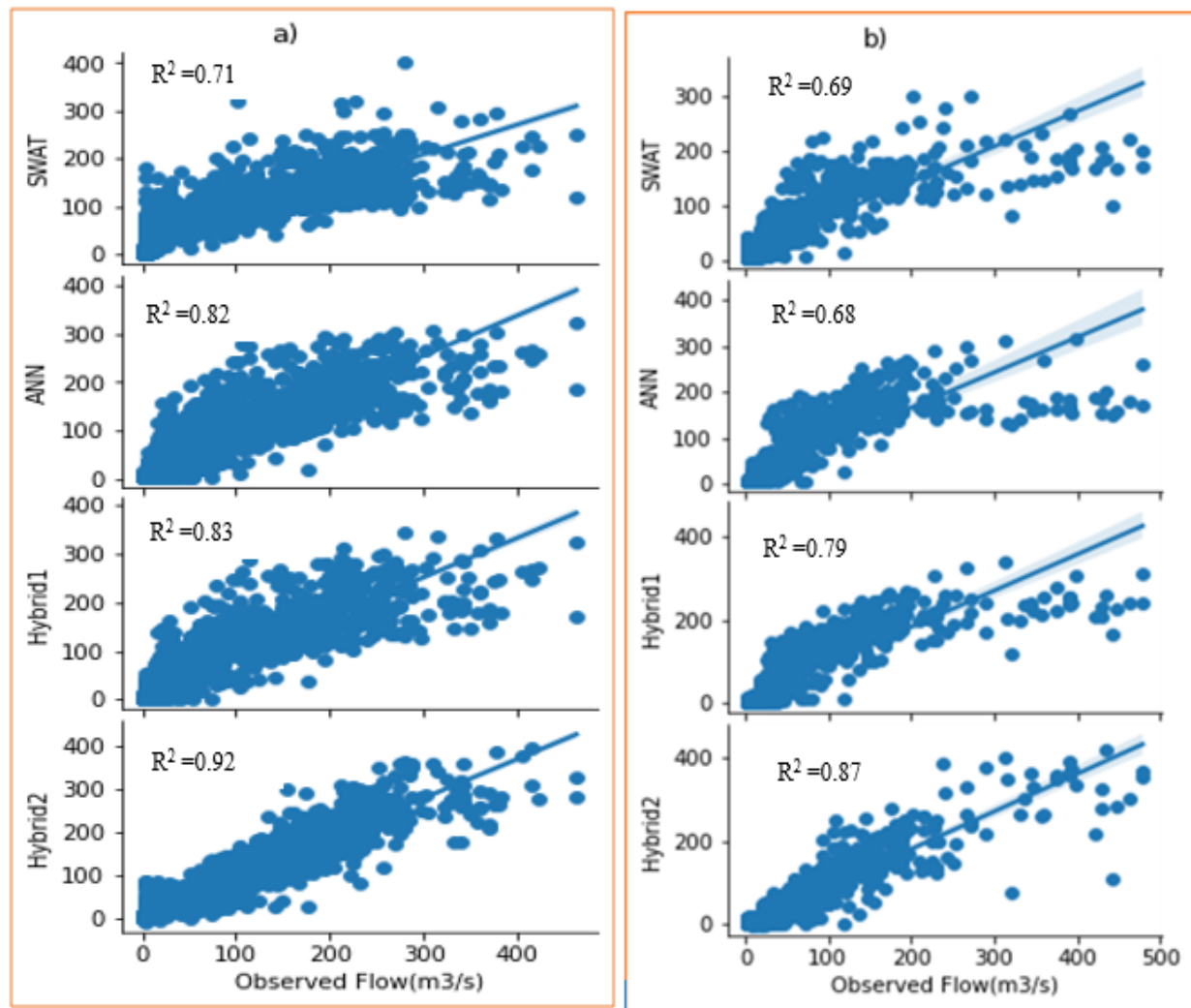


Figure 4.17. Observed and simulated streamflow scatter plots of all models at Hombole station: (a) scatter plot during the calibration period, and (b) scatter plot during the validation period. The y-axis of the graphs represents the simulated flow in (m^3/s) for the indicated models.

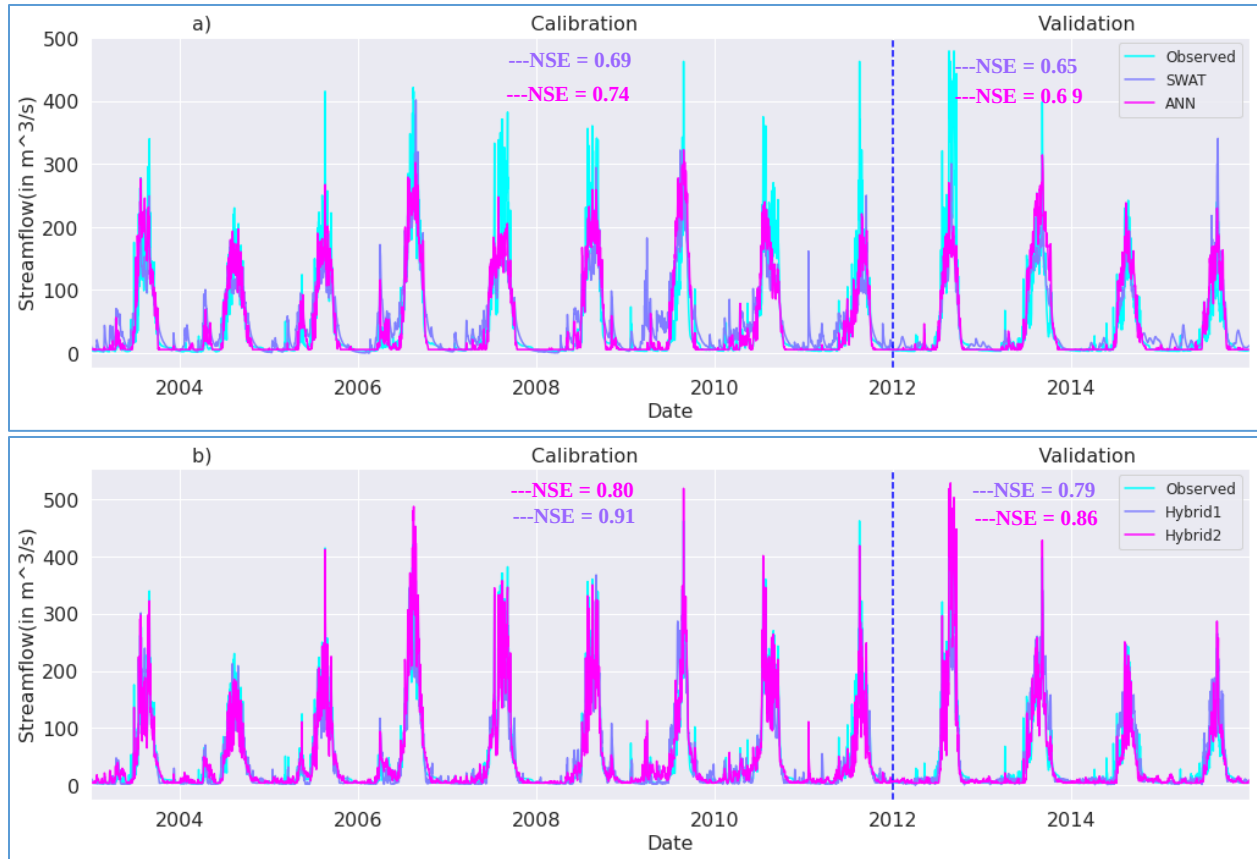


Figure 4.18. Hydrographs of Hombole Station during the calibration and validation period: (a) for observed streamflow, SWAT and ANN models; and (b) for observed streamflow, Hybrid1 and Hybrid2.

The model performance evaluation revealed that, with the exception of Hybrid2, all models underestimated the observed total streamflow at the Akaki watershed, resulting in a higher percent of bias for the validation period than for the calibration period. The SWAT model underestimates the flow during the calibration period at Melka Kunture, whereas ANN, Hybrid1, and Hybrid2 overestimate the flow. Except for ANN, all other models overestimate the flow at this watershed during the validation period. Similarly, with the exception of the SWAT model, all models at Hombole underestimate streamflow; however, during the validation period, SWAT and Hybrid2 underestimate flow, while ANN and Hybrid1 overestimate it.

Table 4.7. Streamflow simulation performance statistics for three models at the Akaki, Melka Kunture, and Hombole hydrometric gauging stations from 2003 to 2011 (calibration period) and 2012 to 2015 (validation period).

	Performance Indicators	Akaki				Melka Kunture				Hombole			
		$Cal_Obs \bar{Q} (m^3/s) = 21$ $Val_Obs \bar{Q} (m^3/s) = 32$				$Cal_Obs \bar{Q} (m^3/s) = 38.1$ $Val_Obs \bar{Q} (m^3/s) = 30.6$				$Cal_Obs \bar{Q} (m^3/s) = 46$ $Val_Obs \bar{Q} (m^3/s) = 39$			
		SWAT	ANN	Hybrid1	Hybrid2	SWAT	ANN	Hybrid1	Hybrid2	SWAT	ANN	Hybrid1	Hybrid2
Calibration period	R ²	0.53	0.58	0.64	0.78	0.65	0.78	0.81	0.87	0.71	0.82	0.83	0.92
	NSE	0.52	0.58	0.64	0.72	0.64	0.69	0.79	0.86	0.69	0.74	0.8	0.91
	PBIAS	13.4	2.4	-1.3	-0.89	-11.8	0.77	1.61	0.85	-9.4	3.8	0.2	1.17
	RSR	0.65	0.64	0.6	0.53	0.6	0.54	0.45	0.35	0.5	0.4	0.42	0.3
	RMSE (m ³ /s)	16.5	16.5	15.4	13.5	42.6	38.1	32.3	25.8	40	30.5	32.2	23.7
	Sim_ $\bar{Q}$ (m ³ /s)	18.8	19.8	20.6	20.6	42.6	37.1	37.5	37.8	50.9	44.1	46	45.2
Validation period	R ²	0.54	0.62	0.62	0.81	0.61	0.67	0.72	0.72	0.69	0.68	0.79	0.87
	NSE	0.50	0.41	0.54	0.72	0.57	0.59	0.7	0.72	0.65	0.69	0.79	0.86
	PBIAS	16.4	41.3	26	-4.9	-20.6	7.37	-14.5	-1.25	16.9	-4.2	-12.2	3.9
	RSR	0.67	0.76	0.68	0.52	0.65	0.63	0.54	0.53	0.58	0.55	0.46	0.37
	RMSE (m ³ /s)	20.6	23.7	21	16.2	41.7	38.6	33.1	32.3	40.9	38.9	33.1	27.2
	Sim_ $\bar{Q}$ (m ³ /s)	29	19.3	24.2	34.4	41.8	28.3	35.1	30.78	41.1	41	44.2	37.1

Where Sim_ $\bar{Q}$ = mean simulated streamflow, Cal_Obs $\bar{Q}$ = calibration period mean streamflow and, Val_Obs $\bar{Q}$ = validation period mean streamflow.

In general, poor model performance was observed in the Akaki watershed, and this may be due to the poor quality of the gauged climate variables, which leads to a poor correlation between climate variables and streamflow in the Akaki watershed. Because the performance of the data-driven model is more sensitive to data quality than physical-based models, the ANN model has less

performance than the physical-based SWAT model at Akaki Watershed. As shown in figure 4.5, the correlation between rainfall and streamflow was poorer in the Akaki watershed than in the Melka Kunture and Hombole. On the other hand, the performance of the models also shows increasing performance with the increase in the watershed area. According to Tessema (2011), the possible explanation for the good model performance of the SWAT model in the Hombole outlet is the aggregation of discharges from various upstream sub-watersheds to a larger area, including the Akaki and Melka Kunture watersheds, which may help in balancing (canceling out) the bias that occurred at various steps; however, this effect is not as significant in a small watershed like Akaki. Shawul et al. (2019) and Woldesenbet et al. (2017) also reported in their studies that the underestimation of flow during high flow conditions and overestimation during low flow conditions of the SWAT model were also observed in this study, and the possible explanation for the inconsistency of the model's simulation of peak events could be due to a lack of sufficient consideration of the rainfall intensity on runoff generation in the CN method, which is used in the SWAT model.

A recent study by Sikorska-Senoner and Quilty (2021) proposed combining a physical-based hydrological model with a data-driven model to achieve high output quality and reported that the application of a hybrid modeling approach gives significant model improvements. As a result, a Hybrid1 model successfully simulates streamflow in this study by combining the benefits of a physical-based model (intended to respect physical connections between hydrological processes) with the capabilities of a data-driven model (which can model complex relationships between input and target variables), and it shows improvement over SWAT and stand-alone ANN models. For instance, the average calibration and validation period RMSE of the SWAT and stand-alone ANN models was reduced by 2% and 10%, respectively, at the Akaki watershed by using the Hybrid1 model. In the Melka Kunture watershed, the average RMSE of the SWAT and stand-alone ANN models was reduced by 22% and 14.7%, respectively; similarly, in the Hombole watershed, the RMSE was significantly reduced by 20% and 6%, respectively, after the application of the Hybrid1.

Ghaith et al. (2019) reported similar improvements in streamflow forecasting at the Guadalupe River Watershed in Texas by integrating the lumped conceptual HYMOD hydrological model with

the ANN model, and that the developed hybrid model reduced the RMSE of the HYMOD model by 30% for the validation period. Likewise, Humphrey et al. (2016) developed hybrid models by combining the Bayesian ANN model and the GR4J conceptual rainfall-runoff model for the South East region of South Australia and reported that the hybrid model had higher accuracy and lower uncertainty for streamflow forecasts than the GR4J model, with a reduction of the RMSE of the GR4J model by more than 45% in many watersheds inside the study area. Konapala et al. (2019) also reported improved performance of streamflow simulations and the capability of hybrid hydrological models in the simulation of streamflow at several catchments in the United States by combining the LSTM and the Sacramento Soil Moisture Accounting Model (SAC-SMA).

In the case of Scheme2 (Hybrid2), which uses a post-processing approach, there was also a promising improvement for all watersheds. The RMSE of the Hybrid2 model at the Akaki watershed is 16.2, which is lower than the original SWAT (20.6). The SWAT model has a PBIAS of 16.4, whereas the Hybrid2 model has a PBIAS that is much closer to zero relatively (-4.9), indicating that the Hybrid2 model underestimates the observations while the SWAT model overestimates.

Similarly, at the Melka Kunture watershed, the Hybrid2 model reduced the RMSE from 41.7 to 32.3, while the PBIAS was reduced from -20.6 to -1.25, whereas the hybrid model reduced the RMSE of SWAT from 40.9 to 27.2 at Hombole, while the PBIAS was reduced from 16.9 to 3.6, a result that is much closer to zero because the hybrid model better simulates the true observed flow than the SWAT model. Therefore, even if the Hybrid2 model underestimates the flow, the smaller magnitude of the PBIAS in the Hybrid2 model is closer to zero, which means it is close to the ideal solution.

4.3.2 Hybrid models performance for extreme events

Even though the performance of the hybrid models used in this study is dependent on the performance of the physical-based SWAT model, the hybrid models significantly improved in streamflow simulation. To determine the significance of the hybrid modeling approach over stand-alone models, the performance of each model was evaluated using the quantile domain. For climate change impact assessments and water resource management planning, a robust hydrological model

should reproduce the observed discharges in the quantile domain associated with low and high flows. Hydrological low-flow periods and droughts have an impact on water supply for drinking, irrigation, industrial needs, hydropower production, and ecosystems; thus, peak flow analysis is useful primarily for flood protection and water resource planning in the study basin (Tegegne et al., 2017).

To get a better understanding of the capability of the models to mimic the extreme events, for each station, a box and whisker plot with extreme data points is presented for both the calibration and validation periods. The height of the box portion is determined by the dataset's interquartile range, which extends from the 25th to the 75th percentiles, with the central red line representing the median values and outliers labeled with a circle dot at the top of the whisker points. SWAT represented the high flow outlier better than the Hybrid1 and stand-alone ANN models in the Akaki watershed during both the calibration and validation periods, and the use of a post-processing approach increased the capability of the SWAT model to a greater extent.

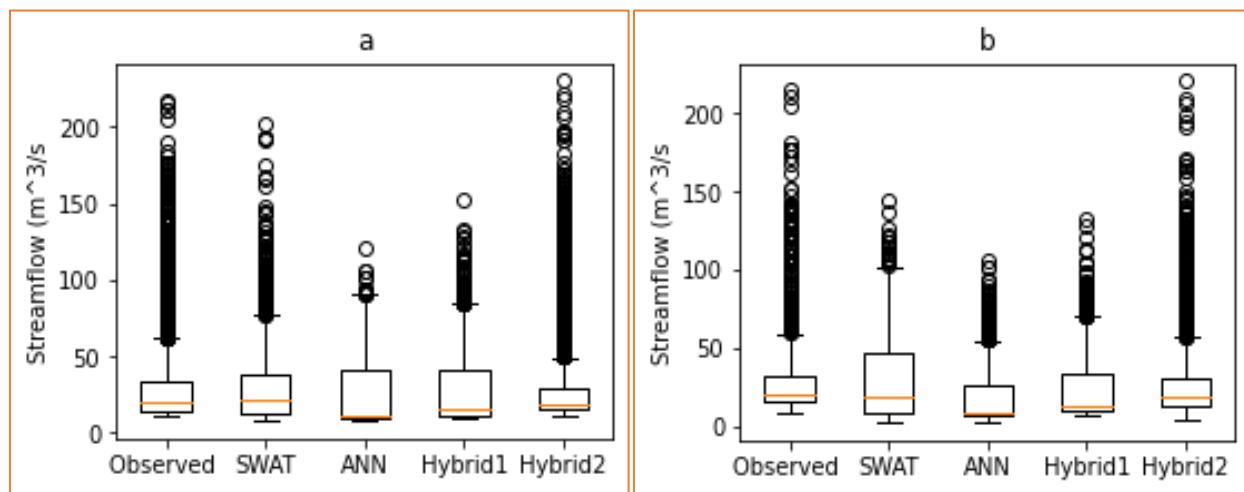


Figure 4.19. Boxplots of observed versus simulated streamflow for the Akaki station, a) for the calibration period and b) for the validation period.

At Melka Kunture and the whole sub-basin at Hombole outlet, results were similar to Akaki watershed; however, for low-flow ANN and hybrid models, performance was better, as represented by the box and whisker plot.

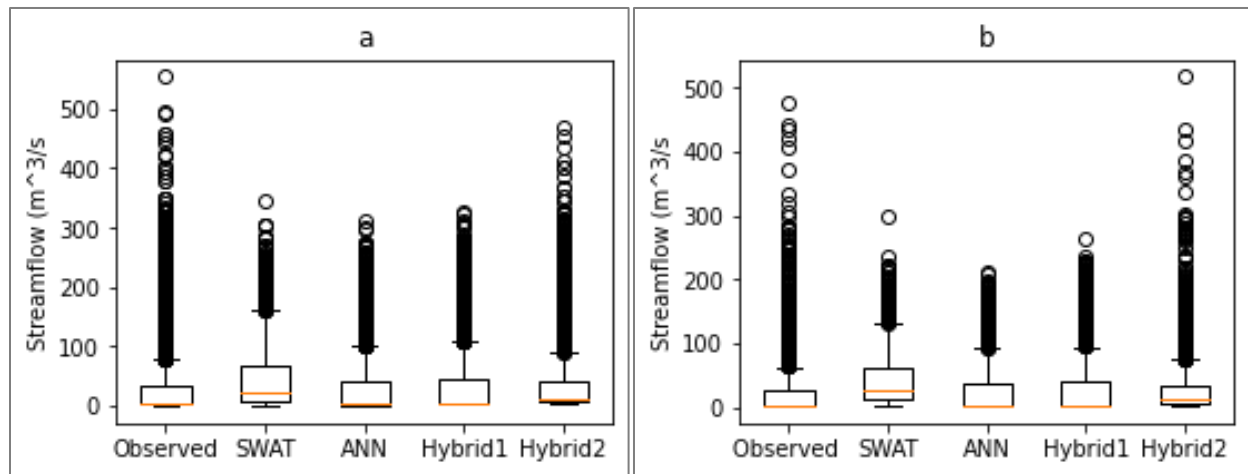


Figure 4.20. Box-plots of the observed streamflow compared with simulated for the Melka Kunture station a) for the calibration period b) for the validation period

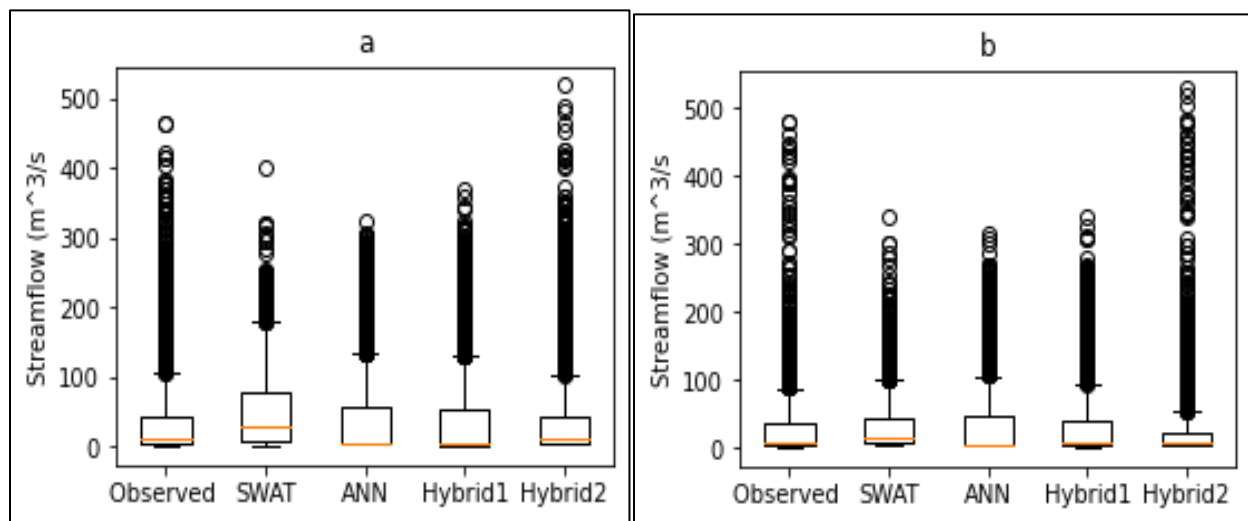


Figure 4.21: Box plots of observed versus simulated streamflow for the Hombole station, a) for the calibration period and b) for the validation period.

In order to see the performance of the models, quantile domain-based model performance was applied using the identified intervals that can be used as broad indicators of hydrologic conditions. There are several types of quantile curve intervals to represent different flow conditions. For the validation period, the quantile curve is typically divided into five zones: high flows (0-10%), moist

flows (10-40%), mid-range flows (40-60%), dry conditions (60-90%), and low flows (90-100%). (Tegegne et al., 2017). The observed and predicted flow intervals by models are given in Table 4.8.

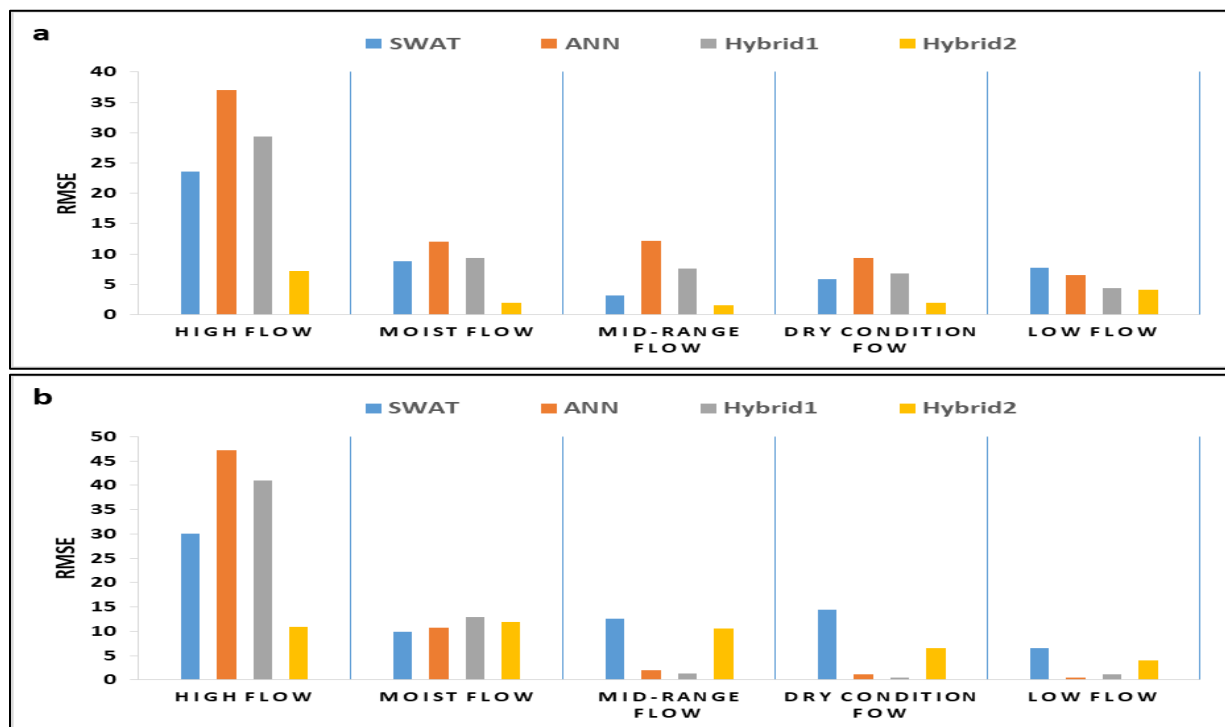
Table 4.8. Classifications of hydrological flow conditions

Stations		Flow conditions intervals				
		High flow (0-10)%	Moist flow (10-40)%	Mid-range flow (40-60)%	Dry conditions (60-90)%	Low flow (90-100)%
Akaki	Observed (m ³ /s)	215.7 - 73.4	73.4 - 23.3	23.3 - 20	20 - 13	13 - 8.9
	SWAT	144.4 - 67.8	67.8 - 27.4	27.4 - 14.1	14.1 - 4.4	4.4 - 1.9
	ANN	105.8 - 53.4	53.4 - 9.8	9.8 - 7.9	7.9 - 6.1	6.1 - 2.5
	Hybrid1	132.8 - 61.6	61.6 - 16	16 - 11.2	11.2 - 8.2	8.2 - 6.7
	Hybrid2	220.7 - 65.2	65.2 - 22	22 - 17.3	17.3 - 8.1	8.1 - 4.3
Melka Kunture	Observed (m ³ /s)	476 - 117	117 - 4.8	4.8 - 2.3	2.3 - 0.7	0.7 - 0.5
	SWAT	296 - 101	101 - 32.6	32.6 - 20.6	20.6 - 6.8	6.8 - 2.7
	ANN	210 - 103.8	103.8 - 5.8	5.8 - 1.5	1.5 - 0.51	0.51 - 0.4
	Hybrid1	262.7 - 147.3	147.3 - 2.9	2.9 - 1.9	1.9 - 1.82	1.82 - 1.5
	Hybrid2	517.9 - 107.1	107 - 18	18 - 10.5	10.5 - 4.9	4.9 - 2.74
Hombole	Observed (m ³ /s)	478 - 125	125 - 14.6	14.6 - 5.8	5.8 - 3.3	3.3 - 2.4
	SWAT	340 - 133	133 - 20.8	20.8 - 11.6	11.6 - 5.2	5.2 - 2.4
	ANN	314 - 155	155 - 7.4	7.4 - 5.7	5.7 - 5.5	5.5 - 5.1
	Hybrid1	341 - 164	164 - 9.93	9.9 - 7.5	7.5 - 3.9	3.9 - 1.2
	Hybrid2	528 - 118	118 - 8.4	8.4 - 5.5	5.5 - 3.4	3.4 - 3.2

After the flow condition is identified, the performance of the models for these five flow conditions is evaluated and presented in terms of RMSE. For a further visual illustration, the flow duration curve and the flow conditions with the corresponding RMSE were presented for all models for the validation period. As per the evaluation result, the SWAT model outperforms Hybrid1 and the stand-alone ANN model in the simulation of high flow, mid-range flow, and dry condition flow in this Akaki watershed, according to the quantile domain-based model performance evaluation results; the SWAT model produces a lower RMSE than the stand-alone ANN and Hybrid1 for high-flow conditions; however, in moist and low flow conditions, the Hybrid1 model performs

better with a smaller RMSE. On the other hand, the post-processing approach based on the SWAT model output significantly improved SWAT model performance and reduced RMSE, as shown in Figure 4.22 a.

The SWAT model simulation represents high flow in the Melka Kuntura watershed better than the other models; however, Hybrid1 and the stand-alone ANN model for moist, mid-range, dry, and low flow conditions show better simulation performance with minimal RMSE. In the case of low-flow extremes, ANN and hybrid models outperformed SWAT and Hybrid2. Even though the SWAT model's performance improved with Hybrid2, the RMSE obtained by Hybrid2 is not as low as that of stand-alone ANN and Hybrid1.



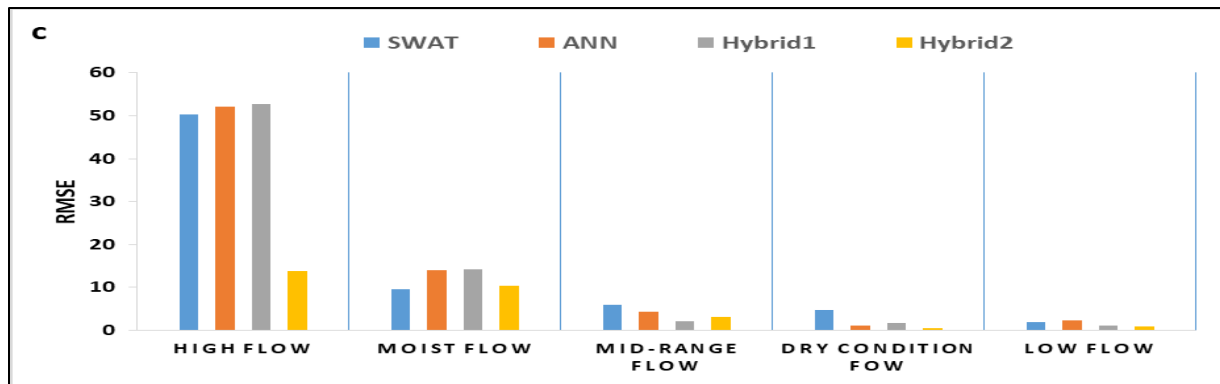


Figure 4.22. Summary of model performance in terms of root mean squared error (RMSE) for daily streamflow simulations, a) Akaki, b) Melka Kunture, and c) Hombole for identified hydrological flow conditions

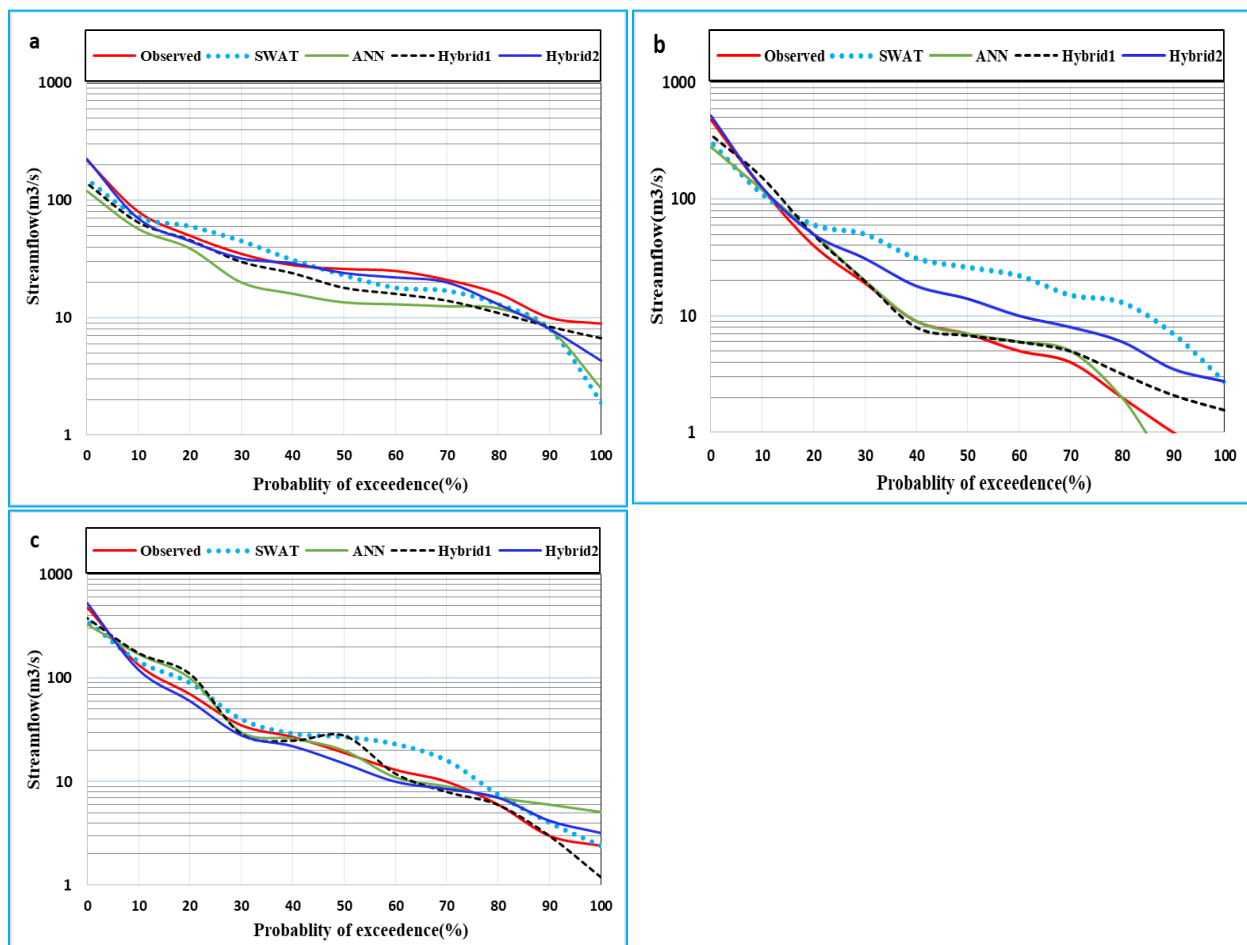


Figure 4.23. Validation period flow duration curves of models for three stations: (a) Akaki, (b) Melka Kunture, and (c) Hombole.

At Hombole SWAT model outperforms stand-alone ANN and Hybrid1 in simulating high and moist flow conditions, while stand-alone ANN outperforms Hybrid1 in simulating dry flow conditions. In the case of mid-range and low flow, Hybrid1 shows a lower RMSE than the others. For dry flow conditions, stand-alone ANN models give better results, but in the case of the Hybrid2 model, the result demonstrates the astonishing performance in all flow conditions except for mid-range flow. The flow duration curve results show that a Hybrid2 model overestimated the peak discharge, whereas a hybrid-1 model underestimated the peak flow in all watersheds, as shown in figure 4.23.

In general, the physical-based SWAT model performed better when simulating high and moist flows, but it performed poorly when simulating mid-range, dry, and low flows. This is because the NSE values were used to calibrate the SWAT model. Since the NSE method gives more weight to high flows than low flows, the calibration process will be more focused on high flows. The calibrated result of the SWAT model by NSE was further calibrated for hybrid models by using RMSE as an objective function, which shifted the emphasis toward lower flows. As a result, the findings of this study of a specific basin with day-to-day flow simulation demonstrate that the hybrid model is also applicable to robust rainfall-runoff models for climate change assessment.

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

In a developing country like Ethiopia, where there is a high demand for water and the frequency of extreme events is increasing, having accurate and relevant hydrological information regarding the volume and variations of streamflow across the time dimension is significant. However, due to an incomplete model structure, insufficient data, measurement errors, and inadequate initial and boundary conditions, both physical-based and data-driven models sometimes show limited performance in streamflow simulation.

As briefly discussed in the literature, recent developments and new hydrological insights suggest that the hybrid modeling approach can overcome deficiencies in both physical-based and data-driven models for capturing streamflow dynamics in watersheds. As a result, this study was motivated to assess the robustness of a hybrid modeling approach for streamflow simulations. In this study, a hybrid modeling approach with two different schemes (Schime1 and Schime2) that integrate a physical-based SWAT model and a data-driven model (ANN) was used to simulate daily streamflow in the Upper Awash Sub-basin, and the models were calibrated at multiple sites, including the Hombole outlet and two interior stations located within the Upper Awash Sub-basin, Akaki and Melka Kunture.

Three different calibration sites were used to test a physical-based SWAT model, a data-driven MLPNN model, and a hybrid model. The models were created using split data record evaluation, which included the calibration and validation periods (roughly two-thirds for calibration and one-third for validation). The two hybrid models are calibrated for streamflow simulation: the Hybrid1 model, which simulates streamflow using a time series of climate variables as well as the SWAT model output via the ANN model, and the Hybrid2 model, which is based on the post-processing approach and relies on estimating the residuals of the SWAT model using the forcing climate variables. The model's performance was evaluated using both statistical and graphical techniques, and the statistical indices R^2 , NSE, PBAIS, RSR, and RMSE were calculated and presented to

evaluate the performance of the models at different calibration sites. The model performance of all four models, namely the two hybrid models (Hybrid1 and Hybrid2), the physical-based SWAT model, and the stand-alone ANN model, was evaluated using the updated model performance evaluation criteria of Marias et al. (2015). In addition to the total time series, the model performance was also evaluated for different hydrological flow conditions based on the flow quantiles.

Based on the model performance evaluation result in the Upper Awash Sub-basin, the following conclusions were drawn: For the Akaki, Melka, Kuntura, and Hombole watersheds, the model efficiency for daily streamflow was obtained for Hybrid1 (NSE = 0.64, 0.79, and 0.8), Hybrid2 (NSE = 0.82, 0.86, and 0.91), SWAT (NSE = 0.52, 0.64, and 0.69), and ANN (NSE = 0.58, 0.69, and 0.74) for the calibration period. Similarly, for the validation period, Hybrid1 (NSE = 0.54, 0.7, and 0.79), Hybrid2 (NSE = 0.72, 0.73, and 0.86), SWAT (NSE = 0.5, 0.57, and 0.65), and ANN (NSE = 0.41, 0.59, and 0.69) were obtained for the Akaki, Melka Kuntura, and Hombole watersheds, respectively. For extreme flow simulation, Hybrid2 and the physical-based SWAT model performed better when simulating high and moist flows, but the SWAT model performed poorly when simulating mid-range, dry, and low flows. However, Hybrid1 and stand-alone ANN models were found to be good at simulating the lower flow extremes.

Generally, model evaluation performance results confirm that hybrid models can combine the strengths of physical-based and data-driven models. Physical-based models encode a scientific understanding of the hydrology system that is difficult to represent in the data-driven framework, and the data-driven methods act to add additional detail and remove biases created by incomplete or incorrect process descriptions in the physical-based models. In conclusion, hybrid modeling approaches can reproduce historical daily streamflow series with acceptable accuracy at all calibration sites, which is very important for effective regional water resource assessment and management in a changing environment.

5.2 Recommendations

In developing countries, the challenge of having adequate and high-quality data may seem like a distant dream. It was a difficult task to collect data for this study, especially from home institutions, and when it was made available, the quality was always questionable. A thorough quality control process is more likely to provide the user with half the data required. As a result, other types/sources of data, such as remotely sensed data (for rainfall, for example), could be investigated to provide broader coverage of rainfall distribution and supplement the existing records. Furthermore, models with few input data must be investigated and adapted for climate-specific regions such as Ethiopia's highlands.

ANN models were integrated with a physical-based SWAT model and both models have been successfully applied in rainfall-runoff modeling across many watersheds found in the world. However, other deep learning modeling approaches, such as LSTM, Bayesian-based ANN, and others, should be tested in order to further improve the hybrid model. The study, on the other hand, only used a physical-based SWAT model; however, other physical-based and less complicated conceptual models should be tested as well. Furthermore, even though multi-site calibration was used in this study, the climate and watershed characteristics of all three calibration sites are similar. Thus, it is highly recommended that in the future, the applicability of hybrid models be tested in various watersheds with a wide range of climate zones.

REFERENCES

- Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S. R. I. N. I. V. A. S. A. N. B., Srinivasan, R., Yang, H., and Kløve, B. (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *Journal of hydrology*, 524, 733-752.
- Abbaspour, K. C., Vaghefi, S. A., and Srinivasan, R. (2017). A guideline for successful calibration and uncertainty analysis for soil and water assessment: a review of papers from the 2016 international SWAT conference. *Water*, 10(1), 6.
- Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., ... and Srinivasan, R. (2007). Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. *Journal of hydrology*, 333(2-4), 413-430.
- Abdella Kemal (2013). The Effect of Climate Change on Water Resources Potential of Omo Gibe Basin, Ethiopia. Dissertation Paper, Munchen, Germany.
- Abebe, A. J., and Price, R. K. (2003). Managing uncertainty in hydrological models using complementary models. *Hydrological sciences journal*, 48(5), 679-692.
- Adane, G.B.; Hirpa, B.A.; Gebru, B.M.; Song, C.; Lee, W.-K.(2021). Integrating Satellite Rainfall Estimates with Hydrological Water Balance Model: Rainfall-Runoff Modeling in Awash River Basin, Ethiopia. *Water* 2021, 13, 800. <https://doi.org/10.3390/w13060800>
- Adombi, A. V. D. P., Chesnaux, R., and Boucher, M. A. (2021). Theory-guided machine learning applied to hydrogeology—state of the art, opportunities and future challenges. *Hydrogeology Journal*, 29(8), 2671-2683.
- Akoko, G., Le, T. H., Gomi, T., and Kato, T. (2021). A Review of SWAT Model Application in Africa. *Water*, 13(9), 1313. <https://doi.org/10.3390/w13091313>
- Aliye, M. A., Aga, A. O., Tadesse, T., and Yohannes, P. (2020). Evaluating the Performance of HEC-HMS and SWAT Hydrological Models in Simulating the Rainfall-Runoff Process for Data Scarce Region of Ethiopian Rift Valley Lake Basin. *Open Journal of Modern Hydrology*, 10(04), 105. <https://doi.org/10.4236/ojmh.2020.104007>
- Aral, M.M. and Gunduz, O. 2006. Large-scale hybrid watershed modeling. In *Watershed Models*, V.P. Singh and D.K. Frevert, eds., CRC/Taylor and Francis Group, Boca Raton, FL, pp.
- Ardabili, S., Mosavi, A., Dehghani, M., and Várkonyi-Kóczy, A. R. (2020). Deep Learning and Machine Learning in Hydrological Processes Climate Change and Earth Systems a Systematic Review. In A. R. Várkonyi-Kóczy (Ed.), *Engineering for Sustainable Future* (pp. 52–62). Springer International Publishing. <https://doi.org/10.1007/978-3-030-36841>

- Arnold, J. G. and Allen, P. M. (1996). Estimating hydrologic budgets for three Illinois watersheds. *J. Hydrol.* 176, 57–77.
- Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... and Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1491-1508.
- Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., ... and Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, 55(4), 1491-1508.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998). Large area hydrologic modeling and assessment part I: model development 1. *JAWRA Journal of the American Water Resources Association*, 34(1), 73-89.
- Basin, R., Patel, A. B., and Joshi, G. S. (2017). Modeling of Rainfall-Runoff Correlations Using Artificial Neural Network-A Case Study of Dharoi Watershed of a Sabarmati, 3(2), 78–87.
- Behmanesh, J. and Ayashm, S. (2015). Rainfall-Runoff Modelling In the Turkey River Using Numerical and Regression Methods. *Journal of Fundamental and Applied Science*, 91-102.
- Belayneh, A., Adamowski, J., and Khalil, B. (2016). Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods. *Sustainable Water Resources Management*, 2(1), 87-101.
- Beyene S.M.(2018) Runoff and Sediment Yield Modeling Using Artificial Neural Network: Case of Muger Watershed, Abbay River Basin, Ethiopia
- Brown, S., Versace, V. L., Laurenson, L., Ierodiaconou, D., Fawcett, J., and Salzman, S. (2012). Assessment of spatiotemporal varying relationships between rainfall, land cover and surface water area using geographically weighted regression. *Environmental Modeling and Assessment*, 17(3), 241-254.
- Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, 58(1-2), 11-27.
- Butcher, J. (2008). Watershed Modeling. <http://www.epa.gov/owow/watershed/wacademy/acad200>. Retrieved August 12, 2014,
- Casari, A., and Zheng, A. (2018). Feature Engineering for Machine Learning: *principles and techniques for data scientists*. “O'Reilly Media, Inc”.
- Cho, K., and Kim, Y. (2022). Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *Journal of Hydrology*, 605, 127297.
- Choong, C. E., Ibrahim, S., and El-Shafie, A. (2020). Artificial Neural Network (ANN) model development for predicting just suspension speed in solid-liquid mixing system. *Flow Measurement and Instrumentation*, 71, 101689.

- Corzo, G. A., Solomatine, D. P., De Wit, M., Werner, M., Uhlenbrook, S., and Price, R. K. (2009). Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin. *Hydrology and Earth System Sciences*, 13(9).
- Dalkılıç, H. Y., and Hashimi, S. A. (2020). Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models. *Water Supply*, 20(4), 1396–1408.
- Daniel, E. B., Camp, J. V., Leboeuf, E. J., Penrod, J. R., Dobbins, J. P., and Abkowitz, M. D. (2011). 56 Watershed Modeling and its Applications : A State-of-the-Art Review, 26–50.
- Darbandi, S., and Pourhosseini, F. A. (2018). River flow simulation using a multilayer perceptron-firefly algorithm model. *Applied Water Science*, 8(3), 1-9.
- Dawson, C. W., and Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1), 80–108.
- Devia, G., Ganasri, B. P., and Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4, 1001–1007. <https://doi.org/10.1016/j.aqpro.2015.02.126>
- Dilnsew A.chekol, Bermand Tishabein, Helmut Eggers and Paul Vlek (2007). Application of SWAT for assessment of spatial distribution of water resources and analyzing impact of 72 different land management practices on soil erosion in Upper Awash River Basin watershed.
- Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
- El Harraki, W., Ouazar, D., Bouziane, A., El Harraki, I., and Hasnaoui, D. (2021). Streamflow Prediction Upstream of a Dam Using SWAT and Assessment of the Impact of Land Use Spatial Resolution on Model Performance. *Environmental Processes*, 8(3), 1165–1186. <https://doi.org/10.1007/s40710-021-00532-0>
- Fanta, S. S., and Sime, C. H. (2021). Performance assessment of SWAT and HEC-HMS model for runoff simulation of Toba watershed, Ethiopia. *Sustainable Water Resources Management*, 8(1), 8. <https://doi.org/10.1007/s40899-021-00596-8>
- Farfán, J. F., Palacios, K., Ulloa, J., and Avilés, A. (2020). A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds. *Journal of Hydrology: Regional Studies*, 27, 100652.
- Farmer, W.H., Vogel, R.M., 2016. On the deterministic and stochastic use of hydrologic models. *Water Resour. Res.* 52, 5619–5633. <https://doi.org/10.1002/2016WR019129>.
- Food and Agriculture Organization (FAO). (2020). Water Accounting in The Awash River Basin.

- Frame, J., Nearing, G., Kratzert, F., and Rahman, M. (2020). Post processing the US National Water Model with a Long Short-Term Memory network. Fufa, T. (2016). Addis Ababa Institute of Technology School of Graduate Studies (Doctoral dissertation, Addis Ababa University Ethiopia).
- Gauch, M., Mai, J., Gharari, S., & Lin, J. (2019). Data-driven vs. physically-based streamflow prediction models. In *Proceedings of 9th International Workshop on Climate Informatics*.
- Gessese, B., Bewket, W., and Brauning, A. (2014). Model Based Characterization and Monitoring of Runoff and soil Erosion in Response to Land use/land cover Changes in the Modjo watershed, Ethiopia.
- Ghaith, M., Siam, A., Li, Z., and El-Dakhakhni, W. (2020). Hybrid hydrological data-driven approach for daily streamflow forecasting. *Journal of Hydrologic Engineering*, 25(2), 04019063.
- Gharde, K.D., Mahesh, K., Mittal, H.K., Singh, P.K. and Dahiphale, P.A. (2016). Sediment Yield Modelling of Kal River in Maharashtra Using Artificial Neural Network Model. *Journal of Recent Sciences*, 120-130.
- Gleason, C. J., Wada, Y., and Wang, J. (2018). A hybrid of optical remote sensing and hydrological modeling improves water balance estimation. *Journal of Advances in Modeling Earth Systems*, 10(1), 2-17.
- Gonfa, Z., and Kumar, D. (2018). Calibration and validation of SWAT model using streamflow and sediment load for Mojo watershed, Ethiopia. *Sustainable Water Resources Management*, 4, 1–13. <https://doi.org/10.1007/s40899-017-0189-1>
- Gunathilake, M. B., Karunanayake, C., Gunathilake, A. S., Marasingha, N., Samarasinghe, J. T., Bandara, I. M., and Rathnayake, U. (2021). Hydrological Models and Artificial Neural Networks (ANNs) to Simulate Streamflow in a Tropical Catchment of Sri Lanka. *Applied Computational Intelligence and Soft Computing*, 2021, e6683389. <https://doi.org/10.1155/2021/6683389>
- Haileysus Belay (2011). Evaluation of Climate Change impacts on hydrology on selected catchments of Abbay Basin, Msc Thesis, Addis Ababa University, Ethiopia
- Hargreaves, G.H. and Z.A. Samani. 1985. Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture* 1:96-99.
- Herath, H. M. V. V., Chadawalawada, J., and Babovic, V. (2021). Hydrologically informed machine learning for rainfall–runoff modelling: towards distributed modelling. *Hydrology and Earth System Sciences*, 25(8), 4373-4401.
- Hu, C., Wu, Q., Li, H., Jian, S., Li, N., and Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*, 10(11), 1543.

- Humphrey, G. B., Gibbs, M. S., Dandy, G. C., and Maier, H. R. (2016). A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. *Journal of Hydrology*, 540, 623–640. <https://doi.org/10.1016/j.jhydrol.2016.06.026>
- Iruela, J. R. S., Ruiz, L. G. B., Capel, M. I., and Pegalajar, M. C. (2021). A tensorflow approach to data analysis for time series forecasting in the energy-efficiency realm. *Energies*, 14(13), 4038.
- Jaiswal, R. K., Ali, S., and Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, 10(1), 48. <https://doi.org/10.1007/s13201-019-1122-6>
- Jajarmizad, M., Harun, S., and Salarpour, M. (2012). A Review on Theoretical Consideration and Types of Models in Hydrology. *Journal of Environmental Science and Technology*, 5(5), 249–261. <https://doi.org/10.3923/jest.2012.249.261>
- Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., and Pulido-Velazquez, D. (2018). A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. Undefined. <https://www.semanticscholar.org/paper/A-Comparison-of-SWAT-and-ANN-Models-for->
- Kassem, A. A., Raheem, A. M., Khidir, K. M., and Alkattan, M. (2020). Predicting of daily Khazir basin flow using SWAT and hybrid SWAT-ANN models. *Ain Shams Engineering Journal*, 11(2), 435–443. <https://doi.org/10.1016/j.asej.2019.10.011>
- Khalid, K., Ali, M. F., Rahman, N. F. A., Mispan, M. R., Haron, S. H., Othman, Z., and Bachok, M. F. (2016). Sensitivity Analysis in Watershed Model Using SUFI-2 Algorithm. *Procedia Engineering*, 162, 441–447. <https://doi.org/10.1016/j.proeng.2016.11.086>
- Konapala, Goutam, et al. "Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US." *Environmental Research Letters* 15.10 (2020): 104022.
- Koycegiz, C., and Buyukyildiz, M. (2019). Calibration of SWAT and Two Data-Driven Models for a Data-Scarce Mountainous Headwater in Semi-Arid Konya Closed Basin. *Water*, 11(1). <https://doi.org/10.3390/w11010147>
- Kumanlioglu, A. A., and Fistikoglu, O. (2019). Performance enhancement of a conceptual hydrological model by integrating artificial intelligence. *Journal of Hydrologic Engineering*, 24(11), 04019047.
- Kwak, J., Han, H., Kim, S., and Kim, H. S. (2022). Is the deep-learning technique a completely alternative for the hydrological model?: A case study on Hyeongsan River Basin, Korea. *Stochastic Environmental Research and Risk Assessment*, 36(6), 1615-1629.

- Lange, H., and Sippel, S. (2020). Machine Learning Applications in Hydrology. In D. F. Levia, D. E. Carlyle-Moses, S. Iida, B. Michalzik, K. Nanko, and A. Tischer (Eds.), *Forest-Water Interactions* (pp. 233–257). Springer International Publishing. https://doi.org/10.1007/978-3-030-26086-6_10
- Latifoğlu, L., Kişi, Ö., and Latifoğlu, F. (2015). Importance of hybrid models for forecasting of hydrological variable. *Neural Computing and Applications*, 26(7), 1669–1680. <https://doi.org/10.1007/s00521-015-1831-1>
- Lee, H., and Moon, Y.-I. (2007). Analysis and development of conceptual rainfall-runoff model structures for regionalisation purposes. *KSCE Journal of Civil Engineering*, 11(1), 57–64. <https://doi.org/10.1007/BF02823373>
- Lu, D., Konapala, G., Painter, S. L., Kao, S. C., and Gangrade, S. (2021). Streamflow simulation in data-scarce basins using bayesian and physics-informed machine learning models. *Journal of Hydrometeorology*, 22(6), 1421-1438.
- Lv, Z., Zuo, J., and Rodriguez, D. (2020). Predicting of Runoff Using an Optimized SWAT-ANN: A Case Study. *Journal of Hydrology: Regional Studies*, 29, 100688. <https://doi.org/10.1016/j.ejrh.2020.100688>
- Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. P. (2010). Environmental Modelling and Software Methods used for the development of neural networks for the prediction of water resource variables in river systems : Current status and future directions. *Environmental Modelling and Software*, 25(8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Malik, M. A., Dar, A. Q., and Jain, M. K. (2022). Modelling streamflow using the SWAT model and multi-site calibration utilizing SUFI-2 of SWAT-CUP model for high altitude catchments, NW Himalaya's. *Modeling Earth Systems and Environment*, 8(1), 1203-1213.
- Mekonnen, B. A., Nazemi, A., Mazurek, K. A., Elshorbagy, A., and Putz, G. (2015). Hybrid modelling approach to prairie hydrology: Fusing data-driven and process-based hydrological models. *Hydrological Sciences Journal*, 60(9), 1473–1489. <https://doi.org/10.1080/02626667.2014.935778>
- Mengistu, A. (2008). Climate variability and change. *Ethiopian journal of Animal production*, 8(1), 94-98.
- Mengistu, D. T., Moges, S. A., and Sorteberg, A. (2016). Revisiting Systems Type Black-Box Rainfall-Runoff Models for Flow Forecasting Application. *Journal of Water Resource and Protection*, 08(01), 65–83.
- Meshesha, Y. B., and Abdi, M. B. (2019). Challenges and opportunities for implementation of Integrated Water Resource Management in Omo-Gibe Basin, Ethiopia. *Journal of Ecology and The Natural Environment*, 11(7), 84-97.

- Monteith, J. L. (1965). "Evaporation and Environment." Symposia of the Society for Experimental Biology 19: 205–34.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900.
- Moriasi, D. N., Gitau, M. W., Pai, N., and Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763-1785.
- Nearing, G., Sampson, A. K., Kratzert, F., and Frame, J. (2020). Post-processing a Conceptual Rainfall-runoff Model with an LSTM
- Neitsch, S., Arnold, J. ., Kiniry, J. and Williams, J. (2011). Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute, 1–647. <https://doi.org/10.1016/j.scitotenv.2015.11.063>.
- Nielsen, R. F., Nazemzadeh, N., Andersson, M. P., Gernaey, K. V., and Mansouri, S. S. (2021). An uncertainty-aware hybrid modelling approach using probabilistic machine learning. In *Computer Aided Chemical Engineering* (Vol. 50, pp. 591-597). Elsevier.
- Noori, N., and Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, 533, 141–151. <https://doi.org/10.1016/j.jhydrol.2015.11.050>
- Özger, M. (2009). Comparison of fuzzy inference systems for streamflow prediction. *Hydrological Sciences Journal*, 54(2), 261-273.
- Parrott, L. (2011). Hybrid modelling of complex ecological systems for decision support: Recent successes and future perspectives. *Ecological Informatics*, 6(1), 44-49.
- Price, R. K., and Solomatine, D. P. (2009.). *A Brief Guide to Hydroinformatics*. 28.
- Priestley, C. H. B., and R. J. Taylor. (1972). "On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters." *Monthly Weather Review* 100 (2): 81–92. [https://doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Quilty, J. M., Sikorska-Senoner, A. E., and Hah, D. (2022). A stochastic conceptual-data-driven approach for improved hydrological simulations. *Environmental Modelling & Software*, 149, 105326.
- Quilty, J., and Adamowski, J. (2021). A maximal overlap discrete wavelet packet transform integrated approach for rainfall forecasting—A case study in the Awash River Basin (Ethiopia). *Environmental Modelling and Software*, 144, 105119. *Water*, 8(5), e1533.

- Raihan, F., Beaumont, L. J., Maina, J., Saiful Islam, A., and Harrison, S. P. (2020). Simulating streamflow in the Upper Halda Basin of southeastern Bangladesh using SWAT model. *Hydrological Sciences Journal*, 65(1), 138–151.
- Raschka, S., Mirjalili, V., 2018. Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow, Second edition, fourth release, Expert insight. Packt Publishing, Birmingham Mumbai.
- Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., ... and Kumar, V. (2019). Process-guided deep learning predictions of lake water temperature. *Water Resources Research*, 55(11), 9173-9190.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5).
- Rezaei, M., Ali, A., Motlaq, A., Mahmoudi, A. R., and Mousavi, H. (2015). River Flow Forecasting using artificial neural network (Shoor Ghaen), 207–215.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The annals of mathematical statistics*, 400-407.
- Robertson, D. E., Pokhrel, P., and Wang, Q. J. (2013). Improving statistical forecasts of seasonal streamflows using hydrological model output. *Hydrology and Earth System Sciences*, 17(2), 579–593. <https://doi.org/10.5194/hess-17-579-2013>
- Saltelli, A., Tarantola, S., & Campolongo, F. (2000). Sensitivity analysis as an ingredient of modeling. *Statistical science*, 377-395.
- Serur, A. B., and Sarma, A. (2016). Evaluation of the ArcSWAT Model in Simulating Catchment Hydrology: In Weyib River Basin, Bale Mountainous Area of Southeastern Ethiopia. *Innovative and Emerging Research in Engineering*, 3, 3–11.
- Shanmuganathan, S. and Samarasinghe, S. (2016). Artificial neural network modelling. Springer, (628).
- Sharma, P. J., Patel, P. L., & Jothiprakash, V. (2021). Data-driven modelling framework for streamflow prediction in a physio-climatically heterogeneous river basin. *Soft Computing*, 25(8), 5951-5978.
- Sharma, N., Zakaullah, Md., Tiwari, H., and Kumar, D. (2015). Runoff and sediment yield modeling using ANN and support vector machines: A case study from Nepal watershed. *Modeling Earth Systems and Environment*, 1(3), 23. <https://doi.org/10.1007/s40808-015-0027-0>

- Shawul, A. A., Chakma, S., and Melesse, A. M. (2019). The response of water balance components to land cover change based on hydrologic modeling and partial least squares regression (PLSR) analysis in the Upper Awash Basin. *Journal of Hydrology: Regional Studies*, 26, 100640.
- Shirke, Y., and Kawitkar, P. R. (2012). Artificial Neural Network based Runoff Prediction Model for a Reservoir, 1(3), 1–5.
- Shrestha, D. L., and Solomatine, D. P. (2009). Assessing uncertainty in rainfall-runoff models: Application of data-driven models. In *Flood Risk Management: Research and Practice*, 1563-1573
- Sikorska-Senoner, A. E., and Quilty, J. M. (2021). A novel ensemble-based conceptual-data-driven approach for improved streamflow simulations. *Environmental Modelling and Software*, 143, 105094.
- Sileshi Gobena. (2010). “Daily Rainfall-Runoff Modelling of Upper Awash Sub-basin Using Conceptual Rainfall Runoff Models." Dissertation. Addis Ababa University, Addis Ababa, Ethiopia.
- Singh, A., Imtiyaz, M., Isaac, R., and Denis, D. (2011). Comparison of soil and water assessment tool (SWAT) and multilayer perceptron (MLP) artificial neural network for predicting sediment yield in the Nagwa agricultural watershed in Jharkhand, India. *Agricultural Water*
- Singh, H. V., Thompson, A. M., and Gharabaghi, B. (2017). Event Runoff and Sediment-Yield Neural Network Models for Assessment and Design of Management Practices for Small Agricultural Watersheds, 22(2), 1–12. [https://doi.org/10.1061/\(ASCE\)HE](https://doi.org/10.1061/(ASCE)HE).
- Singh, V. P. (2018). Hydrologic modeling: Progress and future directions. *Geoscience Letters*, 5(1), 15. <https://doi.org/10.1186/s40562-018-0113-z>
- Solomatine, D. P., Maskey, M., and Shrestha, D. L. (2008). Instance-based learning compared to other data-driven methods in hydrological forecasting. *Hydrological Processes: An International Journal*, 22(2), 275-287.
- Solomatine, D., and Ostfeld, A. (2008). Data-Driven Modelling: Some Past Experiences and New Approaches. *Journal of Hydroinformatics - J HYDROINFORM*, 10.
- Sorooshian, S., Hsu, K., Coppola, E., Tomassetti, B., Verdecchia, M., and Visconti, G. (2008). *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models*. Springer Science and Business Media.
- Sudheer, K., Gosain, A., and Ramasastri, K. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol. Process*, (16):1325–1330

- Szandafa, T. (2021). Review and comparison of commonly used activation functions for deep neural networks. In *Bio-inspired neurocomputing* (pp. 203-224). Springer, Singapore.
- Tamiru, H., and Dinka, M. O. (2021). Application of ANN and HEC-RAS model for flood inundation mapping in lower Baro Akobo River Basin, Ethiopia. *Journal of Hydrology: Regional Studies*, 36, 100855. <https://doi.org/10.1016/j.ejrh.2021.100855>
- Tegegne, G., Park, D. K., and Kim, Y. O. (2017). Comparison of hydrological models for the assessment of water resources in a data-scarce region, the Upper Blue Nile River Basin. *Journal of Hydrology: Regional Studies*, 14, 49-66.
- Tessema, N., Kebede, A., and Yadeta, D. (2021). Modelling the effects of climate change on streamflow using climate and hydrological models: The case of the Kesem sub-basin of the Awash River basin, Ethiopia. *International Journal of River Basin Management*, 19(4), 469–480. <https://doi.org/10.1080/15715124.2020.1755301>
- Tessema, S. M. (2011). Hydrological modeling as a tool for sustainable water resources management: a case study of the Awash River Basin (Doctoral dissertation, KTH Royal Institute of Technology).
- Tessema, S. M., Setegn, S. G., and Mörtberg, U. (2015). Watershed modeling as a tool for sustainable water resources management: SWAT model application in the Awash River basin, Ethiopia. In *Sustainability of Integrated Water Resources Management* (pp. 579-606). Springer, Cham.
- Tufa, F. G., and Sime, C. H. (2020). Streamflow modeling using SWAT model and the model performance evaluation in Toba sub-watershed, Ethiopia. *Modeling Earth Systems and Environment*. <https://doi.org/10.1007/s40808-020-01039-7>
- USDA-SCS. (1972). National Engineering Handbook, Section 4: Hydrology. U.S. Government Printing Office, Washington, DC. von Stosch, M., Davy, S., Francois, K., Galvanauskas, V., Hamelink, J. M., Luebbert, A., and Glassey, J. (2014). Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry. *Biotechnology journal*, 9(6), 719-726.
- Wanas, N., Auda, G., Kamel, M. S., and Karray, F. A. K. F. (1998, May). On the optimal number of hidden nodes in a neural network. In *Conference proceedings. IEEE Canadian conference on electrical and computer engineering (Cat. No. 98TH8341)*
- Wang, J., Bao, W., Gao, Q., Si, W., Sun, Y., 2021. Coupling the Xinanjiang model and wavelet-based random forests method for improved daily streamflow simulation. *J. Hydroinf.*
- Waseem, M., Mani, N., Andiego, G., and Usman, M. (2017). A review of criteria of fit for hydrological models. *International Research Journal of Engineering and Technology (IRJET)*, 4(11), 1765-1772.

- Woldesenbet, T. A., Elagib, N. A., Ribbe, L., and Heinrich, J. (2017). Hydrological responses to land use/cover changes in the source region of the Upper Blue Nile Basin, Ethiopia. *Science of the Total Environment*, 575, 724-741.
- Wu, L., Kaiser, G., Solomon, D., Winter, R., Boulanger, A., and Anderson, R. (2012, May). Improving efficiency and reliability of building systems using machine learning and automated online evaluation. In *2012 IEEE Long Island Systems, Applications and Technology Conference (LISAT)* (pp. 1-6). IEEE.
- Wu, R., Yang, L., Chen, C., Ahmad, S., Dascalu, S. M., and Harris Jr, F. C. (2019). Modeling error learning based post-processor framework for hydrologic models accuracy improvement. *Geoscientific Model Development*.
- Xu, T., and Liang, F. (2021). Machine learning for hydrologic sciences: An introductory overview. *WIREs Water*, 8(5), e1533. <https://doi.org/10.1002/wat2.1533>
- Yang, J., Reichert, P., Abbaspour, K.C., Xia, J. and Yang, H (2008). Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China, *J. Hydrol.*, doi: 10.1016/j.jhydrol.
- Yenigün, K., Bilgehan, M., Gerger, R., and Mutlu, M. (2010). A comparative study on prediction of sediment yield in the Euphrates basin, 5(May), 518–534.
- Zoller M-A, Huber MF. (2019) Benchmark and survey of automated machine learning frameworks, arXiv preprint arXiv: 1904.12054.

APPENDICES

Appendix A

Appendix A.1. Description of meteorological stations

ID	Station Name	Lat	Long	Elevation	PCP	Tmax	Tmin	RH	SR	WS	Rec. Year
1	A.A Obs	9.02	38.75	2386	yes	yes	yes	yes	yes	yes	2000-2015
2	A.A Bole	8.98	38.80	2354	yes	yes	yes	no	no	no	2000-2015
3	Hombole	8.37	38.77	1743	yes	yes	yes	no	no	no	2000-2015
4	Asgori	8.80	38.34	2072	yes	yes	yes	no	no	no	2000-2015
5	Tulubolo	8.66	38.21	2190	yes	no	no	no	no	no	2000-2015
6	Akaki	8.87	38.79	2057	yes	yes	yes	no	no	no	2000-2015
7	Intoto	9.08	38.72	2903	yes	yes	yes	no	no	no	2000-2015

Where PCP is Precipitation, Tmin is Minimum Temperature, Tmax is Maximum Temperature, RH is Relative Humidity, SR is Solar Radiation, and WS is Wind Speed.

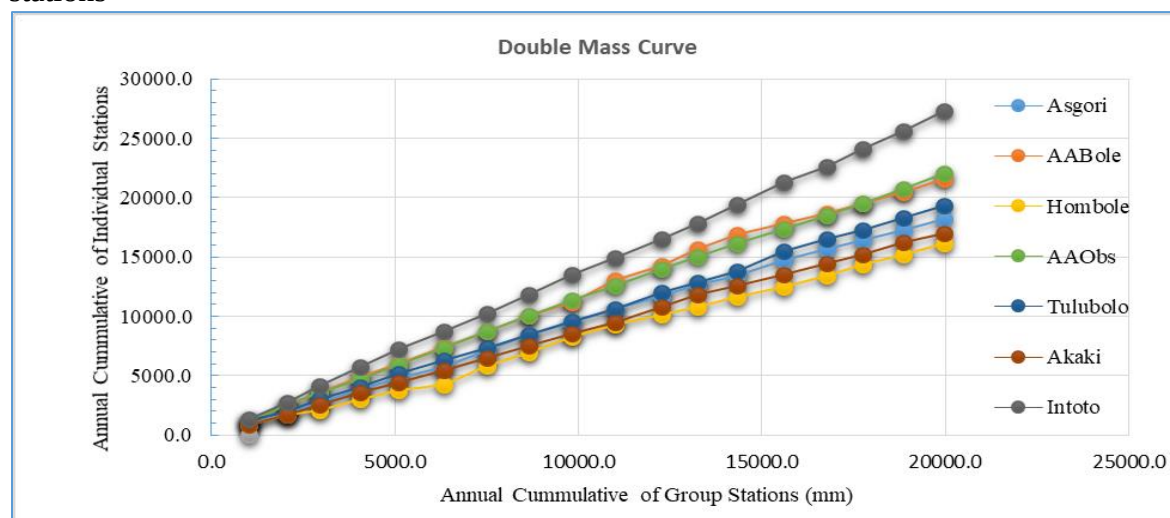
Appendix A.2. Percentage of missing data for each meteorological station

Station Name	Percent of missing for rainfall data	Percent of missing for temperature
Asgori	0.5	5.1
Addis Ababa Bole	3.2	1.2
Hombole	4.6	–
Addis Ababa Obs	0.1	7.2
Tulubolo	1.3	1.5
Akaki	0.9	0.3
Intoto	2.3	0.1
Adis Alem	6.7	3.2

Appendix A.3. Major Soil groups in the study area

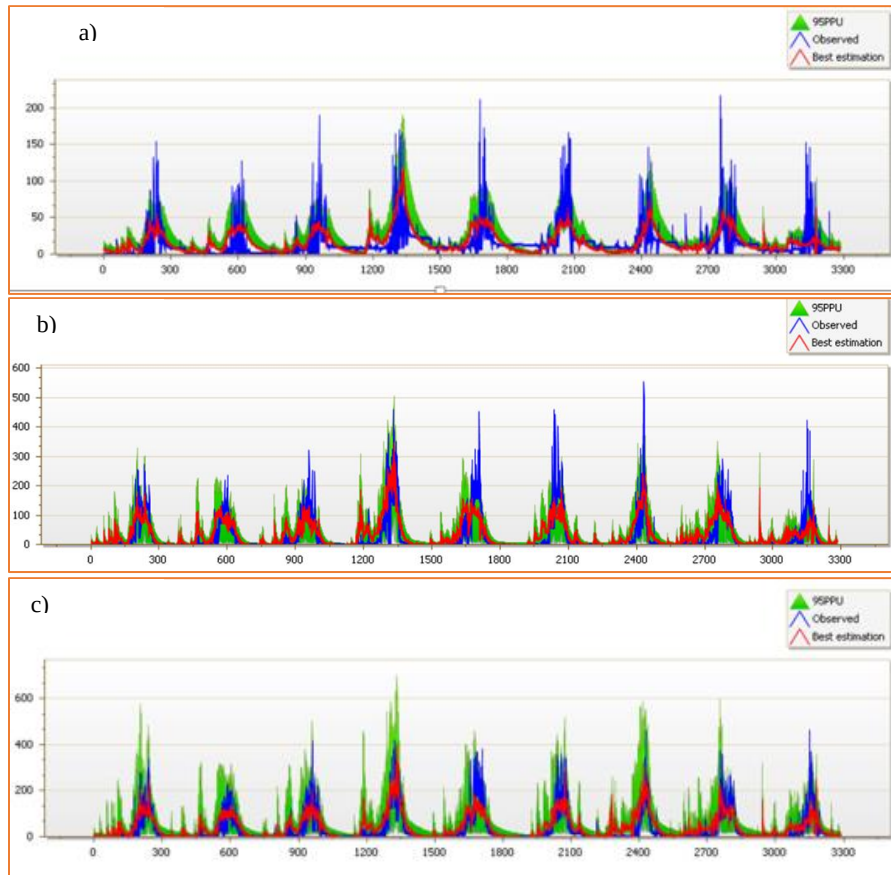
Major Soil Type	Soil code	Area SqKm	Areal coverage (%)	Soil Texture
Lithic Leptosols	LPq	469	6.3	Clay loam
Haplic Luvisols	LVh	19	0.3	Sandy clay loam
Chromic Luvisols	LVx	859	11.5	Clay
Humic Nitisols	Ntu	1056	14.1	Clay loam
Luvic Phaeozems	PHI	7	0.1	Clay
Euritic Vertisols	Vre	4417	59.0	Loam
Eutritic Combisols	Cme	142	1.9	Clay
Vertic Combisols	CMv	504	6.7	Clay
Water	WR	11	0.1	Water

Appendix A.4. Double mass curve representative of the Upper Awash Sub-basin meteorological stations

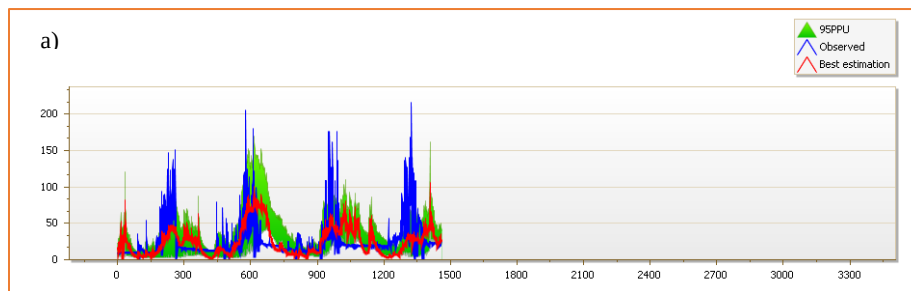


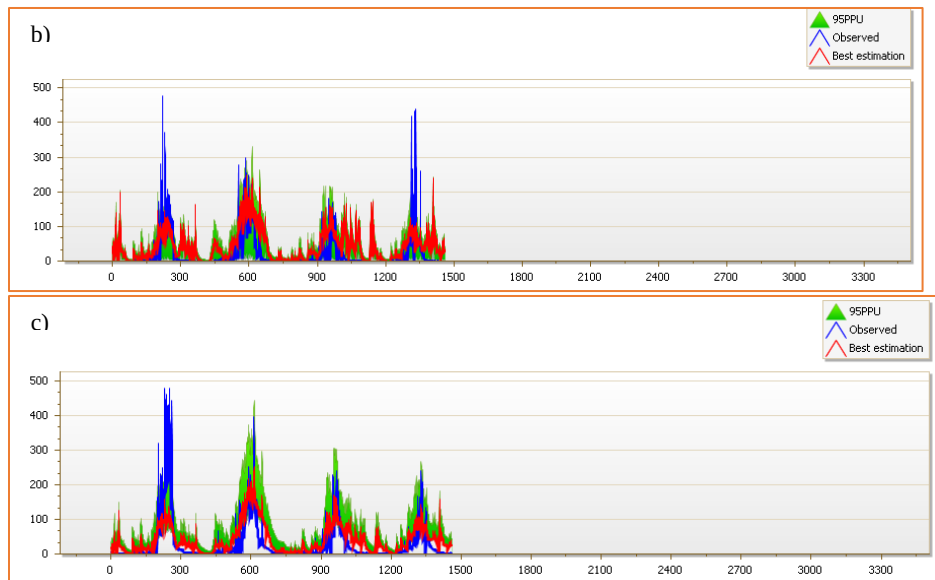
Appendix B

Appendix B.1. Streamflow calibration 95PPU plot: a) Akaki, b) Melka Kunture, c) Hombole



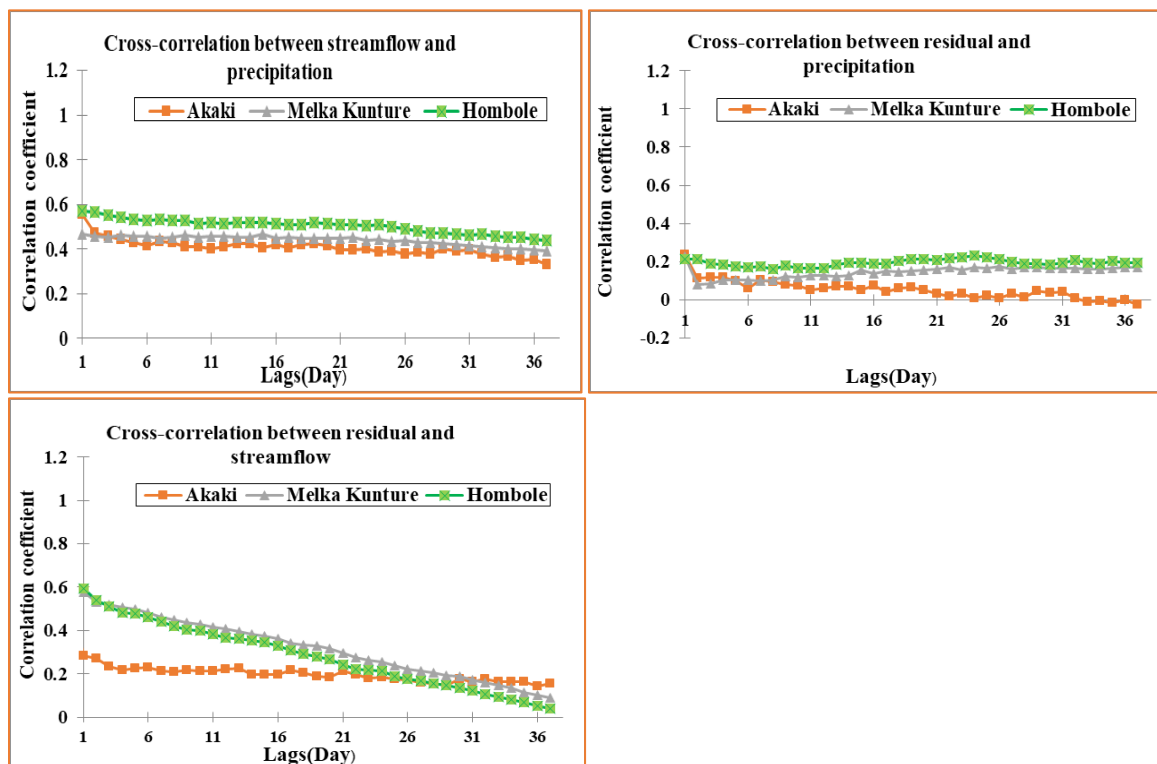
Appendix B.2. Streamflow validation 95PPU plot: a) Akaki, b) Melka Kunture, c) Hombole





Appendix C:

Appendix C 1. Cross-correlation graphs between residual of SWAT with observed streamflow and areal precipitation



Appendix D: Python code to develop data-driven ANN model

The code below was created using Keras, a Python-based deep learning API that runs on top of the machine learning platform TensorFlow. (*KERAS, Keras: The Python Deep Learning Library*, <https://keras.io/>).

The executable codes are given in Courier New font, and they are reusable with a few modifications for the other model inputs and structure, but for demo purposes, the following codes are used to estimate the residual error of the SWAT model by using the preselected input features at Hombole station.

Install and import Necessary TensorFlow and dependencies modules

```
import keras
import numpy as np
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers import Dense
from keras import layers
import matplotlib.pyplot as plt
from keras.callbacks import EarlyStopping
import tensorflow as tf
import seaborn as sns
np.set_printoptions(precision=3, suppress=True)
from tensorflow.keras.callbacks import TensorBoard
import time
#Load the TensorBoard notebook extension.
%load_ext tensorboard
from tensorboard.plugins.hparams import api as hp
```

Get the data

The preprocessing of data, like cleaning and organizing, was completed, and the preprocessed data was imported as a csv file.

```
dataset =pd.read_csv('Data.csv') # 'Data.csv'is the file name of
the processed data.
dataset.tail()
```

Divide the data into training and testing groups.

The data is split into training and test data sets with a percent of 69.25% and 30.75%, respectively.

```
train_dataset, test_dataset = train_test_split( dataset, test_size
=0.3075, shuffle=False)
print (train_dataset)
dataset.describe ()
#Inspect the data (Review the joint distribution of a pairs of
columns from the training set.)

sns.pairplot(train_dataset[['Pt', 'Pt-1', 'Pt-2', 'Qt-1', 'Qt-2', 'Qt-
3', 'Tmax', 'Tmaxt-20', 'Tmin', 'Residual']], diag_kind='kde')
```

Split features from labels

Separate the target value—the "label"—from the features. This label represents the value that will be used to train the model to predict.

```
train_features = train_dataset.copy()
test_features = test_dataset.copy()
train_labels = train_features.pop('Residual')
test_labels = test_features.pop('Residual')
```

Data normalization/standardization

```
normalizer = tf.keras.layers.Normalization(axis=-1)
normalizer = layers.Normalization(input_shape=[1,], axis=None)
normalizer.adapt(np.array(train_features))
print(normalizer.mean.numpy())
first = np.array(train_features[:1])
with np.printoptions(precision=2, suppress=True):
    print('First example:', first)
    print()
    print('Normalized:', normalizer(first).numpy())
```

Hyperparameter optimization

The value of three hyperparameters, including the number of nodes in the first hidden layer, optimizers, and learning rate, were optimized using the grid search strategy.

```
#Make a list of how many nodes to tune between log(m) and 2 N+1.
N=len(train_features.columns)
m=len(train_features.axes[0])
import math
i= int(math.log10(m))
```

```

j= int(2*N+1)

def createList(r1, r2):
    return [item for item in range(r1, r2+1)]

# Driver Code
r1, r2 = i, j
neurons=(createList(r1, r2))
#The hyperparameters and their values to be tested are stored in
a special type called HParam
HP_NUM_UNITS = hp.HParam('num_units', hp.Discrete(neurons))
HP_LEARNING_RATE= hp.HParam('learning_rate', hp.Discrete([0.1, 0
.01, 0.001]))
HP_OPTIMIZER = hp.HParam('optimizer', hp.Discrete(['adam', 'sgd'
]))
#Setting the Metric to RMSE
METRIC_RMSE = 'RootMeanSquaredError'
#Clear any logs from previous runs
!rm -rf ./logs/
#Creating and configuring log files
with tf.summary.create_file_writer('logs/hparam_tuning').as_defa
ult():
    hp.hparams_config(
        hparams=[HP_NUM_UNITS,HP_LEARNING_RATE,HP_OPTIMIZER],
        metrics=[hp.Metric(METRIC_RMSE, display_name='RMSE')]
    )
#afunction that trains and validates the model and returns the r
mse
def train_test_model(hparams):
    #Keras sequential model with Hyperparameters passed from the a
rgument
    model = tf.keras.models.Sequential([
        tf.keras.layers.InputLayer(input_shape = N),

        tf.keras.layers.Dense(hparams[HP_NUM_UNITS], activation=tf.n
n.relu, kernel_initializer = 'uniform'),
        tf.keras.layers.Dense(1),
    ])

    #Compiling the model
    model.compile(
        optimizer=hparams[HP_OPTIMIZER],
        loss='mean_absolute_error',
        metrics=['RootMeanSquaredError'],
    )

```

```

    model.fit(train_features, train_labels, epochs=20) # Run with
20 epoch
    _, accuracy = model.evaluate(test_features, test_labels)
    return accuracy
#A function to log the training process
def run(run_dir, hparams):
    with tf.summary.create_file_writer(run_dir).as_default():
        hp.hparams(hparams)
        rmse = train_test_model(hparams)
        tf.summary.scalar(METRIC_RMSE, rmse, step=2)

#Nested for loop training with all possible combinathon of hyper
parameters
for num_units in HP_NUM_UNITS.domain.values:
    for learning_rate in HP_LEARNING_RATE.domain.values:
        for optimizer in HP_OPTIMIZER.domain.values:
            hparams = {
                HP_NUM_UNITS: num_units,
                HP_LEARNING_RATE: learning_rate,
                HP_OPTIMIZER: optimizer,
            }
            run_name = "run-%d" % session_num
            print('--- Starting trial: %s' % run_name)
            print({h.name: hparams[h] for h in hparams})
            run('logs/hparam_tuning/' + run_name, hparams)
            session_num += 1

#View these logs in TensorBoard:
tensorboard--logdir logs/hparam_tuning

```

Define and build the Keras Sequential Model using optimized parameters.

```

def build_and_compile_model(norm):
    model = keras.Sequential([
        norm,
        layers.Dense(12, activation='relu'),
        layers.Dense(1)
    ])
    model.compile(loss='MSE',
                  optimizer=tf.keras.optimizers.Adam(0.01),
                  metrics=tf.keras.metrics.RootMeanSquaredError())

    return model
model = build_and_compile_model(normalizer)
model.summary()

```

Setting early stopping

Training a model for too many epochs can lead to overfitting of the training dataset, whereas training it for too few may result in an under fit model. Early stopping is a method that allows us to specify an arbitrary large number of training epochs and stop training once the model performance stops improving on a holdout validation dataset.

```
Es = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patience=10, restore_best_weights=True)
```

Fit the model

```
history = model.fit(train_features, train_labels, epochs=2000, validation_data=(test_features, test_labels), callbacks=[Es])
```

Evaluate the model

```
test_results = {}
test_results['model'] = model.evaluate(test_features, test_labels, verbose=1)
def plot_loss(history):
    #Plot training history

    plt.plot(history.history['loss'], label='loss')
    plt.plot(history.history['val_loss'], label='val_loss')
    plt.ylim([0, 2000])
    plt.xlabel('Epoch')
    plt.ylabel('Error [Hombole(residual)]')
    plt.legend()
    plt.grid(True)
plot_loss(history)
#Plot the scatter plot for test dataset
test_predictions = model.predict(test_features).flatten()
a = plt.axes(aspect='equal')
plt.scatter(test_labels, test_predictions)
plt.xlabel('True Values [residual error]')
plt.ylabel('estimated error')
lims = [0, 500]
plt.xlim(lims)
plt.ylim(lims)
_ = plt.plot(lims, lims)
```

Appendix E

Appendix E1. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Akaki station NB: The optimum hyper parameters in figure c and d are marked in green.

Model: "sequential_12"

Layer (type)	Output Shape	Param #
normalization_12 (Normalization)	(None, 5)	11
dense_24 (Dense)	(None, 3)	18
dense_25 (Dense)	(None, 1)	4

Total params: 33
Trainable params: 22
Non-trainable params: 11

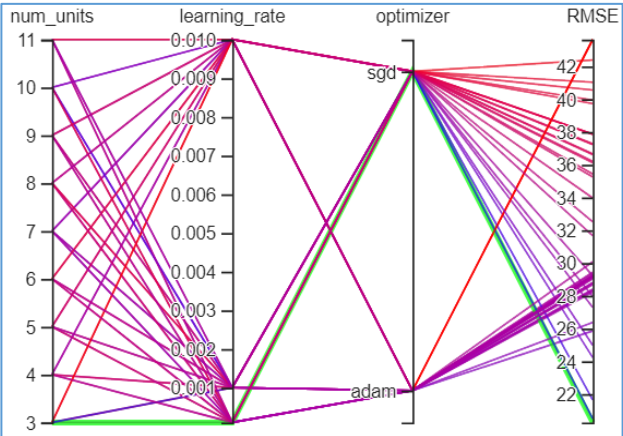
a) ANN model structure

Model: "sequential_14"

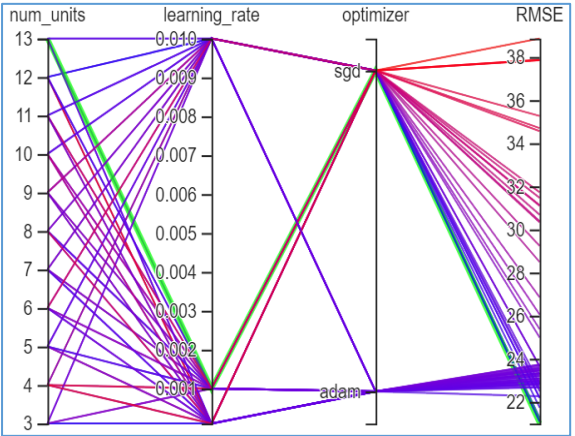
Layer (type)	Output Shape	Param #
normalization_14 (Normalization)	(None, 6)	13
dense_28 (Dense)	(None, 13)	91
dense_29 (Dense)	(None, 1)	14

Total params: 118
Trainable params: 105
Non-trainable params: 13

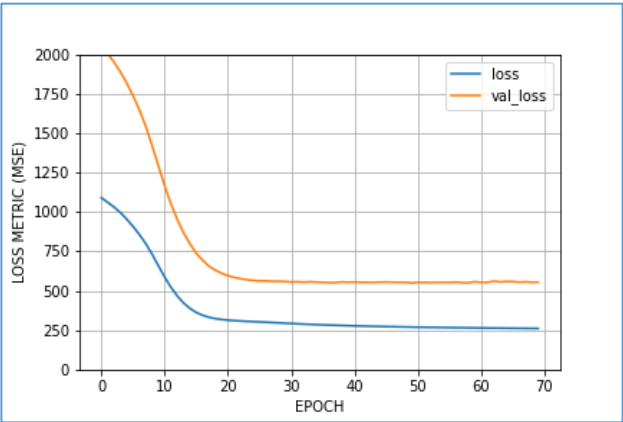
b) Hybrid1 model structure



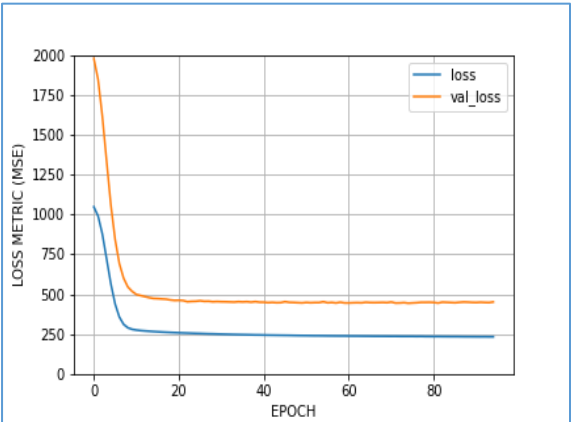
c) Hyperparameter for ANN



d) Hyperparameter for Hybrid1



e) ANN model training history



f) Hybrid1 model training history

Appendix E2. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Melka Kunture station

Model: "sequential_43"

Layer (type)	Output Shape	Param #
normalization_42 (Normalization)	(None, 4)	9
dense_86 (Dense)	(None, 12)	60
dense_87 (Dense)	(None, 1)	13

=====
Total params: 82
Trainable params: 73
Non-trainable params: 9
=====

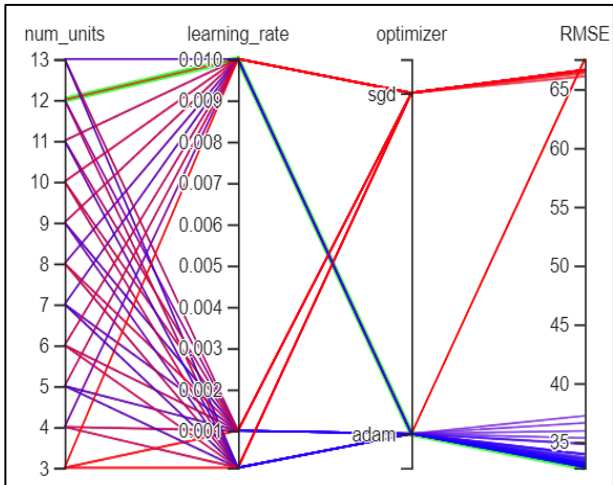
a) ANN model structure

Model: "sequential_40"

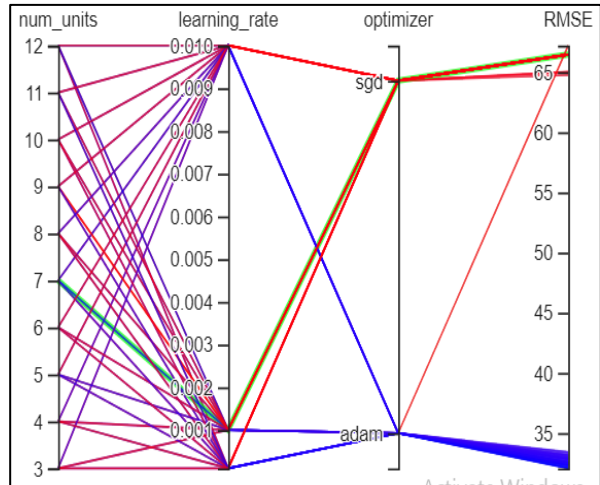
Layer (type)	Output Shape	Param #
normalization_40 (Normalization)	(None, 7)	15
dense_80 (Dense)	(None, 7)	56
dense_81 (Dense)	(None, 1)	8

=====
Total params: 79
Trainable params: 64
Non-trainable params: 15
=====

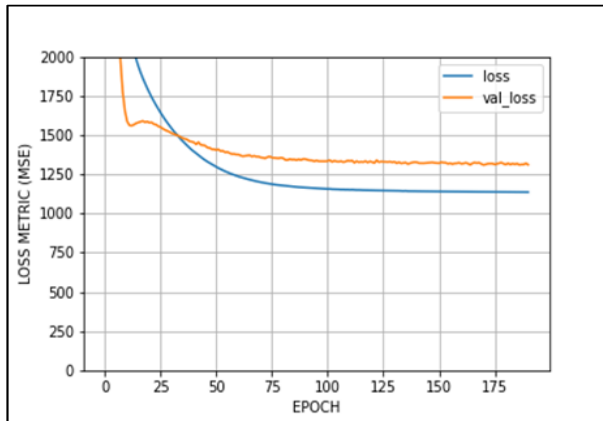
b) Hybrid1 model structure



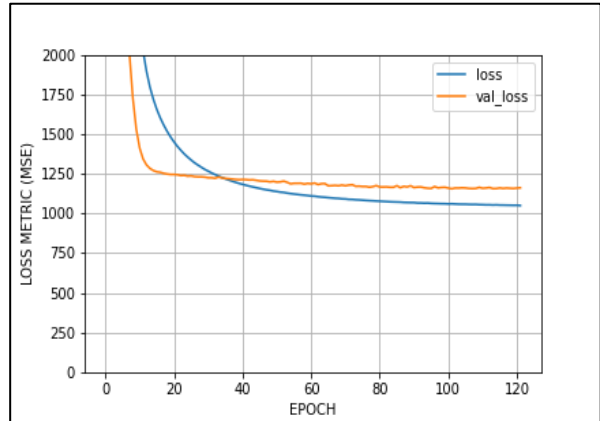
c) Hyperparameter for ANN



d) Hyperparameter for Hybrid1



e) ANN model training history



f) Hybrid1 model training history

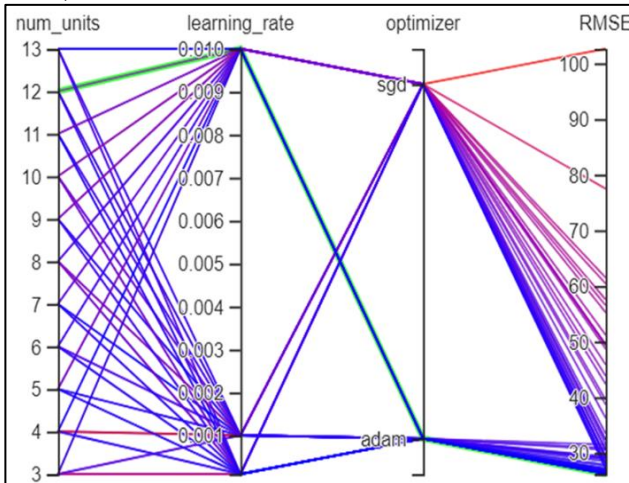
Appendix E3. Stand-alone ANN and Hybrid1 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Hombole station

Model: "sequential_22"		
Layer (type)	Output Shape	Param #
=====		
normalization_22 (Normaliza	(None, 6)	13
tion)		
dense_44 (Dense)	(None, 12)	84
dense_45 (Dense)	(None, 1)	13
=====		
Total params: 110		
Trainable params: 97		
Non-trainable params: 13		

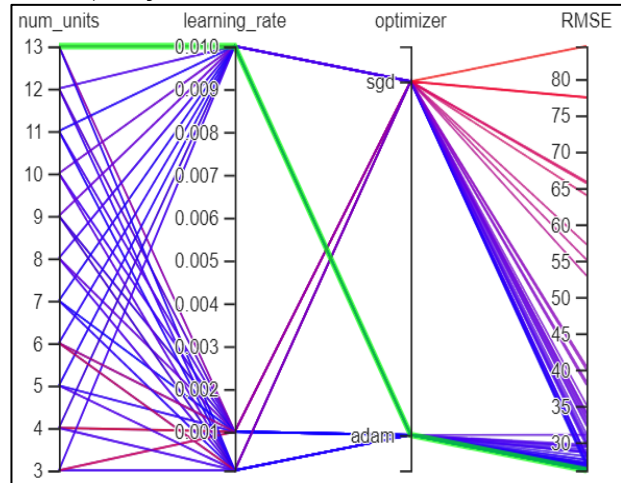
a) ANN model structure

Model: "sequential_40"		
Layer (type)	Output Shape	Param #
=====		
normalization_40 (Normaliza	(None, 7)	15
tion)		
dense_80 (Dense)	(None, 13)	104
dense_81 (Dense)	(None, 1)	14
=====		
Total params: 133		
Trainable params: 118		
Non-trainable params: 15		

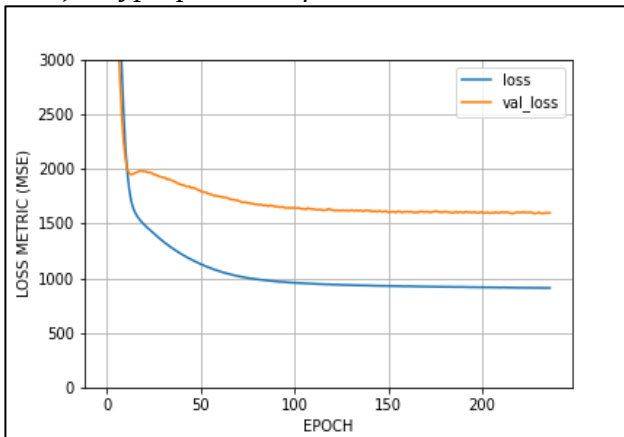
b) Hybrid1 model structure



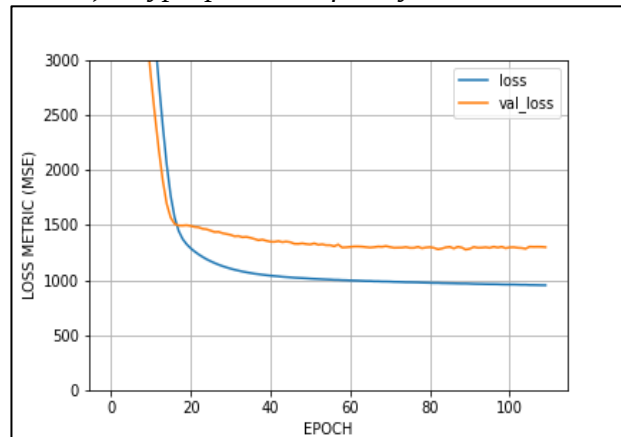
c) Hyperparameter for ANN



d) Hyperparameter for Hybrid1



e) ANN model training history



f) Hybrid1 model training history

Appendix F

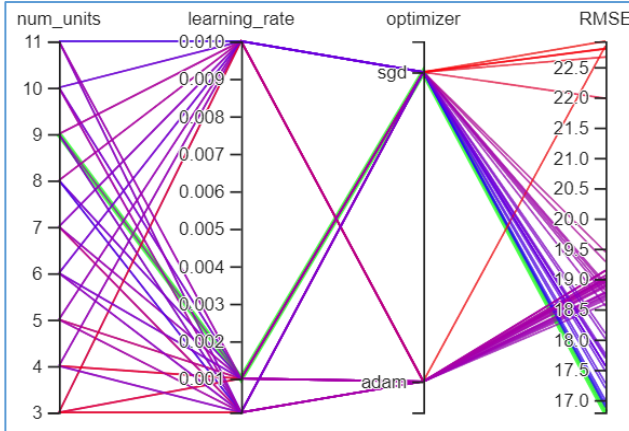
Appendix F1. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Akaki station

Model: "sequential_12"

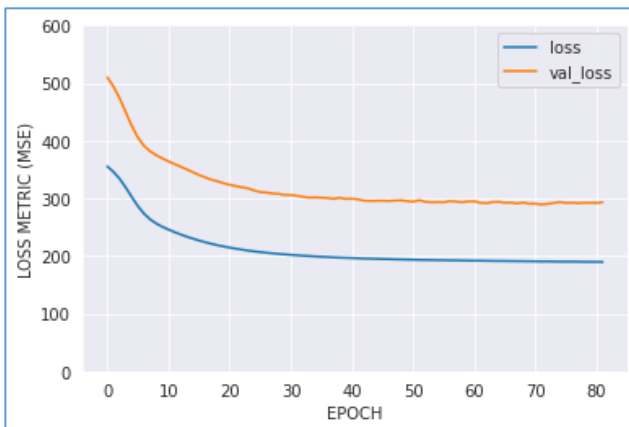
Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 6)	13
dense_24 (Dense)	(None, 9)	63
dense_25 (Dense)	(None, 1)	10

Total params: 86
Trainable params: 73
Non-trainable params: 13

a) Residual model structure



c) Hyperparameter for Residual model



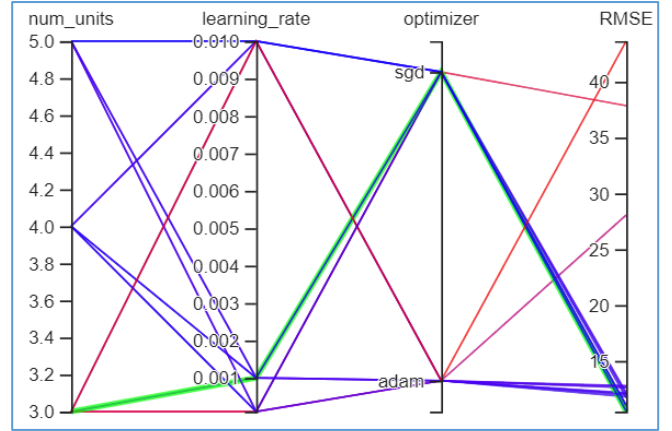
e) Residual model training history

Model: "sequential_3"

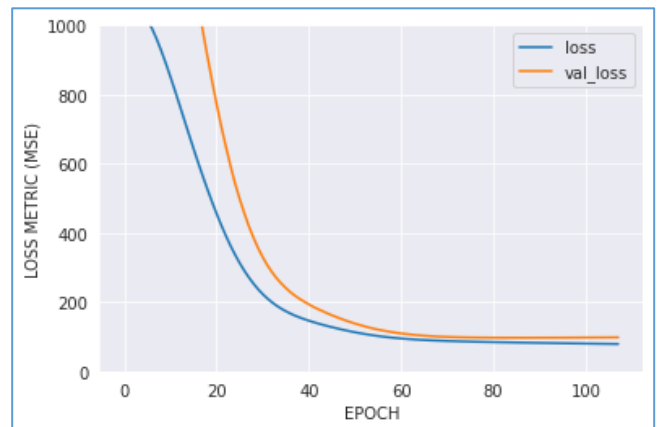
Layer (type)	Output Shape	Param #
normalization_3 (Normalization)	(None, 2)	5
dense_6 (Dense)	(None, 3)	9
dense_7 (Dense)	(None, 1)	4

Total params: 18
Trainable params: 13
Non-trainable params: 5

b) Hybrid2 model structure



d) Hyperparameter for Hybrid2



f) Hybrid2 model training history

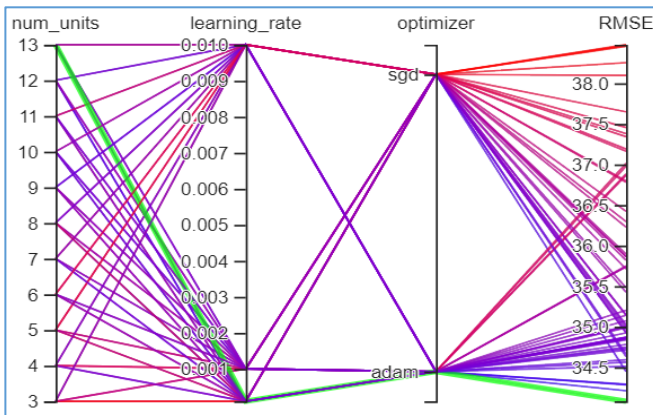
Appendix F2. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Melka Kunture station

Model: "sequential_7"		
Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 6)	13
dense_14 (Dense)	(None, 13)	91
dense_15 (Dense)	(None, 1)	14
Total params: 118		
Trainable params: 105		
Non-trainable params: 13		

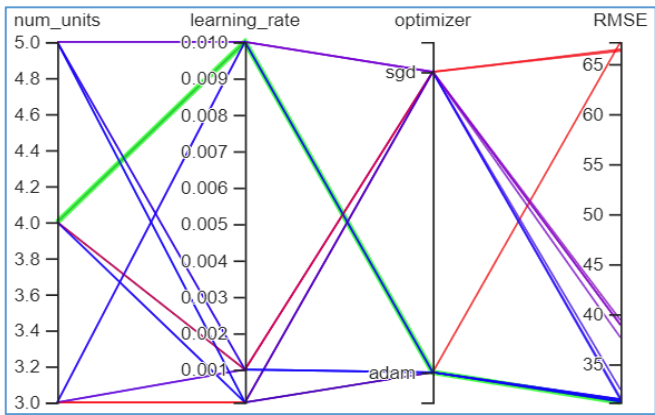
a) Residual model structure

Model: "sequential"		
Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 2)	5
dense (Dense)	(None, 4)	12
dense_1 (Dense)	(None, 1)	5
Total params: 22		
Trainable params: 17		
Non-trainable params: 5		

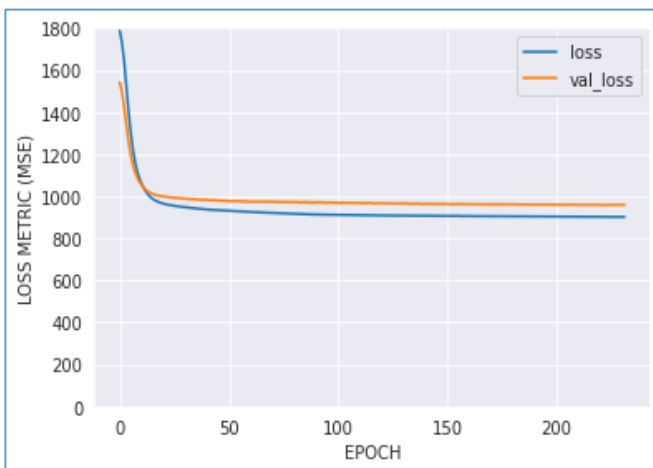
b) Hybrid2 model structure



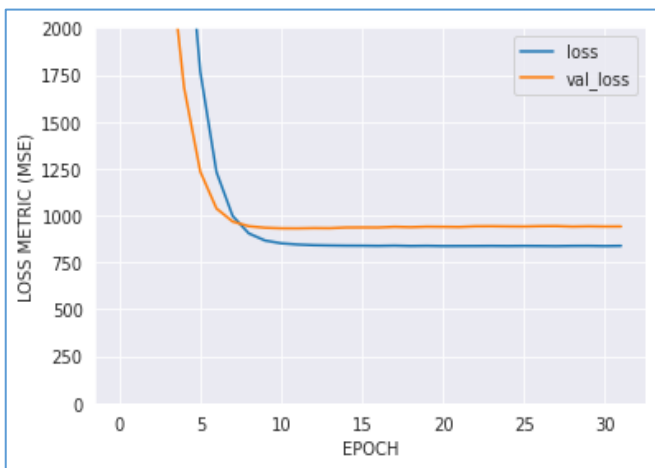
c) Hyperparameter for Residual model



d) Hyperparameter for Hybrid2



e) Residual model training history

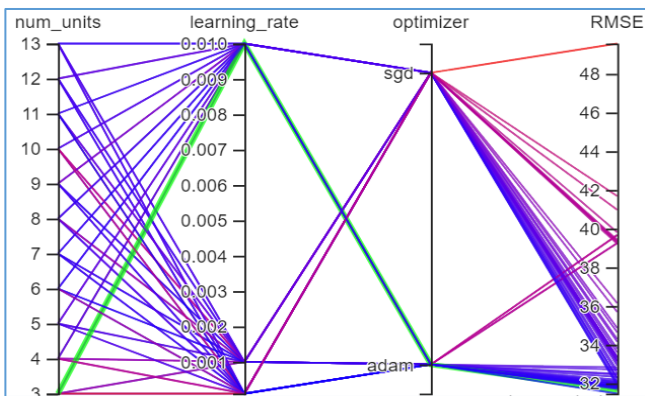


f) Hybrid2 model training history

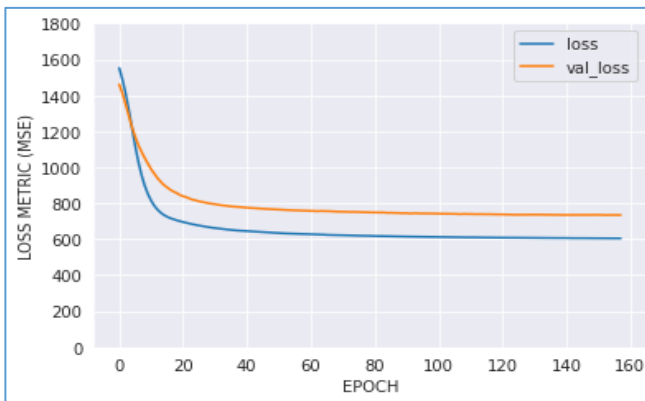
Appendix F3. Residual and Hybrid2 model structure (a and b), optimum parameter (c and d), and training history (e and f) for Hombole station

Model: "sequential"		
Layer (type)	Output Shape	Param #
normalization (Normalization)	(None, 6)	13
dense (Dense)	(None, 3)	21
dense_1 (Dense)	(None, 1)	4
Total params: 38		
Trainable params: 25		
Non-trainable params: 13		

a) Residual model structure



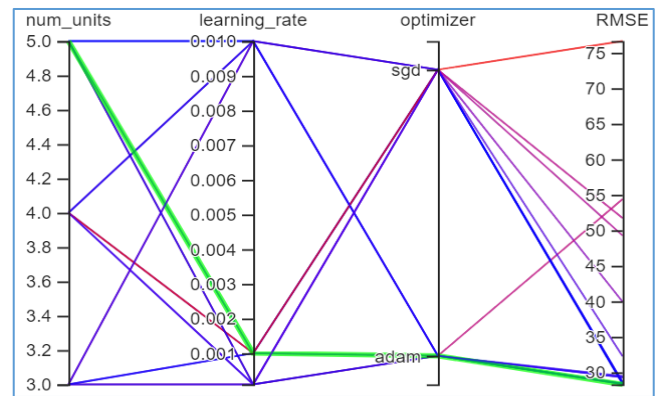
c) Hyperparameter for Residual model



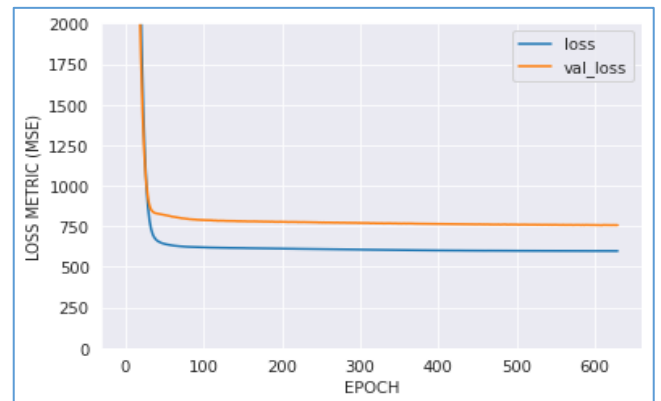
e) Residual model training history

Model: "sequential_2"		
Layer (type)	Output Shape	Param #
normalization_1 (Normalization)	(None, 2)	5
dense_4 (Dense)	(None, 5)	15
dense_5 (Dense)	(None, 1)	6
Total params: 26		
Trainable params: 21		
Non-trainable params: 5		

b) Hybrid2 model structure



d) Hyperparameter for Hybrid2



f) Hybrid2 model training history