Hydro-Climatic Analysis and Future Water Resource Projections: Integrating CMIP6 Models and Hydrological Models in the Upper Awash Sub-Basin, Ethiopia

Yonas Abebe Balcha

Vollständiger Abdruck der von der Fakultät für Bauingenieurwesen und Umweltwissenschaften der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

angenommenen Dissertation.

Gutachter/Gutachterin

1. Prof. Dr.-Ing. habil Andreas Malcherek
2. Associate Prof. Tena Alamirew Agumassie

Die Dissertation wurde am 15. Februar 2024 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Bauingenieurwesen und Umweltwissenschaften am 13. Juni 2024 angenommen. Die mündliche Prüfung fand am 13. Juni 2024 statt.

Acknowledgment

This research project has received financial support from the German Academic Exchange Service (DAAD) and the Ethiopian Ministry of Education (MoE) under the ECBP Homegrown PhD Program, 2019. I am sincerely grateful for the funding provided by these institutions, which has been instrumental in the successful completion of this study.

I extend my heartfelt appreciation to Universität der Bundeswehr München, the host university, not only for providing financial support but also for creating a conducive research environment. Special thanks to my PhD supervisor and Head of the Institute of Hydromechanics and Hydraulic Engineering, Univ.-Prof. Dr.-Ing. Andreas Malcherek, for his consistent mentorship, guidance, patience, and prompt feedback on all my submissions. Vielen Dank!

My heartfelt gratitude also goes to my Ethiopian supervisor, Dr. Tena Alamirew, for his unwavering support and guidance throughout the Ph.D. process and his kindness.

I would also like to express my gratitude to all my colleagues, the technical laboratory staff, and the head of the laboratory, Dr.-Ing. Ivo Baselt, for their unwavering support and guidance throughout the research period.

Lastly, I want to convey my deepest thanks to my wife Beza Yitbarek and other family members for their emotional support and understanding throughout the entire research program.

Zusammenfassung

In dieser Studie wurde eine Untersuchung der zukünftigen Wasserressourcenbedingungen des Untersuchungsgebiets durchgeführt, indem Klimamodellausgaben aus dem neuen CMIP6-Archiv mit hydrologischen Modellen integriert wurden. Es wurden vergleichende Bewertungen zwischen den Ergebnissen hydrologischer Modelle, die datengesteuerte Ansätze nutzen, und solchen, die aus dem Soil and Water Assessment Tool (SWAT) abgeleitet wurden, durchgeführt. Die Motivation für diese Forschung resultiert aus den Diskrepanzen in den Ergebnissen zahlreicher früherer Studien im Teilbecken, die hauptsächlich darauf abzielten, vergangene hydroklimatische Bedingungen zu erläutern.

Um Einblick in die zukünftige Wasserverfügbarkeit im Teilbecken zu erhalten, wurden zwölf Klimamodelle zunächst ausgewählt und auf der Grundlage ihrer Fähigkeit zur Simulation beobachteter historischer Klimazonen bewertet. Ein akribischer Downscaling-Prozess wurde unter der CDO-Plattform implementiert, gefolgt von einer Bias-Korrektur unter Verwendung der QM-Technik für alle Klimamodelldaten. Nach einer rigorosen Bewertung wurden fünf Klimamodelle (ECEARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2 und INM-CM5-0) aufgrund ihrer relativ robusten Potenziale zur Erfassung verschiedener Merkmale der beobachteten Klimaserien identifiziert.

Zeitreihenmodelle, die unter Regimewechsel- und datengesteuerten Ansätzen kategorisiert sind, wurden auf ihr Potenzial hin untersucht. Selbstanregende TAR (SETAR) und Logistisch-Glatte TAR (LSTAR) Modelle, Variationen von TAR-Modellen, wurden auf die Durchflussdaten ausgewählter Stationen im Teilbecken mit lobenswerter Genauigkeit angewendet. Unter den datengesteuerten Modellen (DDMs) zeigte das Adaptive Neuro-Fuzzy Inference System (ANFIS) Modell das größte Potenzial bei der Simulation der Durchflussreihen ausgewählter Pegelstationen im Teilbecken. Schließlich wurde das Nonlinear Auto-Regressive with eXogenous inputs (NARX) Modell aus den DDMs aufgrund seiner minimalen Datenanforderungen und einfachen Modellkonfiguration ausgewählt. Zusätzlich unterzog sich das SWAT-Modell einer Kalibrierung und Validierung.

In der Folge wurde das NARX-Modell zusammen mit dem SWAT-Modell eingesetzt, um zukünftige Wasserressourcenbedingungen im Teilbecken zu analysieren.

Unter Verwendung des trainierten NARX-Modells und des zuvor validierten SWAT-Modells fand die Untersuchung der zukünftigen Wasserressourcenbedingungen am Auslass des Teilbeckens mit Hilfe neuer CMIP6-Klimaszenarien statt. Diese Szenarien wurden durch die Zusammenstellung der Ausgaben der fünf ausgewählten Klimamodelle abgeleitet. Die Analyse prognostizierte einen Anstieg des Frühjahrs- und Sommerabflusses für alle Szenarien (SSP1.26, SSP2.45 und SSP5.85) über nahe (2022-2039), mittlere (2040-2069) und ferne (2070-2099) Zeiträume hinweg. Darüber hinaus deuteten die Ergebnisse aus der Analyse des Standardized Discharge Index (SDI) für simulierte zukünftige Wasserverfügbarkeit auf eine höhere Wahrscheinlichkeit von feuchten Ereignissen im Vergleich zu trockenen hin. Angesichts der bestehenden Landnutzungs- und Wasserressourcenmanagementbedingungen im Teilbecken legen diese Ergebnisse eine erhöhte Wahrscheinlichkeit zukünftiger Überschwemmungen in den niedrig gelegenen Gebieten des Teilbeckens und der aufeinanderfolgenden unterstromgelegenen Teile des Awash-Beckens nahe.

Abstract

In this study, an investigation into the prospective water resource conditions of the study area was conducted by integrating climate model outputs from the new CMIP6 archive with hydrological models. Comparative assessments were made between the outcomes of hydrological models utilizing data-driven approaches and those derived from the Soil and Water Assessment Tool (SWAT). The motivation for this research stems from the discrepancies in the outputs of numerous past studies in the sub-basin that primarily aimed to elucidate past hydro-climatic conditions.

To gain insight into future water availability within the sub-basin, twelve climate models were initially selected and evaluated based on their capacity to simulate observed historical climates. A meticulous downscaling process was implemented under the CDO platform, followed by bias correction using the QM technique on all climate model outputs. Following a rigorous evaluation, five climate models (ECEARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2, and INM-CM5-0) were identified for their relatively robust potential in capturing various characteristics of the observed climate series.

Time series models, categorized under regime switching and data-driven approaches, were scrutinized for potential. Self-Exciting TAR (SETAR) and Logistic-Smooth TAR (LSTAR) models, variations of TAR models, were applied to the flow data of selected stations in the sub-basin with a commendable degree of accuracy. Among the data-driven models (DDMs), the Adaptive Neuro-Fuzzy Inference System (ANFIS) model exhibited the greatest potential in simulating the flow series of selected gauging stations in the sub-basin. Ultimately, the Nonlinear Auto-Regressive with eXogenous inputs (NARX) model from DDMs was chosen for its minimal data requirements and ease of model setup. Additionally, the SWAT model underwent calibration and validation.

Subsequently, the NARX model, in conjunction with the SWAT model, was employed to analyze future water resource conditions in the sub-basin.

Using the trained NARX model and the previously validated SWAT model, the examination of future water resource conditions took place at the sub-basin outlet with the assistance of new CMIP6 climate scenarios. These scenarios were derived by ensembling the outputs of the five selected climate models. The analysis projected an increase in both spring and summer flow for all scenarios (SSP1.26, SSP2.45, and SSP5.85) across near (2022-2039), mid (2040-2069), and far (2070-2099) time periods. Additionally, the outputs indicated a decrease in dry period flow. Results from the Standardized Discharge Index (SDI) analysis on simulated future water availability suggested a higher likelihood of wet events compared to dry ones. Considering existing land use and water resource management conditions in the sub-basin, these findings imply an increased probability of future flooding in the low-lying areas of the sub-basin and successive downstream parts of the Awash basin.

Contents

List of Figures

List of Tables

Acronyms

AIC Akaike Information Criterion

ANFIS Adaptive Network-based Fuzzy Inference System

ANN Artificial Neural Network

AOGCM Atmospheric/Ocean General Circulation Model

ARIMA Autoregressive Intergrated Moving Average

CDF Cumulative Density Function

CDO Climate Data Operator

CMIP-GCMs Climate Model Intercomparision Project Global Climate Models

CMIP6 Climate Model Intercomparision Project Six

DDM Data Driven Models

ECDF Emperical Cumulative Density Function

FAO Food and Agriculture Organization

IDW Inverse Distance Weighted

JJAS June July August September

LSTAR Logistic Smooth Threshold Auto-Regressive

MAM March April May

NSE Nash Sutcliffe Efficiency

PCA Principal Component Analysis

PDF Probability Density Function

QM Quantile Mapping

RMSE Root Mean Square Error

SSP Shared Socioeconomic Pathway

TAR Threshold Auto-Regressive

WRCP World Climate Research Program

Chapter 1

Introduction

1.1 Background

Water is found in most life forms as one of the dominant components, constituting over 55–78% of human body weight [3], and approximately 70% of plant body is made up of water [4]. Additionally, 71% of the Earth's surface is covered by different types of water bodies [5]. Due to this ubiquity, water is a critical factor for the existence of living things and is also one of the crucial components in the hydrosphere that shapes the climate of our planet.

As with most natural resources found on Earth or in its surrounding environment, water is considered both a renewable and non-renewable resource. It is renewable since it can be easily replenished through the action of the hydrological cycle into various forms such as freshwater, seawater, and atmospheric water. It becomes non-renewable if the rate of consumption is much higher than renewal, especially for freshwater, which constitutes only 2.5% of the global water budget. Of this percentage, 68.5% is stored in ice sheets and glaciers, making it inaccessible, and only the remaining 31% is considered an available resource found in surface water, soil moisture, and much of it as groundwater [5].

This limited resource is often widely used to satisfy different human needs, classified under consumptive or non-consumptive water use. Consumptive uses have a tendency to diminish the resource, making it unavailable for other uses, such as water diverted from a stream reach for irrigation, drinking, or industrial application. Non-consumptive water use, on the contrary, doesn't diminish the resource since the water is used at the same location of diversion. Therefore, hypothetically there is no loss in both quantity and quality, as seen with water used for hydropower generation. To meet these different types of uses, it is required to develop water infrastructures like dams, reservoirs, diversion structures, canals, etc. It has been understood that, in addition to the actual water use, the different types of structures built to develop the resource will alter the ecosystem of that location, which will in turn affect the hydrological cycle. This takes place at both local and global scales, and at the global scale, global warming is thought to continuously change the global hydrological cycle. All these factors, in addition to natural climate variability, will have an effect on the temporal and spatial availability of the resource in a given place.

Water security is a term used to define the capability of a society to have access to enough water of sufficient quality to carry out different productive activities [6]. Some of the reasons why water security is undermined are population growth, pollution, land use land cover (LULC) changes, and extreme climate conditions due to climate change. These issues can be resolved through sustainable water management practices, which ensure the availability of good-quality water and avoid its degradation. The endeavor of sustainable management of these resources requires a detailed knowledge of what is going on in the study area, enabling better-informed decisions about future water resource development and management.

A good grasp of water resource conditions in a given area could be achieved through the use of different approaches that focus on different spatial (area, catchment, or basin) or temporal (daily, seasonal, or annual) scales. The first approach is through the analysis of long-term variations in runoff and meteorological elements either by statistical analysis of the relation between runoff and other meteorological variables or by observing past extreme events [7, 8, 9]. These meteorological variables include precipitation, temperature, evapotranspiration, etc. The second approach is through a water balance method over a long period [10, 11]. Lastly, it could be achieved through the study of estimates of changes in the climatic and hydrologic characteristics for large regions using Global Climate Model (GCM) outputs [12]. Another alternative, which is most common nowadays, is the use of deterministic hydrologic models [13, 14].

Nowadays, the use of climate models in trying to understand the future change in water resource potential is a common practice. In trying to observe future hydro-meteorological changes, climate models are important tools whose capabilities can't be underestimated. There are a number of climate models introduced by different institutes throughout the world, and the source of variation is mainly due to model structure, assumptions, calibration processes, and parameterization [15]. In fact, the uncertainty in hydrologic projections is more from GCMs rather than emission scenarios or model parameterization [16]. So, there should be a mechanism to differentiate out models which can closely interpret area-specific conditions.

1.2 Statement of Problem

One of the major sources of water insecurity in Ethiopia is the uneven distribution of the resource, with approximately 70% found in the western part of the country [17]. More than 27 million people in Ethiopia are living in high or extremely high water risk areas, and a further increase in population and economic activity in these areas exacerbates the situation even more [18].

The Awash Basin is one of the most heavily utilized basins found in these water-stressed parts of Ethiopia [17]. The basin encompasses some of the major towns and cities in the country, including the capital city. Additionally, there are numerous industries, and agriculture is highly practiced, whether it is rain-fed or irrigated. From the total water consumed by industries in the country, 55% of it comes from this basin [18].

The basin is also highly affected by hydro-climatic extremes of floods and droughts. Areas affected by flooding range from the upper to lower parts of the basin. According to a recent flood event, around 162,921 people were affected in the lower Awash sub-basin [19]. The most frequently flood-affected areas are Sebeta Hawas, Wolmera, and Egeria Woredas found in the upper part, and Fentale Woreda in the lower part of the Awash basin [20]. According to Gebreyesus et al. (2020) [21], in 2015-2016, close to 90% of the basin was affected by a moderate to severe level of drought, marking one of the worst droughts in decades.

A comprehensive understanding of the water resource conditions in a specific basin is crucial for making informed decisions and effectively managing the resource. While previous studies have delved into various aspects of water resource conditions, particularly in the upper part of the basin, there has been a gap in employing a diverse range of approaches to thoroughly examine the water resource conditions in the Upper Awash sub-basin.

1.3 Research Questions

This study is mainly based on two core research questions:

- What will be the outlook for water availability under future climate conditions?
- How well do deterministic and stochastic stream flow models perform in simulating future water availability?

1.4 Research Objectives

- To comprehensively evaluate and integrate climate model outputs, hydrological models, and data-driven approaches for a thorough analysis of the current and future water resource conditions in the Upper Awash Sub-basin, aiming to provide insights into potential hydrological changes and flood risks in the region.

1.4.1 Specific Objectives

- To evaluate and select a suitable climate model from CMIP6 archive for the sub-basin.
- To study the flow characteristics of some selected rivers
- To evaluate and select suitable stream flow forecasting model for the sub-basin.
- To investigate temporal variation of water availability in the sub - basin.
- To compare the result of the SWAT model with that of the stream flow model output.

1.5 Structure of the Thesis

This thesis comprises seven chapters. **Chapter 1** provides a general background statement on the scientific basis and focus of the study, along with the declaration of research objectives.

Chapter 2 presents detailed information about the Upper Awash Sub-basin, encompassing its location, topography, climate, hydrology, geology, soil, and land use land cover conditions.

Chapter 3 undertakes an evaluation of twelve climate models from CMIP6, employing various techniques. The assessment focuses on the models' ability to simulate observed climate series, specifically for precipitation, maximum temperature, and minimum temperature. This chapter also discusses the future climate conditions of the basin based on the selected climate model outputs.

Chapter 4 delves into the flow characteristics of the series obtained from gauging stations of major rivers in the Upper Awash Sub-basin. The analysis includes the study of stationarity, linearity, trend, and seasonal characteristics of each time series.

Chapter 5 evaluates the potential of selected time series modeling techniques, specifically those belonging to Regime Switching and Data Driven approaches. Two Regime Switching models, SETAR and LSTAR, are fitted to the observed flow series of each gauging station. Additionally, Data Driven techniques involve defining and fitting appropriate NARX and ANFIS models.

Chapter 6 analyzes future water resource conditions in the basin using flow simulations obtained from climate model outputs. SWAT model and NARX models are employed to generate future flow simulations based on climate model outputs. Future water availability is observed, primarily utilizing the Standard Drought Index (SDI) and flood frequency technique.

Chapter 7 summarizes the main results of the thesis, concludes the findings, and provides recommendations for future research.

Chapter 2

Description of the Study Area

2.1 Location and Topography

The study is focused on Upper Awash sub-basin which is found in central Ethiopia and is located between a longitude of $37°57'4''E - 39°17'28''E$ and latitude of $8°4'52''N - 9°19'47''N$ as shown in Figure: 2.1.

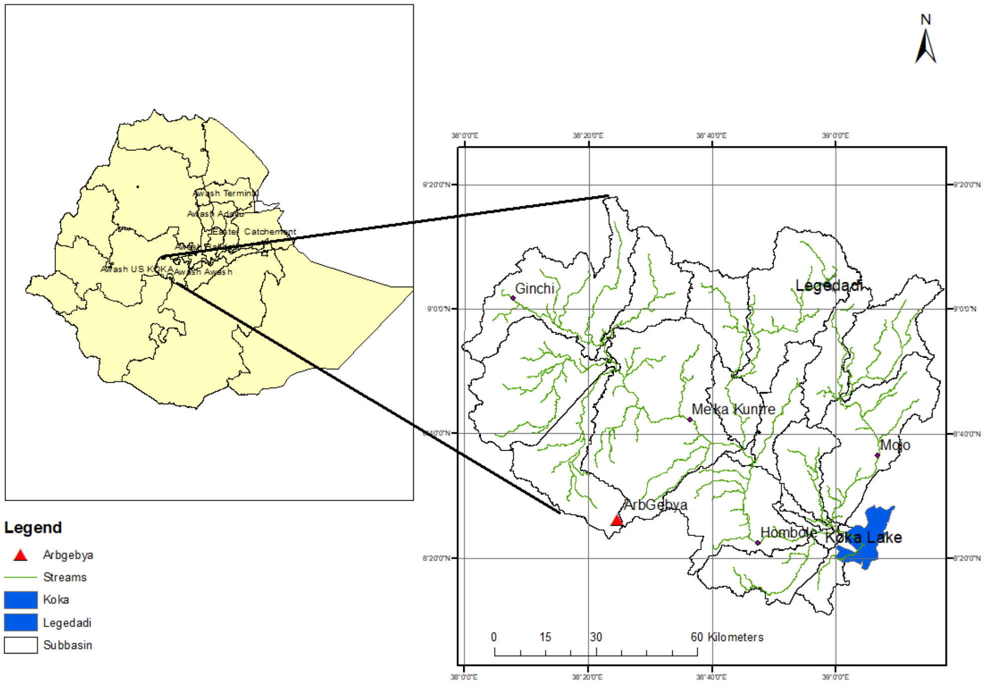


Figure 2.1: Map of Upper Awash sub-basin with major rivers and location of hydrological gauge stations.

Utilizing the Food and Agriculture Organization (FAO) slope gradient classification [22], 96.1% of the drainage basin exhibits sloping to steep terrain, while the remaining portion features a flat to gently sloping landscape. Figure:2.2 illustrates that the northern, western, and south-western sectors of the basin predominantly consist of steep topography. The sub-basin's highest point, near ArbGebya, reaches an elevation of 3561 m, while its lowest point at Koka Lake is situated at an altitude of 1547 m (Figure:2.1 and 2.2).

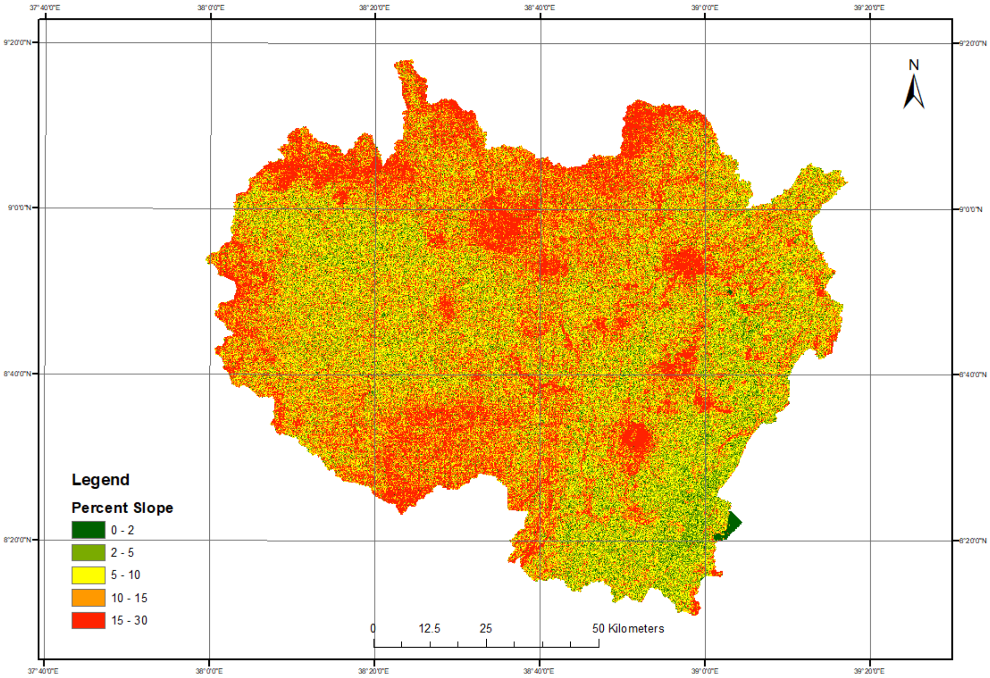


Figure 2.2: Slope map of Upper Awash sub-basin

2.2 Climatology and Hydrology

The annual climate characteristics of the basin are primarily influenced by two air currents—an upward flow from the Atlantic and southern Pacific Ocean to the west and another from the Indian Ocean to the southeast. The convergence of Atlantic-equatorial Pacific westerlies and Indian south easterlies over the basin serves as the main moisture source for precipitation [23] [24]. The basin experiences a major rainfall season lasting from June to September and a shorter rainy season from March to April. The mean annual areal precipitation in the sub-basin is 1020 mm, with station-wise variation ranging from a minimum of 861 mm at Boneya to a maximum of 1223 mm in Addis Ababa. Similarly, the mean annual areal temperature is 17.6°C, with station-wise variation from 16.8°C at Addis Alem to 18.32°C in Debrezeit. A total of 15 rainfall and 10 temperature stations are utilized, and the list of stations is presented in Table:2.1.

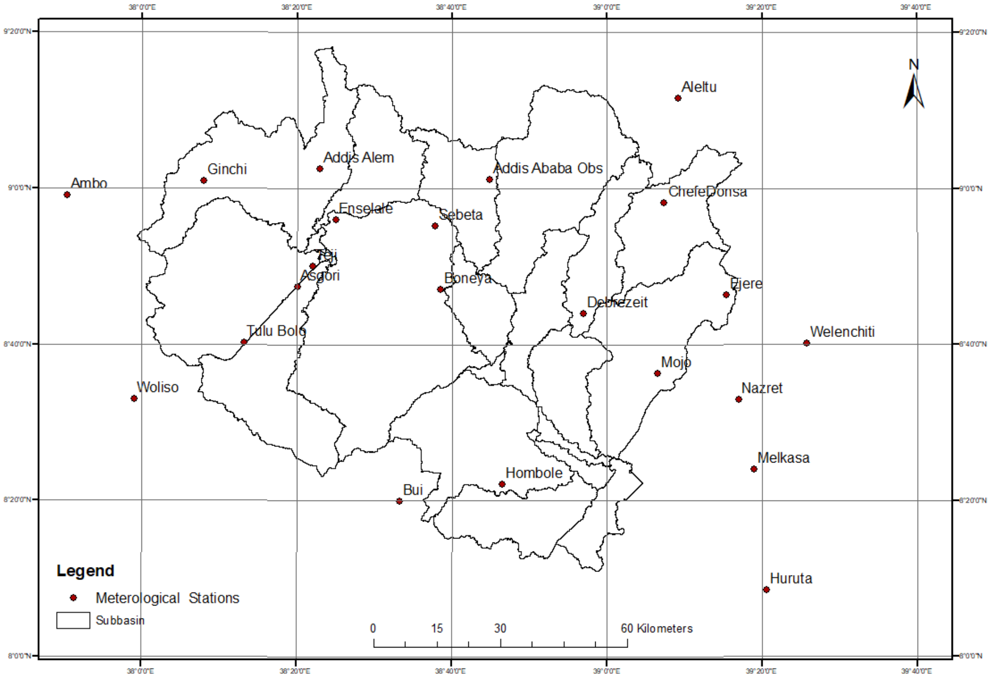


Figure 2.3: Map of Upper Awash sub-basin with location of metrological stations.

Table 2.1: List of names of meteorological stations in the Upper Awash sub-basin with their locations (Universal Transfer Mercator (Universal Transverse Mercator)) and altitude (m) used for analysis.

No.	Station Name	Location (y)	Location (x)	Altitude
1	Addis Ababa Obs*	472248.08	996952.4	2386
2	Addis Alem**	432225.93	999552.91	2372
3	Aleltu	516771.63	1016119.33	2648
4	Ambo**	372449.79	993358.73	2068
5	Asgori*	426775.97	971700.31	2072
6	Boneya	460591.24	971046.06	2251
7	Bui**	450940.01	920899.89	2054
8	ChefeDonsa	513542.53	991537.73	2392
9	Debrezeit*	494500.33	965370.79	1900
10	Ejere	528246.66	969798.67	2254
11	Enselale	435870.79	987532.37	2000
12	Ginchi	404738.43	996808.09	2132
13	Hombole	475209.16	925006.74	1743
14	Huruta*	537697.21	900012.17	2044
15	Melkasa*	534861.76	928532.98	1540
16	Mojo	511901.68	951220.7	1763
17	Nazret	531179.91	945113.49	1622
18	Sebeta	459322.66	986027.98	2220
19	Teji	430354.91	976481.48	2091
20	Tulu Bolo*	414188.26	958456.67	2100
21	Welenchiti	547305.49	958395.38	1458
22	Woliso**	388113.76	945249.62	2058

N.B: The ones without and single star (*) are used for precipitation and the others with one (*) and double star (**) are used to differentiate stations used for temperature analysis

With a total drainage area of 10556.45 km^2, the basin significantly contributes the majority of the annual flow to the larger Awash Basin, as depicted in the top-left map of Figure:2.1. The river originates from two major areas: Ginchi on the left side and the Legedadi reservoir on the right side of the basin. Downstream, the streams emerging from these two areas converge to form the Awash River, which is further augmented by the Mojo River before reaching Koka Lake. The longest stream in the sub-basin spans a length of 227.67 km.

2.3 Land Use Land Cover, Soil, and Geology

The land use land cover (LULC) of the basin primarily consists of agriculture, forest, pasture, shrubs, and urban (settlement). Agriculture dominates the land use, occupying nearly 85% of the total land in the sub-basin, as illustrated in Figure:2.4. Settlement and forest follow as the second and third highest, respectively. According to a recent study, agriculture and settlement are identified as the fastest-growing LULCs in the sub-basin [25].

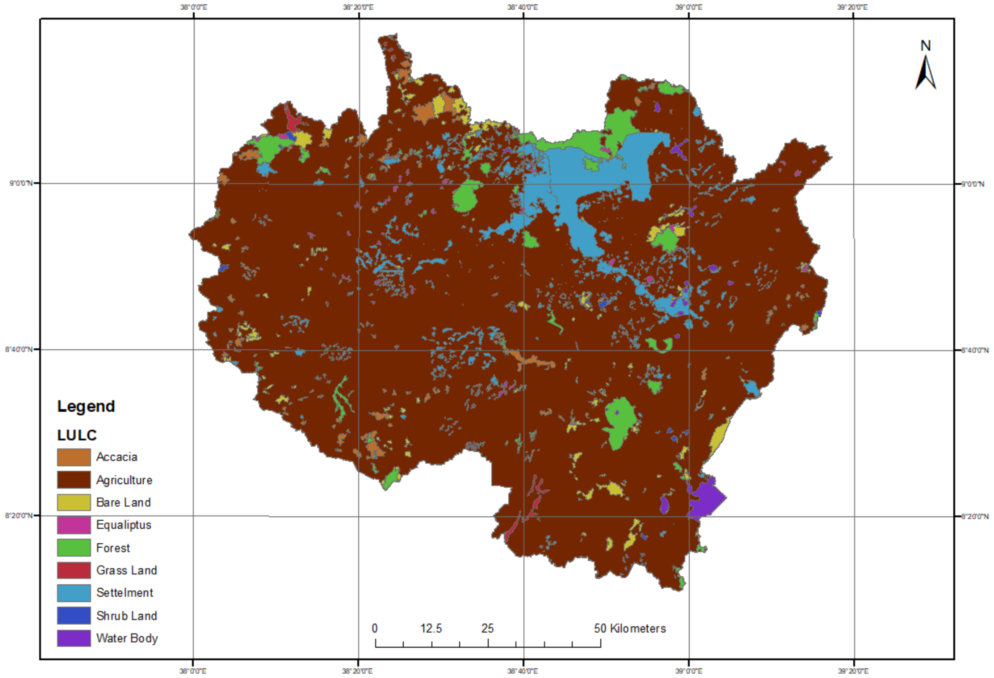


Figure 2.4: Map of major land use land cover types in the Upper Awash sub-basin (Source:AWBA,2017 [1])

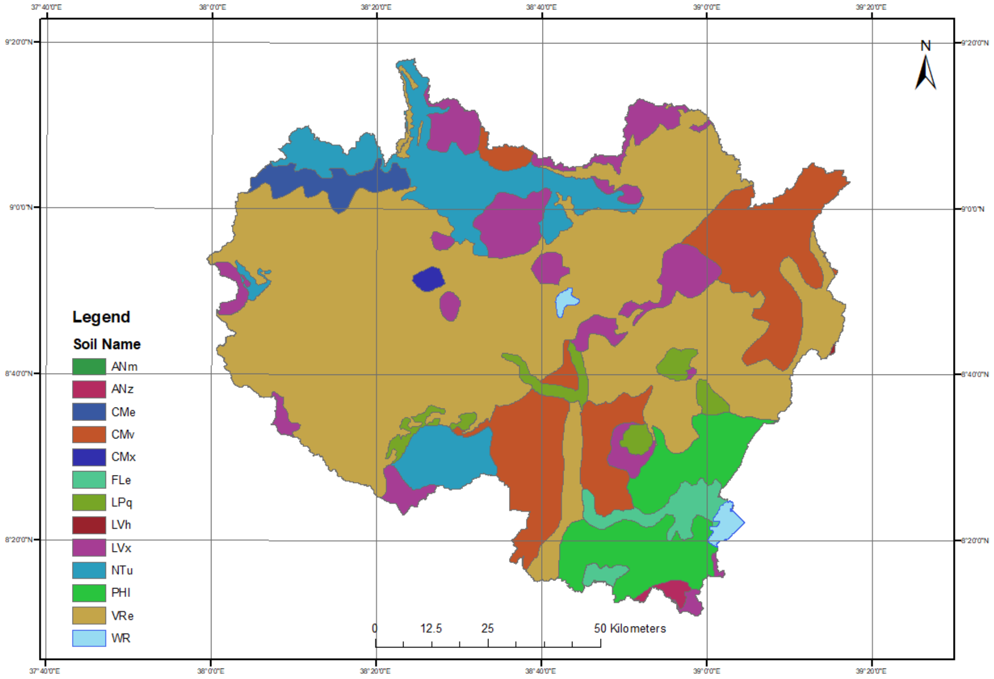


Figure 2.5: Map of major soil types in the Upper Awash sub-basin (Source:AWBA,2017 [1])

The soil map (Figure: 2.5) indicates that over 50% of the sub-basin is composed of Eutric Vertisols (VRe), primarily in the central and western parts. The second most prevalent soil type is Vertic Cambisols (CMv), predominantly found in the eastern and southern regions.

The third dominant soil type, Humic Nitisols (NTu), is mainly located in the northern tip of the sub-basin.

According to Yitbarek et al. (2012) [26], the Upper Awash sub-basin is exclusively confined within the north-central plateau and the adjacent escarpment and rift. The geology of the region transitions from lower tertiary basaltic volcanic plateau in the north to interlayered tertiary basaltic and acidic volcanics in the center, and finally to quaternary volcano-sedimentary sequences to the south, near Koka Lake.

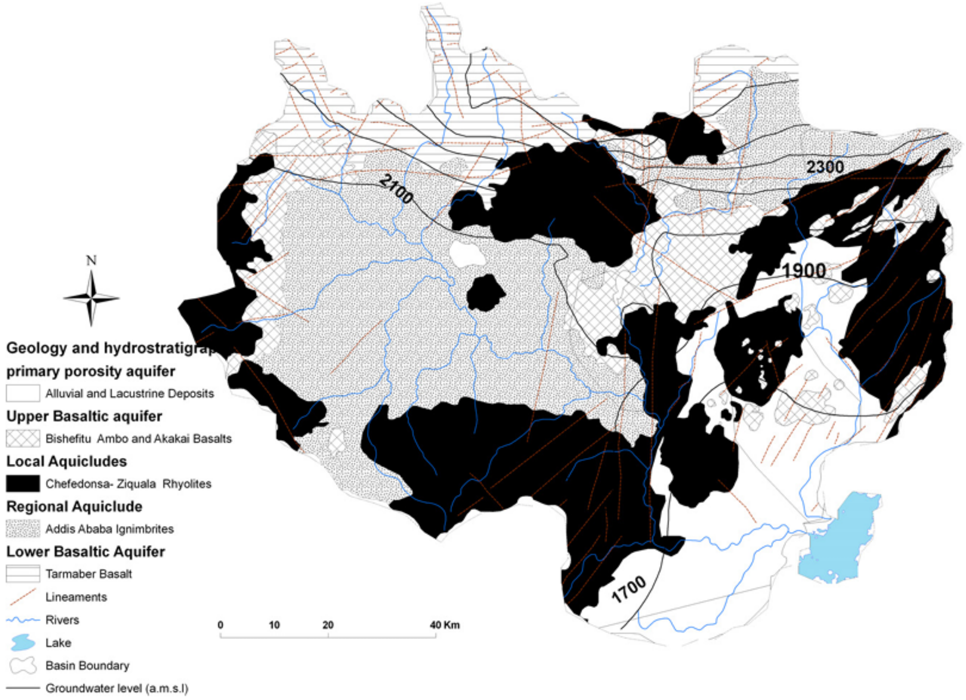


Figure 2.6: Geological and hydrogeological map of Upper Awash sub-basin (Source:Yitbarek et al.(2012))

2.4 Socio – Economic conditions

The sub-basin holds a strategic location as it encompasses the capital city of the country, leading to significant and dynamic economic activity fueled by population growth and migration from various parts of the country. An estimated 6 million people reside in the sub-basin, according to a recent projection in 2017 [27]. It serves as a major industrial hub for the country. The majority of residents are engaged in subsistence agriculture, practicing farming during the two rainy seasons (March-April-May (March April May (MAM)) and June-July-August-September (June July August September (JJAS))).

Chapter 3

Climate Model Selection and Future Climate Over the Sub-basin

3.1 Introduction

Securing water to meet all human needs remains a persistent and challenging priority for countries worldwide. The escalating demand for water is primarily driven by socio-economic growth, evolving consumption patterns, and population expansion [28]. The accessible global freshwater available to humans is less than 1%, categorizing it as a limited resource [29]. Moreover, the natural uneven distribution of rainfall on a global scale adds complexity to ensuring adequate accessibility to this resource.

In recent decades, the impact of climate change on rainfall distribution has been contemplated, contributing to increased variability globally [30]. Anthropogenically induced climate variability manifests as droughts and floods in various regions, as evidenced by recent events like the floods in Mecca, Saudi Arabia, in April 2021 [31], and the drought event in southeastern Alaska, USA, in July 2019 [32].

A study conducted by the National Meteorological Agency (NMA) on climate change adaptation highlights major impacts of climate variability in Ethiopia, including food insecurity, water-borne disease outbreaks, land degradation, and infrastructure damage [33]. The Awash basin, supporting a substantial population with major cities and towns, including the capital city Addis Ababa, is extensively used for water, primarily for irrigation [25]. Most irrigation activities are concentrated in the middle and lower parts of the basin, while intensive rain-fed agriculture dominates the upper basin areas [34, 35]. This upper part of the basin, being mostly highland, is densely populated compared to the middle and lower regions.

The Awash River, the principal drainage for the Awash basin, receives a significant proportion of its average annual flow from the upper basin due to relatively high average annual precipitation in those areas [36]. In recent years, water security issues have become more apparent due to the occurrence of frequent flood and drought events, especially in the middle and lower parts of the basin [37, 38].

A widely used technique for studying the impacts of climate change on water resources is employing outputs from the Coupled Model Intercomparison Project Global Climate Models (Climate Model Intercomparision Project Global Climate Models (CMIP-GCMs)) [39]. CMIP-GCMs are mathematical equations that describe the global climate system with three-dimensional grids, simulating the effects of greenhouse gas emissions on climate. While CMIP-GCMs are often applied at global or regional scales, their coarse resolution requires careful downscaling and bias correction techniques for meaningful insights into local climate conditions, playing a crucial role in decision-making for water resource planning and management [40, 41].

The challenge in predicting future climate conditions goes beyond climate modeling; it extends to effectively utilizing the sheer number of climate model outputs globally. Consequently, models must be evaluated for their performance in simulating the climate characteristics of the intended application area, with the criteria for evaluation depending on the goal of climate model selection [42].

Diverse versions and experiments of climate models generated worldwide make intercomparison between models cumbersome. This led to the establishment of the Climate Model Intercomparison Project (CMIP), consolidating model development efforts worldwide into one framework [43]. The release of new models in the sixth CMIP framework signifies improvements in model setup, resolution, scenarios, and parametrization [44].

The Climate Model Intercomparision Project Six (CMIP6) archive currently houses model output results from over 30 climate research centers worldwide [43]. Each model from these centers differs, resulting in distinct outputs for the same experiment and scenario [45]. Consequently, not all models perform well in a given location, necessitating the evaluation of model capabilities to select those with better performance in describing the local climate [15, 42].

In practice, climate change impact studies often involve the selection of one climate model or a small ensemble of models based on single or multiple criteria. Climate models are typically chosen for their ability to simulate present and near-past climates, with approaches such as the Past–Performance Approach or the Envelop Approach, where an ensemble of models covering a wide range of projections is selected [42, 46].

A clearly defined methodology for evaluating and selecting a single or group of climate models is currently lacking [42]. The selection method and process depend on the study's goal and variables involved [42]. Previous works often employ statistical or data mining techniques, with variations in output (single or ensemble model), analysis period (daily, monthly, seasonal, and annual), and level (station, grid, regional, or spatial averaged).

Common statistical techniques for evaluating climate models include performance indicators, descriptive statistics, and indices [47, 48, 49, 50]. Data mining methods, such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Hierarchical Clustering, Symmetrical Uncertainty (SU), and Canonical Correlation Analysis (CCA), are also applied [46, 51, 52, 53, 15, 54]. The main distinction between these techniques lies in their purpose: data mining is used to detect patterns and relationships, while statistical techniques quantify data [55]. Using a combination of these various techniques, instead of a single method, not only reduces the uncertainty of incorrect selection but also improves it by incorporating various characteristics of time series. Techniques that are often used and aid in incorporating multiple selection methods include Skill Score (SS) and Multi-Criteria Decision Analysis (MCDA) [42, 56].

Previous attempts to understand the climate of the Awash basin, particularly the Upper-Awash basin, were limited. One early work on the Awash basin by Hailemariam [57] used outputs from three randomly selected CMIP3 GCMs (CCCM, GFD3, and GFDL) with two scenarios to predict future runoff conditions. Subsequently, after evaluating the potential of GCMs for their annual cycle, seasonal biases, variability, and trend, three CMIP5 GCMs were selected for studying the impact of climate change by Taye et al. [58]. The impact of climate change on the river basin was further studied with a few selected models from the CMIP5 archive for different scenarios identified from previous works [59]. Similarly, an ensemble of two GCMs [60] and three selected GCMs [61] from the CMIP5 archive was used to characterize river flow in the first case and estimate river nutrient load in the second one. The ensemble mean of five randomly selected GCMs from the CMIP5 archive was used to characterize the hydro-meteorological situation for the Upper-Awash Basin by Emiru et al. [62]. Most recently, similar research, which was published, suggests a group of four climate models from different ensembles of the CMIP6 as an output [63]. Two approaches that combine the envelop and past-performance approach are applied to evaluate outputs of CMIP6 after [42].

Previous climate change projection efforts in the Awash or Upper-Awash Sub-basin focused on utilizing a few randomly selected climate models without thoroughly evaluating their capability to simulate the climate system over the UASB. Therefore, the novelty of this study lies in addressing this shortcoming through the execution of more robust evaluation criteria, primarily relying on the application of statistical and data mining techniques. The main outcome of this study will be identifying model(s) with an overall good performance over the sub-basin. Additionally, predictions of future climate conditions with the newly selected model will be performed later and compared with similar earlier findings in the sub-basin.

3.2 Data Sets

For the evaluation of the climate model, a total of 15 rainfall and 10 temperature stations have been employed. The Table 2.1 provides a detailed list of the utilized stations.

In the initial phase, 12 climate models were identified from the World Climate Research Program (WRCP) CMIP6 archive (https://esgf-node.llnl.gov/search/cmip6/) to be incorporated into the study. The selection was based on the availability of models encompassing all variables (Precipitation, Tmax, and Tmin) for the historical time period. The criteria employed for filtering the models included daily data, a nominal resolution of 100 km, source type: Atmospheric/Ocean General Circulation Model (AOGCM), and the r1i1p1f1 variant. The models were accessed from the aforementioned site on August 17, 2021. The Table 3.1 presents the list of models employed for the analysis.

Table 3.1: List of climate models used in the selection process

Model	Institute	Country
MRI-ESM2-0	Meteorological Research Institute	Japan
ECEARTH3-CC	EC-Earth consortium	Sweden
NorESM2-MM	Norwegian Climate Center	Norway
TaiESM1	Academia Sinica	Taiwan
ECEARTH3.Veg	EC-Earth consortium	Sweden
MPI-ESM1.2.HR	Max Planck Institute for Meteorology	Germany
ECEARTH3	EC-Earth consortium	Sweden
CMCC-ESM2	Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy
GFDL-CM4	NOAA	USA
GFDL-ESM4	NOAA	USA
INM-CM4-8	Institute for Numerical Mathematics	Russia
INM-CM5-0	Institute for Numerical Mathematics	Russia

3.3 Methods

3.3.1 Climate Model Selection

The process commences by extracting GCM outputs for the specified station and time period, facilitated by Climate Data Operator (CDO) (Climate Data Operator) [64]. The GCM outputs are applied to each climate station through the Inverse Distance Weighting (Inverse Distance Weighted (IDW)) interpolation technique under CDO. The comparison between GCM outputs and observed data relies on spatially averaged values for the sub-basin. The Thiessen polygon technique is employed for spatially averaging station point data. Additionally, all historical climate series undergo bias adjustment using the corresponding observed series at each station.

The methodology employed here aims to highlight crucial characteristics of climate time series, crucial for climate impact studies. Drawing from an extensive literature review on current methodologies and aligning with the paper's objectives, a combination of methodologies is utilized. The selection and evaluation are based on historical/observed data and encompass four-time scales: monthly, seasonal (JJAS and MAM), and annual. The choice of these two seasons is driven by their significance as crucial rainy seasons, determining annual water availability in the sub-basin. It is assumed that an analysis at these time scales will capture all statistical characteristics of the time series.

In total, a combination of six methodologies is applied to accentuate diverse characteristics of the time series. The first technique seeks to fit a probability distribution to all climatic variables and assesses which climatic models exhibit a similar distribution to the observed climate variable. This is executed using the gamlss package [65] within the R programming software [66]. The outcome of the distribution analysis is evaluated between the observed and climate model series using (Akaike Information Criterion (AIC)). Subsequently, the climate model or models demonstrating a similar distribution are presumed to better describe the observed series.

The second technique examines trends using the Mann-Kendall (MK) test [67, 68]. In detecting trends, the MK test calculates the statistic 'S' by ranking the data and determining the sign, as described in Equation (3.1) [67].

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{3.1}$$

Where: x_i - is data ranked from i = 1, 2, 3, . . . , n-1 and x_j- is the data ranked from j = i+1, . . . , n.

The sign is the difference between the original data x_i, shortened by one data point, and the data itself without the last data point x_j.

$$sgn(x_j - x_i) = \begin{cases} +1 \; if(x_j - x_i) > 0 \\ 0 \; if(x_j - x_i) = 0 \\ -1 \; if(x_j - x_i) < 0 \end{cases} \tag{3.2}$$

It is observed that when the data point (observation) is more (n $\geq$ 10), the statistic 'S' becomes normally distributed with mean (E(S)) equal to zero and variance calculated as follows:

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{t=1}^{m} t_1(t_1-1)(2t_1+5)}{18} \tag{3.3}$$

Where: n - is the number of data points and t_i are the ties of the sample data series. The test statistic (Z_c) is calculated as:

$$Z_c = \begin{cases} \frac{S-1}{\sigma} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \frac{S+1}{\sigma} \; if \; S < 0 \end{cases} \tag{3.4}$$

Where: σ is standard deviation of the statistic 'S'. The test statistic (Zc) is derived from this analysis, where a positive value signifies an upward trend, while a negative value indicates the opposite. As suggested by [68], this approach is effective for data lacking significant correlation at lag 1. If there is significant correlation at lag 1, the modified MK test is implemented. This modification involves calculating the significance of the trend by adjusting the variance of the MK test statistic ('S') with ESS (Effective Sample Size). Significance is determined based on the p-values at a 0.05 significance level. Thus, if the p-values exceed 0.05, the null hypothesis of no significant trend in the data is accepted; conversely, if the p-values are less than or equal, the opposite holds true.

Furthermore, the results of the MK test can be complemented by the Sen's slope method [69], which computes the magnitude of the trend (slope (T_i)), as illustrated in Equation (3.5) below:

$$T_i = \frac{X_j - X_i}{j - i} \tag{3.5}$$

The median of N values of T_i represents the Sen's slope estimator which is calculated as follows;

If N appears to be odd:

$$Q_{med} = T\frac{(N+1)}{2} \tag{3.6}$$

If N is even then:

$$Q_{med} = \frac{(T_{\frac{N}{2}} + T_{\frac{(N+2)}{2}})}{2} \tag{3.7}$$

Here, a positive Q_{med} value in the time series indicates an upward trend and a negative value the reverse.

In the third approach, four performance metrics are employed to discern which climate model(s) better simulate the observed series. The chosen metrics include the Coefficient of Determination (R^2), (Root Mean Square Error (RMSE)), Mean Absolute Error (MAE), and BIAS. Each performance measure is applied across all time steps, and a rank is assigned to each model based on the magnitude of the respective performance measure. Models demonstrating overall good performance at all four time steps are identified by summing up the ranks at the corresponding time step and subsequently re-ranking.

Another technique applies a time series clustering approach to identify which climate model more effectively captures the stochastic process of the observed series. The Integrated Periodogram algorithm, developed by [70], is implemented using the TSclust package [71] within R. This data mining approach seeks to group time series into clusters based on their Integrated Periodogram (d_{IP}) as a distance (dissimilarity) measure. The periodogram technique facilitates a comparison in the frequency domain, enabling the characterization of each time series in terms of its underlying stochastic behavior. The dissimilarity measure is calculated using Equation (3.8) [70]:

$$d_{IP}(X_T, Y_T) = \int_{-\pi}^{\pi} |F_{X_T}(\lambda) - F_{Y_T}(\lambda)| \, d\lambda \tag{3.8}$$

Where the normalized cumulative periodograms at each data point j are given by:

$$F_{X_T}(\lambda_j) = \frac{1}{C_{X_T}} \sum_{i=1}^{j} I_{X_T}(\lambda_i) \tag{3.9}$$

$$F_{Y_T}(\lambda_j) = \frac{1}{C_{Y_T}} \sum_{i=1}^{j} I_{Y_T}(\lambda_i) \tag{3.10}$$

The weights used for normalizing the periodograms are as shown below where m is the number of data points in the original data series:

$$C_{X_T} = \sum_{i=1}^{m} I_{X_T}(\lambda_i) \tag{3.11}$$

$$C_{Y_T} = \sum_{i=1}^{m} I_{Y_T}(\lambda_i) \tag{3.12}$$

The periodograms calculated at each time stamp (k) of the observed series is given by:

$$I_{X_T}(\lambda_k) = \frac{1}{T} |\sum_{t=1}^{T} X_t e^{-i\lambda_k t}|^2 \tag{3.13}$$

$$I_{Y_T}(\lambda_k) = \frac{1}{T}|\sum_{t=1}^{T} Y_t e^{-i\lambda_k t}|^2 \tag{3.14}$$

Here, $\lambda_k = \frac{2\pi k}{T}$ represents the frequency component corresponding to the input data sequence k ($k = 1, 2, ..., n$), where $n = \frac{T-1}{2}$ depends on the total data length of the observed series (T). Additionally, X_t and Y_t correspond to each pair of time series for which the dissimilarity measure is to be calculated. The term i in Equations (3.13) and (3.14) signifies the imaginary term resulting from the Fourier transformation of each pair of series.

A pairwise matrix of dissimilarity measures is generated through the aforementioned procedure, and clustering is executed using the Agglomerative Hierarchical Clustering technique under the Tsclust package in R, which utilizes the hclust() function from the **stats** package. The classification into a cluster group or merging between clusters follows the complete linkage criteria [70].

The fifth technique employed is (Principal Component Analysis (PCA)), a dimensionality reduction technique that transforms a large number of correlated variables into a much smaller set of uncorrelated variables known as Principal Components (PCs) [72]. In this context, the 13 variables (1 observed + 12 Climate models) are represented in two-dimensional axes called principal components (PCs). The original data points for each variable can be plotted in a two-dimensional space, referred to as a score plot. This is achieved by projecting all data points onto those two PCs using the loading vectors derived from the covariance matrix. The score plot indicates that points closer to the origin are closer to the average, points near each other are similar, and points further outward are considered outliers. The correlation between each variable is better visualized using a loading plot. This plot is created by plotting the eigenvectors and indicates the contribution of each loading to the PCs. The relative length of the vector indicates its contribution to each PC, and the angle between vectors signifies the similarity between variables. In essence, if the angle between two adjacent loading vectors of variables is smaller, they are more correlated; if orthogonal, they are not related; and if in the reverse direction, they are negatively correlated. A combined plot of the score and loading plot is known as a Bi-Plot [73].

Finally, the spatial performance of the models is evaluated by conducting correlation analysis and RMSE as model fit criteria. These two techniques are applied at the station level and later interpreted using a rank-based methodology.

3.3.2 Downscaling, Bias adjustment and Future Scenarios

The outputs of GCMs need to undergo downscaling and bias adjustment before application in real-world scenarios. This is primarily because climate model outputs carry biases resulting from imperfect conceptualization and parametrization, limited data record lengths, quality issues in reference datasets, and insufficient spatial resolution [74, 75]. Consequently, downsizing of GCM outputs is often carried out using either Dynamical or Statistical downscaling approaches. Dynamic downscaling produces Regional Climate Models (RCMs) at a finer scale ($< 50km$) by utilizing GCM boundary conditions, but it is computationally expensive [76]. More frequently, Statistical downscaling is preferred over dynamic downscaling due to its avoidance of computational drawbacks and its capability for bias adjustment. In this context, bias adjustment is achieved by establishing a statistical link between large-scale predictors and a finer-scale predictand.

Moreover, Statistical downscaling approaches can be categorized into two types: Perfect Prognosis (PP) and Model Output Statistics (MOS) [77]. PP methods create day-to-day statistical relationships between large-scale predictors and local-scale predictands during the calibration phase. The calibrated statistical relationship is then employed for downsizing a predictor variable. Examples of PP methods include Regression, Weather Generators, and the Analog Method. In MOS, there is no day-to-day correspondence; instead, a statistical transfer function is applied during the calibration phase and later utilized for downsizing operations [78]. Techniques under MOS include Additive Correction/Scaling, Variance Correction, and Quantile Mapping.

Here both the precipitation and temperature stations are first downscaled to each ground stations shown in Table:2.1 using the IDW technique Equation:3.15 and later bias corrected using Quantile Mapping (QM) approach.

$$V_{St.} = \frac{\left(\frac{V_{GP_1}}{D_1} + \frac{V_{GP_2}}{D_2} + \frac{V_{GP_3}}{D_3} + \frac{V_{GP_4}}{D_4}\right)}{\sum_{i=1}^{4} \frac{1}{D_i}} \tag{3.15}$$

Where: $V_{St.}$ is the variable at station level, V_{GP_i} is the variable at grid point level, and D_i are the distance from the grid point to the specific station.

The QM method, which utilizes empirical Cumulative Density Function (CDF) derived from actual observations and does not necessitate assumptions about the underlying distribution, has been employed in this study for bias adjustment of the climate model outputs [79]. All bias adjustments in this study are conducted in R using the qmap package [75] on the daily climate series. The theoretical assumption of the QM method for bias adjustment is illustrated in Equation (3.16) as follows

$$x_{f,adj} = F_o^{-1}(F_f(x_f)) \tag{3.16}$$

Where: $x_{f,adj}$ is value of bias adjusted future climate variable, $F_o(x)$ is Emperical Cumulative Density Function (ECDF) of Observed variable, F_f^{-1} is inverse ECDF of future GCM, and x_f is the value of the future GCM variable before bias adjustment.

Three future scenarios (Shared Socioeconomic Pathway (SSP)1-2.6, SSP2-4.5, and SSP5-8.5) and three time periods—Near-century (2022-2039), Mid-century (2040-2069), and End of the century (2070-2099)—are employed to comprehend the future climate of the Upper Awash Basin. Each scenario corresponds to sustainable, middle, and worst-case future socioeconomic conditions. The worst-case scenario represents the upper boundary within the available range of scenarios and can be considered as an update of RCP8.5 of CMIP5. The second scenario, SSP2-4.5, signifies the medium pathway for future increases in greenhouse gas emissions, akin to RCP4.5 of CMIP5.

3.4 Results

3.4.1 Selection of Climate Model

All the methods employed are chosen to underscore crucial characteristics of a climate time series that could play a critical role in addressing the future climate impact in the study area. The evaluation criteria encompass distribution fitting, trend analysis, performance measures, DTW-based hierarchical clustering, PCA analysis, spatial correlation, and RMSE.

Initially, a total of twelve climate models, as presented in Table 3.1, were identified based on the common availability of climate models for the three climatic variables (Precipitation, Maximum temperature, and Minimum temperature). These models undergo evaluation for their statistical characteristics using seven different criteria and across different time levels (monthly, monthly average, seasonal, and annual). Monthly series are obtained by summing/averaging precipitation/temperature daily series for each month. The monthly average is computed by taking those monthly values and further averaging them across the total analysis period for each month.

Identification of Distribution

The best possible distributions for the observed and climate model series have been identified using the gamlss package in R [65]. For instance, for the precipitation data of monthly averages, JJAS, MAM, and annual series, the observed data were found to follow Normal (NO), Logistic (LO), Gamma (GA), and Normal (NO) distributions, respectively. Definitions for each distribution type used in fitting the three climate variables are provided in Appendix A. The Probability Density Function (Probability Density Function (PDF)) and Cumulative Distribution Function (CDF) plots for only monthly average precipitation series of both observed and climate models are presented in Figure 3.1 below.

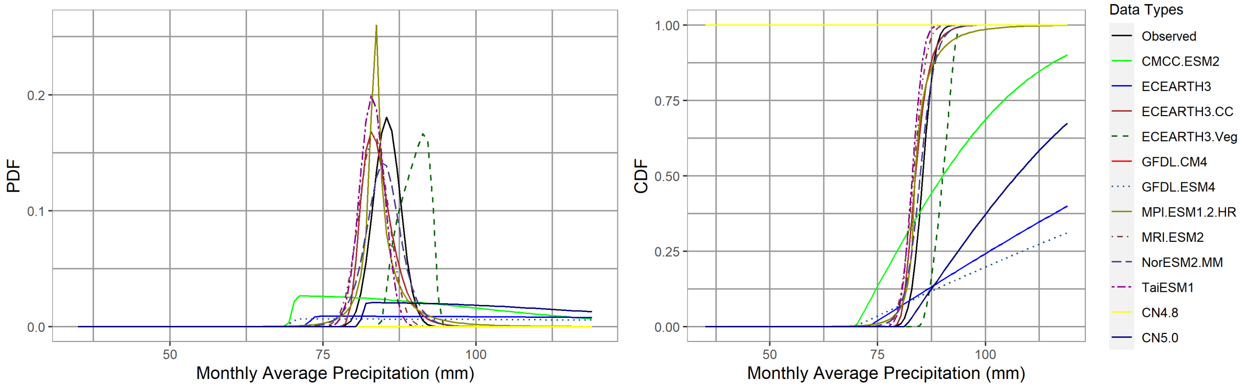


Figure 3.1: Fitted distribution for the monthly average precipitation (black line) series of the observed and 12 climate models (indicated with different colors).

The results of the distribution analysis for all three climatic variables and different time steps of analysis for both observed and climate model series are summarized in Table 3.2. For the precipitation series, it can be deduced that the MRI-ESM2-0 model exhibited similar distributions to the observed series at all time steps. TaiESM1 showed a similar distribution at the Monthly average and Annual time steps. Meanwhile, NorESM2-MM demonstrated a similar distribution only at seasonal time steps, and ECEARTH3-CC only at MAM.

In the case of maximum temperature series, none of the climate models consistently exhibited similarity across different time steps, as shown in Table 3.2. Three models had similar distribution types at the monthly average time step. The GFDL-CM4 model had the same distribution as the observed series twice across those four time steps, and none of the models showed similar distributions at the annual level. For the minimum temperature series, a few models shared a similar distribution type with the observed series. CMCC-ESM2 showed a similar distribution at the monthly average, MPI-ESM1-2-HR, and GFDL-ESM4 at MAM time steps, respectively. No models exhibited similar distributions at the annual time step, as shown in Table 3.2 below.

Table 3.2: Fitted Distributions for all three climatic variables across the four time steps of both observed and climate models series.

	Monthly Average		JJAS		MAM		Annual	
Data Type	AIC	DT	AIC	DT	AIC	DT	AIC	DT
Precipitation								
Observed	**221.62**	**NO**	**359.84**	**LO**	**348.147**	**GA**	**370.72**	**NO**
CMCC-ESM2	210.09	SN2	229.72	SN2	272.3	SEP1	359.19	SN2
ECEARTH3	213.9	SN2	268.16	WEI	277.63	WEI3	362.99	SN2
ECEARTH3-CC	229.69	RG	257.03	GA	285.2	GA	378.79	RG
ECEARTH3-Veg	228.64	SEP1	249.17	SEP3	274.55	RG	381.82	IGAMMA
GFDL-CM4	196.06	WEI3	254.98	WEI	266.21	NO	345.16	WEI
GFDL-ESM4	214.78	SEP3	233.74	SEP2	259.64	NO	364.56	IGAMMA
MPI-ESM1-2-HR	217.1	SEP1	263.67	WEI3	275.74	SEP1	366.19	SEP1
MRI-ESM2	232.6	NO	261.15	LO	271.72	GA	381.7	NO
NorESM2-MM	229.68	LO	277.26	LO	295.01	GA	378.77	LO
TaiESM1	209.12	NO	237.44	IGAMMA	272.38	LO	358.22	NO
INM-CM4-8	233.88	WEI	281.42	SEP2	238.5	WEI3	382.97	WEI3
INM-CM5-0	226.62	SN2	280.22	NO	261.28	WEI2	377.17	SN2
Maximum Temperature								
Observed	**45.97**	**SN2**	**28.93**	**RG**	**51.31**	**WEI**	**13.19**	**PE2**
CMCC-ESM2	32.12	WEI3	52.57	SN2	37.6	LO	32.12	WEI3
ECEARTH3	30.55	NO	44.69	WEI3	53.16	GT	30.55	NO
ECEARTH3-CC	30.06	SN2	39.55	GG	42.85	SN2	30.06	SN2
ECEARTH3-Veg	26.53	NO	42.93	SN2	45.23	NO	26.53	NO
GFDL-CM4	8.59	SN2	37.81	RG	45.1	IGAMMA	8.59	SN2
GFDL-ESM4	27.57	SN2	42.36	WEI3	49.34	LO	27.57	SN2
MPI-ESM1-2-HR	14.35	NO	34.3	SN2	57.13	NO	14.35	NO
MRI-ESM2	10.66	NO	42.33	WEI	42.31	PE2	10.66	NO
NorESM2-MM	37.52	SHASH	58.02	SN2	55.42	WEI3	37.52	SHASH
TaiESM1	2.78	WEI3	31.94	WEI	25.06	NET	2.78	WEI3
INM-CM4-8	14.07	WEI3	14.57	IGAMMA	31.78	PE	14.07	WEI3
INM-CM5-0	11.02	WEI	27.42	LO	35.8	NET	11.02	WEI
Minimum Temperature								
Observed	**42.35**	**SEP1**	**30.9**	**NO**	**34.09**	**NO**	**16.98**	**IGAMMA**
CMCC-ESM2	33.67	SEP1	24.05	SN2	45.87	GA	33.67	SEP1
ECEARTH3	44.68	WEI3	47.64	WEI	23.2	NET	44.68	WEI3
ECEARTH3-CC	40.14	SN2	52.83	WEI	31.7	GU	40.14	SN2
ECEARTH3-Veg	45.88	WEI3	16.64	BCPEo	28.47	PE	45.88	WEI3
GFDL-CM4	3.73	IGAMMA	-6.15	RG	28.03	IGAMMA	3.73	IGAMMA
GFDL-ESM4	31	WEI	15.08	WEI3	38.77	NO	31	WEI
MPI-ESM1-2-HR	21.35	IGAMMA	10.89	SEP3	3.36	NO	21.35	IGAMMA
MRI-ESM2	29.13	RG	9.45	GT	30.08	RG	29.13	RG
NorESM2-MM	50	SHASH	37.67	SEP1	45.7	WEI	50	SHASH
TaiESM1	16.68	NET	12.72	LO	34.01	LO	16.68	NET
INM-CM4-8	16.43	WEI	-6.83	WEI	33.46	SN2	16.43	WEI
INM-CM5-0	12.78	SEP3	-7.68	SN2	40	IG	12.78	SEP3

NB: Those highlighted in grey are models that have shared the same distribution as observed time series across each time period of analysis. DT = Distribution Type.

Trend Analysis

Trend analysis has also been performed similarly at four different levels, as in the distribution analysis. The precipitation series indicates that almost all of the models had similar trends to the observed series across all time steps. Exceptions include ECEARTH3 and ECEARTH3-CC at JJAS and INM-CM4-8 at monthly average and annual time levels, as shown in Table 3.3. Although not exhibiting a significant trend, Sen slope values for the TaiESM1 model indicate a similar trend direction as the observed series but with a different magnitude.

For the maximum temperature series, ECEARTH3-CC does not have a similar trend to the observed series at any of the time steps. Four models (CMCC-ESM2, NorESM2-MM, TaiESM1, and INM-CM5-0) were able to have a similar trend at three-time steps out of four. GFDL-ESM4 is the only model that exhibited similar trend characteristics as the observed series at all time steps, as shown in Table 3.3.

Furthermore, among the ten models, CMCC-ESM2, ECEARTH3-CC, GFDL-CM4, and TaiESM1 exhibited similar trends to the observed minimum temperature series across all time steps, whereas models ECEARTH3, GFDL-ESM4, MRI-ESM2-0, and INM-CM5-0 exhibited the least similarity. Two models, MPI-ESM1-2-HR and NorESM2-MM, showed similarity at all time steps except for the JJAS season, as shown in Table 3.3.

Table 3.3: Results of Trend Analysis for all three climatic variables across the four time steps for both observed and climate models series.

Data Type	Annual		Monthly Ave.		JJAS		MAM	
	Z-val	Sen S.	Z-val	Sen S.	Z-val	Sen S.	Z-val	Sen S.
Precipitation								
Observed	**-0.43**	**-0.91**	**-0.43**	**-0.08**	**-0.07**	**-0.06**	**0.07**	**0.05**
CMCC-ESM2	0.04	0.09	0.04	0.01	-1.39	-0.32	1.82	0.89
ECEARTH3	1.39	3.62	1.39	0.30	2.64	1.06	-0.21	-0.05
ECEARTH3-CC	1.78	5.61	1.78	0.47	3.14	1.00	-0.39	-0.29
ECEARTH3-Veg	0.39	1.51	0.39	0.13	0.64	0.21	-0.36	-0.27
GFDL-CM4	1.68	3.14	1.68	0.26	1.04	0.33	-0.18	-0.05
GFDL-ESM4	-0.25	-1.02	-0.25	-0.09	-1.43	-0.25	-0.86	-0.45
MPI-ESM1-2-HR	0.21	0.55	0.21	0.05	-0.25	-0.08	0.61	0.17
MRI-ESM2	1.57	4.52	1.57	0.38	0.5	0.25	0.46	0.31
NorESM2-MM	1.21	3.29	1.21	0.27	0.14	0.1	0.14	0.08
TaiESM1	-0.86	-1.94	-0.86	-0.16	-1.61	-0.5	0.64	0.24
INM-CM4-8	2.87	4.25	2.87	0.35	1.32	1.04	1.04	0.29
INM-CM5-0	0.71	2.71	0.71	0.23	0.00	-0.01	1.14	0.52
Maximum Temperature								
Observed	**7.36**	**0.03**	**7.36**	**0.03**	**1.27**	**0.02**	**2.38**	**0.05**
CMCC-ESM2	2.13	0.02	2.13	0.02	1.00	0.03	0.34	0.01
ECEARTH3	4.32	0.04	4.32	0.04	2.54	0.05	1.8	0.05
ECEARTH3-CC	1.29	0.01	1.29	0.01	3.23	0.03	1.00	0.02
ECEARTH3-Veg	3.4	0.03	3.4	0.03	2.97	0.04	0.37	0.01
GFDL-CM4	1.32	0.02	1.32	0.02	0.4	0.01	0.79	0.02
GFDL-ESM4	2.48	0.02	2.48	0.02	0.32	0.01	2.38	0.05
MPI-ESM1-2-HR	0.08	0.00	0.08	0.00	0.26	0.01	0.00	0.00
MRI-ESM2	0.00	0.00	0.00	0.00	0.69	0.02	-1.47	-0.01
NorESM2-MM	1.99	0.02	1.99	0.02	0.26	0.01	0.63	0.01
TaiESM1	6.63	0.02	6.63	0.02	1.95	0.03	1.69	0.02
INM-CM4-8	2.48	0.03	2.48	0.03	2.85	0.04	1.74	0.02
INM-CM5-0	2.99	0.02	2.99	0.02	1.56	0.02	1.48	0.02
Minimum Temperature								
Observed	**1.8**	**0.02**	**1.8**	**0.02**	**2.32**	**0.03**	**1.11**	**0.02**
CMCC-ESM2	1.57	0.02	1.57	0.02	2.37	0.02	0.64	0.01
ECEARTH3	2.8	0.06	2.8	0.06	3.38	0.07	3.01	0.03
ECEARTH3-CC	1.41	0.03	1.41	0.03	4.02	0.05	1.69	0.03
ECEARTH3-Veg	4.02	0.04	4.02	0.04	2.64	0.05	1.89	0.01
GFDL-CM4	1.22	0.01	1.22	0.01	2.48	0.02	-0.79	-0.01
GFDL-ESM4	4.76	0.04	4.76	0.04	4.13	0.02	2.46	0.04
MPI-ESM1-2-HR	1.00	0.01	1.00	0.01	0.79	0.01	-0.87	-0.01
MRI-ESM2	5.9	0.04	5.9	0.04	5.58	0.04	2.09	0.03
NorESM2-MM	0.87	0.01	0.87	0.01	1.76	0.02	0.48	0.02
TaiESM1	0.79	0.01	0.79	0.01	6.24	0.02	1.69	0.02
INM-CM4-8	8.95	0.02	8.95	0.02	6.83	0.02	1.06	0.01
INM-CM5-0	2.38	0.02	2.38	0.02	3.73	0.01	2.01	0.03

NB: The ones shaded in gray show that the models having similar trend test outcome. i.e if both observed and model are significant/insignificant then the z-values will be highlighted.

Performance Metrics

Four different performance indicators were implemented to identify models with good prediction capabilities across all time steps. Table 3.4 shows the summarized results for each model across these four time steps against four performance measures. The ranking in the summary table for each performance measure was obtained by ranking the sum of ranks across each time step.

In the precipitation series, ECEARTH3-CC is the model that exhibited relatively poor overall performance in predicting the observed series, while the other two ECEARTH3 models are top-performing. For the case of maximum temperature series, the NorESM2-MM model performed poorly, and TaiESM1 is the top-ranked model in terms of simulating the observed series, as shown in the summary Table 3.4. From Table 3.4, it can also be observed that GFDL-CM4 and ECEARTH3-CC are models that performed well and poorly, respectively, in simulating the observed minimum temperature series.

Table 3.4: Summary of model performances for all three climatic variables across four performance measures for both observed and climate models series.

	Ranks				
Data Type	R^2	RMSE	MAE	BIAS	Sum of Rank
Precipitation					
ECEARTH3	6	2	2	4	14
ECEARTH3-Veg	2	6	1	10	19
GFDL-ESM4	3	1	8	12	24
CMCC-ESM2	10	5	9	1	25
INM-CM4-8	1	10	7	8	26
NorESM2-MM	5	11	6	5	27
MPI-ESM1-2-HR	4	8	5	11	28
MRI-ESM2	9	9	3	7	28
TaiESM1	7	3	10	9	29
GFDL-CM4	8	4	12	6	30
INM-CM5.0	12	12	4	2	30
ECEARTH3-CC	11	7	11	3	32
Maximum Temperature					
TaiESM1	6	1	1	3	11
INM-CM5-0	4	2	2	6	14
ECEARTH3	1	4	9	1	15
INM-CM4-8	7	3	3	4	17
ECEARTH3-CC	2	7	4	10	23
MPI-ESM1-2-HR	8	6	5	5	24
GFDL-ESM4	3	5	6	12	26
ECEARTH3-Veg	9	8	8	2	27
GFDL-CM4	10	9	7	7	33
MRI-ESM2	5	12	11	11	39
CMCC-ESM2	12	10	10	8	40
NorESM2-MM	11	11	12	9	43
Minimum Temperature					
GFDL-CM4	4	1	1	1	7
MPI-ESM1-2-HR	1	2	3	11	17
TaiESM1	3	4	4	6	17
INM-CM4-8	5	3	2	12	22
ECEARTH3-Veg	2	8	8	7	25
INM-CM5-0	10	6	7	4	27
MRI-ESM2	12	7	5	3	27
GFDL-ESM4	9	5	6	10	30
CMCC-ESM2	6	9	9	8	32
NorESM2-MM	8	12	12	2	34
ECEARTH3	7	11	10	9	37
ECEARTH3-CC	11	10	11	5	37

Time Series Clustering

A time-series clustering technique is employed to identify which climate model better captures the observed series in terms of seasonal variation. This technique utilizes a distance measure calculated from the cumulative periodograms of the observed and climate models. Using the distance matrix obtained from the integrated periodograms, a complete-link agglomerative hierarchical clustering was performed on the monthly precipitation, maximum, and minimum temperature series.

From Figure 3.2, it can be observed that for the precipitation series, those twelve models were clustered into four groups. As expected, models from the same modeling institution, such as ECEARTH3 and GFDL families, are clustered together. It can be seen that two models, INM-CM4-8 and MPI-ESM1-2-HR, were clustered with the observed series, indicating their proficiency in simulating seasonal characteristics. The next in line in terms of having closer seasonal characteristics to the observed series are INM-CM5-0 and CMCC-ESM2. There is a clear subgrouping with the rest of the models, indicating less potential in simulating the seasonal behavior of the observed series.

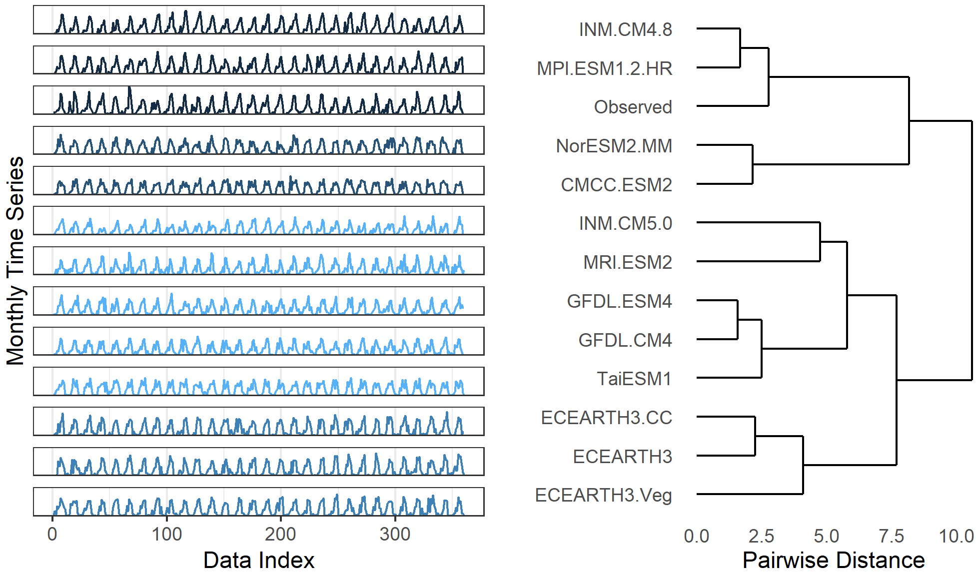


Figure 3.2: Clusters for monthly precipitation series for both observed and climate models series (Left figure: shows the cluster separation with blue color contrast, Right figure: Clustering of climate models and observed data in terms of time series similarity based on seasonal pattern).

Similarly, for the monthly maximum temperature series, a clear clustering between models can be observed. The same models seen in the precipitation series exhibit similar seasonal characteristics, as shown in Figure 3.3.

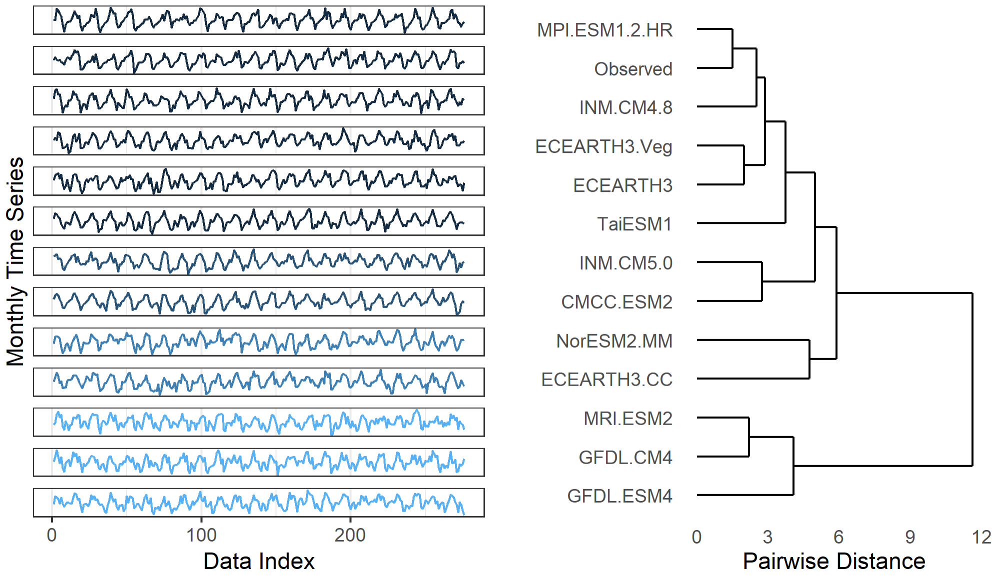


Figure 3.3: Clusters for monthly maximum temperature series for both observed and climate models series (Left figure: shows the cluster separation with blue color contrast, Right figure: Clustering of climate models and observed data in terms of time series similarity based on seasonal pattern).

The clustering pattern for the monthly minimum temperature series differs from the two previous cases, as illustrated in Figure 3.4. In this instance, MRI-ESM2-0 and MPI-ESM1-2-HR have demonstrated proficiency in capturing seasonal variation. Furthermore, the dissimilarity between the observed series and the ECEARTH3 family, as well as NorESM2-MM, is relatively higher compared to the rest of the climate models.

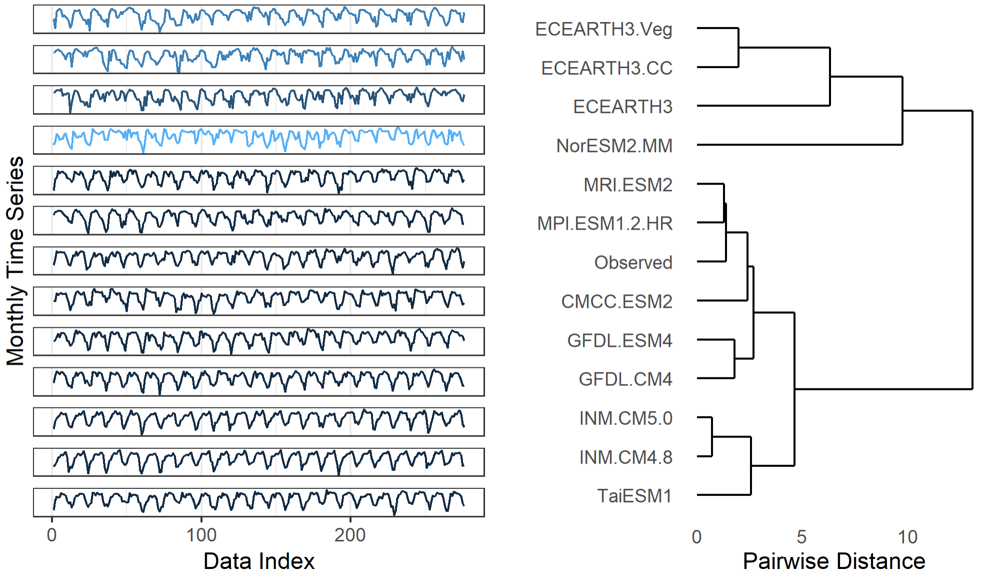


Figure 3.4: Clusters for monthly minimum temperature series for both observed and climate models series (Left figure: shows the cluster separation with blue color contrast, Right figure: Clustering of climate models and observed data in terms of time series similarity based on seasonal pattern).

Principal Component Analysis (PCA)

In the PCA analysis of monthly precipitation data, most climate models align along PC1, with exceptions such as TaiESM1, CMCC-ESM2, and MRI-ESM2. PC1 explains nearly all the variance for models ECEARTH3-Veg and ECAERTH3-CC. The models MPI-ESM1-2-HR and the ECEARTH3 family form a closely correlated group. As depicted in Figure 3.5, the observed series is strongly correlated with MRI-ESM2.

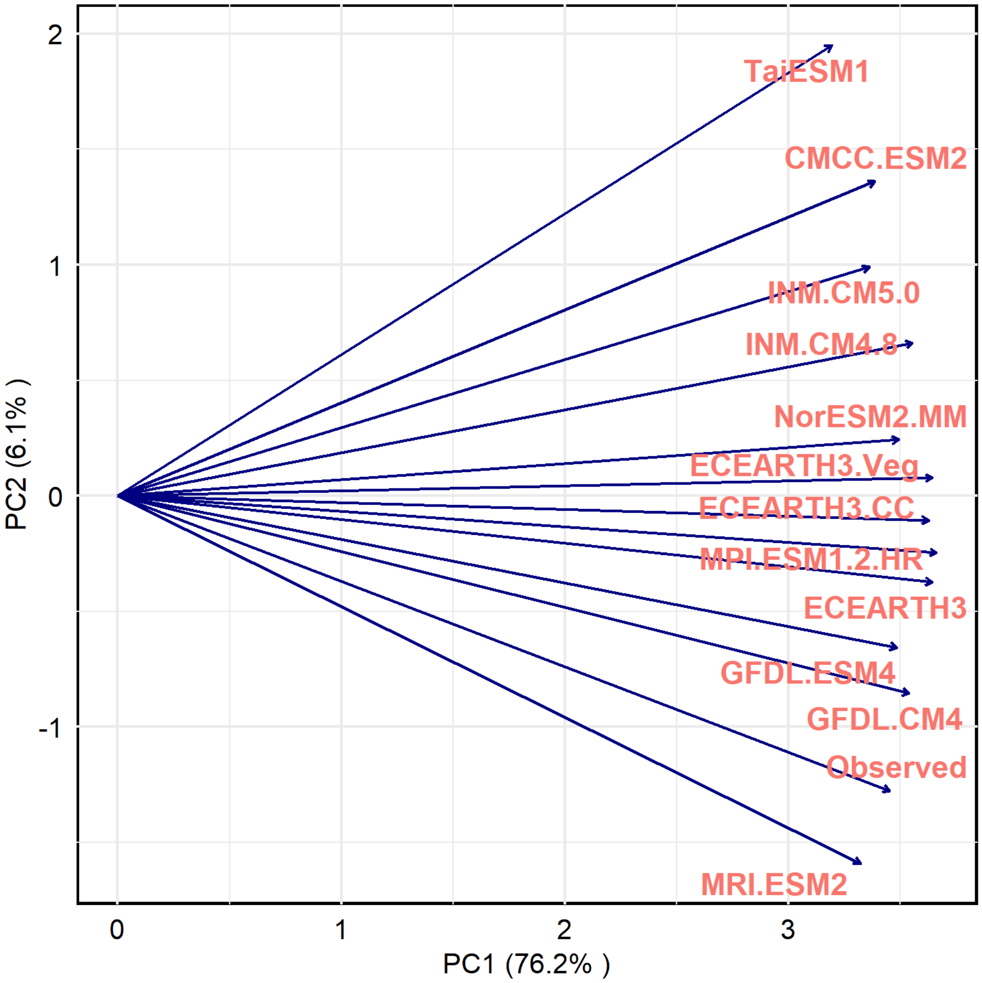


Figure 3.5: PCA Bi-plot for monthly precipitation series of both observed and climate model series.

All the ECEARTH3 families and GFDL-CM4 exhibit a strong correlation with the observed maximum temperature series, as illustrated in Figure 3.6. Conversely, MRI-ESM2-0 and CMCC-ESM2 are the least correlated models in this regard.

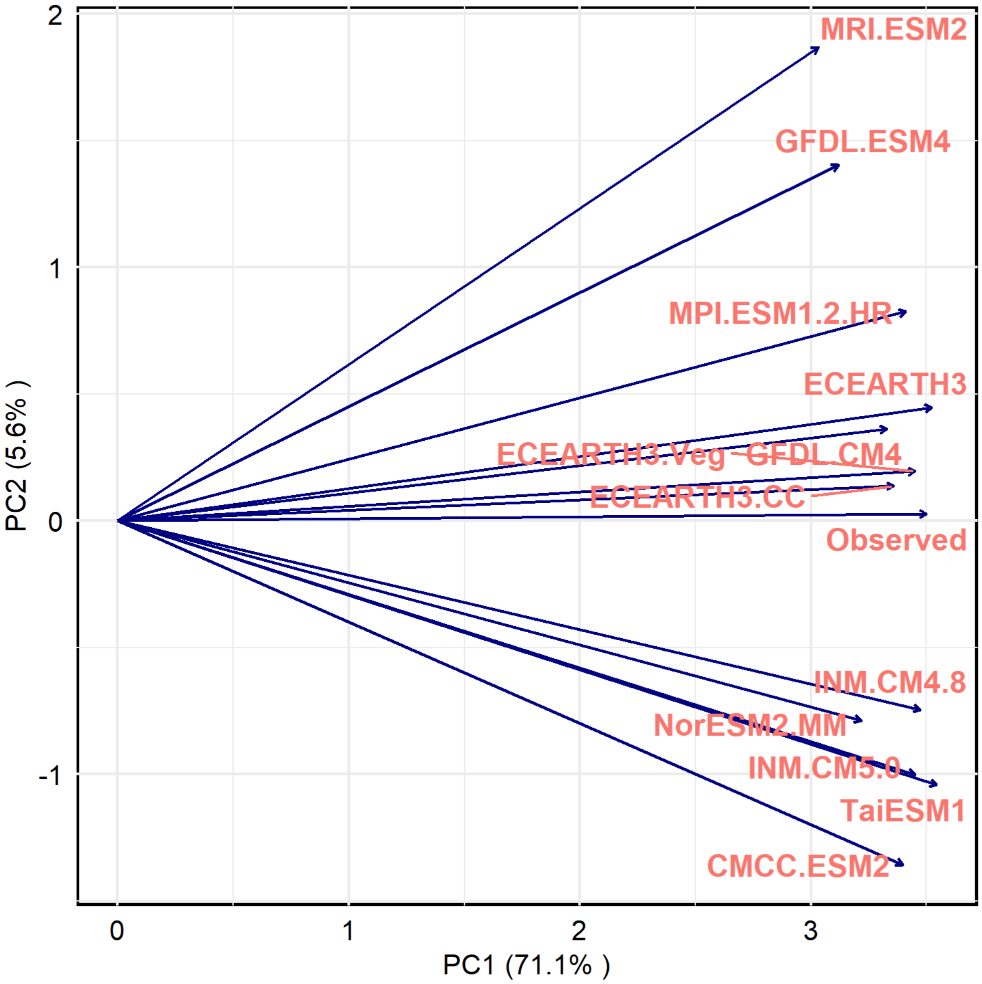


Figure 3.6: PCA Bi-plot for monthly maximum temperature series of both observed and climate model series.

Contrary to the monthly maximum series, the ECEARTH3 families exhibit poor correlation in the monthly minimum series. The most highly correlated model is GFDL-ESM4, and the second-best are groups of models that include GFDL-CM4, NorESM2-MM, and MRI-ESM2-0, as shown in Figure 3.7.

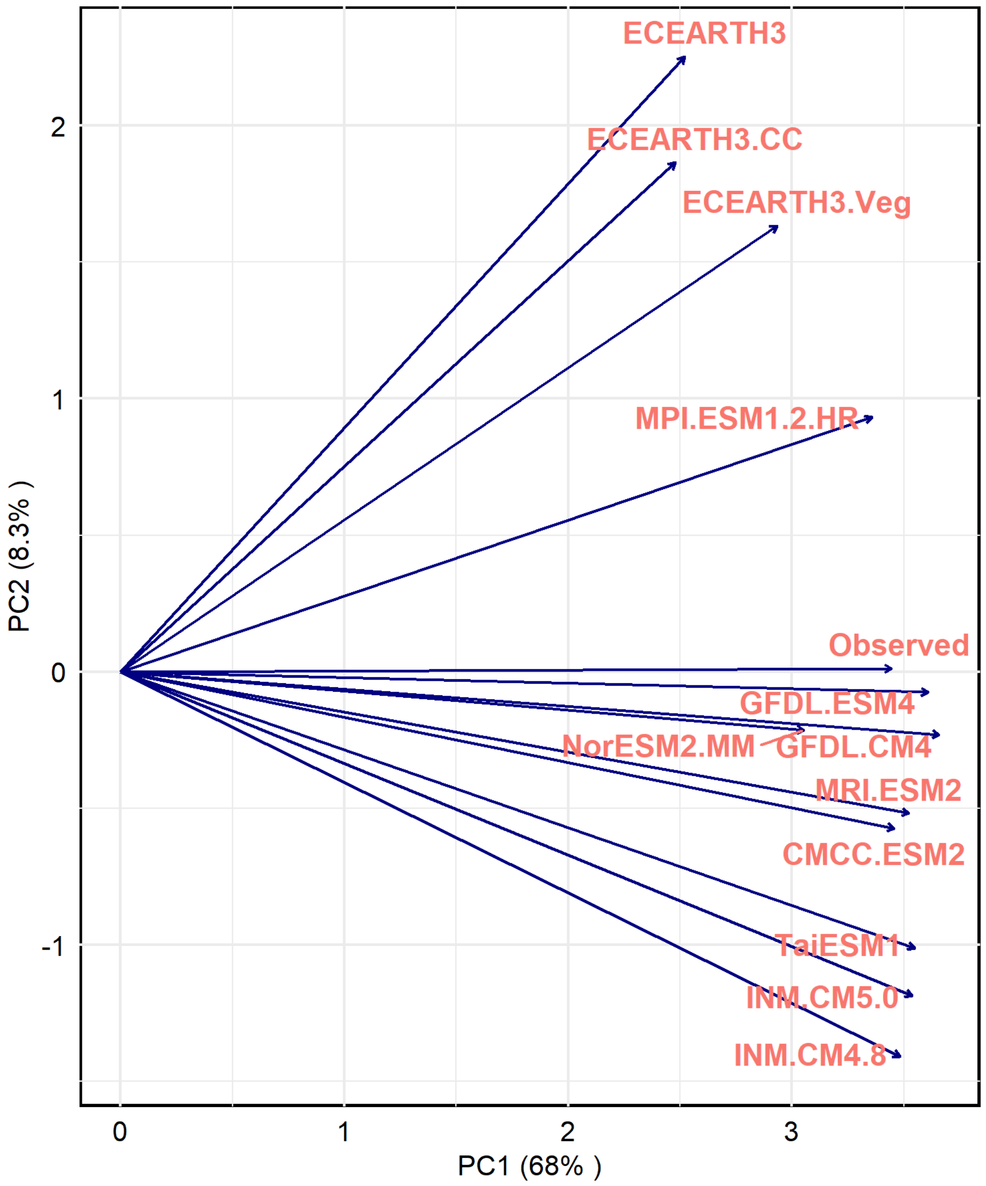


Figure 3.7: PCA Bi-plot for monthly minimum temperature series of both observed and climate model series.

Spatial Performance of Models

The spatial performance of the models was initially evaluated by correlating observed values with climate model outputs at each station for all three climatic variables. For instance, Figure 3.8 presents a box plot of the monthly precipitation correlation coefficient values for each climate model at all stations. From this figure, it can be inferred that models MPI-ESM1-2-HR, ECEARTH3, and MRI-ESM2 exhibit relatively better correlation with observed monthly precipitation across the basin. On the other hand, models CMCC-ESM2 and TaiESM1 are among the poorly performing ones.

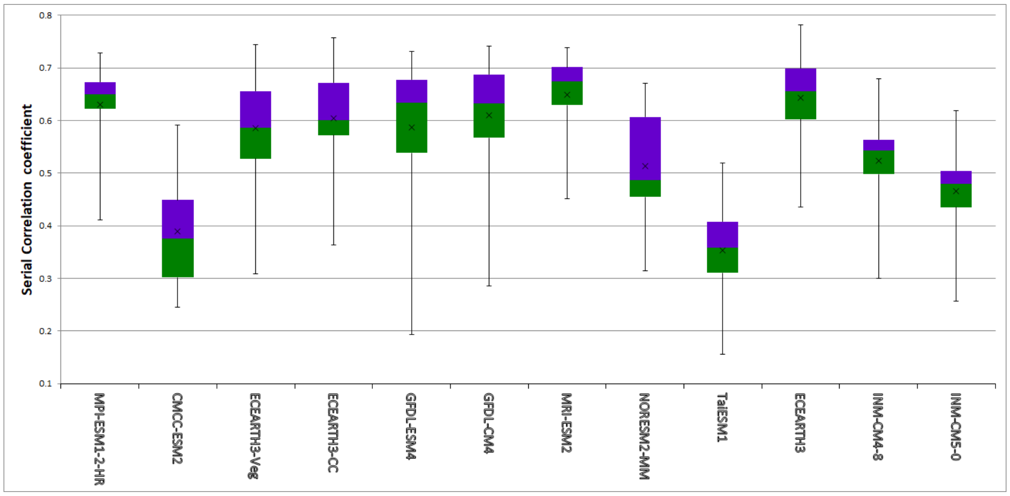


Figure 3.8: Box-plots of monthly precipitation correlation coefficients for each climate modes obtained across all climate stations in UASB.

Similarly, the correlation box plots for the maximum temperature series (Figure 3.9) demonstrate that two models, MPI-ESM1-2-HR and ECEARTH3, exhibit better correlation, while MRI-ESM2 is again the least performing.

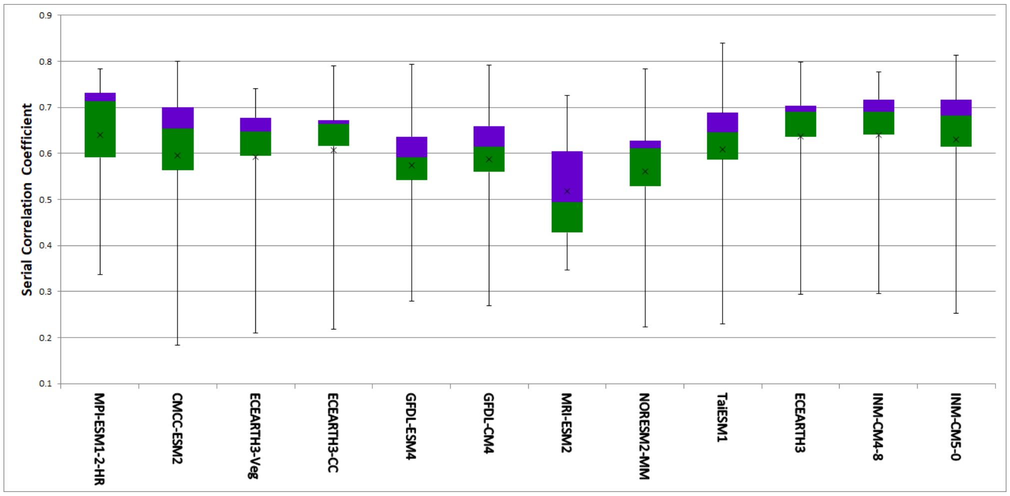


Figure 3.9: Box-plots of monthly Tmax correlation coefficients for each climate modes obtained across all climate stations in UASB.

The correlation coefficient values for the minimum temperature (Figure 3.10) do not align consistently with the precipitation and maximum temperature series. In this case, models belonging to the GFDL family exhibit relatively good performance compared to others, while the ECEARTH families show the lowest correlations.

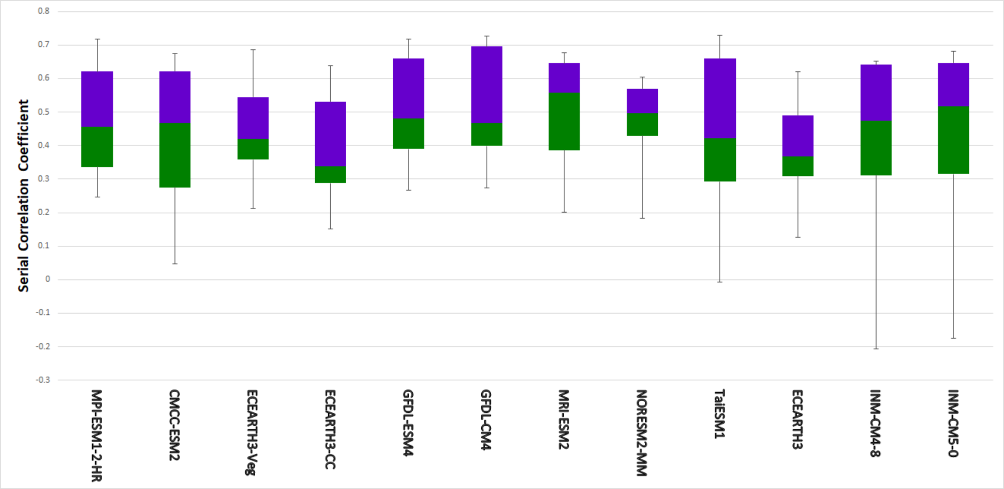


Figure 3.10: Box-plots of monthly Tmin correlation coefficients for each climate modes obtained across all climate stations.

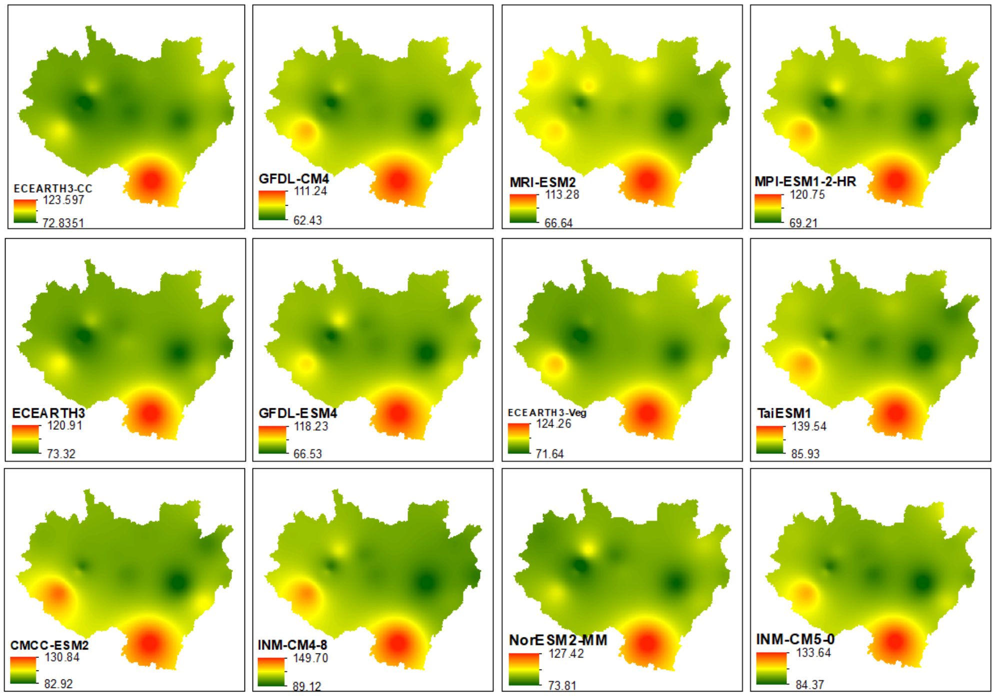


Figure 3.11: Maps of IDW interpolated monthly precipitation RMSE values of those 12 climate models.

In addition to the analysis of the spatial distribution of correlation coefficients, RMSE was used similarly to observe which climate model provides a better fit to the observed series. As shown in Figure 3.11, models GFDL-CM4, MRI-ESM2, and GFDL-ESM4 exhibit lower spatial RMSE values for monthly precipitation compared to the rest. Meanwhile, INM-CM4-8 and TaiESM1 show the least spatial performance. In all models, there appears to be less error in simulating rainfall at stations located at higher elevations.

3.4.2 Model Ranking and Interpretation

In the interpretation and summary of all results, a simple rank-based system is employed. A higher rank (lower in magnitude) is assigned to a specific climate model for the desired effect. For instance, in the interpretation of the result of distribution fitting, if a specific model has been fitted with the same distribution as the observed series, a value of one (unit) is assigned; otherwise, zero is assigned. These values are then summed across the four analysis periods for each model and given a rank. The ranking ranges from one to twelve, corresponding to the total number of climate models used in the study. In this context, a model would receive a higher rank if the sum of ranks is higher.

A similar approach is also used in the interpretation of trend analysis. If a similar trend is obtained for a specific model, a value of one is assigned, and zero otherwise. The ranking procedure is the same as that used in the interpretation of distribution analysis. For the case of performance measures, a higher rank is given to a specific model that has scored well according to the criteria of each measure. Since four performance measures are used for each model, the final ranking is based on the sum of ranks.

To interpret cluster analysis, the result of the distance (dissimilarity) matrix is used. A model with a minimum distance from the observed series is ranked higher than one with a larger distance. In the summary of PCA analysis, the angle between loading vectors of the observed and climate models is used as a ranking criterion. Among those twelve models, a model with a smaller angle is considered much closer to the observed series, so a higher rank is assigned to it. Finally, the areal averaged values of two spatial measures (correlation and RMSE) are given a rank.

From the result of precipitation time-series analysis across each method, model TaiESM1 has performed poorly, as shown in Table 3.5. The top three models that have shown an overall good performance are MRI-ESM2, GFDL-ESM4, and MPI-ESM1-2-HR.

There is no common model that exhibits good performance across all three climate variables, as seen in Table 3.5. The least performance for the maximum temperature series was observed in MRI-ESM2. In the case of the minimum temperature series, the ECEARTH3 family performed poorly, while, on the contrary, the GFDL family is among the top performers, as shown in Table 3.5.

Finally, the same ranking methodology was applied to identify possible models with an overall good performance for all three climatic variables of concern, as shown in Table 3.6. It can be observed that the MPI-ESM1-2-HR model is one of the models with the highest overall ranking when compared across those three climatic variables.

Table 3.6: Summary result for all three climatic variables

Data Type	Precip Rank	Tmax Rank	Tmin Rank	Sum of Rank	Final Rank
MPI-ESM1-2-HR	3	3	2	8	**1**
GFDL-CM4	4	9	1	14	**2**
GFDL-ESM4	2	9	4	15	**3**
INM-CM5-0	8	2	7	17	4
MRI-ESM2	1	12	4	17	4
ECEARTH3	5	1	12	18	6
TaiESM1	12	5	3	20	7
INM-CN4-8	11	3	8	22	8
CMCC-ESM2	9	8	6	23	9
ECEARTH3-Veg	7	6	10	23	9
NorESM2-MM	6	11	8	25	11
ECEARTH3-CC	9	6	11	26	12

Table 3.5: Summary rank based on seven criteria used for evaluating the three climate model time series

Model	PDF Rank	Trend Rank	PM Rank	Cluster Rank	PCA Rank	SpatCorr	SpaRMSE	Sum of Rank	Final Rank
Precipitation									
MRI-ESM2	1	1	8	6	1	1	3	21	**1**
GFDL-ESM4	5	1	3	4	3	5	2	23	**2**
MPI-ESM1-2-HR	5	1	7	1	5	4	4	27	**3**
GFDL-CM4	5	1	10	8	2	3	1	30	4
ECEARTH3	5	10	1	9	4	2	5	36	5
NorESM2-MM	2	1	6	5	8	8	8	38	6
ECEARTH3-Veg	5	1	2	12	7	7	7	41	7
INM-CM5-0	5	1	11	3	10	10	10	50	8
CMCC-ESM2	5	1	4	10	11	11	9	51	9
ECEARTH3-CC	4	10	12	7	6	6	6	51	9
INM-CM4-8	5	12	5	2	9	9	12	54	11
TaiESM1	2	1	9	11	12	12	11	58	12
Maximum Temperature									
ECEARTH3	4	6	3	3	4	3	1	24	**1**
INM-CM5-0	4	2	2	5	8	4	3	28	**2**
INM-CM4-8	4	6	4	4	5	2	4	29	**3**
MPI-ESM1-2-HR	4	9	6	1	6	1	2	29	3
TaiESM1	4	2	1	2	9	7	8	33	5
ECEARTH3-CC	2	12	5	8	1	5	5	38	6
ECEARTH3-Veg	4	6	8	6	2	6	6	38	6
CMCC-ESM2	4	2	11	7	10	8	7	49	8
GFDL-CM4	1	9	9	11	3	9	9	51	9
GFDL-ESM4	2	1	7	10	11	10	10	51	9
NorESM2-MM	4	2	12	9	7	11	12	57	11
MRI-ESM2	4	9	10	12	12	12	11	70	12
Minimum Temperature									
GFDL-CM4	2	1	1	4	2	1	1	12	**1**
MPI-ESM1-2-HR	1	5	2	1	6	6	7	28	**2**
TaiESM1	5	1	3	7	7	4	3	30	**3**
GFDL-ESM4	2	9	8	3	1	3	5	31	4
MRI-ESM2	5	9	7	2	4	2	2	31	4
CMCC-ESM2	2	1	9	5	5	8	8	38	6
INM-CM5-0	5	9	6	6	8	5	4	43	7
INM-CM4-8	5	7	4	8	9	7	9	49	8
NorESM2-MM	5	5	10	11	3	9	6	49	8
ECEARTH3-Veg	5	7	5	9	10	10	10	56	10
ECEARTH3-CC	5	1	12	10	11	11	11	61	11
ECEARTH3	5	9	11	12	12	12	12	73	12

NB: PDFR (PDF Rank), TrendR (Trend Rank), PMR (Performance Measure Rank), ClusterR (Custer Rank), PCAR (PCA Rank), SpatCorr (spatial Correlation), SpaRMSE (Spatial RMSE), SRank (Sum of Rank), FRank (Final Rank).

3.4.3 Future Climate Projections

To evaluate future climate conditions in the study area, the QM technique was applied to bias-adjust the predictions of the top five climate models: MPI-ESM1-2-HR, GFDL-ESM4, INM-CM5-0, MRI-ESM2, and ECEARTH3. Conducting future climate projections involved using an ensemble average of these five models, excluding GFDL-CM4 due to a lack of required scenarios.

Examining the precipitation variable, there is an overall increase in precipitation across three time periods (near, mid, and end of the century) compared to the base period (1980-2009), as detailed in Table 3.7. The mean annual areal precipitation increment for the near, mid, and end of the century is 1.3%, 10.6%, and 19.7%, respectively, under the worst-case socio-economic scenario. In the middle socio-economic scenario, increments of 2.7%, 6.4%, and 7.9% are observed for the same periods. With the exception of the sustainable scenario, a higher magnitude of increment is anticipated by the end of this century. Analyzing the seasonal pattern of change reveals a slightly higher magnitude of increment for spring rainfall compared to summer, with a predicted decrement in summer rainfall, especially for the near time period.

Concerning the mean annual areal maximum temperature in the middle scenario, projections indicate an increment of 1.1°C, 1.6°C, and 2.0°C for the near, mid, and end of the century, respectively (see Table 3.7). Similarly, the worst-case scenario shows a rise of 1.2°C, 2.0°C, and 3.0°C for the same periods. For the minimum temperature series in the middle scenario, an increase of 1.3°C, 2.0°C, and 2.4°C is observed for the near, mid, and end of the century, respectively. The worst-case scenario indicates increments of 1.4°C, 2.5°C, and 4.0°C for the same periods. In conclusion, akin to the precipitation variable, a higher magnitude of temperature increment is projected by the end of this century, with the increment being relatively higher for minimum temperature than for maximum temperature.

Table 3.7: Mean annual areal averaged future climate Change over UASB for three time periods and two scenarios

	Scenarios	Near	Mid	End
Precipitation(%Δ)	SSP5-8.5	(-2.1 - 5.3) 1.3	(6.1 - 16.1) 10.6	(11.8 - 29.4) 19.7
	SSP2-4.5	(-0.7 - 6.8) 2.7	(2.0 - 11.9) 6.4	(2.2 - 15.0) 7.9
	SSP1-2.6	(-1.2 - 9.0) 3.4	(1.4 - 16.6) 8.2	(-0.3 - 14.2) 6.2
Tmax(°C)	SSP5-8.5	(0.8 - 1.5) 1.2	(1.7 - 2.3) 2.0	(2.8 - 3.2) 3.0
	SSP2-4.5	(0.7 - 1.5) 1.1	(1.3 - 2.0) 1.6	(1.7 - 2.3) 2.0
	SSP1-2.6	(0.7 - 1.4) 1.1	(0.9 - 1.7) 1.3	(0.9 - 1.8) 1.3
Tmin(°C)	SSP5-8.5	(1.0 - 1.9) 1.4	(2.2 - 2.9) 2.5	(3.7 - 4.2) 4.0
	SSP2-4.5	(1.0 - 1.7) 1.3	(1.6 - 2.3) 2.0	(2.1 - 2.7) 2.4
	SSP1-2.6	(0.9 - 1.7) 1.3	(1.2 - 2.0) 1.6	(1.1 - 2.0) 1.6

NB: Numbers in the brackets indicate uncertainty limits of mean changes (for Precipitation $(\bar{x}_{predicted} - \bar{x}_{observed})/(\bar{x}_{observed})$ and for Temp $(\bar{x}_{predicted} - \bar{x}_{observed})$) of mean annuals calculated by using the standard deviation of each series (predicted and observed). For example, for predicted series the limits are $= \bar{x}_{predicted} \pm x_{standarddeviation,predicted}$. The same technique was applied in calculating limits of observed.

Table 3.8: Seasonal areal averaged precipitation future climate Change (%Δ) over UASB for three time periods and two scenarios

	MAM			JJAS		
	Near	Mid	End	Near	Mid	End
SSP5-8.5	(-0.15 - 0.46)0.03	(-0.10 - 0.47)0.08	(-0.04 - 0.58)0.15	(-0.12 - 0.02)-0.06	(-0.06 - 0.11)0.01	(0.02 - 0.16)0.08
SSP2-4.5	(-0.13 -0.6)0.08	(-0.08 - 0.45)0.09	(-0.11 - 0.48)0.08	(-0.11 - 0.04)-0.04	(-0.10 - 0.07)-0.02	(-0.06 - 0.08)0.00
SSP1-2.6	(-0.14 - 0.6)0.07	(-0.06 - 0.56)0.14	(-0.07 - 0.43)0.09	(-0.11 - 0.06)-0.04	(-0.09 - 0.09)-0.01	(-0.08 - 0.07)-0.02

The results of the future climate analysis discussed in the preceding paragraphs focus on areal averaged outputs. To gain insights into the spatial patterns of these changes, four representative stations were strategically selected across the basin. Their relative geographic positions (north, east, south, west) within the sub-basin were the main criteria for selection. The aim is to discern any spatial patterns in future climate changes across the sub-basin.

An examination of future precipitation (see Table 3.9) reveals a relatively higher mean change in precipitation and larger uncertainty in mean change predictions for stations Ejere and Hombole. Regarding future temperature changes, stations Deberzeit and Bui exhibit higher changes compared to the other two stations. Additionally, stations Bui and Tulubolo display larger uncertainties in predicting mean changes for both maximum and minimum temperature series. There is no clear spatial pattern observed in the changes in future precipitation across the sub-basin. However, the future changes in temperature values indicate that stations located at lower elevations experience higher relative changes.

Table 3.9: Mean annual future climate Change at selected four meteorological stations in the UASB for three time periods and two scenarios

Stations	Scenarios	Near	Mid	End
			Precip(%Δ)	
AA	SSP5-8.5	(4.3 - 9.7)7.4	(6.0 - 25.9)17.5	(28.6 - 33.2)31.3
	SSP2-4.5	(24.6 - 26.3)25.3	(10.5 - 13.3)11.7	(10.0 - 10.0)10.0
Ejere	SSP5-8.5	(-1.6 - 13.2)3.8	(21.1 - 23.9)22.8	(29.7 - 46.9)36.2
	SSP2-4.5	(11.1 - 37.9)20.9	(5.2 - 28.7)14.0	(6.9 - 24.1)13.4
Hombole	SSP5-8.5	(-23.4 - 63.7)-2.6	(3.4 - 77.7)22.0	(7.5 - 123.3)36.6
	SSP2-4.5	(-14.5 - 93.9)11.4	(-12.5 - 93.7)14.1	(-12.0 - 86.1)12.6
Ginchi	SSP5-8.5	(-3.3 - 5.1)-0.1	(14.2 - 15.1)14.6	(18.0 - 34.6)24.5
	SSP2-4.5	(6.3 - 26.5)14.1	(2.3 - 17.8)8.4	(0.2 - 21.3)8.5
			Tmax($^\circ C$)	
AA	SSP5-8.5	(1.0 - 1.3)1.2	(2.4 - 2.4)2.4	(3.3 - 3.6)3.5
	SSP2-4.5	(0.8 - 0.9)0.8	(2.1 - 2.3)2.2	(2.6 - 2.8)2.7
Bui	SSP5-8.5	(0.6 - 2.1)1.3	(2.4 - 3.4)2.9	(3.8 - 4.6)4.2
	SSP2-4.5	(0.5 - 1.6)1.1	(2.1 - 3.3)2.7	(2.7 - 4.0)3.3
Debrezeit	SSP5-8.5	(1.3 - 1.5)1.4	(2.9 - 3.2)3.0	(4.0 - 4.6)4.3
	SSP2-4.5	(1.0 -1.1)1.1	(2.7 - 2.8)2.7	(3.5 - 3.5)3.5
Tulubolo	SSP5-8.5	(-0.1 - 2.8)1.4	(1.3 - 3.8)2.6	(2.5 - 4.9)3.7
	SSP2-4.5	(-0.3 - 2.5)1.1	(1.0 - 3.8)2.4	(1.5 - 4.4)3.0
			Tmin($^\circ C$)	
AA	SSP5-8.5	(1.5 -1.7)1.6	(2.6 - 2.7)2.7	(3.8 - 4.4)4.1
	SSP2-4.5	(1.1 -1.2)1.2	(2.2 - 2.3)2.2	(2.6 - 3.0)2.8
Bui	SSP5-8.5	(1.3 - 3.8)2.5	(3.5 - 5.4)4.5	(5.4 - 7.0)6.2
	SSP2-4.5	(1.0 - 3.0)2.0	(2.9 - 5.1)4.0	(3.3 - 5.8)4.6
Debrezeit	SSP5-8.5	(1.5 - 1.9)1.7	(3.2 - 3.3)3.2	(4.4 - 4.9)4.7
	SSP2-4.5	(1.2 -1.3)1.2	(2.7 - 2.8)2.8	(3.1 - 3.6)3.3
Tulubolo	SSP5-8.5	(0.9 - 1.8)1.4	(1.9 - 4.1)3.0	(3.2 - 5.2)4.2
	SSP2-4.5	(0.8 - 1.4)1.1	(1.4 - 3.9)2.7	(1.8 - 4.5)3.1

3.5 Discussion

The models incorporated in the new CMIP6 archive demonstrate comparable or improved capabilities in simulating global climate compared to CMIP5 [80]. This improvement is attributed to enhanced model resolution, setup, scenarios, and parameterization in the CMIP6 archive [44]. This study primarily focuses on applying statistical and data mining techniques to identify climate models from the CMIP6 archive that can effectively simulate the climate system of the UASB. The evaluation of these models against observed historical climate series, using the specified evaluation criteria, enables the identification of potential models. All comparisons were based on spatial averaged data for the 1980-2009 period of analysis.

No specific model consistently outperformed others across each evaluation criterion and climatic variable, as detailed in Table 3.5. For instance, Model MRI-ESM2-0 demonstrated proficiency in representing the distributional properties of the precipitation series, while GFDL-CM4 and MPI-ESM1-2-HR excelled in simulating minimum and maximum temperature series, respectively. Model GFDL-ESM4 displayed consistent good performance for both precipitation and maximum temperature series, and models GFDL-CM4, TaiESM1, CMCC-ESM2, and ECEARTH3-CC showed equal potential for minimum temperature series.

The performance measures for model fit indicated that no specific model consistently exhibited superior performance across all three climate variables. MPI-ESM1-HR excelled in capturing the seasonal patterns for all three climate variables. However, in PCA analysis, not all three climate variables were correlated with the same climate model; MRI-ESM2, EC-EARTH3-CC, and GFDL-ESM4 showed better correlations for precipitation, maximum temperature, and minimum temperature series, respectively.

Spatial evaluation criteria revealed that MRI-ESM2 exhibited better spatial correlation, while GFDL-CM4 showed a good model fit for the precipitation series. For maximum temperature series, MPI-ESM1-2-HR and ECEARTH3 ranked highest based on these criteria. Additionally, GFDL-CM4 claimed the top rank for minimum temperature series in both spatial evaluation criteria.

According to recent studies, summer rainfall over Ethiopia is influenced by sea surface temperature conditions over the Gulf of Guinea and the southern Pacific Ocean, while spring rainfall is influenced by sea surface temperature over the North Atlantic Ocean. Equatorial Pacific Ocean temperature, ENSO, and ITCZ are identified as significant factors shaping the climate over the UASB [23, 24].

Many climate models evaluated in this study exhibit varying performances in simulating global climate processes. Notably, EC-EARTH3, MPI-ESM1-2-HR, and TaiESM1 have shown good capabilities in capturing ENSO teleconnections [80]. MPI-ESM1-2-HR has demonstrated proficiency in producing teleconnections over the Indian and North Atlantic Oceans [81]. These models align with the top-ranked models identified in this study.

The top-ranked model, MPI-ESM1-2-HR, likely performed well due to its higher atmospheric and ocean model resolution relative to other models. With a 0.94° atmospheric resolution, 95 verticals, and model top at 0.01 hPa, coupled with a 0.4° horizontal resolution and 40 levels for the ocean model, MPI-ESM1-2-HR outperformed models like CMCC-ESM2, which has a lower atmospheric resolution of $0.9^\circ \times 1.25^\circ$ with 30 verticals and model top at 2 hPa, coupled with a 0.33° ocean resolution [82, 81, 83]. Considering the complex topography of the study area, model resolution is crucial in capturing all significant climate characteristics.

Previous climate change studies in the basin using CMIP5 outputs predicted an increase in temperature for all time periods and scenarios [60, 58, 59]. However, there was no consensus on future precipitation changes. Discrepancies may be attributed to the poor performance of previous climate models in simulating East Africa's climate or inadequate model selection. In contrast, this study predicts an increase in both precipitation and temperature (Table 3.7), aligning with the AR6 IPCC report, which forecasts an increase in annual mean precipitation and maximum temperature values over East Africa for the mid and end of the century [84]. The study's findings on future climate projections, including a slight higher increment in spring rainfall than summer, also align with the AR6 IPCC report, providing a more consistent outlook compared to previous CMIP5 outputs. Notably, the study's results, showing an increase in precipitation and temperature, and the associated uncertainty bands, corroborate with the AR6 IPCC report, indicating a higher uncertainty in future precipitation predictions toward the end of the century and higher uncertainty in temperature predictions for the mid and end of the century. The study concludes that the findings are more consistent with the AR6 IPCC report on future climate conditions over North-East Africa [84].

3.6 Conclusions

The outcomes of the model selection process emphasize the challenge of drawing conclusions about the performance of a single model, as each model exhibits varying performance across different evaluation criteria. Nevertheless, model MPI-ESM1-2-HR demonstrates an overall strong performance when assessed for all time steps and evaluation criteria.

Analyzing the outputs of each evaluation criterion for the three climatic variables, the GFDL-CM4 model emerges as the most robust across three measures (PDF, PCA, and Spatial). In terms of capturing trend behavior, CMCC-ESM2 excels. Model fit performance measures (R^2, RMSE, MAE, and BIAS) highlight that INM-CM4-8 outperforms other models. Notably, MPI-ESM1-2-HR stands out as the best model for capturing the stochastic behavior, including underlying periodic patterns, of the observed series.

The subsequent downscaling and bias adjustment, applied to scenarios from the top five high-ranking models (ECEARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2, and INM-CM5-0) using IDW and QM approaches for all three climate variables (Precipitation, Maximum temperature, and Minimum temperature), resulted in future climate predictions for three socio-economic scenarios (sustainable, middle, and worst) across three time periods (near, mid, and end of the century). The predictions indicate a relative increment compared to the base period (1980-2009) for all three climate variables, with a higher magnitude of increment expected by the end of this century. Seasonal projections reveal a higher magnitude of increment for spring rainfall than for summer. The study's findings align with the new AR6 IPCC report on expected future climate changes in North East Africa.

Considering the intensification of agricultural activity and urbanization in the study area [25], coupled with the projected increase in precipitation, there is an elevated risk of flooding in the UASB in the future.

The results of this study could be further validated by incorporating other ensemble variants in addition to r1i1p1f1.

Chapter 4

Flow Characteristics of Rivers in the Sub-basin

4.1 Introduction

Water is considered a non-renewable resource, necessitating careful management at the basin level through a meticulously planned water resource strategy. The success of any water resource planning and management initiative hinges on a thorough examination of the water resource potential within a given basin. Such studies often probe into both past and future water resource conditions, considering temporal (time series)/spatial scales or a combination of both. To accomplish these objectives, high-quality climatic and hydrological data of sufficient length are imperative.

Studies investigating past and future water resource conditions commonly employ diverse methodologies using hydrological time series. These methodologies may involve the use of raw flow time series or the application of different time series or hydrological models. Before embarking on any modeling effort, it is essential to process raw flow data and ensure it aligns with the assumptions required by these models. Understanding the characteristics of a flow time series significantly aids in selecting a more proficient time series model with superior performance in capturing the underlying stochastic nature [85] [86].

Critical characteristics of a time series encompass its trend, stationarity, linearity, periodicity, and seasonality. Trend refers to a consistent increment or decrement in the flow magnitude over a specified analysis period. Recent studies have employed techniques such as linear regression [87], Mann Kendall [88][89][90], Sen's slope, and ITA [91][92] to discern the trend characteristics of river flow time series. Stationarity, another crucial concept, indicates whether statistical attributes (mean, variance, and autocorrelation) of the time series remain unchanged over a specified time period. Tests like KPSS and ADF [89] are commonly used to identify stationarity. Linearity is equally significant, determining whether observations in a time series exhibit time dependency—a pivotal consideration when fitting time series models. Linearity can be assessed using various methods, including Portmanteau or BDS tests [85][93][94]. Seasonality, indicating the repetition of patterns in a time series, such as peaks or troughs, can be identified through Autocorrelation Function (ACF) (Partial Autocorrelation Function - PACF) plots, periodograms, or by observing different statistical characteristics (mean, standard deviation (SD), coefficient of variation (CV)) [90].

This chapter aims to provide a detailed examination of the flow characteristics derived from major rivers in UASB, offering insights into the hydrology of the sub-basins.

4.2 Data Sets

The investigation into flow characteristics over the UASB centered on four key stations: Akaki, Hombole, Melka Kuntere, and Mojo, as detailed in Table 4.1. Station selection was based on the availability of relatively extensive data records. Akaki, situated in the northeastern part of the UASB, is predominantly a metropolitan area with a total drainage area of 956.75 km^2, featuring the longest stream in the sub-basin stretching to 87.4 km. The second sub-basin, Melka Kuntere, spans most of the northwestern area of the basin and encompasses a total drainage area of 4517.8 km^2, with the longest stream in this sub-basin extending to 120.1 km. The third sub-basin, Mojo, located in the southeastern part of the basin, significantly contributes to the larger Awash basin, covering 1461.707 km^2, and hosting the longest stream with a span of 87.21 km. The fourth station serves as a collection point for the flow from all these stations, featuring a total drainage area of 7693 km^2 and the longest stream measuring 190 km. The table below provides a presentation of the statistical characteristics of the flow series at different time levels for each station.

Table 4.1: Statistics of flow series for each hydrological Stations over UASB.

Station	Periods	Time level	Mean(m^3/s)	Standard Deviation (m^3/s)	Skewness	Kurtosos	ACF(1)
Akaki@Akaki	1981-2008	Daily	7.51	13.74	5.02	32.33	0.69
		1/3 Month	7.51	10.84	3.23	13.67	0.77
		Monthly	7.46	9.74	2.63	8.35	0.53
		Annual	7.46	2.62	0.17	0.00	0.36
Awash@Hombole	1975-2015	Daily	41.00	72.06	2.63	7.91	0.94
		1/3 Month	40.99	68.24	2.22	4.50	0.89
		Monthly	40.70	63.52	1.98	3.34	0.59
		Annual	40.70	10.19	0.27	0.00	-0.16
Awash@Melka	1966-2012	Daily	18.47	28.12	1.94	3.83	0.99
		1/3 Month	18.47	27.69	1.96	4.04	0.93
		Monthly	18.37	26.58	1.90	3.94	0.65
		Annual	18.47	4.82	1.25	2.15	-0.24
Mojo@Mojo Village	1968-1991	Daily	4.98	12.97	6.11	51.02	0.48
		1/3 Month	4.98	8.80	3.21	13.12	0.67
		Monthly	4.95	7.62	2.57	7.77	0.47
		Annual	4.99	1.76	0.67	0.44	0.02

4.3 Methods

The characterization of the flow series in the sub-basin involves the application of various techniques. Before undertaking any calculations or analyses using the flow data from any of the gauge stations, a thorough check for data quality is conducted. Every time series comprises trend, seasonal, and stochastic components, which collectively contribute to understanding its nature [95] [96]. Additionally, stationarity and linearity are two crucial concepts essential for comprehending any time series and subsequently fitting an appropriate model to it. The subsequent subsections delve into a detailed theoretical background of these concepts, accompanied by corresponding statistical tests used to identify their presence in the time series.

4.3.1 Filling Missing Data and Outlier Detection

This data processing stage is crucial before conducting any tests or analysis on the flow series. Raw data often contains missing or erroneous values that may have occurred during various stages of the data retrieval process. In this study, the flow time series underwent an initial visual analysis to identify any missing or outlier data points using graphical techniques on the daily flow series.

For missing data points, the approach involved filling them by taking the average of backward and forward data points relative to the location of the missing data point. If the number of days exceeds three, the daily mean is substituted for that day using available data from other years. Outliers were detected and filled using the 'fill outliers' function in MATLAB [97].

4.3.2 Stationarity Test

Stationarity is a crucial assumption in time series analysis, often signifying that certain statistical properties remain constant over time. The primary metrics used to define stationarity are the first moment ($\mu_x(t)$) and the second moment ($\gamma_x(t)$). There are two general classifications of time series stationarity: Weak and Strict stationarity. Weak stationarity is characterized by both the first and second moments being independent of time. On the other hand, Strict stationarity is a more stringent requirement, demanding that the joint distribution remains the same, in addition to the conditions specified by weak stationarity. Specifically, for a time series (x_t), the distribution of ($x_1, . . ., x_n$) and ($x_{1+h}, . . ., x_{n+h}$) should be identical, where h and n are integers, and $n > 0$. In practice, when a time series is referred to as stationary, it generally implies Weak stationarity, as Strict stationarity is a more demanding assumption.

In this study, a stationarity test is conducted on the flow series from each gauging station after the removal of potential outliers. Two well-known stationarity tests, the Augmented Dickey-Fuller (ADF) test [98] and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [99], are applied.

The ADF test aims to determine the stationarity of the series by identifying the presence of a unit root (i.e., $\rho = 1$) through fitting the data with Equation 4.1.

$$\Delta Y_t = \alpha + \beta t + (\rho - 1)Y_{t-1} + \sum_{i=1}^{K} \theta_i \Delta Y_{t-i} + u_t \tag{4.1}$$

Where: α is constant term, βt is the trend term, the third term is AR(1) process, fourth term is AR(K) process and u_t is a purely random process.

These techniques primarily involve testing for stationarity, particularly to detect suspected trends and autoregressive (AR) processes. Depending on the nature of the time series, the first null hypothesis of no-trend stationarity ($H_0 : (\alpha, \beta, \rho) = (\alpha, 0, 1)$) is examined. If the hypothesis of no-trend stationarity cannot be rejected, the second hypothesis, indicating no stationary AR process ($H_0 : (\alpha, \rho) = (\alpha, 1)$), is then tested. Additionally, the presence of drift ($H_0 : (\alpha, \rho) = (0, 1)$) in the time series can also be identified. The estimated Augmented Dickey-Fuller (ADF) test statistic (Equation 4.2) is compared against the standard Dickey-Fuller t statistic value corresponding to the sample size (T) and the chosen confidence interval.

$$ADF_t = \hat{\rho} - 1/SE(\rho) \tag{4.2}$$

Where: $\hat{\rho}$ represents the estimated coefficient of the lagged differenced variable, and $SE(\rho)$ is the standard error of this coefficient.

Another commonly applied stationary test is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test [99]. Unlike the ADF test, the KPSS test directly examines stationary behavior in the fitted AutoRegressive Integrated Moving Average (ARIMA) process. In this case, the null hypothesis is stationarity. The model used to fit the data includes a time-dependent constant term, trend term, and stationary random walk.

$$y_t = \alpha_t + \beta t + u_t \tag{4.3}$$

$$\alpha_t = \alpha_{t-1} + \epsilon_t \tag{4.4}$$

Using KPSS, a test for both trend and level stationarity can be conducted. The null and alternative hypotheses for the trend stationarity case are shown in Equation 4.5, while for the level stationary case, they are presented in Equation 4.6.

$$H_O : {\sigma_\epsilon}^2 = 0; H_1 : {\sigma_\epsilon}^2 \neq 0 \tag{4.5}$$

$$H_O : \beta = 0 \tag{4.6}$$

The test statistic for the two cases are given by;

$$\hat{\eta_j} = \frac{1}{T^2} \frac{\sum_{t=1}^{T} (S_{t,j})^2}{{S_{Tm}}^2}, j = \mu, Tr \tag{4.7}$$

$$S_{t,j} = \sum_{1}^{t} \hat{u}_{i,j} \tag{4.8}$$

$$S_{Tm}^2 = \sum_{1}^{T} \hat{u}_t^2 + \frac{2}{T} \sum_{1}^{m} (W_{im} \sum \hat{u}_t \hat{u}_{t-i}) \tag{4.9}$$

In Equation 4.7, if $j = \mu, Tr$ means the least square residuals (u_t) are estimated by regressing y both on the constant and trend term. Otherwise, if $j = \mu$ means the residuals are estimated only using the constant term. The m term stands for the truncation parameter and denotes the maximal order up to which the autocovariances are included, and T is the sample size. In Equation 4.9, S_{Tm}^2 is the adjusted variance of the residuals with consideration of autocorrelation. Also, W_{im} is the weight used to ensure the consistency of the variance estimator. According to [100], the weight suggested by [101] is often used, as shown in Equation 4.10.

$$W_{im} = 1 - \frac{i}{m+1} \tag{4.10}$$

Where: $i = 1, 2, \quad . \quad . \quad ., m.$

4.3.3 Linearity Test

Most natural systems exhibit linear relationships, with nonlinearity arising from the complex interactions inherent in these systems. Streamflow, as a component of the Earth's surface processes, is part of the intricate hydrologic cycle, specifically the rainfall-runoff process. This hydrological cycle comprises numerous subsystems, making the estimation of streamflow a challenging task [85]. Linearity is a characteristic that indicates whether a time series originates from random events, often referred to as an independent and identically distributed (iid) process.

A one sided linear process x_t can be expressed with the following equation:

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i} \tag{4.11}$$

where μ and ψ_i are real numbers and a_t is a sequence of iid random variables with mean zero and a well defined density function. Here, $\psi_0 = 1$ and $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

The linearity of a time series can be assessed using Equation 4.11, which describes a linear autoregressive (AR) process. The observed flow series is fitted to this equation after removing any trend and seasonal patterns present. The residuals of the AR model are then examined to determine if they exhibit iid behavior. If the residuals prove to be iid, no further modeling is necessary, and only the estimation of their mean and variance is required. However, if the residuals do not exhibit iid behavior, it suggests the presence of deterministic chaos in the time series x_t, indicating patterns that cannot be described by simple mathematical functions.

This study employs a well-known technique, known as the BDS (Brock, Dechert, and Scheinkman) test [102], to identify possible non-linear behavior in the observed flow series from the four gauging stations. The BDS test is conducted using the tseries package [104] in R.

In the BDS test, the theory of the correlation integral (C) applied to determine whether the linear model fitted to the data series (X_t for t=1,2,. . .,T) is linearly dependent or independent. The correlation integral assesses the degree (probability) to which the data points in a time series are interconnected. The test involves calculating the absolute difference (distance) of each data point to every other data point, creating a distance matrix. A threshold (ϵ)is then set, defining the percentage of data points considered related. Using this threshold, each cell of the distance matrix within the threshold is assigned a value of unity, while those outside receive a value of zero. The correlation integral is subsequently calculated as the sum of values within each cell divided by the total number of possible pairs used in the distance matrix calculation. Equation 4.12 summarizes these steps, describing the correlation integral ($C_{m,\epsilon}$)for an embedding dimension (m) and threshold value (ϵ).

$$C_{m,\epsilon} = \frac{2}{T_m(T_m - 1)} \sum\sum_{m \leq s < t \leq T} I(x_t^m, x_s^m; \epsilon) \tag{4.12}$$

Where, $T_m(T_m - 1)$ is the number of constructed m- histories ($x_t = (X_t, x_{t-1}, ..., x_{t-m+1})$). The term in double summation ($I(x_t^m, x_s^m; \epsilon)$) expresses the assigned values based on the set threshold for every cell in the distance matrix.

The correlation integral represents the probability of dependence between pairs of data points. If similar probabilities can be estimated for pairs of pairs, and these probabilities are independent of each other, the correlation integral for m number of pairs can be expressed as follows:

$$C_{1,\epsilon}^m = Pr(|X_t - X_s| < \epsilon)^m \tag{4.13}$$

The BDS test statistic is calculated based on the expression in Equations:4.12 and 4.13 as follows;

$$V_{m,\epsilon} = \sqrt{T}\frac{C_{m,\epsilon} - C_{1,\epsilon}^{m}}{S_{m,\epsilon}} \tag{4.14}$$

Where $S_{m,\epsilon}$ is the standard deviation of $\sqrt{T}(C_{m,\epsilon} - C_{1,\epsilon}^{m})$. The BDS test statistic converges to normal distribution N(0,1) under fairly moderate regular conditions.

$$V_{m,\epsilon} \xrightarrow{d} N(0,1) \tag{4.15}$$

So, the null hypothesis of iid can be rejected at 5% significance level when ever $V_{m,\epsilon} > 1.96$. In this study, BDS test is applied to residuals obtained from fitting an AR(p) linear model to the data. Also, a distance threshold (ϵ) equal to standard deviation and an embedding dimension (m) up to 5 can be applied.

4.3.4 Trend Test

The trend test is performed with the help of Mann-Kendall [67] or modified Mann-Kendall trend [68] depending on the result of test for presence of auto-correlation. This methodology is already discussed in detail in subsubsection - 3.3.1 of this Thesis.

4.3.5 Periodicity

The seasonal characteristics of the flow series will be studied with the help of spectral density plots (periodograms). The periodogram expresses the relative strength of various frequencies (periods) for explaining the variation in the time series, thereby helping to identify the dominant frequency in the data. A time series can be represented as the superposition of uncorrelated sinusoids, as shown in Equation:4.16 below.

$$X_t = \sum_{j=1}^{k} A_j cos(\omega_j t) + B_j sin(\omega_j t), \quad o < \omega_1 < ... < \omega_k < \pi \tag{4.16}$$

Where: $A_1, B_1, ..., A_k, B_k$ are uncorrelated variables with $E(A_j) = E(B_j) = 0$ and $var(A_j) = var(B_j) = \sigma_j^2$. And $k = n - 1$, n is the data length.
Here, X_t is a zero mean stationary time series with autocovariance function $\gamma(.)$ satisfying $\sum_{h=-\infty}^{\infty}|\gamma(h)| < \infty$. So, spectral density of X_t can be expressed with the function f(.).

$$f(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma(h), \quad -\infty < \lambda < \infty, \quad for \quad all \quad \lambda\epsilon(-\pi, \pi) \tag{4.17}$$

Where; $e^{i\lambda} = cos(\lambda) + i\sin(\lambda)$, h is lag, and $i = \sqrt{-1}$.

In practice the actual data (X_t) is transformed into frequency domain (Periodogram) using Discrete Fourier Transform (DFT) technique as shown in Figure:4.1 using the following expression;

$$I_n(\lambda) = \frac{1}{n}\left(\sum_{t=1}^{n} X_t e^{-it\lambda}\right)^2 \tag{4.18}$$

Where: λ is one of Fourier frequencies ω_k.

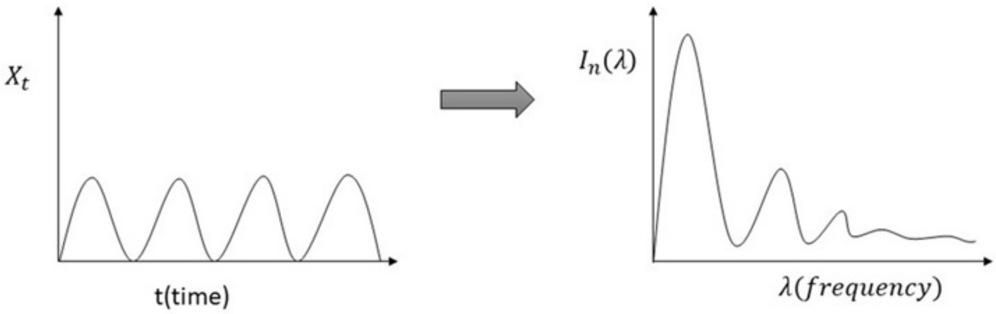


Figure 4.1: Transformation of time series from time into frequency domain.

In this study the periodograms are developed using the parametric method which first fits AR model to the flow series. This procedure is undertaken by the help of astsa package in R [108].

4.3.6 Seasonal Character

Seasonality refers to the variations in time series that can be described by regular short time intervals, such as weekly, monthly, bi-yearly, or quarterly periods. The seasonal nature of time series is understood by observing changes in different flow series statistics. The statistics, including mean, Standard Deviation (SD), Coefficient of Variation (CV), and Autocorrelation Function (ACF), have been applied at both the daily and monthly time levels. The analysis can also be performed on the 1/3 monthly flow series.

To calculate these statistics, the daily and monthly flow series are structured in matrix form, where the rows (i) represent years of analysis, and columns (j) indicate specific days for daily or months for monthly flow series, as shown in the matrix Equation:4.19.

$$\begin{bmatrix} x_{(1,1)} & x_{(1,2)} & . & . & . & x_{(1,j)} \\ x_{(2,1)} & x_{(2,2)} & . & . & . & x_{(2,j)} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ x_{(i,1)} & x_{(i,2)} & . & . & . & x_{(i,j)} \end{bmatrix} \tag{4.19}$$

The mean ($\bar{x_j}$), SD (S_i) and CV_j are calculated as follows;

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j} \tag{4.20}$$

$$S_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)^2} \tag{4.21}$$

$$CV_j = \frac{S_j}{\bar{x}_j} \tag{4.22}$$

In the same manner ACF (x_j, x_{j+k}) for different lags (k) was calculated using the following expression;

$$ACF_j(k) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)(x_{i,j+k} - \bar{x}_{i,j+k})}{S_j S_{j+k}} \tag{4.23}$$

The daily or monthly statistics, calculated using the mentioned techniques, have been summarized and presented primarily through graphical representation in the results section.

In addition to these statistics, data processing techniques have been applied to observe the seasonal autocorrelation of the time series. These data processing steps include trend removal, normalization, and deseasonalization techniques. The trend removal step involves fitting a linear regression model and subsequently removing the linear trend from each time level of the data. The second step is normalizing the time series by either taking the logarithm or applying the Box-Cox transformation [109], as shown in Equation:4.24. Deseasonalization is achieved through standardization techniques, which remove the seasonal mean and standard deviation from each data point in the time series, as shown in Equation:4.25.

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & for \quad \lambda \neq 0 \\ log(x) & for \quad \lambda = 0 \end{cases} \tag{4.24}$$

$$\frac{x_{j,i} - \bar{x}_j}{S_j} \tag{4.25}$$

4.4 Results

4.4.1 Test for Stationarity and Linearity

The results of the stationarity analysis on the flow series of the aforementioned hydrological stations are summarized in Table 4.2. The test were undertaken using the urca package [110] in R. The test results conclude that almost all of the flow series at different time levels exhibit stationary characteristics, except for the annual series of the Mojo gauging station. In the annual series of Mojo, the KPSS trend test indicates non-stationarity, while the ADF trend test shows the opposite. This suggests that differencing the series is necessary to make it stationary.

Table 4.2: Result of stationarity test for four gauge stations along four time levels using KPSS and ADF test

			KPSS test				ADF test	
			Level Stationarity		Trend Stationarity			
Station	Time level	lag Max.	Test Statistic	P-value	Test Statistic	P-value	Test Statistic	P-value
Akaki	Daily	38	0.11	> 0.1	0.10	> 0.1	-26.24	< 0.01
	1/3 Month	21	0.09	> 0.1	0.09	> 0.1	-10.94	< 0.01
	Monthly	16	0.10	> 0.1	0.09	> 0.1	-12.67	< 0.01
	Annual	9	0.13	> 0.1	0.12	> 0.1	-3.14	0.04
Hombole	Daily	42	0.04	> 0.1	0.01	> 0.1	-12.14	< 0.01
	1/3 Month	24	0.07	> 0.1	0.02	> 0.1	-14.87	< 0.01
	Monthly	18	0.12	> 0.1	0.03	> 0.1	-17.22	< 0.01
	Annual	10	0.25	> 0.1	0.13	> 0.1	-7.07	< 0.01
Melkakuntere	Daily	43	0.06	> 0.1	0.02	> 0.1	-8.45	< 0.01
	1/3 Month	24	0.07	> 0.1	0.03	> 0.1	-14.76	< 0.01
	Monthly	18	0.11	> 0.1	0.04	> 0.1	-17.41	< 0.01
	Annual	10	0.29	> 0.1	0.13	> 0.1	-8.78	< 0.01
Mojo	Daily	37	0.29	> 0.1	0.04	> 0.1	-28.56	< 0.01
	1/3 Month	21	0.26	> 0.1	0.04	> 0.1	-10.59	< 0.01
	Monthly	16	0.34	> 0.1	0.06	> 0.1	-11.04	< 0.01
	Annual	8	0.31	> 0.1	0.15	0.04	-4.46	< 0.01

[1] Critical Values: ADF_{Daily}=-2.8618,$ADF_{1/3Month}$=-2.8648,$ADF_{Monthly}$=-2.8674,ADF_{Annual}=-2.9351,$KPSS_{level}$=0.463, $KPSS_{trend}$=0.146.

[2] Significance level(α) is 5%.

Another critical characteristic of time series, essential for identifying a suitable forecasting model, is the linearity assumption. The requirement for linearity is that the residuals (noise) of a model should be independent random variables or iid (independent and identically distributed). If the residuals are independent, there is no need for further modeling other than estimating their mean and variance. However, if there is dependency in the residuals, it indicates that past and recent values are correlated. This information is valuable for identifying a potentially complex stationary time series model.

In this study, the linearity test is performed on the flow time series using the non-parametric BDSL test [102] through the tseries package [104] in R at four time levels, as shown in Table 4.3.

Table 4.3: Result of linearity test for four gauging stations along four time levels using BDS test

		BDS Test							
Station	Time level	m=2		m=3		m=4		m=5	
		T.Stat	P-val	T.Stat	P-val	T.Stat	P-val	T.Stat	P-val
Akaki	Daily	57.39	0.0	62.75	0.0	66.31	0.0	69.11	0.0
	1/3 Month	16.99	0.0	19.91	0.0	21.08	0.0	21.81	0.0
	Monthly	6.03	0.0	5.67	0.0	4.86	0.0	3.83	0.0
	Annual	-3.79	0.0	-1.15	0.25	-0.93	0.35	-1.25	0.21
Hombole	Daily	64.49	0.0	76.07	0.0	83.21	0.0	89.06	0.0
	1/3 Month	18.39	0.0	22.89	0.0	25.90	0.0	28.60	0.0
	Monthly	9.04	0.0	9.65	0.0	8.94	0.0	7.82	0.0
	Annual	-0.71	0.48	-1.24	0.21	-1.07	0.28	-1.23	0.22
Melkakuntere	Daily	86.00	0.0	93.37	0.0	97.02	0.0	100.36	0.0
	1/3 Month	22.85	0.0	27.23	0.0	30.85	0.0	34.79	0.0
	Monthly	8.33	0.0	9.50	0.0	9.48	0.0	8.67	0.0
	Annual	-0.09	0.92	-1.34	0.18	-1.96	0.05	-1.82	0.07
Mojo	Daily	52.17	0.0	57.92	0.0	60.34	0.0	62.11	0.0
	1/3 Month	12.62	0.0	15.62	0.0	17.14	0.0	18.17	0.0
	Monthly	6.35	0.0	6.60	0.0	5.92	0.0	5.02	0.0
	Annual	0.07	0.94	-0.09	0.92	-1.30	0.19	-1.85	0.06

[1] critical values for rejection of null hypothesis is when $|BDSTestStatistics| > 2.0$ [111] and for $\epsilon = standard \quad deviation(\sigma)$.

[2] Significance level(α) is 5%.

The results of the BDS test conducted on residuals from the fitted AR(p) linear model of the flow series at various time intervals for each gauging station in the UASB confirm the existence of ruminant nonlinear dependencies, indicating non-random patterns in the daily, 1/3 month, and monthly series. However, the residuals from the fitted linear model for the annual flow series show a contrasting result, suggesting a more linear behavior with the presence of independent and identically distributed (iid) characteristics.

4.4.2 Trend Characteristics

The trend tests were conducted using the modifiedmk package in R [112] at two time levels. Before applying these trend tests, the data were examined for the presence of any significant autocorrelation characteristics. Depending on this, either the original Mann-Kendall trend test [67] or the modified Mann-Kendall trend test [68] was applied. Tables 4.4 and 4.5 below illustrate the results of the trend tests.

The trend analysis for the annual flow series (Table 4.4) shows that there is no significant trend observed for all stations, with the exception of the Mojo station. In this station, the annual flow seems to decrease at a rate of $0.131m^3/s$. The monthly analysis of trends at the Hombole station (Table 4.5) reveals a significant increasing trend only for four months: June, October, November, and December. Among these, the last three months immediately follow the main rainy season (June-July-August-September(JJAS)). In the Akaki station, there is a significant increasing trend for almost all months, with the exception of the rainy months (July-August-September (JAS)), where there is an insignificant decrement in flow. However, in the Melkakuntere station, it was difficult to observe any clear seasonal trend behavior, as most months showed insignificant test results. In this station, February and August flows exhibited a decreasing trend, while June showed an increase. Although the tests are statistically insignificant, there appears to be a decreasing pattern in flow for most months during the rainy seasons (March-April and JAS). Additionally, the trend test outputs for the Mojo station also indicate similar results to the Melkakuntere station, with only three months (February, April, and August) showing a significant decreasing trend.

Table 4.4: Annual level Trend test on the flow time series of those four guaging staions over UASB.

Station	Z-Value	P-Value	Sen's Slope
Akaki@Akaki	0.849	0.396	0.048
Awash@Hombole	0.752	0.452	0.110
Awash@Melka	-1.32	0.187	-0.071
Mojo@MojoV.	-2.406	0.016	-0.131

[1] critical values for rejection of null hypothesis is when $|Z - value| > 1.96$ and P- value is ≤ 0.05.
[2] Significance level(α) is 5%.

Table 4.5: Monthly level Trend test on flow time series of those four guaging staions over UASB.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Akaki@Akaki												
Z-value	5.642	4.486	4.978	4.330	5.479	4.705	-0.336	-1.600	-1.561	4.592	6.382	6.247
P-value	1.7e−8	7.2e−6	6.4e−7	1.4e−5	4.3e−8	2.5e−6	0.737	0.109	0.118	4.3e−6	1.8e−10	4.2e−10
Sen's S.	0.139	0.139	0.119	0.134	0.139	0.150	-0.042	-0.413	-0.404	0.176	0.196	0.204
Awash@Hombole												
Z-value	1.916	-0.595	0.460	1.112	1.853	2.572	0.236	0.258	0.460	6.968	3.426	2.664
P-value	0.055	0.551	0.645	0.266	0.064	0.010	0.814	0.796	0.645	3.21e−13	0.006	0.008
Sen's S.	0.020	-0.013	0.017	0.056	0.084	0.299	0.139	0.175	0.316	0.244	0.139	0.037
Awash@Melka												
Z-value	0.953	-2.109	-0.128	-0.385	1.540	2.623	-0.092	-2.402	-0.624	-1.540	0.146	0.825
P-value	0.340	0.035	0.898	0.700	0.123	0.009	0.927	0.016	0.533	0.123	0.883	0.409
Sen's S.	0.006	-0.019	-0.0018	-0.008	0.031	0.172	-0.012	-0.472	-0.101	-0.110	0.003	0.005
Mojo@MojoV.												
Z-value	-1.017	-2.108	-1.860	-2.158	-0.273	0.124	-1.265	-2.257	-0.818	1.116	1.067	0.372
P-value	0.309	0.035	0.063	0.031	0.785	0.901	0.206	0.024	0.413	0.264	0.286	0.709
Sen's Sl.	-0.016	-0.032	-0.054	-0.040	-0.009	0.016	-0.199	-0.637	-0.107	0.018	0.014	0.008

[1] critical values for rejection of null hypothesis is when $|Z - value| > 1.96$ and P- value is ≤ 0.05.
[2] Significance level(α) is 5%.

4.4.3 Periodic Patterns

Here, the monthly series was utilized to discern the types of seasonal patterns present in the flow series. The raw periodogram developed from the original data tends to be rough, producing discrete harmonic frequencies. However, a spectral density should be defined over continuous frequencies. Therefore, smoothing techniques or an alternative approach, which involves fitting a higher-order AR model before developing the periodogram, is often recommended. In this study, the latter approach is applied, and the results are shown in Figure 4.2.

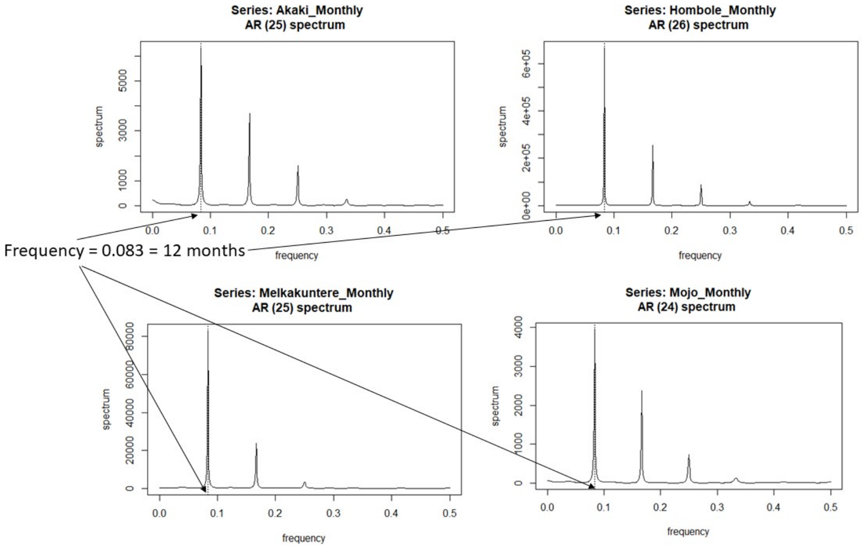


Figure 4.2: Periodograms of the four gauging stations showing the dominant frequency.

As observed in Figure 4.2, the dominant frequency of the flow series obtained at all four gauging stations is 12 months.

4.4.4 Seasonal character

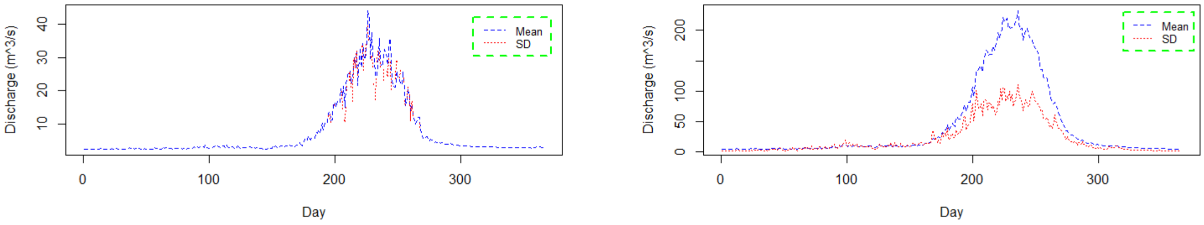
Seasonality in Mean, Standard Deviation and Coefficient of Variation

The seasonal characteristics of stream flow series can be comprehended by examining the annual variations in mean, standard deviation, coefficient of variation, and autocorrelation function (ACF) plots at daily, 1/3 monthly, and monthly levels.

In this study, daily series were utilized to explore the seasonal changes in mean, standard deviation, and coefficient of variation for the four gauging stations across UASB. A common time frame of 365 days was employed for the daily analysis, assuming that the inclusion of an extra day in leap years (366 days) would not significantly impact the interpretation of results. The daily data was organized in matrix form as depicted in Equation 4.19. Subsequently, Equations 4.20, 4.21, and 4.22 were applied to present the data, as illustrated in Figure 4.3.

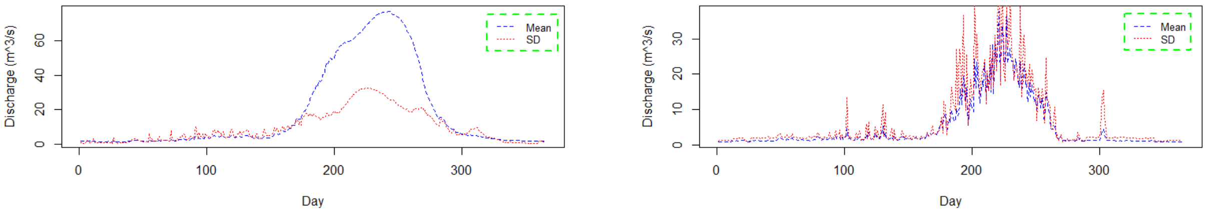
From Figure 4.3, it is evident that both the mean and standard deviation exhibit similar patterns across all four gauging stations. Specifically, they are higher in magnitude during the main wet season (JJAS) and lower in the dry seasons. The observed pattern in the flow from major rivers in the sub-basin is unimodal, corresponding directly to the unimodal characteristics of rainfall in the sub-basin. Additionally, higher variance in flow is noticeable at Akaki and Mojo rivers compared to the other two stations.

The coefficient of variation plots (Figures: 4.4 and 4.5) derived from the daily flow series of Hombole and Melkakuntere stations reveal higher variability in flow during the dry seasons. Conversely, Akaki and Mojo stations exhibit higher variability in flow during the wet seasons. This contrasting behavior in Akaki and Mojo stations during the wet season may be attributed to their smaller catchment sizes compared to the other stations. In basins with smaller sizes, the contribution of base flow relative to direct flow is often considered smaller. Additionally, a steeper elevation drop from the farthest point in the sub-basin to the gauge could contribute



(a) Akaki Station

(b) Hombole station



(c) Melkakuntere station

(d) Mojo station

Figure 4.3: Seasonal variation of daily mean and SD over years for the four gauging stations over UASB.

to this variation. For instance, at Akaki, there is approximately a 1171 m drop in elevation over a distance of 87.4 km, while for Hombole, the drop is 1195 m over a distance of 190 km.

Seasonality in Autocorrelation Structure

The matrix mechanism presented in Equation:4.19 is also employed here to manipulate the data and examine the seasonal variation of autocorrelation across the four gauge points. Two data processing steps, aimed at achieving normality assumption and stationarity, have been implemented. The initial operation involves transforming the data by either taking logarithms or applying Box-Cox transformation to conform to a normal distribution, as described by Equation:4.24. Subsequently, detrending and deseasonalizing the data are performed to attain stationarity. Detrending is executed by fitting a linear regression equation and subtracting the trend from each data point. The final step involves simple deseasonalization using standardization techniques. Two time frames, daily and monthly, are utilized, and autocorrelation coefficients are computed using Equation:4.23.

The autocorrelation plots in Figures:4.6 & 4.7 (a) and (b) depict autocorrelation for the log-transformed and deseasonalized daily time series. On the other hand, Figures:4.8 & 4.9 (a) and (b) illustrate autocorrelation plots for monthly series after detrending, log transformation, and deseasonalization.

Key observations from the daily and monthly autocorrelation function (ACF) plots include:

- Generally, the autocorrelation plots for the daily series (day-to-day autocorrelation) reveal that the dry season flow exhibits higher autocorrelation compared to the wet

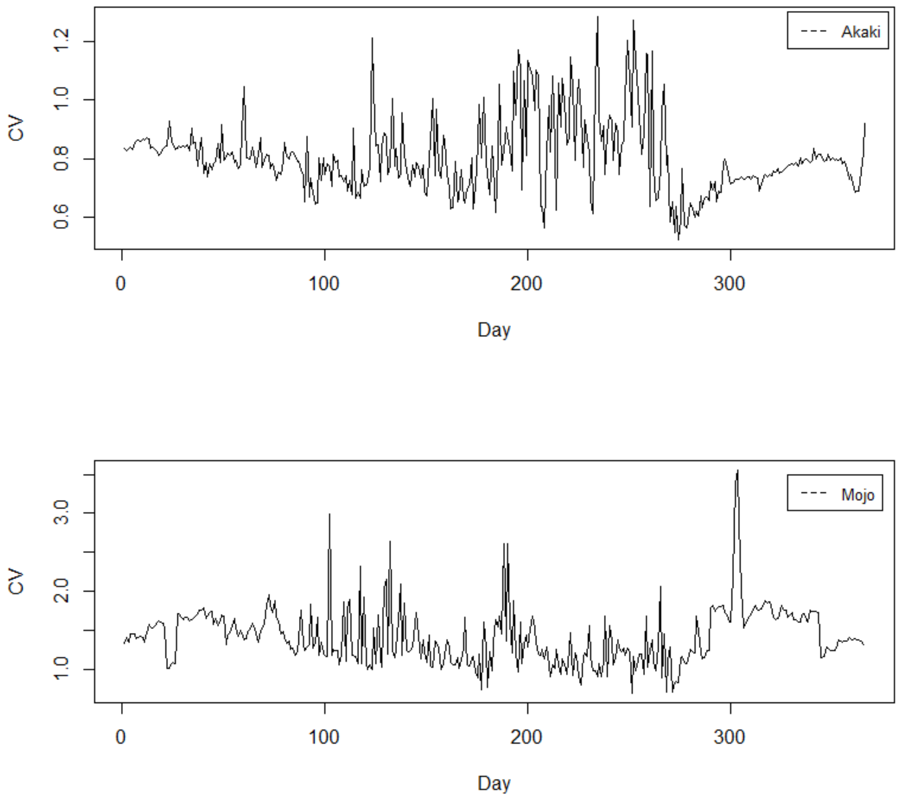


Figure 4.4: Seasonal variation of daily CV over years for Akaki and Mojo gauges.

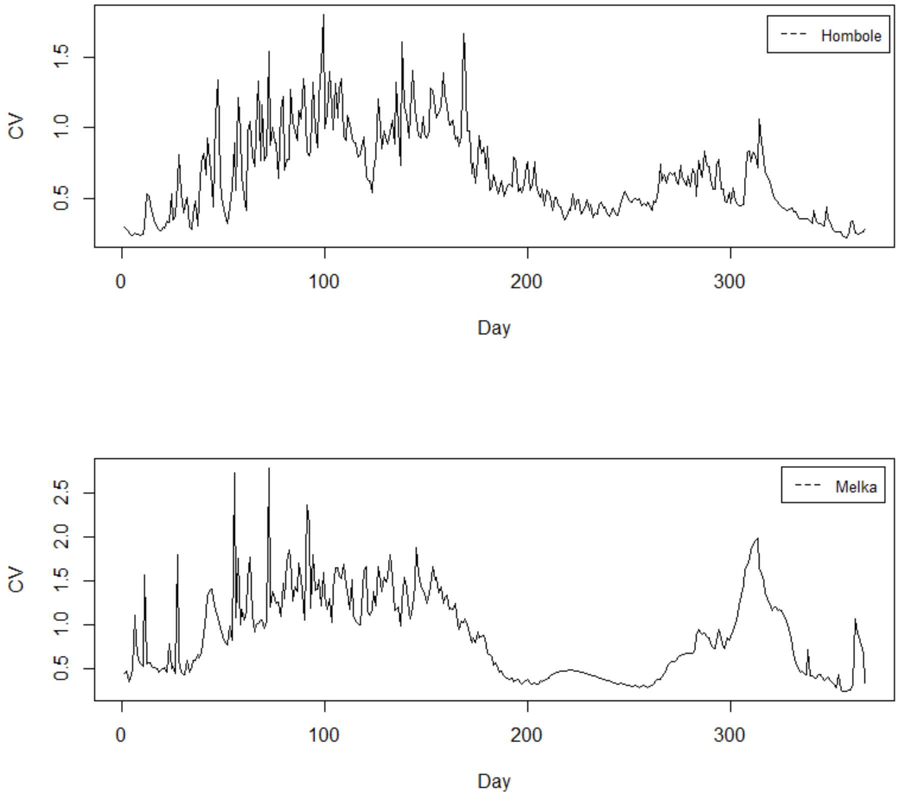
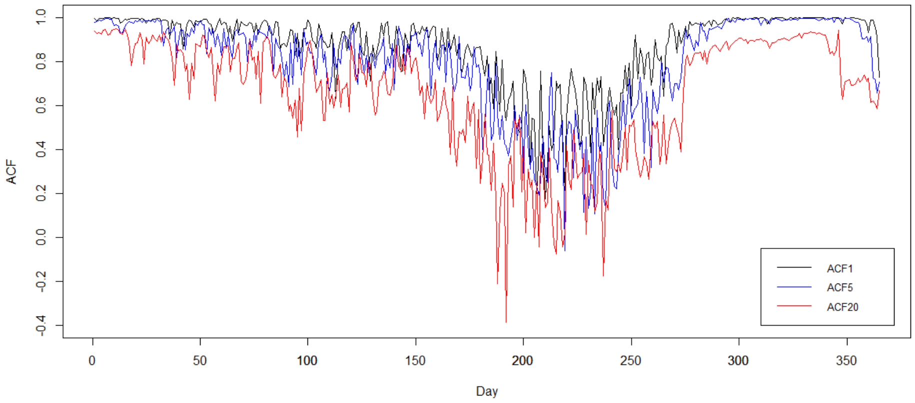
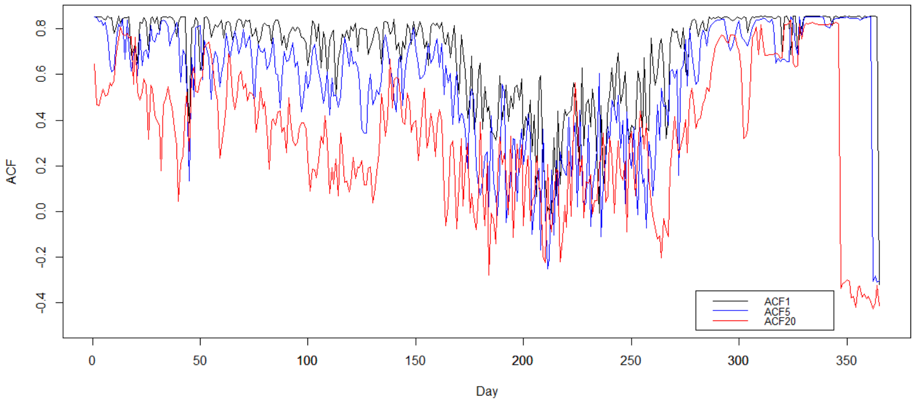


Figure 4.5: Seasonal variation of daily CV over years for Hombole and Melkakuntere.

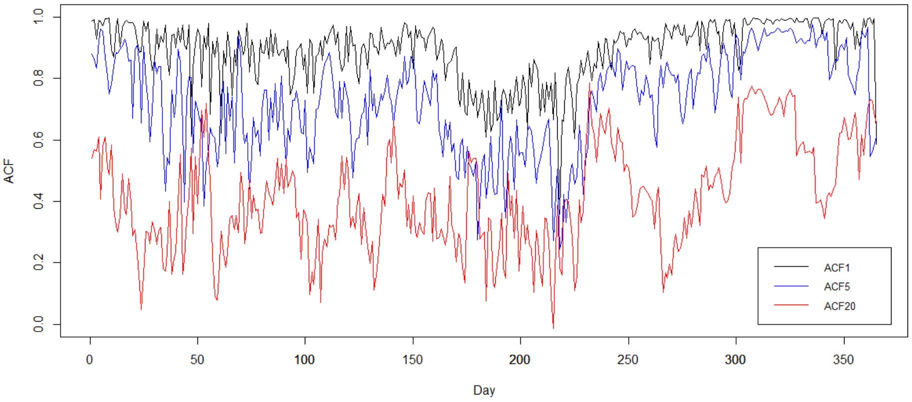


(a) Akaki Station

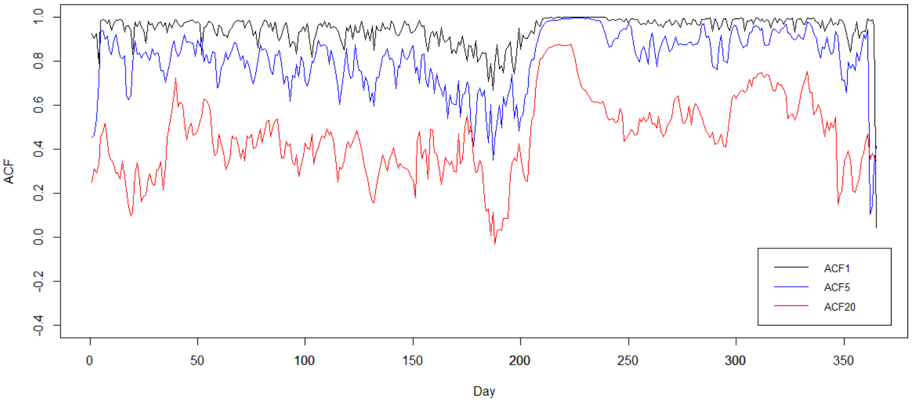


(b) Mojo Station

Figure 4.6: Seasonal variation of daily ACF over years for 1-day, 5-day, and 20-day lag for Akaki and Mojo Village.

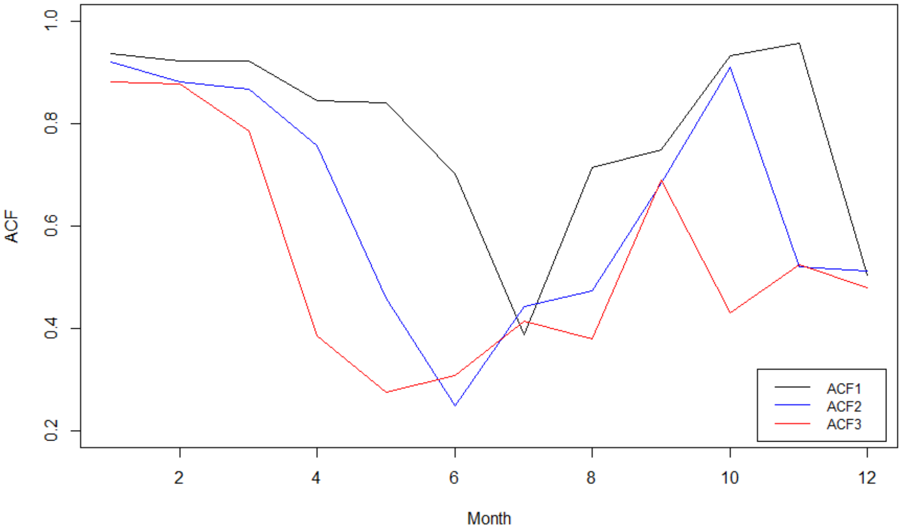


(a) Hombole Station

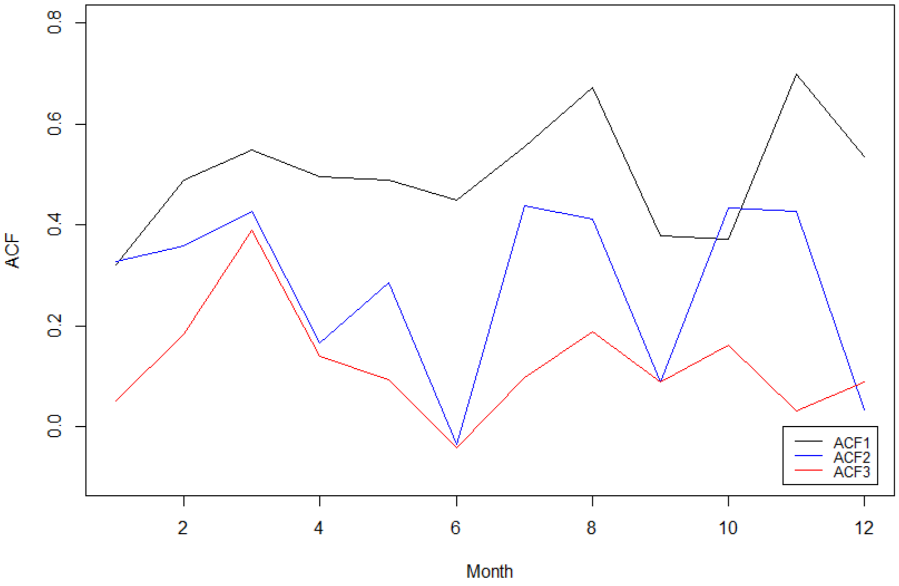


(b) Melkakuntere Station

Figure 4.7: Seasonal variation of daily ACF over years for 1-day, 5-day, and 20-day lag for Hombole and Melkakuntere.

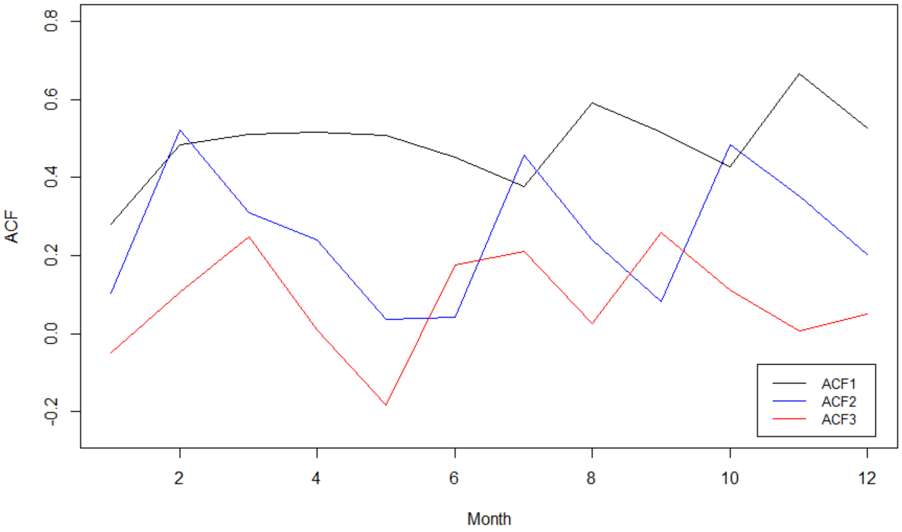


(a) Akaki Station

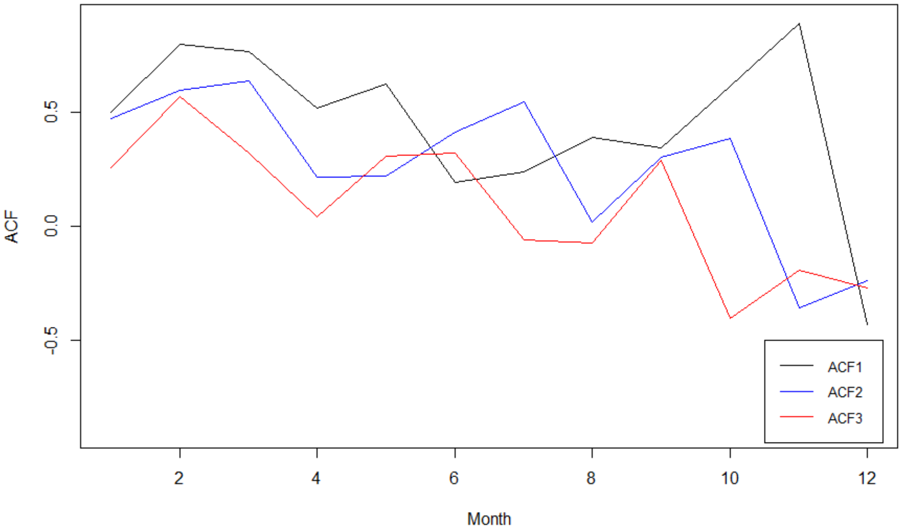


(b) Hombole Station

Figure 4.8: Seasonal variation of monthly ACF over years for 1,2, and 3 - month for Akaki and Hombole.



(a) Melkakuntere Station



(b) Mojo Station

Figure 4.9: Seasonal variation of monthly ACF over years for 1,2, and 3 - month for Melkakuntere and Mojo Village.

seasons. This pattern is particularly noticeable for Akaki@Akaki and Mojo@Mojo stations.

- The monthly series of Akaki@Akaki station mirrors the seasonal autocorrelation structure of the daily series. However, for Awash@Hombole and Awash@Melka stations, the autocorrelation structure differs notably from Akaki@Akaki station. In these two locations along the Awash river, higher autocorrelation is observed for the main rainy months (July and August) and the subsequent months especially for ACF plots of 1 and 2 month lags. Conversely, the short rainy months (April and May) exhibit relatively lower autocorrelation magnitude. For Mojo station higher positive autocorrelation was observed for Jan-Feb-Mar period and low autocorrelation values for subsequent months.

4.5 Discussion

In this chapter, we conducted various statistical analysis on the flow data obtained from each of the four gauge stations in the Upper Awash Sub-basin (UASB). The objective was to comprehend the characteristics of flow time series, gaining insights into the hydrological nature of the sub-basin and aiding any future time series modeling efforts. The types of analyses performed encompassed stationarity, linearity, trend, periodicity, and seasonal characteristics.

The outcomes of the KPSS and ADF stationarity tests conducted on flow data at various time intervals (daily, 1/3 monthly, monthly, and annual) all validate stationary characteristics. However, the same inference cannot be drawn from the linearity test performed using the BDS test. The results indicate that, except for the annual series, all other series exhibit non-linear dependence characteristics. This aligns with similar findings on the linearity characteristics of annual flow series confirmed by Wang et al. (2006) [85] in their studies of four major rivers worldwide.

The trend analysis was conducted using both monthly and annual flow series, revealing no clear correlation between the trend characteristics of each gauging station. Specifically, the analysis on the Hombole River indicates a significant increasing trend for the months of June and the dry months (October, November, and December). This finding aligns with the results reported for Hombole station by Daba et al. (2020) [60]. Similarly, significant increasing trends were observed for all months in the Akaki River except for July, August, and September (JAS). In contrast, no significant trend was observed in the Melkakuntere station across all time periods of analysis. The trend analysis for the Mojo station suggests an increasing trend in flows for the months of February, April, and August. However, it's important to note that the data series for Akaki and Mojo stations are relatively short, making it challenging to draw confident conclusions.

The results of periodic analysis conducted on the monthly flow series affirm the existence of an annual cycle. The dominant frequency observed at all gauging stations is a 12-month duration, with the subsequent dominant frequency being six months, as illustrated in Figure:4.2. This information serves as a crucial input for understanding the seasonal characteristics of the flow series.

The examination of seasonal characteristics in the rivers of the Upper Awash sub-basin involved two primary techniques. Firstly, an analysis of mean, standard deviation (SD), and coefficient of variation (CV) was conducted using daily flow data. Secondly, autocorrelation characteristics were explored using both daily and monthly data. All four plots (Figure:4.3) indicate that both mean and SD exhibit higher magnitudes during the same time periods, signifying their seasonal nature. Moreover, an analysis of the seasonal characteristics of daily flows using SD and CV confirms higher variability during the rainy season in rivers from smaller catchments like Akaki and Mojo. However, the CV plots of Hombole and Melkakuntere rivers demonstrate higher variability in flow during the dry and short rainy seasons. This could be attributed to the hydrological characteristics of these catchments. For instance, Akaki drains a relatively smaller catchment (956.74 km^2) with higher topographic variability, favoring a quick hydrologic response to recorded rainfall events at the outlet. In contrast, Melkakuntere and Hombole have catchment sizes of 1461.7 km^2 and 7693 km^2, respectively, with an elevation drop of 1171 and 1195 m. This, in turn, imparts a smoothing effect on the corresponding outflow recorded at the gauging site of these two rivers. Additionally, the higher flow variability in the short rainy season flow of these stations may be attributed to a smaller proportion of groundwater contribution, given the larger catchment size of Hombole and Melkakuntere compared to Akaki and Mojo.

Seasonality characteristics were also investigated through ACF plots at various daily and monthly intervals. In the daily analysis, calculations were performed for 1, 5, and 20 days, while for the monthly analysis, 1, 2, and 3 months were considered. Analysis at daily time steps revealed that dry season flow exhibited higher autocorrelation than wet seasons, especially for small river basins. Similar findings were observed at the monthly time step, with the exception of a one-month lag. This variation could be attributed to different factors within the basin influencing the runoff generation mechanism. During dry seasons characterized by less rainfall and increased groundwater contribution, river flows tend to be steadier, displaying a more continuous flow pattern, thus being more correlated through years.

4.6 Conclusion

Various statistical analyses were conducted on flow data extracted from major rivers in the Upper Awash Sub-basin (UASB). The results of the stationarity test, performed at different time intervals for all rivers, suggest that the flow series exhibit stationary characteristics. However, the linearity test is not straightforward, revealing linear dependency only in the annual flow series. The remaining time steps (daily, 10-day, and monthly) display nonlinear dependency in the data.

Seasonality analysis, employing the coefficient of variation (CV), revealed higher variability in the wet season for flows from smaller catchments like Akaki and Mojo. Conversely, for larger catchments such as Hombole and Melkakuntere, higher variability was observed during the dry season. Additionally, it was noted that dry season flows across all rivers exhibited higher correlation than wet season flows, with this characteristic being more pronounced in the flow series from smaller catchments like Akaki and Mojo.

Chapter 5

Time Series Modeling and Flow Forecasting

5.1 Introduction

The term ”forecasting” refers to the process of estimating future values of time-dependent data based on available current and past information. Forecasting can be categorized into different types depending on the application or technique employed. In terms of application, it is classified into three categories: short term (1 hour to 2 days), medium term (2-10 days), and long term (more than 10 days). Short-term forecasts are crucial for real-time flood forecasting, medium-term forecasts play a key role in hydro power generation, and long-term forecasts are essential for domestic water supply [113].

In terms of the technique applied, forecasting can be categorized as one-step or multi-step forecasting [114]. One-step (iterated) forecasting involves regressing past information of the independent and/or exogenous variable until the required forecast horizon is reached. A single trained model is used throughout the entire forecasting process. On the other hand, multi-step (direct) forecasting entails regressing a multi-period ahead value of the independent and/or exogenous variable on current and past values of the variable. The regressive model is updated at each iteration step, equivalent to a forecast horizon, with model parameters being updated for every forecast horizon due to varying training data in each iterative step. Additionally, the input data needs to be structured in a specific way depending on the required forecast horizon.

Future predictions regarding water availability in a given basin are crucial for effectively managing existing resources and planning for any future water resource development. Consequently, hydrological forecasting has attracted significant attention from both water resources planners and researchers. Models employed for hydrological forecasting can be broadly classified into univariate and multivariate types based on the number of variables or factors utilized in the forecasting process [115]. Univariate models involve a single variable, such as using precipitation to forecast streamflow, whereas multivariate models incorporate multiple variables or factors. Depending on the assumptions and techniques applied, hydrological forecasting can also be categorized into conventional and data-driven techniques. Conventional forecasting often adheres to the traditional Box and Jenkins (1970) approach for forecasting future values [116], while data-driven approaches heavily rely on various assumptions in the fields of statistics and machine learning (AI).

Conventional forecasting techniques, such as those applying the regression concept, necessitate the fulfillment of specific statistical assumptions in the raw data, such as stationarity, linearity, and normality [115]. One prominent example in this category is different versions of an Autoregressive Integrated Moving Average (Autoregressive Intergrated Moving Average (ARIMA)) model. Often, time series obtained from natural processes face challenges in meeting these assumptions. Consequently, data must undergo various manipulations to satisfy these conditions. Some of the data manipulations performed include transformations (e.g., taking logarithms to ensure data follows a normal distribution), standardization (to remove scale effects and enhance data stationarity), differencing (to induce stationarity in the data), and decomposition (to model each component of a time series separately) [117] [95] [118].

Data-driven approaches are particularly attractive due to their lower complexity, reduced processing time, a smaller number of inputs, and higher accuracy in streamflow forecasting compared to physically-based models [119] [120] [121]. However, it's important to note that there is no single data-driven model (Data Driven Models (DDM)) capable of working in all circumstances [122]. Instead, an effective DDM should strike a balance between system physics and data availability [123]. Various AI techniques widely recognized in this context include Artificial Neural Networks (ANN), Support Vector Machine (SVM), Fuzzy Logic approach, Evolutionary algorithms (Genetic Programming (GP), Genetic Algorithm (GA)), swarm intelligence algorithms (Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO)), and Wavelet Artificial Intelligence (W-AI) [122].

Several authors have demonstrated the effectiveness of these DDMs in simulating streamflow and generating future forecasts. These models can be employed independently or in combination with other methods based on the characteristics of the streamflow data and the desired level of accuracy. A recent study by Wang et al. (2009) [124] applied four AI techniques along with ARMA for long-term monthly streamflow forecasting in two rivers in China. The study concluded that, although each model exhibited different performance levels during training and validation, the top-performing models were Adaptive Network-based Fuzzy Inference System (ANFIS), SVM, and GP. However, there hasn't been a similar significant effort in the Upper Awash Basin, except for one by Edossa and Babel (2011) [125]. In their study, a three-layer neural network model with a backpropagation algorithm was utilized to forecast weekly streamflow in the Melka-Sedi River, Awash Basin, Ethiopia. A major limitation of this study is the absence of information regarding the appropriate ANN network design and selection.

The aim of this study is to assess the efficacy of various DDMs in forecasting the monthly flow series of selected rivers in the UASB. The flow characterization conducted in the previous chapter revealed that all four monthly flow series obtained from each river exhibit stationary and non-linear characteristics. Unfortunately, modeling such non-linear processes is not as straightforward as the Box-Jenkins approach (Box-Jenkins, 1970 [116]). In this study, two primary approaches have been proposed for forecasting the monthly flow series from three gauge stations in the sub-basin: Regime Switching Models and Data-Driven approaches (Artificial Neural Network (ANN) and ANFIS). As discussed earlier, there has been limited research evaluating the potential of different time series or data-driven techniques in the sub-basin.

5.2 Data Sets

The model fitting and time series simulation were primarily conducted using monthly flow time series data from three major rivers in the UASB: Akaki, Melkakuntere, and Hombole river. Additionally, areal climatic data, formed by a Thiessen polygon technique for each river catchment, was incorporated. Detailed information about the climate data and gauging stations is discussed in chapters 2, 3, and 4.

5.3 Methods

5.3.1 Regime Switching Models

Since traditional ARMA models or their different versions cannot be applied straightforwardly, and in order to capture the non-linear dependency of a monthly flow series, Regime Switching Models are suggested. Regime switching models divide the data series into a number of partitions called states (regimes) and attempt to fit an AR(p) model to each state. The data partitioning can be done by different techniques, and here, two of the most widely applied techniques, Threshold method and Markov chain technique, are discussed.

Threshold Auto-Regressive Models

This class of time series models has been popular since it was initially proposed by Tong [126]. Unlike the classical linear ARMA model, these models assume that a given time series consists of multiple regimes (states), and each of these individual regimes within a given time series can be fitted to an AR model. The regimes (states) are parts (components) of a time series, each having a different mean, variance, and auto-correlation character. A regime that occurs at a given time (t) is identified with the help of a threshold variable (r) as shown in Equation:5.1.

$$x_t = \begin{cases} \phi_{0,1} + \sum_{i=1}^{p} \phi_{i,1} x_{t-i} + \sigma_1 \epsilon_t, & if \quad x_{t-d} \leq r_1 \\ \phi_{0,2} + \sum_{i=1}^{p} \phi_{i,2} x_{t-i} + \sigma_2 \epsilon_t, & if \quad r_1 < x_{t-d} \leq r_2 \\ . \quad . \quad . \\ \phi_{0,m} + \sum_{i=1}^{p} \phi_{i,m} x_{t-i} + \sigma_m \epsilon_t, & if \quad r_{m-1} < x_{t-d} \end{cases} \tag{5.1}$$

Where:

x_t is m-regime TAR model in which m indicates the number of regimes (states).

x_{t-d} is the threshold variable which is set to make x_t shift from one state to another.

ϵ_t is iid (0,σ^2)

$\phi_{0,i}$ are real valued parameters for some i.

d is a positive integer denoting the delay ($d > 0$).

r_k is the threshold value at k^{th} state.

p is the order of the AR series.

The above equation can also be written as:

$$x_t = \phi_{0,j} + \phi_{1,j}x_{t-1} + \quad . \quad . \quad . \quad + \phi_{p,j}x_{t-p} + \sigma_j \epsilon_t, \quad if \quad r_{j-1} < x_{t-d} \le r_j \tag{5.2}$$

where: $j = 1, . \quad . \quad ., m$

In practice, if the type of threshold variable used is exogenous (i.e., a variable other than the time series to be modeled), then the model is known as TAR; otherwise, if it is endogenous, then it's called SETAR (Self-Exciting TAR). For example, in a TAR model, other variables (e.g., snowfall, soil moisture, reservoir release, etc.) might be used to express the point at which the mean fluctuates in the time series. But in the SETAR model, the threshold variable is defined from the time series to be modeled itself.

In addition to the type of threshold variable used, the way a time series is set to be switched from one regime to another can be controlled by using transition functions. The most commonly used transition functions, for example, are Logistic (Equation:5.3) and Exponential (Equation:5.4) functions. Variations of the TAR model that apply transition functions to define a threshold variable are known as Smooth TAR (STAR) models. This is because the transition from one regime isn't abrupt but rather smooth due to the nature of the applied transition functions.

$$G(S_t|\gamma, c) = \frac{1}{1 + exp(-\gamma(S_t - c))} \tag{5.3}$$

$$G(S_t|\gamma, c) = 1 - exp(-\gamma(S_t - c)^2), \quad \gamma > 0 \tag{5.4}$$

Where: S_t is the threshold variable and γ and c are scale and location (mean) parameters.

Therefore, depending on the type of threshold variable and transition functions, we may have different variations of TAR models. For example, LSETAR and ESETAR are sets of models that use endogenous threshold variables with logistic and exponential transition functions, respectively. A two-regime type equation that applies smooth transitions between regimes on a given time series can be represented by Equation:5.5.

$$\begin{aligned} x_t = (\phi_{0,1} + \phi_{1,1}x_{t-1} + \quad . \quad . \quad . \quad + \phi_{p,1}x_{t-P} + \sigma_1\epsilon_t)[1 - G(S_t|\gamma, c)] + \\ (\phi_{0,2} + \phi_{1,2}x_{t-1} + \quad . \quad . \quad . \quad + \phi_{p,2}x_{t-P} + \sigma_2\epsilon_t)G(S_t|\gamma, c) \end{aligned} \tag{5.5}$$

Where: $0 \le G(S_t|\gamma, c) \le 1$

Estimation of the AR parameters in each of the regimes is performed through a non-linear least squares method [127]. As a demonstration, parameter estimation for a two-regime TAR model is achieved through the minimization of the conditional least squares estimation objective function (Equation:5.6).

$$L_T(\Theta) = \sum_{t=p+1}^{T} [x_t - E_\Theta(x_t|F_{t-1})]^2 \tag{5.6}$$

Where: x_t is the time series to be modeled, $\Theta = (\phi'_{i,1}, \phi_{i,2}, r, d)$ are coefficient vector of equations describing the two regimes. In which $\boldsymbol{\phi_{i,1}} = (\phi_{0,1}, \phi_{1,1}, \quad . \quad . \quad ., \phi_{p,1})'$ and $\boldsymbol{\phi_{i,2}} = (\phi_{0,2}, \phi_{1,2}, \quad . \quad . \quad ., \phi_{p,2})'$, also $F_{t-1} = (x_{t-1}, x_{t-2}, \quad . \quad . \quad . \quad, x_1)$.

The minimization of this objective function is achieved through a minimization of $S(r, d)$, which is the minimum residual sum of squares of the model with a given fixed threshold r and delay d. To minimize $S(r, d)$, the sum of squares of residuals of the linear equation used in each regime is minimized. Furthermore, for a given r and d values, let $S_1 = \{t|x_{t-d} \leq r\}$ and $S_2 = \{t|x_{t-d} > r\}$ for the two regimes. Then, the parameters $\boldsymbol{\phi_{i,1}}$ and $\boldsymbol{\phi_{i,2}}$ are estimated by ordinary least squares linear regression of the following two equations.

$$x_t = \phi_{0,1} + \sum_{i=1}^{p} \phi_{i,1} x_{t-i} + e_t, t \in S_1 \tag{5.7}$$

$$x_t = \phi_{0,2} + \sum_{i=1}^{p} \phi_{i,2} x_{t-i} + e_t, t \in S_2 \tag{5.8}$$

In this study, the appropriate order and delay term have been identified with help of PACF plots and threshold nonlinearity test [128]. The threshold non-linearity helps to determine whether or not a given time series with certain order (p) and delay (d) has a non-linear character. After identifying the order and possible delay, a grid search algorithm under tsDyn package [129] in R is employed to identify the number of regimes and threshold value. Moreover, model parameters in each regime can also be estimated under this package.

Markov Switching Models

Markov Switching Models (MSMs) were introduced in the 1980s by James Hamilton [130] in his paper "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle" published in the Journal of Economic Dynamics and Control in 1989. Hamilton proposed the idea of allowing the parameters of a time series model to switch between different regimes or states over time, rather than assuming that they remain constant. This allows MSMs to capture the non-stationarity and regime-specific behavior of time series data, making them useful for modeling economic and financial time series, among other applications. Since their introduction, MSMs have become widely used in economics, finance, and other fields.

For a given time series x_t a two state Markov Switching Model can be written by the following equation;

$$x_t = \begin{cases} \phi_{0,1} + \sum_{i=1}^{p} \phi_{i,1} x_{t-i} + \sigma_1 \epsilon_t, & if \quad S_t = 1 \\ \phi_{0,2} + \sum_{i=1}^{p} \phi_{i,2} x_{t-i} + \sigma_2 \epsilon_t, & if \quad S_t = 2 \end{cases} \tag{5.9}$$

Where: S_t represents the state of the process at a time t, $\phi_{i,j}$ are real numbers, $\sigma_i > 0$ and ϵ_t is iid $\sim (0, \sigma^2)$.

It can be observed that the two-state MSM (Equation:5.9) is very similar to a TAR model, with the only exception being how the state transition is defined. In MSMs, the state transition occurs through transition probabilities (Equation:5.10) defined based on probabilities calculated from the data, i.e., from past information.

$$P(S_t = 1|S_{t-1} = 1) = p_{11}, \qquad P(S_t = 2|S_{t-1} = 2) = p_{22} \tag{5.10}$$

Where: $0 < \eta_j < 1$.

The above probability terms can also be written as a probability (transition) matrix as follows:

$$\begin{bmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{bmatrix} \tag{5.11}$$

In general a k-state Markov switching autoregressive model can be written as:

$$x_t = \phi_{0,S_t} + \sum_{i=1}^{p} \phi_{i,S_t} x_{t-i} + \sigma_{S_t} e_t, Where: S_t \epsilon (1, . \quad . \quad . \quad , k) \tag{5.12}$$

The probability of switching to a given state (S_t) at time t (ξ_{jt}) is calculated by application of Baye's rule as follows:

$$\xi_{jt} = \frac{\xi_{1t-1} p_{1j} f(R_t | S_t = j, \Im_{t-1}; \mathbf{\Theta}) + \xi_{2t-1} p_{2j} f(R_t | S_t = j, \Im_{t-1}; \mathbf{\Theta})}{f(R_t | \Im_{t-1}; \mathbf{\Theta})} \tag{5.13}$$

Where: The $\xi_{j_{t-1}}$ term is the unconditional probability term, p_{ij} are the components of transition probability matrix in Equation:5.11, and $f(R_t | S_t = j, \Im_{t-1}; \mathbf{\Theta})$ is the conditional probability density of an observation at time t(R_t) as shown in Equation:5.14 below. And the $\Im_{t-1}$ expresses all the past information available up to $t-1$ and $\mathbf{\Theta}$ is the vector holding the parameters of the model to be estimated.

$$\begin{aligned} f(R_t | \Im_{t-1}; \mathbf{\Theta}) = Pr(R_t | \mathbf{\Theta}) = \xi_{1t-1} p_{11} f(R_t | S_t = 1, \Im_{t-1}; \mathbf{\Theta}) + \\ \xi_{1t-1} p_{12} f(R_t | S_t = 2, \Im_{t-1}; \mathbf{\Theta}) + \\ (1 - \xi_{1t-1}) p_{21} f(R_t | S_t = 1, \Im_{t-1}; \mathbf{\Theta}) + \\ (1 - \xi_{1t-1}) p_{22} f(R_t | S_t = 2, \Im_{t-1}; \mathbf{\Theta}) \end{aligned} \tag{5.14}$$

$$f(R_t | S_t = j, \Im_{t-1}; \mathbf{\Theta}) = \frac{1}{\sqrt{2\pi\sigma}} exp \left[- \frac{(R_t - \mu_j - \phi_{1j} R_{t-1} - \phi_{2j} R_{t-2} -, \quad ... \quad , -\phi_{pj} R_{t-p})^2}{2{\sigma_j}^2} \right] \tag{5.15}$$

The parameters of the MSM model ($\Theta = [\phi_{ij}, \sigma_j]$) are estimated using the past information obtained from the available data using the MLE (Maximum Likelihood Estimation) technique. For this the initial estimate of $\xi_{1t-1} = \xi_{10}$ is required to initialize the maximization algorithm.

$$\begin{aligned} \ln f(R_1, R_2, ..., R_T | R_0; \theta) = \sum_{t=1}^{T} \ln f(R_t | \Im_{t-1}; \mathbf{\Theta}) = \\ \sum_{t=1}^{T} \sum_{i=1}^{2} \sum_{j=1}^{2} \xi_{it-1} p_{ij} \frac{1}{\sqrt{2\pi\sigma}} exp \left[- \frac{(R_t - \mu_j - \phi_{1j} R_{t-1} - \phi_{2j} R_{t-2} -, \quad ... \quad , -\phi_{pj} R_{t-p})^2}{2{\sigma_j}^2} \right] \end{aligned} \tag{5.16}$$

5.3.2 Data Driven Techniques

In this study, two widely known data-driven techniques, namely the ANN and ANFIS models, are employed to simulate the flow series of three rivers in UASB. The main reason for proposing the application of these methods is their high performance and simplicity in modeling the non-linear characteristics present in monthly flow series.

Artificial Neural Network (ANN)

The concept of artificial neural networks (ANN) originated approximately 80 years ago with McCulloch and Pitts (1943), drawing inspiration from their efforts to understand the functioning of the human brain. Neural networks are comprised of individual building blocks called nodes (neurons), featuring links that transfer signals between nodes, adjustable parameters (weights) assigned to each link, and an activation (transfer) function determining the output from each node. These models adhere to specific rules governing computation progression and adaptation (learning) [131].

Each ANN can be custom-tailored to the specific problem at hand, characterized by its architecture, including the connection between nodes, weight determination, and type of activation function applied. The architecture may be defined based on the number of layers, nodes in each layer, and the direction of information flow. For example, Figure:5.1 illustrates a three-layer ANN architecture. In terms of information flow and structure, ANNs can be categorized as feedforward or recurrent network types. In a feedforward network, information flows from the input layer to each node through the hidden layer and finally to the output layer. Conversely, in a recurrent network, information can also travel backward by taking the output of the previous forward run as the new input. Nodes within a given layer can be linked based on requirements. The number of layers and nodes is typically adjusted through a trial-and-error approach.

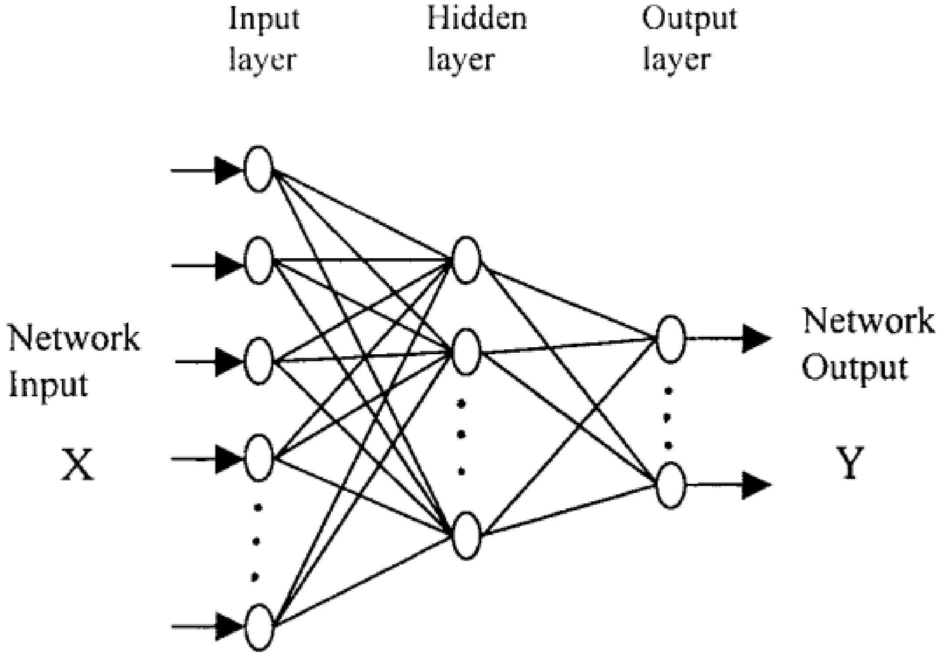


Figure 5.1: Three layer Feed-forward ANN architecture with X indicating input and Y as output from the system (Source: ASCE paper on ANN [2]).

The way at which an input vector to each node is converted to an output can be seen in the Figure:5.2. Let $\mathbf{X} = \mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ be the input vector to a certain node (j) and the weights of each link be as $\mathbf{W_j} = \mathbf{w_{1j}}, \mathbf{w_{2j}}, ..., \mathbf{w_{nj}}$ then the output y_j can be obtained by the following equation:

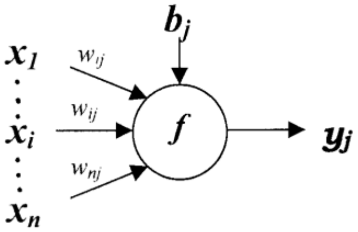


Figure 5.2: ANN at node level (Source: ASCE paper on ANN [2]).

$$y_j = f(\mathbf{X} \cdot \mathbf{W_j} - \mathbf{b_j}) \tag{5.17}$$

Where: f is the transfer (activation) function and b_j is the bias term assigned to each node.

The output from each node is influenced by the type of transfer (squashing) function used, typically reflecting the response of a node to an input signal. This function ensures that node values fall within the range of 0 and 1. The Sigmoid function is the most common type of transfer function and is defined as follows:

$$f(t) = \frac{1}{1 + e^{-t}} \tag{5.18}$$

The primary reason why this equation is the preferred choice lies in certain mathematical properties that make it highly suitable for modeling non-linear relationships. This function possesses bounded, monotonic, and non-decreasing properties, and its derivatives can be easily defined. Other applied transfer functions include bipolar sigmoid and hyperbolic tangent functions.

The term "Model Training" refers to the process through which weights and biases in an ANN model are adjusted. Typically, the model training process involves Feed-Forward and Back-Propagation steps. In the Feed-Forward step, assumed initial weights and biases, along with the input series, are introduced to the network system. Initial output series is generated with these assumed weights and biases. Subsequently, in the Back-Propagation step, an optimization algorithm is employed to refine these assumed weights and biases. This process, operating backward from the output layer to the hidden layer, aims to optimize the model coefficients. The steps involved in setting these model coefficients are briefly explained using a simple three-layer ANN depicted in Figure:5.3.

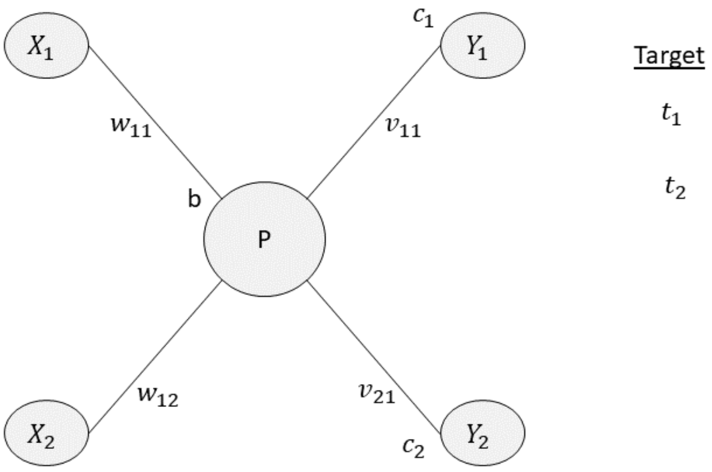


Figure 5.3: Three layer ANN composed of two inputs variables ($X_1 \quad and \quad X_2$), one hidden layer node, two outputs ($Y_1 \quad and \quad Y_2$), weights (w_{11},w_{12},v_{11},v_{21}),biases (b,c_1,c_2), and targets (t_i).

Let's assume its the weight term (v_{ij}) to be set, here i indicates the node of output layer and j is for hidden layer, then

1. Using the initial weights and bias the simulation is performed and the Y_i are estimated.
2. The simulations Y_i are compared with targets t_i using a cost function (C_0) such as Sum of Squared Residuals (SSR) as shown in Eq.5.19.

$$C_0 = \sum_{i=1}^{2} (t_i - Y_i)^2 \tag{5.19}$$

3. Using the cost function as an objective function and v_{ij} a minimization optimization is defined as follows:

$$\frac{\partial C_0}{\partial v_{ij}} = \frac{\partial C_0}{\partial Y_i} \frac{\partial Y_i}{\partial Z_i} \frac{\partial Z_i}{\partial v_{ij}} \tag{5.20}$$

where: $Z_i = Pv_{ij} + c_i$ and $Y_i = f(Z_i)$. Here, the $f(.)$ term indicates the activation function used. Therefore, the partial differential equation above when evaluated comes down to as;

$$\frac{\partial C_0}{\partial v_{ij}} = \sum_{i=1}^{2} 2(t_i - Y_i) f'(Z_i) P \tag{5.21}$$

4. Using the Equation:5.21 which indicates the sensitivity of cost function to the variation of v_{ij}, the previous value of v_{ij} can be updated by the following relationship,

$$v_{ij_{new}} = v_{ij_{old}} - SS \tag{5.22}$$

$$SS = \left(\frac{\partial C_0}{\partial v_{ij}}\right) LR \tag{5.23}$$

Where: SS - is to mean step size and LR is Learning Rate to be defined so that to accelerate the search in parameter space as shown in Figure:5.4.

5. Using the updated $v_{ij_{new}}$, a new value for Y_i is predicted there by updating equation 5.19.
6. Steps 4 and 5 are repeated until local minima in the search surface is obtained.

The described steps outline how the weight of links to one of the output nodes is adjusted. Similar steps are executed for other weights and bias terms in the network. As mentioned earlier, the technique discussed in the above steps is known as Backpropagation. Additionally, there are several algorithms that perform similar tasks, employing the concept of backpropagation or similar assumptions. Some of these techniques include Conjugate Gradient, Radial Basis Function, Recurrent Neural Networks, etc.

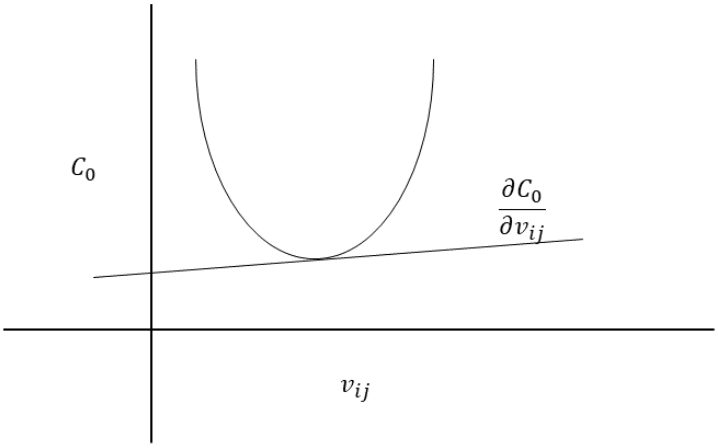


Figure 5.4: Hypothetical search surface for the optimal (minimum) point by decreasing the search gradient ($\frac{\partial C_0}{\partial v_{ij}}$).

In this study, a special type of ANN model called Non-linear Auto-Regressive with eXogenous inputs (NARX) has been applied due to its good performance in simulating long-term dependencies present in a time series [132]. This model has shown promising results in simulating univariate time series with the help of a delay term together with a feedback loop mechanism in the network [133] [134].

Forecasting of future flow series is performed with the help of exogenous variables like rainfall, temperature, evapotranspiration, etc. So, the equation with rainfall as an exogenous variable can be given as:

$$\bar{F}(t+k) = \Phi_o(F(t), F(t-1), \quad . \quad . \quad ., F(t-d_F), R(t), R(t-1), \quad . \quad . \quad ., R(t-d_R)) \tag{5.24}$$

Where: $\bar{F}(t+k)$ is forecast for k time step ahead, Φ_o is a transfer function, $F(t), F(t-1), \quad . \quad . \quad ., F(t-d_F)$ are lagged flow series, and $R(t-1), \quad . \quad . \quad ., R(t-d_R)$ are lagged rainfall series.

The network architecture for the NARX model is shown in the figure below with a feedback loop.

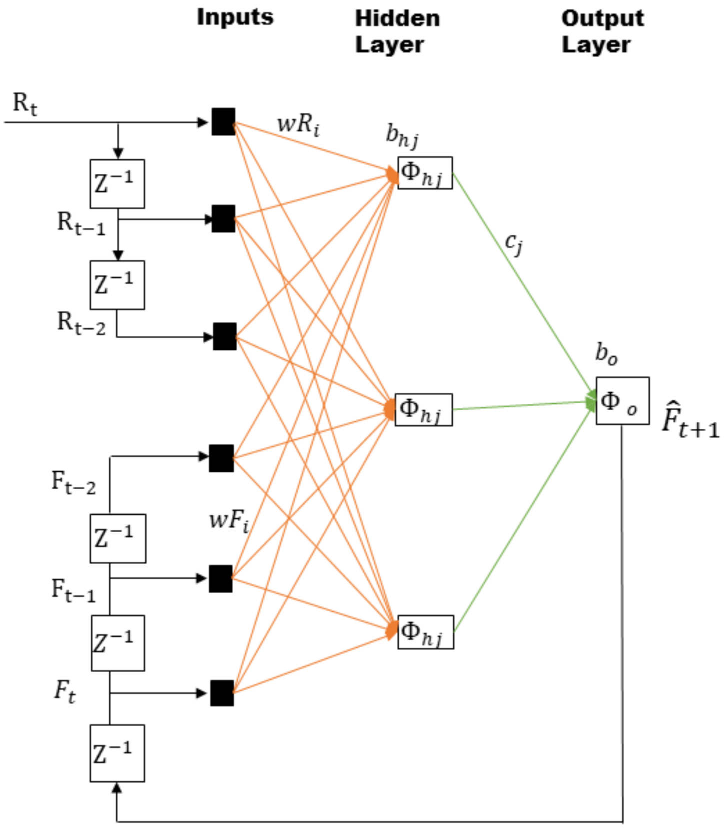


Figure 5.5: NARX model with two inputs and delay =2 and an output of one-time step ahead forecast

The forecasted flow for a one-time step ahead (t+1) and for j = 1 to 3, which is equal to number of hidden nodes.

$$\bar{F}(t+1) = \Phi_o(b_o + \sum_{j=1}^{3} c_j O_{hj}) \tag{5.25}$$

Where; b_o is the bias of the output node, c_j is the weight of connection between hidden and output node, and O_{hj} is output of the hidden node j.

The output from any hidden node j is calculated as shown below;

$$O_{hj} = \Phi_{hj}(b_{hj} + \sum_{i=0}^{2} wR_i R(t-i) + \sum_{i=0}^{2} wF_i F(t-i)) \tag{5.26}$$

Where: ϕ_{hj} is the transfer function at hidden node, b_{hj} is the bias term at hidden nodes, wR_i is weights for rainfall delay terms, wF_i is weights for flow delay terms, $R(t-i)$ is rainfall value for a previous time steps based on the delay, and $F(t-i)$ is flow value for a previous time steps based on the delay.

Adaptive Neuro-Fuzzy Inference System (ANFIS) Model

Another class of data-driven models with greater potential for simulation is known as the ANFIS model, first introduced by Shing and Jang in 1993 [135]. This technique merges the network adaptation mechanisms of the neural network concept with fuzzy inference to create an efficient predictive model. Unlike the neural network discussed above, here the links used to connect each node don't hold weights. In ANFIS, nodes are of two types: adaptive and fixed nodes. The difference between them is that adaptive nodes hold parameters, whereas fixed nodes don't. For easy distinction, square-shaped nodes are often used for adaptive nodes, while circles are used for the fixed nodes (Figure:5.6). The training inputs to the model are fuzzified by assigning an appropriate fuzzy membership function.

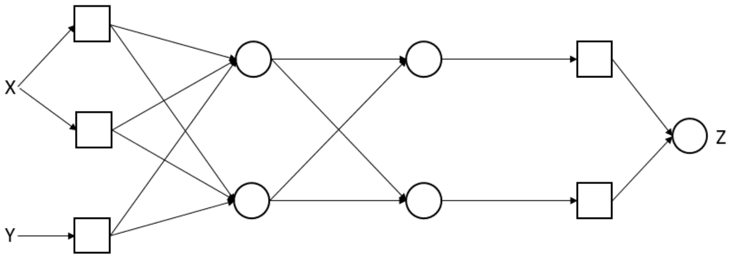


Figure 5.6: Adaptive Network with two inputs (X and Y) and outputs (Z) .

Fuzziness, used to express something that is not clear or vague, is a concept in mathematics used to define sets that don't follow the classical bivalent logic in set theory. In classical bivalent set theory (Crisp Set Theory), the membership of a given number in a set is assigned as true or false, with nothing in between. However, in Fuzzy Set Theory, the membership of a given value in a set is determined based on the degree of closeness. The definition of closeness relies on an appropriate distance measure, such as Euclidean distance, etc. Both cases can be further illustrated with the help of Figure:5.7.

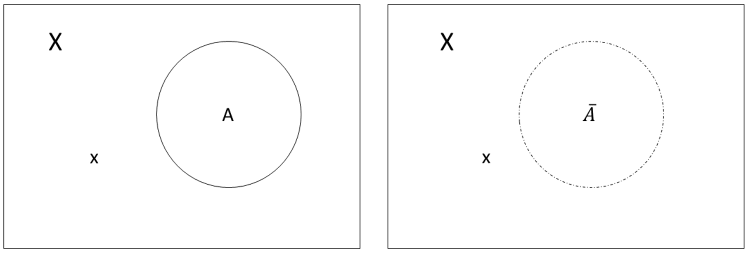


Figure 5.7: Universal set (X) which is a real number holding value x, Crisp Set A (Left), and Fuzzy Set ($\bar{A}$) (right) .

As shown in Figure:5.7, in Crisp Set Theory, the membership of x in set A ($\mu_A(x)$) could be expressed as follows:

$$\mu_A(x) = \begin{cases} 1, & \text{if } x \epsilon A \\ 0, & \text{otherwise} \end{cases} \tag{5.27}$$

However, in Fuzzy Set Theory, the membership of x in set A is defined by values in between, including 0 and 1 (hence the hidden lines in the boundary of the fuzzy set $\bar{A}$) in Figure:5.7) as shown by the mathematical expression in Equation:5.28.

$$\bar{A} = \left\{ (x, \mu_{\bar{A}}(x)) : x \epsilon X; \mu_{\bar{A}}(x) \epsilon [0, 1] \right\} \tag{5.28}$$

Here, the $\mu_{\bar{A}}(x)$ which shows the membership of value x and lying between 0 and 1, is termed as membership function. This membership function is a function that expresses the degree of membership of a given set of values to a given set A. This function is often represented with a convex-shaped function, as shown in Figure:5.8.

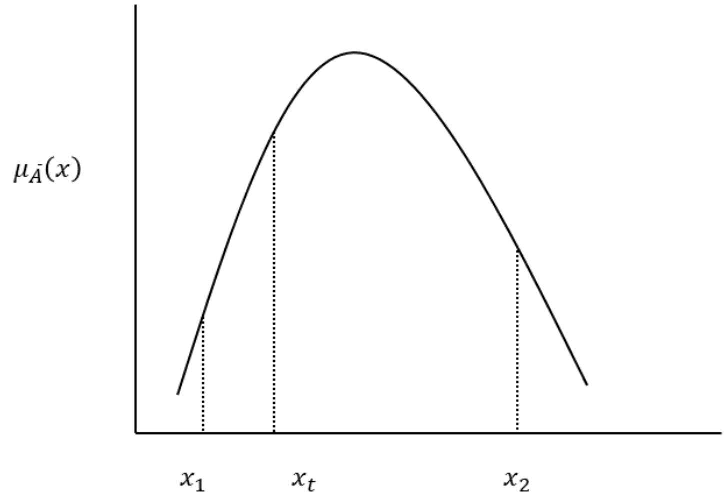


Figure 5.8: Function of a Convex fuzzy set.

Where; $\mu_{\bar{A}}(x_t) \geq min(\mu_{\bar{A}}(x_1), \mu_{\bar{A}}(x_2))$
Properties of Membership Function

The membership function for the fuzzy set $\bar{A}$ depicted in Figure:5.9 consists of three components: Core $(x|\mu_{\bar{A}}(x) = 1)$, Boundary $(x|0 < \mu_{\bar{A}}(x) < 1)$, and Support $(x|\mu_{\bar{A}}(x) > 0)$.

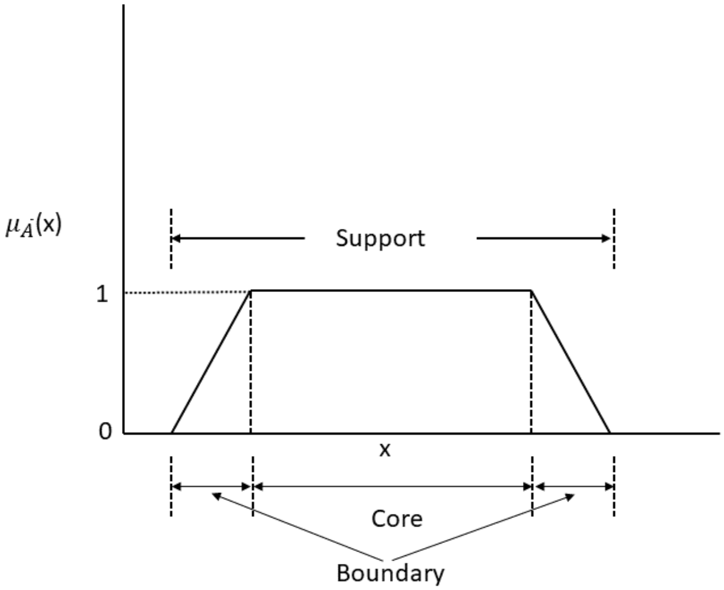


Figure 5.9: Components of membership function for a given normal fuzzy set $\bar{A}$.

Standard Operations of fuzzy sets

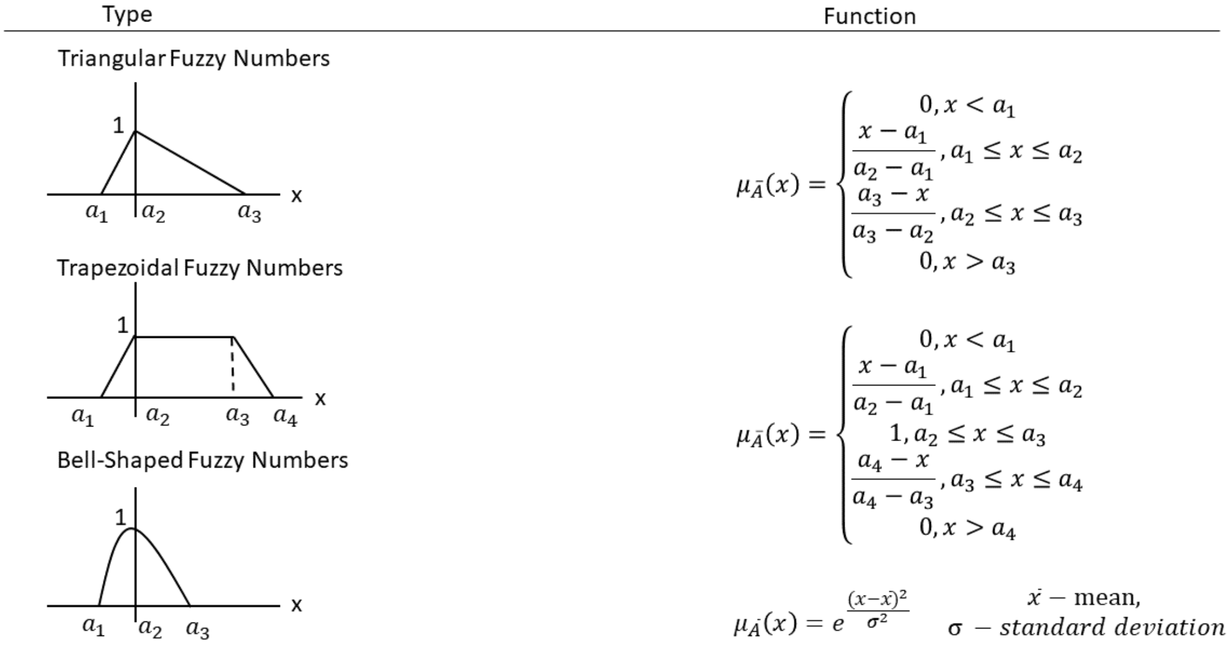
i. Complement: $\mu^{c}_{\bar{A}}(x) = 1 - \mu_{\bar{A}}$

ii. Union: for two fuzzy sets $\bar{A}$ and $\bar{B}$, their union can be expressed as,
   $\bar{C} = \bar{A} U \bar{B}$
   $\mu_{\bar{C}}(x) = max(\mu_{\bar{A}}(x), \mu_{\bar{B}}(x))$

iii. Intersection: and the for the same fuzzy sets as in ii.
   $\bar{C} = \bar{A} n \bar{B}$
   $\mu_{\bar{C}}(x) = min(\mu_{\bar{A}}(x), \mu_{\bar{B}}(x))$

Fuzzy Numbers

Depending on the application or condition, there may be a need for different types of fuzzy numbers. The most common ones are triangular, trapezoidal, and bell-shaped fuzzy numbers (Figure:5.10).



$$\mu_{\bar{A}}(x) = \begin{cases} 0, x < a_1 \\ \dfrac{x-a_1}{a_2-a_1}, a_1 \le x \le a_2 \\ \dfrac{a_3-x}{a_3-a_2}, a_2 \le x \le a_3 \\ 0, x > a_3 \end{cases}$$

$$\mu_{\bar{A}}(x) = \begin{cases} 0, x < a_1 \\ \dfrac{x-a_1}{a_2-a_1}, a_1 \le x \le a_2 \\ 1, a_2 \le x \le a_3 \\ \dfrac{a_4-x}{a_4-a_3}, a_3 \le x \le a_4 \\ 0, x > a_4 \end{cases}$$

$$\mu_{\bar{A}}(x) = e^{\frac{(x-\dot{x})^2}{\sigma^2}} \qquad \dot{x} - \text{mean}, \quad \sigma - \textit{standard deviation}$$

Figure 5.10: Different types of fuzzy numbers and their mathematical notations.

Defuzzification

Once the fuzzy operations have been carried out, it may be necessary to transform them into the actual variable values that the fuzzy membership values entail. This process of converting the fuzzy membership values into their corresponding crisp values is known as defuzzification. The most common types of defuzzification techniques are explained below.

1. Maximum Membership Principle

   Let's assume $\bar{C}$ is a fuzzy set obtained after fuzzy operations are done on two fuzzy sets. Then Z^* is a value obtained by taking the corresponding z value at a point of maximum membership value (Figure:5.11).

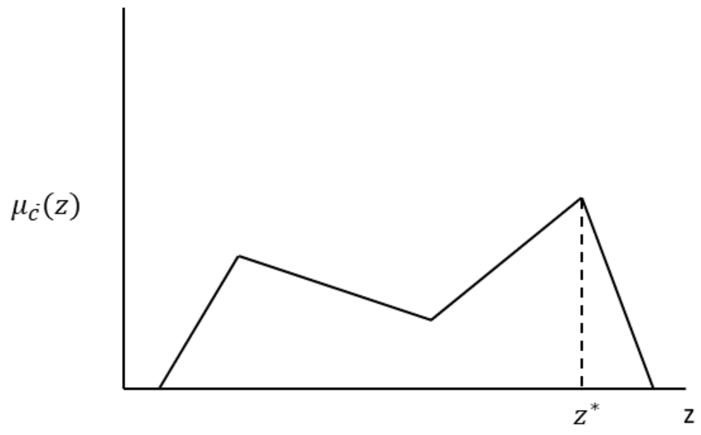


Figure 5.11: Defuzzification of fuzzy membership function for a given fuzzy set $\bar{c}$ using Maximum Membership Principle

2. Centroid Method

   This technique works by finding the centroid of the fuzzy membership function curve. Therefore, for the same figure in Figure:5.11, the equation can be written as:

$$z^* = \frac{\int \mu_{\bar{c}}(z) z dz}{\int \mu_{\bar{c}}(z) dz} \tag{5.29}$$

3. Weighted Average Method

   This one works for those which has symmetrical membership function (Figure:5.12) and is given as follows:

$$z^* = \frac{\sum \mu_{\bar{c}}(\bar{z}_i)\bar{z}_i}{\sum \mu_{\bar{c}}(\bar{z}_i)} \tag{5.30}$$

   Where: $\bar{z}_i$ is the centroid of each symmetric membership function.

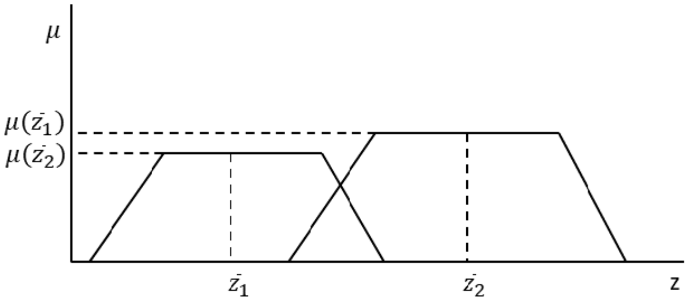


Figure 5.12: Defuzzification of fuzzy membership function for a given fuzzy set $\bar{c}$ using Weighted Average Method

4. Mean-Max Method

   This is also known as the middle of maximum and the defuzzified value of variable Z^* is given by:

$$Z^* = \frac{a+b}{2} \tag{5.31}$$

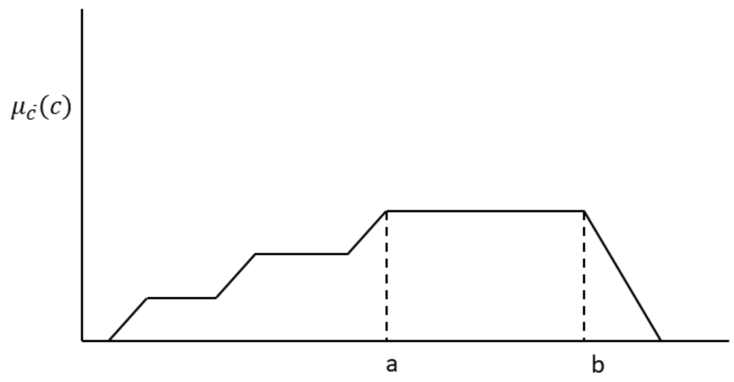


Figure 5.13: Defuzzification of fuzzy membership function for a given fuzzy set $\bar{c}$ using Mean-Max Method

Fuzzy Rule Based System

This is where the application of fuzzy sets comes into solving real-world problems where there is a cause-and-effect phenomenon present between two or more variables. For example, the relationship between temperature and pressure, rainfall and runoff, etc. Such kinds of relationships can be expressed with the following conditional expression:

$$\text{If} \quad \textit{Antecedent} \quad \text{Then} \quad \textit{Consequent} \tag{5.32}$$

Here, the antecedent part expresses the cause variable, and the consequent part takes the effect. Depending on the condition of variables, both the antecedent and consequent can be expressed with fuzzy membership functions, and such kinds of expressions are known as fuzzy rule-based systems. Using the above notation, a conditional statement for a scenario consisting of two input variables (X and Y) and k rules can be written as follows:

$$\begin{array}{ccccccccccc}
\text{If} & X & is & \bar{A_1}^1 & and & Y & is & \bar{B_1}^1 & \text{Then} & Z & is & \bar{C}^1 \\
\text{If} & X & is & \bar{A_2}^2 & and & Y & is & \bar{B_2}^2 & \text{Then} & Z & is & \bar{C}^2 \\
 & & & & & & & & & & & . \\
 & & & & & & & & & & & . \\
 & & & & & & & & & & & . \\
\text{If} & X & is & \bar{A_n}^k & and & Y & is & \bar{B_n}^k & \text{Then} & Z & is & \bar{C}^k
\end{array} \tag{5.33}$$

Where: X and Y are inputs and whereas Z is an output variable. In addition, $\bar{A_n}^k$ and $\bar{B_n}^k$ are the antecedent side n^{th} linguistic label for the k^{th} rule of input variables. Similarly, $\bar{C}^k$ is the consequent side k^{th} linguistic label of the output variable. Here, the linguistic label identifies which category of data range is used. For example, the entire data range may be categorized into low, medium, and high. Therefore, the data will have three linguistic labels.

In the above conditional statement (Equation:5.33), the consequent statements are connected by "AND," which is a conjunction condition. However, depending on the type of problem, "OR" (disjunctive) could also be used. For these two types of logical connectives, the value of membership functions for the consequent side is evaluated, as shown by Equations:5.34 and 5.35.

$$\begin{gathered}\bar{C}^k = \bar{A_n}^k \quad and \quad \bar{B_n}^k \quad and \quad . \quad . \quad . \\ \mu_{\bar{C}^k}(X) = min\Big[\mu_{\bar{A_n}^k}(X), \mu_{\bar{B_n}^k}(X), \quad . \quad . \quad .\Big]\end{gathered} \tag{5.34}$$

$$\begin{gathered}\bar{C}^k = \bar{A_n}^k \quad or \quad \bar{A_n}^k \quad or \quad . \quad . \quad . \\ \mu_{\bar{C}^k}(X) = max\Big[\mu_{\bar{A_n}^k}(X), \mu_{\bar{A_n}^k}(X), \quad . \quad . \quad .\Big]\end{gathered} \tag{5.35}$$

Graphical Techniques of Fuzzy Inference

There are three common types of fuzzy rule inference systems that have been widely used to solve various kinds of practical problems. The major differences between these inference techniques are mainly in the definition of rules on the consequent side and the type of defuzzification method.

Mamdani Inference System (Type - II)

Let's consider two input variables, X and Y, with corresponding linguistic labels $\bar{A}^K$ and $\bar{B}^K$, two rules, and an output (Z) also with linguistic labels $\bar{C}^K$ as follows:

$$\begin{aligned} Rule-1 \quad &: \quad \text{If} \quad X_1 \quad is \quad \bar{A_1}^1 \quad and \quad Y_1 \quad is \quad \bar{B_1}^1 \quad Then \quad Z_1 \quad is \quad \bar{C}^1 \\ Rule-2 \quad &: \quad \text{If} \quad X_2 \quad is \quad \bar{A_2}^2 \quad and \quad Y_2 \quad is \quad \bar{B_2}^2 \quad Then \quad Z_1 \quad is \quad \bar{C}^2 \end{aligned} \tag{5.36}$$

The steps followed by these techniques are summarized and illustrated in Figure:5.14. It can be observed that the minimum membership function value is used on the consequent side to indicate at which membership function value the output should be evaluated since the type of logical connectives used in the consequent side is conjunctive. Here, the final membership function value is obtained by summing the two output membership functions on the consequent side. Finally, to come up with a defuzzified (crisp) value, either the Centroid Method, Maximum Membership Value Method, etc., could be used.

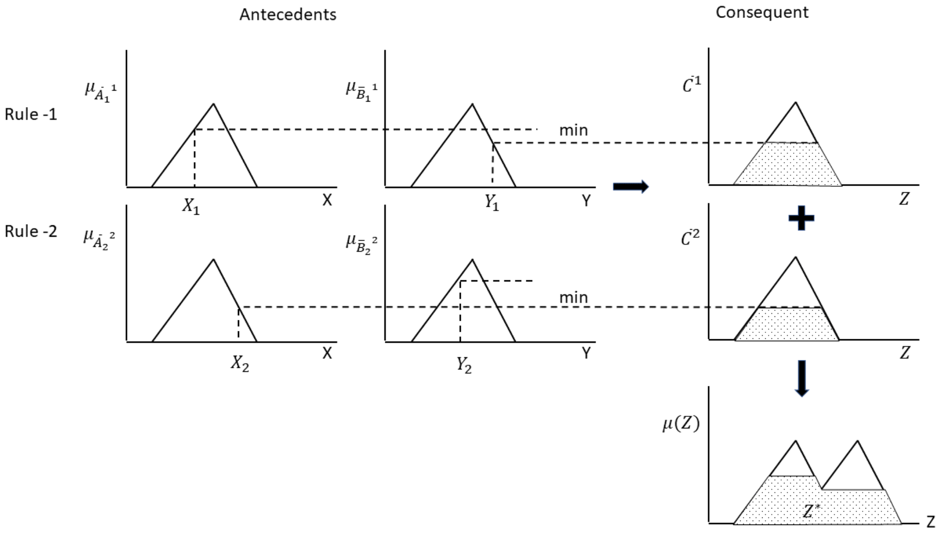


Figure 5.14: Mamdani Inference System for two inputs, one output, one membership function for each, and two rules.

Takagi-Sugeno Inference System (Type- III)

Here, unlike the previous system, on the consequent side, a linear function is used. This is better demonstrated in Figure:5.15 for the same problem specified in the type-I case. The crisp output in this case is a weighted sum of those two linear functions on the consequent side.

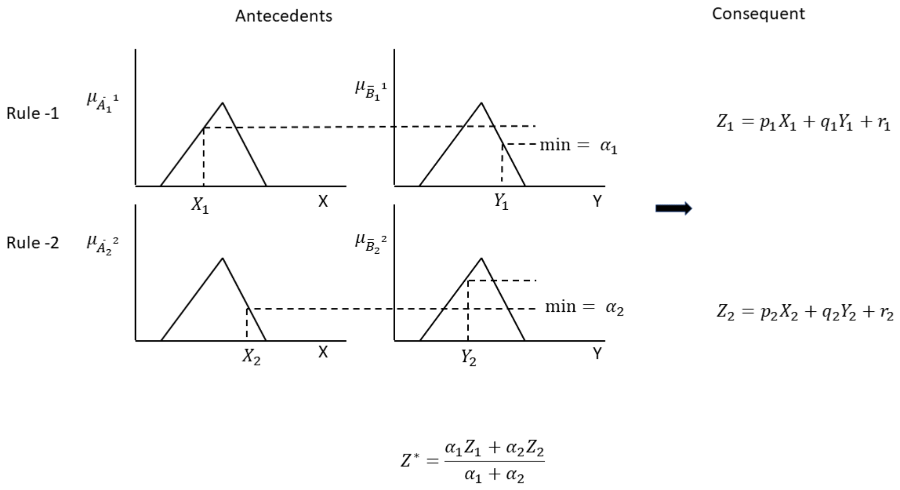


Figure 5.15: Takagi-Sugeno Inference System for two inputs, one output, one membership function for each, and two rules.

Tsykamoto Inference System (Type-I)

Similarly, in this system, the transformation takes place in the same manner as in the Type-III case, with the only exception of a monotonic function being used to define the membership function on the consequent side (Figure:5.16).

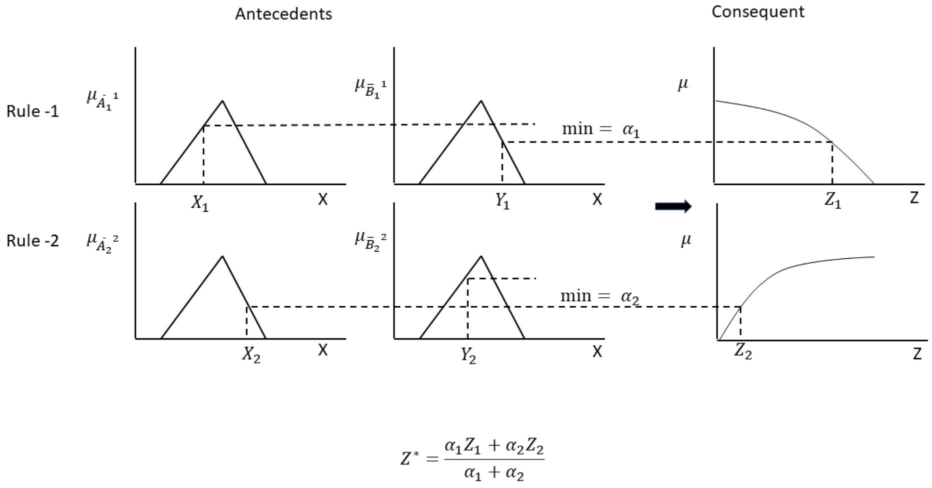


Figure 5.16: Tsykamoto Inference System for two inputs, one output, one membership function for each, and two rules.

ANFIS Steps (using Takagi-Sugeno System)

Unlike in ANN, in the ANFIS network, the number of layers is fixed at five, as shown in Figure:5.17. At each layer, various transformations on the input data are performed to generate the output.

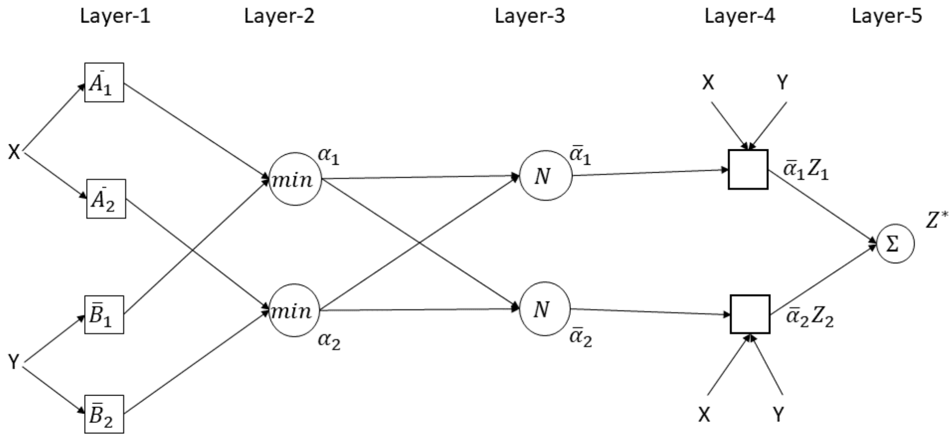


Figure 5.17: ANFIS layers

The operations performed at each layer is discussed as follows:

1. Layer-1

   This layer contains nodes with membership functions and is where the fuzzification occurs. The output from each node in this layer can be expressed as follows:

   $$O^1_{\bar{A}_i} = \mu_{\bar{A}_i}(X) \quad or \quad O^1_{\bar{B}_i} = \mu_{\bar{B}_i}(Y) \tag{5.37}$$

   Various types of membership functions can be used here, with the most commonly applied one being a bell-shaped function (Equation:5.38) that spans values from 0 to 1.

   $$\begin{aligned} \mu_{\bar{A}_i}(X) &= exp\Big\{-\big(\frac{X - c_i}{a_i}\big)^2\Big\} \\ \mu_{\bar{B}_i}(Y) &= exp\Big\{-\big(\frac{Y - c_i}{a_i}\big)^2\Big\} \end{aligned} \tag{5.38}$$

   Where: $\Big\{a_i, c_i\Big\}$ are antecedent side parameter sets.

2. Layer-2

   At this step, a minimization fuzzy operation is performed on the incoming signals from nodes in the first layer. Therefore, the procedure here can be described with the following equation:

   $$\alpha_i = min\Bigg[\mu_{\bar{A}_i}(X), \mu_{\bar{B}_i}(Y)\Bigg] \tag{5.39}$$

3. Layer-3

   This layer is a normalizing layer where weighting factor is calculated for each α_i as follows:

$$\bar{\alpha}_i = \frac{\alpha_i}{\alpha_1 + \alpha_2} \tag{5.40}$$

4. Layer-4

   The outputs of nodes in this layer are corresponding node function multiplied by the weighing factor obtained from layer-3:

$$O_i^4 = \bar{\alpha}_i Z_i \tag{5.41}$$

   Where: $Z_i = p_i X + q_i Y + r_i$, and p_i, q_i, r_i are consequent parameter sets.

5. Layer-5

   This is the final layer in the network where the weighted outputs from every node of layer-4 are summed up together to give a final defuzzified output as shown below:

$$O_1^5 = \sum_i \bar{\alpha}_i Z_i \tag{5.42}$$

   Where: O_1^5 is the final output at layer-5 and 1 indicates the first output.

ANFIS Training (Learning)

The same as the ANN network discussed above, the ANFIS network is also trained with a training data set using a gradient-based learning procedure. Since every node's output depends on the input signal from the previous output node and its parameters, this relationship could be described as:

$$O_i^k = O_i^k(O_i^{k-1}, . \quad . \quad . \quad , O_{(\#k-1)}^{k-1}, a, c, . \quad . \quad . \quad) \tag{5.43}$$

Where; O_i^k is the output from k^{th} layer and i^{th} node, and $\#k$ is the number of nodes in previous layer. In addition, a and c are the node function parameters.

Let's assume that the training data consists of P number of inputs, the error at layer L for the $p^{th}(1 \leq p \leq P)$ input and output could be written as:

$$E_p = \sum_{m=1}^{\#(L)} \left(T_{m,p} - O_{m,p}^L\right)^2 \tag{5.44}$$

Where: $T_{m,p}$ and $O_{m,p}^L$ are Target and outputs for the p^{th} training input and m^{th} component of the output vector. Here,$\#L$ stands for the number of layers.

Therefore, if α is a parameter for given ANFIS network, error at a certain node is given by:

$$\frac{\partial E_p}{\partial \alpha} = \sum_{O^* \epsilon S} \frac{\partial E_p}{\partial O^*} \frac{\partial O^*}{\partial \alpha} \tag{5.45}$$

Where: S is a set of output nodes in the network whose outputs depends on α.

The overall error for whole training set (P) is therefore will be given as:

$$\frac{\partial E}{\partial \alpha} = \sum_{p=1}^{P} \frac{\partial E_p}{\partial \alpha} \tag{5.46}$$

Update of the past value of the parameter will be calculated as follows:

$$SS_\alpha = -\kappa \frac{\partial E}{\partial \alpha} \tag{5.47}$$

Where: SS_α is the step size and κ is the learning rate.

However, since the gradient method is slow and can be trapped in local minima, Jang (1993) [135] suggested a Hybrid learning rule. This approach suggests combining the gradient descent method with least square estimation (LSE) to speed up and increase the efficiency of parameter searching. Here, the learning process starts by running the network until layer-4 with initial parameters and training data in the forward pass. At layer five, the LSE is applied to fit parameter sets of the consequent side. Then, in the backward pass, gradient descent would be used to optimize the antecedent parameter sets. The advantage of the Hybrid learning rule is that other optimization techniques could be put into use in the antecedent part, such as conjugate-gradient descent, non-linear optimization, etc.

5.4 Results

In this section, the methodologies discussed above have been applied to fit different types of models to monthly flow data obtained from three major rivers over UASB. The primary objective of this section is to evaluate the potential of two time series modeling techniques in terms of their ability to simulate the observed flow data. The techniques applied are categorized into regime switching and data-driven techniques.

5.4.1 Regime Switching Models

In this section, the potential of Threshold Autoregressive (Threshold Auto-Regressive (TAR)) models, which apply a threshold method, will be evaluated. Two major types of TAR models, called SETAR and Logistic Smooth Threshold Auto-Regressive (LSTAR), have been fitted to the monthly flow data. The modeling procedure started by first transforming the monthly time series into a normal distribution with the help of the Box-Cox technique and then removing the seasonal effect by fitting a Fourier Harmonic regression. These steps have been shown graphically for the Hombole station in Figure:5.18, and the fitted parameters of the two-period harmonic regression equation for all three gauging stations are given in Table:5.1.

Table 5.1: Parameters of Harmonic Regression Equation (with period=2)

Station	Constant	Cos(1)	Sin(1)	Cos(2)	Sin(2)
Hombole	1.764	-0.682	-0.340	0.070	0.240
Melka	1.428	-1.162	-0.626	0.090	0.350
Akaki	0.511	-0.682	-0.271	0.091	0.295

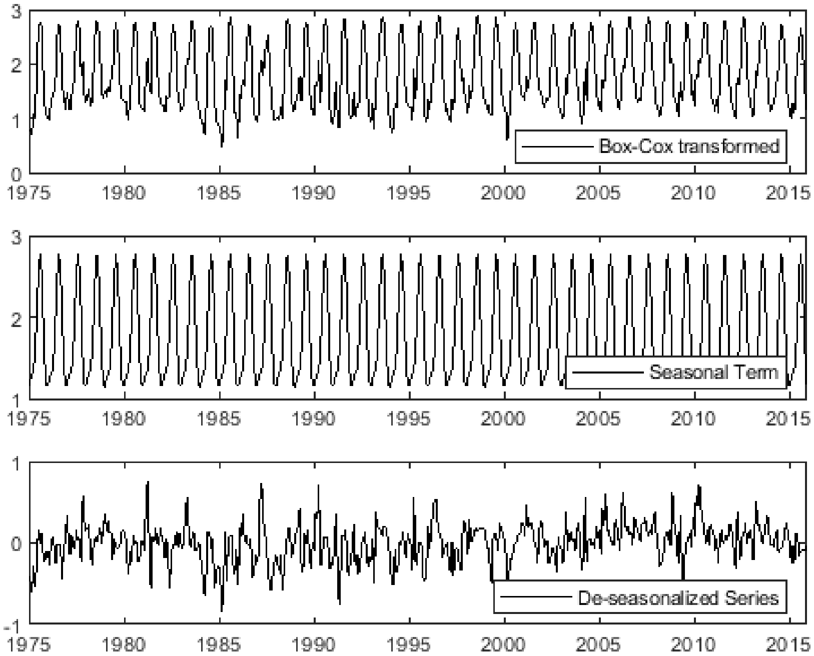


Figure 5.18: De-seasonlizing of monthly series of Hombole station.

The actual models were fitted to the residuals, which are obtained by removing the seasonal part from the transformed series. After model fitting, its performance was measured by reversing the whole procedures used to obtain the residual series, i.e., adding the seasonal term and back-transforming. In all cases, modeling was performed after dividing the total data length into a training and test set. Since the total data length at each gauging station is different, the proportion of data used for model fitting and testing varies.

Procedures followed for Fitting a TAR Model;

Two types of TAR models have been fitted to the time series: SETAR and LSTAR. In both cases, the general modeling procedure followed is:

1. The model order (p) was first identified with the help of PACF plots.
2. Delay term (d) is specified with the help of threshold non-linearity test [128].
3. The grid search algorithm under the tsDyn package ([129]) helps to identify threshold value (Thr) with low pooled AIC.
4. Model parameters of the AR terms in each regime where again estimated with the help of tsDyn package.
5. Model adequacy is proven with help of pacf plots and Ljung-Box test.
6. At end, forecasts where done and compared with original series to measure model performance.

Hombole Station

The monthly flow data at this station was identified to have an order (p) of 7 with the help of pacf plots, as shown in Figure:5.19. After running a threshold test for different combinations of (p, d), a delay value of d=2 was selected for the threshold model.

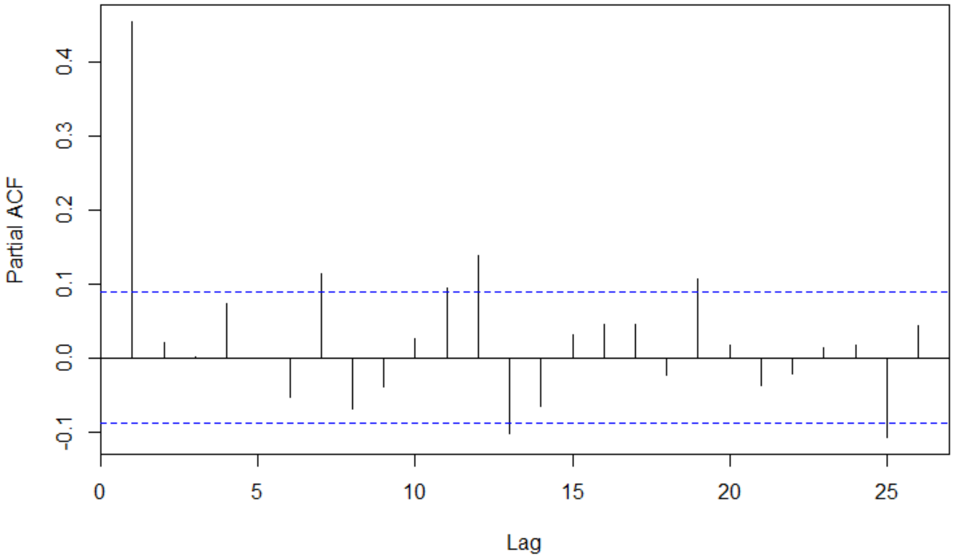


Figure 5.19: PACF of Monthly Box-Cox transformed and deseasonalized Hombole flow series to identify order(p).

Next, a grid search algorithm was employed to determine which combinations of (p, d) would yield low pooled AIC values, utilizing the tsDyn package (Figure:5.20). The grid search identified a model with (d, mL, mH) = (1, 7, 4) (Table:5.2) that provided the lowest pooled AIC value for various combinations of delay and order among the values defined previously. Here, mL and mH represent the orders of the two regimes (Low and High) in the model.

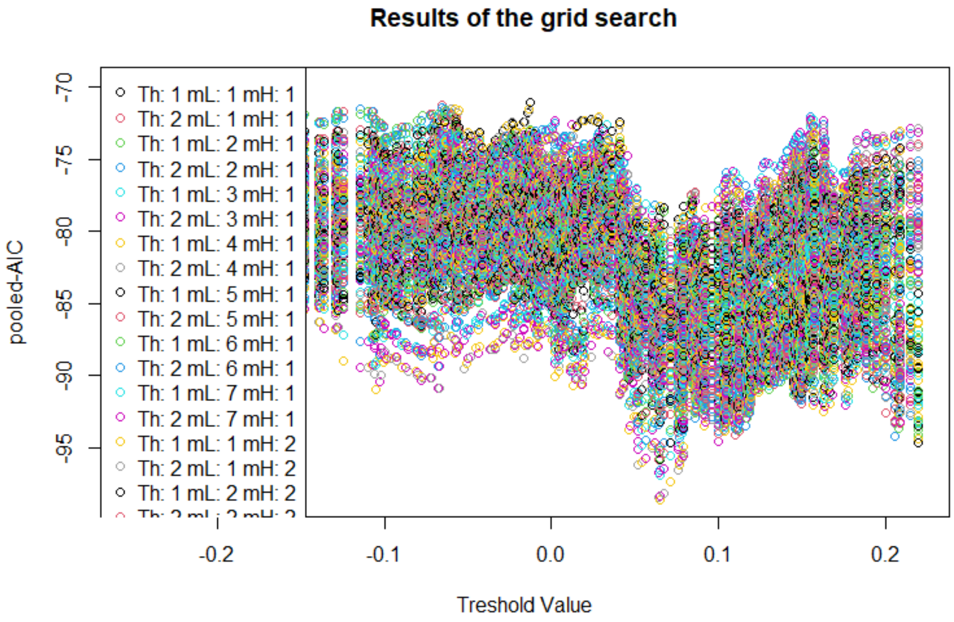


Figure 5.20: Grid search for Hombole flow series to identify which combination of (mL,mH,d) will yield low pooled AIC for SETAR model Fitting.

Table 5.2: Top ten best outcomes with low AIC values for the Grid search on Hombole Station)

Trials	thDelay	mL	mH	th	pooled-AIC
1	1	7	4	0.06477938	-98.60532
2	1	7	4	0.0639956	-98.40885
3	1	7	4	0.06500946	-98.31959
4	1	7	4	0.06710211	-98.17887
5	1	7	4	0.07111458	-97.32338
6	1	7	5	0.06477938	-96.88736
7	1	7	5	0.0639956	-96.72643
8	1	7	5	0.06500946	-96.57525
9	1	7	4	0.07605536	-96.52041
10	1	7	5	0.06710211	-96.43385

However, when the model was later tested for adequacy using pacf and the Ljung-Box test, it was found to be a poor fit. Therefore, the next option, (1, 7, 5) as shown in Table:5.2, was considered, and this model proved to be appropriate. The pacf plots demonstrated that almost all partial autocorrelation values fell within the 95% confidence band (Figure:5.21).

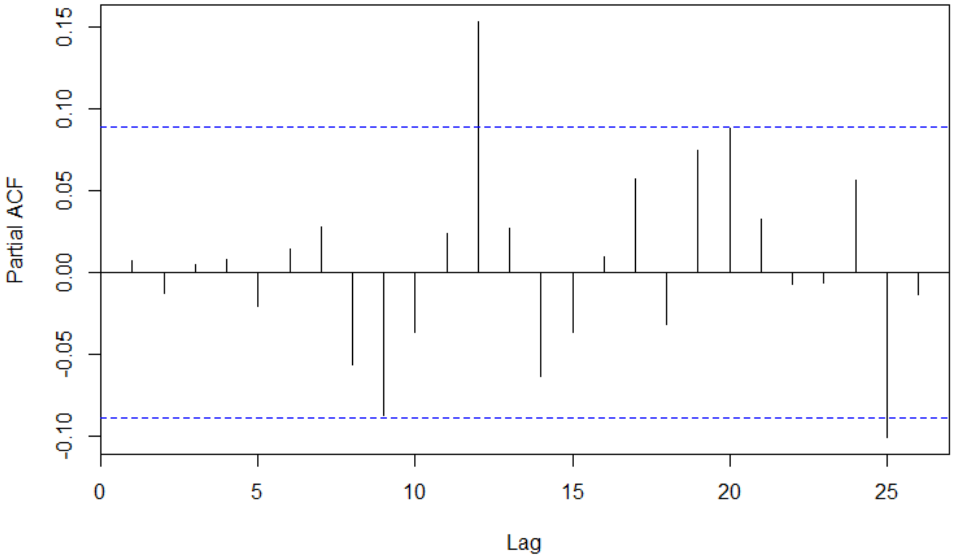


Figure 5.21: PACF plot with confidence band (95%) for the residual of SETAR (1,7,5) model fitted to deseasonalized series of monthly Hombole station.

From Figure:5.22, it can be observed that the threshold values suggested by the grid search algorithm show a clear shift in the mean of the flow series starting from April 2003, where most of the data points are lying above the threshold line. The parameters of the fitted two-regime SETAR model are shown in Table:5.3.

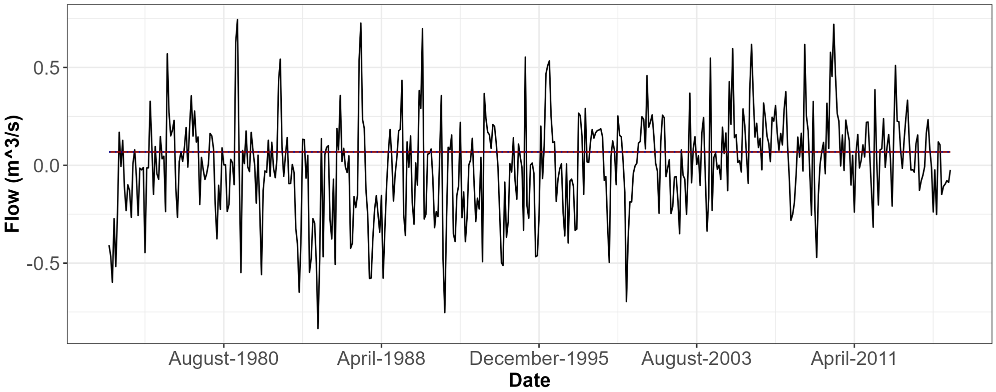


Figure 5.22: Transformed and deseasonalized monthly Hombole station flow series with threshold lines (SETAR threshold (Red line) = $0.065m^3/s$ and LSTAR threshold (dotted blue line)= $0.068m^3/s$) overlaid.

Table 5.3: SETAR(1,7,5) model parameters and fit statistics for monthly flow series of Hombole station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.009	0.017	0.522	0.602
phiL.1	0.412	0.054	7.605	0.000
phiL.2	0.011	0.082	0.129	0.897
phiL.3	0.065	0.064	1.029	0.304
phiL.4	0.059	0.065	0.908	0.364
phiL.5	0.013	0.064	0.204	0.838
phiL.6	-0.207	0.062	-3.315	0.001
phiL.7	0.204	0.058	3.505	0.000
const.H	0.004	0.028	0.137	0.891
phiH.1	0.541	0.08	6.772	0.000
phiH.2	0.009	0.108	0.086	0.931
phiH.3	-0.2	0.076	-2.65	0.008
phiH.4	0.118	0.073	1.618	0.106
phiH.5	0.032	0.067	0.478	0.633

Threshold = 0.0648
Residual variance= 0.044
AIC = -1500
MAPE = 265.3%

Similar procedures were followed to identify the best LSTAR model, and a two-regime model with a threshold value of 0.068 m^3/s and smoothing parameter $(\gamma) = 2000$ was found to have the lowest AIC. Figure:5.23 shows the plot of the transition function with the threshold variable (S_t), and the red dotted line indicates the threshold value = 0.068 m^3/s.

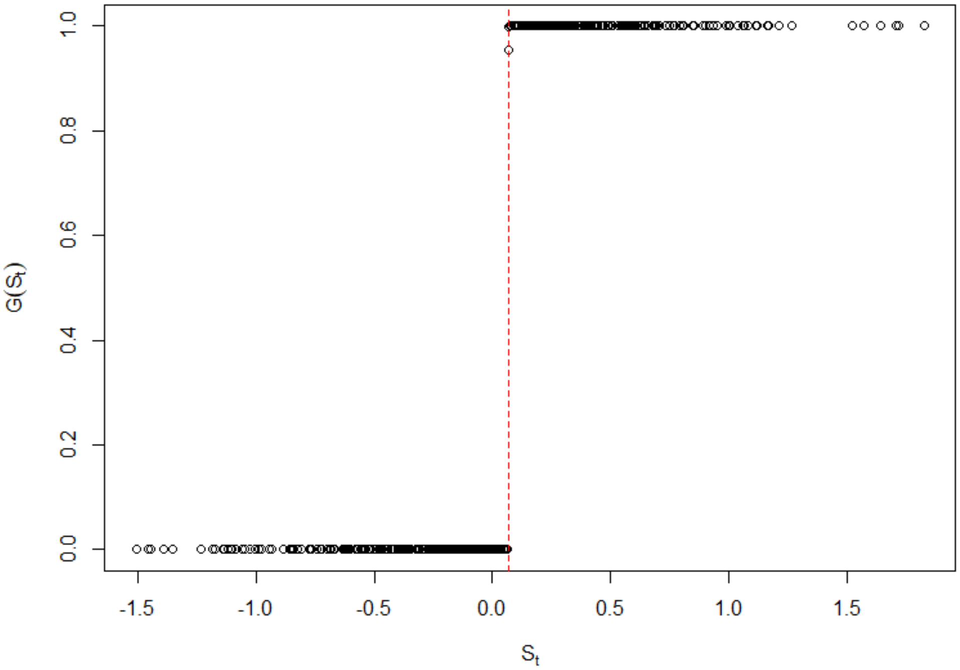


Figure 5.23: Transition function $(G(S_t))$ of transformed and deseasonalized monthly Hombole flow series (S_t) with the red dotted line indicating the threshold value = 0.068 m^3/s used in the fitted model.

Again, LSTAR model parameters were identified using the tsDyn package in R and are shown in Table:5.4. The model adequacy was evaluated using the pacf plot of model residuals shown in Figure:5.24, and it can be concluded that the selected model is appropriate.

Table 5.4: LSTAR(1,7,7) model parameters and fit statistics for monthly flow series of Hombole station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.01	0.016	0.582	0.561
phiL.1	0.412	0.053	7.729	0.000
phiL.2	0.013	0.08	0.166	0.868
phiL.3	0.066	0.063	1.052	0.293
phiL.4	0.057	0.064	0.897	0.37
phiL.5	0.011	0.063	0.18	0.857
phiL.6	-0.205	0.061	-3.341	0.001
phiL.7	0.204	0.057	3.564	0.000
const.H	-0.006	0.033	-0.176	0.86
phiH.1	0.13	0.095	1.368	0.171
phiH.2	-0.007	0.135	-0.053	0.957
phiH.3	-0.264	0.098	-2.7	0.007
phiH.4	0.059	0.098	0.6	0.549
phiH.5	0.004	0.098	0.045	0.964
phiH.6	0.239	0.099	2.4	0.016
phiH.7	-0.203	0.089	-2.28	0.023
gamma	2000	10504	0.19	0.849
th	0.068	0.008	8.522	$< 2.2e-16$

[2] Significance level(α) is 5%.
residual variance = 0.045
AIC = -1493, MAPE = 264.2%

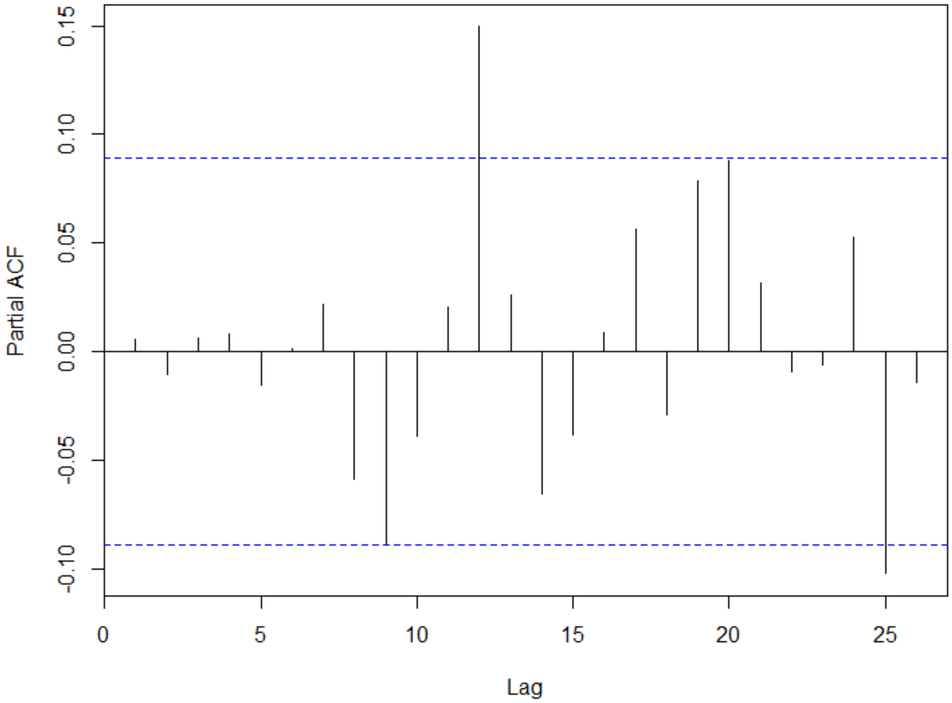


Figure 5.24: PACF plot with confidence band (95%) for the residual of LSTAR (1,7,7) model fitted to deseasonalized series of monthly Hombole station.

After fitting model parameters, one-step ahead forecasts were performed using both SETAR and LSTAR models with a few years of data used to initialize the models. Both models seem to have reasonable performance when compared against the original flow series, as shown in Figures:5.25 and 5.26. However, it can be observed that both models face severe difficulty in estimating the peak values.

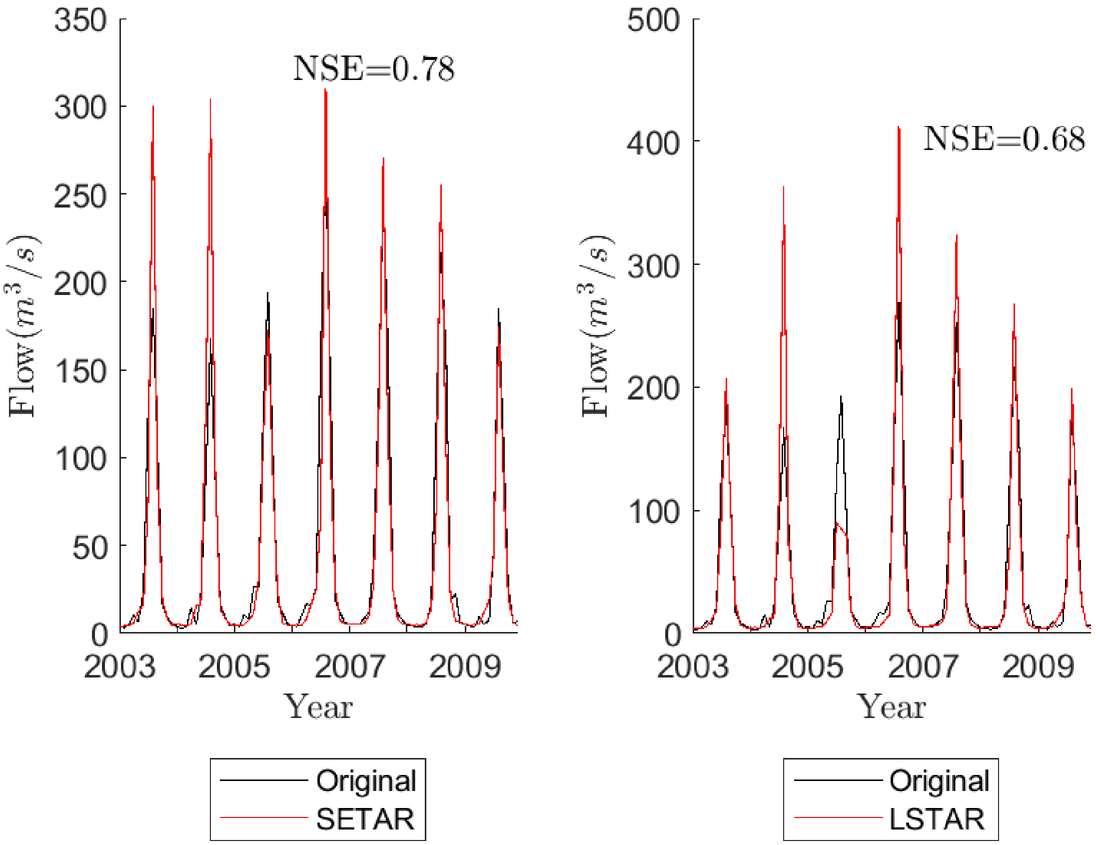


Figure 5.25: Monthly plot of original and forecasted SETAR (Left) and LSTAR (right) model outputs for the period (01/01/2003-12/01/2009) in Hombole station.

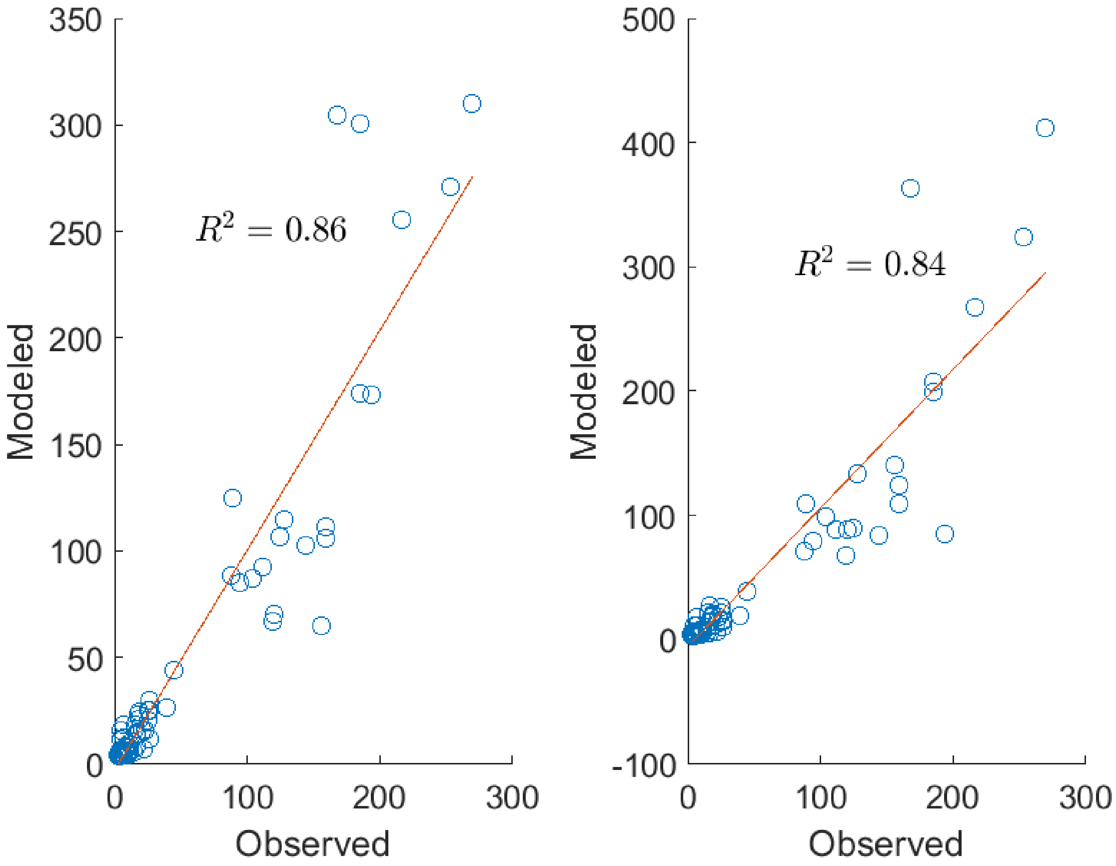


Figure 5.26: Scatter plot of original and forecasted SETAR (Left) and LSTAR (right) model outputs fo the test period for the monthly flow series in Hombole station.

Melka Kuntere Station

In the Melka Kuntere station, an order (p) of 9 from the PACF plot and a delay of 9 from the threshold test were obtained and later used to identify appropriate SETAR and LSTAR threshold values using the same grid search algorithm. Therefore, the threshold values for SETAR and LSTAR were obtained as $-0.288 m^3/s$ and $-0.356 m^3/s$, respectively. These threshold values were later overlaid over the residual series, as shown in Figure:5.27. It can be observed from this plot that, unlike the Hombole station, here there is no clear observable shift in mean, with the exception of a few months at the start of the series. The plot shows that the threshold values separate the entire data series into two, with data points having higher variance located above the threshold line and lower variance below the line.

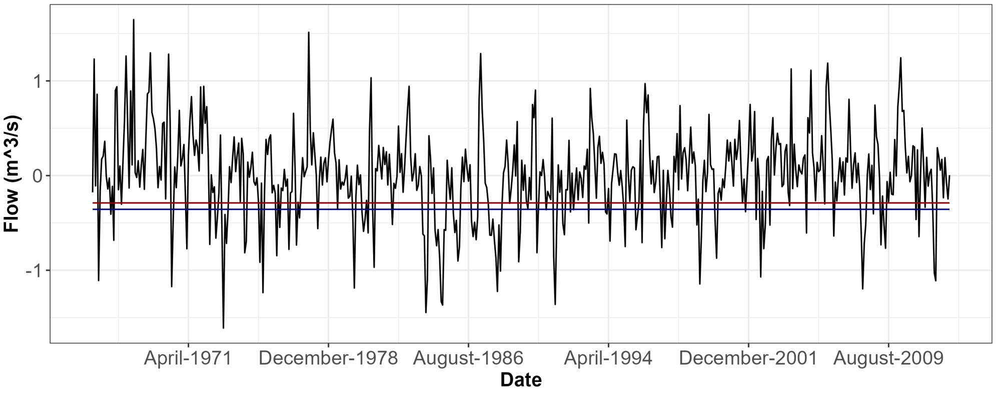


Figure 5.27: Transformed and deseasonalized monthly MelkaKuntere station flow series with threshold lines (SETAR threshold (Red line) = $-0.288m^3/s$ and LSTAR threshold (dotted blue line)= $-0.356m^3/s$) overlaid.

In the Melka Kuntere station, the fitted model parameters for each model are shown in Tables:5.5 and 5.6. These model parameters were later used to generate a forecast in the same manner as for the Hombole station as shown in Figures: 5.28 and 5.29 . The resulting forecasts are not very promising and have performed poorly compared to the Hombole station. The models completely miss the peaks, with only seasonal patterns being captured.

Table 5.5: SETAR(6,6,1) model parameters and fit statistics for monthly flow series of Melkakuntere station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.005	0.044	0.11	0.913
phiL.1	0.347	0.089	3.876	0.000
phiL.2	0.163	0.103	1.591	0.112
phiL.3	0.067	0.105	0.638	0.524
phiL.4	-0.238	0.099	-2.42	0.016
phiL.5	0.082	0.094	0.872	0.384
phiL.6	0.099	0.088	1.123	0.262
const.H	0.002	0.02	0.101	0.919
phiH.1	0.431	0.043	10.005	< 0.000

[2] Significance level(α) is 5%.
Threshold =-0.2885
residual Variance= 0.178
AIC = -956
MAPE = 213.7%

Table 5.6: LSTAR(1,9,1) model parameters and fit statistics for monthly flow series of Melka-kuntere station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.019	0.06	0.309	0.757
phiL.1	0.578	0.079	7.352	0.000
phiL.2	0.021	0.059	0.362	0.718
phiL.3	0.059	0.045	1.307	0.191
phiL.4	-0.086	0.045	-1.899	0.058
phiL.5	0.031	0.045	0.7	0.484
phiL.6	0.038	0.044	0.857	0.392
phiL.7	-0.006	0.044	-0.139	0.890
phiL.8	-0.07	0.044	-1.599	0.110
phiL.9	0.093	0.041	2.273	0.023
const.H	-0.009	0.069	-0.129	0.898
phiH.1	-0.235	0.091	-2.571	0.010
gamma	2000	9934.3	0.201	0.840
th	-0.356	0.059	-6.069	0.000

[2] Significance level(α) is 5%.
residual variance = 0.175
AIC = -954, MAPE = 298.6%

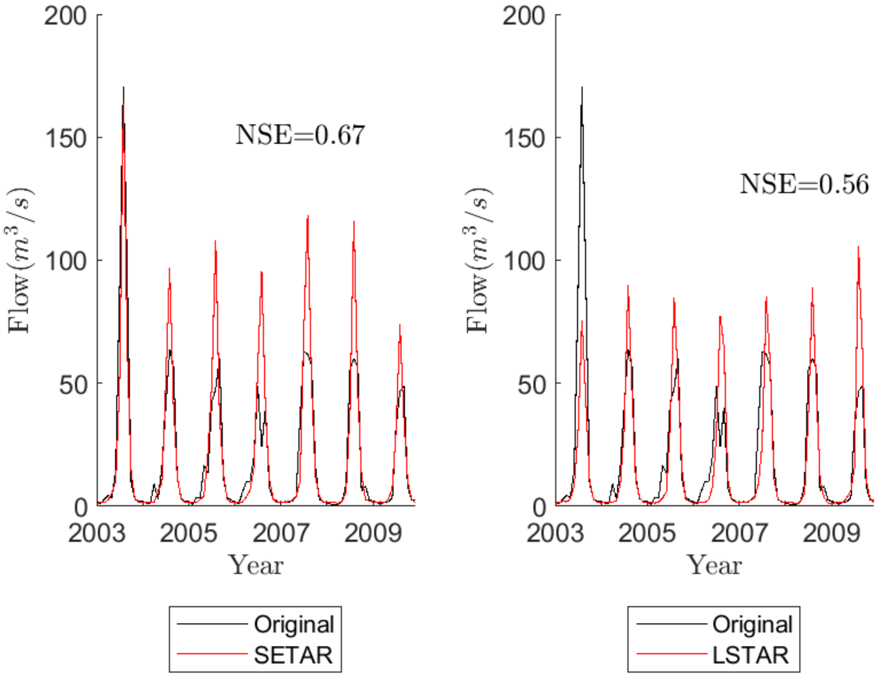


Figure 5.28: Original and Forecasted SETAR and LSTAR model output for Melka station.

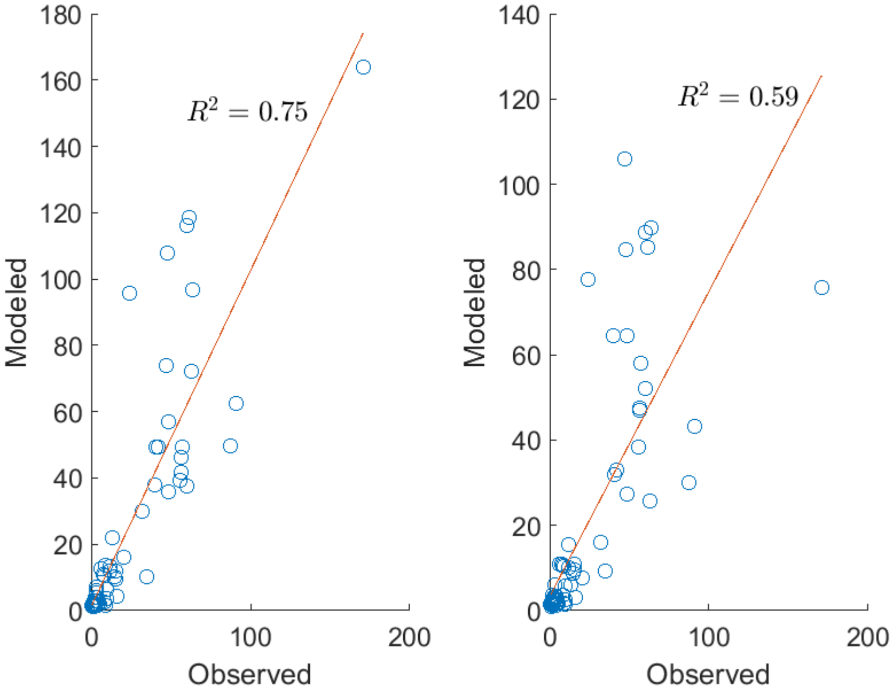


Figure 5.29: Original and Forecasted SETAR (left) and LSTAR (right) model trend fit line for Melka station.

Akaki Station

In this station, the monthly flow data was identified to have p=7 and d=3. These were later used to identify possible threshold values for both SETAR and LSTAR models using the grid search algorithm. The identified threshold values with low pooled AIC values are shown in Figure:5.30. From the plot, it can be seen that the threshold value obtained from the SETAR model shows the separation of data much better than the LSTAR model.

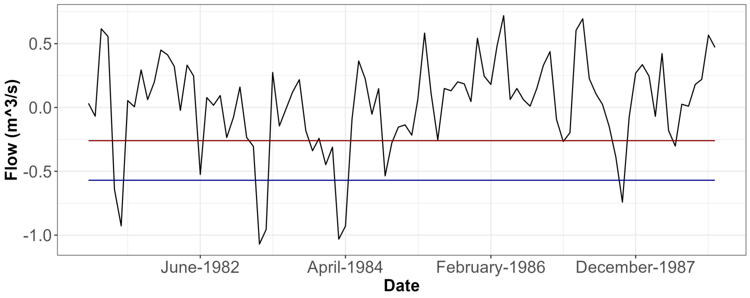


Figure 5.30: Transformed and deseasonalized monthly Akaki station flow series with threshold lines (SETAR threshold (Red line) = $-0.26 m^3/s$ and LSTAR threshold (dotted blue line)= $-0.57 m^3/s$) overlaid.

Later, both models were fitted using model parameters obtained with the help of the grid search algorithm. The corresponding parameters used in forecasting are shown in Tables:5.7 and 5.8.

Table 5.7: SETAR(1,1,7) model parameters and fit statistics for monthly flow series of Akaki station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.003	0.098	0.032	0.974
phiL.1	0.439	0.177	2.482	0.015
const.H	0.01	0.042	0.236	0.814
phiH.1	0.579	0.134	4.322	0.000
phiH.2	-0.052	0.169	-0.306	0.761
phiH.3	0.012	0.137	0.09	0.928
phiH.4	-0.04	0.131	-0.306	0.760
phiH.5	0.067	0.124	0.544	0.588
phiH.6	-0.224	0.117	-1.913	0.059
phiH.7	0.266	0.102	2.606	0.011

[1] critical values for rejection of null hypothesis is when $|Z - value| >$ 1.96 and P- value is $\leq$ 0.05.

[2] Significance level(α) is 5%.
Threshold =-0.26
residual Variance= 0.082
AIC = -220, MAPE = 132.5%

Table 5.8: LSTAR(1,1,7) model parameters and fit statistics for monthly flow series of Akaki station)

Sub-models	Estimate	Standard Error	TStatistic	P-Value
const.L	0.095	0.128	0.741	0.458
phiL.1	-0.057	0.229	-0.249	0.803
const.H	-0.11	0.132	-0.837	0.403
phiH.1	0.724	0.249	2.906	0.004
phiH.2	0.003	0.133	0.019	0.984
phiH.3	0.04	0.112	0.353	0.724
phiH.4	-0.026	0.108	-0.241	0.809
phiH.5	0.113	0.105	1.07	0.285
phiH.6	-0.237	0.101	-2.349	0.019
phiH.7	0.273	0.088	3.1	0.002
gamma	36.539	3238.5	0.011	0.991
th	-0.572	0.029	-19.447	$<$ 0.000

[1] critical values for rejection of null hypothesis is when $|Z - value| >$ 1.96 and P- value is $\leq$ 0.05.

[2] Significance level(α) is 5%.
residual variance = -0.572
AIC = -229, MAPE = 155.5%

Even though the data series at this station is very short, both models have satisfactorily simulated the original time series as shown in Figures:5.31 and 5.32. As it can be seen, the LSTAR model has a lower performance than its SETAR counterpart. This is due to the threshold value estimated for the LSTAR model poorly partitioning the data series into two regimes as compared to its SETAR counterpart.

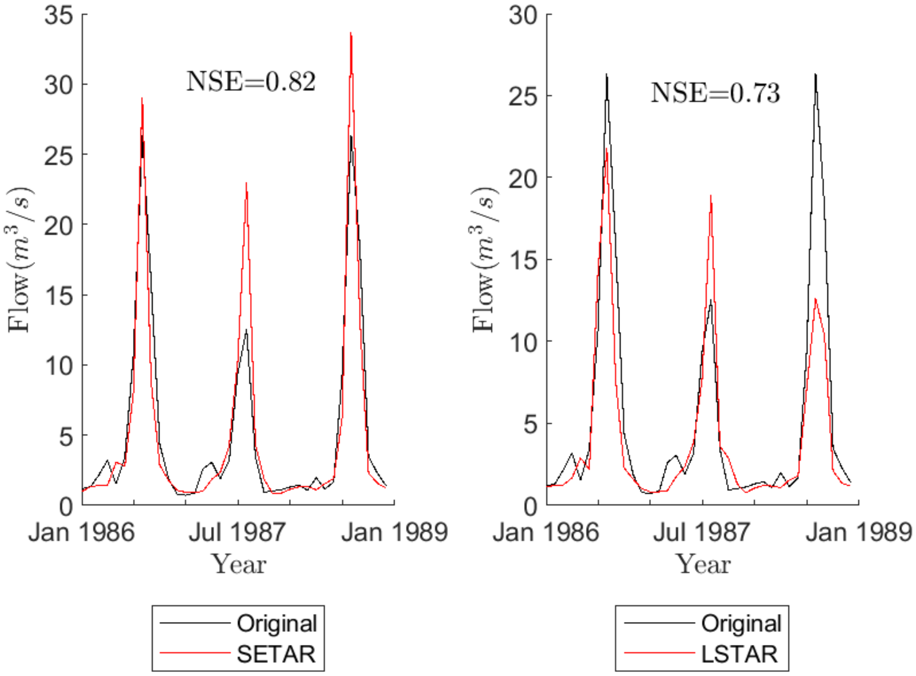


Figure 5.31: Original and Forecasted SETAR and LSTAR model output for Akaki station.

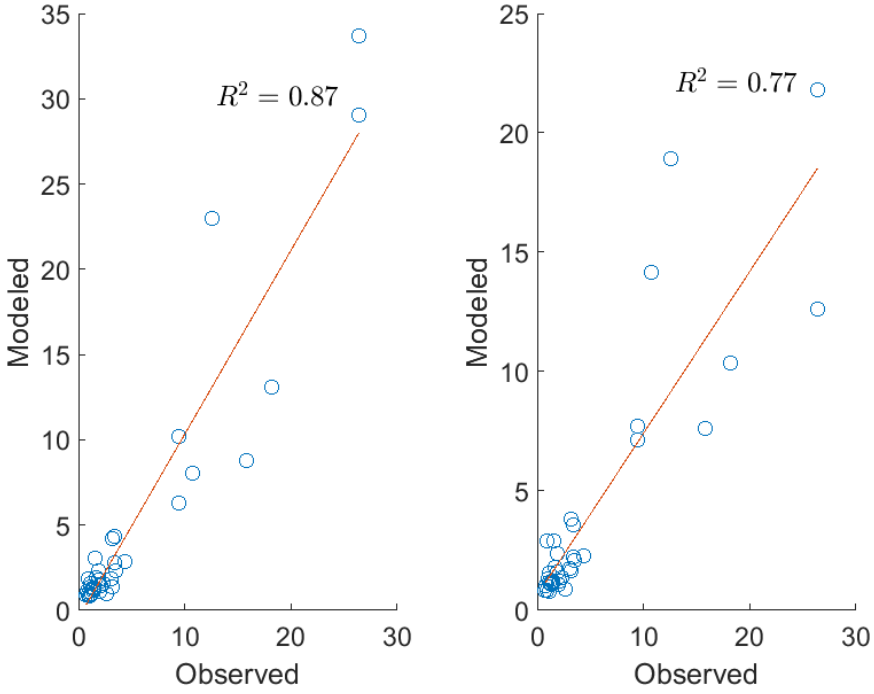


Figure 5.32: Original and Forecasted SETAR (left) and LSTAR (right) model trend fit line for Akaki station.

5.4.2 Data Driven Approaches

NARX Model Training, Validation, and Uncertainty

The model building started with the identification of model inputs at each gauging station, including monthly rainfall, mean monthly temperature, monthly potential evapotranspiration (Eto), mean monthly 30-day Antecedent Precipitation Index (API), and monthly flow data. All climate inputs are areal weighted values obtained through the use of a Thiesson polygon developed for a catchment corresponding to each gauging station. After identifying possible model inputs, the next task was to analyze which input combination gives the best-performing model. In this study, nine possible input combinations were identified based mainly on intuition and a trial-and-error approach, as shown in Table:5.9. Each of these nine input combinations corresponds to a single model.

Table 5.9: NARX input combinations

Model	Rf	mTemp	Eto	API	Flow
1	x	x	x	x	x
2	x	x	x		x
3	x	x			x
4	x		x		x
5	x				x
6		x	x	x	x
7		x		x	x
8			x	x	x
9				x	x

Both model inputs and outputs are normalized (standardized) before being given as input to the NARX model in the training and testing phases. Additionally, Table:5.10 shows the general NARX model settings used while training and testing the model at all three gauges.

Table 5.10: General NARX settings used in model generation

Input transfer function:	Sigmoid
Output transfer function:	ReLu
Training Algorithm:	Levenberg Marquardt with Bayesian regularization
Delay:	1
Number of Hidden layers:	1
Number of Hidden nodes:	Varying
epoch	1000

Hombole Station

The primary approach employed for selecting an appropriate model began with defining the network architecture. In this context, a two-layer NARX model comprising a hidden layer and an output layer was chosen. This decision stemmed from the observation that increasing the number of layers did not enhance model accuracy; instead, it heightened computational time. Consequently, each stage of model development concentrated on determining an optimal number of nodes in the hidden layer. The input data underwent partitioning into training and validation sets, with 75% of the data series allocated for model training and the remaining 25% for testing.

The initial step in model construction involved identifying the best architecture for each of the nine models. This was achieved iteratively by varying the number of hidden nodes and subsequently assessing performance. The chosen performance metric at this stage was RMSE.

For each model, a specific number of nodes was selected, yielding a relatively lower RMSE value in both the training and testing phases. Ultimately, the model with the lowest RMSE value among the nine was considered the most suitable for forecasting the monthly flow series at this station. The results of this analysis for each of the nine models, corresponding to different input combinations, are depicted in Figure:5.33. Notably, Model 5 emerges as the preferred alternative, demonstrating a lower RMSE value in the testing phase. Subsequent input combinations displayed signs of overfitting. Further analysis revealed that the optimal number of nodes for the hidden layer of Model 5 was nine.

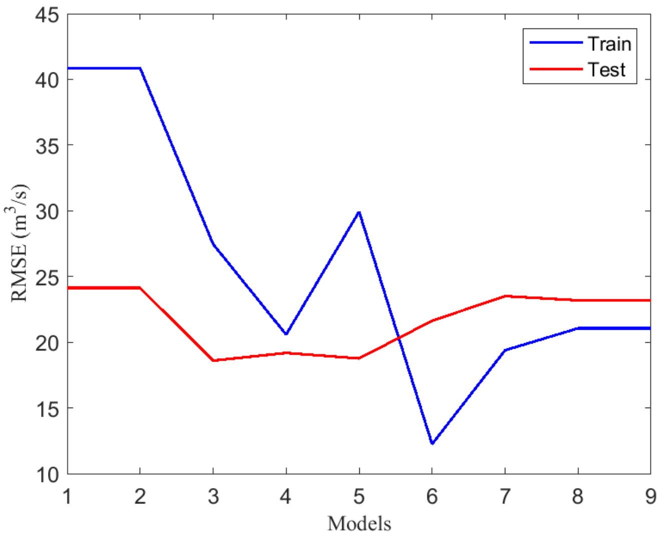


Figure 5.33: Hombole station NARX model performance using different input combinations for Train and Test period

The results of model training and validation (testing) are visually presented in Figures:5.34 and 5.35. Based on the Nash Sutcliffe Efficiency (NSE) values, it can be inferred that the model excelled in capturing seasonal variations and low flows (base flow). However, conversely, the model did not exhibit a comparable level of performance in predicting peak flows. This discrepancy is further evident in the uncertainty plot depicted in Figure:5.36.

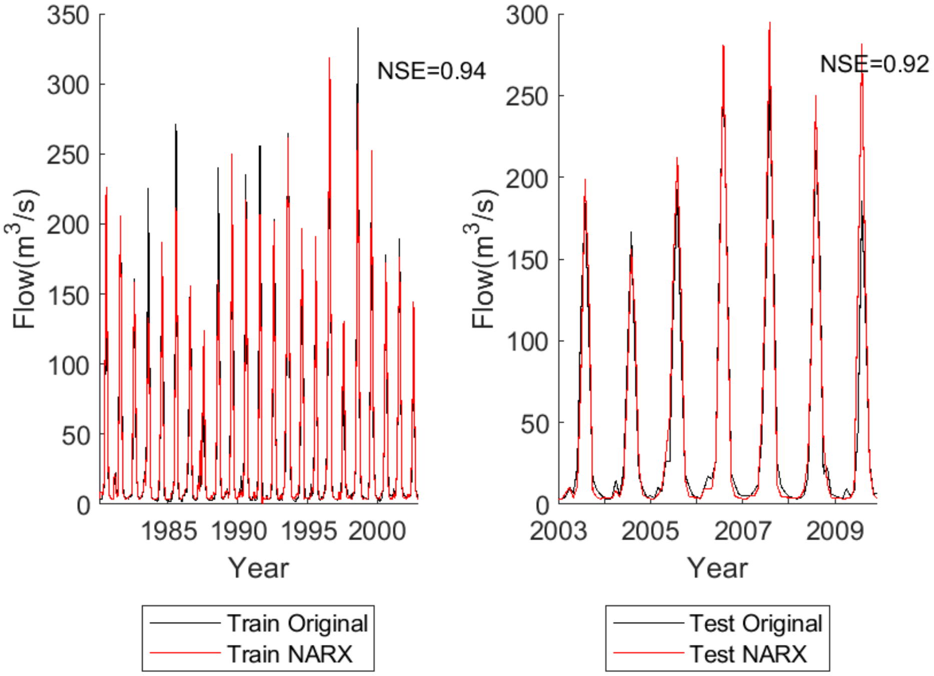


Figure 5.34: Hombole station NARX simulation for Train and Test period and its performance.

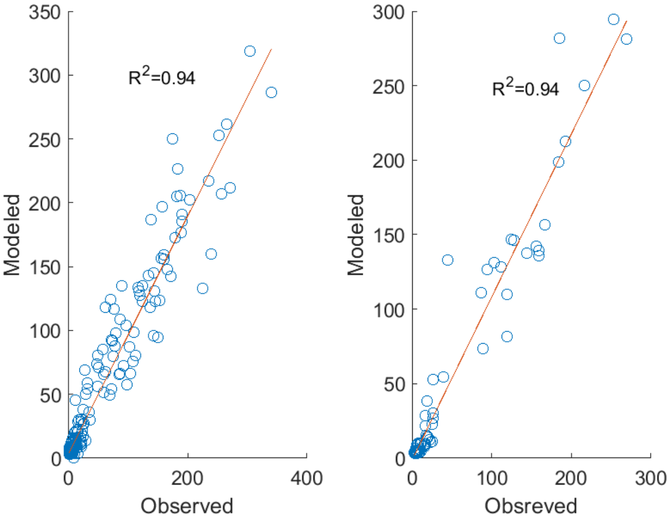


Figure 5.35: Hombole station Observed versus NARX output scatter plot for train (left) and test (right) period.

The analysis of uncertainty in the NARX model involves generating 300 sample datasets with the same time span as the model's input variables. These samples are created using a bootstrapping technique implemented with the datasample() function in MATLAB [97]. The bootstrapping method involves creating a resampled dataset from the original dataset with replacement.

The results of the uncertainty analysis are illustrated in Figure 5.36. It is noticeable that the model predictions closely align with the bootstrapped mean and fall within the 95% confidence interval. Notably, as the prediction horizon increases, predictability diminishes, particularly evident from August 2006 onward. Additionally, the uncertainty band widens as the prediction horizon extends.

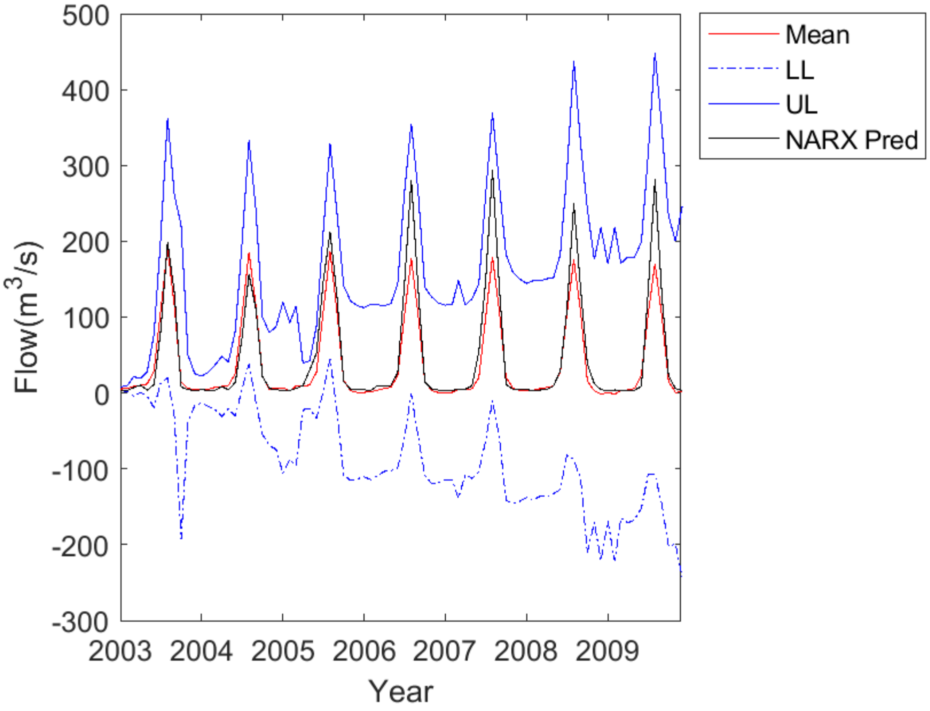


Figure 5.36: Hombole station NARX bootstrapped (Mean and 95% Lower Limit(LL) - Upper Limit (UL) versus Prediction for test period.

MelkaKuntere Station

At this station, the same set of input combinations, detailed in Table 5.9, was employed using the methodology previously described to evaluate candidate models across different numbers of hidden nodes. The analysis of model performance reveals that Model No. 5 demonstrates superior performance, boasting the best results with a set of 7 hidden nodes, as illustrated in Figure:5.37.

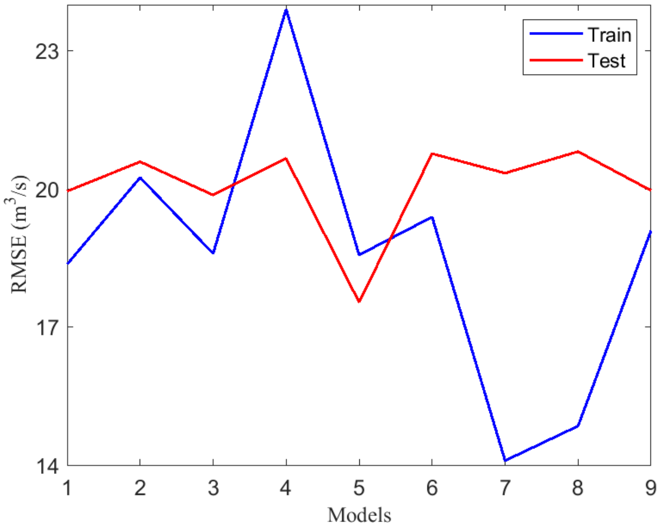


Figure 5.37: Melka Kuntere station NARX model performance using different input combinations for Train and Test period

Following the selection process, the model was executed with the fitted parameters derived from the training phase, and the results are illustrated in Figure:5.38 and 5.39. This depiction reveals that, similarly, the model performed well in capturing low flows and seasonal characteristics overall. However, there is a noticeable pattern of missing some peak values, as highlighted in the figures.

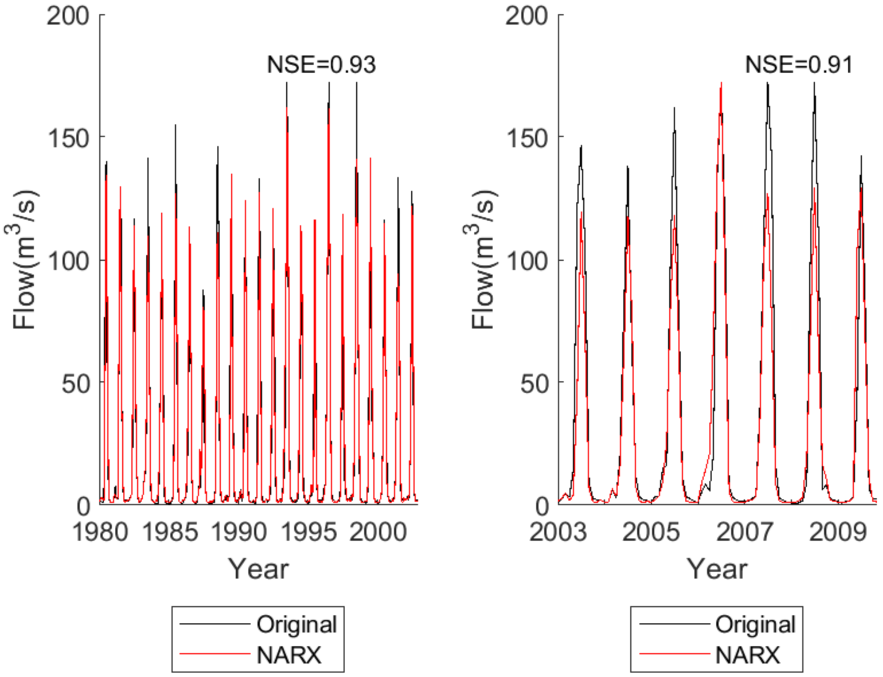


Figure 5.38: Melka station NARX simulation for Train and Test period and its performance.

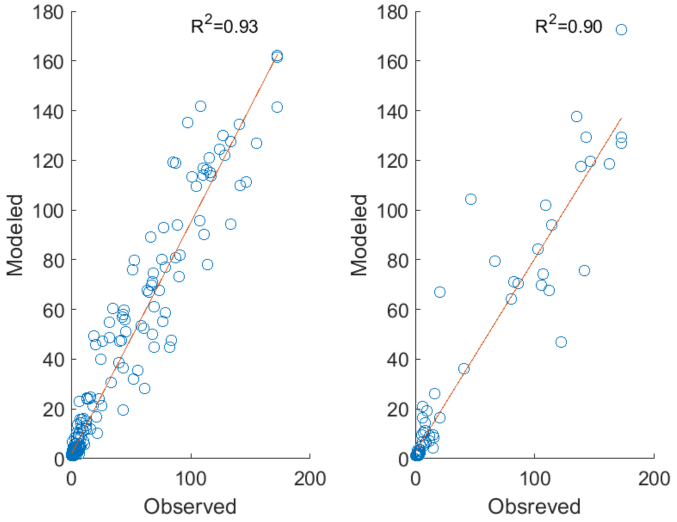


Figure 5.39: Melka station Observed versus NARX output scatter plot for train (left) and test (right) period.

The result of the uncertainty analysis is presented in the plot shown in Figure:5.40. This plot indicates that the trained model consistently produces results when exposed to different datasets, as evidenced by the minimal difference observed between the current and mean of the bootstrapping simulation.

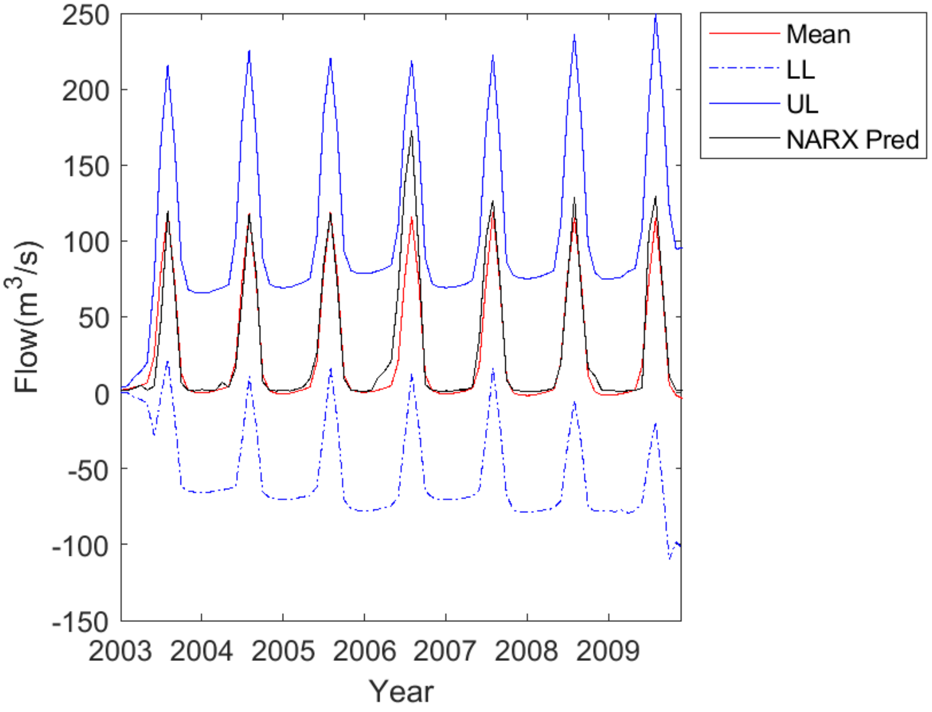


Figure 5.40: Melkakuntere station NARX bootstrapped (Mean and 95% Lower Limit(LL) - Upper Limit (UL) versus Prediction for test period.

Akaki Station

The same approach was applied here as well, and it was determined that model no. 7, encompassing mean monthly temperature, API, and flow, emerged as the best-performing model among the various suggested combinations as shown in Figure: 5.41. Despite the limited data length used for model training and validation, it is evident from Figures:5.42 and 5.43 that this model yielded satisfactory results.

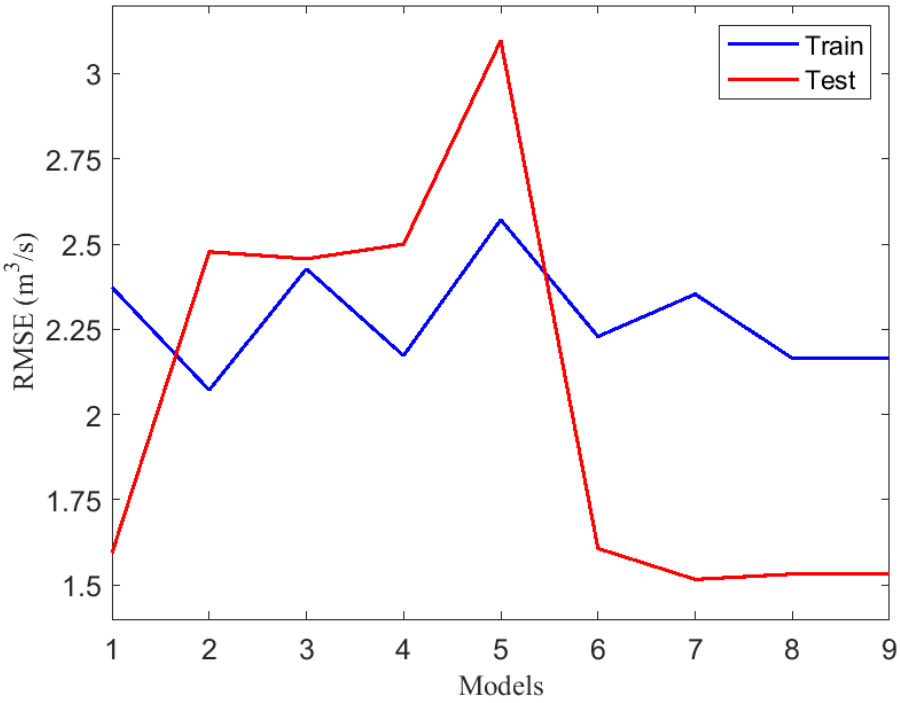


Figure 5.41: Akaki station NARX model performance using different input combinations for Train and Test period.

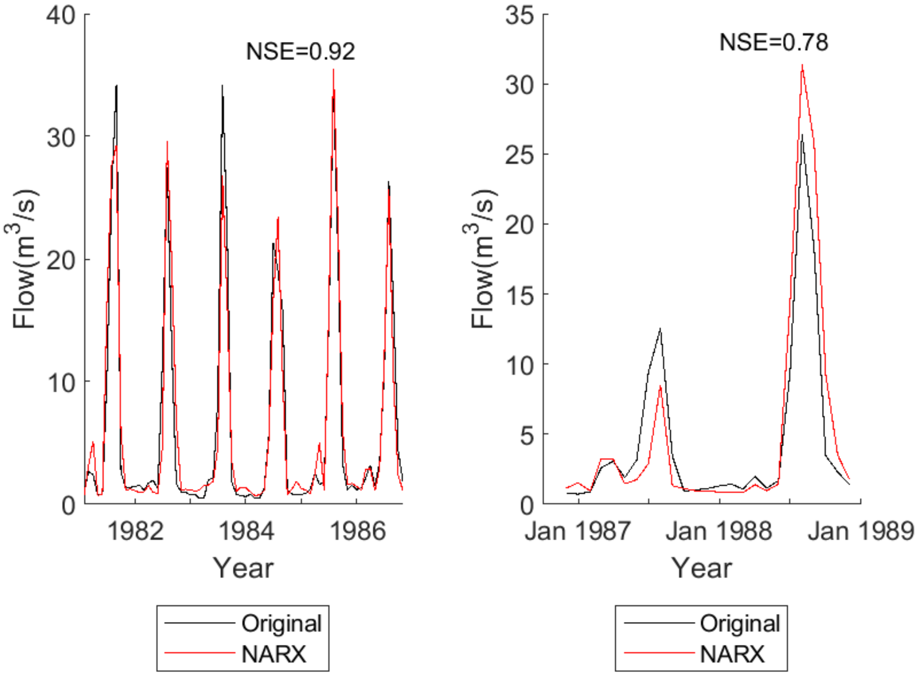


Figure 5.42: Akaki station NARX simulation for Train and Test period and its performance.

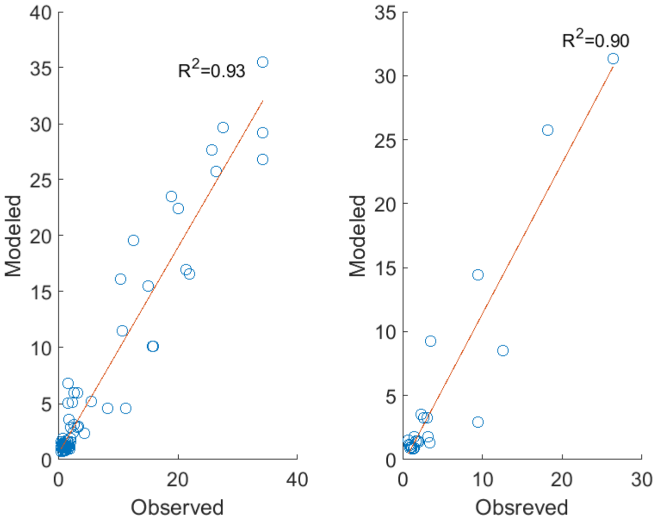


Figure 5.43: Akaki station Observed versus NARX output scatter plot for train (left) and test (right) period.

The model bootstrapping performed for the short validation period yielded satisfactory results for the mean prediction. However, it is noteworthy that the uncertainty band is notably higher when compared to the previous two stations. This outcome is predictable, given that the data used for this station has a shorter duration.

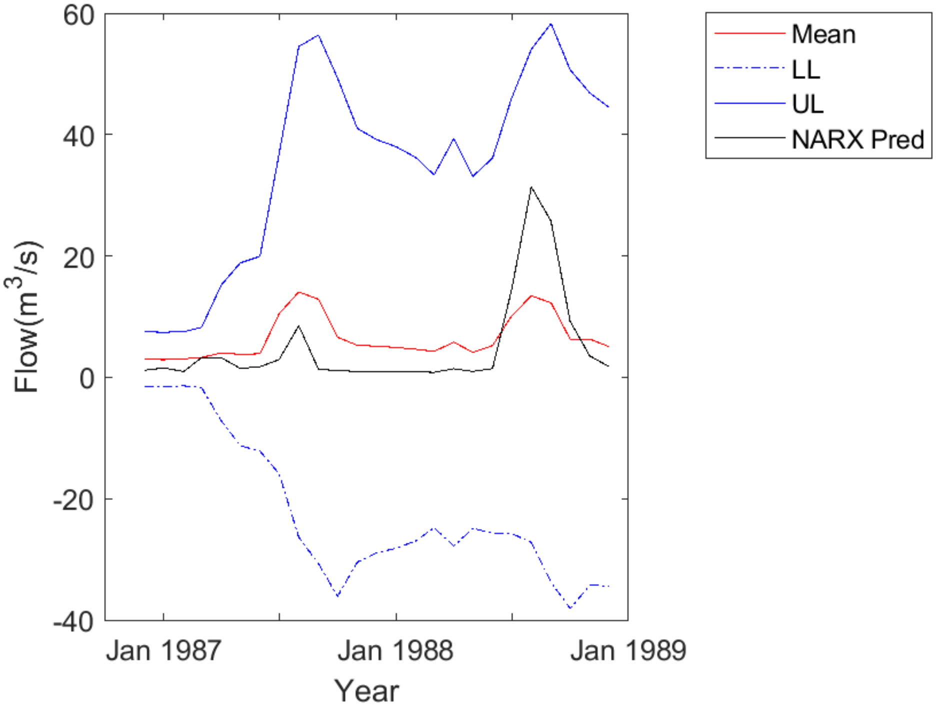


Figure 5.44: Akaki station NARX bootstrapped (Mean and 95% Lower Limit(LL) - Upper Limit (UL) versus Prediction for test period.

ANFIS Model Training, Validation, and Uncertainty

The development of the ANFIS model commenced with the identification of critical environmental factors influencing runoff production in the sub-basin. These factors encompass catchment morphology, climate, soil and geology, and land use land cover. However, obtaining these inputs as continuous time-dependent data (time series) can be challenging. Therefore, this study primarily utilized climatic factors and runoff data for forecasting. Monthly rainfall, mean monthly 30-day Antecedent Precipitation Index (API), monthly mean temperature, monthly potential evapotranspiration (ETo), and flow were selected as inputs for the study. A cross-correlation matrix was then created between these variables at different lags and a 12-month ahead flow to generate candidate model inputs.

Following this step, potential models were developed using various combinations of these candidate model inputs. The Fuzzy C-Mean (FCM) clustering technique was employed to generate linguistic labels for each variable. The sensitivity of the ANFIS structure to each model was analyzed by varying the number of clusters and epochs. The cluster numbers ranged from a minimum of 2 to a maximum of 13, corresponding to the total number of inputs. The number of epochs during both the training and testing phases varied from 100 to 300. The objective of the sensitivity analysis was to identify the optimal model structure for each scenario. Consequently, the final model selected was the one exhibiting the highest performance among the models considered. Other settings used in the ANFIS modeling are detailed in Table:5.11.

Table 5.11: General ANFIS settings used in model generation

Method for MF		ANFIS Parameters	
Cluster method	FCM	Max. Number of epochs	varying
no. of clusters	varying	Error Goal	0
Partition matrix exponent	2	Initial step size	0.01
Maximum number of iterations	100	Step size decrease rate	0.9
minimum Improvement	0.00001	Step size increase rate	1.1
Type of membership function	Gaussian	Optimization Algorithm	Hybrid (backpropagation+least Square)

The selected model was trained by partitioning the total time series into training and testing phases. Similar to the approach employed in previous NARX model, 75% of the data was utilized for training, with the remaining 25% reserved for testing.

Hombole Station

For this station, the methodologies discussed previously were rigorously adhered to, and the model candidates featuring various combinations of environmental factors are presented in Table:5.12. Additionally, Table:5.13 provides a comprehensive list of models along with their respective performance metrics obtained during the sensitivity analysis.

Table 5.12: Argument combinations used to build the ANFIS model

Inputs	Different Input Combinations (models)								
	1	2	3	4	5	6	7	8	9
Flow(t-12)	X	X	X	X	X	X	X	X	X
Flow(t-24)	X	X	X	X	X	X	X	X	X
Flow(t)	X	X	X	X	X	X	X	X	X
API(t+12)	X	X	X	X	X				
API(t-24)	X	X	X	X	X				
API(t-12)	X	X	X	X	X				
API(t)	X	X	X	X	X				
Rf(t+12)	X					X	X	X	X
Rf(t-24)	X					X	X	X	X
Rf(t-12)	X					X	X	X	X
Rf(t)	X					X	X	X	X
Eto(t+12)	X	X	X			X	X		
TMP(t+6)	X	X		X		X		X	

Observing Table:5.13 and the accompanying Figure 5.45, it is evident that model number 4 stands out as the optimal choice, boasting the lowest RMSE during the test period. The input combination used to construct this model comprises $[Flow(t-12) \quad Flow(t-24) \quad Flow(t) \quad API(t+12) \quad API(t-24) \quad API(t-12) \quad API(t) \quad TMP(t+6)]$.

Table 5.13: Best performing models for different Cluster and Epoch number for Hombole

Models	Train R	Test R	Train RMSE	Test RMSE	no. of clusters	Max. Number of epochs
1	0.955	0.93	0.282	0.372	3	200
2	0.975	0.935	0.214	0.351	4	200
3	0.974	0.934	0.221	0.395	5	100
4	0.972	0.949	0.229	0.313	5	200
5	0.967	0.94	0.247	0.343	4	100
6	0.971	0.931	0.232	0.344	4	300
7	0.968	0.944	0.242	0.334	5	100
8	0.966	0.935	0.248	0.329	6	100
9	0.96	0.918	0.268	0.391	4	200

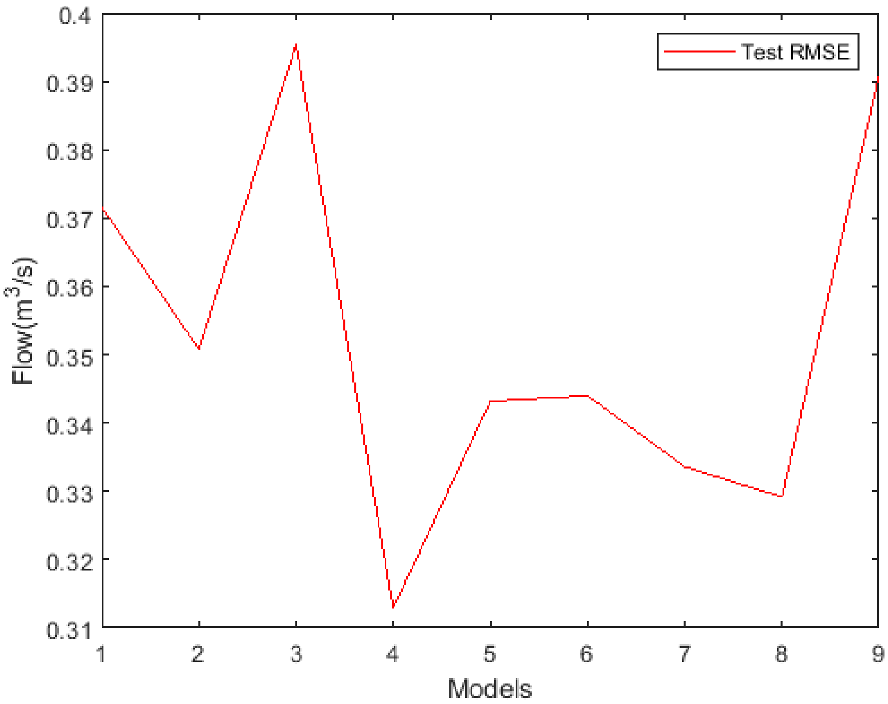


Figure 5.45: Performance of ANFIS model for different input combinations of variables to predict 12 month ahead flow.

After identifying the appropriate model, a simulation was executed using the trained model, and the results are visually represented in Figures: 5.46 and 5.47. Despite some exceptions in the peaks for certain years, the model, on the whole, demonstrated the capability to adequately capture seasonal variations and low flows.

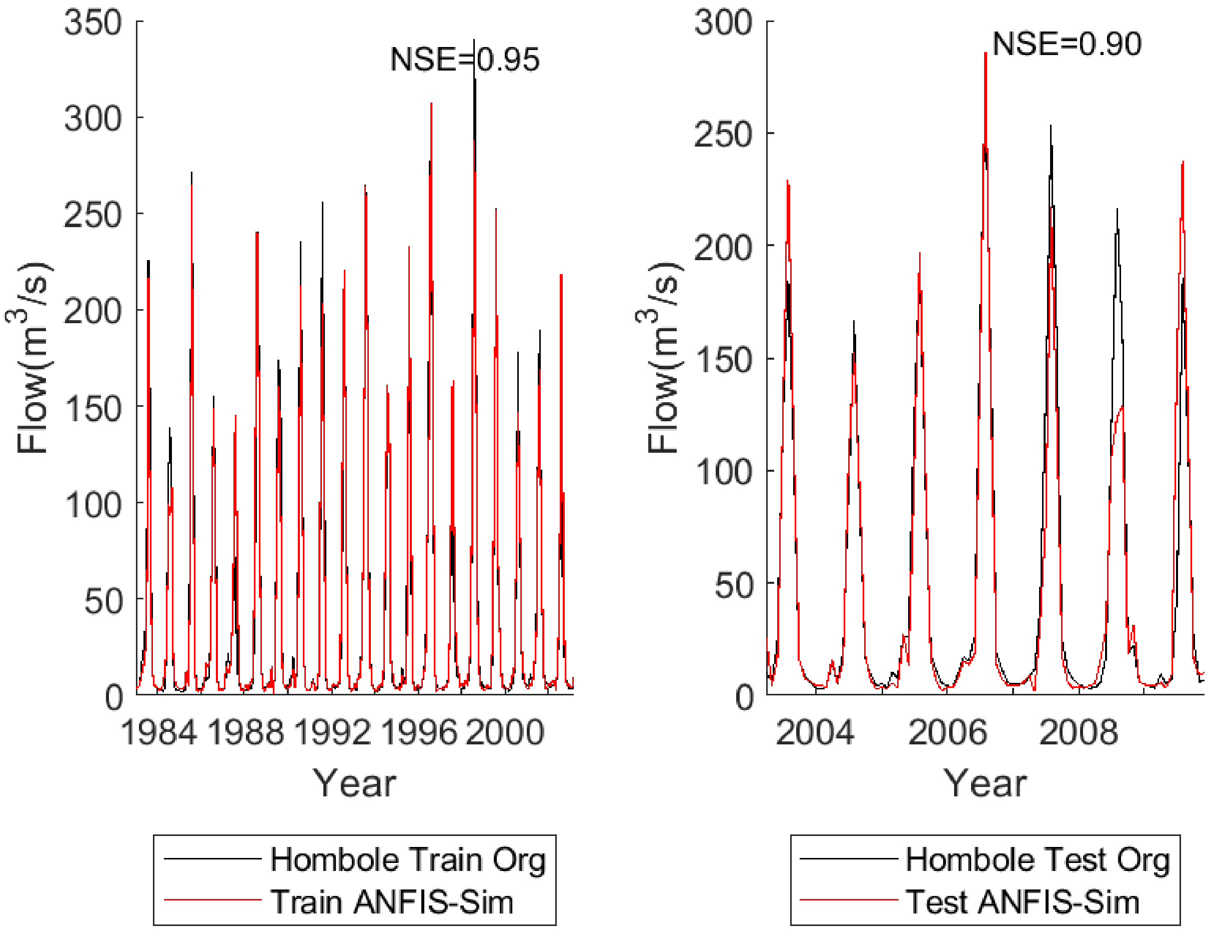


Figure 5.46: Hombole station ANFIS simulation for Train and Test period and its performance.

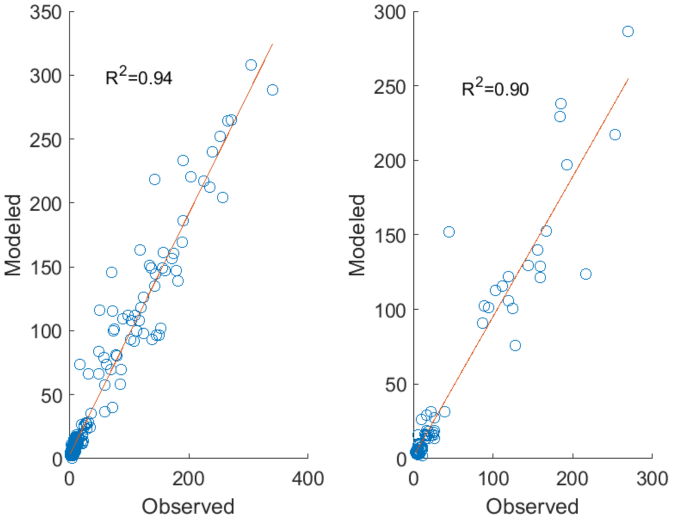


Figure 5.47: Hombole station Observed versus ANFIS output scatter plot for train (left) and test (right) period.

The uncertainty of the ANFIS model predictions was similarly assessed through a bootstrapping technique. In this analysis, 300 different input sets of the selected model were generated by bootstrapping the input series. The input series were bootstrapped with replacement using the datasample() function in MATLAB. Subsequently, a 95% confidence limit was constructed using the mean and standard deviation calculated from these 300 simulations. The plot of the result of bootstrapping in Figure:5.48 indicates a high similarity between the mean of the bootstrapped results and the model predictions. This consistency underscores the model's reliability in predicting future values for inputs with similar statistical characteristics (stationarity and distribution).

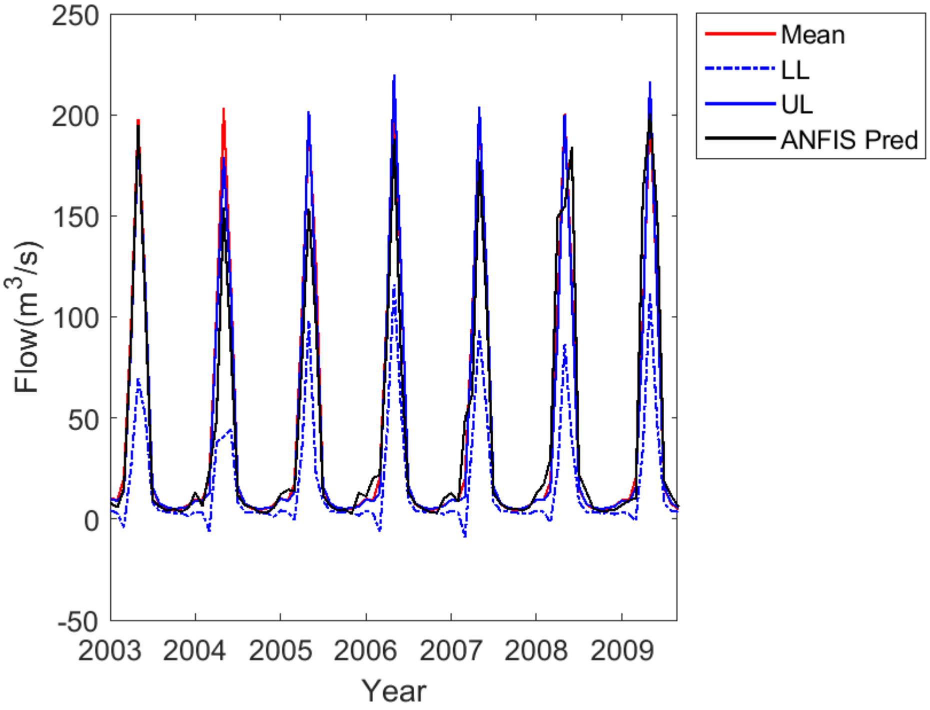


Figure 5.48: Hombole station ANFIS bootstrapped (Mean, Lower Limit (LL), and Upper Limit(UL)) versus Prediction for test period.

MelkaKuntere Station

From the correlation analysis between multiple lagged series, preliminary input combinations were identified for this station as well, and these are outlined in Table:5.14.

Table 5.14: Argument combinations used to build the ANFIS model for Melkakuntere

Inputs	Different Input Combinations									
	1	2	3	4	5	6	7	8	9	10
Flow(t-12)	X	X	X	X	X	X	X	X	X	X
Flow(t-24)	X	X	X	X	X	X	X	X	X	X
Flow(t)	X	X	X	X	X	X	X	X	X	X
API(t+12)	X	X	X	X	X					
API(t-24)	X	X	X	X	X					
API(t-12)	X	X	X	X	X					
API(t)	X	X	X	X	X					
Rf(t+12)	X					X	X	X	X	X
Rf(t-24)	X					X	X	X	X	
Rf(t-12)	X					X	X	X	X	
Rf(t)	X					X	X	X	X	X
Eto(t+12)	X	X	X			X	X			
TMP(t+6)	X	X		X		X		X		

Similar procedures were followed, mirroring the methodology employed at the Hombole station, by varying cluster and epoch numbers to identify a model that yields the lowest RMSE among the candidates listed in Table:5.14. The results of this analysis are summarized in Table:5.15.

Table 5.15: Best performing models for different Cluster and Epoch number for Melkakuntere

Models	Train R	Test R	Train RMSE	Test RMSE	no. of clusters	Max. Number of epochs
1	0.972	0.899	0.204	0.488	2	100
2	0.972	0.896	0.204	0.496	3	100
3	0.971	0.896	0.207	0.498	2	100
4	0.97	0.894	0.21	0.493	2	100
5	0.998	0.906	0.055	0.489	9	100
6	0.968	0.894	0.216	0.497	2	100
7	0.968	0.896	0.216	0.496	2	100
8	0.966	0.894	0.222	0.493	2	100
9	0.966	0.896	0.223	0.493	2	100
10	0.953	0.915	0.263	0.429	2	100

The results summarized in Table:5.15 indicate that Model 10 yields a lower Test RMSE result and is consequently selected. The input combinations used for this model are as follows: $[Flow(t-12)Flow(t-24)Flow(t)Rf(t-12)Rf(t)]$. Employing these input combinations, the model was trained with 75% of the total series, and the remaining 25% was used to test the model. The results of the final simulations are presented in Figures:5.49 and 5.50.

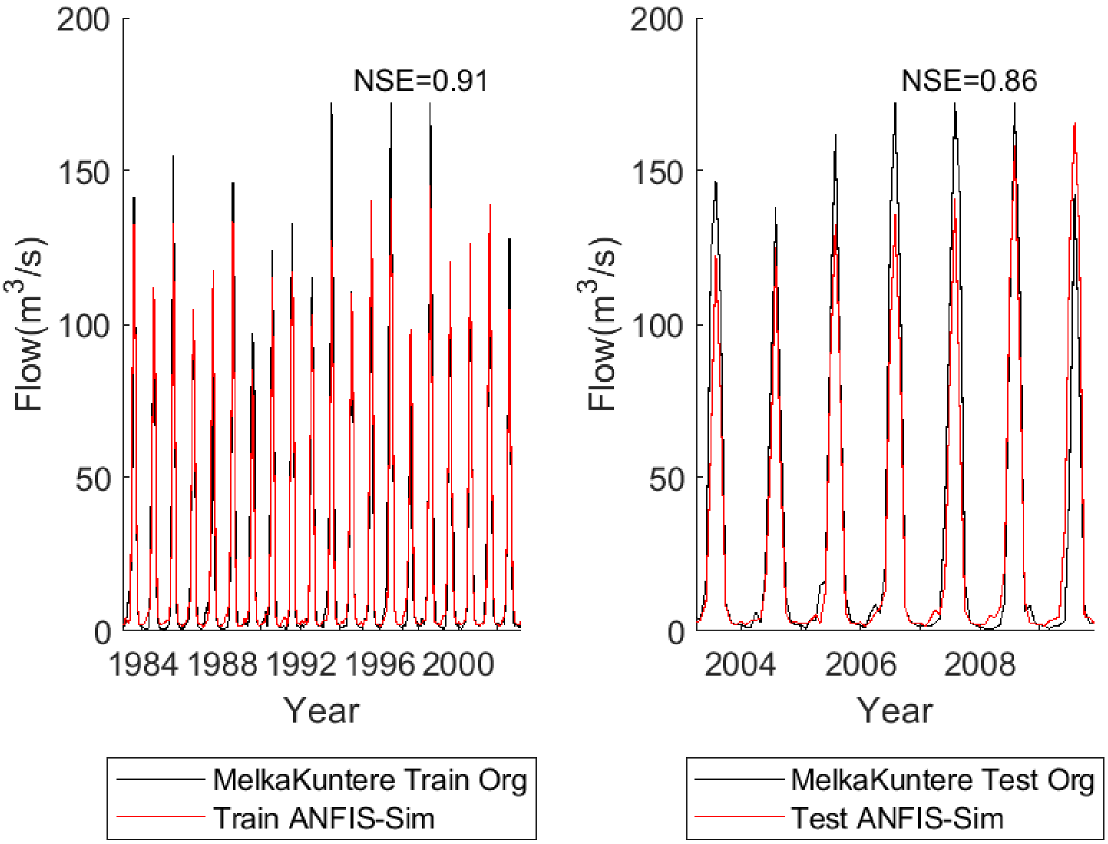


Figure 5.49: Melkakuntere station ANFIS simulation for Train and Test period and its performance.

The model successfully captured the seasonal behavior and low flows, but in contrast to the Hombole station, it struggled to accurately represent high flows.

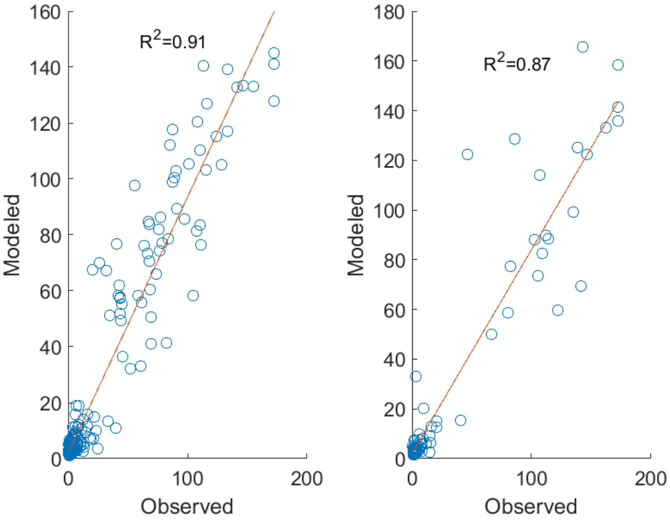


Figure 5.50: Melkakuntere station Observed versus ANFIS output scatter plot for train (left) and test (right) period.

The results of bootstrapping indicate that the ANFIS simulation consistently falls well within the 95% confidence limit for almost the entire time series length. However, as the years progress, particularly in the later years (2008-2009), it becomes apparent that the model predictions start to deviate more from the upper confidence limit.

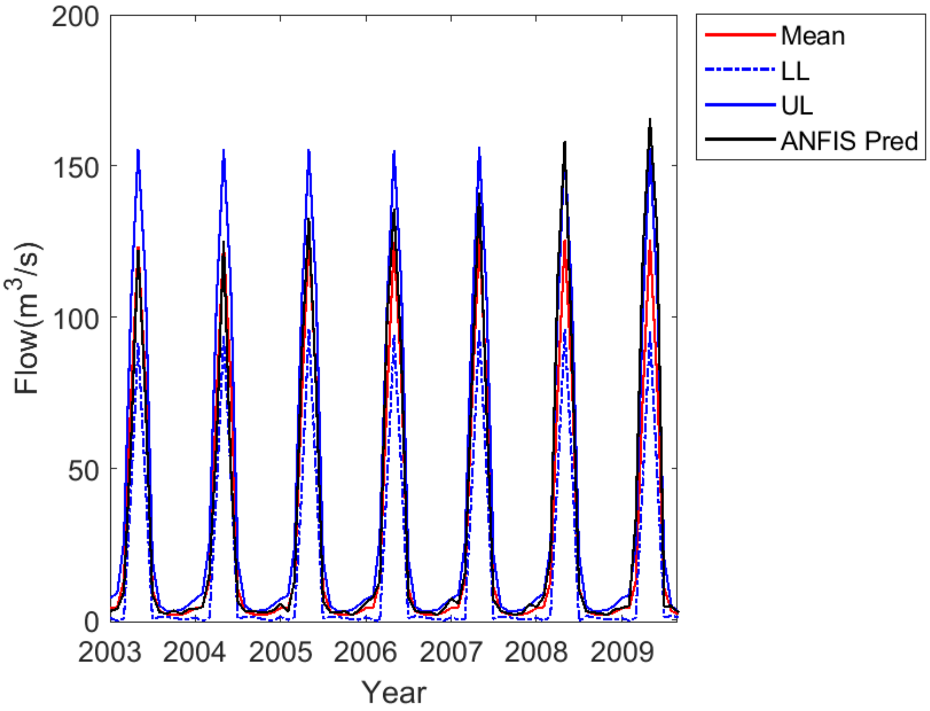


Figure 5.51: Melkakuntere station ANFIS bootstrapped (Mean, Lower Limit (LL), and Upper Limit(UL)) versus Prediction for test period.

Akaki Station

Despite the station having a very limited dataset, similar procedures were employed to identify an appropriate model. Consistent with the approaches discussed for the other two stations, input combinations were defined, as illustrated in Table:5.16.

Table 5.16: Argument combinations used to build the ANFIS model for Akaki station

Inputs	Different Input Combinations										
	1	2	3	4	5	6	7	8	9	10	11
API(t+12)	x	x	x	x	x	x	x	x			x
Flow(t-12)	x	x	x	x	x	x	x	x	x	x	x
Rf(t+12)	x	x	x	x	x	x			x	x	
API(t)	x	x	x	x	x	x	x	x			x
API(t-12)	x	x	x	x	x	x	x	x			x
Flow(t)	x	x	x	x	x		x	x	x	x	x
Rf(t)	x	x	x	x					x	x	
Rf(t-12)	x	x	x						x	x	
Eto(t+12)	x	x					x		x		x
TMP(t+12)	x						x		x		

The performance of the 11 models outlined in Table:5.16 was assessed using various combinations of cluster and epoch numbers. The results of this analysis are presented in Table:5.17, revealing that Model number 7 exhibits the lowest RMSE.

Table 5.17: Best performing models for different Cluster and Epoch number for Akaki

Models	Train R	Test R	Train RMSE	Test RMSE	no. of clusters	Max. Number of epochs
1	1	0.88	0	0.467	15	100
2	1	0.85	0	0.505	12	100
3	0.997	0.874	0.086	0.486	2	300
4	0.999	0.875	0.036	0.536	5	200
5	0.976	0.832	0.235	0.622	2	100
6	0.994	0.854	0.121	0.631	3	100
7	1	0.918	0	0.348	15	300
8	1	0.805	0.005	0.6	13	100
9	1	0.869	0	0.483	15	200
10	0.986	0.831	0.178	0.592	2	100
11	1	0.88	0	0.468	12	300

The corresponding input combinations for the selected Model number 7 are $[API(t+12)\quad Flow(t-12)\quad API(t)\quad API(t-12)\quad Flow(t)\quad Eto(t+12)\quad TMP(t+12)]$. The resulting simulations for both the training and testing phases are depicted in Figures:5.52 and 5.53. It is evident that the selected model exhibits excellent performance in the training phase and relatively very good performance in the test phase.

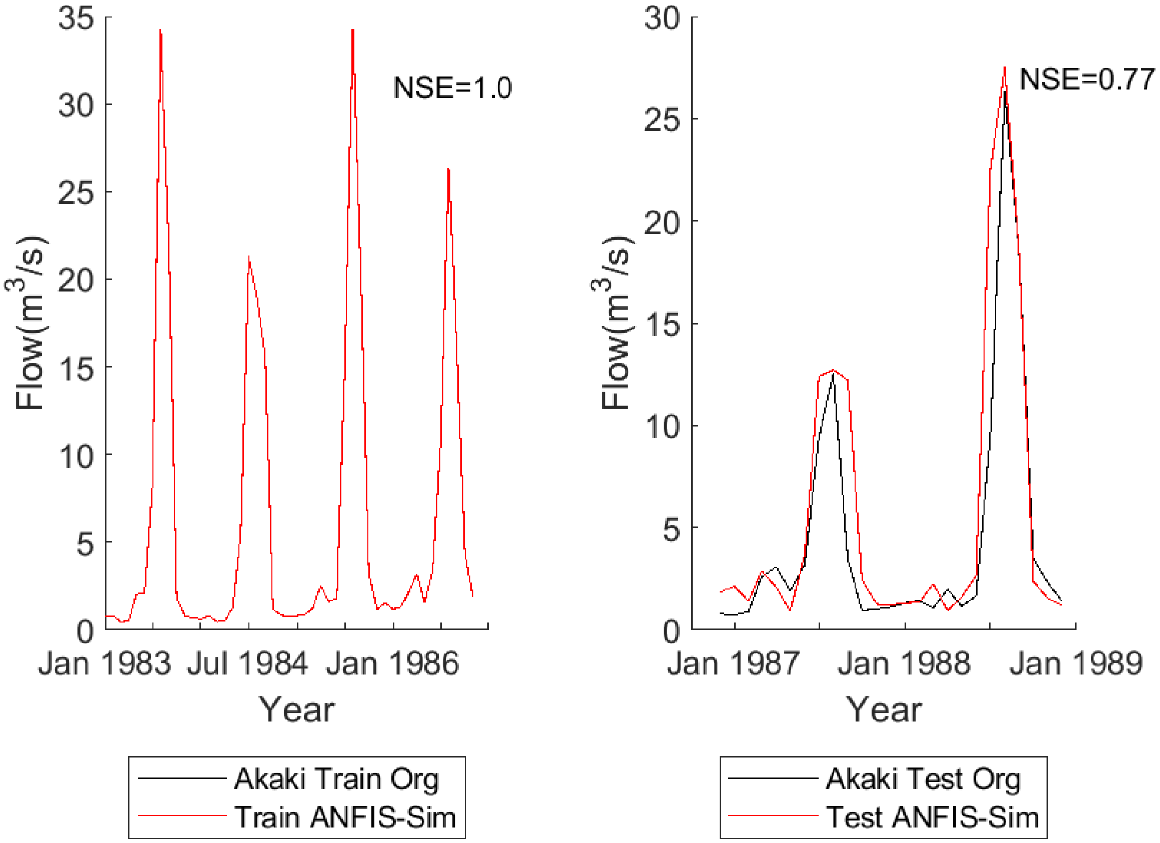


Figure 5.52: Akaki station ANFIS simulation for Train and Test period and its performance.

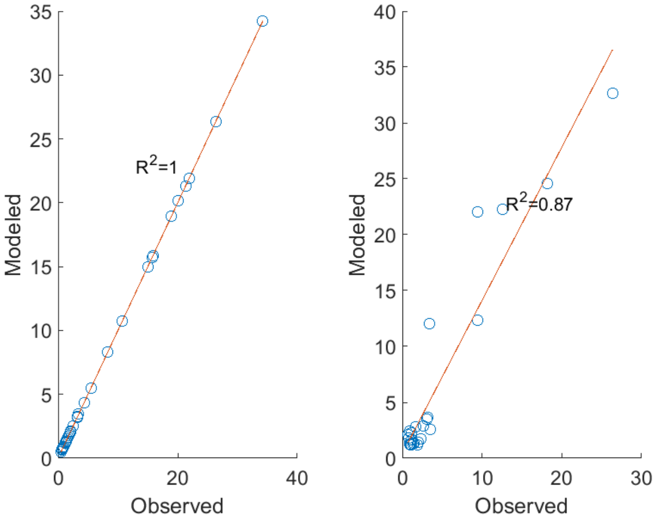


Figure 5.53: Akaki station Observed versus ANFIS output scatter plot for train (left) and test (right) period.

In this station as well, the mean of the bootstrapped 300 simulations closely aligns with the simulation conducted using actual data, signifying model consistency. However, the 95% uncertainty bands in Figure:5.54 indicate higher uncertainty in predictions compared to similar analyses conducted on the Hombole and Melkakuntere stations. As previously mentioned, this heightened uncertainty can be attributed to the relatively shorter data length used in the simulation process.

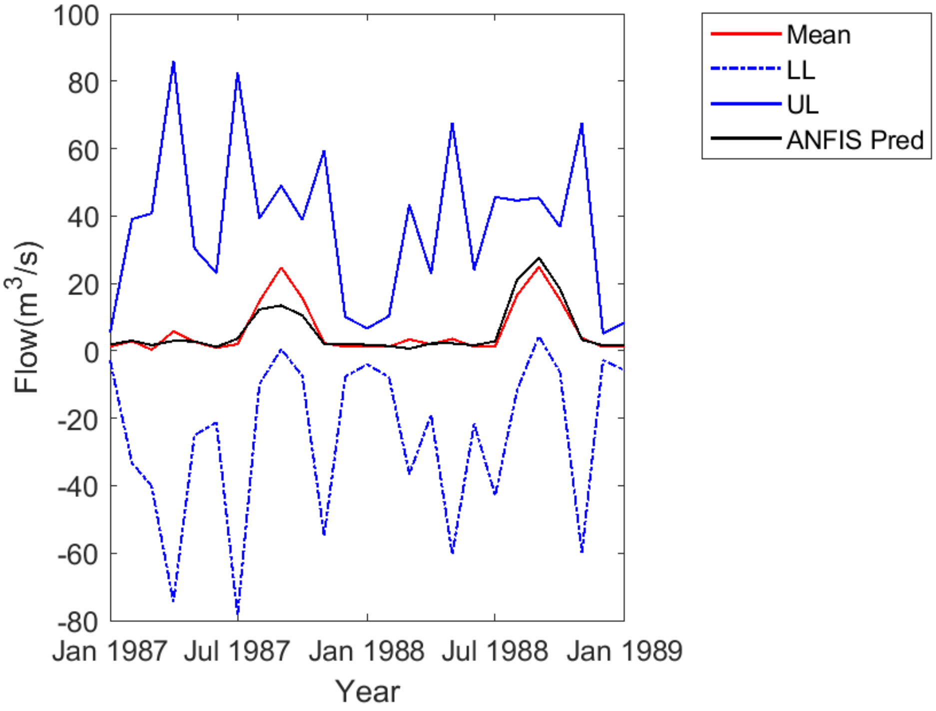


Figure 5.54: Akaki station ANFIS bootstrapped (Mean, Lower Limit (LL), and Upper Limit(UL)) versus Prediction for test period.

5.5 Discussion

The primary motivation behind this section of the study arises from the curiosity to measure the performance of different time series models in forecasting flow series from selected rivers over the Upper Awash Sub-Basin (UASB). One of the major challenges encountered during this analysis was related to data quality, including factors such as having sufficient data length, fewer missing data, and handling outliers. Consequently, the initial step involved addressing this issue using appropriate tools and methods.

In the UASB, four major rivers were identified: Akaki, Melkakuntere, Hombole, and Mojo. While Hombole and Melkakuntere exhibited relatively good data length and quality, Akaki and Mojo faced challenges. Particularly, the flow data quality in the Mojo river was one of the poorest, making time series analysis difficult. Therefore, this study focused solely on Akaki, Melkakuntere, and Hombole rivers.

The type of time series forecasting techniques applied in this study mainly fall into two broad categories: regime-switching and data-driven techniques. From the regime-switching techniques, the potential of two threshold autoregressive methods known as SETAR and

LSTAR models were investigated. In order to apply these techniques, a box-cox transformation was performed on the original flow series, followed by the removal of the seasonal term through fitting a Fourier Harmonic regression equation with a period of two. Finally, the actual TAR model was fitted to the remaining residual series. The biggest challenges in fitting such kind of models is the estimation of the threshold value. In this study, the grid search algorithm available under tsDyn package [129] was used. By overlaying the threshold values over the flow series, it can be understood that for the Hombole station, regime change has occurred due to a shift in mean. However, for both Akaki and Melkakuntere station, there is no observable mean shift in the flow series. It can be concluded that the estimated threshold values are due to the high variability of the residual flow series. Therefore, it can be observed that the residual flow data at these stations was divided into low and high variance series. The forecast performance was done after combining the outputs of TAR models with a seasonal term. Both the SETAR and LSTAR forecasts have shown that the forecast performance at Hombole and Akaki were relatively higher than Melkakuntere river. This is due to the high variability of the residual flow series of Melkakuntere, which makes the determination of regime difficult. It was also observed that the forecasts made using the LSTAR model were of lower performance than the SETAR model. This is due to the challenging task of obtaining a transition function in the LSTAR model which fits well to the residual flow series, especially for Akaki and Melkakuntere.

The second type of time series forecasting technique implemented was the NARX model, which is one of the different versions of the ANN model. As it's known, ANN models have good characteristics that make them perform well when simulating non-linear dependencies present in time series. The NARX model, in addition, has high potential in describing long-term dependency that is present in monthly flow series [132]. Model definition using NARX can be performed using a number of variables thought to have an impact on flow. In this study, monthly rainfall, mean monthly temperature, monthly evapotranspiration, and mean monthly 30-day antecedent precipitation index have been selected as input variables for generating candidate models. An optimal model architecture is defined for each candidate model using different combinations of those input variables and flow as an output. In the configuration of the NARX model, the most important parameters are delay, the number of hidden layers, and the number of nodes in each layer. In each model, delay (d) and the number of hidden layers were kept to a value of one. And the optimal model was found by varying the number of nodes in the hidden layer and measuring its performance using RMSE. For both Hombole and Melkakuntere station, the NARX model developed with rainfall as an input and flow as an output variable gave the best result among other input combinations. Whereas, for Akaki mean temperature and API were used as an input and flow as output. The number of nodes used in the hidden layer varies for each station and are found to be 14, 7, and 9 for Akaki, Melkakuntere, and Hombole, respectively. All the models trained for each gauging station have produced a satisfactory forecast with NSE value greater than 0.75, with the forecast at Hombole being better than the others. In model setup of the NARX model, Levenberg Marquardt with Bayesian regularization (trainbr) gave the best forecast performance than other training algorithms such as gradient descent. The potential of this training algorithm has been proved through the consistent model prediction shown in the uncertainty plots developed for each gauging station.

In the same manner as the methodology used for the NARX model, ANFIS candidate models were also built using different combinations of input variables. The input variables used to build a model came from a correlation matrix. The input variables lagged at different numbers of months ranging from 3 - 24 were correlated with a 12-month ahead flow series to form a correlation matrix. Candidate models were developed using different combinations of these lagged input variables. The definition of the linguistic labels in the antecedent side was done with the help of FCM algorithm and Gaussian membership functions. From preliminary analysis, this algorithm has shown better performance than grid partitioning technique. The model architecture of each candidate model was assessed using RMSE as a performance measure. The important parameters used to build ANFIS model are the number of clusters and epochs. By varying these values and through a trial and error approach, an appropriate model has been identified for each candidate model. Here also, RMSE has been used to evaluate model performance. The final models selected at each gauge station hold a different number and type of input variable combinations. In Hombole station flow, API, and mean temperature have been used in the final model. And, in Melkakuntere only two variables mean temperature and flow were used. In Akaki station, flow, API, Eto, and mean temperature were applied.

5.6 Conclusion

The results clearly indicate that data-driven techniques exhibit superior forecast capabilities compared to TAR models for each gauging station. The simplicity and ease of model setup and training, facilitated by the functionalities in the Deep Learning Toolbox in MATLAB [97], make these techniques practical and straightforward. However, a significant challenge lies in identifying the optimal set of model inputs that yield satisfactory performance. Future efforts to enhance forecast accuracy should explore models that can effectively distinguish between deterministic and stochastic components.

In the analysis of these data-driven techniques, it becomes evident that the ANFIS model outperforms the NARX model, showcasing a relatively superior performance.

Chapter 6

Future Water Availability in the Sub-Basin

6.1 Introduction

Water is an indispensable resource, serving myriad purposes from sustaining all forms of life on Earth to driving cutting-edge industrial applications. The absence of water in our world is unimaginable. Over 71% of the Earth's surface is composed of water, and within this vast expanse, 96.5% resides in oceans [136]. Ethiopia boasts an estimated water potential of around 122 BCM, with the Awash basin alone holding 4.9 BCM, accounting for approximately 0.04% of the country's total potential [17]. The Upper-Awash Sub-Basin (UASB) emerges as a key player, contributing significantly to the main Awash River due to its comparatively higher magnitude of rainfall within the broader Awash Basin [137] [60].

The UASB holds significant strategic importance for the country due to its geographic location and vibrant socio-economic activity. The basin, housing the capital city and serving as a hub for most large industries, witnesses intense socio-economic activity. The basin is home to over 6 million people, with the majority residing in the capital city and nearby towns [27]. In these areas, people are predominantly engaged in industries, both public and private sectors, while farming remains the dominant economic activity outside the cities and towns. The economic disparity results in high growth and a related population increase in cities and towns. Despite various water resource developments and future plans in the sub-basin, mainly focused on domestic water consumption, any development in these areas is expected to create water stress downstream, emphasizing the importance of studying past, present, and future water resources in the sub-basin.

Several approaches exist for studying water resources in a given area. The first involves analyzing long-term variations in runoff and meteorological elements, either through statistical analysis of the relationship between runoff and other meteorological variables or by observing past extreme events [7] [8] [9]. These meteorological variables include precipitation, temperature, and evapotranspiration. The second approach employs a water balance method over an extended period [10] [11]. The third approach involves studying estimates of changes in climatic and hydrologic characteristics for large regions using Global Climate Model (GCM) outputs [12]. Another common alternative nowadays is the use of deterministic hydrologic models [13] [14].

Historically, numerous efforts have been made to study water resource conditions in the basin. The earliest, by Hailemariam K. (1999) [57], used a water balance model to explore the impact of climate change on runoff estimation in the Awash basin using climate models under the CMIP3 archive. The study concluded a decrease in future water resource potential for all scenarios and time periods. Subsequently, Taye et al. (2018) [58] investigated water availability using climate models under CMIP5 until the end of the 21st century. This study estimated changes in water resource availability using a change factor (CF) developed by subtracting evaporation from precipitation at the monthly level. An increase in water availability was predicted for recent and mid-time periods for all scenarios, while a decrease was predicted for the far-time period. Heyi et al. (2020) [138] applied the HBV model together with the HadCM3 climate model, an output of earlier works under CMIP3, indicating a decrease in rainfall during JJAS and an increase in MAM for two scenarios and all time periods. A recent study by Chelkeba et al. (2023) [139] applied the SWAT model with four CORDEX RCMs from the CMIP3 archive on the Akaki Catchment, aligning with the results obtained by Taye et al. (2018) [58].

In this study, similar efforts have been undertaken to study future availability conditions in UASB by using SWAT and NARX models together with an ensemble of climate model outputs obtained from the new CMIP6 archive. It has been discussed in Chapter Two that models incorporated under CMIP6 are thought to show comparable or improved capabilities in terms of simulating global climate compared to the older CMIP5. Therefore, the outputs of this study will further shed light on the future water resource condition over the sub-basin with the help of these new climate model outputs.

6.2 Data Sets

As mentioned previously, two rainfall-runoff models will be employed for simulating streamflow. The first model, the Soil and Water Assessment Tool (SWAT), will utilize a 30m Digital Elevation Model (DEM) obtained from https://earthexplorer.usgs.gov/, along with grid-based Land Use Land Cover (LULC) data (year 2013) and soil data. These LULC and soil data are derived from the Ministry of Water and Energy (MoWE) report (AWBA, 2017) [1]. On the other hand, the NARX (Nonlinear AutoRegressive with eXogenous inputs) model will exclusively use flow, rainfall, and mean temperature data at a monthly time step. The climate dataset employed in the SWAT model comprises daily records of precipitation, maximum temperature, minimum temperature, relative humidity, solar radiation, and wind speed. A total of 15 rainfall and 10 temperature stations obtained from the National Meteorological Agency (NMA) of Ethiopia are utilized, and the list of stations, along with additional information, is presented in Table 2.1. Similarly, daily river flow data for the gauging station at Hombole, considered as the outlet of the sub-basin, was gathered from the Ministry of Water and Energy (MoWE). All of the time series datasets were collected within a common time frame of 1980 – 2009.

6.3 Methods

6.3.1 Surface Runoff Simulation in SWAT Model

The Soil and Water Assessment Tool (SWAT) is a hydrological model initially introduced by Dr. Jeff Arnold at the United States Department of Agriculture (USDA) for Agricultural Research Services (ARS) in the early 1990s [140]. Widely acknowledged, SWAT is designed to simulate the movement of water, sediment, and nutrients at the scale of a river, watershed, or basin. It has become a cornerstone in various environmental and water resource management applications.

SWAT operates as a semi-distributed, process-based model that partitions a watershed into multiple sub-basins, further breaking them down into smaller hydrological response units (HRUs). Executing simulations on a daily time step, the model replicates a spectrum of hydrological processes, encompassing precipitation, runoff, evapotranspiration, infiltration, soil erosion, and nutrient transport.

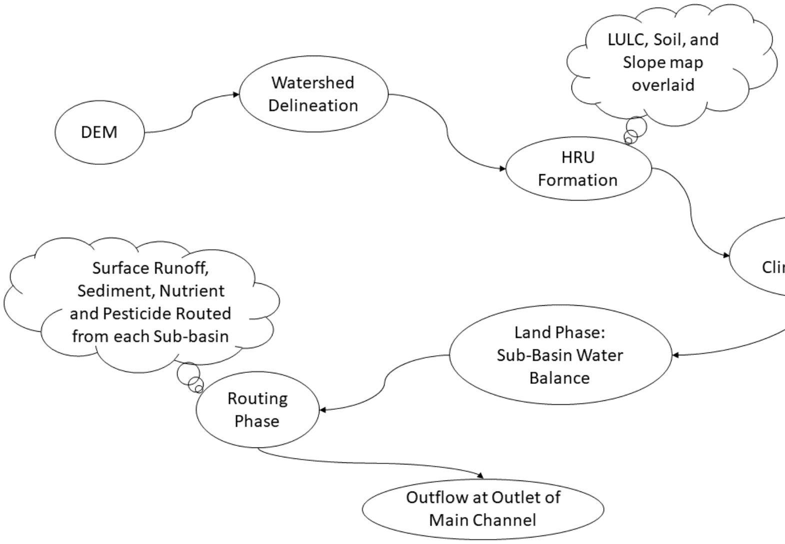


Figure 6.1: General steps followed by SWAT Hydrological Model

The initial phase of SWAT modeling involves watershed delineation utilizing a Digital Elevation Model (DEM). This delineation is followed by subdividing the watershed into sub-basins, which are further partitioned into Hydrological Response Units (HRUs) based on topography, land use, and soil types (Figure: 6.1). In the third step, daily weather data, such as precipitation, temperature, wind speed, relative humidity, and solar radiation, are incorporated into the hydrological process. The fourth step sees SWAT simulating the movement of rainfall through various pathways, encompassing surface runoff and infiltration into the soil. This stage also considers water loss through evaporation from the soil surface and transpiration by plants. The model captures the movement of water into the soil, accounting for factors like soil type and land cover at the HRU level. Following the estimation of various processes in this step, a water balance is executed at the sub-basin level, as expressed by Equation 6.1. The surface runoff generated from each HRU is aggregated to form streamflow within sub-basins, ultimately routed downstream through the main channel.

$$SW_t = SW_o + \sum_{i=1}^{t} \left(R_t - Q_{surf} - E_a - W_{sweep} - Q_{GW} \right) \tag{6.1}$$

Where: SW_t - is soil water content at end of time period t, SW_o - is initial soil water content, R_t - rainfall at time period t, Q_{surf} - Surface runoff, E_a - actual Evapotranspiration, W_{sweep} - amount of water entering vadose zone, and Q_{GW} - amount of Return Flow.

SWAT also possesses the capability to model the transport of nutrients (such as nitrogen and phosphorus) and sediment throughout the watershed, taking into account factors like land use, erosion, and nutrient application.

In this study, the Hargreaves method has been employed for evapotranspiration estimation, while default settings have been utilized for other hydrological variables. Post-simulation, SWAT necessitates calibration and validation using observed data to ensure that simulated results align with real-world conditions. Once calibrated and validated, SWAT becomes a powerful tool for scenario analysis, enabling the assessment of the impacts of land use changes, climate change, or management practices on both water quality and quantity.

SWAT generates a variety of outputs, including daily and monthly streamflow, sediment yield, nutrient concentrations, and more. These outputs are invaluable for making informed decisions regarding watershed management and environmental protection.

6.3.2 SWAT Model Sensitivity, Uncertainty, Calibration and Validation

Once the SWAT model is set up, the subsequent task is to evaluate its accuracy. This is achieved by first identifying model parameters that have a significant effect on the required output. This crucial step is performed before initiating any calibration process. In the context of SWAT, this identification is often carried out using a tool known as SWAT-CUP [141]. SWAT-CUP employs various algorithms for model uncertainty, sensitivity, and calibration, and in this study, one of the most widely applied algorithms, SUFI-2, has been utilized.

When employing the SUFI-2 algorithm in SWAT-CUP, two types of model sensitivity analysis can be conducted: One At a Time (OAT) for local sensitivity or All At a Time (AAT) for global sensitivity [142]. In the former, all parameters are held constant except one, and the impact on the objective function (model output) is observed through 3 to 5 model runs. In the latter, all parameters are varied, and 500 to 1000 or more runs may be required. This algorithm establishes a regression equation (Equation: 6.2) between an objective function and parameters to identify less sensitive parameters that can be excluded from the subsequent calibration process.

$$g = \alpha + \sum_{i=1}^{m} \beta_i b_i \tag{6.2}$$

Where: g - is an objective function, α - is a regression constant, β - is a coefficient of parameter, b - is the parameter, m -is the number of parameters.

The identification of less sensitive parameters in the regression equation above is facilitated by a t-test. The t-test is computed by dividing the coefficient of a parameter by its standard error. Consequently, a parameter with a $>> |t|$ and $<< p$ value is considered relatively more sensitive compared to others.

Once the model parameters that are more sensitive to the objective function are identified, the subsequent step involves the calibration of these parameters. Calibration is a technique employed to minimize the difference between model simulation and observed values using an optimization function that satisfies a specific objective function. In SWAT-CUP, under the SUFI-2 algorithm, 10 different objective functions are available for use. In this study, the Nash-Sutcliffe efficiency (NS) (Equation: 6.3) has been chosen as the objective function [141].

$$Maximize : NS = 1 - \frac{\sum_i (Q_o - Q_s)_i^2}{\sum_i (Q_{o,i} - \bar{Q}_o)^2} \tag{6.3}$$

Where; Q - is a variable (e.g discharge) and the bar over it means average. Whereas the o and s indicate measured and simulated variable respectively. In addition, if there are more than one variable used, then the objective function would be a weighted sum of j variables ($g = \sum_j w_j NS_j$).

In SWAT-CUP, all three procedures occur sequentially, beginning with sensitivity analysis followed by calibration and then uncertainty analysis. The calibration procedure employs a Latin Hypercube Sampling technique to sample a set of parameter values from initial absolute ranges at each iteration step.

Upon completion of the calibration procedure, it is often necessary to conduct uncertainty analysis. Model uncertainty, in theory, arises from various sources, including assumptions in model building, parameterization, definition of the objective function in calibration, the optimization algorithm used, and potential non-uniqueness (equi-finality). Therefore, in SWAT-CUP, uncertainty analysis is performed at the end of the calibration procedure by utilizing a 95% prediction uncertainty (95PPU). SWAT-CUP establishes the 95PPU band and employs two statistics—p-factor and r-factor—to assess model output (Equation: 6.4). The p-factor indicates the percentage of measured data within the 95PPU band, while the r-factor illustrates the thickness of the 95PPU band.

$$r - factor_j = \frac{\frac{1}{n_j} \sum_{t_i=1}^{n_j} \left(x_s^{t_i,97.5\%} - x_s^{t_i,2.5\%} \right)}{\sigma_{oj}} \tag{6.4}$$

Where: $x_s^{t_i,97.5\%}$ and $x_s^{t_i,2.5\%}$ are the upper and lower boundary of the 95PPU at time step t and simulation i, n_j is number of data points, σ_{oj} is standard deviation of the j^{th} observed variable.

The final stage of model verification is the validation stage, and in SWAT, this is accomplished using a new set of parameters obtained at the final stage of the calibration procedure, along with an independent observed dataset. It is essential that the observed dataset used for validation differs from the one employed in the calibration procedure. Additionally, the model setup used during this stage should be consistent with the setup applied during the calibration period. However, caution must be exercised when selecting the observed dataset, ensuring that it possesses statistical similarity to the one used during the calibration period.

6.3.3 Defining Water Availability in the Sub-Basin

One of the techniques employed to study past, present, or future water resource conditions involves understanding the characteristics of drought in a given locality. Drought can be characterized by its duration, intensity, accumulated deficit water volume (severity), or spatial extent. Two widely used approaches for characterizing drought are the application of drought indices and the threshold method [143].

Drought indices are typically classified as either single-value or time series indices, sampled at regular or irregular intervals. Examples of single-value indices include the 95th percentile, while the mean annual n-day minimum flow (MAM(n-day)) is an example of a time series index.

The second technique for characterizing a drought event is the Threshold Level Method. In this approach, resultant indices that define an event above or below a given threshold are applied. These thresholds can be either constant or variable over a specified time period of analysis.

In this study, the Standardized Streamflow Index (SSI), a time series drought index, along with frequency analysis, was used to evaluate the outputs of SWAT and NARX model simulations in terms of future water resource conditions in the sub-basin. Both analyses were conducted using monthly time series data from the Hombole gauge station. The equation used to calculate SSI is given by Equation 6.5 [144]. SSI was computed for 3 and 12-month time steps after fitting a gamma distribution to the monthly flow series using the SPEI package in R [145].

$$SSI = W - \frac{C_o + C_1W + C_2W^2}{1 + d_1W + d_2W^2 + d_3W^3} \tag{6.5}$$

Where $W = \sqrt{-2ln(p)}$ for $p \leq 0.5$, p is the probability of exceeding a given value x, and $p = 1 - F(x)$. Also, if $p > 0.5$, p will be replaced by $1 - p$ and the sign of SSI will be reversed. $C_o = 2.515517$; $C_1 = 0.802853$; $C_2 = 0.010328$; $d_1 = 1.432788$; $d_2 = 0.189269$; and $d_3 = 0.001308$ are constants.

The SSI values computed were classified into certain drought classification using classification standard developed by Mckee, et. al (1993) [146] and Hughes and Saunders [147] (Table: 6.1).

Table 6.1: SSI/SPI Drought Classification

SPI/SSI values	Drought Classifications
≥ 2.0	extremely wet
1.5 to 1.99	very wet
1.0 to 1.49	moderately wet
-0.99 to 0.99	near normal
-1.0 to -1.49	Moderate drought
-1.5 to -1.99	Severe drought
≤ -2.0	Extreme drought

Another technique applied in this study to examine extreme flow conditions involves observing the return level versus return period plot. This was executed by fitting Generalized Extreme Value distributions (GEV) (Equation: 6.6) to the maximum annual series using the Statistics and Machine Learning Toolbox in MATLAB [148]. Subsequently, return levels for a given return period (probability) were estimated from the fitted probability distributions. Finally, the return periods were plotted against return levels, providing insights into how flow extremes obtained from different scenarios behave for a given return period.

$$p_{GEV}(x) = \frac{1}{\sigma}t(x)^{\xi+1}e^{-t(x)} \tag{6.6}$$

Where:

$$t(x) = \begin{cases} \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}, & for \quad \xi \neq 0 \\ e^{-}\left(\frac{x-\mu}{\sigma}\right), & for \quad \xi = 0 \end{cases} \tag{6.7}$$

Assuming that $\mu = 0$ and $\sigma = 1$, then when, $\xi = 0$ then P_{GEV} is Gumbel distribution (Type I), $\xi > 0$ then P_{GEV} is Frechet distribution (Type II), and $\xi < 0$ then P_{GEV} is Weibull distribution (Type III). The cumulative distribution can expressed as follows,

$$F(x) = e^{t(x)} \tag{6.8}$$

At same time, the inverse of P_{GEV} distribution is given by,

$$F^{-1}(p) = \begin{cases} \mu + \frac{\sigma}{\xi}\big[(lnp)^{-\xi} - 1\big], & for \quad \xi \neq 0 \\ \mu - \sigma ln(-lnp), & for \quad \xi = 0 \end{cases} \tag{6.9}$$

6.4 Results

6.4.1 SWAT Model Run, Calibration and Validation

The SWAT model was executed with the prepared inputs and subsequently calibrated using observed data from the Hombole station, employing the SUFI-2 algorithm in SWAT-CUP [141]. The total period of analysis in the modeling procedure spans 30 years (01/01/1980 - 12/01/2009). The initial five years of data were designated for model warm-up, with the remaining data used for calibration and validation. A significant portion of the total dataset, covering the period from 01/01/1985 to 01/12/2000, was utilized for model calibration, while the remainder was allocated for validation.

During the model sensitivity analysis under SWAT-CUP, a total of 14 parameters were identified as the most sensitive for streamflow prediction. The top three most sensitive parameters were identified as CN2, Sol_K, and Sol_BD, as illustrated in Figure 6.2. Subsequently, their possible ranges and best fits were determined, as shown in Table 6.2.

Table 6.2: SWAT model parameters ranges and fits identified during model sensitivity and calibration stage for Hombole Station

No.	Parameter Name	Fitted Value	Min value	Max value
1	CN2.mgt	-0.08	-0.335	-0.031
2	GW_DELAY.gw	50.228	-64.751	278.471
3	GW_REVAP.gw	0.247	0.096	0.288
4	RCHRG_DP.gw	0.053	0.01	0.508
5	ALPHA_BF.gw	0.596	0.068	0.69
6	REVAPMN.gw	3.943	0.564	6.856
7	GWQMN.gw	15.361	9.635	28.915
8	SOL_BD(..).sol	0.232	0.114	0.361
9	SOL_K(..).sol	-1.272	-1.348	0.084
10	SOL_AWC(..).sol	0.611	0.077	0.631
11	USLE_K(..).sol	0.014	0.001	0.014
12	OV_N.hru	-0.065	-0.15	-0.049
13	SURLAG.hru	11.978	3.527	17.179
14	ESCO.bsn	0.465	0.362	1.066

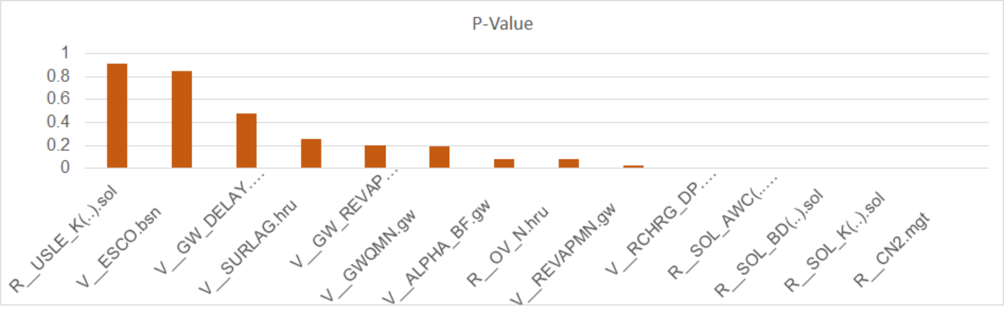


Figure 6.2: SWAT model sensitive parameters in calibration procedure at Hombole station.

The model performance measures in Table 6.3 indicate that the model parameters have been adequately calibrated, and the results are deemed very good based on commonly used performance measures such as Nash-Sutcliffe efficiency (NS) and R2. These results meet the minimum model requirements suggested in Moriasi, 2007 [149].

Table 6.3: SWAT-CUP model calibration and validation period performance summary at Hombole station

Period	p-factor	r-factor	R2	NS	bR2	MSE	SSQR	PBIAS
Calibration	0.9	1.19	0.89	0.88	0.8098	600	180	-12.3
Validation	0.87	1.31	0.83	0.82	0.6771	760	220	-19.4

Following the calibration of the model, its performance was tested with a second dataset that was not used during the calibration period. Model simulations for both the calibration and validation periods, compared against their corresponding observed data, are illustrated in Figure 6.3 and 6.4. It is evident from both plots and the corresponding Nash-Sutcliffe efficiency (NSE) performance measure that the model performed reasonably well.

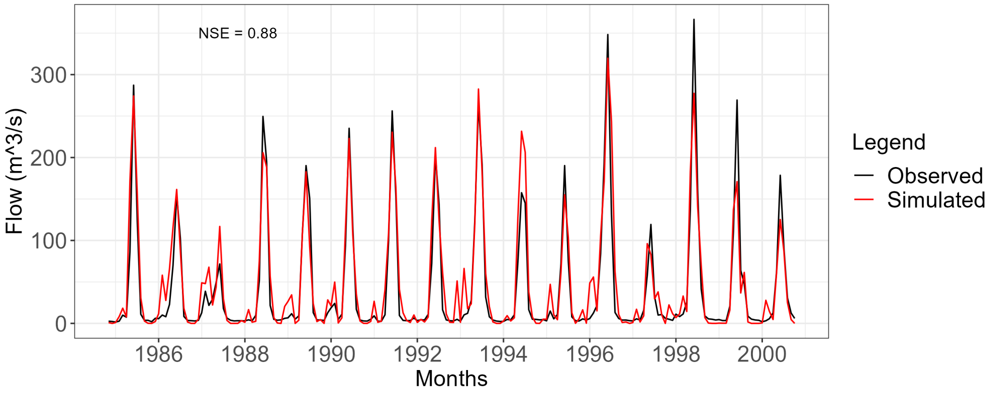


Figure 6.3: Hombole SWAT simulated stream flow for Calibration period plotted together with observed flow.

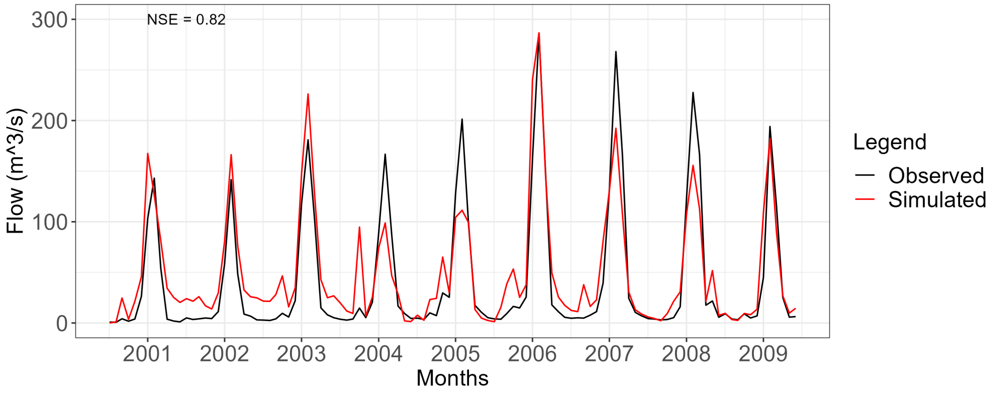


Figure 6.4: Hombole SWAT simulated stream flow for Validation period plotted together with observed flow.

6.4.2 Future Water Availability Scenarios

After successfully training/calibrating, and validating both the NARX and SWAT models for the Upper Awash basin, the next step involves projecting future water availability. This projection will be executed using an ensemble of selected climate model outputs from the CMIP6 archive. In chapter 3, a meticulous ranking of climate models was conducted to pinpoint those capable of accurately capturing the climatic characteristics of the Upper Awash Sub-basin. Among the initially identified 12 models, five exhibited relatively good performance. Consequently, this section focuses on employing an ensemble of scenario outputs from these top-performing climate models, in conjunction with the hydrological models, to observe and analyze future water availability in the sub-basin.

SWAT Simulation Results

The mean monthly flow obtained from SWAT simulations aligns with the uni-modal characteristics of precipitation in the sub-basin. Notably, significant flows are still recorded for the months of July, August, and September, with low flows in December, January, and February, mirroring the patterns observed during the control period (1980-2009), represented by the blue color in Figure 6.5. Over the 30-year control period, the minimum and maximum mean monthly flows were 4.09 m^3/s and 203.8 m^3/s, occurring in January and August, respectively. In contrast, for the near period, the predicted values are 0.42 m^3/s and 255.89 m^3/s, for the mid period, they are 0.51 m^3/s and 259.96 m^3/s, and for the far period, they are 0.68 m^3/s and 406.48 m^3/s, for minimum and maximum flow conditions, respectively. Across all three periods of climate model simulations, the minimum mean monthly flows are predicted for the month of February, while the maximum flows occur in August. Additionally, there is a noticeable tendency for a decrease in minimum mean monthly flows and, conversely, a significant increase in maximum flows, particularly for the far time period, compared to the control period.

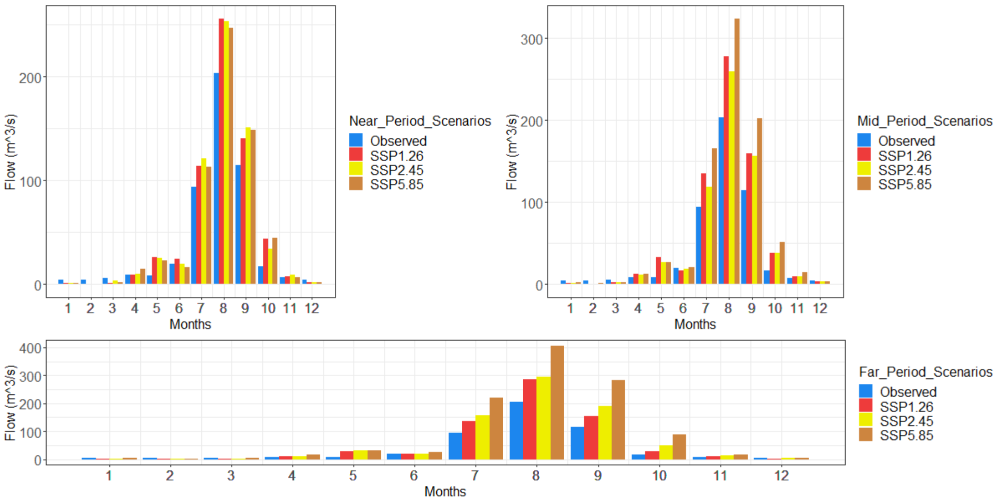


Figure 6.5: Future monthly average water availability scenarios at Hombole station obtained using an ensemble of five CMIP6 climate model outputs under SWAT model.

The decrease in flow during low-flow seasons is evident in Figure 6.6, and this trend is predicted to continue for all three time periods of analysis. The plots indicate an increase in mean monthly flow during the summer period (July-August-September) with varying degrees of proportions. Additionally, all three time periods of future simulations generally show a doubling of mean monthly flows in May and October. However, a decrease in June mean monthly flow is anticipated, especially for the near and mid-time periods, particularly in the SSP2.45 scenario.

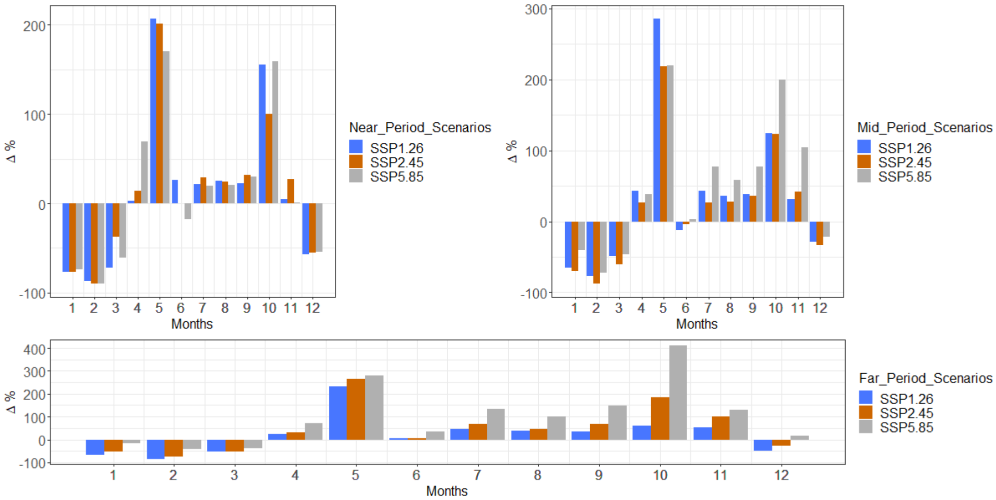


Figure 6.6: Mean monthly changes between flows obtained from control (1980-2009) and SWAT model simulated scenario periods.

The box plots depicting SWAT-generated annual mean flows for all scenarios in Figure 6.7 reveal a clear increasing trend in each scenario when compared against the observed values. This trend is evident by observing the means of the box plots.

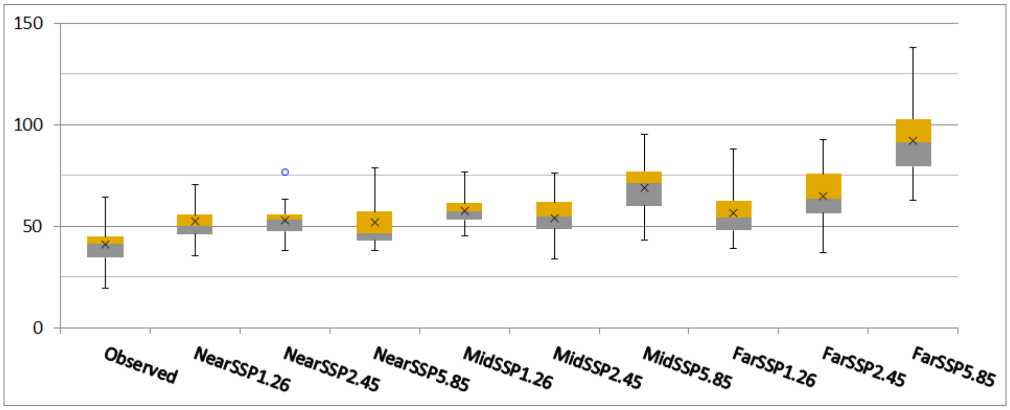


Figure 6.7: Boxplots of future annual mean flows of observed and scenarios obtained using an ensemble of five CMIP6 climate model outputs under SWAT model at Hombole station.

The change in annual mean flow from the control period for all scenarios indicates a relative increase of more than 20%, as shown in Table 6.4. Notably, a significant increase is predicted for the far time period under the SSP5.85 scenario.

Table 6.4: Percentage changes of SWAT model simulated annual mean flow of scenarios from control period (1980 - 2009) for three time periods.

	Near (m^3/s)	Mid (m^3/s)	Far (m^3/s)	Near (Δ Q in m^3/s)	Mid (Δ Q in m^3/s)	Far (Δ Q in m^3/s)
SSP1.26	52.26	57.63	56.59	27.29	40.37	37.85
SSP2.45	52.64	54.01	64.95	28.22	31.57	58.22
SSP5.85	51.71	69.09	92.02	25.96	68.29	124.14
Observed	41.05					

NARX Simulation Results

Similar procedures were followed to simulate future flows at the Hombole station using the trained NARX model from the previous chapter. It is important to note that in this chapter, the best-fitting NARX model was found to require only rainfall data to predict future flow conditions. Additionally, in this section, rainfall time series for each corresponding time period of analysis and scenarios are obtained from the previous climate modeling endeavor.

As shown in Figure 6.8, the outputs of flow prediction obtained from the NARX model also exhibit bi-modal mean monthly flows, with high flows in July, August, and September, and low flows in December, January, and February. Similar to the SWAT model simulations, the long-term flow analysis indicates that minimum mean monthly flows occur in February, while maximum flows are observed in August. The respective minimum | maximum mean monthly flow values for each time period of analysis are as follows: for the near period, 3.68 m^3/s | 228.52 m^3/s; for the mid period, 3.64 m^3/s | 248.45 m^3/s; and for the far period, 3.54 m^3/s | 253.58 m^3/s.

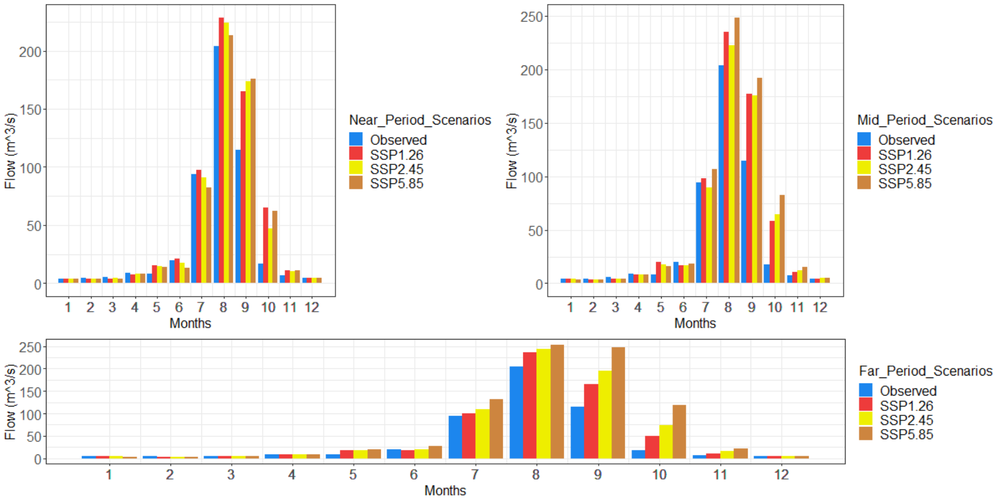


Figure 6.8: Future monthly average water availability scenarios at Hombole station obtained using an ensemble of five CMIP6 climate model outputs under NARX model.

An important finding, similar to SWAT simulations, is a clear indication of a decrease in mean monthly flows during low-flow seasons compared to the control period (Figure 6.9). Similarly, flows in May and October show an increase of more than 50%. There is also a predicted increase in mean monthly flow during the summer rainy seasons, particularly in August and September, especially at higher proportions for the SSP5.85 scenario. Similar to SWAT model simulations, there appears to be a decrease in mean monthly flow for the month of June, especially in two scenarios (SSP2.45 and SSP5.85) for the near and mid time periods. Additionally, similar to SWAT model simulations, mean monthly flows in May and October show a dramatic increase.

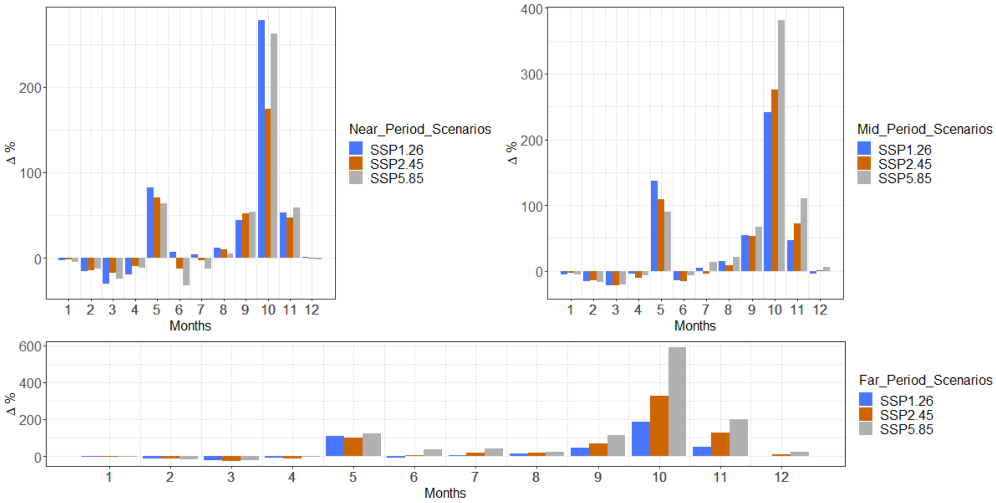


Figure 6.9: Mean monthly changes between flows obtained from control (1980-2009) and NARX model simulated scenario periods.

When observing the annual mean box plot in Figure 6.10, it can be generally concluded that there is an increasing trend in the long-term minimum, maximum, and mean of annual mean flows in each scenario compared to the observed values.

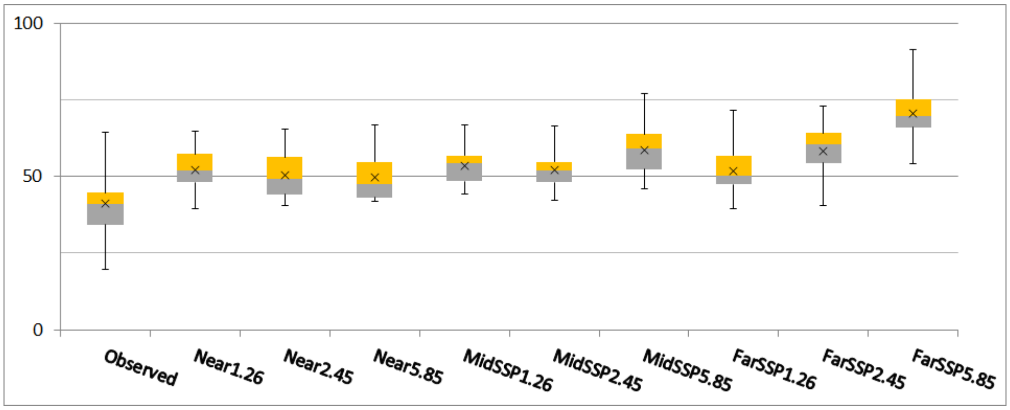


Figure 6.10: Boxplots of future annual mean flow scenarios at Hombole station obtained using an ensemble of five CMIP6 climate model outputs under NARX model.

Similar to the climate model scenario analysis performed with the SWAT model, change estimation was carried out on annual mean flow from the control period for all scenarios, as shown in Table 6.5. The results of the analysis in the table show a similar trend as before, with all scenarios indicating an increase of more than 20%, and a significant increase being projected for the far time period under the SSP5.85 scenario.

Table 6.5: Percentage changes of NARX model simulated annual mean flow of scenarios from control period (1980 - 2009) for three time periods.

	Near (m^3/s)	Mid (m^3/s)	Far (m^3/s)	Near (Δ Q in m^3/s)	Mid (Δ Q in m^3/s)	Far (Δ Q in m^3/s)
SSP1.26	52.25	53.49	51.87	27.27	30.29	26.35
SSP2.45	50.34	51.99	58.26	22.63	26.65	41.93
SSP5.85	49.75	58.73	70.38	21.18	43.07	71.44
Observed	41.05	41.05	41.053			

Additionally, flow frequency analysis was conducted by fitting Generalized Extreme Value (GEV) distributions to the annual maxima of the control period, mid, and far scenarios. The annual maxima obtained from the SWAT model suggest that higher magnitudes of floods can be expected for a given return period in scenarios compared to the control period. However, the same cannot be said about flow simulations obtained from the NARX model. The difference in the two plots arises from how the two models behave in simulating peaks during the training phase. In general, the NARX model encounters difficulty in handling extreme events from climate scenario outputs since the model has been trained on observed data.

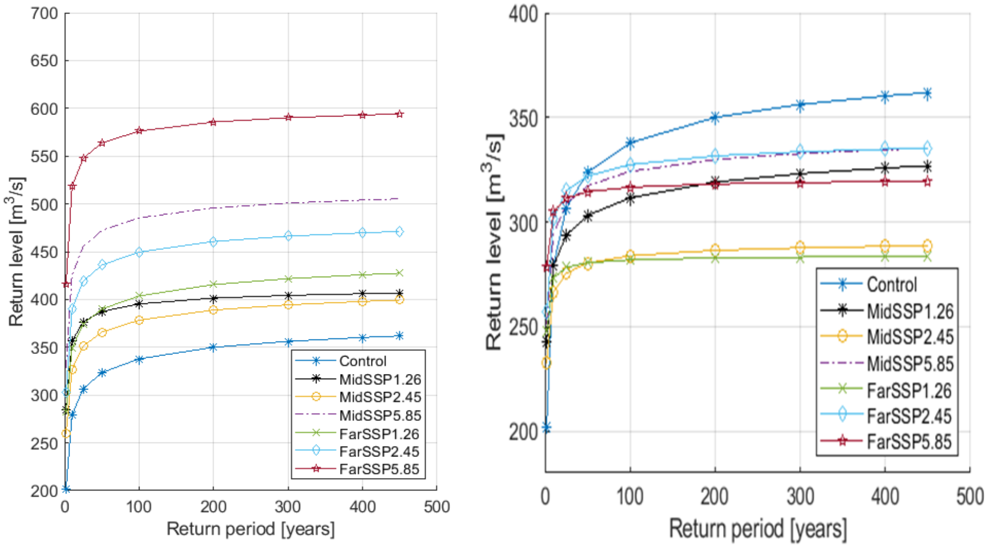


Figure 6.11: Return period versus return level plots for simulations of SWAT (left) and NARX (right) model for different climate scenarios

6.4.3 Comparison of SWAT and NARX Model Future Simulations

While both models have performed satisfactorily when evaluated against data in the control period, the results of simulations using climate model outputs are not similar in magnitude, as shown in the previous two sub-sections. However, both models were able to produce similar trends in future water resource projections, as discussed above and additionally shown in Figure 6.12.

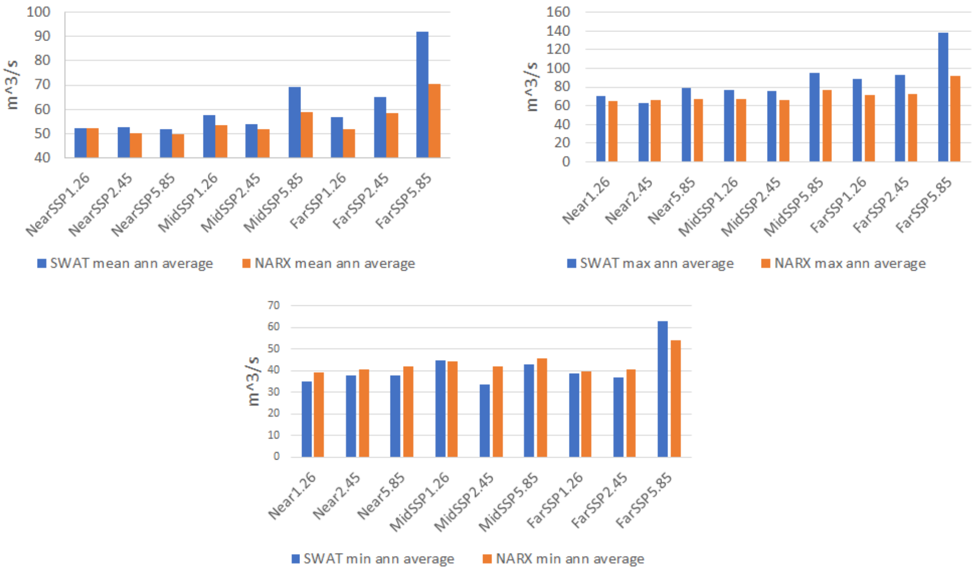


Figure 6.12: Mean (top left), maximum (top right), and minimum (bottom middle) of annual mean flow simulations obtained from SWAT and NARX models.

In general, SWAT model predictions resulted in higher predictions for the mean and maximum of annual mean flows for different scenarios. In contrast, low flow predictions were lower than NARX model simulations for most scenarios. This difference might be due to the fact that the calibrated SWAT model tends to predict the base flow component relatively poorly compared to the NARX model, as already shown in Figure 6.4 and 5.34.

6.4.4 Present and Future Water Resource Conditions

In this specific section, the utilization of the Standardized Streamflow Index (SSI) is contemplated for the comprehensive examination of water resource conditions within the sub-basin. The SSI, distinguished by its conformity to the identical calculation steps as the Standard Precipitation Index (SPI) and its endorsement by the World Meteorological Organization (WMO), shall be deployed for analytical purposes. The gauge station situated at Hombole is deliberately selected for scrutiny, given its proximity to the terminus of the Upper Awash Basin.

Present Conditions

The analytical process commenced with the computation of the Standardized Streamflow Index (SSI) derived from the 30-year streamflow time series data collected at the designated station. The application of SSI calculations, spanning both 3 and 12 months, was undertaken to elucidate the prevailing water resource availability within the sub-basin, with a specific focus on both agricultural and hydrological considerations (refer to Figure 6.13).

Upon examination of Figure 6.13 and the concise presentation in Table 6.6, it is discernible that, out of a total of 360 months within the 30-year span, 50 and 47 months were identified as drought months (i.e., $SSI \leq -1$) for the 3 and 12-month SSI calculations, respectively. Specific time periods, such as Oct (1985), Aug-Nov (1987), and Sep-Nov (1997), were notably characterized by the occurrence of the most severe 3-month droughts, with the maximum duration extending to 4 months in Aug-Nov (1987). Conversely, Sep (1987)-Jul (1988), Sep-Oct (1997), and Jun (1998) marked the most extreme 12-month dry periods in the basin. The lengthiest duration of an extreme dry period spanned a consecutive 11 months (Sep (1987)-Jul (1988)) for the 12-month category, as evident from the SSI plot (indicating Hydrological Drought).

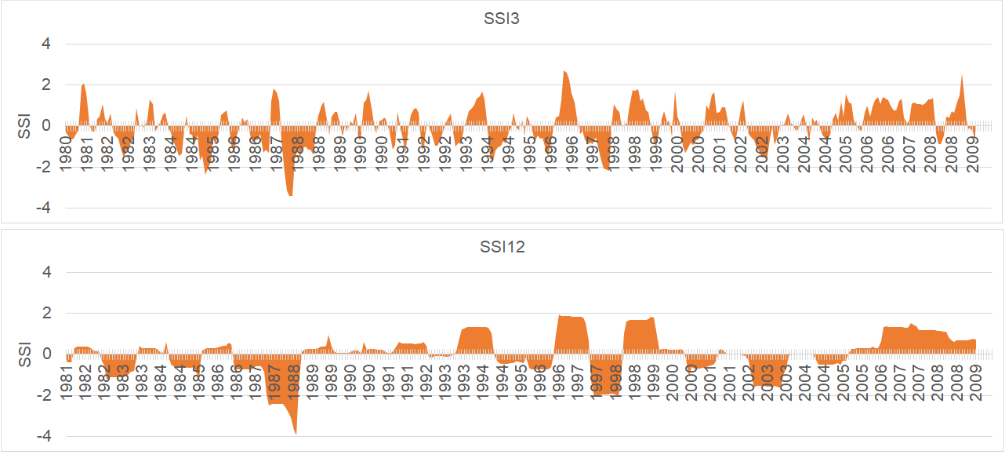


Figure 6.13: Standardized Streamflow Index (SSI) for Upper Awash Basin at Hombole guage station.

Over the course of the 30-year analytical timeframe, a cumulative total of 61 and 60 months were identified as wet months for the 3 and 12-month SSI analyses, respectively.

The most pronounced instances of wet months in the 3-month SSI analysis were observed in Apr (1981), Jun-Aug (1996), and Jan (2009). The lengthiest consecutive periods of extreme wetness spanned from Jun-Aug (1996), encompassing a mere 3 months in total. In parallel, Aug (1996) - Jun (1997), Sep (1998) - Jul (1999), and Jun (2007) emerged as the most precipitation-rich months for the 12-month SSI analysis. Conversely, for the 12-month analysis, two distinct occurrences of 11 consecutive months each were noted. The first transpired from Aug (1996) - Jun (1997), while the second extended from Sep (1998) - Jul (1999). Both plots collectively illustrate the 1980s as the epoch of most severe drought years, juxtaposed against the contrasting precipitation abundance characterizing the initial decades of the 2000s.

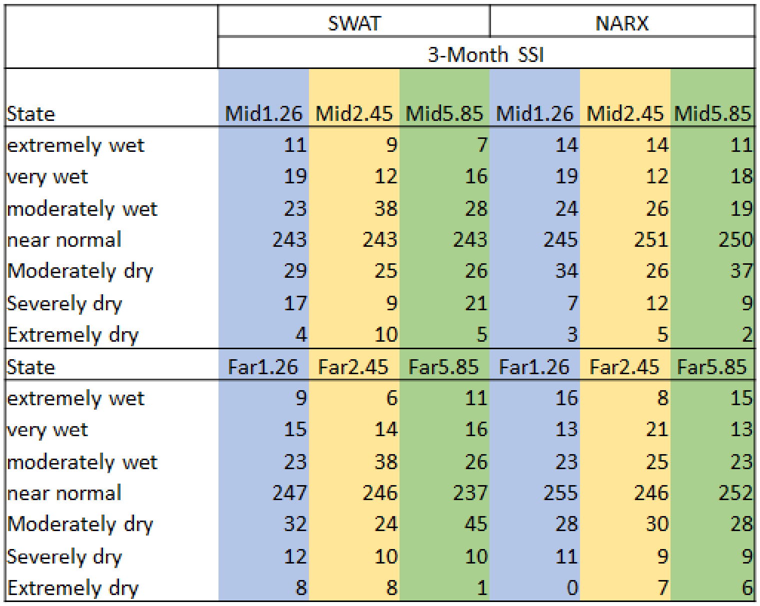
Table 6.6: Summary table of SSI for Observed in Upper Awash Basin at Hombole guage station.

State	3-Months	12-Months
Extremely wet	5	0
Very wet	14	23
Moderately wet	42	37
Near normal	235	230
Moderately dry	31	13
Severely dry	10	20
Extremely dry	9	14

Future Conditions

In the exploration of potential future water resource scenarios, we employed streamflow simulations generated through the combination of various scenarios derived from an ensemble of climate model outputs. The analyses incorporated outputs from both SWAT and NARX models, with a specific focus on mid and far time periods, as the near time period did not meet the requisite minimum of 30 years for SSI analysis.

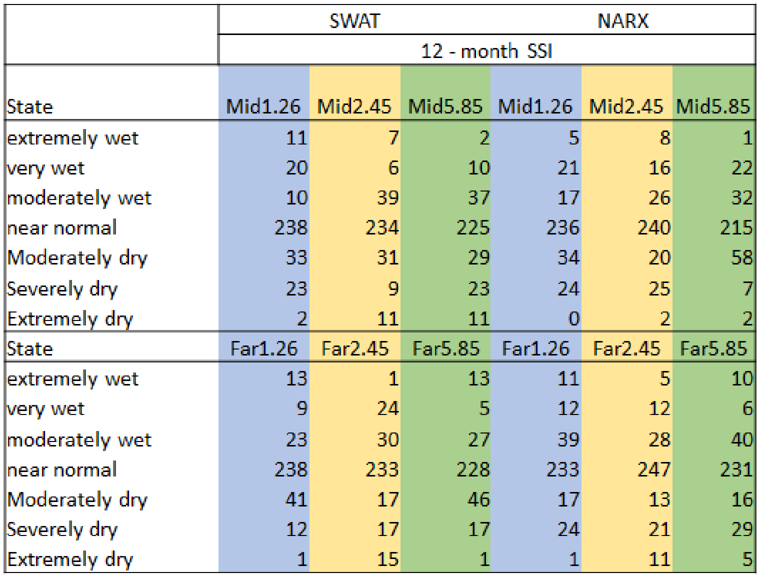
The application of the Standardized Streamflow Index (SSI) for 3 and 12-month time frames provided insights into potential future conditions, as succinctly summarized in Table 6.14 and 6.15. Notably, the summary table indicates a relatively stable outlook, with no discernible significant increase or decrease in the number of months for each prospective state when compared to the existing conditions outlined in Table 6.6. This analytical endeavor stands as a valuable tool for comprehending and foreseeing future dynamics in water resource availability within the sub-basin.



	SWAT			NARX		
	3-Month SSI					
State	Mid1.26	Mid2.45	Mid5.85	Mid1.26	Mid2.45	Mid5.85
extremely wet	11	9	7	14	14	11
very wet	19	12	16	19	12	18
moderately wet	23	38	28	24	26	19
near normal	243	243	243	245	251	250
Moderately dry	29	25	26	34	26	37
Severely dry	17	9	21	7	12	9
Extremely dry	4	10	5	3	5	2
State	Far1.26	Far2.45	Far5.85	Far1.26	Far2.45	Far5.85
extremely wet	9	6	11	16	8	15
very wet	15	14	16	13	21	13
moderately wet	23	38	26	23	25	23
near normal	247	246	237	255	246	252
Moderately dry	32	24	45	28	30	28
Severely dry	12	10	10	11	9	9
Extremely dry	8	8	1	0	7	6

Figure 6.14: Summary table of 3 month SSI of outputs obtained from SWAT and NARX models for Upper Awash Basin at Hombole guage station.

When comparing the number of months classified as extremely wet conditions for both the 3 and 12-month SSI calculations in both the mid and far periods, a relative increment is evident in contrast to the existing conditions. Conversely, concerning extremely dry conditions, there is a relative decrease across almost all scenarios. The exceptions in this case include mid2.45 for the 3-month period and far2.45 for the 12-month SSI. This observed trend remains consistent for the NARX simulations as well.



	SWAT			NARX		
	12 - month SSI					
State	Mid1.26	Mid2.45	Mid5.85	Mid1.26	Mid2.45	Mid5.85
extremely wet	11	7	2	5	8	1
very wet	20	6	10	21	16	22
moderately wet	10	39	37	17	26	32
near normal	238	234	225	236	240	215
Moderately dry	33	31	29	34	20	58
Severely dry	23	9	23	24	25	7
Extremely dry	2	11	11	0	2	2
State	Far1.26	Far2.45	Far5.85	Far1.26	Far2.45	Far5.85
extremely wet	13	1	13	11	5	10
very wet	9	24	5	12	12	6
moderately wet	23	30	27	39	28	40
near normal	238	233	228	233	247	231
Moderately dry	41	17	46	17	13	16
Severely dry	12	17	17	24	21	29
Extremely dry	1	15	1	1	11	5

Figure 6.15: Summary table of 12 month SSI of outputs obtained from SWAT and NARX models for Upper Awash Basin at Hombole guage station.

In the 3-month SSI analysis, the longest successive extreme wet months are projected to occur for a duration of 4 months, expected in Sep-Dec (2078) for SWAT and Jun-Sep (2068) for NARX simulations simultaneously. Conversely, extremely dry conditions are anticipated to last for 4 months (Feb-May 2073) for SWAT and 3 months (Sep-Nov 2075) for NARX.

For the 12-month SSI analysis, the most prolonged extreme wet conditions are predicted to last for a total of 13 months, spanning from Jul 2096 to Jul 2097 for SWAT, and 9 months (Oct-Jun 2079) for NARX. Additionally, extreme dry months are projected to persist for a total of 8 months (Oct 2075 - May 2076) for both SWAT and NARX simulations.

Summarizing the severity of events (SSI values ≤ -1) in each time period for various scenarios in Table 6.7, it is evident that there is a decrease in severity as the analysis period progresses from mid to far time periods. Moreover, for the annual SSI analysis, the worst-case scenario (SSP5.85) exhibits higher severity than the other two scenarios.

Table 6.7: Severity analysis based on SSI analysis done on streamflow simualtions of SWAT and NARX model

			3-month				12-month			
			Obs	SSP1.26	SSP2.45	SSP5.85	Obs	SSP1.26	SSP2.45	SSP5.85
SWAT	Mid	Duration	50	50	44	52	47	58	51	63
		Severity	-74.82	-77.64	-68.78	-77.82	-84.31	-82.56	-82.28	-97.84
	Far	Duration		52	42	56		54	49	64
		Severity		-76.53	-67.32	-74.13		-71.77	-81.88	-86.19
NARX	Mid	Duration		44	43	48		58	47	67
		Severity		-59.56	-66.1	-65.01		-80	-71.86	-89.76
	Far	Duration		39	46	43		42	45	50
		Severity		-51.41	-66.53	-61.05		-63.53	-80.66	-81.51

6.5 Discussion

In this chapter, an assessment of the potential impact of prospective climate changes on the water resource potential of the Upper Awash Sub-Basin (UASB) was conducted through the application of two hydrological models in conjunction with various climate scenarios. The initial model employed was a semi-distributed, physically-based Soil and Water Assessment Tool (SWAT) model. This model underwent calibration (Nash-Sutcliffe Efficiency (NSE) = 0.88) and validation (NSE = 0.82) at the Hombole gauge station, utilizing monthly data. The second model employed was a non-linear data-driven model known as Non-linear Auto-Regressive with eXogenous inputs (NARX). This model underwent training (NSE = 0.93) and validation (NSE = 0.92) utilizing the same dataset used for SWAT model calibration, as detailed in Chapter 5. Climate model scenarios were derived from an ensemble of five climate model outputs in the CMIP6 archive (ECEARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2, and INM-CM5-0).

The results of simulations from both the SWAT and NARX models exhibit a prevailing trend of increased future water availability across all combinations of time periods and climate scenarios when compared to the control period. This trend is particularly evident when examining changes at the annual level. Additionally, a distinct observation from the mean monthly flow series indicates a reduction in flow magnitudes during the low flow season (Dec-Jan-Feb) and a significant increase during the high flow season (Jul-Aug-Sep). Furthermore, an augmentation in mean monthly flow for both May and October has been projected, contributing significantly to higher flows in the MAM (March-April-May) and OND (October-November-December) seasonal periods. This finding aligns with a recent study conducted in the East African region by Gebrecherkos et al. (2023) [12]. It is noteworthy that previous studies by Taye et al. (2018) [58] and Chelkeba et al. (2023) [139] presented similar projections for recent and mid-time periods, but projections for the far time period from both studies indicated a decrease.

A noticeable distinction is apparent in the magnitude of flow simulations derived from the SWAT and NARX models, particularly noticeable for mid and far time periods. Nevertheless, both models exhibit a clear consensus regarding the general trend of future water resource conditions in the sub-basin. The calibration and validation results of the SWAT model underscore its relative inefficiency in simulating low flows, occasionally tending to overestimate them when compared to the NARX model. Conversely, the NARX model displays poor performance in capturing peaks, potentially attributed to its sensitivity to outliers. The combination of these characteristics has led to future climate projections obtained from the SWAT model indicating relatively higher magnitudes of flow compared to the NARX model. Both SWAT and NARX flow simulations affirm that the minimum mean monthly flow occurs in February and the maximum in August, consistent with the control period. However, minimum mean monthly flows derived from the NARX model surpass those from the SWAT model significantly, suggesting the former's enhanced capability in simulating low flows.

For a more detailed understanding of future water resource conditions in the sub-basin, Standardized Streamflow Index (SSI) was generated using data obtained from both the control and scenario periods of SWAT and NARX model simulations. The analysis of 12-month SSIs reveals that the longest successive extreme events occurred at the same time periods between observed and simulated data. For instance, the longest extreme dry event during the control period spanned from September to July, lasting for a successive 11 months, while for SWAT and NARX simulations, it endured for 8 months (October-May). Similar findings were observed for the analysis of the longest successive wet events. In the control period, the longest successive wet months persisted for a total of 11 months (two events, one lasting from Aug - Jun, and the second from Sep-Jul). Conversely, for the simulations, it extended for a length of 13 months (July-July). The 12-month SSI analysis suggests that the duration of dry events will shorten, while wet events will become longer. However, in the case of the 3-month SSI analysis, temporal congruence between observed and simulated outputs was only observed for results obtained from the NARX model. Here, the longest extreme dry events occurred from September to November (3 months), with June to September (4 months) being the wettest. This implies that no significant change in the duration of dry and wet events is expected for the 3-month SSI.

In general, the SSI analysis conducted on streamflow outputs from climate models suggests that extreme events will be more pronounced for the hydrological period (12 months) than the agricultural period (3 months). Consequently, it is anticipated that water resource conditions will become less severe in the long run, as indicated by predictions for far time periods.

6.6 Conclusion

The outputs of streamflow projections obtained using CMIP6 climate model outputs indicate an increase in future streamflow across all time periods and scenarios. Previous studies employing climate model outputs from the CMIP5 archive yielded similar results for recent and mid-time periods but demonstrated a disagreement for the far time period.

Regarding seasonal changes, both spring and summer stream flows have exhibited an increase, with the latter demonstrating a relatively higher magnitude of increment. Additionally, the results of this study indicate a decrease in the flow during drier periods (Dec-Jan-Feb).

Both the SWAT and NARX models demonstrate agreement regarding the direction of future water resource conditions in the sub-basin. However, the NARX model exhibits less capability in simulating peak flows compared to the SWAT model, while it performs well in representing low flows.

The results of the 12-month Standardized Streamflow Index (SSI) analysis indicate that the duration of dry events will shorten, and wet periods will become longer.

Chapter 7

Summary, Conclusion, and Recommendations

7.1 Summary and Conclusion

As the primary contributor to streamflow in the greater Awash basin and amidst the escalating competition for this resource driven by population growth and diverse economic activities, it is imperative to monitor the temporal status of this vital resource. Understanding the future water resource conditions of a given region is facilitated through the combined use of CMIP-GCM outputs with hydrological models. Despite similar studies conducted in the sub-basin, there exists no consensus on the direction of future hydro-climatic conditions in the area. Additionally, there are no studies that have tested future climate conditions using new sets of climate model outputs, specifically from CMIP6. In this study, a meticulous and appropriate modeling effort has been employed to comprehensively examine both future climate and hydrological conditions over the Upper Awash Sub-Basin (UASB). This chapter presents some of the key conclusions drawn from the thesis.

7.1.1 Climate model selection and future projections

Globally, various institutions develop Climate Model Intercomparison Project (CMIP) - Global Climate Models (GCMs), and the outputs from these models for a given region may differ due to several reasons. The primary differences in their outputs arise from variations in spatial resolution, model complexity, parametrization, initial conditions and forcing data, and research focus. Consequently, there is a need to evaluate these models for their capacity to simulate regional meteorological conditions.

In this study, a total of 12 climate models were preliminarily selected from the CMIP6 archive and evaluated for their potential to simulate the regional climate of the Upper Awash Sub-Basin (UASB). This evaluation was conducted after downscaling and bias correction of all climate models using distance-weighted average remapping and Quantile Mapping techniques, respectively. The climate models were assessed against their ability to replicate climate data obtained from ground observations. Five techniques, evaluating the models on their capability to capture the distributional, trend, magnitude, seasonal, and temporal characteristics, were applied. Additionally, the models' performance in simulating the spatial distribution of meteorological characteristics in the study area was assessed. Out of the twelve initially identified models, the top five, demonstrating the best performance, were selected for further analysis. These models are ECEARTH3, GFDL-ESM4, MPI-ESM1-2-HR, MRI-ESM2, and INM-CM5-0.

Future climate projections were then made using ensembles of these five climate model outputs over the UASB for three socio-economic scenarios (sustainable, middle, and worst) across three time periods (near, middle, and end of the century). The projections covered areal weighted precipitation, maximum temperature, and minimum temperature. The results of future climate predictions demonstrated relative increments compared to the base period (1980–2009) for all three climate variables. The magnitude of increment was notably higher toward the end of the century compared to the near and middle periods. Additionally, the analysis of seasonal projections indicated that the magnitude of increment for spring rainfall exceeded that for the summer. It is noteworthy that the outputs of this study align with the new AR6 IPCC report on expected future climate changes in Northeast Africa.

7.1.2 Stream flow characterization and time series modeling

To assess water resource conditions under future climate change projections, the use of appropriate hydrological models is crucial. In this study, both deterministic and stochastic models were considered. Among the deterministic models, the Soil and Water Assessment Tool (SWAT) was chosen due to its solid theoretical foundation and widespread acceptance in Ethiopia and globally. Selecting a stochastic model was more complex than deterministic models, as it required a careful examination of the statistical characteristics of the time series. These characteristics provide insights into the runoff generation mechanism and guide the selection of an appropriate model. Stationarity and linearity, crucial features of a time series, influence the choice of an adequate time series model.

The monthly time series from the three gauging stations (Akaki, Melka Kuntere, and Hombole) in the Upper Awash Sub-Basin (UASB) were identified as having both stationary and non-linear characteristics, posing challenges for the application of conventional time series models like ARIMA. Consequently, two classes of modeling techniques with the potential to capture the non-linear character in the flow data were suggested: Regime Switching and Data-Driven models. Within the regime-switching models, the suitability of SETAR and LSTAR models was evaluated. For data-driven models, the potential of NARX and ANFIS was assessed. For each type of model, appropriate parameters were fitted using required procedures, and model uncertainties were estimated using bootstrapping techniques.

The forecasts for each gauging station indicated that data-driven techniques exhibited higher forecast capabilities than TAR models. Model setup and training were relatively straightforward using various functions under the Deep Learning Toolbox in MATLAB. However, a significant challenge in these models was identifying the right set of model inputs that would deliver acceptable performance. Future efforts to enhance forecast performance should consider testing models that separately model the deterministic and stochastic components. From the analysis using data-driven techniques, it was evident that the Adaptive Neuro-Fuzzy Inference System (ANFIS) model demonstrated relatively better performance than the NARX model.

7.1.3 Future Water resource availability

The assessment of future water resource availability in the sub-basin involved combining climate model outputs with a well-calibrated and validated hydrological model. In this phase of the study, the Soil and Water Assessment Tool (SWAT) model was prepared, calibrated, and validated at the Hombole gauging station. The calibration period yielded an Nash-Sutcliffe Efficiency (NSE) of 0.9, and for validation, it was 0.87. The top three sensitive parameters for the SWAT model were identified as curve number, soil hydraulic conductivity, and bulk density. Additionally, the Nonlinear Auto-Regressive with eXogenous inputs (NARX) model, trained for the same gauge station, was employed. It was observed that the NARX model exhibited good potential in simulating low flows, while SWAT performed relatively well in capturing peak flows.

Following the calibration, future streamflow projections for each of the future scenarios (SSP1.26, SSP2.45, and SSP5.85) and three time periods (near, mid, and far) were generated using the SWAT model. Mean monthly streamflow simulations for each scenario were compared against a control period (1980-2009). Both the SWAT and NARX models showed an increase in mean monthly flows during both spring and summer, with varying degrees of proportion for all scenarios and time periods. However, the dry period (Dec-Jan-Feb) exhibited a relative decrease.

While both the NARX and SWAT models yielded similar outputs in terms of future projections on water resource availability, the magnitude of these projections differed due to the models' distinct assumptions. SWAT projections were notably higher in magnitude than those of the NARX model. However, it was observed that the magnitude of future projections is lowest for the near period (2022-2039) and highest for the far time period (2070-2099).

The Standardized Streamflow Index (SSI) analysis performed on the future projections using the outputs of both NARX and SWAT models indicated that wet events would be more likely in the future than dry events, especially in the hydrological period of analysis.

7.2 Recommendations

The regional climate in Ethiopia, influenced by rugged topography and proximity to the equator, has been the subject of various hypotheses. Previous attempts to observe future climate conditions in the Upper Awash Sub-Basin (UASB) using climate models from different institutions have resulted in divergent projections, with no common understanding of future climate trends. This study utilizes models from the new CMIP6 archive to investigate future climate conditions in the UASB. Notably, the CMIP6 models, when compared to previous CMIP3 and CMIP5 efforts, exhibit different projections, particularly for the far time period. The reasons for these differences remain unclear and warrant further investigation.

The findings of this study indicate an anticipated increase in future rainfall and streamflow in the area. This projected change can have significant socio-economic implications for both the sub-basin and the larger Awash basin, given the expected population growth and economic activities. The heightened potential for flooding in both the upper and lower parts of the Awash basin, attributed to increased farming and urban development in the UASB, emphasizes the need for sustainable land and water resource management practices. Addressing this issue requires careful attention.

A major limitation encountered in this study is the quality of data obtained from river gauging stations. Many stations in the Awash basin have fragmented data without accompanying metadata, highlighting the importance of improving data quality for these stations.

Two types of models, belonging to the deterministic and stochastic classes, were employed in this study. Future efforts to enhance forecast performance should consider testing models that separately address the deterministic and stochastic components, providing a more comprehensive understanding of the forecasting capability.

Bibliography

[1] *Awash Basin Sedimentation Modeling Project.* AWBA,MoWE, 2017.

[2] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. Artificial neural networks in hydrology. i: Preliminary concepts. *Journal of Hydrologic Engineering*, 5(2):115–123, 2000.

[3] Mohammad Zakir Hossain. Water: The most precious resource of our life. *Global Journal of Advanced Research*, 2(9):1–11, 2015.

[4] Gederts Ievinsh. Water content of plant tissues: So simple that almost forgotten? *Plants*, 12(6):1238, 2023.

[5] Where is earth's water? `https://www.usgs.gov/special-topics/water-science-school/science/where-earths-water`. Accessed: 2023-10-09.

[6] Binaya Kumar Mishra, Pankaj Kumar, Chitresh Saraswat, Shamik Chakraborty, and Arjun Gautam. Water security in a changing environment: Concept, challenges and solutions. *Water*, 13(4):490, 2021.

[7] Sergio M Vicente-Serrano, Fernando Domínguez-Castro, Conor Murphy, Jamie Hannaford, Fergus Reig, Dhais Peña-Angulo, Yves Tramblay, Ricardo M Trigo, Neil Mac Donald, M Yolanda Luna, et al. Long-term variability and trends in meteorological droughts in western europe (1851–2018). *International journal of climatology*, 41:E690–E717, 2021.

[8] Qiaoqian Han, Shuanghu Zhang, Guoxian Huang, and Rui Zhang. Analysis of long-term water level variation in dongting lake, china. *Water*, 8(7):306, 2016.

[9] Alka Abraham and Subrahmanya Kundapura. Evaluating the long-term trends of the climatic variables over three humid tropical basins in kerala, india. *Arabian Journal of Geosciences*, 15(9):811, 2022.

[10] Edoardo Daly, Salvatore Calabrese, Jun Yin, and Amilcare Porporato. Hydrological spaces of long-term catchment water balance. *Water Resources Research*, 55(12):10747–10764, 2019.

[11] Kunyang Wang, Shinichi Onodera, Yuta Shimizu, and Mitsuyo Saito. Evaluation of long-term variations in water balance controlled by the density of land use changes in an urban-forest mixed basin of western japan using swat. *Authorea Preprints*, 2020.

[12] SH Gebrechorkos, Meron Teferi Taye, B Birhanu, D Solomon, and T Demissie. Future changes in climate and hydroclimate extremes in east africa. *Earth's Future*, 11(2):e2022EF003011, 2023.

[13] Tracy J Baker and Scott N Miller. Using the soil and water assessment tool (swat) to assess land use impact on water resources in an east african watershed. *Journal of hydrology*, 486:100–111, 2013.

[14] MR Knebl, Z-L Yang, Keith Hutchison, and DR Maidment. Regional scale flood modeling using nexrad rainfall, gis, and hec-hms/ras: a case study for the san antonio river basin summer 2002 storm event. *Journal of Environmental Management*, 75(4):325–336, 2005.

[15] Thair S Khayyun, Imzahim A Alwan, and Ali M Hayder. Selection of suitable precipitation cmip-5 sets of gcms for iraq using a symmetrical uncertainty filter. In *IOP Conference Series: Materials Science and Engineering*, volume 671, page 012013. IOP Publishing, 2020.

[16] Katrina E Bennett, Arelia T Werner, and Markus Schnorbus. Uncertainties in hydrologic and climate change impact analyses in headwater basins of british columbia. *Journal of Climate*, 25(17):5711–5730, 2012.

[17] Seleshi Bekele Awulachew, Aster Denekew Yilma, Makonnen Loulseged, Willibald Loiskandl, Mekonnen Ayana, and Tena Alamirew. *Water resources and irrigation development in Ethiopia*, volume 123. Iwmi, 2007.

[18] Strategies for water-wise development in ethiopia. `https://www.wri.org/insights/strategies-water-risk-insecurity-ethiopia`. Accessed: 2023-10-10.

[19] Ethiopia flood response plan - 2020 kiremt season floods. `https://www.unocha.org/publications/report/ethiopia/ethiopia-flood-response-plan-2020-kiremt-season-floods-september-2020`. Accessed: 2023-10-10.

[20] Addisalem Bitew Mitiku, Berhanu Woldyohanes, Sharew Anteneh Mekonnen, Kirubeal Mekonnen, and Abel Tadesse. The need for integrated flood management approach from social, economic and environmental perspectives: The case of upper awash river basin. *International Journal of Water Resources and Environmental Engineering*, 12(3):57–70, 2020.

[21] Marta Gebreyesus, Aberham Chernet, Moges Molla, Tefera Ashine, and Goitom Kelem. Drought characterization using reconnaissance drought index (rdi): In the case of awash river basin, ethiopia. *Int J Environ Sci Nat Res*, 26(3):70–77, 2020.

[22] R Jahn, HP Blume, VB Asio, O Spaargaren, and P Schad. *Guidelines for soil description.* FAO, 2006.

[23] Asmaa Alhamshry, Ayele Almaw Fenta, Hiroshi Yasuda, Reiji Kimura, and Katsuyuki Shimizu. Seasonal rainfall variability in ethiopia and its long-term link to global sea surface temperatures. *Water*, 12(1):55, 2020.

[24] Meron Teferi Taye, Ellen Dyer, Katrina J Charles, and Linda C Hirons. Potential predictability of the ethiopian summer rains: understanding local variations and their implications for water management decisions. *Science of the Total Environment*, 755:142604, 2021.

[25] Alemayehu A Shawul and Sumedha Chakma. Spatiotemporal detection of land use/land cover change in the large basin using integrated approaches of remote sensing and gis in the upper awash basin, ethiopia. *Environmental Earth Sciences*, 78(5):1–13, 2019.

[26] Andarge Yitbarek, Moumtaz Razack, Tenalem Ayenew, Engida Zemedagegnehu, and Tilahun Azagegn. Hydrogeological and hydrochemical framework of upper awash river basin, ethiopia: with special emphasis on inter-basins groundwater transfer between blue nile and awash rivers. *Journal of African Earth Sciences*, 65:46–60, 2012.

[27] *Population projection of Ethiopia for All Regions at Wereda Level from 2014-2017.* CSA, 2013.

[28] CARE Water. Leaving no one behind. *The United Nations World Water Development Report*, 2019.

[29] Peter H Gleick. *Water in crisis*, volume 100. New York: Oxford University Press, 1993.

[30] LN Thakural, Sanjay Kumar, Sanjay K Jain, and Tanvear Ahmad. The impact of climate change on rainfall variability: a study in central himalayas. In *Climate Change Impacts*, pages 181–192. Springer, 2018.

[31] Fathy Shaaban, Abdullah Othman, Turki M Habeebullah, and Waleed A El-Saoud. An integrated gpr and geoinformatics approach for assessing potential risks of flash floods on high-voltage towers, makkah, saudi arabia. *Environmental Earth Sciences*, 80(5):1–15, 2021.

[32] Shield Claire and Gutzmer Denise. Droughtscape. pages 3–6, 2019.

[33] NMA. Climate change national adaptation programme of action (napa) of ethiopia. *National Meteorological Services Agency (NMA), Ministry of Water Resources, Federal Democratic Republic of Ethiopia, Addis Ababa*, 2007.

[34] Kebede Nanesa. Awash river's the ongoing irrigation practices, future projects and its impacts on the environment of awash river basin. *Irrigation Drainage Systems Engineering*, 10(6), 2021.

[35] Zeyede Aregahegn and Mulate Zerihun. Study on irrigation water quality in the rift valley areas of awash river basin, ethiopia. *Applied and Environmental Soil Science*, 2021, 2021.

[36] Adey Nigatu Mersha, Charlotte de Fraiture, Ilyas Masih, and Tena Alamirew. Dilemmas of integrated water resources management implementation in the awash river basin, ethiopia: irrigation development versus environmental flows. *Water and Environment Journal*, 35(1):402–416, 2021.

[37] Desalegn Chemeda Edossa, Mukand Singh Babel, and Ashim Das Gupta. Drought analysis in the awash river basin, ethiopia. *Water resources management*, 24(7):1441–1460, 2010.

[38] Prabin Rokaya Shenduli, Schalk Jan Van Andel, Surafel Mamo, and Ilyas Masih. Improving hydrological prediction with global datasets: experiences with brahmaputra, upper awash and kaap catchments. *E-proceedings of the 37 th IAHR World Congress*, 2017.

[39] Paul N Edwards. History of climate modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 2(1):128–139, 2011.

[40] Nishi Bhuvandas, Prafulkumar V Timbadiya, Prem L Patel, and Prakash D Porey. Review of downscaling methods in climate change and their role in hydrological studies. *Int. J. Environ. Ecol. Geol. Mar. Eng*, 8:713–718, 2014.

[41] Enrique Soriano, Luis Mediero, and Carlos Garijo. Selection of bias correction methods to assess the impact of climate change on flood frequency curves. *Water*, 11(11):2266, 2019.

[42] Arthur F Lutz, Herbert W ter Maat, Hester Biemans, Arun B Shrestha, Philippus Wester, and Walter W Immerzeel. Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *International Journal of Climatology*, 36(12):3988–4005, 2016.

[43] Carbon Brief. How do climate models work?, 2018.

[44] Brian Ayugi, Jiang Zhihong, Huanhuan Zhu, Hamida Ngoma, Hassen Babaousmail, Karim Rizwan, and Victor Dike. Comparison of cmip6 and cmip5 models in simulating mean and extreme precipitation over east africa. *International Journal of Climatology*, 41(15):6474–6496, 2021.

[45] RL Wilby. Evaluating climate model outputs for hydrological applications. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, 55(7):1090–1093, 2010.

[46] Rajab Homsi, Mohammed Sanusi Shiru, Shamsuddin Shahid, Tarmizi Ismail, Sobri Bin Harun, Nadhir Al-Ansari, Kwok-Wing Chau, and Zaher Mundher Yaseen. Precipitation projection using a cmip5 gcm ensemble model: a regional investigation of syria. *Engineering Applications of Computational Fluid Mechanics*, 14(1):90–106, 2020.

[47] S Zahra Samadi, Gummeneni Sagareswar, and Maryam Tajiki. Comparison of general circulation models: methodology for selecting the best gcm in kermanshah synoptic station, iran. *International Journal of Global Warming*, 2(4):347–365, 2010.

[48] AJ Pitman, SE Perkins, et al. Reducing uncertainty in selecting climate models for hydrological impact assessments. *IAHS PUBLICATION*, 313:3, 2007.

[49] Tamiru Haile Alemseged and Rientjes Tom. Evaluation of regional climate model simulations of rainfall over the upper blue nile basin. *Atmospheric research*, 161:57–64, 2015.

[50] Victor Ongoma, Haishan Chen, and Chujie Gao. Evaluation of cmip5 twentieth century rainfall simulation over the equatorial east africa. *Theoretical and Applied Climatology*, 135(3):893–910, 2019.

[51] Mohamed Salem Nashwan and Shamsuddin Shahid. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research*, 230:104632, 2019.

[52] Ali Ahmadalipour, Arun Rana, Hamid Moradkhani, and Ashish Sharma. Multi-criteria evaluation of cmip5 gcms for climate change impact analysis. *Theoretical and applied climatology*, 128(1-2):71–87, 2017.

[53] Arun Rana, Shilpy Madan, and Lars Bengtsson. Performance evaluation of regional climate models (rcms) in determining precipitation characteristics for gothenburg, sweden. *Hydrology Research*, 45(4-5):703–714, 2014.

[54] Renate AI Wilcke and Lars Bärring. Selecting regional climate scenarios for impact modelling studies. *Environmental Modelling & Software*, 78:191–201, 2016.

[55] Jerome H Friedman. Data mining and statistics: What's the connection? *Computing science and statistics*, 29(1):3–9, 1998.

[56] K Srinivasa Raju and D Nagesh Kumar. Ranking general circulation models for india using topsis. *Journal of Water and Climate Change*, 6(2):288–299, 2015.

[57] Kinfe Hailemariam. Impact of climate change on the water resources of awash river basin, ethiopia. *Climate Research*, 12(2-3):91–96, 1999.

[58] Meron Teferi Taye, Ellen Dyer, Feyera A Hirpa, and Katrina Charles. Climate change impact on water resources in the awash basin, ethiopia. *Water*, 10(11):1560, 2018.

[59] Mahtsente T Tadese, Lalit Kumar, and Richard Koech. Climate change projections in the awash river basin of ethiopia using global and regional climate models. *International Journal of Climatology*, 40(8):3649–3666, 2020.

[60] Mekonnen H Daba and Songcai You. Assessment of climate change impacts on river flow regimes in the upstream of awash basin, ethiopia: based on ipcc fifth assessment report (ar5) climate change scenarios. *Hydrology*, 7(4):98, 2020.

[61] Gianbattista Bussi, Paul G Whitehead, Li Jin, Meron T Taye, Ellen Dyer, Feyera A Hirpa, Yosef Abebe Yimer, and Katrina J Charles. Impacts of climate change and population growth on river nutrient loads in a data scarce region: The upper awash river (ethiopia). *Sustainability*, 13(3):1254, 2021.

[62] Nega Chalie Emiru, John Walker Recha, Julian R Thompson, Abrham Belay, Ermias Aynekulu, Alen Manyevere, Teferi D Demissie, Philip M Osano, Jabir Hussein, Mikias Biazen Molla, et al. Impact of climate change on the hydrology of the upper awash river basin, ethiopia. *Hydrology*, 9(1):3, 2022.

[63] Selamawit Haftu Gebresellase, Zhiyong Wu, Huating Xu, and Wada Idris Muhammad. Evaluation and selection of cmip6 climate models in upper awash basin (uba), ethiopia. *Theoretical and Applied Climatology*, pages 1–27, 2022.

[64] Uwe Schulzweida. Cdo user guide, October 2021.

[65] R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, 54:507–554, 2005.

[66] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[67] Henry B. Mann. Nonparametric tests against trend. *Econometrica*, 13(3):245–259, 1945.

[68] Sheng Yue and ChunYuan Wang. The mann-kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water resources management*, 18(3):201–218, 2004.

[69] Pranab Kumar Sen. Estimates of the regression coefficient based on kendall's tau. *Journal of the American statistical association*, 63(324):1379–1389, 1968.

[70] David Casado De Lucas. *Classification techniques for time series and functional data*. PhD thesis, Universidad Carlos III de Madrid, 2010.

[71] Pablo Montero and José A. Vilar. TSclust: An R package for time series clustering. *Journal of Statistical Software*, 62(1):1–43, 2014.

[72] Robert I Kabacoff. *R in action: data analysis and graphics with R*. Manning Publishings Co., 2015.

[73] K Hartmann, J Krois, and B Waske. E-learning project soga: Statistics and geospatial data analysis. *Department of Earth Sciences, Freie Universitaet Berlin*, 33, 2018.

[74] Yeli Sarvina, Thomas Pluntke, and Christian Bernhofer. Comparing bias correction methods to improve modelled precipitation extremes. *Jurnal Meteorologi dan Geofisika*, 19(2):103–110, 2019.

[75] Lukas Gudmundsson, John Bjørnar Bremnes, Jan Erik Haugen, and Torill Engen-Skaugen. Downscaling rcm precipitation to the station scale using statistical transformations–a comparison of methods. *Hydrology and Earth System Sciences*, 16(9):3383–3390, 2012.

[76] Hossein Tabari, Santiago Mendoza Paz, Daan Buekenhout, and Patrick Willems. Comparison of statistical downscaling methods for climate change impact analysis on precipitation-driven drought. *Hydrology and Earth System Sciences*, 25(6):3493–3517, 2021.

[77] Douglas Maraun. Bias correcting climate change simulations-a critical review. *Current Climate Change Reports*, 2(4):211–220, 2016.

[78] M Iturbide, J Bedia, S Herrera, J Baño-Medina, J Fernández, MD Frías, R Manzanas, D San-Martín, E Cimadevilla, AS Cofiño, et al. climate4r: An r-based open framework for reproducible climate data access and post-processing. *Environ. Modell. Softw*, 2018.

[79] J Boé, L Terray, F Habets, and E Martin. Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 27(12):1643–1655, 2007.

[80] Yann Y Planton, Eric Guilyardi, Andrew T Wittenberg, Jiwoo Lee, Peter J Gleckler, Tobias Bayr, Shayne McGregor, Michael J McPhaden, Scott Power, Romain Roehrig, et al. Evaluating climate models with the clivar 2020 enso metrics package. *Bulletin of the American Meteorological Society*, 102(2):E193–E217, 2021.

[81] JH Jungclaus, Nils Fischer, Helmuth Haak, K Lohmann, J Marotzke, D Matei, U Mikolajewicz, D Notz, and JS Von Storch. Characteristics of the ocean simulations in the max planck institute ocean model (mpiom) the ocean component of the mpi-earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2):422–446, 2013.

[82] Oliver Gutjahr, Dian Putrasahan, Katja Lohmann, Johann H Jungclaus, Jin-Song von Storch, Nils Brüggemann, Helmuth Haak, and Achim Stössel. Max planck institute earth system model (mpi-esm1. 2) for the high-resolution model intercomparison project (highresmip). *Geoscientific Model Development*, 12(7):3241–3281, 2019.

[83] Annalisa Cherchi, Pier Giuseppe Fogli, Tomas Lovato, Daniele Peano, Doroteaciro Iovino, Silvio Gualdi, Simona Masina, Enrico Scoccimarro, Stefano Materia, Alessio Bellucci, et al. Global mean climate and main patterns of variability in the cmcc-cm2 coupled model. *Journal of Advances in Modeling Earth Systems*, 11(1):185–209, 2019.

[84] V Masson-Delmotte, P Zhai, A Pirani, SL Connors, C Péan, S Berger, N Caud, Y Chen, L Goldfarb, MI Gomis, et al. Contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change, 2021.

[85] W Wang and JK Vrijling. Phajm van gelder, and j. ma (2006), testing for nonlinearity of streamflow processes at different timescales. *J. Hydrol*, 322(1):247–268.

[86] Marcella Corduas. Nonlinearity tests in time series analysis. *Journal of the Italian Statistical Society*, 3(3):291–313, 1994.

[87] Cecilia Svensson, W Zbigniew Kundzewicz, and Thomas Maurer. Trend detection in river flow series: 2. flood and low-flow index series/détection de tendance dans des séries de débit fluvial: 2. séries d'indices de crue et d'étiage. *Hydrological Sciences Journal*, 50(5), 2005.

[88] Mohammad Firuz Ramli, Ahmad Zaharin Aris, Nor Rohaizah Jamil, Jabir Haruna Abdulkareem, et al. Runoff irregularities, trends, and variations in tropical semi-arid river catchment. *Journal of Hydrology: Regional Studies*, 19:335–348, 2018.

[89] Martins Y Otache, Mohammad Bakir, and Zhijia Li. Analysis of stochastic characteristics of the benue river flow process. *Chinese Journal of Oceanology and Limnology*, 26:142–151, 2008.

[90] Sintayehu Yadete Tola and Amba Shetty. Extreme hydroclimatic variability and impact of local and global climate system anomalies on extreme flow in the upper awash river basin. *Theoretical and Applied Climatology*, pages 1–21, 2023.

[91] Richarde Marques Da Silva, Celso AG Santos, Madalena Moreira, João Corte-Real, Valeriano CL Silva, and Isabella C Medeiros. Rainfall and river flow trends using mann–kendall and sen's slope estimator statistical tests in the cobres river basin. *Natural Hazards*, 77:1205–1221, 2015.

[92] Rawshan Ali, Alban Kuriqi, Shadan Abubaker, and Ozgur Kisi. Long-term trends and seasonality detection of the observed flow in yangtze river using mann-kendall and sen's innovative trend method. *Water*, 11(9):1855, 2019.

[93] Ruey S Tsay and Rong Chen. *Nonlinear time series analysis*, volume 891. John Wiley & Sons, 2018.

[94] Margherita Gerolimetto, Bisaglia Luisa, et al. Testing for (non) linearity in economic time series: a monte carlo comparison. 2014.

[95] Milan Stojković, Stevan Prohaska, and Jasna Plavšić. Stochastic structure of annual discharges of large european rivers. *Journal of Hydrology and Hydromechanics*, 63(1):63–70, 2015.

[96] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting.* Springer, 2002.

[97] The MathWorks Inc. Matlab version: 9.13.0 (r2022b), 2022.

[98] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a):427–431, 1979.

[99] Denis Kwiatkowski, Peter CB Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178, 1992.

[100] Pierre Perron. Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of economic dynamics and control*, 12(2-3):297–332, 1988.

[101] Maurice S Bartlett. Smoothing periodograms from time-series with continuous spectra. *Nature*, 161(4096):686–687, 1948.

[102] William A Broock, José Alexandre Scheinkman, W Davis Dechert, and Blake LeBaron. A test for independence based on the correlation dimension. *Econometric reviews*, 15(3):197–235, 1996.

[103] Ruey S Tsay. Nonlinearity tests for time series. *Biometrika*, 73(2):461–466, 1986.

[104] Adrian Trapletti, Kurt Hornik, Blake LeBaron, and Maintainer Kurt Hornik. Package 'tseries', 2022.

[105] Kung-Sik Chan, Brian Ripley, Maintainer Kung-Sik Chan, and Sik Chan. Package 'tsa'. *R package version*, 1, 2022.

[106] James Bernard Ramsey. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(2):350–371, 1969.

[107] Daniel MacRae Keenan. A tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72(1):39–44, 1985.

[108] David Stoffer and Maintainer David Stoffer. Package 'astsa'. *blood*, 8:1, 2022.

[109] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[110] Bernhard Pfaff, Eric Zivot, Matthieu Stigler, and Maintainer Bernhard Pfaff. Package 'urca'. *Unit root and cointegration tests for time series data. R package version*, pages 1–2, 2016.

[111] Christopher Brooks. Testing for non-linearity in daily sterling exchange rates. *Applied financial economics*, 6(4):307–317, 1996.

[112] Sandeep Kumar Patakamuri, Nicole O'Brien, and Maintainer Sandeep Kumar Patakamuri. Package 'modifiedmk'. *Cran. R-project*, 2020.

[113] Patrice Yapo, Soroosh Sorooshian, and Vijai Gupta. A markov chain flow model for flood forecasting. *Water resources research*, 29(7):2427–2436, 1993.

[114] Massimiliano Marcellino, James H Stock, and Mark W Watson. A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1-2):499–526, 2006.

[115] Zhenghao Zhang, Qiang Zhang, and Vijay P Singh. Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal*, 63(7):1091–1111, 2018.

[116] GEORGE EP Box, Gwilym M Jenkins, and G Reinsel. Time series analysis: forecasting and control holden-day san francisco. *BoxTime Series Analysis: Forecasting and Control Holden Day1970*, 1970.

[117] K Hamidi Machekposhti, Hossein Sedghi, Abdolrasoul Telvari, and Hossein Babazadeh. Modeling climate variables of rivers basin using time series analysis (case study: Karkheh river basin at iran). *Civ Eng J*, 4(1):78–92, 2018.

[118] Gabriel G Vazquez-Amabile and Bernard A Engel. Fitting of time series models to forecast streamflow and groundwater using simulated data from swat. *Journal of hydrologic engineering*, 13(7):554–562, 2008.

[119] Jan Franklin Adamowski. Peak daily water demand forecast modeling using artificial neural networks. *Journal of Water Resources Planning and Management*, 134(2):119–128, 2008.

[120] Amel Fouchal and Doudja Souag-Gamane. Long-term monthly streamflow forecasting in humid and semiarid regions. *Acta Geophysica*, 67:1223–1240, 2019.

[121] Stefano Galelli and Andrea Castelletti. Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling. *Hydrology and Earth System Sciences*, 17(7):2669–2684, 2013.

[122] Priyanka Sharma and Deepesh Machiwal. Streamflow forecasting: overview of advances in data-driven techniques. *Advances in Streamflow Forecasting*, pages 1–50, 2021.

[123] Dimitri P Solomatine and Avi Ostfeld. Data-driven modelling: some past experiences and new approaches. *Journal of hydroinformatics*, 10(1):3–22, 2008.

[124] Wen-Chuan Wang, Kwok-Wing Chau, Chun-Tian Cheng, and Lin Qiu. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of hydrology*, 374(3-4):294–306, 2009.

[125] Desalegn Chemeda Edossa and Mukand Singh Babel. Application of ann-based streamflow forecasting model for agricultural water management in the awash river basin, ethiopia. *Water resources management*, 25:1759–1773, 2011.

[126] Howell Tong. Threshold models in non–linear time series analysis," new york: Springer–verlag. 1983.

[127] R.S. Tsay and R. Chen. *Nonlinear Time Series Analysis*. Wiley Series in Probability and Statistics. Wiley, 2018.

[128] Ruey S Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American statistical association*, 84(405):231–240, 1989.

[129] Antonio Fabio Di Narzo, Jose Luis Aznarte, and Matthieu Stigler. Package 'tsdyn'. 2023.

[130] James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, pages 357–384, 1989.

[131] Daniel Pena, George C Tiao, and Ruey S Tsay. *A course in time series analysis*. John Wiley & Sons, 2011.

[132] Eugen Diaconescu. The use of narx neural networks to predict chaotic time series. *Wseas Transactions on computer research*, 3(3):182–191, 2008.

[133] Hang Xie, Hao Tang, and Yu-He Liao. Time series prediction based on narx neural networks: An advanced approach. In *2009 International conference on machine learning and cybernetics*, volume 3, pages 1275–1279. IEEE, 2009.

[134] José Maria P Menezes Jr and Guilherme A Barreto. Long-term time series prediction with the narx network: An empirical evaluation. *Neurocomputing*, 71(16-18):3335–3343, 2008.

[135] J-SR Jang. Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665–685, 1993.

[136] USGS. How much water is there on, in, and above the earth? `https://www.usgs.gov/special-topics/water-science-school/science/how-much-water-there-earth#overview`, 11/13/2019. Accessed: 10/05/2023.

[137] Solomon Mulugeta, Clifford Fedler, and Mekonen Ayana. Analysis of long-term trends of annual and seasonal rainfall in the awash river basin, ethiopia. *Water*, 11(7):1498, 2019.

[138] Eshetu Ararso Heyi, Megersa Olumana Dinka, and Girma Mamo. Assessing the impact of climate change on water resources of upper awash river sub-basin, ethiopia. *Journal of Water and Land Development*, 2022.

[139] Bekan Chelkeba, Fekadu Fufa Feyessa, and Wakjira Takala Dibaba. Climate change in the upper awash subbasin and its possible impacts on the stream flow, oromiyaa, ethiopia. *Water Science*, 37(1):179–197, 2023.

[140] Susan L Neitsch, Jeffrey G Arnold, Jim R Kiniry, and Jimmy R Williams. Soil and water assessment tool theoretical documentation version 2009. Technical report, Texas Water Resources Institute, 2011.

[141] Karim C Abbaspour. Swat-cup 2012. *SWAT calibration and uncertainty program—A user manual*, pages 1–100, 2013.

[142] Karim C Abbaspour, Saeid Ashraf Vaghefi, and Raghvan Srinivasan. 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, 2017.

[143] Georgios Sarailidis, Lampros Vasiliades, and Athanasios Loukas. Analysis of streamflow droughts using fixed and variable thresholds. *Hydrological processes*, 33(3):414–431, 2019.

[144] Sergio M Vicente-Serrano, Juan I López-Moreno, Santiago Beguería, Jorge Lorenzo-Lacruz, Cesar Azorin-Molina, and Enrique Morán-Tejeda. Accurate computation of a streamflow drought index. *Journal of Hydrologic Engineering*, 17(2):318–332, 2012.

[145] Santiago Beguería and Sergio M. Vicente-Serrano. *SPEI: Calculation of the Standardized Precipitation-Evapotranspiration Index*, 2023. https://spei.csic.es, https://github.com/sbegueria/SPEI.

[146] Thomas B McKee, Nolan J Doesken, John Kleist, et al. The relationship of drought frequency and duration to time scales. In *Proceedings of the 8th Conference on Applied Climatology*, volume 17, pages 179–183. California, 1993.

[147] Benjamin Lloyd-Hughes and Mark A Saunders. A drought climatology for europe. *International Journal of climatology: a journal of the royal meteorological society*, 22(13):1571–1592, 2002.

[148] The MathWorks Inc. Statistics and machine learning toolbox, 2022.

[149] Daniel N Moriasi, Jeffrey G Arnold, Michael W Van Liew, Ronald L Bingner, R Daren Harmel, and Tamie L Veith. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3):885–900, 2007.

[150] Mikis Stasinopoulos, Bob Rigby, and Calliope Akantziliotou. Instructions on how to use the gamlss package in r second edition, 2008.

Appendices

Appendix A

Metrological Data

Table A.1: Mean Monthly Temperature ($^{\circ}C$)

Month	Chefe	Ginchi	Debre	Ejere	Enselale	Guranda	Bole	Obs	Asgori
Maximum									
1	23.17	25.49	26.46	24.79	23.2	22.8	23.76	23.92	27.56
2	24.57	26.75	27.77	26.34	24.47	24.08	24.76	24.93	28.54
3	25.59	27	28.27	27.7	24.82	24.61	25.3	25.18	28.92
4	25.67	26.43	28.15	27.8	24.79	24.66	24.81	24.55	28.22
5	26.07	25.91	28.84	28.38	24.86	24.66	25.21	25.09	28.54
6	26.02	22	27.42	29.58	22.68	23.21	23.53	23.42	27.5
7	23.1	16.79	24.4	27.78	17.84	18.9	21.2	21.06	24.86
8	23.1	16.89	23.99	27.69	17.79	18.83	21.04	21	24.7
9	24.37	20.09	25.23	27.72	20.15	20.94	21.88	21.76	25.49
10	23.65	22.16	26.21	25.86	21.36	21.69	22.96	22.82	26.03
11	23.11	24.38	26.06	24.81	22.45	22.3	23.08	23.17	26.45
12	22.63	24.83	25.84	24.17	22.53	21.2	23.12	23.2	26.74
Minimum									
1	6.12	8.45	9.71	7.4	6.53	6	7.38	9.03	6.11
2	7.37	9.86	11.13	8.59	8.12	7.49	8.94	10.24	7.39
3	8.02	10.6	12.74	9.36	9	8.35	10.57	11.64	8.83
4	8.8	10.76	13.53	10.23	9.39	8.81	11.31	12.21	9.91
5	9.1	9.94	13.1	11.24	8.7	8.17	11.19	12.42	9.32
6	11.58	10.38	12.83	14.19	10.01	10.04	10.73	11.47	10.01
7	11.93	10.64	13.45	14.26	10.5	10.61	11.07	11.34	10.64
8	10.98	10.1	13.48	13.22	9.67	9.7	11.17	11.43	10.85
9	7.09	8.59	12.59	10.09	7.07	6.49	10.43	11.25	9.64
10	4.91	7.53	10.4	7.06	5.47	4.76	8.31	10.23	5.86
11	4.72	7.38	9.15	6.17	5.07	4.44	6.4	8.71	4.62
12	4.97	7.63	8.98	6.25	5.36	3.89	6.42	8.33	4.45

Table A.2: Mean Monthly Wind Speed (m/s)

Month	Chefe	Ginchi	Debre	Ejere	Enselale	Guranda	Bole	Obs	Asgori
1	2.5	1.98	1.36	2.23	2.21	2.35	0.7	0.7	2.48
2	2.52	2.04	1.59	2.27	2.27	2.4	0.73	0.73	2.53
3	2.5	2.05	1.58	2.32	2.32	2.42	0.78	0.78	2.53
4	2.55	2.13	1.64	2.39	2.4	2.48	0.69	0.69	2.57
5	2.65	2	1.62	2.53	2.32	2.34	0.68	0.68	2.57
6	2.41	1.44	1.31	2.63	1.57	1.76	0.44	0.44	1.96
7	2.14	1.47	1.15	2.45	1.48	1.58	0.35	0.35	1.69
8	1.98	1.29	1	2.27	1.32	1.44	0.33	0.33	1.56
9	2.07	1.33	0.97	2.25	1.5	1.64	0.45	0.45	1.78
10	2.56	1.94	1.45	2.37	2.24	2.36	0.73	0.73	2.49
11	2.64	2.01	1.53	2.36	2.3	2.44	0.74	0.74	2.59
12	2.57	2.06	1.46	2.29	2.33	1.6	0.7	0.7	2.56

Table A.3: Mean Monthly Precipitation (mm)

Month	Chefe	Ginchi	Debre	Ejere	Enselale	Guranda	Bole	Obs	Asgori
1	15.87	26.9	15.51	19.22	12.61	11.96	12.7	14.41	17.19
2	25.41	40.77	30.61	36.09	24.06	26.25	31.29	37.77	34.29
3	50.64	74.67	54.58	58.67	43.7	54.58	62.83	65.09	58.35
4	61.77	93.45	62.51	61.29	63.48	74.38	90.96	91.8	89.22
5	52.83	91.25	59.03	53.12	52.06	63.32	70.67	86.32	70
6	99.7	136.51	89.32	83.09	99.96	117.29	111.22	141.9	135.25
7	215.58	229	202.64	221.77	197.29	240.32	227.64	277.16	247.8
8	240.16	239.77	208.93	237.61	170.35	260.7	229.25	289.93	241.9
9	99.74	138.96	101.35	107.48	85.09	150.74	124.48	173.61	100.87
10	19.58	36.35	32.09	33.93	16.74	31.22	29.32	36.7	20.74
11	3.35	10.58	5.96	10.51	8.74	5.67	4.7	8.41	6.48
12	6.32	11.12	4.06	14.16	5.29	5.16	6.64	11	4.67

Table A.4: Mean Monthly Solar Radiation Mj/m^2

Month	Chefe	Ginchi	Debre	Ejere	Enselale	Guranda	Bole	Obs	Asgori
1	24.72	21.47	20.76	25.06	22.22	22.81	20.84	21.18	23.51
2	26.83	22.37	22.82	27.32	23.39	24.27	21.6	21.84	25.26
3	28.11	22.59	22.6	28.91	23.94	25.04	21.86	21.98	26.23
4	28.22	22.77	22.1	28.87	24.15	25.23	20.83	20.81	26.4
5	28.03	22.68	21.96	28.84	24.03	24.89	20.63	20.51	26.09
6	25.61	18.18	19.25	27.51	19.76	21.12	17.86	17.71	22.52
7	21.47	12.25	16.87	24.88	14.15	15.83	14.75	14.63	17.5
8	23.17	13.26	17.57	26.23	15.77	17.58	15.22	15.17	19.39
9	26.74	21.27	20.15	28.36	22.58	23.49	17.64	17.7	24.49
10	27.15	23.67	22.4	27.62	24.33	24.96	21.09	21.29	25.7
11	25.96	23.41	21.9	25.88	24.13	24.6	21.21	21.53	25.19
12	24.78	22.41	20.94	24.73	23.1	22.48	19.77	20.13	24.02

Appendix B

Distributions

Here a description of the distribution types under gamlss package [150] used in sub-sub-section 3.1.1 is listed as follows:
Normal (or Gausian) Distribution (NO(μ,σ)): The Normal distribution is parameterized its pdf is defined by ;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right] \tag{B.1}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ ¿ 0. Where μ is the mean and σ is standard deviation of Y.
Logistic distribution (LO(μ,σ)): The Logistic distribution is parameterized and its pdf is defined by;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma}\left\{exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}\left\{1+exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{-2} \tag{B.2}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ ¿ 0. Where μ is the mean and σ is standard deviation of Y.
Gumbel Distribution (GU(μ,σ)): pdf is defined by;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma}exp\left[\left(\frac{y-\mu}{\sigma}\right) - exp\left(\frac{y-\mu}{\sigma}\right)\right] \tag{B.3}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ ¿ 0.
Reverse Gumbel Distribution (RG(μ,σ)): pdf is defined by;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma}exp\left\{-\left(\frac{y-\mu}{\sigma}\right) - exp\left[-\frac{(y-\mu)}{\sigma}\right]\right\} \tag{B.4}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ ¿ 0.
Power Exponential Distributions (PE):
PE(μ,σ,ν) pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu) = \frac{\nu exp\left[-\left|\frac{z}{c}\right|^{\nu}\right]}{2c\sigma\Gamma\left(\frac{1}{\nu}\right)} \tag{B.5}$$

PE2(μ,σ,ν) pdf is defined by:

$$f_Y(y|\mu,\sigma,\nu) = \frac{\nu exp\big[-\big|z\big|^{\nu}\big]}{2c\sigma\Gamma\Big(\frac{1}{\nu}\Big)} \tag{B.6}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and ν ¿ 0. And $c^2 = \Gamma(1/\nu)\big[\Gamma(3/\nu)\big]^{-1}$.

Normal Exponential t- distribution (NET(μ,σ,ν,τ)): is a four parameter distribution and pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{c}{\sigma}\begin{cases} exp\big\{-\frac{z^2}{2}\big\}, when \quad \big|z\big| \leq \nu \\ exp\big\{-\nu\big|z\big| + \frac{\nu^2}{2}\big\}, when \quad \nu < \big|z\big| \leq \tau \\ exp\big\{-\nu\tau\log\Bigg(\frac{\big|z\big|}{\tau}\Bigg) - \nu\tau + \frac{\nu^2}{2}\big\}, when \quad \big|z\big| > \tau \end{cases} \tag{B.7}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞, ν ¿ 1, and τ ¿ ν. And $c = (c_1 + c_2 + c_3)^{-1}$. Where, $C_1 = \sqrt{2\pi}\big[1 - 2\Phi(-\nu)\big]$, $C_2 = \frac{2}{\nu}exp\big\{-\frac{\nu^2}{2}\big\}$ and $C_3 = \frac{2}{(\nu\tau-1)\nu}exp\Bigg\{-\nu\tau + \frac{\nu^2}{2}\Bigg\}$, Where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

Sinh-Arcsinh (SHASH(μ,σ,ν,τ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{c}{\sqrt{2\pi\sigma}(1+r^2)^{1/2}}e^{-z^2/2} \tag{B.8}$$

Where $\quad z = \frac{1}{2}\big\{exp\big[\tau sinh^{-1}(r)\big] - exp\big[-\nu sinh^{-1}(r)\big]\big\}$
and $\quad c = \frac{1}{2}\big\{\tau exp\big[\tau sinh^{-1}(r)\big] + \nu exp\big[-\nu sinh^{-1}(r)\big]\big\}$
and $r = (y-\mu)/\sigma$ for -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ¿0 ,ν ¿ 0, and τ¿0.

Skew Exponential Power type I distribution (SEP1(μ,σ,ν,τ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{2}{\sigma}f_{Z_1}(z)F_{Z_1}(\nu z) \tag{B.9}$$

for -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ¿0 ,−∞¡ν¡∞, and τ¿0, and where $z = (y-\mu)/\sigma$ and f_{Z_1} and F_{Z_1} are the pdf and cdf of $Z_1 \sim PE2(0, \tau^{1/\tau,\tau})$ a power exponential type 2 distribution with $f_{Z_1} = \sigma^{-1}exp[-|z|^{\tau}/\tau]$, where $\alpha = 2\tau^{(1/\tau)-1}\Gamma(1/\tau)$.

Skew Exponential Power type II distribution (SEP2(μ,σ,ν,τ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{2}{\sigma}f_{Z_1}(z)\Phi(\omega) \tag{B.10}$$

for -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞ and σ¿0 ,−∞¡ν¡∞, and τ¿0, and where $z = (y-\mu)/\sigma$ and $\omega = sign(z)|z|^{\tau/2}$ and f_{Z_1} is the pdf of $Z_1 \sim PE2(0, \tau^{1/\tau,\tau})$ and $\Phi(\omega)$ is the cdf of standard normal variable evaluated at ω.

Skew Exponential Power type III distribution (SEP3(μ,σ,ν,τ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{c}{\sigma}\Bigg\{exp\big[-\frac{1}{2}|\nu z|^{\tau}\big]I(y<\mu) + exp\big[-\frac{1}{2}|\frac{z}{\nu}|^{\tau}\big]I(y \geq \mu)\Bigg\} \tag{B.11}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞, σ ¿ 0,ν ¿ 0, and τ ¿ 0 and where $z = (y-\mu)/\sigma$ and $c = \nu\tau/\big[(1+\nu^2)2^{1/\tau}\Gamma(\frac{1}{\tau})\big]$.

Gamma Distribution (GA(μ,σ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}exp\left\{-\frac{[log(y)-\mu]^2}{2\sigma^2}\right\} \tag{B.12}$$

for y¿0, where μ¿0 and σ¿0.

Inverse Gaussian Distribution (IG(μ,σ)): its pdf is defined by;

$$f_Y(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2 y^3}}exp\left[-\frac{1}{\sqrt{2\mu^2\sigma^2 y}}(y-\mu)^2\right] \tag{B.13}$$

for y¿0, where μ¿0 and σ¿0.

Weibull Distribution (WEI1, WEI2, WEI3): their parametrization and pdf is defined by;

First parametrization (WEI1 (WEI(μ,ν))) is as follows,

$$f_Y(y|\mu,\sigma) = \frac{\sigma y^{\sigma-1}}{\mu^\sigma}exp\left[-\left(\frac{y}{\mu}\right)^\sigma\right] \tag{B.14}$$

for y¿0, where μ¿0 and σ¿0.

Second parametrization (WEI2)

$$f_Y(y|\mu,\sigma) = \sigma\mu y^{\sigma-1}e^{-\mu y^\sigma} \tag{B.15}$$

for y¿0, where μ¿0 and σ¿0.

Second parametrization (WEI3)

$$f_Y(y|\mu,\sigma) = \frac{\sigma}{\beta}\left(\frac{y}{\beta}\right)^{\sigma-1}exp\left\{-\left(\frac{y}{\beta}\right)^\sigma\right\} \tag{B.16}$$

for y¿0, where μ¿0, σ¿0 and where $\beta = \mu/\Gamma(\frac{1}{\sigma}+1)$.

Box-Cox Power Exponential Distribution (BCPE(μ,σ,ν,τ)): its parametrization and pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \frac{y^{\nu-1}f_T(z)}{\mu^\nu\sigma F_T(\frac{1}{\sigma|\nu|})} \tag{B.17}$$

for y¿0, μ ¿0, σ¿0 , τ¿0, $-\infty$¡ν¡∞, $z=(y-\mu)/\sigma$, and where f_T and F_T are pdf and cdf of a random variable T.

Generalized t Distribution (GT(μ,σ,ν,τ)): its parametrization and pdf is defined by;

$$f_Y(y|\mu,\sigma,\nu,\tau) = \tau\left\{2\sigma\nu^{1/T}B(1/\tau,\nu)\left[1+|z|^\tau/\nu\right]^{\nu+(1/\tau)}\right\}^{-1} \tag{B.18}$$

For -∞ ¡ y ¡∞, where -∞ ¡ μ ¡ ∞, σ ¿ 0,ν ¿ 0, and τ ¿ 0 and where $z=(y-\mu)/\sigma$.