BE CIVIL ENGINEERING

PROJECT REPORT

AI-SUPPORTED INVESTIGATION OF HYDROLOGICAL SUSTAINABILITY OF MALAKAND-3 HYDROPOWER PLANT

Project submitted in partial fulfillment of the requirements for the degree of

BE Civil Engineering

PROJECT ADVISOR: Dr. Muhammad Amjad

SUBMITTED BY

CSM AHMED SAEED (Syn Ldr)	**CMS 325505**
GC FAWAD KHAN	**CMS 325528**
GC MUHAMMAD ADAM	**CMS 325513**
GC SAIFULLAH	**CMS 325525**
GC HISHAM ABRAR	**CMS 325456**

MILITARY COLLEGE OF ENGINEERING,

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD, PAKISTAN.

2023

This is to certify that the

BE Civil Engineering Project entitled

AI-SUPPORTED INVESTIGATION OF HYDROLOGICAL SUSTAINABILITY OF MALAKAND-3 HYDROPOWER PLANT

SUBMITTED BY

CSM AHMED SAEED (Syn Ldr)	**CMS 325505**
GC FAWAD KHAN	**CMS 325528**
GC MUHAMMAD ADAM	**CMS 325513**
GC SAIFULLAH	**CMS 325525**
GC HISHAM ABRAR	**CMS 325456**

Has been accepted towards the partial fulfilment of requirements for BE Civil Engineering Degree

_____

Assistant Professor Dr. Muhammad Amjad

(Project Advisor)

ACKNOWLEDGEMENT

Acknowledgements Praise be to ALLAH ALMIGHTY Whose Divine Will is the reason for what we achieved. We are extremely grateful to our esteemed Project Advisor Dr. Muhammad Amjad whose unprecedented guidance, assistance and oversight helped us in completion of the project. It was his immense knowledge and technical expertise that kept us on the track during the execution of this project.

We are also thankful to our parents and families for their love, care, prayers and encouragement throughout the whole endeavor. The project would not have been possible without the praiseworthy effort, encouragement and guidance of various people who have remained involved with us throughout the duration of the project or have been of vital value to our endeavor at many different stages.

ABSTRACT

The increasing demand for renewable energy and the effects of climate change necessitate accurate hydrological monitoring and forecasting for hydropower plants. This study focuses on the Malakand-3 hydropower plant in Khyber Pakhtunkhwa, Pakistan, and aims to investigate the hydrological sustainability of its input water source in the view of changing climate conditions. The study area encompasses the Swat River Catchment characterized by its complex topography as well as its potential for irrigation and hydroelectric power generation. The research addresses the limitations of previous studies by incorporating a broader range of data, including discharge, precipitation, and temperature trends for the past decade. A hydrological model was developed and calibrated to simulate the Swat River streamflow during the recent past. Furthermore, the study utilized ARIMA and neural network forecasting methods to forecast future streamflow trends. The ARIMA model, specifically ARIMA (4,1,1) (0,1,0) [365], was found to be the most suitable for forecasting. The results indicated an increase in flood frequency and rising water levels in Swat River, which indicates advantages for future hydropower production but challenges for the working of Malakand-3 hydropower plant due to forecasted floods. The study highlights the significance of accurate forecasting for optimizing power plant operations and flood management. The findings also demonstrate the potential for hydrologically sustainable development and for installation of additional hydroelectric power plants along the Swat River. This work intends to contribute to enhancing the understanding of water flow dynamics in the study area and provides valuable insights for decision-making in the sectors such as renewable energy, water resources management, and disaster management.

TABEL OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Background

The increasing global demand for energy has led to a significant rise in the deployment of renewable energy, particularly from water resources, which are considered highly sustainable. However, the changing climate and rising temperatures have resulted in an increase in water-related hazards, emphasizing the importance of accurately forecasting these trends. Energy production forecasting plays a crucial role in optimizing the marketing of renewable energy and facilitating systems integration. However, the production of renewable energy is heavily dependent on environmental conditions such as precipitation, wind power, and geothermal heat. Consequently, the fluctuating nature of these conditions makes it challenging to forecast the feed-in of renewable energy into the power grid. Addressing the need for energy forecasting is essential to ensure the production of green energy without causing environmental hazards.

The significance of energy production forecasting becomes even more pronounced as the contribution of green energy to the power grid continues to grow. Several countries have implemented laws prioritizing green energy over fossil fuel-based energy sources due to environmental concerns. In certain remote areas, energy from local small hydropower plants becomes the sole source of power.

For small hydropower plants, production forecasting is pivotal for sustainable development as it ensures an adequate supply of energy resources. At the same time, accurate demand forecasting for grid utilization is necessary. Forecasting energy production in the short term is an essential tool for maintaining a consistent power supply, planning reserve power resources, facilitating energy transactions between power stations, and assessing available resources. Energy production forecasting is a dynamic process that requires regular updates of weather measurements, discharge data, and previous energy production to ensure proper calibration of the forecasting system. However, controlling energy production is not solely dependent on uncontrollable power

resources; it also relies on managerial strategies implemented at power plants, which are influenced by technical limitations, price fluctuations, user demand, and other factors. Moreover, the availability of comprehensive information is not always guaranteed, making renewable energy forecasting challenging to manage. Additionally, forecasting energy production based on weather conditions necessitates forecasting the weather itself, which introduces errors that impact the final results. As the forecasting process requires extensive data, it is crucial for relevant departments to maintain accurate and up-to-date information, employing new processes to achieve error-free data forecasting.

1.2. Problem statement

Pakistan is highly vulnerable to the impacts of changing climate and has witnessed several climatic disasters during the last few decades. Khyber Pakhtunkhwa has experienced intense weather events because of climate change like the 2010 mega floods. With each passing year the meteorological conditions are getting intense with long heat spells during summers and short winters.SO there is a need for the forecasting of Malakand-3 hydropower plant as it is one of the largest hydropower plant in terms of production (81-83MW) while keeping in mind the changes in climate condition as it is directly affecting the inflow to the hydropower plant.

1.3. Literature Review

1.3.1. Climate Change of Pakistan

Pakistan has been significantly affected by the impacts of climate change, as evidenced by numerous extreme weather events over the past decade. These events, including floods, droughts, glacial lake outbursts, hurricanes, and heat waves, have resulted in substantial loss of life, property damage, and hindered economic growth. One notable example is the devastating super flood of 2010, which resulted in the loss of 1,600 lives and affected an area of 38,600 square kilometers (km2). The economic damage caused by this event alone was estimated to be around $10 billion.[17] Similarly, the Karachi heat wave (June 2015) led to the death of more than 1,200 people.[18]

Geography and Climate of Pakistan:

Pakistan spans a vast area of approximately 796,000 km2, encompassing diverse climates with varying temperatures and rainfall patterns. In the eastern regions of the southern half of the country, rainfall is primarily received during the southwest summer monsoon season, which occurs from June to September. Conversely, the northern and western regions of the southern half experience rainfall mainly during the western winter season, spanning from December to March. The summer monsoon contributes to about 60% of the total annual rainfall in Pakistan.

The climate across the country varies from arid to semi-arid, with approximately three-quarters of the nation receiving less than 250 millimeters (mm) of rainfall annually. The annual precipitation levels range between 760 mm and 2,000 mm. The northern region of Pakistan is characterized by some of the world's tallest mountains, including K-2, which stands at an impressive height of 8,611 meters (m). Additionally, this region is home to vast glaciers such as Siachen, stretching over 70 kilometers (km), and Biafo, spanning 63 km. [16] In winter, temperatures in the region drop to minus 50 °C and remain around 15 °C during the warmest months from May to September..[19] The western and southern parts of the country represent the plains of the Indus Valley and the Balochistan Plateau. The trans boundary Indus Basin covers 520,000 km^2 or 65% of the country's total area, including the entire Punjab, Khyber Pakhtunkhwa, most of Sindh and eastern Balochistan. [18] The Indus Basin Irrigation System is the largest continuous irrigation system in the world, accounting for 95% of the country's total irrigation system. Average annual rainfall in the Indus Valley is about 230 mm. The temperature difference between the upper and lower parts of the basin is evident. Average winter temperatures (December to February) range from 140°C to 200°C in the lower layers and 20°C to 230°C in the upper layers, with average summer temperatures. (March to June), the average monthly temperatures are 420°C to 440°C in the lower layers and 230°C to 490°C in the upper layers. The Balochistan Plateau is a vast wilderness area of mountains in the southwestern part of the country with an average altitude of about 600 m. Rivers flow through the region seasonally, but much of the northwest is vast desert. A desert in the central part of the country, such as Thar and Cholistan. Precipitation in this area is less than 210 mm per year or 20-30 mm per month.

Demography Pakistan is currently one of the most populous countries in the world, with an estimated population of approximately 184.5 million people. It is important to note that the population figures mentioned are based on data available up until my knowledge cutoff in September 2021.Pakistan has experienced a relatively high average annual population growth rate of around 2% in the past:.[20] Pakistan has a population density of approximately 231 persons per square kilometer. It is worth noting that this figure is based on the information available up until September 2021. To obtain the most recent data, it is recommended to consult updated sources such as national statistical agencies or international organizations like the United Nations.Around 37% of Pakistan's population resides in urban areas. It is important to note that urbanization rates may vary over time, and recent data should be consulted for the most accurate figures.According to the information you provided, approximately 47% of the urban population in Pakistan are slum dwellers.[21] The poverty rate estimated at $2 per day purchasing power parity exceeds 50% of the total population with stark provincial disparities.[22] The southern sub-regions of all provinces in Pakistan are known for having a significantly higher incidence of severe poverty compared to their northern counterparts, except for the province of Khyber Pakhtunkhwa. In Khyber Pakhtunkhwa, severe poverty rates are equally high in both the northern and southern sub-regions.[23] Pakistan faces significant challenges due to high levels of poverty and limited access to resources, which contribute to its classification as a country with low human development. It is ranked 146 out of 187 countries in terms of human development, falling below the average Human Development Index compared to other South Asian nations. The vulnerability of Pakistan to climate change threats is highlighted in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) for the Asia region. Being an agriculture-dependent economy, Pakistan's sensitivity to climate change stems from various factors, including its geographical characteristics, demographic trends, socioeconomic conditions, and limited adaptive capacity. These factors collectively contribute to the vulnerability profile of the country, exacerbating the cycle of poverty. [24] According to the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) for South Asia, climate change projections indicate that the region is likely to experience above-average warming compared to the global mean. These changes in temperature are expected to impact glaciers' melting rate and precipitation patterns, particularly

affecting the timing and intensity of monsoon rainfall. As a result, sectors reliant on water, such as agriculture and energy, will be significantly affected in terms of productivity and efficiency.

1.3.2. Malakand-3 Hydro Power Plant

The Malakand-3 hydropower project was initiated after the construction of the Swabi Salinity Management and Reclamation Program (SCARP) auxiliary tunnel in the Upper Swat Canal (USC) system. The USC system was completed in 1918 and originates from the Swat River at Amandara Head Works. It provides irrigation to around 121,400 hectares of land in the Peshawar Valley. The water from Amandara is transported through six kilometers of canals to the base of the Malakand hills. To redirect the water to Dargai Nullah, a 3.5 km long tunnel called the Benton Tunnel was constructed through the Malakand Hills. Two 20 MW power plants were built at Javan and Dargai between the exit of the Benton Tunnel and the Burificator in 1937 and 1953, respectively, using this water. To meet the increased demand for irrigation water, the Amandara Head Works was expanded under the Swabi SCARP, and an auxiliary tunnel was built parallel to the existing Benton Tunnel to accommodate the planned increase in drainage. The Malakand III hydropower project was designed to utilize this additional discharge for hydropower generation before supplying the Machai branch. The project was intended to operate in conjunction with the existing power plants at Jabban and Dargai, considering the variations in water availability in the Swat River. It is worth noting that apart from fulfilling hydraulic requirements, the feasibility of Malakand-3 also considered the irrigation needs. As a result, the 'Visai' irrigation project, designed for gravity irrigation of approximately 25,000 acres of land, is being constructed by the Ministry of Irrigation from the outlet of the sediment separation unit of Malakand III. The Malakand III project was completed in 2008 and connected to the national grid through the 132 kV Dargai grid station. Under a 25-year Power Purchase Agreement (PPA) with the National Transmission & Dispatch Company (NTDC), the electricity generated is sold to them. The selling price to NTDC was determined by the National Electric Power Regulatory Authority (NEPRA). The project achieved commercial operation on November 1, 2008, and has been successfully meeting the required targets. It has generated sales of over Rs. 28 billion rupees to date.[13]

Salient Features of the Hydropower Plant: [13]

- Location: Dargai (Malakand Agency)
- Capacity: 81 MW
- Rated Flow: 51 Cumecs
- Net Head: 181 m
- Annual Energy: 549 GWh / annum
- Total Project Length: 10 km
- Headrace channel: 2.62 km
- Tunnels (5 Nos.): 3.33 km
- Cross Drainage Structures:20
- Sediment Excluder: 140 m Long X 45 m Wide X 11.25 m Deep
- Forebay: 60 m long, 14 m wide and 14.20 m deep
- Spillway: 886 m long (726m Spillway Cascade & 160m long Spillway Tunnel)
- Penstock: 832 m Long with 4 m Diameter, running through 120 m tunnel.
- Powerhouse: 46 m Long X 36 m Wide (in 32 m deep excavation).
- Turbines: 3 No. Vertical Shaft Francis turbines.
- Tailrace: 1005 m long concrete lined channel.
- Wasteway: 2385 m long stone pitched channel.
- Irrigation from Bezai Scheme: 20,000 Acres

1.3.3. Forecasting with AI

1.3.3.1. Time Series Forecasting

Time series forecasting is a method of examining and analyzing time series data recorded or collected over a period of time. This technique is used to forecast values and make predictions for the future.[1] Top four types of forecasting methods are; 1) Straight-line, 2) Moving average, 3) Simple linear regression,4) Multiple linear regression.

Table 1 Time Series Types

TECHIQUES	USES	MATH INVOLVED	DATA NEEDED
Straight line	Constant growth rate	Minimum level	Historical data
Moving average	Repeated forecasts	Minimum level	Historical data
Simple linear regression	Compare 1 independent with one dependent variable	Stratistical knowledge requied	A sample of relevant observations
Multiple linear regression	Compare more than one independent variable with one dependent variable	Stratistical knowledge requied	A sample of relevant observations

While developing models using time series data, it is crucial to grasp the underlying patterns present in the data over time. These patterns can be categorized into four components:

1. Trend: The trend component captures the gradual changes or shifts observed in the time series data over an extended period. Trend patterns exhibit long-term increases or decreases in the data.
2. Level: The level component represents the baseline value or the average level of the time series. It is often depicted by a straight line that indicates the central tendency of the data.
3. Seasonality: Seasonality refers to the recurring and predictable patterns that repeat within a fixed unit of time, such as daily, weekly, monthly, or yearly cycles. These patterns may be influenced by factors like weather, holidays, or other regular events.
4. Noise: The noise component, also known as residual or error, accounts for the random fluctuations or irregularities present in the data. It represents the unpredictable variations that cannot be explained or predicted by the model. These fluctuations may arise from measurement errors, random events, or unaccounted factors.

By understanding and incorporating these four components—trend, level, seasonality, and noise—into time series modeling, it becomes possible to capture the various dynamics and patterns exhibited by the data over time. This understanding helps in developing accurate forecasting models and gaining insights into the underlying behavior of the time series.[1]

Time series data finds applications in various fields, including but not limited to signal processing, pattern recognition, earthquake prediction, weather forecasting, econometrics, automated control, and electroencephalography. A time series data is characterized by a sequence of observations collected over time, typically denoted as y1, y2, and so on. What distinguishes time series data from other data types is the consideration of time dependence when analyzing the data. In formal terms, a time series is a collection of pairs, where each pair consists of a time value representing when the parameter was measured and the corresponding parameter value itself. It is essential that the time values are unique, ensuring that each observation is associated with a distinct point in time. However, multiple observations may have the same parameter value. The inherent time dependency in time series data enables the analysis of patterns, trends, and relationships over time. This characteristic is particularly valuable in understanding how the parameter of interest changes and evolves with respect to time. By employing various techniques of time series analysis, meaningful insights can be derived, facilitating prediction, forecasting, and decision-making in their respective fields of application.[2]

If the observations are made at specific times and occur at regular intervals, the nature of the time series can be classified as both continuous and discrete. However, when the observations are not evenly spaced, the time series becomes non-uniform. Furthermore, time series can be categorized based on the type of values recorded. They can either be one-dimensional or multidimensional, representing multiple parameters at any given time. It is also worth noting that the parameters within a time series can be non-numeric, allowing for the representation of variables beyond numerical values. [3]. Additionally, it is worth noting that a time series process can be categorized as either deterministic or probabilistic. A deterministic process implies that the parameters of the time series can be accurately predicted, whereas a probabilistic process indicates that past values only partially determine the future, making precise forecasting unattainable. When designing time series models, an important consideration is the assumption of stationarity in the process. A stationary stochastic process typically exhibits the following characteristics:

Constant Mean: Mean of the process remains constant over time.

Constant Variance: The variance of the process remains constant over time.

Autocorrelation Function (or Spectral Density Function): The correlation between observations at different time points follows a consistent pattern.

Strictly stationary processes maintain these properties, ensuring that the parameters such as mean and variance do not change over time. Additionally, the joint probability distributions of strictly stationary processes remain constant over time.

Mean of a stationary process represents the level at which the process fluctuates consistently over time. By assuming stationarity, time series models can effectively capture and analyze patterns and relationships in the data. This allows for reliable predictions, accurate forecasting, and informed decision-making based on historical information. The concept of stationarity ensures the stability of statistical properties, such as the mean, variance, and autocorrelation, which are essential for meaningful analysis of time series data..[4]

$\mu = E[yt\ ] = Z\ yp(y)dyS$

The variance of a stationary process is the spread about this level.

$\sigma\ 2\ y = E[(yt - \mu)\ 2\ ] = Z\ (y - \mu)\ 2\ p(y)dy$

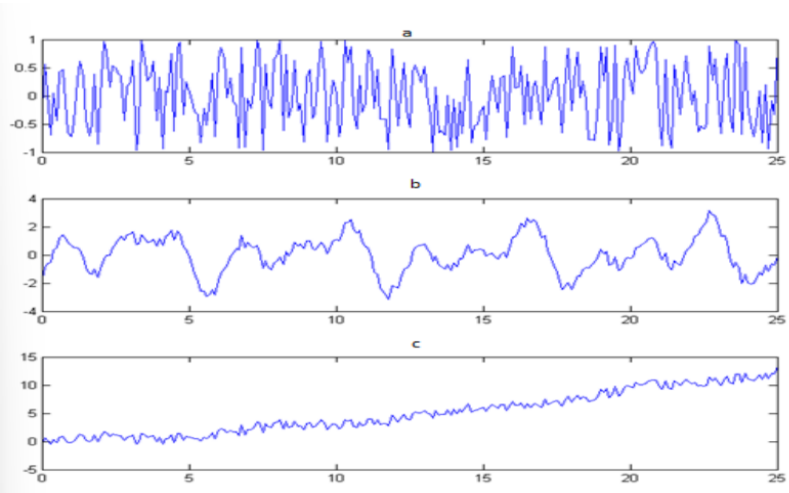

Figure 1 Stationary process

On the contrary, the additive model for a time series y1, y2, ..., yn assumes that each value in the series can be decomposed into four distinct components: trend, long-term seasonal influence, short-term seasonal influence, and deviations (or errors).

Yt = Tt + Zt + St + Rt

Where t = 1, 2, . . ., n .[5]

Here Tt is a trend, a monotone function.

In this context, Zt represents the long-term seasonal influence, while St represents the short-term seasonal influence. Together, they can be combined as a single variable Gt = Zt + St, which characterizes the long-term behavior of the time series. The random variable Rt summarizes all the deviations or errors in the series. Formally, the task of time series forecasting can be framed as a function minimization problem. Let y(t) be a function representing the original process. The objective is to find a function ˇy(t) that can effectively describe the process while minimizing the difference between y(t) and ˇy(t). This can be achieved by minimizing the residual, |yt - ˇyt|, where yt represents the actual value of the time series at a given time t and ˇyt represents the corresponding predicted value. The residuals provide information about the accuracy of the predictive model, and their accuracy can be assessed through confidence intervals using least-squares estimates. Ideally, the optimal solution would be when y(t) equals ˇy(t), resulting in the minimization (or optimization) of the time series forecasting problem.

1.3.3.2. ARIMA Models

ARIMA model is the combination of AR, MA and ARMA.

- Autoregressive model (AR) considers a current value as an aggregate of a previous deviations and a shock a_t.

Let $\check{y}t, y\check{t}-1, \ldots$ be deviations from the mean the process fluctuates about:

$\check{y}t = yt-\mu.$

Then the autoregressive process is defined as[4]:

$y\check{t} = \varphi_1 y\check{t}-1 + \varphi_2 y\check{t}-2 + \ldots + \varphi p y\check{t}-p + a_t$

The autoregressive process, denoted as AR(p), involves a parameter p that represents the order of the process. The process is characterized by the presence of a random shock, denoted as at, and can be referred to as "regressive" because it exhibits a tendency to return to its previous value, and "automatic" because it reverts to its own past values. This process captures dependencies between values and provides a framework for modeling errors or residuals.

To simplify notation, an autoregressive operator is defined to represent the process.

$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p$

or $\varphi(B)\check{y}t = a_t$

The model has $p + 2$ parameters: $\mu, \varphi 1, \ldots, \varphi p, \sigma_2{}^a$, where $\sigma_2{}^a$ is the variance of white noise.

- Moving average (MA) model consider the current value $\check{y}t$ is dependent linearly on a finite number of q previous a's:

$\check{y}t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$

Such a process is called a moving average process of order p. As in autoregressive models, for short, a moving average operator is defined [4]

The autoregressive (AR) process can be represented by the autoregressive operator θ(B) = 1 - θ1B - θ2B^2 - ... - θqB^q, where B is the backward shift operator. The predicted value of the time series, denoted as y̌t, is obtained by multiplying the autoregressive operator θ(B) with the random shock at, resulting in y̌t = θ(B)at. The AR model consists of q + 2 parameters: μ, θ1, θ2, ..., θq, and σ^2a. Despite its name, the term "autoregressive" might be somewhat misleading, but it is widely used in the field.

On the other hand, the Autoregressive Moving Average (ARMA) model combines both autoregressive (AR) and moving average (MA) models, and is considered to provide better fit for real-time series compared to individual models. The ARMA model is given by the equation: y̌t = φ1y̌t−1 + φ2y̌t−2 + ... + φpy̌t−p + at + θ1at−1 - θ2at−2 - ... - θqat−q, which can also be written as φ(B)y̌t = θ(B)at. The ARMA model involves p + q + 2 unknown parameters: μ, φ1, φ2, ..., φp, θ1, θ2, ..., θq, and σ^2a. Typically, stationary time series can be adequately described by models with p and q values less than 2.[4]

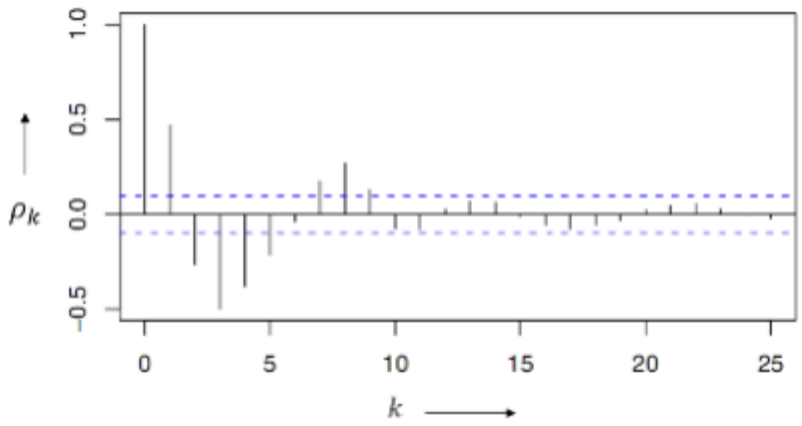

Figure 2: ARMA Model

In many real-life cases, time series data may not exhibit stationarity and may not exhibit variations around a constant mean value. To analyze such nonstationary processes, we can utilize the Autoregressive Integrated Moving Average (ARIMA) model.To apply the ARIMA model, we first consider the dth difference of the process to transform it into a stationary form. The ARIMA model combines the Autoregressive (AR) and Moving Average (MA) models with this differencing operation.To illustrate this, let φ(B) represent the generalized autoregressive operator defined as φ(B) = φ(B)(1 - B)^d, where ϕ(B) corresponds to the stationary operator and d represents the order of differencing required to achieve stationarity. One condition for this model is that φ(B) = 0^d should have a root of 1.

By employing the ARIMA model, we can capture the dynamics of nonstationary time series data and create a suitable model to explain their behavior.[4].

Thus we have

$\varphi(B)z_t = \phi(B)(1-B)^d z_t = \theta(B)a_t$

We can rewrite this in a following way:

$\phi(B)\omega_t = \theta(B)a_t$ where $\omega_t = \nabla^d z_t$.

In practice, the order of differencing in the ARIMA model, denoted as d, typically takes values of 0, 1, or 2. ARIMA models are also applicable for seasonal time series data.The primary objective

of model identification is to determine appropriate values for the parameters p, d, and q in the ARIMA model, as well as selecting a subset of all possible ARIMA models that can effectively describe a given time series. To achieve this, it is necessary to make the given time series stationary by differencing it, often denoted as zt.

The autocorrelation function (ACF) can be used in this process. The ACF of the ARIMA process satisfies the difference equation φ(B)ρk = 0, where φ(B) represents the parameter of the autoregressive (AR) process, ρk represents the autocorrelation, q is the order of the moving average (MA) process, and k is the lag. In stationary processes, the roots of the equation φ(B) = 0 lie outside the unit circle, resulting in the autocorrelation function decaying for larger lag values.Typically, if the partial autocorrelation function (PACF) does not decay rapidly, it suggests that the process is likely to be nonstationary. This observation also applies to nonlinear models. To achieve stationarity, the appropriate differencing order, d, should be selected for the ARIMA model. $\omega_t = \nabla^d z_t$. such that the autocorrelation function decays rapidly. According to Box and Jenkins [4], Typically, the differencing order, denoted as d, is chosen to be 0, 1, or 2. To determine the appropriate value for d, the first 20 autocorrelation values of the initial time series can be observed, focusing on the difference between the first and second orders. Once d is determined, the next step is to select the parameter p.When considering the order of q, the following points should be kept in mind:

- The autocorrelation function of an autoregressive (AR) process with p orders decreases smoothly, while the partial autocorrelation functions "vanish" after lag p.

- Conversely, for a moving average (MA) process with q orders, the autocorrelation function "vanishes" after lag q, while the partial autocorrelation functions drop smoothly.

- For ARIMA processes, the autocorrelation function after the first q - p values can be represented as a combination of exponential and decaying harmonic curves. Similarly, the partial autocorrelation function follows this pattern with p - q values.

An initial attempt can be made with p = 1 and q = 0 to build the model. If the prediction results are not satisfactory, adjustments can be made by trying different values, such as p = 2 and q = 1, before estimating a new model. The Box-Jenkins approach provides a diagram to determine parameters

for models like AR(2), MA(2), and ARMA(1,1). In general, the parameters of an ARMA(p, q) model can be estimated using the first p + q + 1 autocovariances.

For estimation of a built model Maximum Likelihood estimation [6] and Least Squares method [7] are used.

1.3.3.3. Machine Learning Model

A machine learning model is a computational file or algorithm that is capable of recognizing recurring patterns or patterns associated with a specific timeline. By providing data inputs, the model learns and gains insights about these patterns, making it a part of the broader field of artificial intelligence. The process of training and evaluating the model helps it understand the data and make predictions based on the learned patterns. Machine learning models are considered predictive tools, enabling efficient and streamlined experimentation and prediction-making.

One of the advantages of machine learning models is their ability to process and analyze large volumes of data efficiently, leveraging the available time and memory capacity. This capability allows for the handling of thousands of experiments, making the overall process more efficient and manageable.

In the context of time series forecasting, traditional methods such as ARIMA (AutoRegressive Integrated Moving Average) and ETS (Exponential Smoothing State Space) models have been widely used. These methods were considered classic models and were typically applied to monthly, quarterly, and yearly data. However, these traditional methods faced limitations when dealing with high-frequency time series data, and their application was often slow and imprecise.

Furthermore, traditional methods like ARIMA and ETS were not designed to model multiple time series together to identify similarities or common patterns between different series. This limitation hindered their ability to capture complex relationships and make accurate predictions for diverse datasets.

Machine learning techniques, on the other hand, offer the potential to overcome these limitations. By leveraging advanced algorithms and approaches, machine learning models can handle high-frequency time series, capture interdependencies between variables, and provide more accurate

predictions by leveraging the patterns discovered in the data. These models can effectively learn from historical data to make predictions on unseen or future observations, enabling more robust and flexible forecasting capabilities.[8]

Machine Learning Models are either supervised or unsupervised. Supervised learning uses the input-output functions for learning of the model.

Supervised models are further subdivided into:

1. Regression model
2. Classification model

Regression model: In this model we have a continuity in the output by providing the input data. It consist of models such as linear regression, decision tree, neural network etc.

Linear regression draws a line that fits the data. Two variables are used that form a linear equation for the observed data. Independent and dependent variables. A scatterplot is formed between these two variables to show the relationship between them. If there is no relationship between these two variables, i.e. no uptrend and downtrend, the model does not make sense. The relationship between these two variables is given by a coefficient between 1 and minus 1. The equation formed is $Y = a + bX$,

where

X is the independent variable

Y is the dependent variable

b is the slope

a is the intersection point.

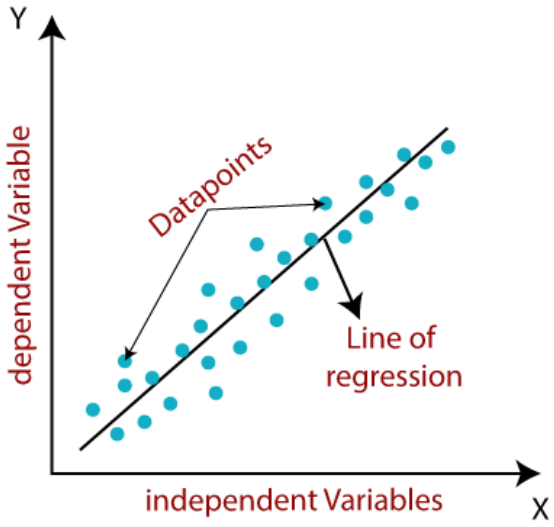

Figure 3 Linear Regression

Linear regression is one of the most commonly used approaches for modeling relationships between two or more variables. Multiple linear regression plots the best-fit plane and polynomial regression plots the best-fit curve. You can use this to find the global temperature and yield. Decision trees are used for research, planning, and machine learning. Nodes are used in decision trees. The higher the number of nodes, the more accurate the decision tree.[9] The final node that makes the final decision is called the leaf of the tree. They are easy to make, but very difficult to achieve accuracy. A regression tree is a decision tree in which a target variable is given continuous values. Use separate selection and exit criteria. Decisions are explained, events to occur are identified, and outcomes are predicted. A regression tree model divides the data into nodes, branches, and leaves. The method consists in dividing the prediction space into regions. Experimental predictions typically use the mean of the data. A prediction space is formed by dividing the space using such trees, and this method is called the decision tree method. These methods are simple and aid in interpretation.

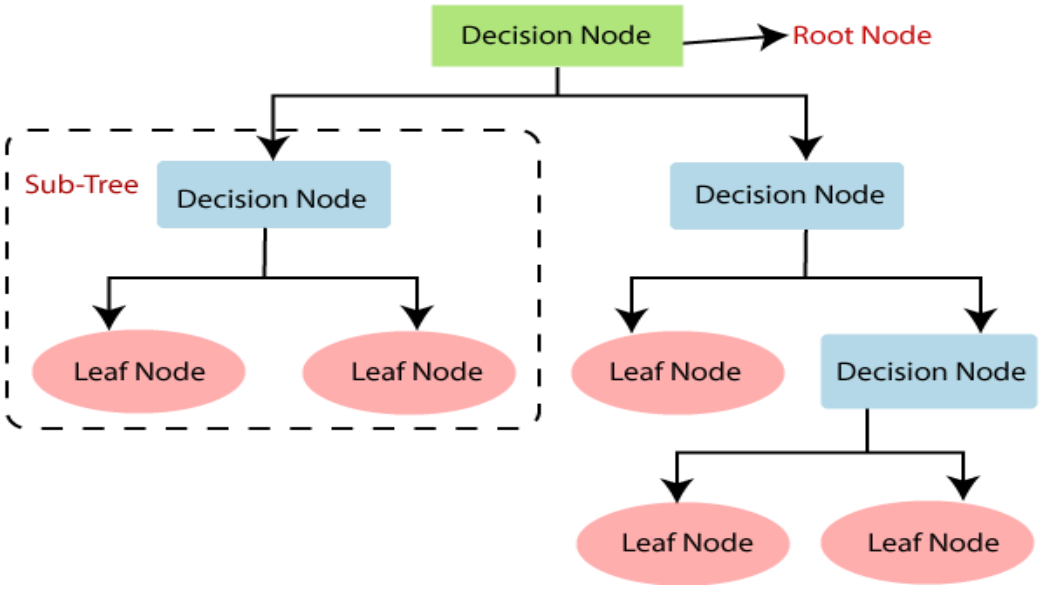

Figure 4 Decision Tree

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are named after neurons in the human brain for their functionality. This is a network of formulas. Multiple input variables can be stored to process equation networks, thus providing results as multiple output variables. This is the relationship between the network of vector inputs and vector outputs. It is a machine learning and deep learning algorithm tool. Its node layer consists of an input layer, a hidden layer, and an output layer. Each node or artificial neuron connects to another node using a threshold. When the threshold for one specified node is exceeded, the node is activated, and data is sent to the next tier. Otherwise the data will not be passed to other layers.[10]

Information is first entered into the input layer. Input nodes process data. Analyze the data and pass it to the next layer while classifying it.

These layers receive data from input layers or other hidden layers. An artificial neural network can have many hidden layers. Hidden layers analyze and process output data from other layers and pass the data.

The final result is provided by the output layer. Consists of single or multiple nodes. Returns the answer in binary form I-e (1 or 0).

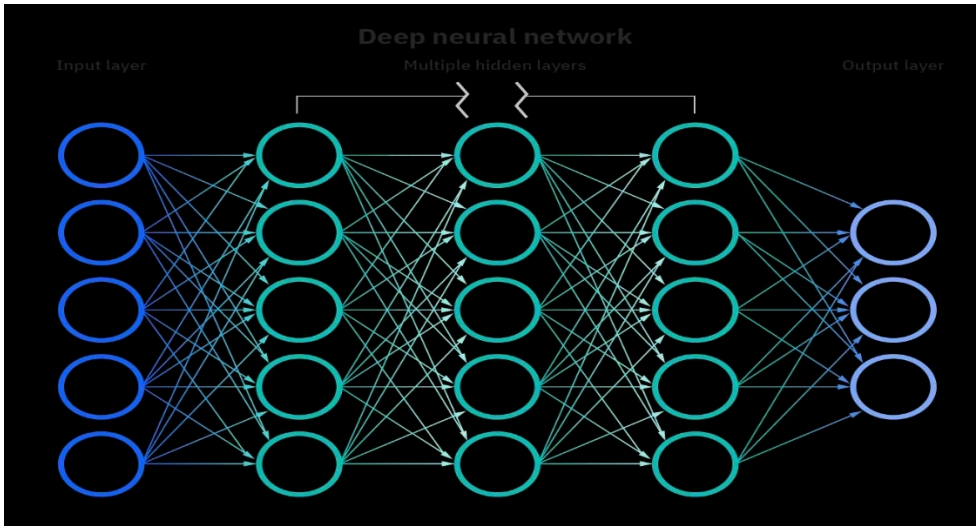

Figure 5 Neural Network

On the other hand we have classification models such as Support Vector Machine. In classification models, the output is discrete.

Support vector machines are placed into a supervised learning classification model. This supervised learning algorithm is used for both classification and regression problems. It's basically a machine learning tool. The SVM algorithm creates decision boundaries that can divide the n-dimensional space into classes, allowing new data inputs to be placed into appropriate categories for data prediction. This decision boundary is called a hyperplane. SVM selects extreme vectors or points to help create hyperplanes. The extreme vectors are called support vectors. There are linear SVMs and nonlinear SVMs.[11]

Linear SVM: For linear separation of data, use a linear SVM that divides the dataset into two classes by a boundary (straight line). Such data are called linearly separable data.

Nonlinear SVM: For nonlinear segregated data, use a nonlinear SVM in which the dataset cannot be divided into two classes by a boundary (straight line). Such data are called nonlinear data.

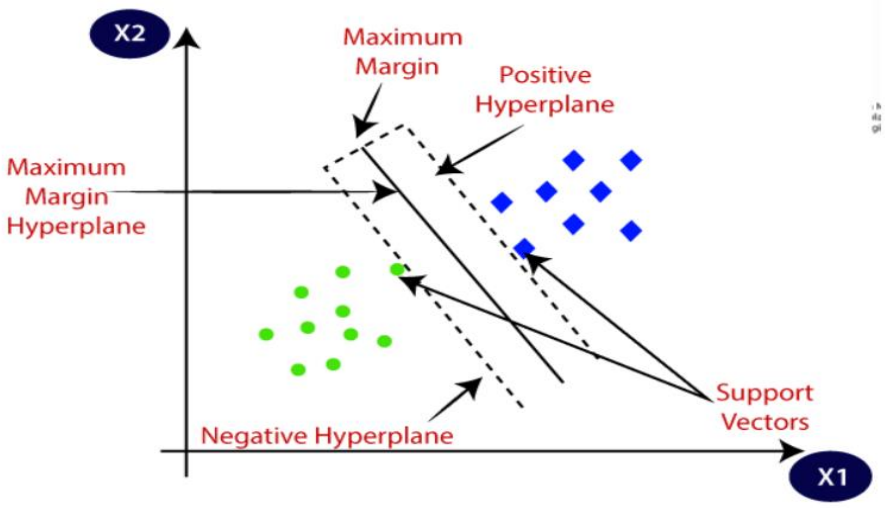

Figure 6 support vector machine

1.3.3.4. Related Works

ARIMA

In 2003, Zhang proposed a hybrid model called ARIMA-NN, which combines the strengths of Autoregressive Integrated Moving Average (ARIMA) and Neural Network (NN) models to improve prediction accuracy. The model assumes that the time series data consists of both linear and non-linear components. By incorporating both models, ARIMA-NN aims to capture the complex patterns and relationships in the data more effectively. The study conducted by Zhang and Min (2005) demonstrated that the ARIMA-NN hybrid model outperformed each individual model in terms of prediction accuracy. The research was published in the European Journal of Operational Research.

Similarly, in their hybrid model, Babbu and Reddy (2014) utilized a moving average filter to decompose the given time series data into two subsets. The ARIMA and NN models were then separately applied to each subset, and the forecast results from both models were combined to generate the final prediction. This approach aimed to leverage the strengths of both models and

enhance the accuracy of forecasting. The study conducted by Babbu and Reddy was published in the Applied Soft Computing journal.

Feedforward Neural Networks

From the comparative study held by Faraway et al. [8] In a study conducted by Faraway and Chatfield (1998), it was observed that neural networks (NNs) with a multilayer structure and a large number of nodes can fit the training data well. However, when it came to forecasting, these models did not perform as effectively. The research, which used airline data as a case study, was published in the Applied Statistics journal. This suggests that while NNs may have the ability to capture complex patterns in the data during training, they may not generalize well for future predictions.

Additionally, Peter Zhang and Min Qi conducted research on the forecasting capabilities of neural networks for seasonal and trend time series data. They found that NNs were not effective in handling seasonality or trends when working with raw, unpreprocessed data. This study was published in the European Journal of Operational Research in 2005. It highlights the limitations of NNs in capturing and modeling the patterns related to seasonality and trends in time series data.

Recurrent Neural Networks

Failure prediction using ANN-based models for PV plants maintenance tasks was studied by F.A. Olivencia Poloa, J.F. Bermejoa, J.F. Gómez Fernándezd, and A.C. Márquezd in their paper titled "Failure mode prediction and energy forecasting of PV plants to assist dynamic maintenance tasks by ANN-based models" (Renewable Energy, 2014).

The determination of representative time horizons for short-term wind power prediction using artificial neural networks was investigated by E. Izgia, A. Oztopalb, B. Yerlib, M.K. Kaymakb, and A.D. Sahinb in their article titled "Determination of the representative time horizons for short-term wind power prediction by using artificial neural networks" (Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2014).A research conducted by students from Stanford University, E. Busseti, I. Osband, and S. Wong, explored deep learning techniques for

time series modeling in their technical report titled "Deep learning for time series modeling" (CS 229, Stanford University, December 2012).

K. Hatalis, P. Pradhan, S. Kishore, and R.S. Blum applied nonlinear recurrent neural networks for multi-step forecasting of wave power in their paper titled "Multi-step forecasting of wave power using a nonlinear recurrent neural network" (Proceedings of the PES General Meeting — Conference Exposition, IEEE, June 2014, National Harbor, MD).

Support Vector Machines

Papadimitriou et al. [11]SVM model showed a good ability for short-term forecasting and predicting peaks of a time series process. (T. Papadimitriou, P. Gogas, and E. Stathakis. Forecasting energy markets using support vector machines. Energy Economics, 44:135–142, 2014.)

1.3.4. Forecasting In Pakistan

We evaluate three statistically based forecasting options for their practical feasibility in developing models for seasonal runoff forecasting in the study area.;[15]

-a statistical approach using the **Bayesian joint probability (BJP) model** with predictors accounting for antecedent basin conditions and climate drivers

-an ESP approach using the **snowmelt runoff model (SRM)**

-a hybrid approach using option with an additional predictor – **ARIMA model.**

BJP forecasting models

The Bayesian Joint Probability (BJP) approach, introduced by Wang et al. in 2009, is a statistical seasonal forecasting model used to optimize the transformation of available information for discharge forecasting scenarios. This approach incorporates various factors such as previous catchment conditions, large-scale climate forcing through climate indices, and hydrological models. The BJP model utilizes the conditional multivariate normal distribution to simulate the relationship between predictors and predictions. The predictors and predictor data are transformed

using the Log(Sinh) or Yeo-Johnson transformation techniques. The model parameters are derived using the Markov Chain Monte Carlo (MCMC) method, which takes into account parameter uncertainties, especially in cases with limited data. To generate probabilistic (ensemble) predictions, samples are generated from the estimated conditional multivariate normal distribution. Several studies, including Robertson et al. (2013), Robertson and Wang (2012), Schepen et al. (2012), and Wang and Robertson (2011), have contributed to the development and application of the BJP model in seasonal forecasting..[15]

SRM forecasting model

The Snow Melt Runoff Model (SRM), developed by Martinec et al. in 2008, has been widely used in various studies related to water resource management. WAPDA (Water and Power Development Authority) obtained an Excel-based version of the SRM and utilized data from 2003 to 2010 to calibrate the model for a specific case study on Jhelum inflow to Mongla Reservoir in 2012. The model's performance was then validated against actual inflow data recorded between 2007 and 2011. Unlike the probabilistic predictions offered by the Bayesian Joint Probability (BJP) model, the SRM is a deterministic model that provides a single prediction for a given set of inputs, without considering probabilistic outcomes. Given the unreliable nature of seasonal weather forecasts in the region, an Ensemble Streamflow Prediction (ESP) approach is employed to estimate the range of potential Kharif seasonal inflows. This involves running the SRM model starting with observed snow cover in late March and using historical precipitation (P) and temperature (T) data from previous years to generate multiple Kharif seasonal scenarios. To simulate the reduction of snow cover over time, a modified depletion curve approach is incorporated, which takes into account the degree-day series of different years.

The SRM, with its deterministic nature, provides a specific prediction for water runoff based on given inputs. By incorporating the ESP approach and utilizing historical records, it enables the exploration of various scenarios and potential outcomes for the Kharif season's water influx..[15]

ARIMA MODEL

Given the importance of the Kabul River, projections of the Kabul River's flow up to 2030 were necessary.[16] Potential future changes in flow intensity of the Kabul River Basin should be analyzed to prepare for repeated natural flood vulnerability and to avoid financial losses and

casualties. The goal of this research is to use an effective learning algorithm that can accurately predict and evaluate different patterns of water levels based on different time periods. Another purpose of this analysis is to assist upstream reservoir engineers by providing better forecasting tools for forecasting expected water levels using automated ARIMA models. The goals achieved are important for responsible authorities as they help them efficiently plan socio-economic development activities to meet future needs, provide water retention structures in case of floods, and develop strategies against water disasters. It helps relief workers reduce irreparable human and economic losses.[15]

1.4. Literature Gap

Pakistan built the Warsak dam in 1960 on the Kabul River, which generates 243 MW hydropower [25] The Warsak dam, constructed on the Kabul River in Pakistan, plays a crucial role in electricity generation as it relies on the water level in the river. To forecast hydroelectric consumption in Pakistan, a 53-year dataset was utilized. The autoregressive integrated moving average (ARIMA) model with (p,d,q) values of (9,1,7) was selected as the methodology for forecasting. The results indicated an annual increase of 1.65% in hydroelectric consumption, with a cumulative rise of 23.4% expected by 2030 across Pakistan.

The water flow data of the Kabul River in Swat, a city in the Khyber Pakhtunkhwa (KP) province, were collected from 1961 to 2005 by WAPDA, a government department. Subsequently, data collection was discontinued, but a private consultant named AGES collected the data for the period of 2006 to 2010. To analyze the dataset and extract useful insights about the time series, a seasonal decomposition was performed. The analysis revealed that the highest recorded water flow value was 110 cumecs in 1991, followed by the second-highest value of 107.78 cumecs in 2005. The lowest reading of 59.97 cumecs occurred in 1982.

In this study, the water flow data of the Kabul River were collected and analyzed using the time series method with the assistance of Eviews software. Eviews is an interactive program well-suited for conducting detailed data analyses. [26]. Eviews offers a convenient forecasting capability through its Automated ARIMA feature, which saves time compared to traditional programming languages. The term "automated ARIMA" refers to the selection of the appropriate model among

AR, MA, ARMA, ARIMA, and seasonal ARIMA models. It does not imply that only the ARIMA model is considered. In this study, the ARIMA model was utilized for time series forecasting.

While there are various tools available for linear time series forecasting, ARIMA is widely recognized as one of the most suitable methods based on the existing body of knowledge. [27]. The ARIMA model combines three components: autoregressive (AR), integrated (I), and moving average (MA). Each component serves a specific purpose in capturing the patterns and characteristics of a time series. By combining these three components, the ARIMA model provides a comprehensive framework for analyzing and forecasting time series data, taking into account the relationships, stationarity, and error structure of the series. [28]. ARIMA is one of the most powerful and successful linear statistical models for time series forecasting Out of 600, 480 observations were taken as train data. Overall, 225 models were run where the best ARMA model came as (2,4)(2,2) based on AIC value, which was equal to 9.216.

A comparison was conducted between the forecasted and actual data for a period of ten years, specifically from 2000 to 2010. The dataset used for this analysis consisted of data provided by WAPDA for the first five years and data provided by AGES for the remaining five years. This dataset was also used as the test data.

Using this test data as a reference, forecasts were generated for the subsequent 20 years. Upon examination, it was observed that the actual and forecasted values closely align with each other, with only a few instances of deviation between the two.

Overall, the forecasted values demonstrated a strong level of accuracy and reliability, as they closely matched the actual values for the given time period. This suggests that the forecasting model used was effective in capturing the underlying patterns and trends in the data, allowing for accurate predictions beyond the test period.

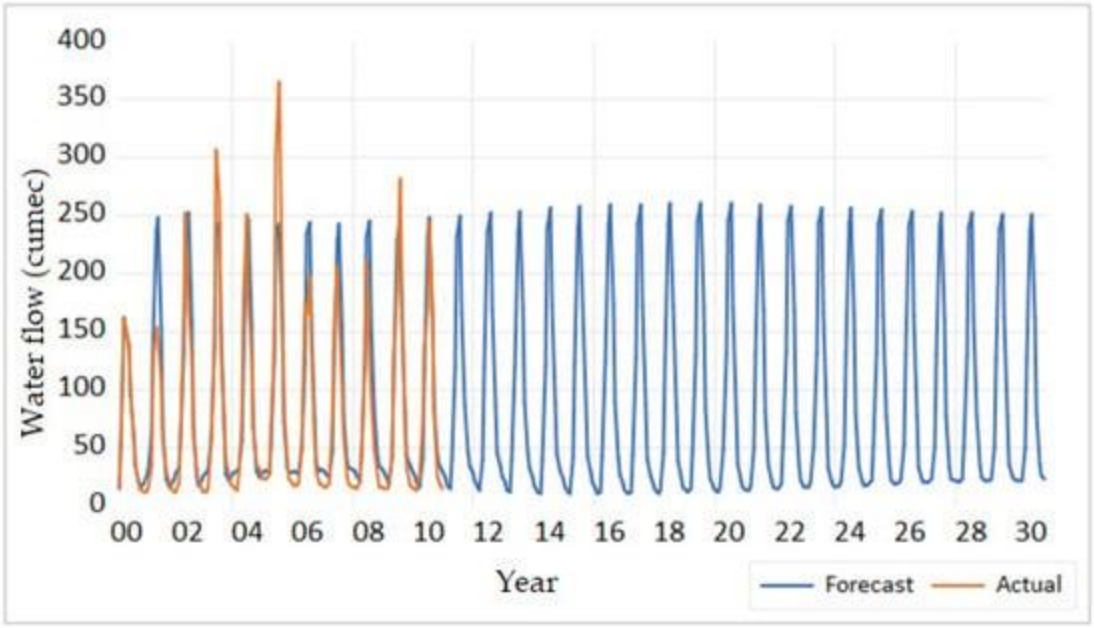

Figure 7 Comparison of Actual and Forecasted Value of Kabul River

.

In spite of that, the gap between our project Malakand-3 hydropower plant and Kabul River is as given below

- The data collected for forecasting for Kabul Rive from (2000-2010) was based only one factor i.e. discharge while ignoring certain important factors such as temperature and precipitation.
- The forecasted data collected for Kabul River was collected from the lower region of SWAT and not from the main source so there might be errors in data which would have also affected the forecasted data provided
- They used Eviews software, which is old and using old techniques for processing which proved to have low accuracy and precision.

1.5. Objectives

The purpose of this study is to investigate the effects of climate change on hydropower generation. The specific objectives are:

- Investigation of effects of climate change on hydrological sustainability of Malakand-3 HPP
- The introduction of Artificial Intelligence (AI) in predicting the streamflow in Swat River that formulates input to Malakand-3 HPP

The innovation in the research is the use of ARIMA model and neural network model for forecasting for longer time period. The introduction of AI in this field will help a lot in regulation of useful hydropower plant such as Malakand-3 hydropower plant. The result of these improved algorithms can help to better predict the hydrological sustainability of hydropower generation.

1.6. Scope of Study

A worldwide Hydro-meteorological (temperature, rainfall, humidity, domestic and irrigation uses, surface wind, thunderstorm, dust storm) alteration can affect precipitation and Temperature. The changes in precipitation and Temperature have been very uncertain and, in some cases, (2010 Event) they can be extremely intense in a few territories. Mostly the rainfall takes place from July to September (Monsoon) and the temperature approaches its extreme range both in summers and winters.

However very little research has been done on forecasting of hydropower plants for longer time period. The scope of this study is to introduce Artificial Intelligence in this field which will not only help in the forecasting but also can alert us about any destructive events in the future which will be very helpful socially and economically.

CHAPTER 2

DATA AND METHODS

2.1. Study Area

Upper swat canal system

Figure 8 Upper Swat Canal

The Swat River, which originates in the Swat Kohistan Glacier, was opened in 1885 by Raj engineers to build the Lower Swat Canal, which drains the river in the territory of the independent Dir State. However, due to the irregular topography of the land, including today's Charsada, Mardan and Swabi districts, the canal could only irrigate a small portion of the land, leaving considerable potential untapped. Punjab Irrigation Inspectorate Officer John Benton recently achieved spectacular success by making a ruling in favor of Mangla's unique terrain. As a result, the upper reaches of the Jhelum Canal flowed across the Jhelum without barrages or weirs thrown. Now the man is being asked to develop a system to channel the waters of the Swat River beyond the control of the Lower Swat Canal to parts of Mardan and Swabi. Given the topography, it was clear that the river north of the Malakand Pass would need to be used to irrigate the target area. Amandara, just outside the town of Batkhela, was chosen as a starting point after doing some research. Here the Swat River turns sharply to the right and its flow begins to hug the left bank in a narrow channel, the same situation Benton allowed at Mangla. In this narrow canal, a three-wing superstructure was sufficient to raise the water level and feed the canal through the regulator. However, just south of the planned headworks was the mountain barrier of the Malakand Pass. He proposed a tunnel to channel Swat water into the heart of the Yousafzai Plain.

As early as 1909, when headline and channel control work and channel excavation were underway at Amandara, mining engineer G.L. Bill set out to survey the shortest route for the tunnel. British mining engineers with experience in building railway tunnels across the subcontinent were familiar with the types of rock they worked with.

Over time, a fine network of smaller tributaries drained from the main channel, increasing its area. As the 20th century drew to a close, population growth and changing cropping patterns increased the need for irrigation. For this purpose, Pakistani irrigation engineers designed and built a side tunnel that runs alongside the Benton Tunnel. Located at the lower end of the Mardan district, the new outlet will operate an 81-megawatt power plant and expand the current irrigated area upstream of the Swat Canal to approximately 2.5 million acres.

Amandara head works

It is located in Amandara (Chakdara). The main purpose is to divert water from the Swat River to the USC system where many hydroelectric power plants are being built, and the main purpose of this is irrigation.

Figure 9 Amandara head works

Designed water flow is 56657 cs.

Instruments are installed in these headworks where the flows to USC and Cusec are measured. We also collected data from this observatory for our project.

Channel

Design flow i.e. Q=3657 cs

Length L = 19750 feet

The canal was originally unpaved, but was later paved to stop seepage. Construction of the canal began in 1907 and was completed in 1914. The first engineer, Sir John Benton. The research was conducted by Hr tekon. It is given from Amandara and ends at the foot of the Benton Tunnel near Akhtar Ghundai. Water is supplied to the Dargai district through the Benton Tunnel. The Benton Tunnel was innovated by NWFP Chief His Commissioner Sir Johns Staved.

The water is then split into two tunnels at the 540-foot-long Bifurcater Tunnel, which splits the water evenly.

Auxiliary tunnel

Q=1800cs

Length = 11 + 081 feet

Basically, it is the main tunnel that carries water from the upper reaches of the Swat Canal to the Malakand Sri Hydroelectric Power Station. It has a water intake or outflow rating of 1800 cs and opens with Malakand-3 channels.

The water reaches there with great pressure, and a drainage channel is installed there, where the speed of the water is slowed, the sediment settles, and in a magical way the excess water is removed, and the required amount of water is obtained. 81 MW of power is generated by pressurizing two turbines.

Auxiliary tunnels are basically built to open access to water in mountains such as the Malakand Hills.

2.1.1. DEM Of Swat River Catchment

A Digital Elevation Model (DEM) is a digital representation of the Earth's surface, showing elevation at different locations. It accurately depicts the topography by capturing height and slope variations. DEM data is obtained through remote sensing techniques, like satellite imagery or radar. Elevation values are assigned to pixels or grid cells, forming an elevation grid. Higher resolution DEMs provide more accuracy but require more storage and processing power. DEM finds applications in terrain analysis, flood modeling, slope stability assessment, viewshed analysis, and visualization. Processed DEM datasets provide slope, aspect, hillshading, and contour lines for terrain interpretation. Integration with spatial datasets enables comprehensive analysis. Overall, DEMs are essential for understanding Earth's surface and supporting various disciplines.The purpose of finding the catchment area through DEM (Digital Elevation Model) model is to accurately delineate the drainage basin of the Swat river and its tributaries. By using a

DEM model, we can obtain a detailed digital representation of the terrain and its elevation, which allows us to identify the ridges, valleys, and stream channels that make up the catchment area. This information is essential for understanding the hydrological processes that occur in the catchment area, such as surface runoff, infiltration, and groundwater recharge.

The purpose of using DEM model to find the catchment area of Swat River is to provide a spatially explicit analysis of the Swat River basin that can be used for the catchment area information to develop flood forecasting models, design water resources management plans, and evaluate the impacts of climate change on the hydrology of the region. The accuracy and precision of the catchment area information derived from a DEM model are crucial for ensuring the validity and reliability of these applications.The DEM model was used to analyze the Swat river catchment area, obtained from the Earth Data website and had a resolution of 12.5 meters. This means that the model provided elevation data for each cell or pixel in the dataset, with a spatial resolution of 12.5 meters. The DEM model used was likely derived from the Shuttle Radar Topography Mission (SRTM) dataset, which was collected by NASA's Space Shuttle Endeavour in February 2000. The SRTM data was processed to remove errors and voids caused by factors such as vegetation cover, water bodies, and urban areas, resulting in a highly accurate and reliable DEM model.

Stepwise Procedure:

We obtained our DEM model for the Swat River catchment area from the Earth Data website and used ArcGIS software to analyze the catchment area using the following methodology:

- Extracted Swat River catchment area DEM from Khyber Pakhtunkhwa (KP) dataset using ArcMap's extraction tool based on catchment area boundary.
- Used ArcMap's fill tool to remove small imperfections in elevation data by filling sinks in surface raster.
- Generated flow directions from filled DEM using flow direction tool, recording flow out of each cell in surface raster.
- Calculated flow accumulation data based on flow direction output.
- Marked pour points/outlets to create pour point data for catchment area.
- Delineated watershed boundary using raster representing flow direction.

- Defined streams in watershed to determine contributing area and converted watershed and streams to vector format.
- Extracted catchment area streams and DEM for further analysis.
- Calculated catchment area's watershed area using information from DEM model and extracted vector data.

To comprehensively analyze the hydrological processes and spatial characteristics of the Swat River catchment area, we employed a combination of extraction, fill, flow direction, flow accumulation, delineation, and vector conversion tools within ArcGIS software. This approach ensured a remarkably accurate and precise assessment of the catchment area.

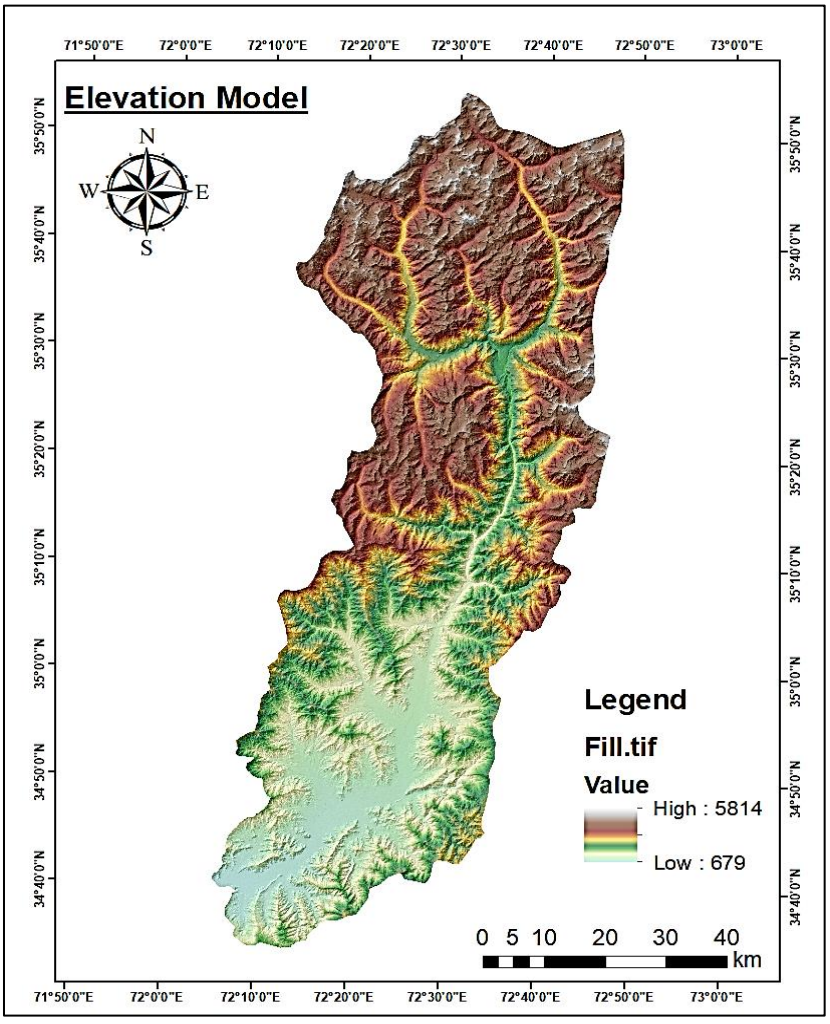

Figure 10 Elevation Model OF Swat River

The elevation model of the Swat River catchment was generated using ARCGIS, providing valuable information about the vertical distribution of land surface within the study area. This model allows for the identification of high and low elevation areas, aiding in the understanding of the topographic characteristics and potential hydrological processes within the catchment.

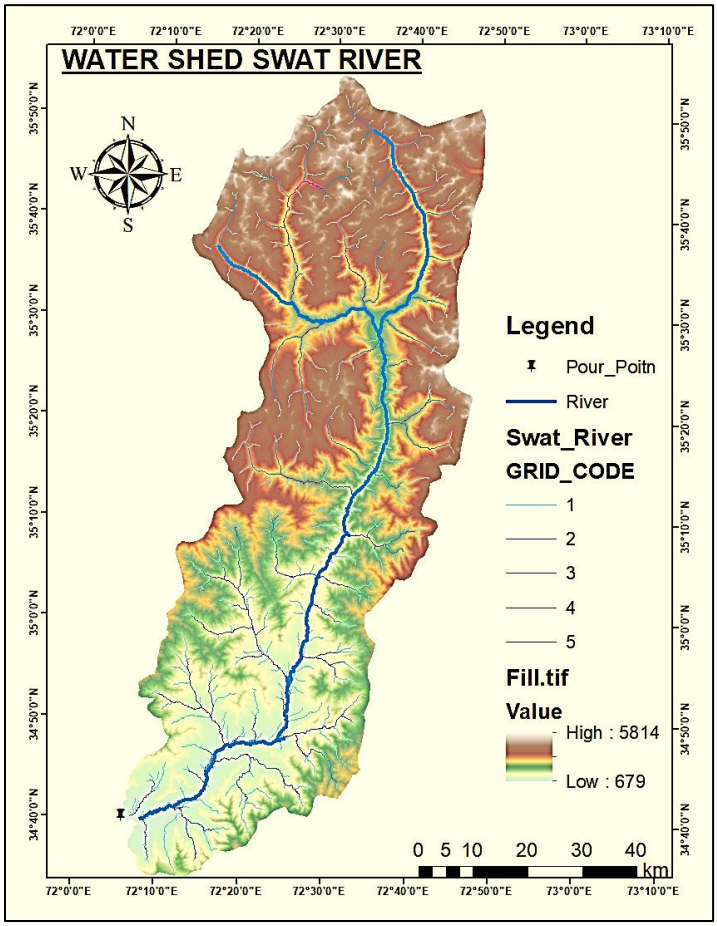

Figure 11 Water Shed of Swat River

The watershed model of Swat River, created using ARCGIS, delineates the boundaries of the drainage area contributing to the river. It provides a spatial representation of the land surface features that contribute surface runoff to the river network. This model facilitates the identification

of sub-catchments and their corresponding hydrological characteristics, enabling a more detailed analysis of the water flow patterns and basin-scale hydrological processes.

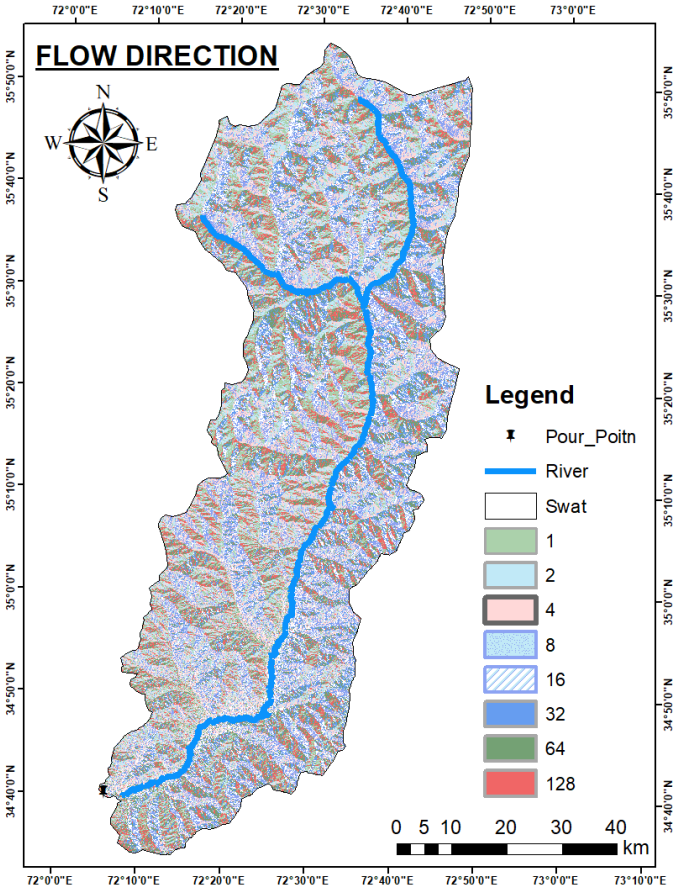

Figure 12 Flow Direction Of Swat River

The flow direction model generated in ARCGIS depicts the pathways of water flow within the Swat River catchment. It identifies the direction of surface runoff based on the topographic slope, assisting in understanding the connectivity and routing of water through the landscape. This model serves as a fundamental component for hydrological analysis, enabling the identification of flow accumulation areas and contributing to the accurate simulation of water movement within the catchment.

2.2. Data Collection

2.2.1. ERA5- Land

In this study, the temperature and precipitation data were obtained from the ERA5-Land dataset. ERA5-Land, provided by the European Centre for Medium-Range Weather Forecasts (ECMWF), encompasses a high-resolution, global coverage of land surface variables, making it a valuable resource for the research conducted.

By accessing the temperature and precipitation data from ERA5-Land, the climatic conditions and their variability over the study area could be investigated. The dataset provided reliable and consistent information, derived from a combination of satellite observations, ground-based measurements, and numerical weather prediction models.

Subsequently, the temperature and precipitation data were utilized in the hydrological model. These datasets were integrated into the model, contributing to a more comprehensive analysis of the hydrological processes in the study area.

By incorporating the temperature and precipitation data from ERA5-Land into the hydrological model, a more accurate representation of the climatic conditions and their impacts on the hydrological system was achieved. The integration of these datasets enhanced the reliability of the model's simulations and the robustness of the study's findings.

The utilization of data from ERA5-Land in the hydrological model strengthens the scientific basis of this study, allowing for a more informed understanding of the complex interactions between climate and hydrology in the study area.

2.2.2. Observed Data

We are working on a forecasting project. Data is very important in our project. We visited various sites and collected data from them.

The research method consisted of a field survey followed by an interview.

Note that the Swat River is the main water source to the Malakand-3 Hydropower Station. Upstream of his three hydropower plants in Malakand are the Swat River and the USC in the Batkhela district of Malakand. Data on the Swat River

The Sanskrit name means "clear blue water".

The water in the river is very clear. The Swat River's main source lies in the Hindu Kush Mountains and is fed year-round by glaciers. This river rises from the high valley of Swat Kohistan where the Usho and Gavral (also known as Utral) rivers meet at a column. From its confluence, the Swat River flows through the narrow canyons of the Kalam Valley to the city of Madhyan. From there, the river gently flows for 160 km through the flat areas of the Lower Swat Valley to Chakdara. At the southernmost tip of the Swat Valley, the river enters a narrow gorge and joins the Panjikora River at Karangi before entering the Peshawar Valley. Finally, we end at the Kabul River near Charsadda. River flow is measured at various points. A head unit is installed for both flow control and flow measurement. Data were collected from the Department of Irrigation for the width discharge at the Chakdara Bridge.

Swat water is used for hydroelectric power generation at the Jabang Hydroelectric Power Station (completed in 1938) and the Dargai Hydroelectric Power Station (completed in 1952). Malakand-3 Hydro Power Station (completed in 2008). The Momand Dam, at the lower end of the Swat River before entering the Peshawar Valley, has a capacity of 740 MW and is currently under construction.

Sources:

1. Surveys conducted by group members between November 1, 2022 and November 5, 2022

2. Interview with SDO Nizam, Department of Irrigation

3. Track the Swat River online – Wikipedia

2.2.3. Layout Plans

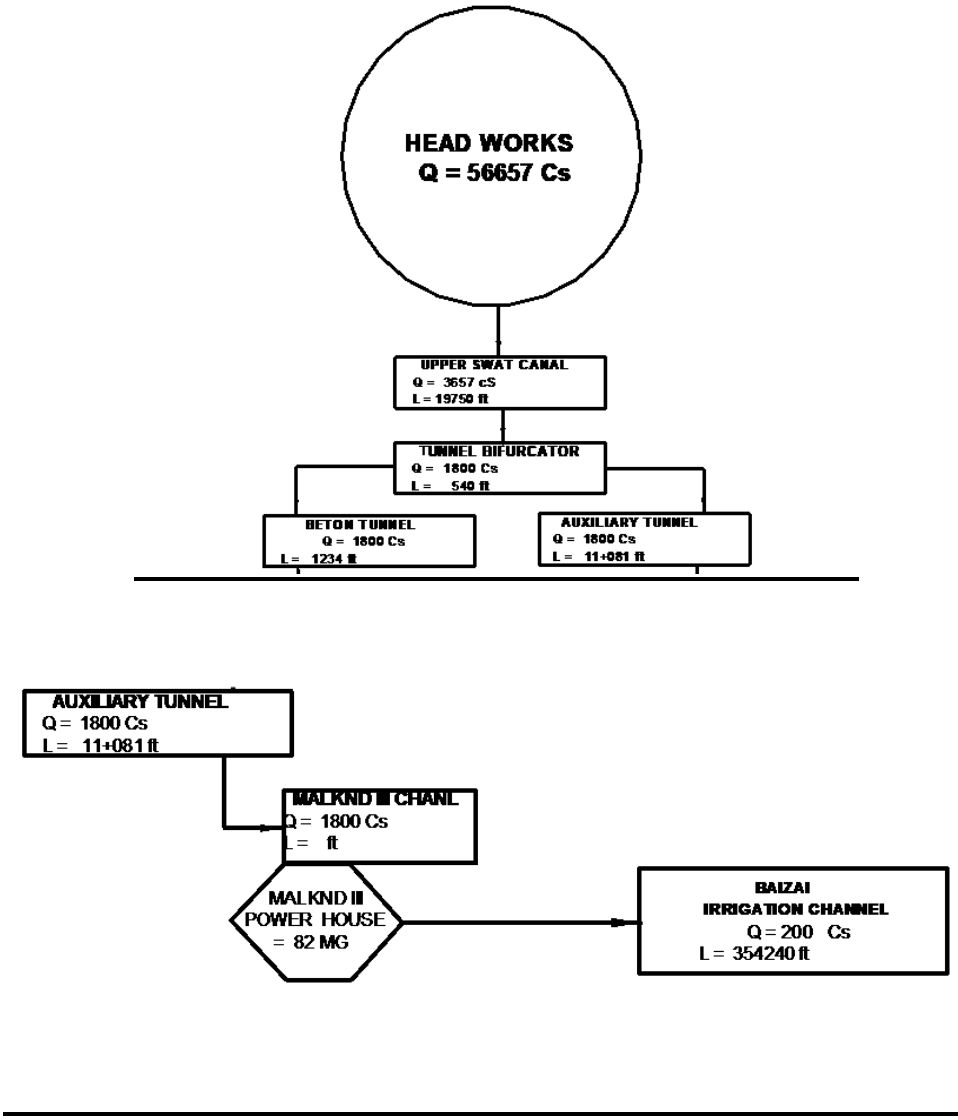

Figure 13 Flow chart of layout

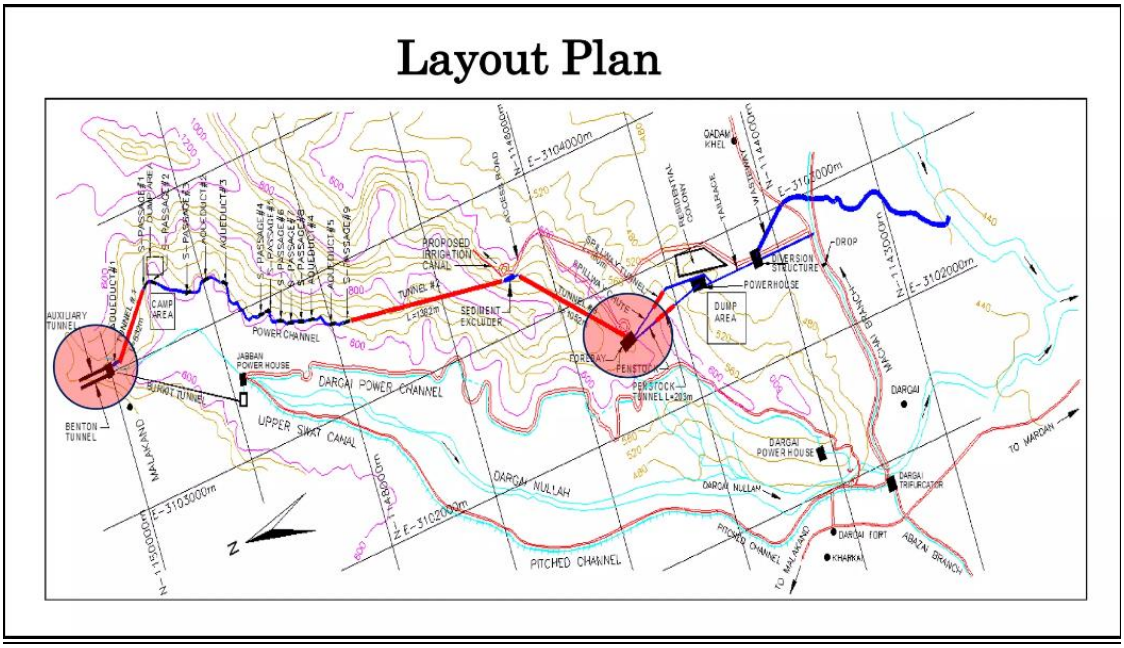

Figure 14 Contour layout

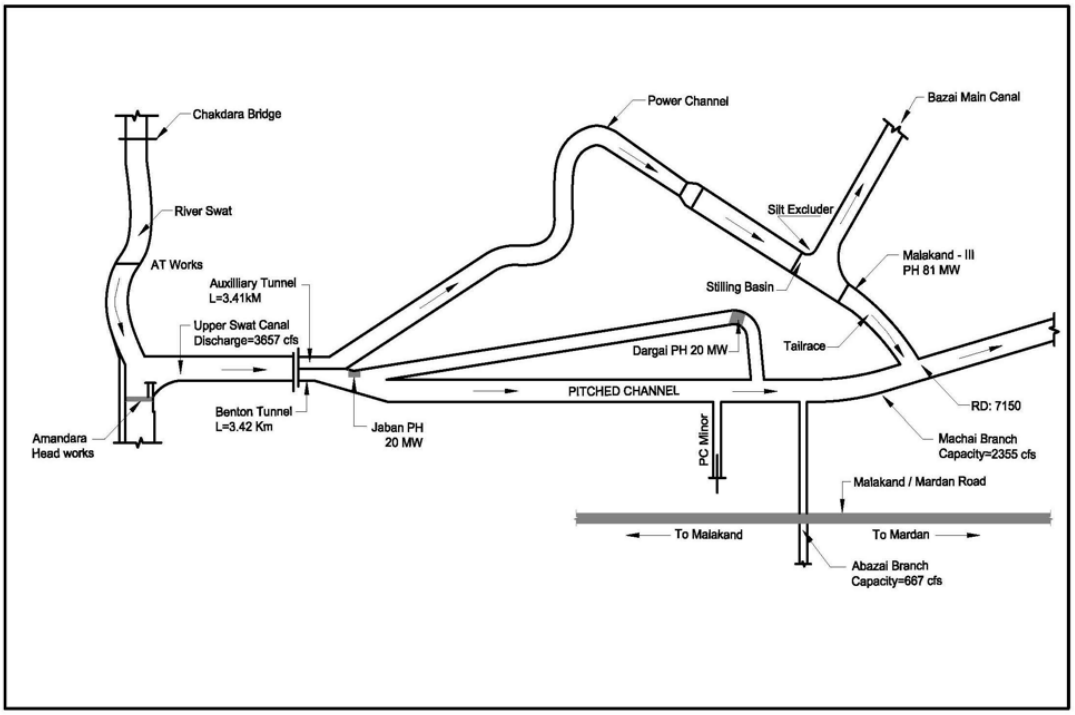

Figure 15 General layout

2.3. Preprocessing of data

After the data was obtained, the process of digitizing it in Excel format was initiated since the original data was in written form. The focus of the analysis was on the Swat River discharge, as it is the primary source of power generation at the Malakand-3 hydropower plant. Any changes in the Swat River discharge have a direct impact on the power generation at the plant.

During the data preprocessing stage, thorough efforts were made to address the visible errors present in the handwritten data. These errors were carefully corrected to ensure the accuracy of the data. Additionally, missing data points were encountered, and assumptions were made based on the surrounding data to fill in these gaps. This process resulted in a complete and reliable dataset, which served as the basis for further analysis and forecasting.

The data preprocessing step played a crucial role in ensuring the quality and reliability of the dataset used in the analysis. By meticulously addressing errors and filling in missing data, biases and inaccuracies in the results were minimized. This comprehensive approach to data preprocessing establishes a strong foundation for robust and reliable forecasting of Swat River discharge, ultimately contributing to accurate predictions of power generation at the Malakand-3 hydropower plant.

2.4. ERA5-Land Data:

The fifth-generation reanalysis data (ERA5-Land) released by the European Centre for Medium-Range Weather Forecasts (ECMWF) is a global dataset that provides high-resolution (a spatial resolution of ~10 km) information on land surface conditions covering a wide range of variables related to land processes such as temperature, precipitation, soil moisture, humidity, etc. ERA5-Land is a faction of state-of-the-art fifth-generation reanalysis of ECMWF namely ERA5.

ERA5-Land data is archived at Climate Data Store (CDS) of Copernicus (https://cds.climate.copernicus.eu/#!/home). For this study, monthly averages of temperature and precipitation for ERA5-Land were obtained from CDS. The data, in its origin, is a continuous gridded data that spanned over Pakistan for the period of 1950-2022.

Reading and Processing ERA5-Land data:

ERA5-Land monthly averages for precipitation and temperature, obtained from CDS, processed to get an R-compatible format (i.e., ".Rdata") and were trimmed for Kabul River Basin.

To run hydrological model to simulate surface runoff, the precipitation and temperature datasets of ERA5-Land were given input (as a "ptq" file) to HBV Light. Similarly, "EVAP" file was prepared by processing ERA5-Land temperature data into potential ET data using "thornwaite" package of R-statistical language.

CORDEX Climate Projections:

Coordinated Regional Downscaling Experiment, known as CORDEX (https://cordex.org/), archives climate projections data for different spatial segments (domains) of the globe such as "South Asia", "Middle East North Africa", etc. Climate data projections are available for various regional circulation models (RCMs) while the data for these RCMs is downscaled from Global Circulation Models (GCMs) having relatively coarser spatial resolution. This study utilized climate projections (for RCP 8.5) obtained from RCMs having mean resolution of ~0.44 (i.e., ~45 km) over Middle East North Africa (MENA) domain.

Reading and Processing CORDEX Climate Projections:

The climate projections data (originally in ".nc" format) over MENA was processed to R-compatible format and trimmed for Kabul River Basin. The data contained daily estimates for two parameters (total precipitation and 2m-temperature) for the period 2023-2100. Later, the daily precipitation and temperature data was converted into monthly data to make it compatible for HBV Light.

A "ptq" file to be given as input to HBV Light was prepared using climate projections for precipitation and temperature along with "EVAP" file generated by using temperature data.

2.5. Hydrological modeling

HBV light is the new version of the HBV model developed by the Swedish Meteorological and Hydrological Institute (SMHI). It provides an easy-to-use Windows version for research and

education. The basic equations for HBV light are in accordance with the SMHI-version HBV-6, however, instead of using initial states, the new version uses a warming-up period. In the original version, only integer values are allowed for the routing parameter MAXBAS. This limitation has been removed in the new version. To keep the program as simple as possible, several functions found in the HBV-6 software were not implemented in the HBV light software. The HBV-light version provides two options which do not exist in the HBV-6 version. The first one is the possibility to include observed groundwater levels in the analysis and the second is the possibility to use a different response routine with a delay parameter.

The compulsory input files required for runoff simulations are:

1. The PTQ-file (ptq.txt) which contains time series of daily precipitation [mm/day], temperature [^{o}C] and the observed runoff [mm/time].

2. The evaporation-file (EVAP.txt) which contains values for the potential evaporation [mm/time].

The total period for simulation is required to be split into calibration and validation periods and saving the results for both the periods is optional. For each of the two periods, the detailed results-files contain data for input, as well as the resulting output variables; the output variables include the time series of simulated runoff (Qsim), simulated rain (Qsim_rain), and simulated snow (Qsim_snow), among the other variables. Whereas the summary (of the results) files contain:

1. Annual water balance (mm/year), which includes the sums of simulated runoff, observed runoff, precipitation, and other variables like evapotranspiration.

2. Goodness of fit, which includes coefficient of determination, model efficiency or Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), mean difference etc.

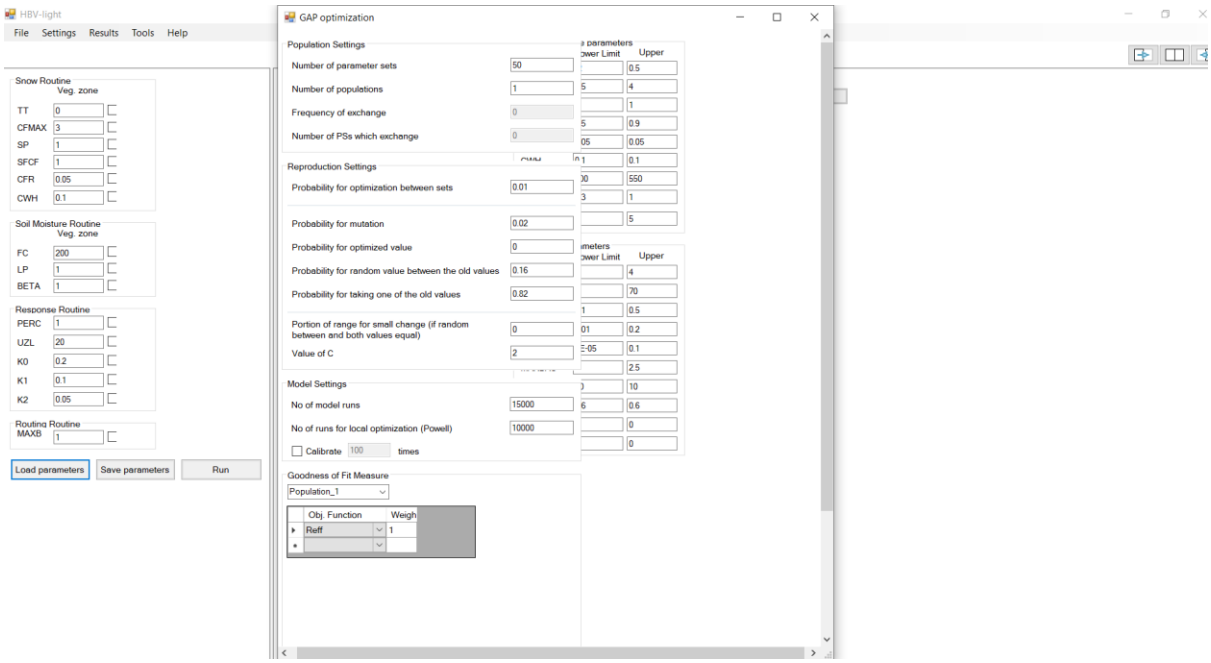

Figure 16 Calibration: Parameter optimization, ss1

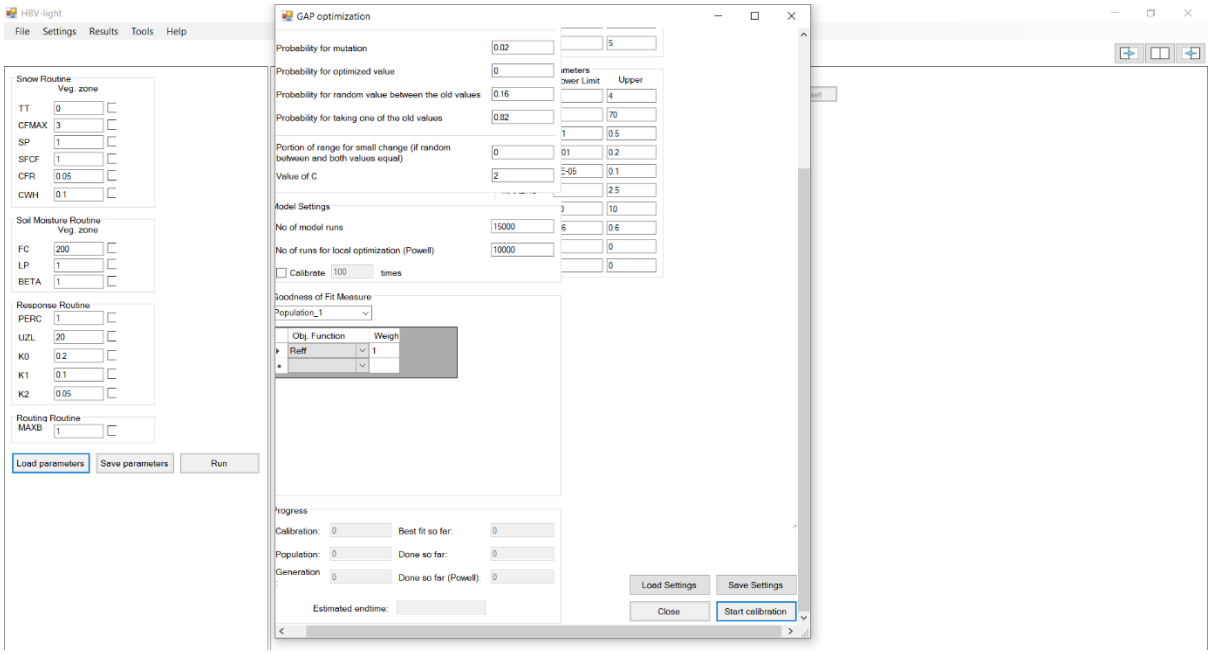

Figure 17 Calibration: Parameter optimization, ss2

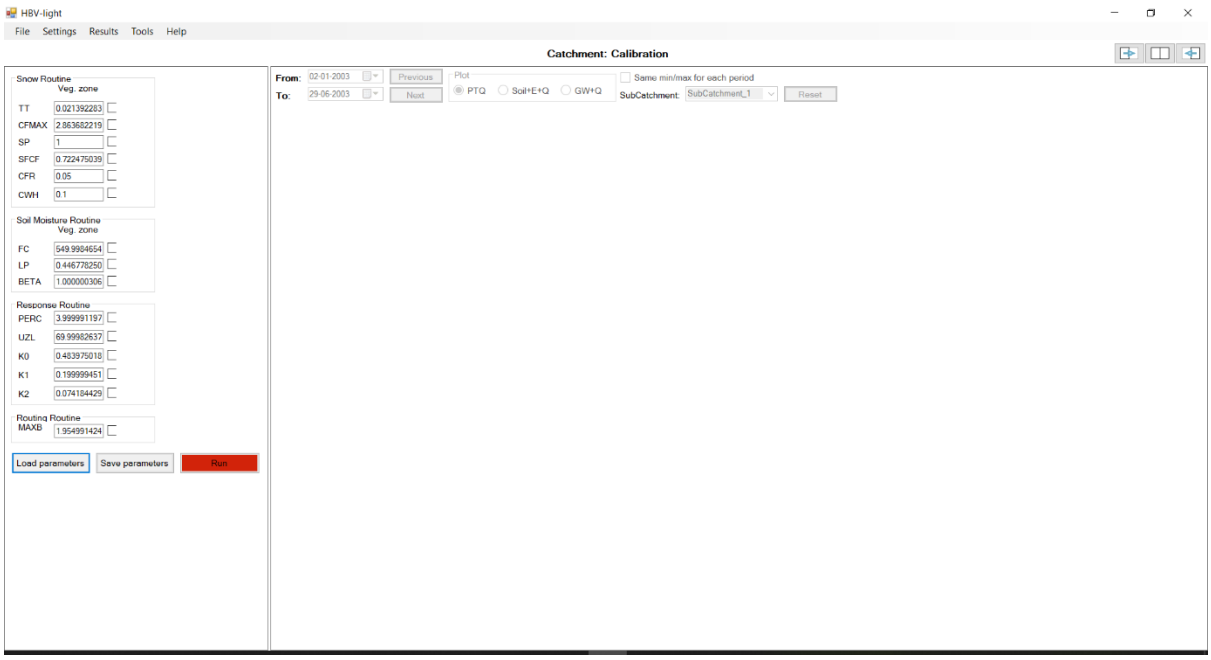

Figure 18 Calibration: Parameter loading

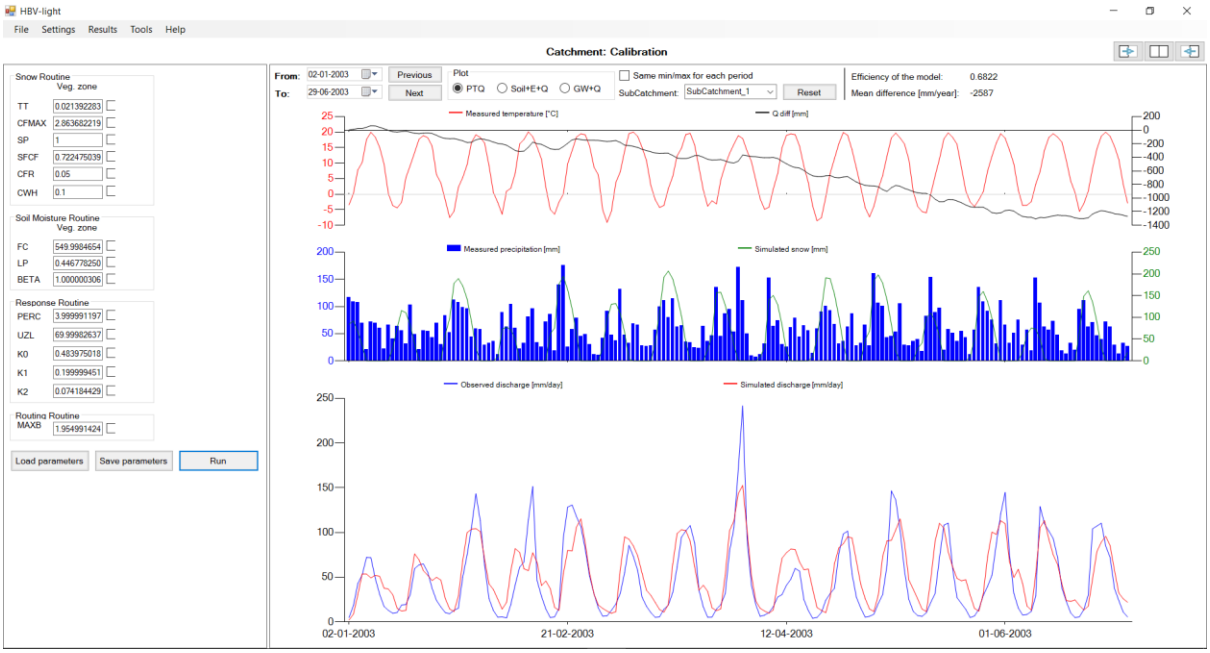

Figure 19 Calibration: Model run

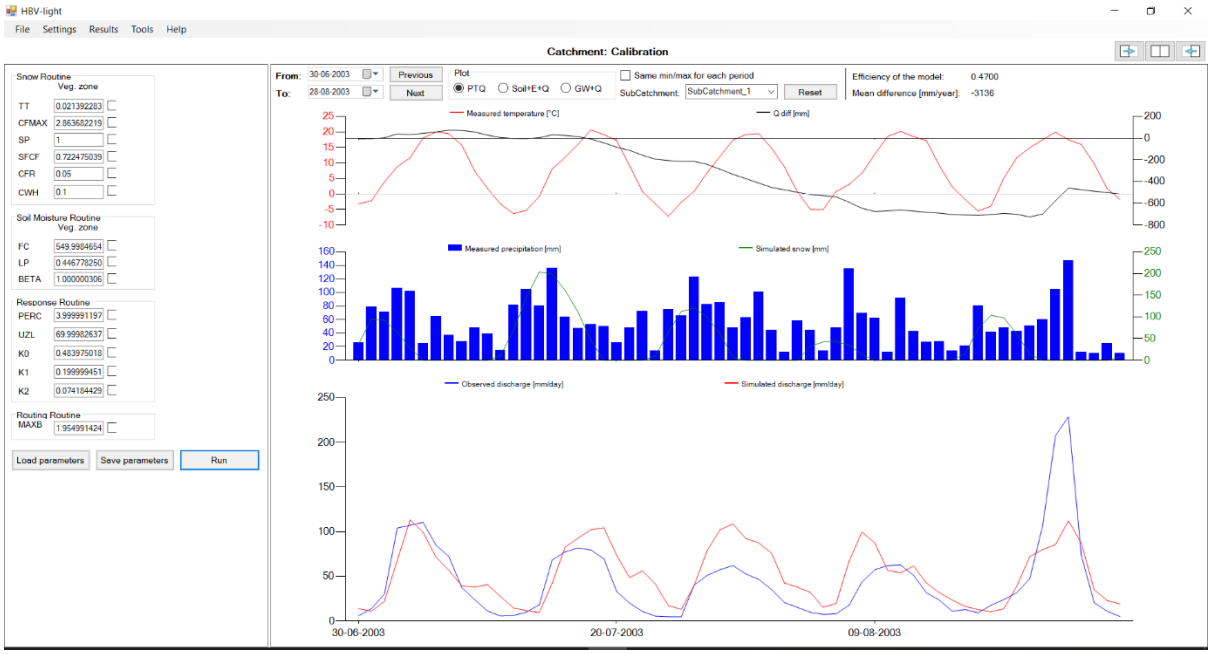

Figure 20 Validation: Model run

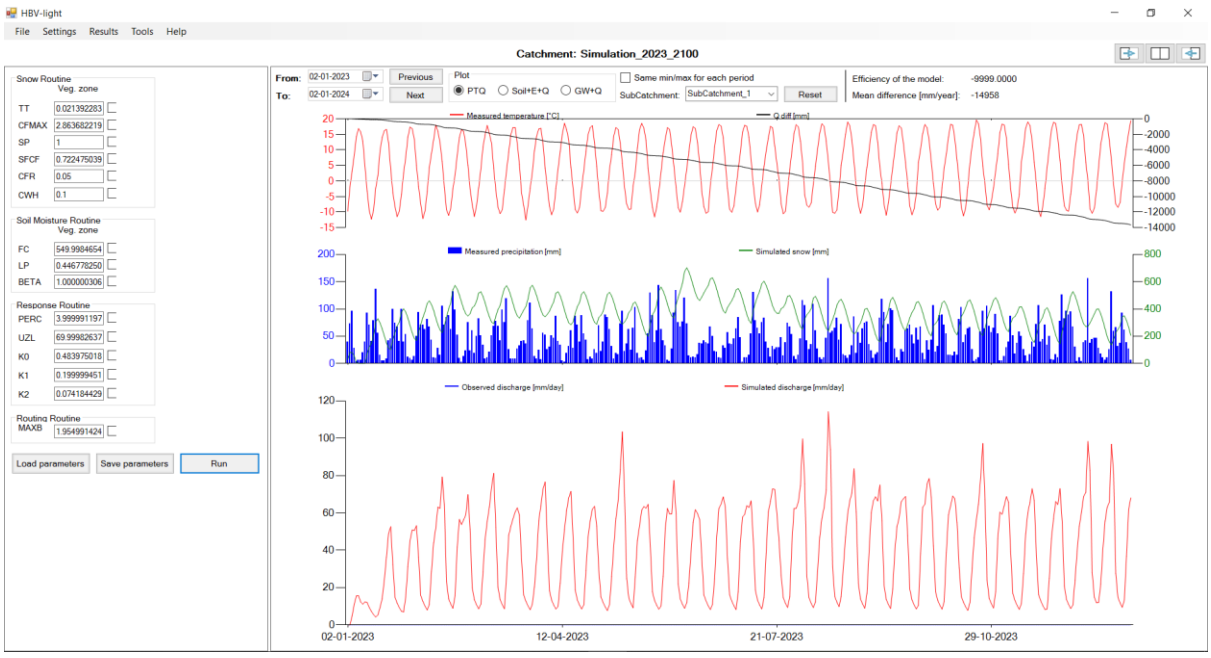

Figure 21 Simulation: Model run

2.6. Analysis and Forecasting

2.5.1. ARIMA Forecasting

The methodology employed in this study involved the following steps for the analysis of the daily discharge of SWAT River:

- Data Collection: Daily discharge data of SWAT River spanning from 2010 to 2018 was collected. Additionally, monthly discharge data from 1950 to 2018 obtained from the Hydrological model was also included in the analysis.
- Preliminary Analysis: The collected data underwent preliminary analysis to understand its characteristics. Time series plots were generated using the 'autoplot' function from the 'fpp2' package in R. Seasonal plots and sub-series plots were utilized to examine the presence of seasonal patterns in the data. Decomposition using the 'decompose' function facilitated the examination of the trend, seasonal, and residual components.
- Data Transformation: To mitigate the influence of trends and seasonality, the anomalies of the daily discharge time series data were calculated by taking the first difference. The resulting time series of anomalies was assessed for stationarity.
- Model Selection: An appropriate ARIMA model was selected using an automated selection procedure based on artificial intelligence (AI). The 'auto.arima' function was employed, which determined the optimal ARIMA model by considering the AIC and BIC criteria. The AI algorithm autonomously determined the order of differencing as 1 for the daily discharge time series data and seasonal differencing as 1 for the monthly discharge data. Autocorrelation of the model residuals was examined using the 'checkresiduals' function.
- Forecasting: The chosen ARIMA model was utilized to generate forecasts for the daily and monthly discharge time series data. Forecasts were produced using the 'forecast' function, specifying the desired number of forecast periods. The resulting forecasts were visualized through the 'autoplot' function. Additionally, the forecast results were saved in CSV and Excel formats using the 'write.csv' and 'write.xlsx' functions.
- Interpretation: The forecast results were interpreted to gain insights into the future behavior of SWAT river discharge. Visualizations of the forecasted data aided in comprehending

the anticipated changes in discharge over time. Furthermore, the 80% probability range of high discharge values was depicted to facilitate an understanding of potential risks associated with such occurrences.

ARIMA Models selected by AI:

For the daily discharge data of the SWAT river, the AI-based selection process identified the ***ARIMA(4,1,1)(0,1,0)[365]*** model as the most suitable. This model captures the underlying patterns and dynamics in the daily discharge series. The selected model has a seasonal component with a periodicity of 365 days, indicating the presence of annual seasonality in the data.

The ARIMA(4,1,1) component of the model refers to the autoregressive (AR) and moving average (MA) components. The autoregressive term (p=4) takes into account the relationship between the current discharge value and the previous four values, considering the lagged effect. The moving average term (q=1) captures the influence of the residual errors on the current discharge value.

The (0,1,0) component represents the differencing order, denoting that the model includes a first-order difference to make the series stationary. This differencing step eliminates the trend present in the original series, making it suitable for modeling.

For the monthly discharge data of the hydrological model, the AI-based selection process identified the ***ARIMA(2,1,2) (0,1,2)([12]*** model as the most appropriate. This model is specifically tailored to capture the patterns and characteristics exhibited in the monthly discharge series.

The ARIMA(0,1,2) component of the model indicates that it includes an autoregressive moving average term with two lagged values. This implies that the current monthly discharge is influenced by the residual errors and the two previous monthly discharge values.

The seasonal component (2,1,2) denotes the presence of seasonal patterns with a periodicity of 12 months, representing the annual cycle in the discharge data. This component accounts for the seasonal variation and captures the effects of the past two seasonal differences and the residual errors on the current monthly discharge value.

The (0,1,0) component indicates the first-order differencing applied to the series to remove the trend and make it stationary for modeling purposes.

Main Coding

```
#load the forecasting package
library(fpp2)
library(ggplot2)

#load the data

# Declare this as time series data
Y <- ts(mydata[,2], start = c(2010, 1), end = c(2018, 365), frequency = 365)

#data has a strong trend. Investigate transformation
#take the first difference to remove the trend.
DY <- diff(Y)

#Series appears trend-stationary, use to investigate seasonality


fit_add = decompose(Y,type = 'additive')
plot(fit_add)



################################################
#fit a ARIMA model#
################################################

fit_arima <- auto.arima(Y,d=1,D=1,stepwise = TRUE, approximation = TRUE,trace = TRUE)
print(summary(fit_arima))
checkresiduals(fit_arima, type="correlation")



################################################
################################################
#forecast with ARIMA model#
################################################
fcst <- forecast(fit_arima,h=7300)
autoplot(fcst,include=360)
print(summary(fcst))

write.csv(fcst, file = "fcst.csv", row.names = TRUE)
install.packages("xlsx")
library(xlsx)
write.xlsx(fcst, file = "forecast_results.xlsx", sheetName = "Sheet1", row.names = TRUE)

################################################
#now we will run this model for reading which we got from Hydrological model and see thr result
################################################
#load the data

# Declare this as time series data
SY <- ts(HBook1[,1], start = c(1950, 2), end = c(2018, 12), frequency = 12)

"""""""""""""""""""""""""""""""""""""""""""""""""""""""""
```

```
#data has a strong trend. Investigate transformation
#take the first difference to remove the trend.
DSY <- diff(SY)

#Series appears trend-stationary, use to investigate seasonality

fitS_add = decompose(SY,type = 'additive')
plot(fitS_add)

################################################
#fit a ARIMA model#
################################################

fit_arimaSH <- auto.arima(SY,d=1,D=1,stepwise = TRUE, approximation = TRUE,trace = TRUE)
print(summary(fit_arimaSH))
checkresiduals(fit_arimaSH)

################################################
#forecast with ARIMA model#
################################################
fcstSH <- forecast(fit_arimaSH,h=240)
autoplot(fcstSH,include=12)
print(summary(fcstSH))

write.csv(fcstSH, file = "fcstH.csv", row.names = TRUE)
install.packages("xlsx")
library(xlsx)
write.xlsx(fcstSH, file = "forecast_resultsSH.xlsx", sheetName = "Sheet1", row.names = TRUE)
```

2.5.2. Neural Network Forecasting

The methodology employed in this study involved the following steps for the analysis of the daily discharge of SWAT River:

- Load data from Excel sheet using the readtable function which reads the data into a table variable. The data was stored in a table variable for easy manipulation.
- Convert the date column to a datetime data type. The date column was in string format in the original data file, so it was necessary to convert it to a datetime data type using the datetime() function. This makes it easier to work with the dates and allows to perform time-series analysis.
- Split data into training and testing sets in order to evaluate the performance of our model. the first 3200 rows of data was used as the training set, and the remaining data as the testing set.
- Sort the table by the date column using the sortrows function.
- Extract input and output variables from the training data. The inputs variable is a vector of years (already extracted from the date column), and the targets variable is a vector of river discharge values.

- A feedforward neural network model with one hidden layer was created using the fitnet() function. The number of neurons in the hidden layer was set to 10 through trial and error. The neural network was designed to predict future river discharge values based on the input variables.
- The neural network model was trained using the train() function, which utilized back propagation to update the weights of the network. This process aimed to minimize the difference between the predicted and actual output values.
- Predictions for the next 20 years were made using the trained model. A for loop was implemented to generate predictions for each year, and the results were stored in a vector.
- To visualize the performance of the model, the actual and predicted river discharge values were plotted using the plot() function. Additionally, the river discharge values from the testing set were included for comparison.
- A seasonal plot of the actual data was created using the subplot() function. This plot illustrated the river discharge values for each month across the entire dataset.
- Residuals were calculated by subtracting the predicted values from the actual values to assess the model's performance. The autocorrelation of the residuals was determined using the autocorr() function, and the results were plotted. This analysis aimed to evaluate the correlation and normal distribution of the residuals.

In summary, future river discharge values were predicted through a combination of data loading, manipulation, neural network modeling, and data visualization techniques. The performance of the model was evaluated using various methods, including residual analysis and comparison with other models like ARIMA. The results demonstrated the neural network's ability to accurately forecast river discharge values, enabling informed decision-making for the power plant's potential and flood prevention measures.

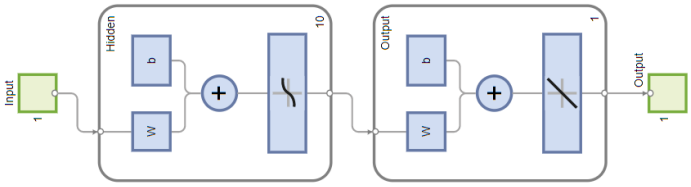

Figure 22 Neural Network Model

Main Coding

```
% Load data from Excel sheet using the readtable function which
% reads the data into a table variable
data = readtable('data.xlsx');


%converts the date column to a datetime data type
data.Date = datetime(data.Date, 'InputFormat', 'MM/dd/yyyy');


% Split data into training set
train_data = data(1:3200, :);
test_data = data(3201:end, :);

%sorts the table by the date column using the sortrows function.
data = sortrows(data, 'Date');
test_inputs = year(test_data.Date)';
```

```
% Extract input and output variables from the training data
% The inputs variable is a vector of years (already extracted from the date column),
% and the targets variable is a vector of river discharge values.
% inputs extracts the year values from the Date column in train_data and transposes
% the resulting row vector to a column vector using the transpose operator '
inputs = year(train_data.Date)';

% extracts the Discharge column from train_data and transposes
% the resulting row vector to a column vector using the transpose operator '
targets = train_data.Discharge';



% Create neural network model using the fitnet function with a specified number of hidden layers
% here its 10
hidden_layer_size = 10;
%creates a feedforward neural network with one hidden layer using the fitnet() function.
net = fitnet(hidden_layer_size);



% Train neural network model  using the train function with the input and output
net = train(net, inputs, targets);
```

```
% Use trained model to make predictions for next 20 years
%here we use the trained neural network model to make predictions for the next 20 years starting from today
%sets the start date for making predictions to the current date and time
start_date = datetime('today');
%calculates the end date for making predictions 20 years into the future using the calyears() function.

end_date = start_date + calyears(20);

%calculates the number of years to predict and generates a vector of dates corresponding to each year
num_years = year(end_date) - year(start_date) + 1;
dates = datetime(year(start_date):year(end_date), 1, 1);



%for loop to make a prediction for each year and stores the predictions in a vector
predictions = zeros(1, num_years);
for i = 1:num_years
    input = year(dates(i));
    prediction = net(input);
    predictions(i) = prediction;
end
```

```
% Plot actual and predicted river discharge values
figure;
%plots the actual and predicted river discharge values on a graph using the plot function
plot(data.Date, data.Discharge, 'b', 'LineWidth', 1);


hold on;%hold on command is used to allow multiple plots on the same figure
plot(test_data.Date, test_data.Discharge, 'g', 'LineWidth', 1);
plot(dates, predictions, 'r', 'LineWidth', 1);
legend('Actual','Test', 'Predicted');
xlabel('Years');
ylabel('River Discharge');
title('Forecasted Data');
% Plot seasonal plot of actual data
figure;
for i = 1:12
    %subplot(3, 4, i);
    idx = month(data.Date) == i;
    plot(year(data.Date(idx)), data.Discharge(idx), 'g','LineWidth', 2);
    title('Seasional Plot');
    xlabel('Years');
    ylabel('River Discharge');
end
```

```
% Calculate and plot residuals
%calculates the residuals by subtracting the predicted values from the actual values, using the test_data set.
% The min function is used to make sure that the lengths of the test_data.
% Discharge and predictions arrays are the same before subtracting them
residuals = test_data.Discharge(1:min(length(test_data.Discharge),length(predictions)))-predictions(1:min(length(
% residuals array is then transposed to a row vector
residuals = residuals(:)';


figure;
%The autocorr function calculates the autocorrelation of the residuals and plots the results.
autocorr(residuals);
xlabel('Lag');
ylabel('Autocorrelation');
title('Residuals');
```

CHAPTER 3

RESULTS AND DISCUSSION

3.1. Primary Analysis

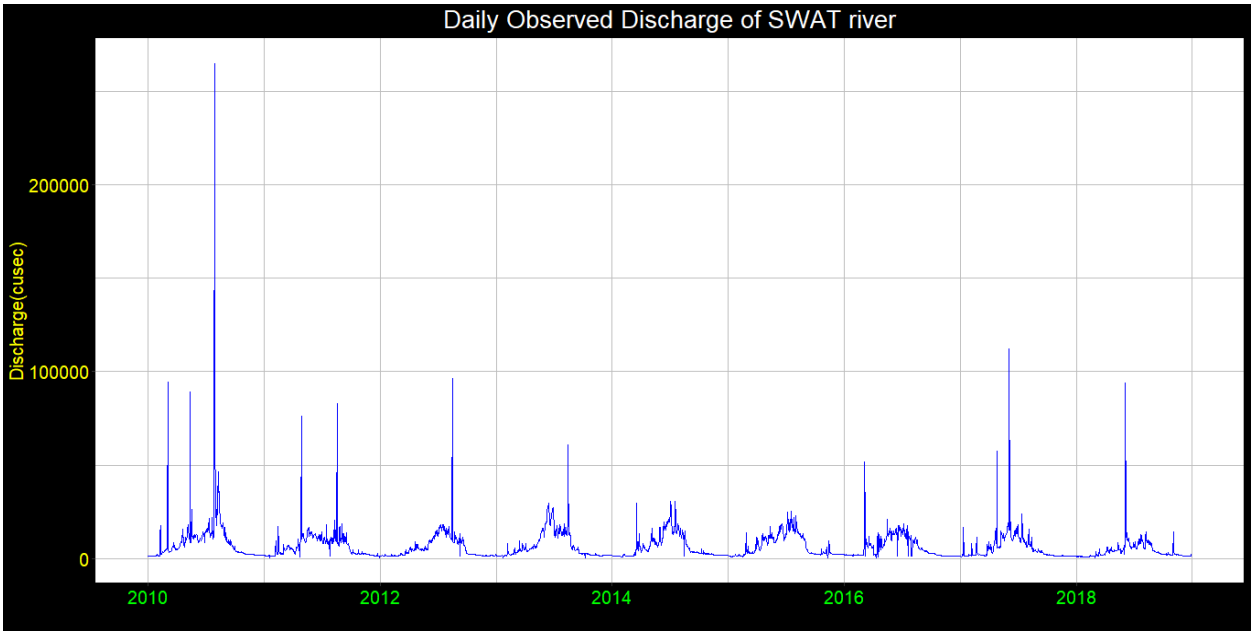

Figure 23 Daily Observed Discharge of SWAT River

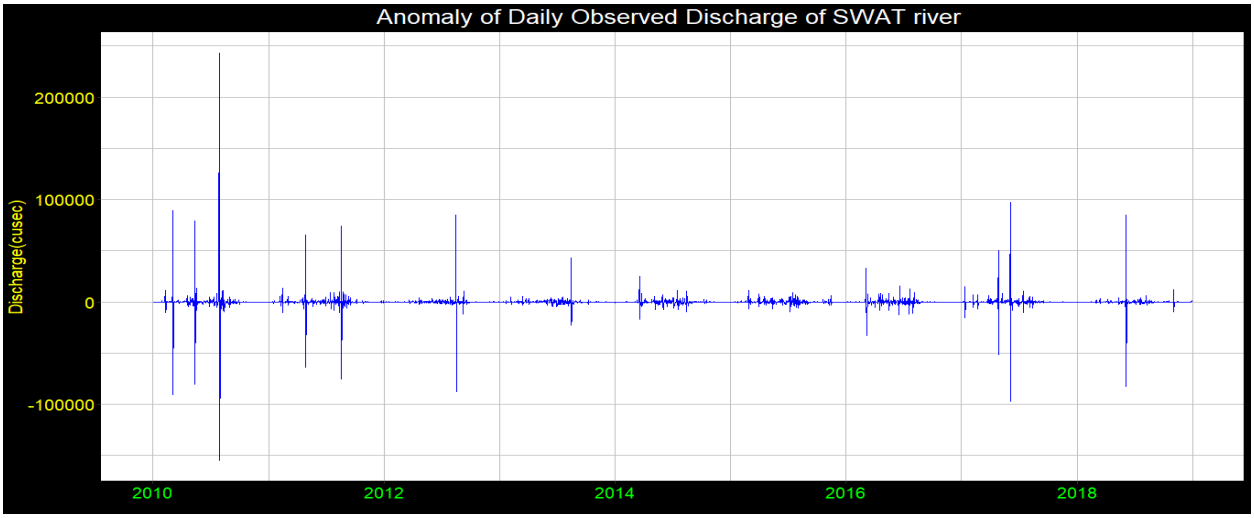

Figure 24 Anomaly of Daily Discharge of SWAT river

In the period spanning from 2010 to 2018, the daily observed discharge data of the Swat River exhibited a significant anomaly in 2010, marked by a notably high reading. This exceptional spike in discharge is indicative of a severe flood event that occurred during that year. Subsequently, the remaining years displayed a consistent pattern of discharge, demonstrating well-defined trends. These regular trends in the data reflect the typical behavior of the river's flow, undisturbed by any extraordinary events. This observation further underscores the significance of the flood event in 2010, as it stands out as an exceptional occurrence amidst an otherwise predictable discharge pattern. The findings from this graph provide valuable insights into the hydrological dynamics of the Swat River and contribute to a comprehensive understanding of its behavior during the studied timeframe.

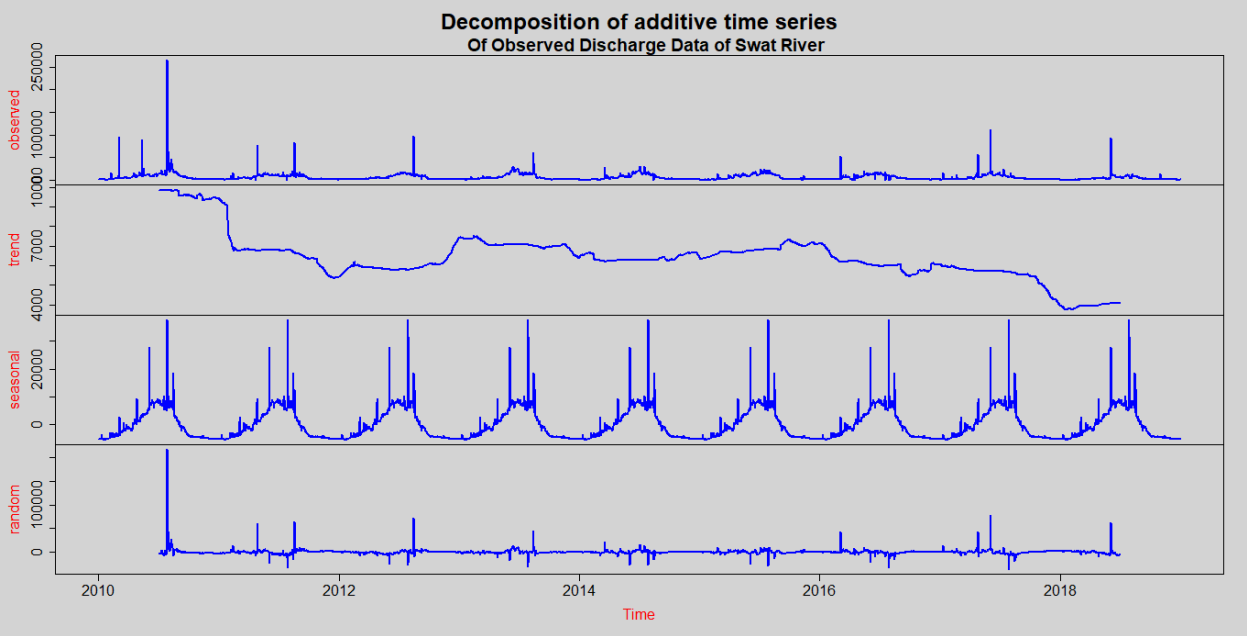

Figure 25 Decomposition of additive time series

The decomposition analysis of the upper data provides a comprehensive understanding of its underlying components, including trend and seasonality, among other factors. The trend component reveals a clear pattern in the data, indicating a notable decrease in discharge following the flood event in 2010. This decline suggests a gradual recovery phase for the Swat River after the devastating flood. However, an intriguing observation arises from the data post-2018, as it

exhibits a tendency towards an upward trend once again. This potential resurgence in discharge levels implies the need for continued monitoring and analysis to comprehend the underlying factors influencing this change.

Examining the seasonality component, it becomes evident that the discharge levels exhibit a distinct cyclic pattern throughout the studied period. The data highlights that the mid-year months consistently experience higher discharge compared to the beginning and end of each month. This recurring pattern of heightened discharge during the mid-year months indicates the presence of seasonal influences on the river's flow dynamics. The identification of this seasonality contributes to a more comprehensive understanding of the river's behavior and assists in future predictions and management of water resources.

3.2. Model Selection

Interpreting an ACF graph involves examining the correlation values at different lags. The ACF value at a particular lag provides information about the strength and nature of the relationship between the current observation and the observation at that lag. A positive ACF value indicates a positive correlation, meaning that as the current observation increases, the lagged observation tends to increase as well. Conversely, a negative ACF value suggests a negative correlation, indicating that as the current observation increases, the lagged observation tends to decrease.

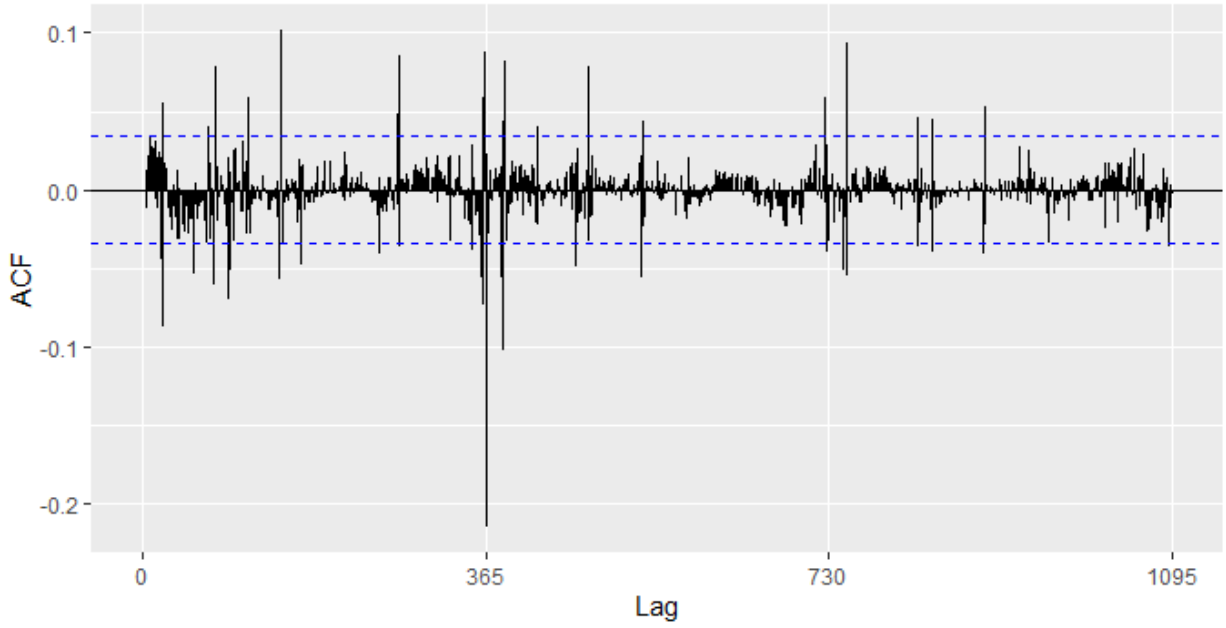

Figure 26 AutoCoorelation Faactor of ARIMA(4,1,1)(0.1.0)

Based on the information gotten from the ACF graph of ARIMA model, the majority of the ACF values of the ARIMA model fall within the range of 0.1 and -0.1, it suggests that the lags have relatively strong correlations with the current observation. This pattern indicates that there are significant dependencies or strong autocorrelations in the data beyond the immediate lag.

In much cases where the ACF values are small and not statistically significant, it implies that the data points are relatively dependent. This characteristic makes the observed discharge data suitable for modeling and forecasting using the ARIMA model selected by AI i.e. *ARIMA(4,1,1)(0,1,0)*

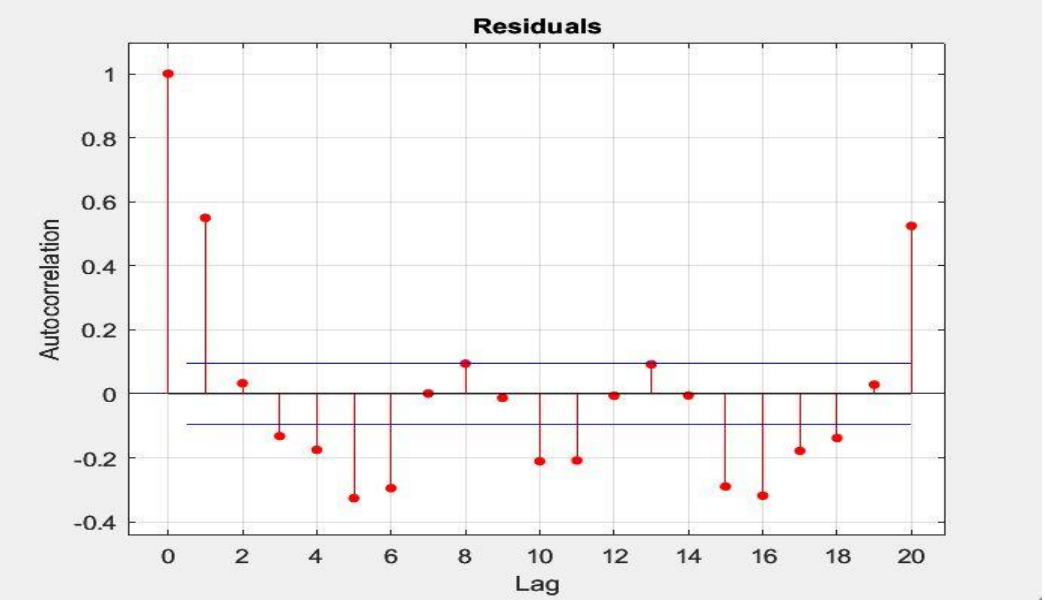

Figure 27 AutoCoorlation factor of Neural Network

The Autocorrelation Function (ACF) of the neural network model revealed several values exceeding the range of +0.1 and -0.1, suggesting a lack of fitness for accurate forecasting. This indicates that the neural network model did not adequately capture the temporal dependencies present in the data. The deviations from the desired range in the ACF signify a lack of significant autocorrelation at shorter lags, which is crucial for capturing the patterns and trends necessary for reliable forecasting. Consequently, alternative modeling approaches or adjustments may be required to improve the forecasting performance and ensure the model's suitability for capturing the underlying dynamics of the data.

The findings from the ACF analysis highlight the limitations of the neural network model in accurately capturing the temporal dependencies and emphasize the importance of selecting ARIMA model that accounts for the specific characteristics of the data making it better option for selecting it for forecasting.

3.3. Forecasted Results

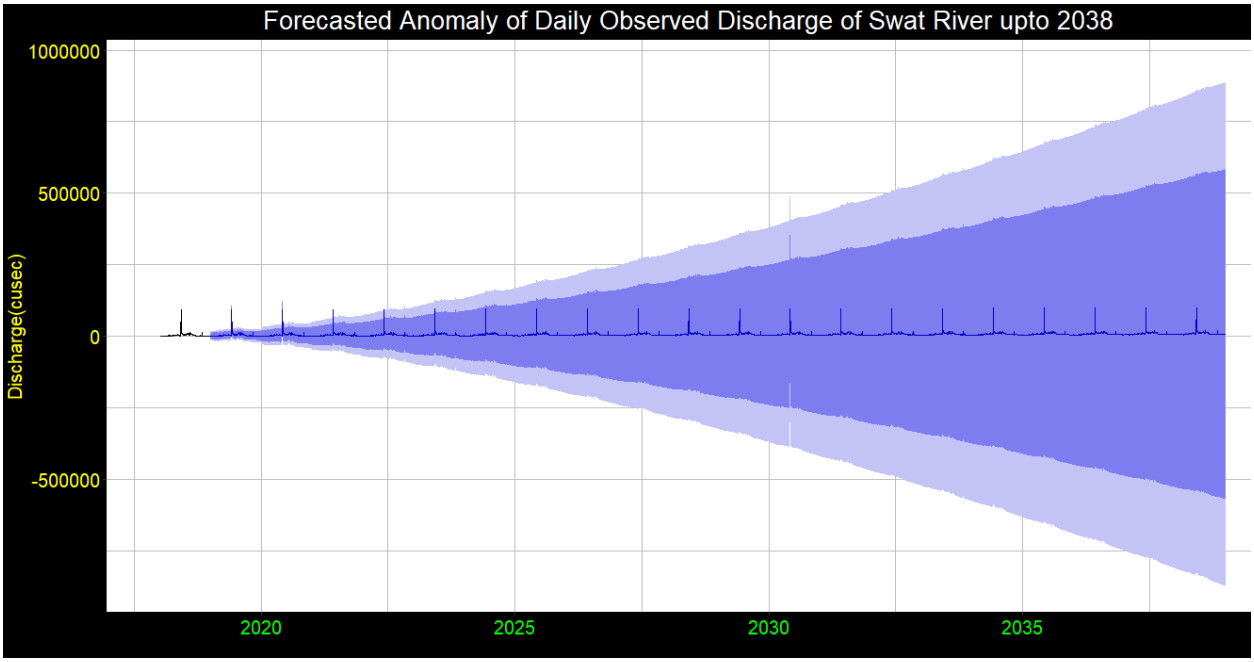

Figure 28 Forecasted Anomaly of Daily Observed Discharge of SWAT River

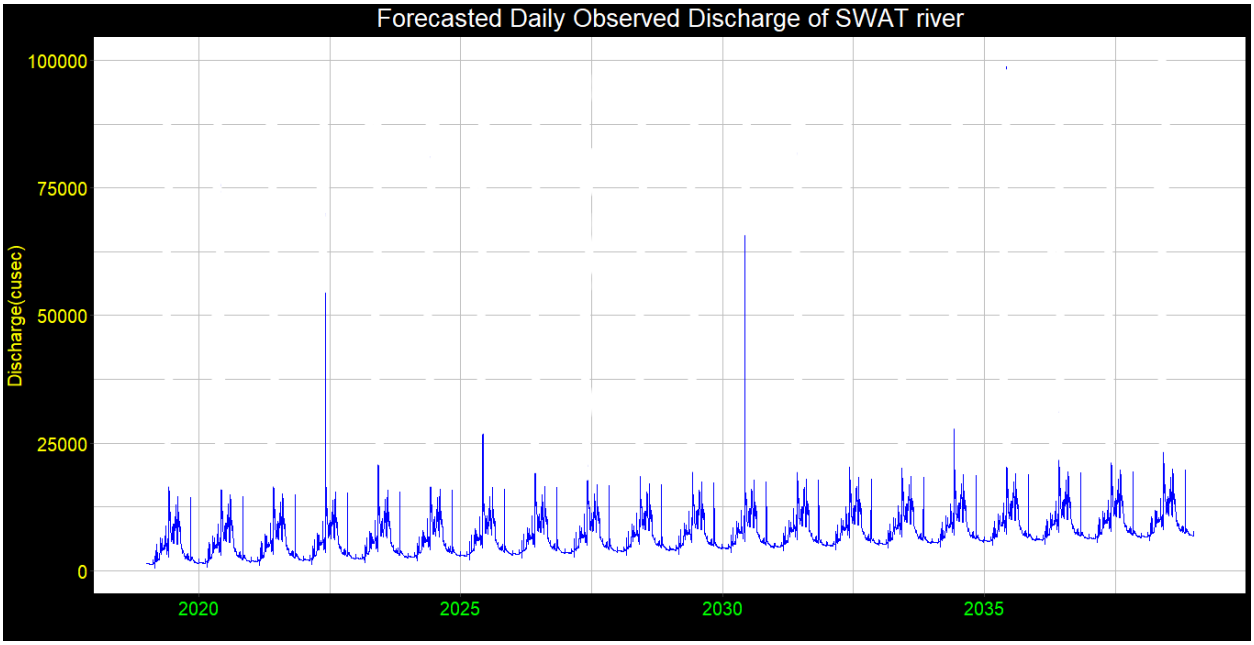

Figure 29 Forecasted Daily Observed Discharge of SWAT river

The analysis of the forecasted data reveals intriguing insights through two distinct graphs: the anomaly graph and the graph depicting the actual predicted values.

In the anomaly graph, the forecasted values exhibit a predicted increase in discharge, accompanied by certain fluctuations or noises. The dark blue portion of the graph represents the 90% possibility range, indicating a higher level of confidence in the forecasted anomalies. The light blue portion corresponds to the 80% possibility range, providing additional context regarding the uncertainty surrounding the predicted anomalies.

Upon further examination, the graph displaying the actual predicted values sheds light on the nature of the noises observed in the anomaly graph. It becomes evident that the anomalies coincide with two significant flood events. The first flood occurrence transpired at the beginning of 2023, while the second flood was predicted for the year 2030.

These findings highlight the predictive capability of the anomaly graph and its ability to capture and flag notable events such as floods. The visualization of the flood occurrences in the graph depicting the actual predicted values provides valuable validation of the anomaly predictions and demonstrates the usefulness of the forecasting model.

Understanding and accurately identifying such events are crucial for effective decision-making and management of river systems. The integration of anomaly detection techniques into the forecasting process can aid in providing early warnings for potential flood events and inform proactive measures to mitigate their impact.

3.4. Hydrological Model Results

The application of the hydrological model to the monthly data set has provided valuable insights into the observed increase in discharge. The primary analysis, followed by the fitting of the ARIMA model and subsequent forecasting, has strengthened our understanding of the discharge patterns over time. The hydrological model serves as a powerful tool for validating and corroborating the findings obtained from the observed daily data.

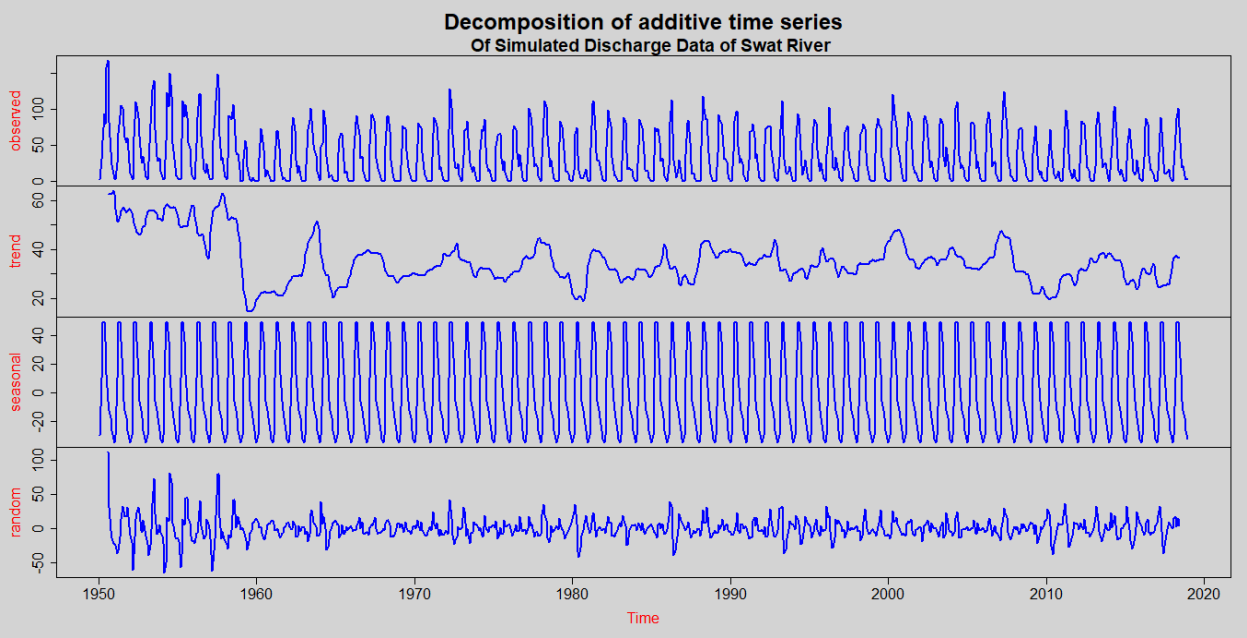

Figure 30 Decomposition of additive time series of Hydrological model

The hydrological model has confirmed that the observed data indeed exhibits an increase in discharge, aligning with the results derived from the primary analysis. This validation reassures the accuracy and reliability of the observed data, bolstering confidence in the conclusions drawn from the analysis.

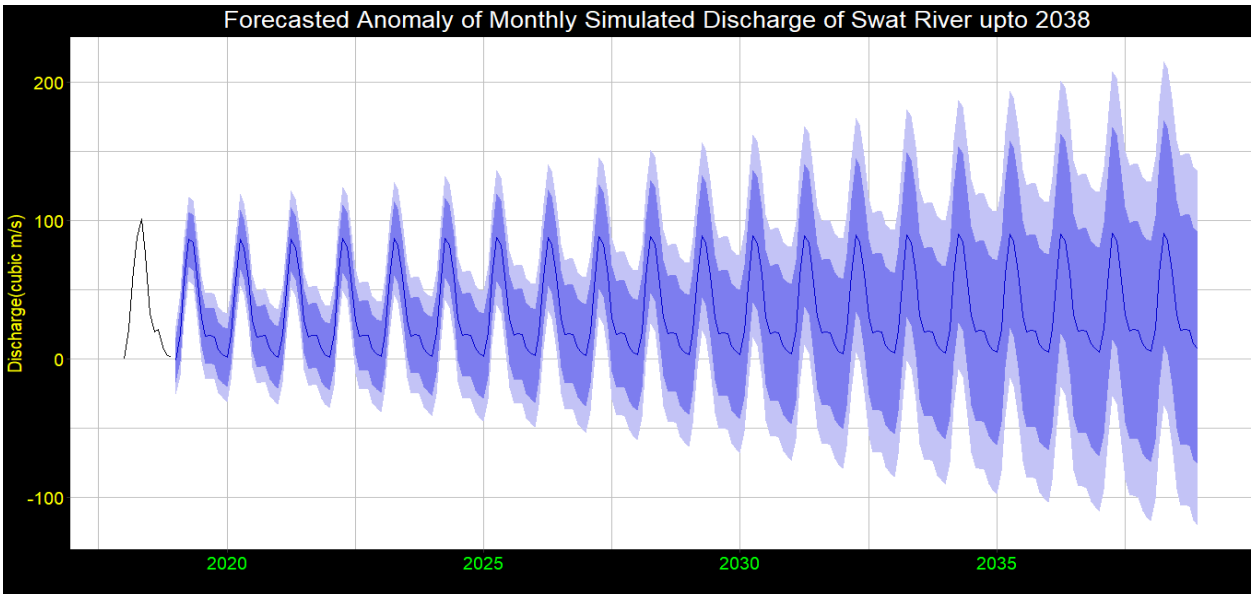

Figure 31 Forecasted Anomaly of Monthly Simulated Discharge

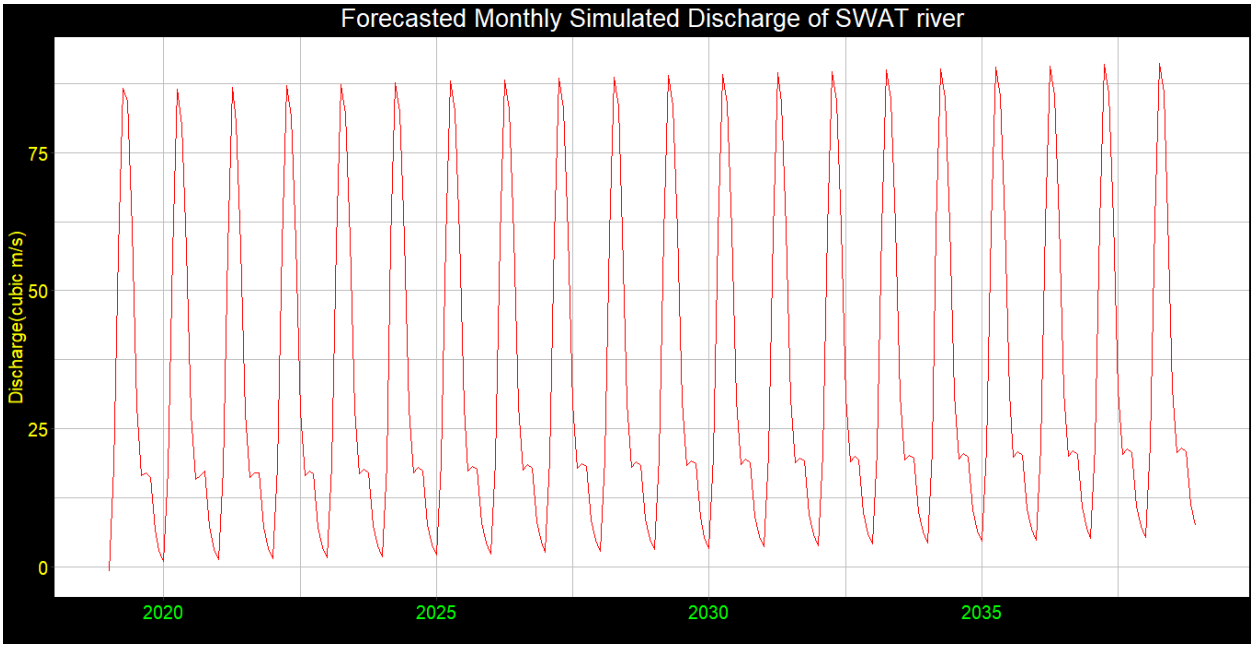

Figure 32 Forecasted Monthly Simulated Discharge of SWAT river

Although the hydrological model does not explicitly indicate the occurrence of floods due to the monthly resolution of the data, it has played a vital role in verifying the overall discharge trends. The hydrological model provides a broader perspective on the hydrological processes and their influence on the observed data, enhancing our understanding of the underlying dynamics.

CHAPTER 4

CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions

- For our dataset ARIMA model works better than neural network as ARIMA gave us better autocorrelation pattern than neural network.
- The discharge of the Swat River is predicted to increase in the next 20 years thus the hydropower plant will be able to generate power effectively.
- The frequency of floods is increasing in the next 20 years so arrangements should be made for smooth working of Malakand-3 HPP during the flood events.
- Furthermore, the findings indicate a rising water level in the Swat River, which serves as the upper stream for both the Swat Canal and the Malakand-3 hydropower plant.

4.2.1. Future Implications of Conclusions

Increase In Floods Frequency

The projected escalation in flood occurrence and water capacity within the Swat River can be attributed to multiple factors, with a particular emphasis on its geographical location and environmental elements. Analyzing the Digital Elevation Model (DEM) of the Swat River reveals its formation through the confluence of the Ushu and Gabral rivers, which join each other in the Kalam Valley, subsequently expanding the riverbed of the Swat River. Tracing the origins of these two rivers leads to the mountains, primarily the Hindu Kush ranges, which house substantial quantities of snow and glaciers. The mounting global warming crisis intensifies temperatures, thereby hastening the thawing process of the snow and glaciers, ultimately heightening the frequency of flood events.

The sustainability of the Malakand-3 hydro power plant has faced significant challenges due to the occurrence of floods and subsequent water level increases. These extreme weather events have resulted in the overflow of the Swat Upper Canal, which, in turn, has led to the closure of the tunnels leading to the power plant. As a result, power generation has been halted for an extended period, lasting approximately 6-8 months.

The impact of such interruptions on the sustainability of the Malakand-3 hydro power plant cannot be understated. The closure of the tunnels and the subsequent cessation of power generation have

disrupted the consistent and reliable supply of electricity to the grid. This not only hampers the power plant's ability to meet the energy demands of the region but also affects the overall stability of the power supply system.

Hydrological Sustainability of Malakand-3 Hydropower Plant

The Malakand III hydroelectric power plant is an exemplary hydrological sustainable project due to its utilization of water from the Upper Swat Canal (USC), which, in turn, draws water from the Swat River. The canal infrastructure designed to extract water from the USC for Malakand III is engineered to handle a discharge rate of 1800 cusec (cs), while the USC itself is designed to accommodate a discharge rate of 3657 cusic (cs). As the water levels in the Swat River increase, the current average rate of water discharge in Swat River is it 10, 000 cusec while the ARIMA forecast results shows that the amount of discharge will increase in the future, thus it ensures a sufficient and continuous water supply for future diversion towards the USC. Consequently, an ample amount of water will be available for diversion to not only Malakand III but also for Malakand I and Malakand II hydroelectric power plants.

This illustrates the project's sustainable nature and indicates the potential for further installation of hydroelectric power plants along the same river in the future.

4.2. Recommendations

- Provide water escapes in the Upper Swat Canal (USC) in the view of increasing discharge in the future as well as widen the USC at certain areas in this regard.
- There is a further potential of installation of more hydropower plants on the Swat River
- Installation of filtration tank before the tunnels on USC is recommended due to increasing water stage in the future.
- Establishing a website based-on automated process of this study for prediction of future trends in streamflow and floods in Swat River.

REFERENCES

1. *Time series forecasting method and components.* https://www.simplilearn.com.
2. Sinha, A. and A.K. Sinha, *A Next-Gen Power Generation Using Simulation And Machine Learning Forecasting.*
3. Chatfield, C., *The analysis of time series: an introduction*. 2003: Chapman and hall/CRC.
4. Box, G.E., et al., *Time Series Analysis: Forecasting and Control. Wiley.* Hoboken, NJ, 2008.
5. Falk, M., et al., *A First Course on Time Series Analysis: Examples with SAS [Version 2012. August. 01]*. 2012.
6. Sowell, F., *Maximum likelihood estimation of stationary univariate fractionally integrated time series models.* Journal of econometrics, 1992. **53**(1-3): p. 165-188.
7. Lai, T.L. and C.Z. Wei, *Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems.* The Annals of Statistics, 1982. **10**(1): p. 154-166.
8. Islam, M., et al. *Detection of Facebook Addiction Using Machine Learning*. in *International Conference on Image Processing and Capsule Networks*. 2022. Springer.
9. *Tree regression.* https://www2.stat.duke.edu/~rcs46/lectures_2017/08-trees/08-tree-regression.
10. *Neutral network.* https://aws.amazon.com/what-is/neural-network/.
11. *Support vector machine.* https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
13. *Malakand-3 hydropower plant.* https://pedokp.gov.pk/.
15. *Forecasting in Pakistan.* https://www.google.com/url?sa=t&source=web&rct=j&url=https://hess.copernicus.org/articles/22/3533/2018/&ved=2ahUKEwjvv8LRoJb7AhU7h_0HHbbBCUQQFnoECBQQAQ&usg=AOvVaw20XBQ4Oq4UGFP9Za1ukUg1.
16. *Forecasting in Kabul river.* https://www.mdpi.com/.
17. Ali, A., *Indus basin floods: Mechanisms, impacts, and management.* 2013.
18. *FAO Aquastat Data Portal.* http://www.fao.org/nr/water/aquastat/basins/indus/index.stm.

19. Mustafa, U., M.B. Baig, and G.S. Straquadine, *Impacts of Climate Change on Agricultural Sector of Pakistan: Status, Consequences, and Adoption Options*, in *Emerging Challenges to Food Production and Security in Asia, Middle East, and Africa*. 2021, Springer. p. 43-64.
20. Studies, N.I.o.P., M.I.I.f.R.D. Demographic, and H. Surveys, *Pakistan demographic and health survey*. 2012: National Institute of Population Studies.
21. Muzammil, S., *Occupational health in Pakistan: challenges and future needs.* Occupational and Environmental Medicine, 2020. **77**(1): p. 56-56.
22. Ismail, S., *Military expenditure and economic growth in South Asian countries: Empirical evidences.* International Journal of Economics and Financial Issues, 2017. **7**(3): p. 318-325.
23. Naveed, A. and N. Ali, *Clustered deprivation: District profile of poverty in Pakistan*. 2012: Sustainable Development Policy Institute.
24. Kant, S., *Trend and variability of hourly intensity of rainfall over eastern and northern part of Uttar Pradesh during 1969-2014.* MAUSAM, 2018. **69**(4): p. 577-588.
25. Coghlan, A., *A little book of R for time series.* Published under Creative Commons Attribution, 2015. **3**.
26. Zhang, G.P., *Time series forecasting using a hybrid ARIMA and neural network model.* Neurocomputing, 2003. **50**: p. 159-175.
27. Zhang, G.P. and M. Qi, *Neural network forecasting for seasonal and trend time series.* European journal of operational research, 2005. **160**(2): p. 501-514.
28. Weka, W., *3: data mining software in Java.* University of Waikato, Hamilton, New Zealand (www. cs. waikato. ac. nz/ml/weka), 2011. **19**: p. 52.
29. Cox Jr, J.E. and D.G. Loomis, *Improving forecasting through textbooks—A 25 year review.* International Journal of Forecasting, 2006. **22**(3): p. 617-624.
30. Bowerman, B.L., R.T. O'Connell, and A.B. Koehler, *Forecasting, time series, and regression: an applied approach*. Vol. 4. 2005: South-Western Pub.
31. Hyndman, R.J. and A.B. Koehler, *Another look at measures of forecast accuracy.* International journal of forecasting, 2006. **22**(4): p. 679-688.

32. Witten, I.H., E. Frank, and M. Hall, *Algorithms: The Basic Methods Data Mining: Practical Machine Learning Tools and Techniques (pp. 85-146).* Burlington, MA: Morgan Kaufmann, 2011.