# QUANTIFYING AND PROJECTING THE STREAMFLOW VARIATIONS IN SWAT RIVER


By

| | |
|---|---|
| (Leader - CMS 358930 | CSUO Fahad Hussain) |
| (Member 1 - CMS 358947 | GC Mohammad Usman) |
| (Member 2 - CMS 358949 | SGT Saad Qazi) |
| (Member 3 - CMS 358959 | BSS Omer Munsif) |

Department of Water Resources Engineering and Management

Military College of Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan


(2024)

# QUANTIFYING AND PROJECTING THE STREAMFLOW VARIATIONS IN SWAT RIVER

By


(Leader - CMS 358930 CSUO Fahad Hussain)

(Member 1 - CMS 358947 GC Muhammad Usman)

(Member 2 - CMS 358949 SGT Saad Qazi)

(Member 3 - CMS 358959 BSS Omer Munsif)


A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Bachelor of Science in
Civil Engineering


Supervisor: Dr. Muhmmad Amjad

Military College of Engineering

National University of Sciences & Technology (NUST)

Islamabad, Pakistan


(2024)

# THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of UG Thesis written by Mr. Fahad Hussain (Registration No. 358930), Mr. Mohammad Usman (Registration No.358947), Mr. Saad Qazi (Registration No. 358949), Mr. Omer Munsif (Registration No. 358959), of Military College of Engineering has been vetted by undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of Bachelor's degree. It is further certified that necessary amendments as point out by GEC members and foreign/ local evaluators of the scholar have also been incorporated in the said thesis.

Signature: ____________________

Name of Supervisor: Dr. Mohammad Amjad

Date: 14/05/24

Signature: ____________________

Name of HOD: Dr. Mohammad Amjad

Date: 14/05/24

## CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled **QUANTIFYING AND PROJECTING THE STREAMFLOW VARIATIONS IN SWAT RIVER** was conducted by Mr. Fahad Hussain under the supervision of Dr. Mohammad Amjad.

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the Department of Water Resources and Environment) in partial fulfillment of the requirements for the degree of Bachelors of Civil Engineering.

Department of Construction Engineering and Management, Military College of Engineering, Risalpur.

Student Name: Fahad Hussain Signature: ____________

Mohammad Usman Signature: ____________

Saad Qazi Signature: ____________

Omer Munsif Signature: ____________

Supervisor Name: Dr. Mohammad Amjad Signature: ____________

## AUTHOR'S DECLARATION

I Fahad Hussain, Mohammad Usman, Saad Qazi and Omer Munsif hereby state that our UG thesis titled **QUANTIFYING AND PROJECTING THE STREAMFLOW VARIATIONS IN SWAT RIVER** is our own work and has not been submitted previously by us for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my UG degree.

Name of Student: Fahad Hussain

Mohammad Usman

Saad Qazi

Omer Munsif

Date: 14 May, 24

# PLAGIARISM UNDERTAKING

I solemnly declare that research work presented in the thesis titled is **QUANTIFYING AND PROJECTING THE STREAMFLOW VARIATIONS OF RIVER SWAT** solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of UG degree, the University reserves the rights to withdraw/revoke my UG degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: ____________________

Name: Fahad Hussain

Student Signature: ____________________

Name: Mohammad Usman

Student Signature: ____________________

Name: Saad Qazi

Student Signature: ____________________

Name: Omer Munsif

# DEDICATION

To Almighty Allah for giving us the strength to undertake this project. And to our parents, whose dedication and teachings have been a beacon of light all our life.

# ACKNOWLEGMMENTS

We are thankful to **Allah Almighty** for guiding us and for every new thought which He set up in our minds. Nothing would have been possible except by hiswill and command. All praises are for Him who has been our help the entire way. We are extremely grateful to our parents who have dedicated their entire efforts to our better upbringing and have supported us through all the difficult times in our lives. We would also like to express the deepest of gratitude to our supervisor **Dr. Muhammad Amjad** for helping us throughout our project and for guiding us through the difficult processes. We can safely say that without his help and dedication undertaking such an arduous project would have been nothing short of impossible.

Finally, we would like to express our gratitude to all the individuals who have rendered valuable assistanceto our work and study.

All Syndicate Members

# TABEL OF CONTENTS

## LIST OF FIGURES



## LIST OF TABLES

# ABSTRACT

The changing climate is seriously imposing threats to existence of mankind and other creatures. This study focuses River Swat basin in Khyber Pakhtunkhwa, Pakistan, and aims to forecast its input source by considering changing climate conditions that affect water flow. The study area encompasses the Upper Swat Canal system, specifically the Swat River and its tributaries, characterized by complex topography and used for irrigation. 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. Data collection involved field surveys, interviews, and online resources, and preprocessing techniques were applied to ensure data accuracy. A hydrological model was developed and calibrated to simulate the water cycle and discharge in the study area. The ARIMA model, specifically ARIMA (4,1,1) (0,1,0) [12], was found to be the most suitable for forecasting. The results indicated an increase in flood frequency and rising water levels in the Swat River. The study highlights the significance of accurate forecasting and flood management. The study focuses on use of different models and their intercomparison. Overall, this research contributes to enhancing the understanding of water flow dynamics in the study area and provides valuable insights for decision-making for public benefit.

# CHAPTER 1

# INTRODUCTION

## 1.1. Background

The Swat River in Khyber Pakhtunkhwa, originating from the Hindukush range, is a essential freshwater source supporting agriculture, biodiversity, and hydropower. However, it faces unprecedented challenges because of climate change, main to shifts in precipitation patterns and circulation float uncertainty. This poses risks to agriculture, extended floods and droughts, and socio-economic influences. To cope with this, we intention to quantify weather exchange's effect using advanced hydro logical modeling and historical records analysis. This venture seeks to understand the interplay between climate alternate and the Swat River's hydrodynamics, guiding sustainable aid control in an era of fast environmental exchange.

## 1.2. Problem statement

Observed changes in temperature, precipitation patterns and snow melt dynamics are increasingly influencing the flow patterns of River swat. Due to sudden climate change the intensity & frequency of Floods has also been increased that is very alarming. Glaciers are melting on higherrates which is point of consideration. It demands a complete scientific enquiry quantifying the impact of climate change on the stream flow of river swat including the mechanism and predicting the future scenarios. This paper will address the question by critically analyzing the issue and using advanced hydro logical modeling, coupled with the historical data which will release the output for the future course of action of River Swat.

## 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]

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 km2 or 65% of the country's total area, including the entire Punjab, Khyber Pakhtunkhwa, most of Sindh and eastern Baluchistan. [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 Baluchistan 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]

### 1.3.2. 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 forthe future.[1] Top four types of forecasting methods are 1) Straight-line, 2) Moving average, 3) Simple linear regression,4) Multiple linear regression.

When developing models using time series data, it is essential to understand the underlying patterns present in the data over time. These patterns can be divided into four components:

a. Trend: The trend component captures gradual changes or shifts observed in time series data over a longer period. Trend patterns show long-term increases or decreases in data.

b. Level: The level component represents the base value or average level of the time series. It is often represented by a straight line that indicates the central tendency of the data.

c. Seasonality: Seasonality refers to recurring and predictable patterns that repeat themselves within a fixed unit of time, such as daily, weekly, monthly, or annual cycles. These patterns can be affected by factors such as weather, holidays or other regular events.

d. Noise: The noise component, also known as residual or error, accounts for the random fluctuations or irregularities present in the data. It represents unpredictable variations that cannot be explained or predicted by a model. These fluctuations may come from measurement errors, random events, or unaccounted for factors.

If observations are made at specific times and occur at regular intervals, the nature of time series can be classified as either continuous or discrete. However, if 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 be either unidimensional or multidimensional, representing multiple parameters at a given moment.

It is also worth noting that the parameters within a time series can be non-numeric, allowing the representation of variables outside of numerical values. [3]. Additionally, it is worth noting that a time series process can be categorized as deterministic or probabilistic. A deterministic process implies that the parameters of a time series can be accurately predicted, while a probabilistic process implies that past values only partially determine the future, making accurate predictions unattainable.

#### 1.3.2.1. ARIMA Models

The ARIMA model, which stands for Autoregressive Integrated Moving Average, is a powerful tool for analyzing and forecasting time series data, especially when the data is non-stationary. Hereis a summary of its key components and steps:

***ARIMA Components***

- Autoregressive (AR) model (p): Captures dependencies between current and past values in timeseries data.
- **Moving average (MA) model (q)**: Models the relationship between the current value and pastforecast errors.
- Integration (I) Component (d): Achieves stationarity by differentiating time series data.

- The ARIMA model combines these components into a single framework.

***Model Equation***

The ARIMA model equation is given by: $\phi(B)\omega t=\theta(B)at$ϕ(B)ωt=θ(B)at
Here, $\phi(B)$ϕ(B) and $\theta(B)$θ(B) are the autoregressive and moving average operators, $\omega t$ωt is the difference time series and $at$at is the random shock.

***Stationarity and differentiation:***

- The ARIMA model requires the time series data to be stationary.
- Differentiation is used to achieve stationarity, denoted by the parameter $d$d.

- Typical values for $d$d are 0, 1 or 2.

***Model identification:***

- Involves determining appropriate values for $p$p, $d$d and $q$q.
- It is performed by analyzing autocorrelation and partial autocorrelation functions (ACF and PACF).
- The Box-Jenkins approach provides a systematic method for parameter identification.

***Parameter estimation:***

- Maximum likelihood estimation and the method of least squares are commonly used for parameter estimation.
- Parameters are estimated using autocovariances.

***Refinement of the model:***

- Initial parameter values (e.g., $p$=1 p=1 and $q$=0 q=0) can be adjusted based on prediction performance.
- Different combinations of $p$p and $q$q can be tried to improve the fit of the model.

In summary, ARIMA models offer a flexible approach to modeling and forecasting time series data by incorporating autoregressive, moving average, and differential components. Parameter selection and model specification are critical steps in building an effective ARIMA model.

## 1.4. Literature Gap

a. The absence of inter comparison between traditional hydrological modeling and ARIMA based.
b. Deficient studies regarding streamflow assessment and projections on swat river basin in Pakistan.
c. Less forecasting of stream flow as climate change can introduce additional uncertainties into stream flow forecast

## 1.5. Objectives

a. The objectives of this study are:

b. To investigate the stream flow variations in Swat River in the past.

c. To forecast the stream flow for near future.

d. To investigate the overall effects of climate change on stream flow variation of Swat Riverbasin.

## 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 summer and winter.

# CHAPTER 2

# DATA AND METHODS

## 2.1. Study Area

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 irregulartopography of the land, including today's Charsada, Mardan and Swabi districts, and the canal could only irrigate a small portion of the land leaving considerable potential untapped. 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 ina 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.

### 2.1.1 DEM of Swat River Catchment

#### *2.1.1.1. Step-wise Procedure:*

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

a. Extracted DEM of Swat River basin from Khyber Pakhtunkhwa (KP) dataset using ArcMap basinboundary-based extraction tool.

b. Used ArcMap's fill tool to remove small imperfections in the elevation data by filling dips in thesurface raster.

c. Generated flow directions from the populated DEM using the flow direction tool, recording theflow from each cell in the surface raster.

d. Calculated flow accumulation data based on downstream output.

e. Labeled freezing points/outfalls to generate freezing point data for the watershed.

f. Delineated watershed boundary using a grid representing the flow direction.

g. Defined watershed flows to determine contributing area and converted watersheds and flows tovector format.

h. Extracted watershed flows and DEM for further analysis.

i. Calculated basin area using information from the DEM model and extracted vector data.

To comprehensively analyze the hydrological processes and spatial characteristics of the Swat river basin, we used a combination of extraction, filling, flow direction, flow accumulation, delineation and vector conversion tools within ArcGIS software. This approach provided a remarkably accurate and precise assessment of the watershed.

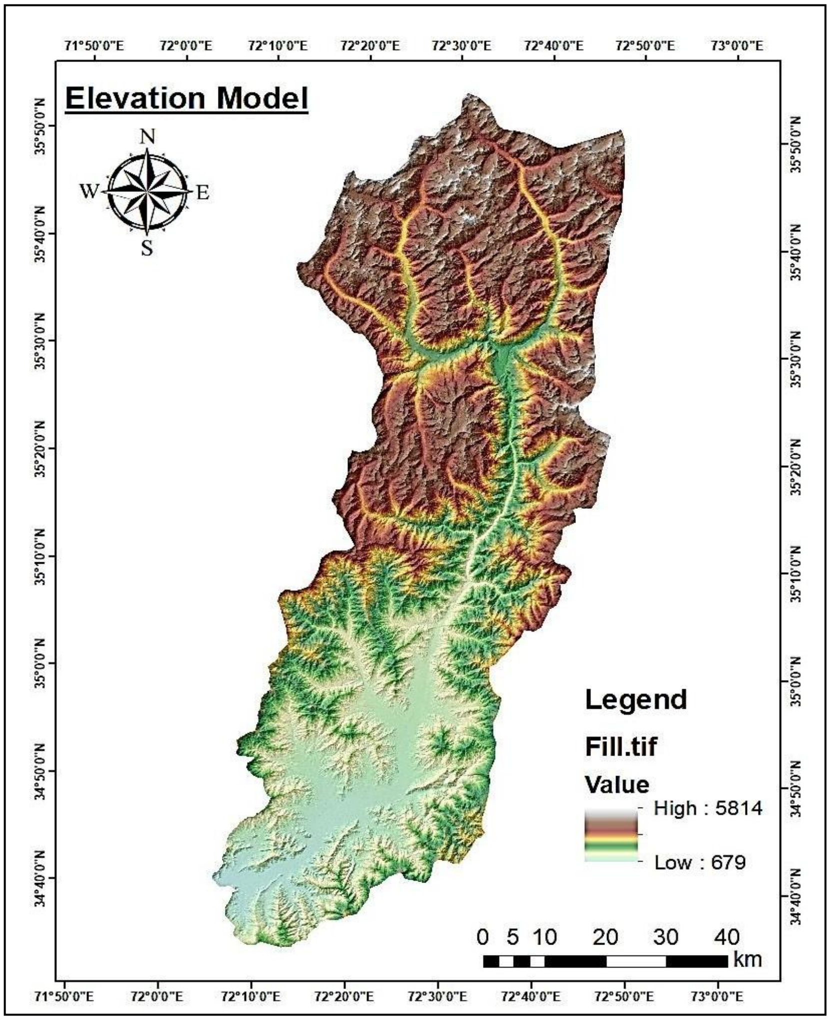


*Figure 1: Elevation model of River Swat*

The elevation model of the River Swat 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.

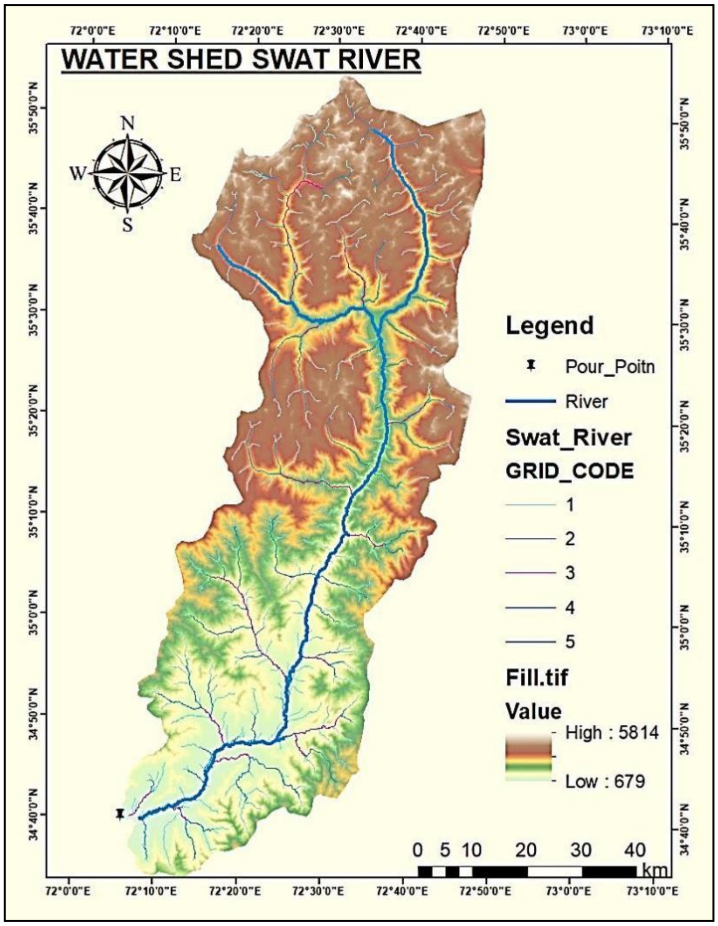


*Figure 2: Watershed of River Swat*

The watershed model of River Swat, 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.

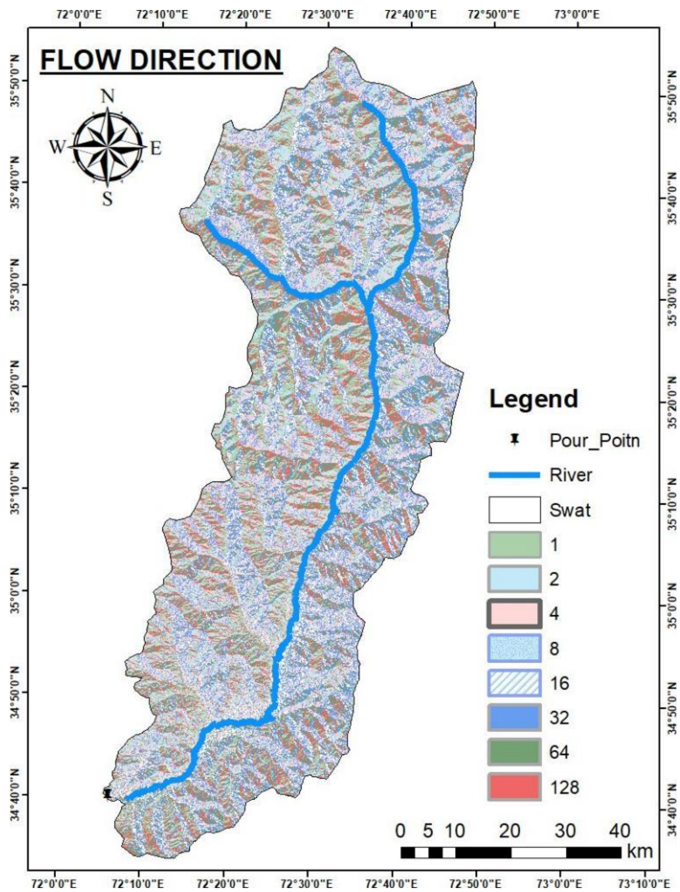


*Figure 3: Flow direction of River Swat*

The flow direction model generated in ARCGIS depicts the pathways of water flow within the River Swat 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. Thismodel serves as a fundamental component for hydrological analysis, enabling the identification offlow accumulation areas and contributing to the accurate simulation of water movement within the catchment.

## 2.2. Data Collection

### 2.2.1. 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 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. Surveys were conducted by group members between November 15, 2023, and on November 19, 2023, and an interview with SDO Nizam from the Department of Irrigation was conducted.

## 2.3. ERA5-Land Data:

The fifth-generation reanalysis data (ERA5-Land) released by the European Center for Medium-Range Weather Forecasts (ECMWF) is a global dataset that provides high-resolution (~10 km spatial resolution) 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 fraction of the state-of-the-art fifth generation reanalysis of ECMWF, namely ERA5.

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

***Reading and processing ERA5-Land data***:

ERA5-Land monthly precipitation and temperature averages obtained from CDS, processed into an R-compatible format (i.e. ".Rdata") and adjusted for the Kabul basin.

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

***CORDEX climate projections:***

The Coordinated Regional Downscaling Experiment, known as CORDEX (https://cordex.org/), archives climate projection data for different spatial segments (domains) of the globe, such as "South Asia", "Middle East North Africa", etc. Data Projection Climate are available for various regional circulation models (RCMs), while the data for these RCMs are downscaled from global circulation models (GCMs) that have a relatively coarser spatial resolution. This study used climate projections (for RCP 8.5) obtained from RCMs with a mean resolution of ~0.44 (i.e., ~45 km) over the Middle East and North Africa (MENA) region.

***Reading and processing CORDEX climate projections:***

Climate projection data (originally in ".nc" format) over MENA was processed into an R-compatible format and trimmed for the Kabul basin. The data contained daily estimates for two parameters (total precipitation and 2m-temperature) for the period 2023-2100. Later, daily precipitation and temperature data were converted to monthly data to be compatible with HBV Light.

The "ptq" file to be input to HBV Light was prepared using climate projections for precipitation and temperature along with the "EVAP" file generated using the temperature data.

### 2.4. Hydrological modeling

HBV light is a simplified version of the HBV hydrological model developed by the Swedish Meteorological and Hydrological Institute (SMHI), designed for easy use in research and education. While based on the basic equations of the HBV-6 model, HBV light brings several modifications and improvements.

Key features of the HBV light include:

a. **Warm-up period**: Instead of using initial states, HBV light uses a warm-up period, which simplifies the model setup process.

b. **Flexibility of routing parameters**: Unlike the previous version, which limited routing parameters (MAXBAS) to integer values only, HBV light removes this limitation and offers moreflexibility in parameterization.

c. **Simplified Functionality**: In order to maintain simplicity, HBV light omits some features present in HBV-6, focusing on the basic features necessary for outflow simulations.

d. **Additional options:** HBV light introduces two new options that are missing from HBV-6: the ability to incorporate observed groundwater levels into the analysis and the use of a different response routine with a delay parameter.

For outflow simulations, HBV light requires two mandatory input files:

- PTQ file (ptq.txt): Contains daily time series of precipitation, temperature and observed runoff data.
- evaporation-file (EVAP.txt): Provides values for potential evaporation.

The simulation period must be divided into calibration and validation periods with the possibility of saving results for each period. Detailed results files include both input and output variables, such as simulated runoff, rainfall, and snow time series data. The summary files offer annual water balance information and goodness-of-fit metrics including coefficient of determination, Nash-Sutcliffe efficiency (NSE), Kling-Gupta efficiency (KGE), and mean difference.

In summary, HBV light simplifies the HBV hydrologic modeling process, providing an accessible tool for performing runoff simulations and analyzing hydrologic processes while maintaining core functionality and introducing new capabilities for enhanced analysis.

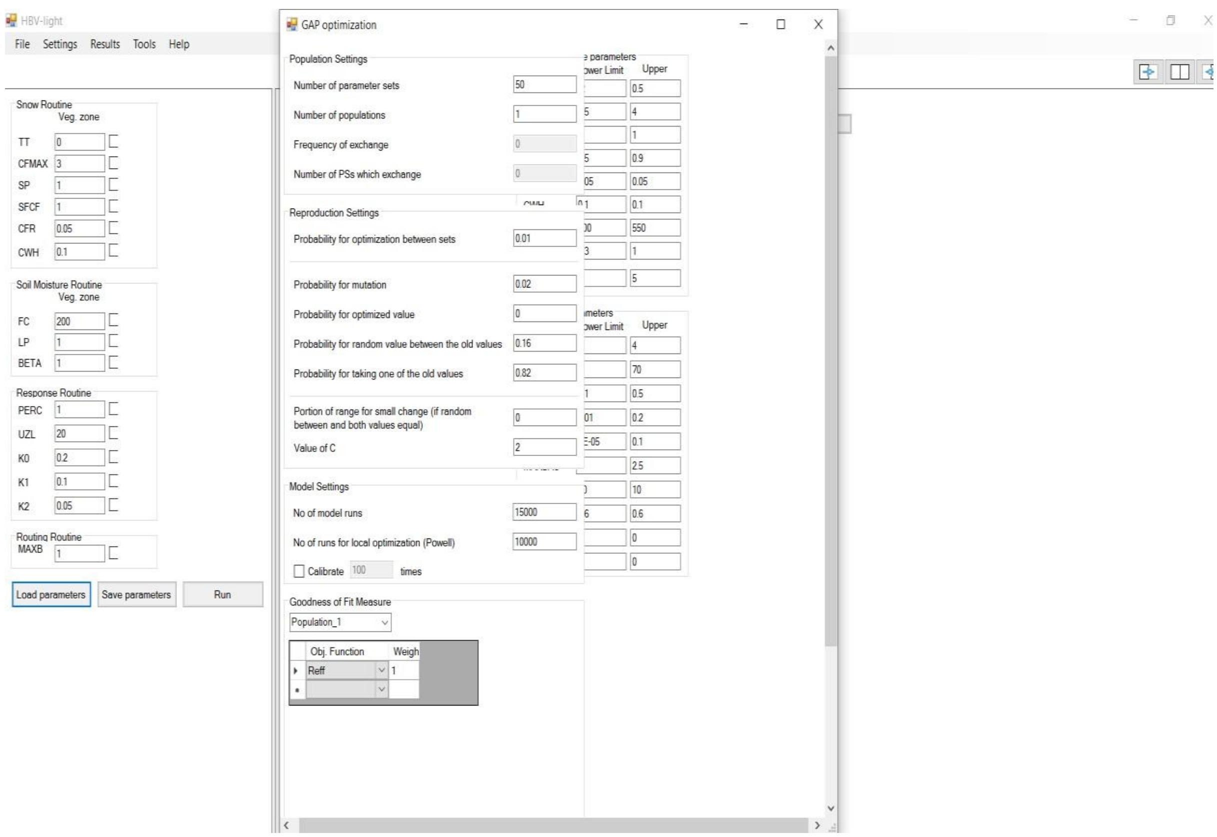


*Figure 4: Calibration – parameter optimization*

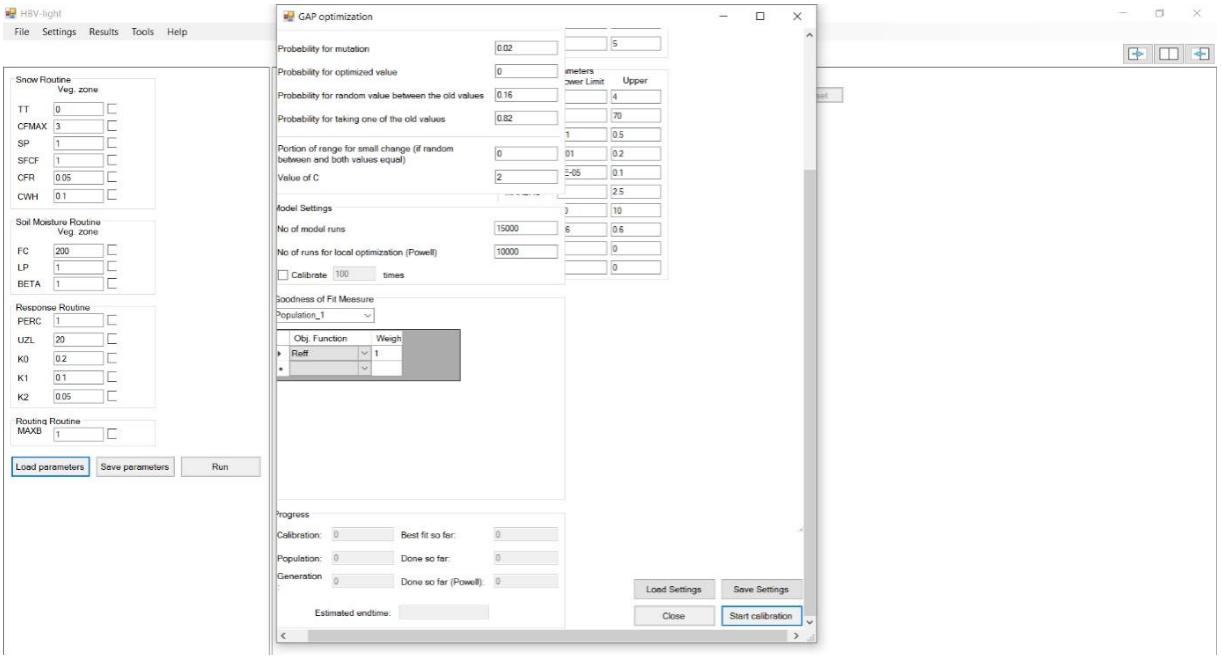

*Figure 5: Calibration – parameter optimization*

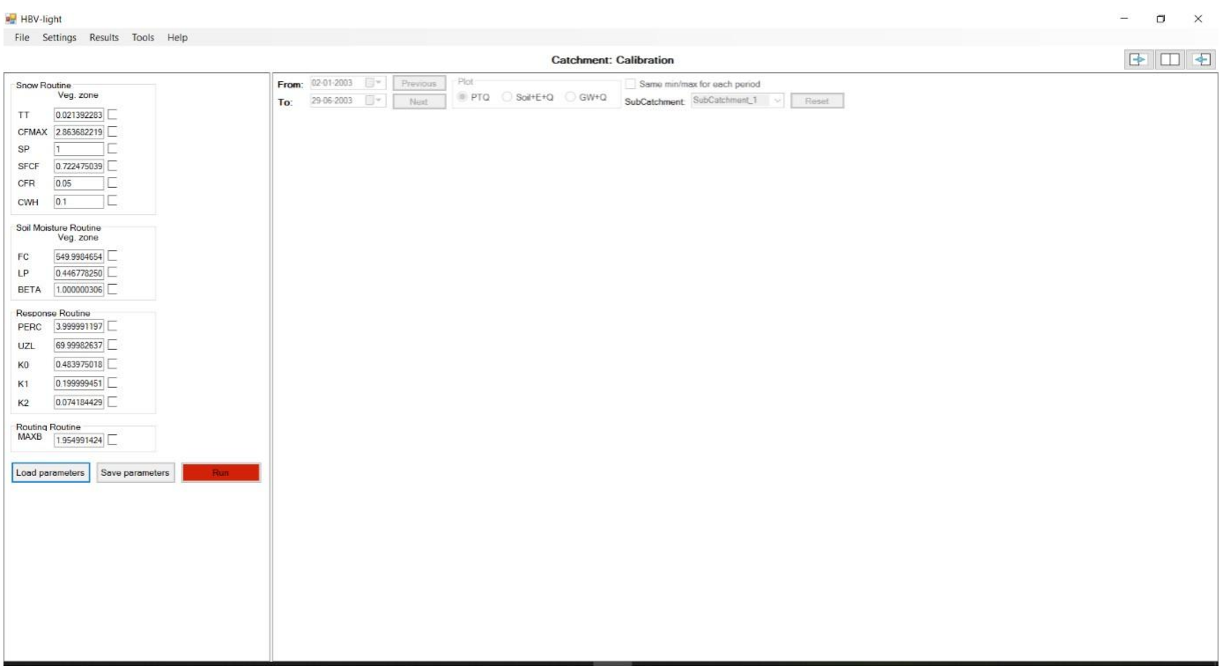

*Figure 5: Calibration – parameter loading*

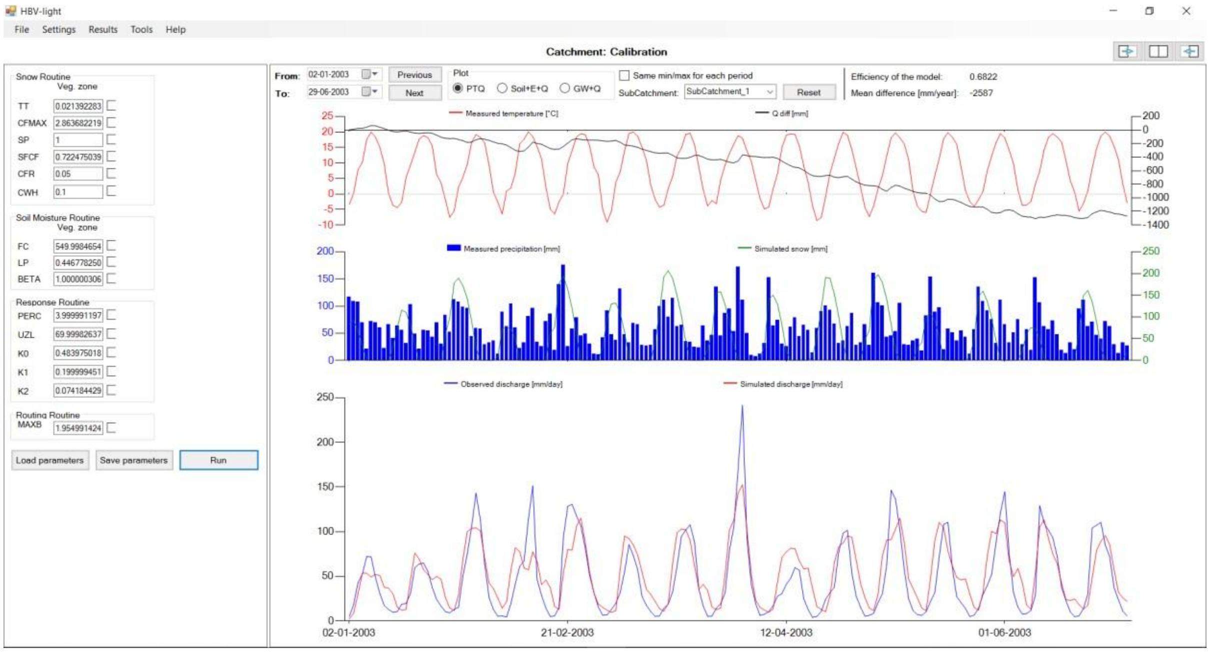


*Figure 6: Calibration – model run*


*Figure 7: Validation – model run*

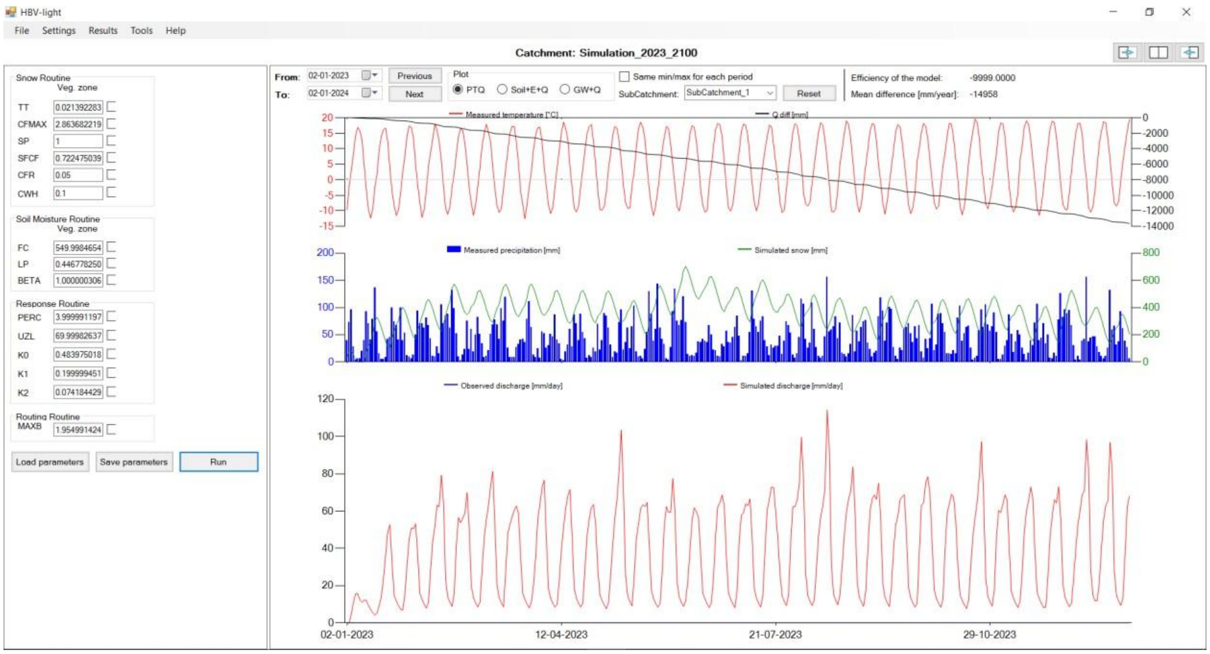


*Figure 8: Simulation – model run*

### 2.5. Analysis and Forecasting

#### 2.5.1. ARIMA Forecasting

The methodology adopted in this study involved several sequential steps for analyzing the dailydischarge of Swat River:

1. **Data Collection**: Daily discharge data from 2010 to 2018 and monthly discharge data from 1950 to 2018, obtained from the Hydrological model, were collected for analysis.

2. **Preliminary Analysis**: Initial analysis was conducted to understand the characteristics of the collected data. Time series plots, seasonal plots, and sub-series plots were generated using the 'autoplot' function from the 'fpp2' package in R. Additionally, decomposition was

performed using the 'decompose' function to examine trend, seasonal, and residual components.

3. **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 anomaly time series was assessed for stationarity.

4. **Model selection**: An appropriate ARIMA model was selected using an automated procedure based on artificial intelligence (AI). The 'auto.arima' function determined the optimal ARIMA model by considering criteria such as AIC and BIC. The AI algorithm autonomously determined the differentiation order for daily flow data and the seasonal difference for monthly flow data. Autocorrelation of model residuals was examined using the 'checkresiduals' function.

5. **Forecast**: The selected ARIMA model was used to generate forecasts for both daily and monthly flow time series. Forecasts were generated using the "forecast" function, which specified the desired number of forecast periods. The resulting forecasts were visualized using the “autoplot” function, and the forecast results were saved in CSV and Excel formats.

6. **Interpretation:** The forecast results were interpreted to understand the future flow behavior of Swat rivers. Visualization of predicted data helped to understand expected changes over time. In addition, an 80% probability range of high discharge values was shown to assess the potential risks associated with such events.

### 2.5.2. 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)


#Series appears trend-stationary, use to investigate seasonality

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

# CHAPTER 3

# RESULTS AND DISCUSSION

## 3.1. Primary Analysis

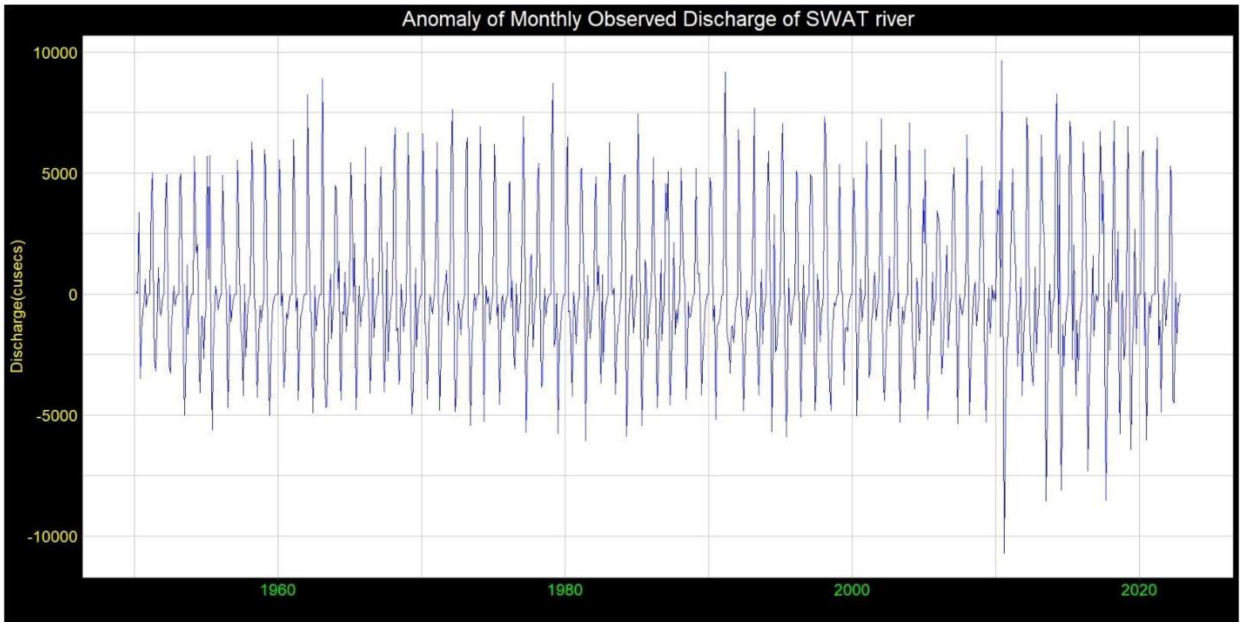


*Figure 9: Monthly observed discharge of Swat River*

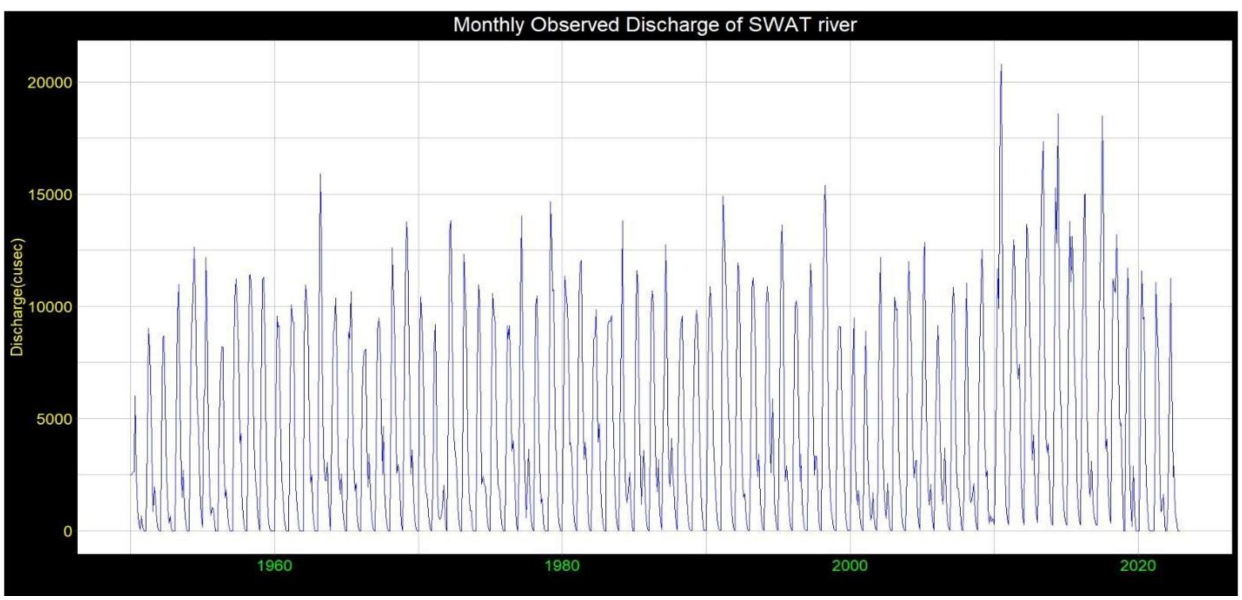


*Figure 10: Anomaly of monthly discharge of Swat River*

Between 2010 and 2018, the daily observed discharge data of the Swat River in 2010 showed a significant anomaly characterized by a remarkably high value. This exceptional increase in flow is indicative of a severe flood that occurred that year. Subsequently, the remaining years showed a consistent discharge pattern, demonstrating well-defined trends. These regular trends in the data reflect typical river flow behavior, undisturbed by any extraordinary events. This observation further underscores the significance of the 2010 flood as an exceptional event in the midst of an otherwise predictable flow. Findings from this plot provide valuable insights into the hydrological dynamics of the Swat River and contribute to a comprehensive understanding of its behavior over the time frame studied.

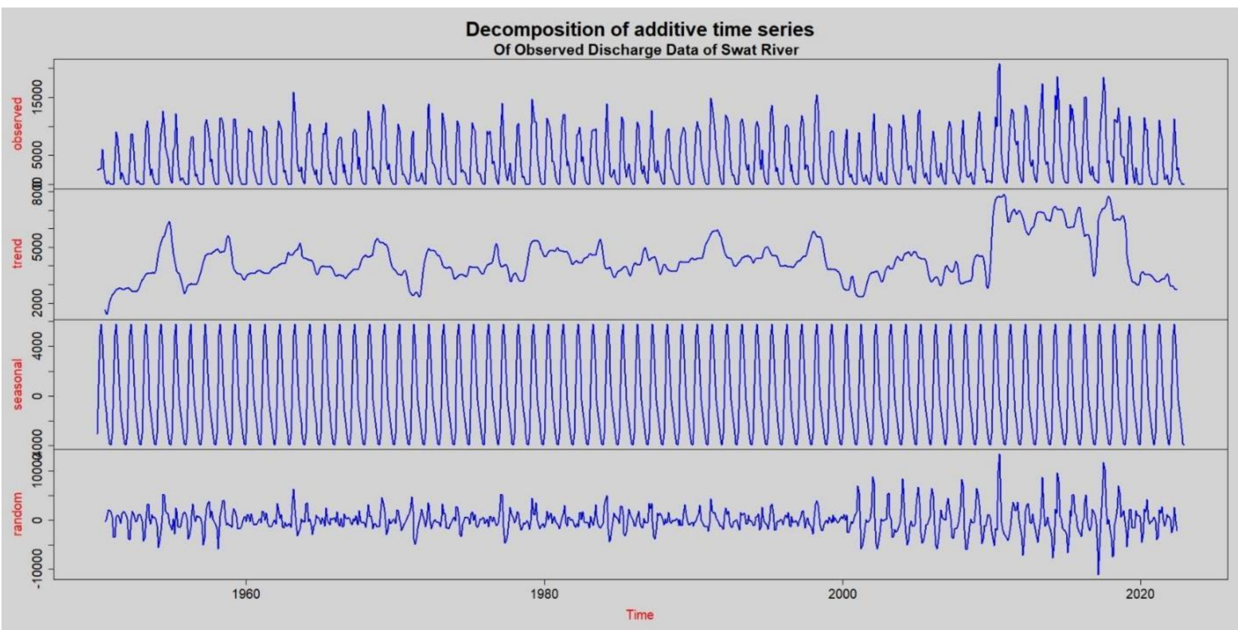


*Figure 11: Decomposition of additive time series*

Decomposition analysis of upstream data provides a comprehensive understanding of its underlying components, including trend and seasonality and other factors. The trend component reveals a clear pattern in the data that indicates a significant drop in flow after the flood in 2010. This drop indicates a gradual recovery phase of the Swat River after the devastating flood. However, there is an interesting observation from the data after 2018, as they again show an upward trend. This potential rebound in discharge levels implies the need for continuous monitoring and analysis to understand the underlying factors driving this change.

Examining the seasonality component, it is clear that discharge levels show a distinct cyclical pattern during the study period. The data highlights that the mid-year months consistently show higher discharges compared to the beginning and end of each month. This recurring pattern of increased flow during the mid-year months indicates the presence of seasonal influences on river flow dynamics. Identifying this seasonality contributes to a more comprehensive understanding of the river's behavior and helps with future forecasting and water resource management.

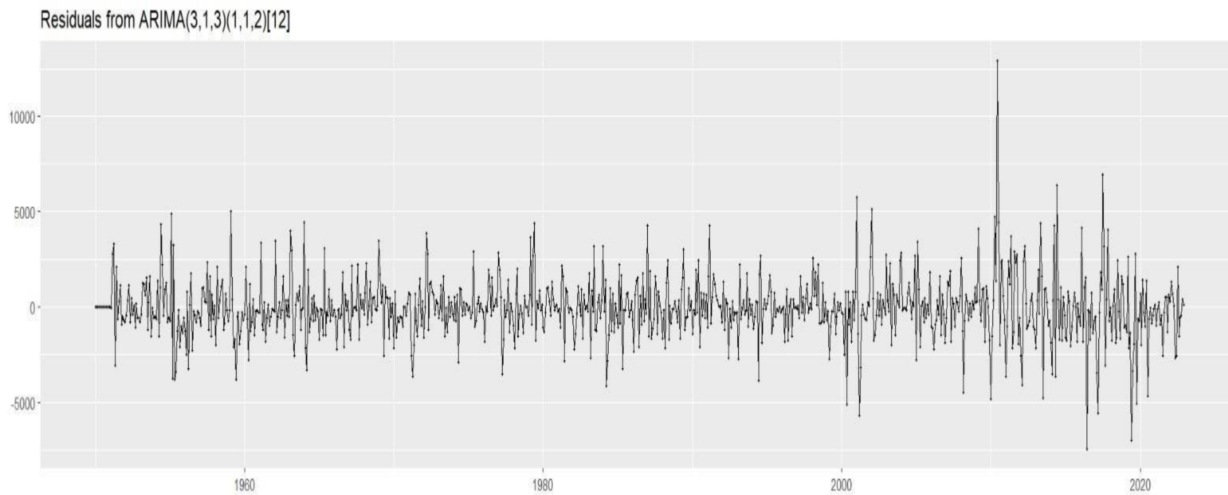


*Figure 12: Autocorrelation factor of ARIMA (4,1,1)(0.1.0)*

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)*.

### 3.3. Forecasted Results

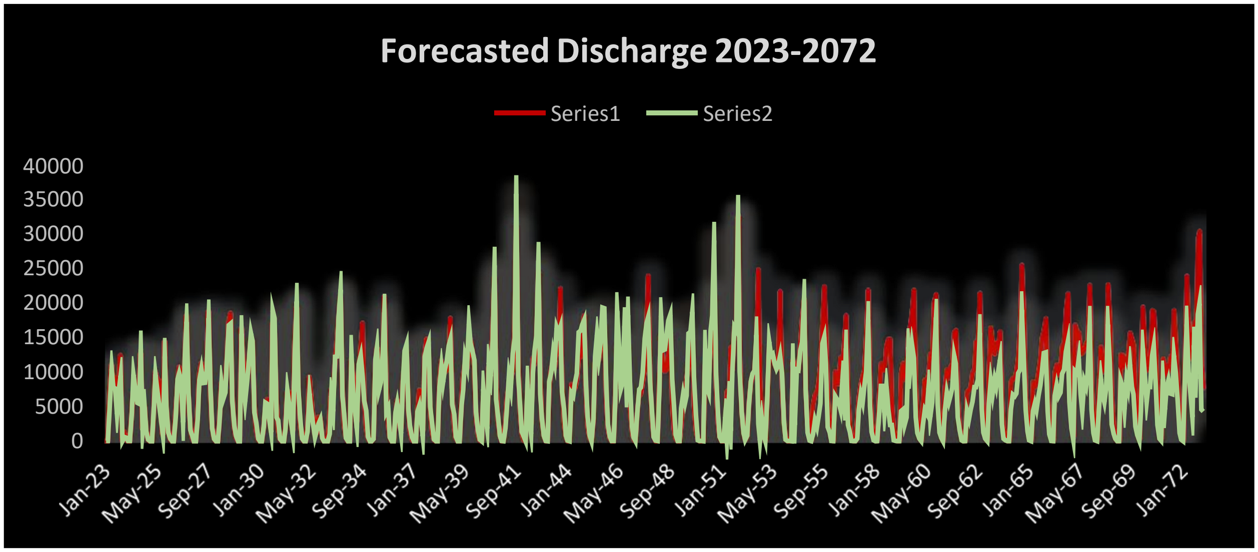


*Figure 13: Forecasted monthly discharge of Swat River till 2072*

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.

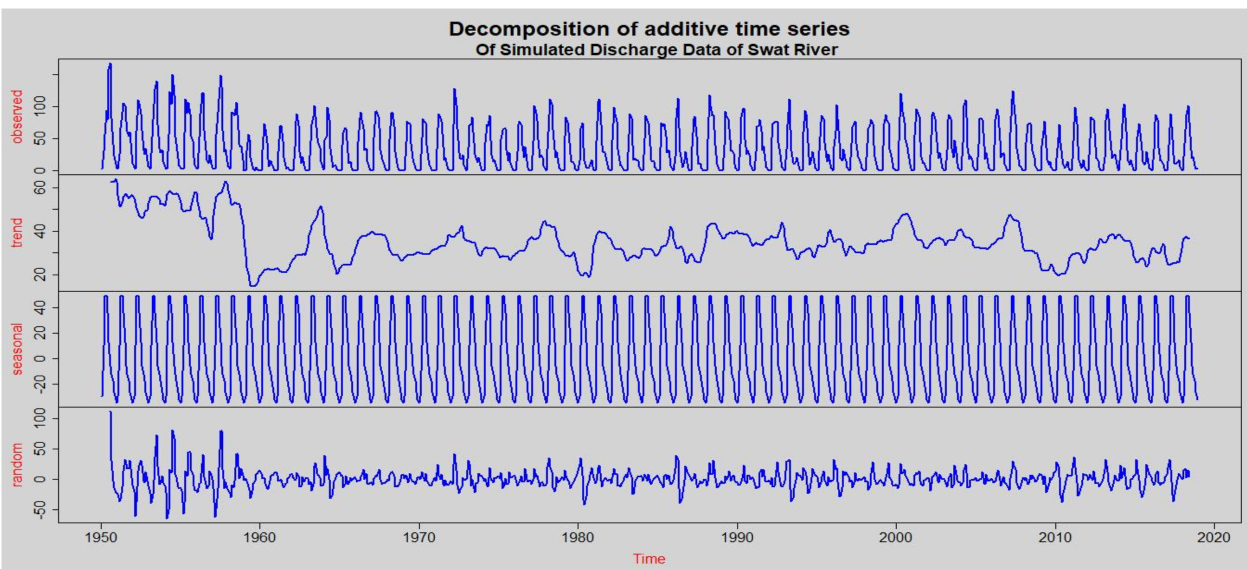


*Figure 14: Decomposition of additive time series of hydrological model*

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 the confidence in the conclusions drawn from the analysis.

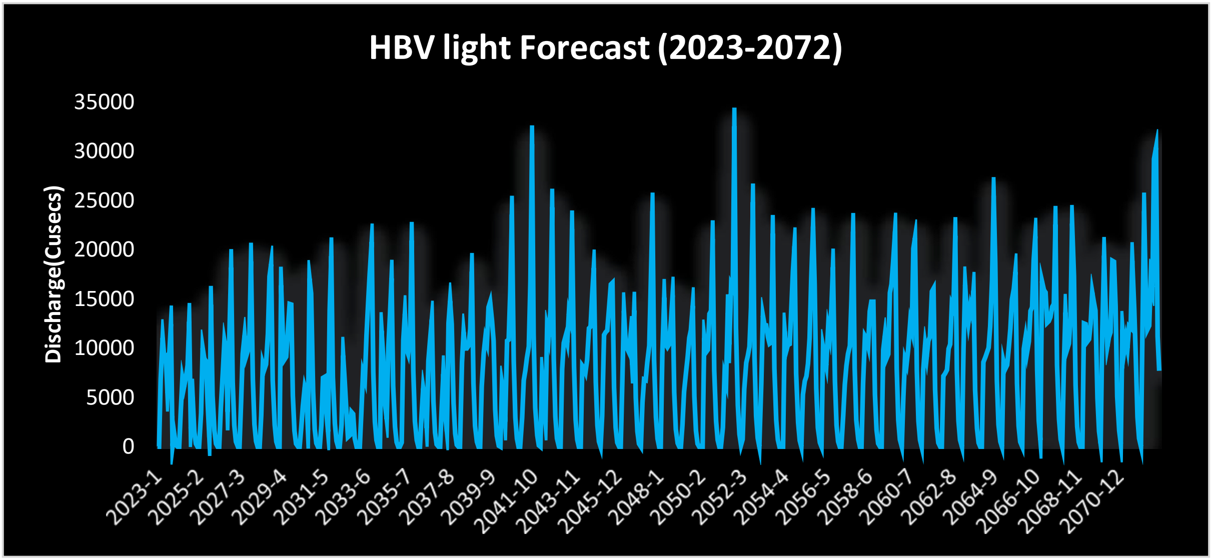


*Figure 15: Forecasted Simulated discharge.*

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.

## 3.5. Intercomparison b/w Different Models

### *3.5.1. Kling Gupta efficiency*

The Kling-Gupta efficiency (KGE) is a goodness-of-fit indicator widely used in the hydrological sciences for comparing simulations with observations. It was created by hydrologists Harald Kling and Hoshin Vijai Gupta. Its creators intended to improve widely used metrics such as the coefficient of determination and the efficiency coefficient of the Nash–Sutcliffe model.

### *3.5.2. Nash–Sutcliffe Efficiency*

The Nash–Sutcliffe Model Efficiency (NSE), also known as the Nash–Sutcliffe Coefficient of Efficiency (NSE), is a statistical measure used to evaluate the accuracy of hydrologic or hydraulic models. It is commonly used in hydrology, especially in the area of river flow modeling.

NSE compares simulated values from the model with values observed from real data. It quantifies how well the model predictions match the observed data.

### *3.5.3. Coefficient of variation*

The coefficient of variation (CV) is a statistical measure used to measure the relative variability ofa data set regardless of the units of measurement. This is particularly useful when comparing the variability of datasets with different means or units. Mathematically, it is calculated as the ratio ofthe standard deviation to the mean, expressed as a percentage:

$CV$=Standard Deviation Mean×100%CV=Standard Deviation×100%

A low coefficient of variation indicates that the data points tend to be close to the mean of the data set, while a high coefficient of variation indicates that the data points are spread over a wider range of values relative to the mean.

For example, if you have two data sets with different means and standard deviations, you can use the coefficient of variation to determine which data set has the higher relative variability, regardless of the scale of the data.

*Table 1: Intercomparison between forecasts of different models*

| PERIOD | | 2010-2018 | 2023-2073 |
|---|---|---|---|
| Mean (cusecs) | OBS | 6335.83 | - |
| | SIM | 6319.38 | 7825.62 |
| | ARIMA | - | 6483.72 |
| SD (cusecs) | OBS | 5340.37 | - |
| | SIM | 5320.26 | 6411.27 |
| | ARIMA | - | 5838.5 |
| CV (cusecs) | OBS | 0.84 | - |
| | SIM | 0.82 | 0.81 |
| | ARIMA | - | 0.9 |
| R b/w OBS & SIM | | 0.88 | - |
| R b/w SIM & ARIMA | | - | 0.89 |
| NSE OF SIM | | 0.76 | - |
| KGE OF SIM | | 0.88 | - |

# CHAPTER 4

# CONCLUSIONS AND RECOMMENDATIONS

## 4.1. Conclusions

The results and discussion section provide an overview of the primary analysis, decomposition of the time series data, model selection, auto-correlation factor (ACF) analysis of the ARIMA and neural network models, and the forecasted results. The primary analysis reveals a significant flood event in 2010 and consistent discharge patterns in subsequent years. The decomposition analysis shows a declining trend after the flood event and a potential upward trend in discharge levels post-2018. The presence of seasonal influences on discharge is observed. The ACF analysis suggests that the ARIMA model captures the weak autocorrelations in the data, making it suitable for modeling and forecasting. In contrast, the neural network model fails to adequately capture the temporal dependencies. The forecasted results show predicted increases in discharge with fluctuations or noise, and the anomaly graph successfully identifies flood events. The integration of anomaly detection aids in early warning and flood management. The application of the hydrological model validates the observed data, confirming the increase in discharge and enhancing understanding of the overall discharge trends.

The results of the project can be summarized as follows:

• The goal of moving from manual forecasting to AI techniques was pursued, evaluating two key models: ARIMA and HBV-Light forecasting. After a comprehensive analysis, the ARIMA model was found to be the best fit for the data set.

• Examination of ARIMA-derived data predicts a significant increase in flood frequency, suggesting the potential occurrence of more major floods in the future.

### 4.2. Recommendations

- The observed discharge was of very short period by which the accuracy of further forecasting got affected. Concerned authorities should formulate the proper system for discharge recording so that the affective Analysis could be done which may help in policy making.
- The Observed discharge is recorded in very raw form manually on Notebooks which is difficult to carry on with an. lysis. If observed data is legitimate and accurate, further analysis would be more accurate and beneficial.
- There is pattern in forecasting which is showing extreme discharge after every 10-11 years in 2041 & 2051 which is alarming and requires proper safety management. The Concerned authorities PDMA & NDMA should take necessary steps to avoid affects as we faced in past floods which very disastrous for our country.

# 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.