

Government of the People's Republic of Bangladesh  
Ministry of Water Resources  
Water Resources Planning Organization (WARPO)



Establishment of Water Quality Index (WQI) through Principal Component Analysis for the Dhaka-based Rivers

Inception Report  
October, 2023

**WARPO**  
পানি সম্পদ পরিকল্পনা সংস্থা

Water Resources  
Planning Organization



Department of Chemical  
Engineering, BUET

**Government of the People's Republic of Bangladesh**  
**Ministry of Water Resources**  
**Water Resources Planning Organization (WARPO)**

---

**Establishment of Water Quality Index (WQI) through Principal  
Component Analysis for the Dhaka-based Rivers**

**Inception Report**  
**October, 2023**



**Water Resources  
Planning Organization**



**Department of Chemical  
Engineering, BUET**

## Table of Content

**Table of Content i**

**List of Figures iii**

**List of Tables iv**

<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Background .....	1
1.2 Research Motivation .....	7
1.3 Research Objectives.....	9
1.4 Research Team Composition .....	9
1.5 Report Outlines .....	10
<b>Chapter 2: Water Quality Analysis (WQA): Theoretical and Statistical Approaches ....</b>	<b>11</b>
2.1 What is Water Quality Analysis .....	11
2.2 Different Approaches for WQA .....	13
2.3 Different Statistical Approaches for WQA .....	15
2.3.1 Principal Component Analysis (PCA).....	15
2.3.2 Factor Analysis (FA) .....	16
2.3.3 Cluster Analysis (CA) .....	16
2.3.4 Time Series Analysis .....	17
2.4 Why is principal component analysis superior? .....	17
<b>Chapter 3: Principal Component Analysis (PCA): Theory and Literature .....</b>	<b>19</b>
3.1 Introduction.....	19
3.2 Theory Behind PCA.....	19
3.2.1 Basic Statistics Behind PCA .....	19
3.2.2 Basic Linear Algebra Behind PCA.....	20
3.2.3 Performing PCA.....	22
<b>Chapter 4: Water Quality Index: Theory and Literature .....</b>	<b>24</b>
4.1 Introduction.....	24
4.2 Overview of Water Quality Index Modeling .....	24
4.2.1 Parameter Selection.....	25
4.2.2 Sub-indexing .....	26
4.2.3 Parameter Weighting .....	26
4.2.4 Aggregating Function.....	27

4.3 Existing WQI Models .....	29
4.3.1 The Horton Index .....	29
4.3.2 National Sanitation Foundation (NSF) WQI.....	30
4.3.3 The Scottish Research Development Department (SRDD) Index .....	31
4.3.4 Canadian Council of Ministers of Environment (CCME) WQI .....	32
4.3.5 Bascaron Index (BWQI).....	33
4.3.6 Fuzzy Interface System (FIS) Based Index.....	33
4.3.7 Malaysian Index (MWQI).....	34
4.3.8 West Java Index (WJ-WQI) .....	34
4.3.9 WQI Evaluation.....	34
<b>Chapter 5: Approach and Methodology .....</b>	<b>37</b>
5.1 Introduction.....	37
5.2 Site Selection .....	37
5.3 Data Collection .....	41
5.4 ArcGIS Modeling.....	41
5.5 Principal Component Analysis.....	42
5.6 Development of the Water Quality Index Model.....	44
5.7 Summary of Methodology .....	44
<b>Chapter 6: Work Plan and Progress .....</b>	<b>46</b>
6.1 Agreement Signing.....	46
6.2 Field Visit and Sample Collection .....	47
6.3 Future Work Plan and Reporting .....	49
<b>Chapter 7: References .....</b>	<b>52</b>
<b>Chapter 8: Appendices .....</b>	<b>65</b>
Appendix A: Research Proposal .....	65
Appendix B: Terms of Reference (ToR) .....	91
Appendix C: Site Selection for Sample Collection.....	94
Appendix D: Field Visit Images .....	102

## List of Figures

Figure 1-1: Peripheral Rivers of Dhaka City .....	1
Figure 1-2: The pollution scenario of Buriganga, Turag, Balu and Shitalakhya rivers .....	2
Figure 1-3: Motivation of the research project .....	8
Figure 2-1: General flow diagram of water quality analysis process .....	12
Figure 2-2: Implementation framework for prediction of water quality.....	17
Figure 4-1: General model structure of the WQI model.....	25
Figure 5-1: Sampling site locations for river .....	40
Figure 5-2: Steps involving in PCA using R.....	43
Figure 5-3: Flowchart for overall methodology.....	45
Figure 6-1: Contract signing between WARPO and Dept. of ChE, BUET .....	46
Figure 6-2: Members present in sample collection at Buriganga (BG4).....	47
Figure 6-3: Members present in sample collection at Turag river (TR1) .....	48
Figure 6-4: Members present in sample collection at at Balu river (BL1) .....	48
Figure 6-5: Members present in sample collection at at Shitalakhya river (SL1) .....	49
Figure 8-1: Field visit at Buriganga (BG4).....	102
Figure 8-2: Field visit at Turag (TR1).....	103
Figure 8-3: Field visit at Balu (BL1) .....	104
Figure 8-4: Field visit at Shitalakhya (SL1) .....	105

## List of Tables

Table 1-1: Seasonal variation in Bangladesh .....	3
Table 1-2: Parameters used in previous studies and their findings .....	4
Table 1-3: Research team composition .....	9
Table 2-1: Types of Water Quality Parameters .....	12
Table 4-1: Different parameter weight value used in different studies .....	31
Table 4-2: Different WQI Model Evaluation Classes .....	34
Table 4-3: WQI parameters used in different WQI models .....	36
Table 5-1: Site selection reasons .....	38
Table 5-2: Different considerable parameters and analyzing instruments/methods .....	41
Table 6-1: Activity Schedule starting from June, 2023.....	50
Table 8-1: Sampling sites with coordinates .....	94

# Chapter 1: Introduction

## 1.1 Background

River water is one of the most useful resources to people because of its multipurpose uses including domestic, industrial and agricultural in addition to transportation, tourism and recreational purposes. Dhaka division is one of the eight administrative divisions of Bangladesh which includes the Dhaka city, the capital of Bangladesh. The current population of the city is about 20 million, which turns Dhaka city into one of the most densely populated megacities in the world. Considering the importance of rivers to humans, Dhaka city has grown on the bank of Buriganga from the era of the Mughal empire and is surrounded by four rivers (Figure 1-1). There are Balu and Sitalakhya on the eastern side, Turag and Buriganga on the western side. These rivers receive water from the Jamuna (Brahmaputra river) in the wet season, and in the dry season upper reaches of these rivers are slowly replenished by the release of groundwater into the rivers. The lower reaches of the rivers are also influenced by the tidal variations travelling upstream from the Bay of Bengal. In the monsoon season, the river levels reach around 6.5m MSL (mean sea level) and drop to about 2.5m MSL in the dry season (BWDB, 2023). However, the unplanned development of the city and the establishment of different industries on the banks of the rivers over time resulted in a decrease in the water quality of the rivers.

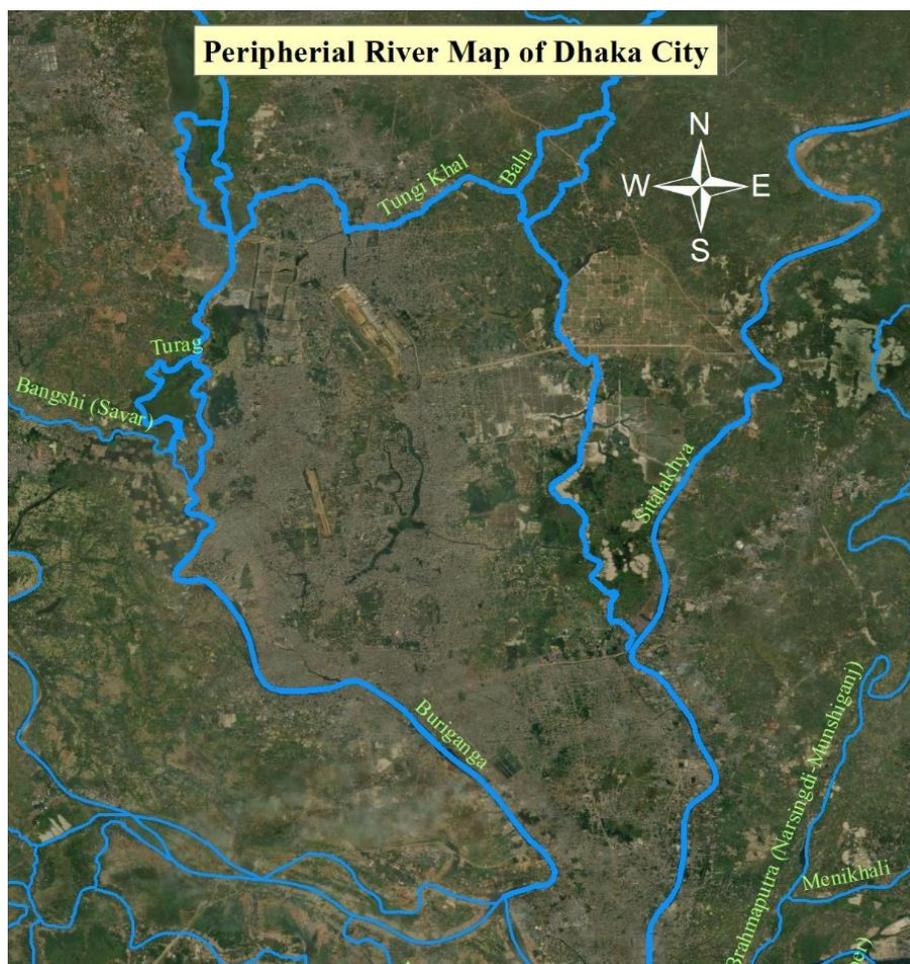
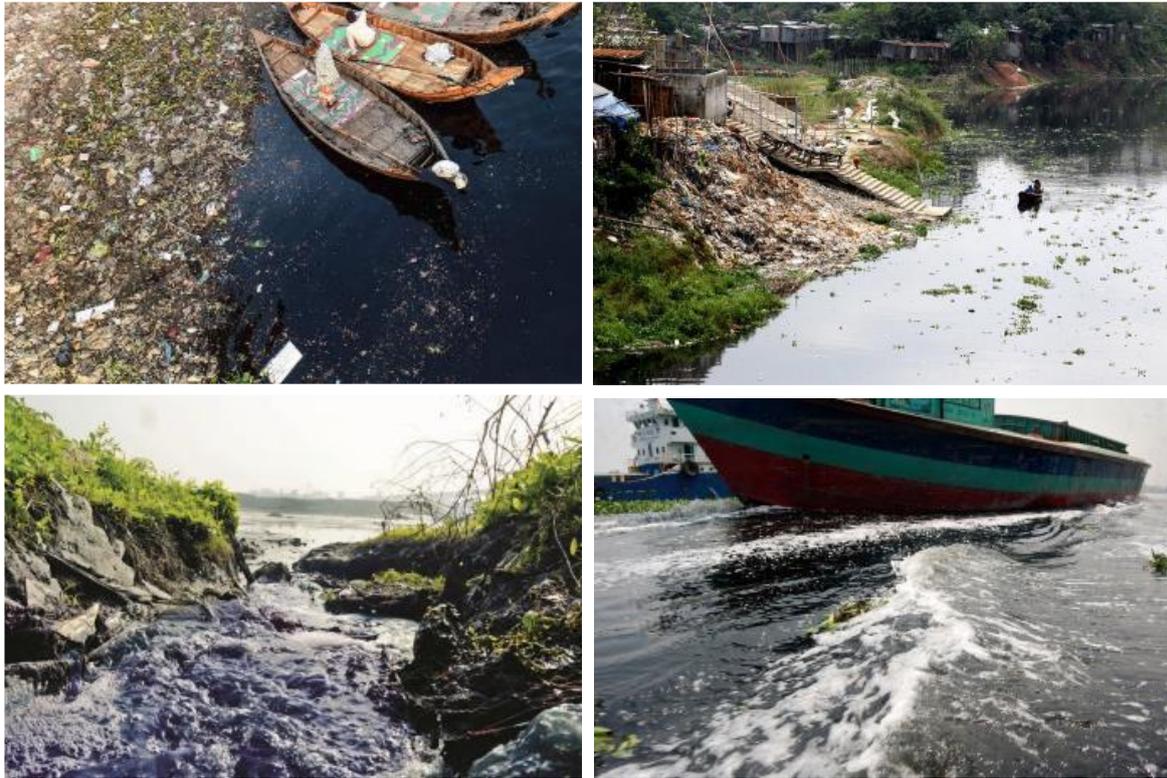


Figure 1-1: Peripheral Rivers of Dhaka City



*Figure 1-2: The pollution scenario of Buriganga, Turag, Balu and Shitalakhya rivers*

There are a few studies available on the Dhaka-based river water quality assessment, which were collected and reviewed. Mir Mostafa Kamal et al. reported the alarmingly low level of DO in the Buriganga river back in the 2000s (Kamal et al., 1999). Many other researchers studied water quality parameters e.g., dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), pH, turbidity, conductivity, total dissolved solids (TDS), nitrate and phosphate in Buriganga river and those studies found the DO, BOD, COD, TDS, turbidity, nitrate and phosphate are at an alarming level and a discussion on the possible sources of the pollution are presented in some of the papers (Akbor et al., 2017; Fatema et al., 2018). A study indicates the pollution sources of Buriganga rivers are both point and non-point, making the river water quality highly unhealthy from the perspective of aquatic ecosystems (Salman et al., 2018). Another study on the water quality assessment of the Buriganga river conducted by Md. Ashiqur Rahman et al. discussed the effect of different water quality parameters and concluded most of the parameters are not acceptable according to the allowable limit during both dry and wet seasons (M. A. Rahman & Al Bakri, 2010). Shaikh et al. studied 10 parameters for 10 different sampling stations of the Buriganga river to evaluate the water quality and found unsatisfactory results for each station (Sayed Ahammed et al., 2016). Also, there have been researches done on other peripheral rivers of Dhaka city regarding water quality assessment (Naushad Alam et al., n.d.; Nazma Sultana et al., 2019; A. Rahman et al., 2012).

One of the most important objectives of water quality assessment studies is assessing the seasonal variation which has been addressed in most of the previous works. Several government and international agencies reported four seasons in Bangladesh. Dry/Winter, Pre-Monsoon, Monsoon and Post-Monsoon seasons. Table 1-1 indicates the above-mentioned four seasons in different literature from government and international agencies.

*Table 1-1: Seasonal variation in Bangladesh*

Season Name	Duration	Month Span	Agency Names	Reference
Dry/Winter	3 months	December-February	WARPO	(Ministry of Water Resource, 2001)
Pre-Monsoon	3 months	March-May	BMD	(Khatun et al., n.d.)
Monsoon	4 months	June-September	FAO	( <i>Country Profile-Bangladesh</i> , 2014)
Post-Monsoon	2 months	October-November	World Bank	( <i>Country Profile Bangladesh COUNTRY OVERVIEW</i> , 2011)

But in contrast to these reports, several scholars have studied the water quality of rivers in Bangladesh mentioning different timespan for different seasons. For example, a study on water quality assessment of Sitalkhya river was conducted mentioning the seasonal variation of two seasons, dry and wet (Naushad Alam et al., n.d.). However, there was no specific month or duration mentioned about the seasonal span. Other studies were conducted in the peripheral rivers of Dhaka city, especially Buriganga and Turag, and most of them mentioned dry and wet seasons (Fatema et al., 2018; Kamal et al., 1999; A. Rahman et al., 2012; S. Rahman & Hossain, 2008). Some of them specified the seasonal span, for instance, Fatema et al. mentioned two seasons, dry (November – January) and wet (June – August) (Fatema et al., 2018) from which existence of a total of four seasons in Bangladesh can be inferred and these two seasons can be pre-monsoon and post-monsoon mentioned in the *Development of an Assessment System to Evaluate the Ecological Status of Rivers in the Hindu Kush-Himalayan Region* (ASSESS-HKH) project (Moog & Scientific Conference Rivers in the Hindu Kush-Himalaya - Ecology & Environmental Assessment 2008 Kāthamāṇḍū; Dhulikhel, n.d.). In contrast, Rahman et. al. mentioned in their study about seasons covering 12 months of the year – dry season (December to May) and the wet season (June to November) (A. Rahman et al., 2012). Another Study based on the Buriganga river mentioned three conventional seasons – summer, winter and autumn and specified the months (Sayed Ahammed et al., 2016). Since there is no specified seasonal span, these studies differed from each other, but all these studies concluded a generalized result that water quality highly depends on seasonal variation. In this current work, we are aiming to establish a seasonal variation in water quality in accordance with the national and international reports regarding the seasons.

Table 1-2 represents a summary of the parameters used in different studies along with their findings.

*Table 1-2: Parameters used in previous studies and their findings*

Study Name	Parameters Studied	No. of Sampling Stations and Seasons	Approach and Major Findings	References
<b>Assessment of Pollution of the River Buriganga, Bangladesh, Using a Water Quality Model</b>	DO, BOD <sub>5</sub> , COD, NH <sub>3</sub> -N, NH <sub>4</sub> -N, E. Coli, Total Coliform, TSS, NO <sub>3</sub> -N, Orth.PO <sub>4</sub> -P, Cr	Sampling was done in 1 Station during 1 Season (Dry)	Low DO and higher values of other parameters were found with an occasional rise to DO and low concentration of other parameters due to local rainfall and during the spring time period.	(Kamal et al., 1999)
<b>Investigation of Water Quality Parameters at Different Points in the Buriganga River, Bangladesh</b>	EC, TDS, Salinity, pH, DO, BOD, COD, Hardness, Acidity, Free CO <sub>2</sub> , Alkalinity, Turbidity	Sampling was done in 8 Station during 1 Season (Dry)	High BOD, COD, EC, TDS, Salinity, Alkalinity and Turbidity with low DO indicating very critical water condition of river Buriganga.	(Akbor et al., 2017)
<b>Water Quality Assessment of the River Buriganga, Bangladesh</b>	Temperature, DO, pH, Conductivity, Phosphate, Mn, Fe, Pb, Cd	Sampling was done in 3 Station during 2 Season (Wet and Dry)	Significant temperature difference depends on the months, not on the sampling stations. Differences in pH, DO, Phosphate etc. are significant between dry and wet seasons.	(Fatema et al., 2018)
<b>Water Quality Assessment of the Buriganga River, Dhaka, Bangladesh</b>	pH, Temperature, Alkalinity, Color, Turbidity, Chloride, Iron, BOD	Sampling was done in 4 Station	All the water quality parameters measured fall below the standard and acceptable reason. Parameter values change from Hazaribag to Postogola.	(Salman et al., 2018)
<b>A Study on Selected Water Quality Parameters along the River</b>	Temperature, pH, EC, DO, BOD, COD, PO <sub>4</sub> -P, NH <sub>3</sub> -N, Pb, Cr	Sampling was done in 5 Station during 2 Season (Wet and Dry)	The observed DO was lower than the standard level for all sampling stations. Low DO could	(M. A. Rahman & Al Bakri, 2010)

<b>Buriganga, Bangladesh</b>			be linked to higher turbidity.	
<b>Water Quality Assessment of Balu River, Dhaka Bangladesh</b>	pH, EC, Fe, Cd, Pb, Cr	Sampling was done in 1 Station during 1 Season (Dry)	A significant amount of heavy metals was found in the samples and the overall quality of the river is found to be threatening.	(Nazma Sultana et al., 2019)
<b>Study of the seasonal variations in Turag river water quality parameters</b>	pH, EC, Salinity, TDS, DO, BOD, COD, Free CO <sub>2</sub> , Hardness, As, Cd, Pb, Zn, Hg	Sampling was done in 3 Station during 2 Season (Wet and Dry)	This study was conducted to compare the seasonal effect on water quality parameters in dry and wet seasons. Higher BOD and COD were found during the dry season. TDS, EC and Hardness also increased during the dry season which fell below the standard limit.	(A. Rahman et al., 2012)
<b>Risk and Water Quality Assessment overview of River Sitalakhya in Bangladesh</b>	DO, pH, EC, TDS		Studied the parameter trend over the years in the Shitalakhya river. DO tends to decrease and TDS and EC are increasing each year.	(Naushad Alam et al., n.d.)
<b>An Investigation Into The Water Quality Of Buriganga - A River Running Through Dhaka</b>	DO, BOD, COD, pH, Temperature, Conductivity, Turbidity, TDS, Nitrate, Phosphate	Sampling was done in 10 Station during 3 Season (Summer, Winter and Autumn)	This study measured water quality parameters of selected sites and compared them with other literature values. Based on the parameter values probable point sources of pollution were indicated and during the dry season, most of the parameters deteriorated than the wet season.	(Sayed Ahammed et al., 2016)

Paul Whitehead et al. (2018) and his team performed a baseline survey of water chemistry and total coliforms of the Turag-Tongi-Balu river system and showed DO close to zero in the dry season, high organic loading together with extreme levels of Ammonium-N and total coliform in the water (Whitehead et al., 2018). However, most of the studies are scattered, non-

comprehensive, dealing with several parameters and limited sites of the river. All these mentioned surface water quality assessment studies focused on a singular evaluation of parameters individually, no study was conducted to combine the important parameters and represent them as a single entity to evaluate water quality. These studies emphasized more insights, causes and consequences of river water pollution and thus, these studies could not reach the policymakers. In order to introduce long-term and effective river water management, focused and structured research is necessary to understand the actual condition of the study area (baseline). In order to recognize the Dhaka-based river water quality, a good source of data on the water quality is indispensable. Although the regular measurement of water quality parameters is a laborious job because of spatial and temporal variations of water environment quality along the river, it will provide a representative and reliable estimation of surface water quality. The long-term monitoring of many profiles with different reach would generate a large and complex database, which needs a good approach to interpretation (Kabir et al., 2022). To identify and mitigate the pollution source, the correct interpretation of the collected data is crucial. Moreover, random site selection of study areas is also an issue in water quality assessment. The geographical information system (GIS) is very helpful in evaluating the spatial distribution of water quality parameters over the study area (Oseke et al., 2021). The collected data can be analyzed by the application of different multivariate statistical techniques, such as Principal Component Analysis (PCA), Cluster Analysis (CA), and multiple linear regression. These techniques assist in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems, allow the identification of possible factors that influence water environment systems and offer a valuable tool for the reliable management of water resources. Assessment of river water quality or optimization of the monitoring procedure of river water is linked to clustering of sampling locations, river water quality parameters, identification of possible sources of pollution, or modeling the contribution of the identified sources to the formation of the total concentration of the monitored chemical tracers.

Most of the time, the research on water quality assessment fails due to the thousands of output information because water quality regulations and monitoring programs generate a vast number of multidimensional water quality parameter sets, pollution characteristics, and numerical data about various water sources that are understandable only to the scientists (Oketola et al., 2013). This information, however, should be beneficial to the authorities of the water sector in decision-making to restore the water quality. As a result, a technique known as the Water Quality Index (WQI) has been developed to solve this issue. The Water Quality Index is a quantitative representation that is used to determine the ecological health of a body of water. The purpose of the Water Quality Index is to categorize waters according to their physical, chemical, and biological attributes, thereby establishing their potential uses and controlling their allocation decisions (Pesce & Wunderlin, 2000). By providing a single dimensionless value, the WQI helps to reduce the multivariate nature of the data that describes the quality of water (Tharmar et al., 2022) and specify the impact of seasonal variation in water quality. Because of its generalized structure, it has been a popular choice since the 1960s when it was first introduced (M. G. Uddin et al., 2021). More than 35 WQI models have been developed or introduced by different countries or agencies for the assessment/evaluation of surface water quality till now (Abbasi & Abbasi, 2012a; Dadolahi-Sohrab et al., 2012; Kannel et al., 2007; Stoner et al., n.d.).

Thus, proper site selection, proper data collection, correct analysis of the collected data, and indicative results all should be a part of a river water quality assessment project. Otherwise, the research would be limited to the researcher only. In this research, a comprehensive study will be carried out to understand four main river water characteristics using ArcGIS, numerical Principal Component Analysis (PCA) and WQI analysis.

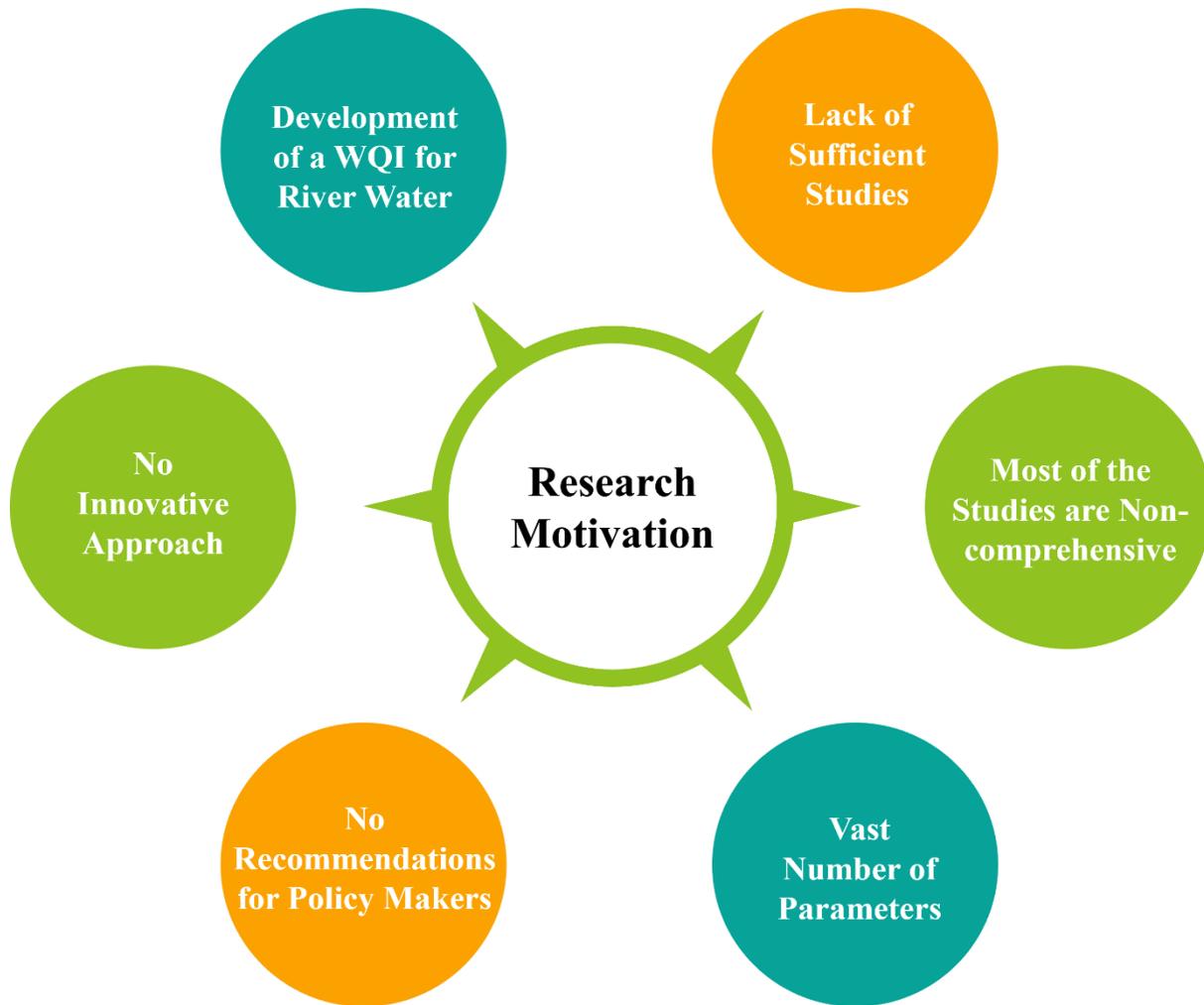
## 1.2 Research Motivation

The water quality of rivers, a vital component of urban ecosystems, plays a pivotal role in sustaining both environmental health and human well-being. Despite the undeniable importance of this subject, the current state of research on water quality in Dhaka-based rivers reveals several critical gaps that necessitate immediate attention and comprehensive investigation.

- 1) **Lack of Sufficient Studies:** One of the primary motivations for this research stems from the apparent lack of sufficient studies addressing the water quality of Dhaka-based rivers. Existing literature reveals a paucity of in-depth examinations into the multifaceted factors influencing water quality, highlighting the urgent need for a more thorough investigation.
- 2) **Most of the Studies are Non-comprehensive:** Existing studies on water quality in the region often fall short in providing a comprehensive analysis. Many have focused on a limited set of parameters, neglecting the intricate interplay of various factors that collectively define water quality. This research aims to address this limitation by adopting a more holistic approach, considering a broader spectrum of influential variables.
- 3) **Dealt with a Vast Number of Parameters:** Contrary to the non-comprehensive nature of some studies, others have attempted to tackle water quality by incorporating an extensive array of parameters. However, this approach, while well-intentioned, often leads to information overload and lacks a systematic method to distill the most crucial factors. Our research seeks to streamline this process through the application of Principal Component Analysis (PCA), providing a more nuanced understanding of the interrelationships among diverse parameters.
- 4) **No Recommendations for Policy Makers:** Despite numerous studies on water quality, a notable gap exists in translating research findings into actionable recommendations for policymakers. Our research aspires not only to identify areas of concern but also to offer concrete suggestions and guidelines for policymakers to formulate effective strategies for water quality management in the Dhaka region.
- 5) **No Innovative Approach:** The absence of an innovative approach in current studies underscores the need for novel methodologies that can offer fresh perspectives and insights. By integrating Principal Component Analysis into the establishment of a Water Quality Index (WQI), our research introduces an innovative framework, fostering a more sophisticated understanding of the complex dynamics influencing river water quality.
- 6) **Development of a WQI for River Water:** Finally, our research is motivated by the aspiration to develop a Water Quality Index (WQI) for Dhaka-based rivers. Inspired by the success of the Air Quality Index (AQI), which has become a widely accepted and accessible

tool, a similar metric for water quality is crucial for raising public awareness, aiding policymakers, and facilitating informed decision-making in the context of environmental management.

In summary, the identified gaps and limitations in existing research on water quality in Dhaka-based rivers underscore the significance of this study. By addressing these shortcomings and introducing an innovative approach, our research aims to contribute substantially to the understanding and management of water quality, with the ultimate goal of establishing a practical Water Quality Index for the region. Figure 1-3 points out the major motivation behind this research study.



*Figure 1-3: Motivation of the research project*

### 1.3 Research Objectives

The main objective of the research is to establish a comprehensive method to determine the Water Quality Index (WQI) for the Dhaka-based rivers. On a pilot basis, the research would concentrate on the selected reaches of the peripheral rivers of Dhaka city - Buriganga, Turag, Balu and Shitalakshya rivers.

The following specific objectives have been set to achieve the goal of this research work:

- To understand the baseline water quality parameters of the Buriganga, Turag, Balu and Shitalakshya rivers.
- To identify the most critical water quality parameters for the selected rivers using principal component analysis (PCA).
- To establish a method/approach, consistent with the internationally adopted approach, to calculate the water quality index (WQI) for the selected rivers.
- To categorize and classify the selected rivers considering seasonal variation and based on their respective WQI values.

### 1.4 Research Team Composition

This is a collaborative research project with the participation of the Department of Chemical Engineering (DChE), BUET and WARPO. BUET team consists of Principal Investigator, 2 Research Assistants and 1 professional with relevant background. WARPO team consists of 4 professionals with relevant backgrounds. The Research Team composition is shown in Table 1-3.

*Table 1-3: Research team composition*

Designation	Designation, Institution and Position in the Research Team
<b>Prof. Dr. Md. Shahinoor Islam</b>	Professor, DChE, BUET & Principal Investigator
<b>Kazi Saidur Rahman</b>	Senior Scientific Officer, WARPO & Principal Research Coordinator
<b>Md. Al-Amin Kabir Bhuiyan</b>	Scientific Officer, WARPO
<b>A. M. Zoraf</b>	Scientific Officer, WARPO
<b>Shuvro Bhowmick</b>	Scientific Officer, WARPO
<b>Hridoy Roy</b>	Lecturer, DChE, BUET
<b>Bimol Nath Roy</b>	Research Assistant, DChE, BUET
<b>Foysal Mahmud</b>	Research Assistant, DChE, BUET

## 1.5 Report Outlines

The first chapter includes the background of the research, objectives and organization of the report. This chapter describes the necessity of this research on national level. The research team composition is also described in this chapter.

The second chapter includes an extensive literature review which describes the concept of water quality analysis and different types of statistical approaches for water quality analysis. Moreover, an in-depth discussion on different statistical approaches for water quality analysis has been thoroughly analyzed.

The third chapter describes the basic theory and concept of basic statistics and linear algebra required for principal component analysis (PCA), factory analysis (FA), cluster analysis (CA), and time series analysis. In addition, a chronological discussion has been made which establishes the feasibility of PCA over other methods. Along with the theoretical understanding, this chapter also includes how to perform PCA in any data set.

The fourth chapter comprises the theories and literature reviews of water quality index (WQI). Overview of general model structure and most commonly used WQI models which are discussed briefly along with their evaluation process.

The fifth chapter includes the approaches and methodology of the study. The methodology comprises of site selection, data collection, ArcGIS modeling, sample testing and analysis, PCA and WQI development.

The sixth chapter mostly covers the progress of the research activities in line with the work plan, This chapter also includes the agreement signing, field visit and sample collection images, future work plan and activity schedule. The last chapter includes the references. And the final part of the report is the appendices which include the research proposal, terms of references, sampling sites etc..

## **Chapter 2: Water Quality Analysis (WQA): Theoretical and Statistical Approaches**

### **2.1 What is Water Quality Analysis**

Water quality analysis (WQA), also called hydrochemical analysis, is the assessment and evaluation of the physical, chemical, and biological properties of water to determine its overall quality and appropriateness for specific uses. It includes the collection, measurement, and interpretation of numerous parameters and indicators to evaluate the safety level of the water sources. Water quality analysis is crucial for a variety of reasons, including human health, environmental impact, regulatory compliance, industrial and agricultural processes, early prediction of potential pollutions, etc. Contaminants such as bacteria, viruses, heavy metals, pesticides, and chemical pollutants can have adverse effects on human health as well as industrial operations (Chowdhary et al., 2020). Water quality analysis helps to identify and quantify these contaminants to assess the potential risks associated with water consumption. Industrial wastewater discharges, agricultural runoff after pesticide and fertilizer usage, and improper municipal solid waste disposal can introduce toxic elements into water bodies, leading to ecological imbalances and danger to aquatic life (Garg et al., 2022). Industries and agricultural activities often require a significant amount of water usage. In industries, substandard water quality may cause product and safety hazards resulting in extreme economic loss. Water quality analysis is therefore crucial in these sectors to ensure that the water used is suitable for specific processes, such as manufacturing, irrigation, or livestock consumption. WQA helps identify any potential issues that could affect the efficiency of industrial processes or harm agricultural productivity. Governments and regulatory bodies enforce water quality regulations and guidelines to safeguard public health, the ecosystem, and the climate. Regular monitoring and analysis help identify potential violations and ensure maintaining water quality within acceptable limits. Improvement in water management and pollution control requires accurate predictions of water quality. A lack of information management contributes to the occurrence of numerous water pollution incidents. Therefore, the results of WQA help inform decision-making regarding water treatment, resource management, pollution control, and environmental conservation.

Water quality analysis starts with different laboratory testing and field measurements. Analytical techniques employed in water quality analysis include spectrophotometry (Hudson et al., 2008), chromatography (Kasiske et al., 1978), titration (Belle-Oudry, 2008),

microbiological assays (Bonadonna et al., 2019), and molecular techniques (Girones et al., 2010). Figure 2-1 shows the general flow diagram of the WQA process (Roy, 2018).

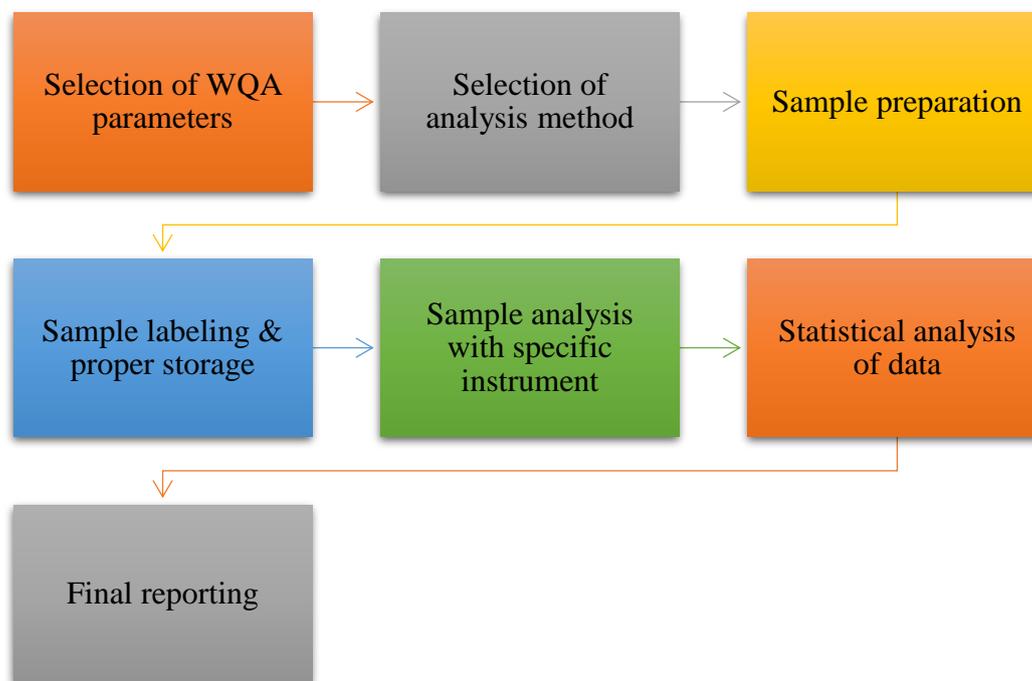


Figure 2-1: General flow diagram of water quality analysis process

Water quality indicators consist of physical (Duclos Alege & Gnauck, 2006; Huey & Meyer, 2010; Pinto et al., 2013; Toming et al., 2016; Uthicke et al., 2012), chemical (Kannel et al., 2007; Maskooni et al., 2020; Sánchez et al., 2007; Ustaoglu et al., 2020), and biological (Akay & Dalkıran, 2020; Al-Afify et al., 2019; Bigham et al., 2019; Summers & Summers, 2020) variables that provide explicit signals about the status and changes of the water system. The evaluation of these indicators is done through field monitoring of water bodies that provide important data for identifying water quality trends, providing relevant water authorities with information on water quality, and recommending future actions. Several essential freshwater quality indicators are provided in Table 2-1.

Table 2-1: Types of Water Quality Parameters

Physical Indicators	Chemical Indicators	Biological Indicators
Temperature	pH	Escherichia coli (E. coli)
Suspended Solids	Dissolved Oxygen	Coliforms
Turbidity	Electrical Conductivity	Biological Oxygen Demand (BOD)
Solar Radiation	Nutrients	Aquatic Macroinvertebrates
Color	Heavy Metal	

Physical indicators, e.g., temperature can be useful in measuring the degree of hotness or coldness of water, which can affect the solubility of substances and influence aquatic organisms (Bozorg-Haddad et al., 2021). Turbidity indicates the clarity or cloudiness of water caused by suspended particles, which can affect light penetration and restrain photosynthesis in the bottom layer of water bodies (Donald et al., 2019). Color can be useful as an indicator for the visual appearance of water, which can be influenced by natural substances or contaminants. The acidity or alkalinity of water is measured using a chemical indicator, such as pH. The solubility of minerals and the survival of aquatic organisms are both impacted by it, making it crucial. Conductivity, on the other hand, provides an estimate of its dissolved ion content and salinity. Dissolved Oxygen (DO) determines the amount of oxygen available in water to support aquatic life. Low DO concentrations may represent pollution or an excess of organic matter (Cooper, 1993). Certain levels of nitrogen, phosphorus, and other nutrients can promote excessive plant growth (eutrophication) and cause water quality issues. Biologically accessible nutrients and hazardous chemicals in excess can result in a number of disorders, e.g., toxic algal blooms, oxygen depletion, fish mortality, biodiversity loss, and the devastation of aquatic plant beds (Anderson et al., 2021). Again, effluents from municipal, agricultural, and industrial streams may contain harmful heavy metals such as lead (Pb), chromium (Cr), nickel (Ni), cadmium (Cd), iron (Fe), arsenic (As), and mercury (Hg) (Khan et al., 2023). Total Coliforms and *Escherichia coli* (*E. coli*) are biological indicators that suggest the potential presence of harmful pathogens in water. Aquatic Macroinvertebrates examine the diversity and abundance of small organisms living in water bodies that can provide insights into the overall ecological health and water quality. Globally, microbial indicators have been used to determine if a water body has been contaminated by human refuse. Typically, microorganisms found in high concentrations in human feces are utilized. In the United States, the most common indicators are total coliform, fecal coliform, *Escherichia coli*, and enterococci (Shibata et al., 2004).

It is particularly noteworthy that the specific indicators analyzed can differ based on the desired application of water (e.g., water for drinking, recreational purposes, industrial usage) and the regulatory standards applicable in a particular region. Therefore, water quality analysis often involves multiple indicators to obtain an accurate picture and extensive understanding of water quality and detect probable issues or origin of pollution.

## 2.2 Different Approaches for WQA

The most important natural resource is water. Accepting the significance and scarcity of resources required to meet biological needs and support all forms of economic and growth activity is of the utmost importance. One of the crucial concerns in the management of water resources is water quality. Water can be classified into two groups depending on the source: one is the groundwater, and another is the surface water. A variety of contaminants, including heavy metals, pesticides, fertilizers, toxic chemicals, and oils, may present as a result of domestic, industrial, and agricultural operations, which could result in the contamination of water (Omer & Omer, 2019). According to Swamee and Tyagee (2007) (Swamee & Tyagi, 2007), there are several metrics for each of the three major categories for assessing water quality - physical, chemical, and biological. By field monitoring of rivers, these three categories are evaluated. This information is used to identify patterns, notify water authorities about the

quality of the water, and suggest future courses of action. The normal method of conducting this assessment is to consider the planned applications, human health, and natural water quality (Gazzaz et al., 2012; Pesce & Wunderlin, 2000). Water quality could be evaluated either spatially or temporally, as per Rosemond et al. (2009). However, measuring the various water quality indicators and comparing the measured values to threshold values are two essential components of water quality assessment. The threshold value denotes the maximum or minimum variable concentration that is safe for consumption by the general public.

Researchers have been focused on hydrochemical analysis in various ways for decades. The Piper diagram has been widely used to examine the groundwater facies in order to advance study on topics like disclosing the evolution of phreatic water and comprehending the hydrochemical properties as well as the mechanism of groundwater production (Wang et al., 2019; Qingchun Yang et al., 2016). Traditionally, hydrochemical evaluation of groundwater was based on laboratory study, but for the last 20 years, geosciences research has been conducted in a variety of domains utilizing geospatial approaches, which have the advantage of monitoring and integrating many thematic levels with ease, precision, and on time (Ali & Ali, 2013; Ali & Pirasteh, 2004). Ahmad Ali et al. (2017) (Ali & Ali, 2018) analyzed the chemical variations in groundwater under diverse natural and anthropogenic activities, including the spatial distribution of hydrochemical parameters. The adequacy of groundwater quality for future domestic, agricultural, industrial, and drinking uses was also examined in this study. Basic techniques in hydrochemical research for compiling and presenting data on water quality include graphical and numerical interpretations. There are a large number of methods for categorization, correlation, analysis, and illustration (Sharma et al., 2012; A. K. Singh et al., 2008). In the past, water quality reports were generated using trend analysis utilizing a single variable on particular sample locations. Trend analysis can be done using both parametric and nonparametric tests.

Since the 1960s, water quality indices (WQIs) have been used as a technique to assess the condition of river water quality. WQI has grown in importance and popularity as a vital and widely used tool for evaluating the water quality of water bodies around the world, particularly rivers, due to its simplicity of use and scientific foundation. Since the introduction of the WQI idea, numerous scholars have created and developed numerous indices. Additionally, WQIs have been seen as a crucial component of broader environmental or natural resource indices such as the Environmental Performance Index (EPI 2010) (*2010 ENVIRONMENTAL PERFORMANCE INDEX*, n.d.). Canadian Council of Ministers of the Environment Water Quality Index (2001) and Oregon Index (Dhany Sutadian et al., n.d.) are the widely used indices for the purpose. The study of water quality has given considerable attention to the use of various multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), and discriminant analysis (DA), which can assist in the interpretation of complex data matrices to better understand the water quality and ecological status of the studied systems and allow the identification of potential factors/sources that influence water quality (Q. Yang et al., 2015; Qingchun Yang et al., 2015)

## 2.3 Different Statistical Approaches for WQA

Water quality data usually involves a large number of measurements. These data serve as a foundation for plant operation, modeling the process, treatment planning, and economic assessments. In national and global contexts, there is a growing demand for methods to integrate several water quality-related variables into a single index. Globally, many scientific methods have been used to assess and monitor water quality. For detecting the origins of physical and chemical features and their associations, as well as disseminating information about the overall quality of the water, multivariate statistical techniques have recently gained popularity (Kazi et al., 2009). Multiple authors have incorporated water quality variables into indices, formally known as WQI. Increasingly, computational methods based on artificial intelligence have been applied to environmental problems with inherent uncertainties and subjectivities (Li et al., 2019). However, conducting different statistical analyses on the same data may significantly increase the statistical power and acceptance of the method.

Even though many WQIs have been designed, there is no widely accepted way to carry out the WQI development process globally (Dhany Sutadian et al., n.d.). Many researchers found it ambiguous to define WQI and convey it in a concise and unified manner. Due to the multifaceted nature of the elements or indicators that affect water quality and their wide range of value, this challenge became prominent. Since having several measurements in the data may affect the prediction and accuracy, popular multivariate statistical techniques, e.g., principal component analysis (PCA), factor analysis (FA), cluster analysis (CA), and time series analysis are widely used to group the important variables and are reviewed in this study.

### 2.3.1 Principal Component Analysis (PCA)

PCA is predominantly used to identify trends and patterns in a complex dataset by converting a set of potentially correlated observations into an entirely novel set of linearly uncorrelated variables, known as principal components. PCA is primarily employed for exploratory data analysis and the development of models for forecasting (Jolliffe & Cadima, 2016). It aims to explain as much variance as possible with the lowest possible number of principal components while maintaining a minimum information loss (Zou et al., 2006). When there are countless predictors in relation to the number of data, PCA has the potential to scale down the number of factors significantly (Chan et al., 2022).

In terms of mathematics, PCA consists of the following five main steps: (1) ensure that all of the measures are given the same weight in the statistical evaluation, defining the elements  $x_1, x_2, x_3, \dots, x_p$  with zero means and unit variance, (2) estimating the covariance matrix  $C$ , (3) determining the eigenvalues  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$  and the associated eigenvectors  $a_1, a_2, a_3, \dots, a_p$ , (4) discarding elements that merely contribute a minor amount to the variance in the datasets, and (5) developing the factor loading matrix to determine the principal parameters (Chauhan & Sharma, 2003).

PCA is the most basic form of multivariate analysis based on true eigenvectors (McGarigal et al., 2000). It has a close relation to factor analysis, which is related to canonical correlation analysis (CCA) (Matthew et al., 1994). PCA defines a new orthogonal coordinate system that

optimally describes the variance in a single dataset, whereas CCA defines coordinate systems that optimally describe the cross-covariance between two datasets. Recent years have seen the application of the PCA model to a wide range of sustainability problems, such as the assessment of groundwater quality monitoring, pollution source determination in ambient soil and air, heavy metal detection, climate change analysis, etc. (Olsen et al., 2012; Ying Ouyang, 2005).

Ouyang et al. (2006) analyzed river water quality using PCA to assess seasonal correlations along the main stem in Florida, United States (Y. Ouyang et al., 2006). PCA and FA were conducted using statistical analysis system (SAS) software. Results from the analysis indicated that the importance of any water quality indicator, e.g., temperature, total alkalinity, color, and BOD depends significantly on the particular season, except dissolved organic content (DOC) and conductance. Bengraïne and Marhaba (2006) demonstrated the usage of PCA to identify the components linked to the variations in hydrochemistry using 19 independent parameters in the Passaic river, New Jersey (Bengraïne & Marhaba, 2003). PCA was used since the units or the variables differ substantially in this study. Moreover, to solve the issue of variables loading substantially on one or more of the axes, a varimax rotation was also carried out in that study.

### 2.3.2 Factor Analysis (FA)

FA is a statistical technique used to investigate the latent factors or dimensions underlying a data set. It is similar to PCA, but it concentrates on identifying the common factors that explain the correlations between observed variables, as opposed to maximizing variance. The FA can be expressed as the following equation (S. K. Singh et al., 2016).

$$Z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \dots + a_{fm}f_{mi} + e_{fi}$$

Where  $Z_{ji}$  is the component score,  $a$  is the component loading,  $f$  is the factor score,  $e$  is the residual term accounting for errors or other sources of variation,  $i$  is the sample number and  $m$  is the total number of variables.

The primary objective of factor analysis is to reduce the contribution of insignificant variables in order to further consolidate the data structure generated by PCA (Wunderlin et al., 2001). FA assumes that the observed variables are influenced by a smaller number of latent factors, which are not explicitly observable but can be inferred from the intercorrelation patterns between the observed variables. The objective of factor analysis is to identify these latent factors and comprehend their contribution to the observed data. The applications of PCA and PFA in environmental management and protection research were greatly illuminated by recent works (Y. Ouyang et al., 2006).

### 2.3.3 Cluster Analysis (CA)

Similar to factor analysis, cluster analysis also groups the data or variables into homogeneous clusters and after getting clusters, the correlation between homogeneous clusters can be identified. The hydrogeochemical parameters are first categorized using cluster analysis based

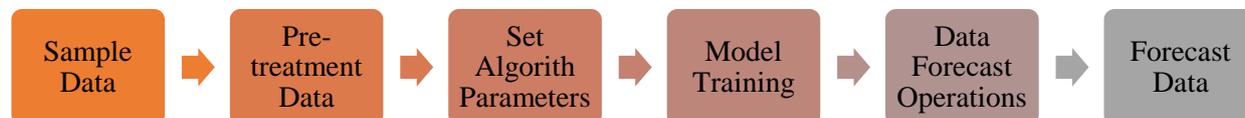
on their average likeness. Typically, the analysis uses the squared Euclidean distance and Ward's linkage approach (Bu et al., 2020).

### 2.3.4 Time Series Analysis

Analyzing water quality data collected over different time domains, time series analysis is applied. A time series is a collection of data points that are usually recorded at regular intervals over a period of measurements. It examines patterns, trends, and seasonality in the data and can identify long-term or periodic changes. It is possible to analyze and interpret temporal patterns in water quality parameters using techniques such as autocorrelation analysis, moving averages, and seasonal decomposition.

The monthly variance of water quality standards was utilized by Parmar and Bhardwaj (2014) to compare statistical parameters, e.g., mean, median, mode, standard deviation, kurtosis, skewness, and coefficient of variation (Parmar & Bhardwaj, 2014). Tan et al. (2012) studied the least squares support vector machine LS-SVM algorithm to build a nonlinear time series forecasting model to predict water quality (Tan et al., 2012). Figure 2-2 shows the framework for the LS-SVM approach, which was adopted to have a high prediction accuracy and to be better suitable for small sample sizes in real-time water quality data forecasts.

*Figure 2-2: Implementation framework for prediction of water quality*



## 2.4 Why is principal component analysis superior?

The quality of water has been analyzed by different researchers using various parameters. Quality, however, is a general phrase that is difficult to define using precise data. For instance, good quality water cannot simply be defined as having a pH value of 7.0 or higher (Mahapatra et al., 2012). Monitoring any potential increasing or declining trends in water quality is therefore essential to understand how the quality of the water is changing. In order to accomplish this, the responsible parties should be involved in implementing suitable monitoring programs that include all necessary criteria to identify the natural and anthropogenic mechanisms regulating water quality (Benkov et al., 2023). A better understanding of the condition of the environment can be obtained by characterizing the spatial variation and source allocation of water quality measures, which can also assist policymakers in setting priorities for sustainable water management (Huang et al., 2010). To accurately assess WQI, it is important to determine which parameters are appropriate. According to agency guidelines like the World Health Organization (WHO) (2006), physical, chemical, and biological indicators are typically used to determine the quality of water. These indicators

include pH, electrical conductivity (EC), total dissolved solids (TDS), total suspended solids (TSS), hardness, turbidity, and contaminant concentrations.

PCA is a well-established statistical method for the investigation of water quality (Acquavita et al., 2015; Navarro et al., 2010; Platikanov et al., 2014) by analysis of obtained observed data. It is a statistical tool for data reduction that can be used to evaluate controls on groundwater composition, interpret observed relationships between variables, and produce simpler relationships that put insight into the underlying structure of the variables. It can also be used to aggregate the effects of many variables into a small subset of factors (Liu et al., 2003). By analyzing the primary sources of data variation, the latent factors (principal components) generated by the original variables (the water quality indicators) are used to interpret the data ("Principal Component Analysis for Special Types of Data," 2002). With the least amount of original data loss possible, PCA offers information on the important parameters (K. P. Singh et al., 2004).

The original variables are combined in weighted linear combinations to form the principal components (PCs). PC gives data on the most important factors that describe the entire dataset and allows for data reduction with the least amount of original information lost (Filik Iscen et al., 2008).

In a recent study by Ersan Batuar et al. (2021) (Batur & Maktav, 2019) the findings of multiple linear regression (MLR), artificial neural network (ANN), and support vector machines (SVM) data mining methods were compared with those obtained using PCA-based response surface regression (RSR) approach for calculating surface water quality parameters. The investigation proved that the PCA-based RSR method is more accurate at estimating lake water quality parameters than MLR, ANN, and SVM data mining models. In 2008, T.G. Kazi et al. showed the origins or variables that cause changes in water quality were found using PCA (Kazi et al., 2009). In their studies of the regional variability of surface water quality and source apportionment, Shrestha and Kazama (Shrestha & Kazama, 2007), Huang et al. (Huang et al., 2010), and Juahir et al. (2011) (Juahir et al., 2011) divided the analyzed water bodies into three categories: High pollution site (HP), Sites with moderate pollution (MP) and low pollution (LP).

PCA aids in the analysis of complicated data matrices so that the water quality and ecological status of the system under study can be better understood. These technologies make it easier to identify potential water quality influencing elements, and they can help with dependable water resource management as well as rapid pollution problem solutions (Adams et al., 2001).

## Chapter 3: Principal Component Analysis (PCA): Theory and Literature

### 3.1 Introduction

Various multivariate statistical methods are used for extracting important and concerning data. In the case of establishing a water quality index – Cluster Analysis (CA), Factor Analysis (FA), Principal Component Analysis (PCA), etc., are used to reveal the relationships within the dataset. The application of PCA in WQI is to reveal “hidden” factors controlling water quality (Franco et al., 2021; Kumar et al., 2022; Subramanian & Baskar, 2022). One of the major drawbacks of water quality assessment is that handling large datasets and interpreting large datasets is often difficult (Jolliffe & Cadima, 2016). Thus, dimensionality reduction in a way such that interpreting the dataset is much easier is what we are looking for. Among various tools and techniques, PCA is the oldest and the most widely used one to deal with this situation (Jolliffe & Cadima, 2016). PCA identifies new variables, the principal components, which are linear combinations of the original variables (Ringnér, 2008). PCA can be used to reduce dimensionality by preserving most of the information that is required in water quality assessment studies to accurately evaluate the water quality.

### 3.2 Theory Behind PCA

For understanding PCA, basic knowledge of linear algebra and statistics is important (Ringnér, 2008). To understand the basic method of performing PCA, the following statistical concepts and some basic theories of linear algebra are discussed.

#### 3.2.1 Basic Statistics Behind PCA

Understanding or having insights about three statistical variables that describe the trend in the dataset would help to get the idea of PCA. These three variables are – a) the mean, b) the variance and c) the covariance.

- a) The mean: The mean defines the middle of the data or the average value per sample of the data. Mathematically,

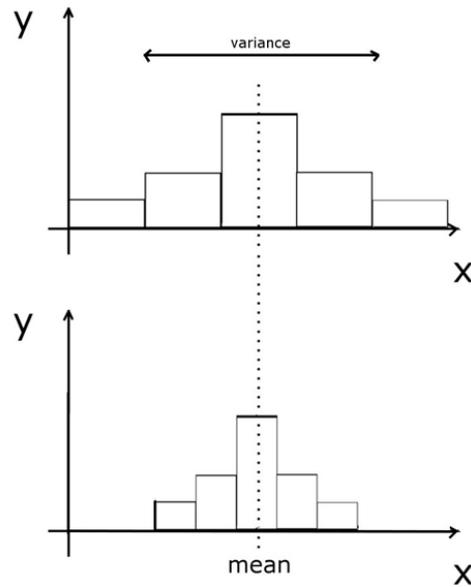
$$\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$$

Where  $n$  is the number of elements in the set  $S$  and  $\bar{S}$  is the mean value of  $S$ .

- b) The variance: Variance represents the spread of the data. Higher variance means a higher range of datasets. For the dataset  $S$  with  $n$  elements, variance can be expressed as –

$$Var(S) = \frac{1}{n-1} \sum_{i=1}^n (S_i - \bar{S})^2$$

Note that,  $n - 1$  is used for sample data while  $n$  is used for the whole population. The following illustration describes the variance in a graphical way. Both the datasets have the same mean but the dataset with higher variance is more spread.



- c) The covariance: covariance defines the degree of codependence of two variables. If two variables are  $X$  and  $Y$ , then the covariance of  $X$  and  $Y$  is given by –

$$Cov(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

Now, for the covariance of  $X$  with itself,

$$Cov(X, X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X}) = Var(X)$$

With these statistical insights, we proceed to the next section.

### 3.2.2 Basic Linear Algebra Behind PCA

Now that we know what the required statistical ideas are, we need to use them in our system. We have used the one-dimensional array or data to explain the statistical concepts. But in real life, datasets are made of different samples and parameters. Most of the time, 2D arrays are used which is also applicable for water quality assessments. There will be different samples that will be analyzed to determine different water quality parameters. Thus, we have to deal

with 2D arrays or what we can call matrices. Suppose we have  $m$  number of parameters for each sample and  $n$  number of samples. If we build a dataset of all the values, then the dataset will look like the following –

$$D = \begin{bmatrix} D_{11} & D_{12} & \dots & \dots & D_{1n} \\ D_{21} & D_{22} & \dots & \dots & D_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ D_{m1} & D_{m2} & \dots & \dots & D_{mn} \end{bmatrix} = [D_{ij}]$$

where,  $D_{ij}$  represents the  $i$ th parameter of the  $j$ th sample and  $D$  is the dataset. Now if we think about applying the statistical concepts in this dataset for each parameter we will have,

$$\bar{D}_i = \frac{1}{n} \sum_{j=1}^n D_{ij}$$

where,  $\bar{D}_i$  is the mean of  $i$ th parameter. The covariance of taking any two of the parameters can be expressed as –

$$Cov(D) = \frac{1}{n} DD^T$$

where,  $D^T$  is the transpose of matrix  $D$ .  $Cov(D)$  is called the covariance matrix of  $D$  having a dimension of  $m \times m$  and is given by,

$$Cov(D) = \begin{bmatrix} Cov(D_1, D_1) & Cov(D_1, D_2) & \dots & \dots & Cov(D_1, D_m) \\ Cov(D_2, D_1) & Cov(D_2, D_2) & \dots & \dots & Cov(D_2, D_m) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ Cov(D_m, D_1) & Cov(D_m, D_2) & \dots & \dots & Cov(D_m, D_m) \end{bmatrix} = [Cov(D_p, D_q)]$$

where,  $Cov(D_p, D_q)$  is the covariance of  $p$ th and  $q$ th parameter.

Now that we have implemented the statistical concepts in the matrix system, we need basic insights into eigenvectors and eigenvalues, the two most important concepts of linear algebra. Eigenvectors are those which does not change direction in any transformation. For any square matrix  $A$ , there will be –

$$Ax = \lambda x$$

Where  $\lambda$  is a scalar and  $x$  is a vector. Here, multiplying a matrix by a vector produces the same result as multiplying a scalar by that vector. The scalar value  $\lambda$  is called the eigenvalue and the vector  $x$  associated with  $\lambda$  is called the eigenvector. Different eigenvectors are found for different eigenvalues. Now, it can be written that,

$$Ax - \lambda x = 0$$

$$Ax - \lambda Ix = 0$$

where  $I$  is the identity matrix. Thus,

$$|A - \lambda I| = 0$$

and solving this equation for  $\lambda$  will determine the eigenvalues and using the first equation, eigenvectors can be determined. With these insights on statistics and linear algebra, we can proceed to perform PCA for any dataset.

### 3.2.3 Performing PCA

As mentioned before, a matrix  $D$  with  $m \times n$  dimension is taken. Performing PCA includes the following steps –

1. Determine the mean of each parameter/variable of the matrix  $D$ .

$$\bar{D}_i = \frac{1}{n} \sum_{j=1}^n D_{ij}$$

2. Subtract the corresponding mean of any parameter from each row. A new matrix  $D'$  will be found which will be called a shifted data matrix. This is done to standardize the data.  $D'$  can be expressed as,

$$D' = \begin{bmatrix} D_1 - \bar{D}_1 \\ D_2 - \bar{D}_2 \\ \dots \\ \dots \\ D_m - \bar{D}_m \end{bmatrix}$$

3. Calculate the covariance matrix,  $C$  for  $D'$

$$C = \frac{1}{n} D' D'^T$$

4. Calculate the eigenvalues and eigenvectors of the covariance matrix,  $C$
5. Check for length of eigenvector, eigenvector should be of unit length.

$$length = \sqrt{\text{sum of the square of the elements of eigenvectors}}$$

if  $length \neq 1$ , then rescale them with the following expression,

$$Scaled\ element = \frac{\text{original element}}{length}$$

6. Place the eigenvectors side by side with the eigenvector of the largest eigenvalue to the left and the descending.

$$W = [e_1 \quad e_2 \quad \dots \quad \dots]$$

where,  $e_1$  is the largest eigenvalue.

7. Transpose the obtained matrix  $W$  and express the data in new rotated coordinate by,

$$D_{PCA} = W^T D'$$

The coordinate transformation that we generated with this method has rotated the axes so that the first coordinate axis lines up with the data in such a way that most of the variation is in that coordinate. The data were correlated in the original coordinates, but they are not correlated in the new rotated coordinates. The aim of PCA is to derive a new set of coordinates for the data that are uncorrelated and that are in the order of the degree of variation in that coordinate. How many of the new coordinates (also called components) are kept is an arbitrary choice; however, the intention is to keep only enough components to capture the essence of the data. A typical method of selecting the components to keep is to sum all of the eigenvalues and then keep only those components with the largest eigenvalues, which sum up to no less than 90% of the total. This is done because the larger the eigenvalue, the greater the amount of variation of the data in the direction of the corresponding eigenvector. In PCA, we choose to keep only those components that carry most of the variation of the data.

## Chapter 4: Water Quality Index: Theory and Literature

### 4.1 Introduction

Evaluating water quality involves the collection and analysis of large datasets consisting of a large number of water quality parameters that can be difficult and create ambiguity in water quality assessment. Thus, researchers came up with the idea of the Water Quality Index (WQI) which can be used to measure the combined influence of numerous water quality parameters (Pesce & Wunderlin, 2000). In general, WQI is a grading system for comparing water from different sources (Stoner et al., n.d.). WQI models were established to analyze natural and artificial activities based on fundamental groundwater chemistry markers. WQI can also assess the quality of river water in general as well as in terms of its intended use for drinking, recreation, aquatic, agriculture, etc. (Al-Shujairi, 2013).

### 4.2 Overview of Water Quality Index Modeling

Modeling of WQI can be categorized into four main steps (Abrahão et al., 2007; Lumb et al., 2011; Sutadian et al., 2018), namely –

1. Selection of water quality parameters that are most important for evaluating water quality
2. Sub-indexing the water quality parameters for making a unitless comparison between the parameters
3. Weighting the water quality parameters based on the significance of the water quality
4. Using an aggregating function to determine the WQI by aggregating the sub-indices into a single value

Although some models have been developed with fewer steps, most of the models considered these four steps. The general model structure for developing WQI models is illustrated in Figure 4-1.

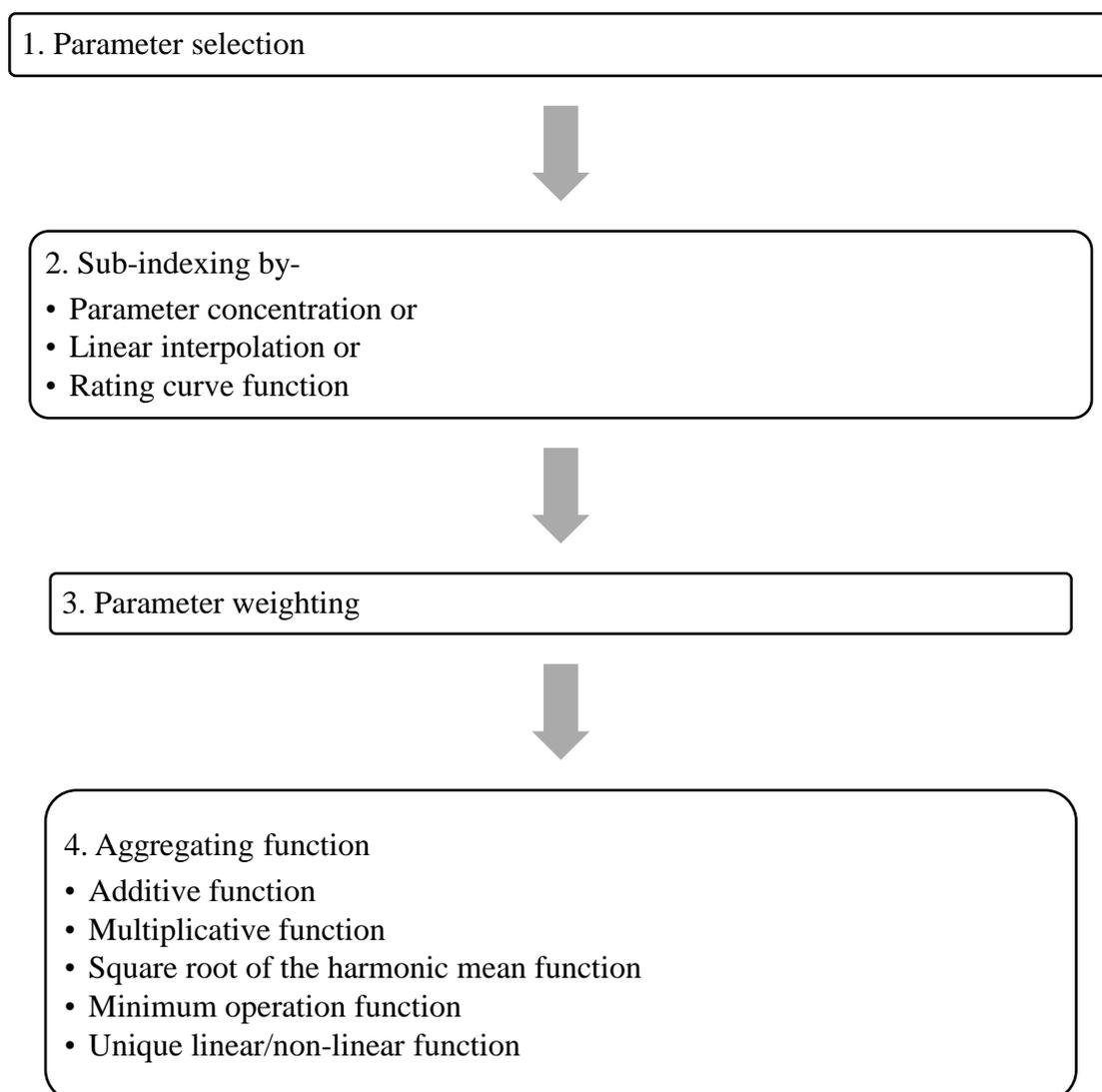


Figure 4-1: General model structure of the WQI model

### 4.2.1 Parameter Selection

Parameter selection is a rudimentary step for modeling WQI for any given instance. Parameters selection can be executed based on the availability of data, expert opinion or the environmental significance of a water quality parameter. WQI has been modeled with as high as 26 parameters such as in the Bascaron Index and Dojildo Index (Koçer & Sevgili, 2014; M. G. Uddin et al., 2021), but most of the models deal with 8 to 11 parameters. However, models such as CCME-WQI were developed considering only 4 parameters (*CCME WATER QUALITY INDEX USER'S MANUAL 2017 UPDATE*, 2017). Many of the researchers did not include some parameters, for example, suspended solids, microbiological contamination in their model because of high analytical cost and limited analytical laboratory facilities (Ma et al., 2020; Naubi et al., 2016). Galal et al. reported that in most of the developed WQI models the Delphi

technique was followed to select the water quality parameters based on gathering expert opinions by interviewing or surveying (M. G. Uddin et al., 2021). The Delphi technique, developed by Dalkey and Helmer in the 1950s is a widely acknowledged method for selecting important opinions of one's interest from a vast set of options available generally by using questionnaires to collect data from an expert panel (Hsu & Sandford, 2019). However, some critics suggest that this technique produces data uncertainty and low model accuracy (Goodman, 1987; Linstone, 1985). Generally, there are no specific rules for selecting water quality parameters to include in the WQI modeling. Recently, statistical approaches have been taken to select water quality parameters and in the field of water quality assessments (Benkov et al., 2023; Tharmar et al., 2022). A statistical approach for selecting water quality parameters that will be followed in this study is principal component analysis (PCA). Recent studies on water quality assessment suggest that PCA can be a better tool to use for evaluating surface water quality (G. Uddin et al., 2022). As there are a copious amount of variables available for evaluating water quality, the relationship among the variables is supposed to give better interpretations than the variables themselves. Also, PCA can be used to find out the dominant variables and their impact on the water quality (Tharmar et al., 2022). By this method, a minimum number of useful parameters can be selected to include in the WQI model with higher efficiency compared to the number of parameters. Lowering the number of parameters is important due to the time consumption for determining all the parameters analytically and economically.

#### **4.2.2 Sub-indexing**

Parameters for evaluating water quality assessment are immensely disparate in the unit. Thus, sub-indexing is done to get rid of the units and make all the parameters comparable in one format. Most of the model uses standard guideline values to achieve sub-indices (Abbasi & Abbasi, 2012a; Liou et al., 2004). However, some of the models, for example, CCME-WQI skipped this step and worked with the aggregating function directly (*CCME WATER QUALITY INDEX USER'S MANUAL 2017 UPDATE*, 2017).

There are several ways for sub-indexing. The most used sub-indexing method is the linear interpolated function where parameters are converted according to their water quality standard value from 0 to 100 (Effendi et al., 2015; Lobato et al., 2015) with the help of any given equation or graphical representation. However, models such as the Horton index and Said index use the measured parameter concentration directly as the sub-index which is the simplest way to do so. Another way of sub-indexing is using rated curving functions used in MRWQI and OWQI (Gazzaz et al., 2012; Hasan et al., 2015). This method is also based on the water quality parameter standard guideline values imposed by the legal authorities. Rather than taking a linear relation, this method takes logarithmic or non-linear regression to achieve sub-indices.

#### **4.2.3 Parameter Weighting**

The main purpose of this step is to apply the relative importance of the water quality parameters (Sarkar & Abbasi, 2006). Most of the models uses an unequal weighting technique where in

some of the models, the total sum of all the parameter weight is equal to 1. Some models, for example – the Horton index and Bascaran Index use integer value of the parameter weight causing the total sum to be greater than 1. A model such as OWQI uses equal weight for all the parameters. However, the CCME-WQI model does not require parameter weighting to evaluate WQI. The analytical hierarchy process (AHP) method, developed by Thomas Satty, is a decision-making technique that can be an effective method for evaluating parameter weighting (M. G. Uddin et al., 2021). This method can reduce uncertainty and give a better accuracy of the weighting process which can be found with AHP (Sutadian et al., 2017).

#### 4.2.4 Aggregating Function

The final process of modeling WQI is the aggregating process. It is used to aggregate all the sub-indices into a single index for evaluating water quality (Dhany Sutadian et al., n.d.). This single index is called the WQI. There are different aggregating functions to achieve this step. These are discussed below –

- i. Additive function: The additive aggregation function is one of the simplest ways for aggregating the sub-indices used in several models. The additive function can be expressed as –

$$WQI = \sum_{i=1}^n w_i S_i$$

Where,  $S_i$  is the sub-index of the  $i^{\text{th}}$  parameter and  $w_i$  is the parameter weight value. This method has a major issue called ‘eclipsing,’ which means that the final index value does not represent the whole scenario (Dhany Sutadian et al., n.d.). Rather, it is dominated by the larger values of the sub-indices and neglect relatively lower values. There are variations in the additive function. Some models used the modified additive function which can be expressed as (Bordalo et al., 2006; Carvalho et al., 2011) –

$$WQI = \frac{1}{100} \left( \sum_{i=1}^n S_i \right)^2$$

$$WQI = \frac{1}{100} \left( \sum_{i=1}^n S_i w_i \right)^2$$

Where,  $S_i$  is the sub-index of the  $i^{\text{th}}$  parameter and  $w_i$  is the parameter weight value. Another modified form of additive function is proposed by Bascaron (Dhany Sutadian et al., n.d.) and adopted and modified by some other models (Pesce & Wunderlin, 2000; Sánchez et al., 2007). In this function, the total value of final aggregation is divided by the total weight, i.e., the sum of the weights of the parameters. This modified function is expressed as –

$$WQI = \frac{\sum_{i=1}^n C_i P_i}{\sum_{i=1}^n P_i}$$

Where,  $C_i$  is the sub-index of the  $i^{\text{th}}$  parameter called the normalization factor and  $P_i$  is the parameter weight value. If a total weight of 1 is taken, then this equation becomes identical to the primary additive function.

- ii. Multiplicative function: Another common aggregation function suggested by Brown and later included in the different models (Almeida et al., 2012; Devendra Swaroop Bhargave & Asce, 1985) is the multiplicative function which can be expressed as follows –

$$WQI = \sum_{i=1}^n S_i^{w_i}$$

Ambiguity arises in this method when the parameter weight is close to zero (Liou et al., 2004). Then this method fails to provide the actual water quality.

- iii. Minimum operator function: The minimum aggregating method considers the minimum sub-index as the final WQI.

$$WQI = \text{Min}(S_1, S_2, S_3, \dots, S_n)$$

This method omits the eclipsing and ambiguity problem but fails to provide an overall composite picture of water quality (Swamee & Tyagi, 2000).

- iv. Combined function: Liou et al. proposed a combined model for aggregating functions to avoid eclipsing and ambiguity problems (Liou et al., 2004). This aggregating function was proposed for evaluating river status and expressed as –

$$WQI = C_{temp} C_{pH} C_{tox} \left[ \left( \sum_{i=1}^l I_i w_i \right) \times \left( \sum_{j=1}^m I_j w_j \right) \times \left( \sum_{k=1}^n I_k w_k \right) \right]^{1/3}$$

Here,  $C_{temp}$ ,  $C_{pH}$  and  $C_{tox}$  are the three scaling factors which are the sub-indices for temperature, pH and toxicity, respectively. In this study, the results from the principal component analysis were used and the results were clustered into three groups. At first, the variables found from PCA are aggregated in the additive method for each principal component. Then the geometric mean is taken and finally, WQI is determined by a multiplicative function with three scaling factors and aggregated indices.

- v. The square root of the harmonic mean function: Another proposed aggregating function is the square root of the harmonic mean function (Cude, 2001; Dojlido et al., 1994). This function can be expressed as –

$$WQI = \sqrt{\frac{n}{\sum_{i=1}^n \frac{1}{S_i^2}}}$$

Where,  $S_i$  is the sub-index of  $i^{\text{th}}$  parameter. A study by Swamee and Tyagi (Swamee & Tyagi, 2000) stated that this aggregation method might cause a problem where all the sub-indices are acceptable but the overall index is not.

- vi. Unique linear/non-linear aggregating function: Said et al. suggested a new aggregating function that deals directly with the parameter value of some preselected parameters and sub-indexing is not required (Said et al., 2004). The WQI equation is proposed as –

$$WQI = \frac{(DO)^{1.5}}{(3.8)^{TP}(Turb)^{0.15}(15)^{FCol/10000} + 0.14(SC)^{0.5}}$$

Where DO is the dissolved oxygen (%), T is turbidity (NTU), TP is total phosphate (mg/L), FCol is fecal coliform bacteria (counts/100 mL) and SC is the specific conductivity (MS/cm). The major limitation of this method is the study is based on a specific region and might cause inaccuracy for other regions.

### 4.3 Existing WQI Models

Galal et al. studied 107 cases and stated that almost 85% of the cases used seven models naming The Horton index, National Sanitation Foundation WQI (NSF-WQI), Scottish Research Development Department Index (SRDD index), Canadian Council of Ministers of the Environment WQI (CCME-WQI), Bascaron index (BWQI), Fuzzy Interface System (FIS) based index, Malaysian WQI (MWQI) (M. G. Uddin et al., 2021). Each of these models has a minimum of four applications, with the highest 36 applications of CCME-WQI and NSF-WQI at the second ranking based on the application on the case studies with a number of 18 applications. Another mentionable WQI model is the West Java WQI (WJ-WQI) model, not because of the number of applications but because it is one of the most recent WQI models developed considering the reduction of uncertainty of the existing models (Sutadian et al., 2018). These above-mentioned models are briefly described in the following sections.

#### 4.3.1 The Horton Index

Horton established the very first WQI model in 1965 for defining the water quality of rivers (Gupta & Gupta, 2021; Horton, 1965; Mukate et al., 2019; Tirkey et al., 2013). This model includes the four general steps for WQI modeling. Parameter selection was based on the environmental impact, data reliability and relative influence on other parameters (Abbasi & Abbasi, 2012c). Water quality parameters such as DO, pH, coliforms, specific conductance, carbon chloroform extract, and alkalinity (based on percent entry of population upstream served by treatment) were considered (Shah & Joshi, 2017). In this model, sub-indexing was done using a linear scaling function and the values ranged from 0 to 100. Where 0 represents the worst quality and 100 represents the excellent quality (Horton, 1965). Parameter weighting was based on the Delphi method. The expert panel suggested weighting values from 1 to 4 for different parameters.

An additive function is used as the final aggregating equation for determining the WQI value, expressed as the following –

$$WQI = \left[ \frac{\sum_{i=1}^8 w_i S_i}{\sum_{i=1}^8 w_i} \right] \times m_1 \times m_2$$

Where,  $m_1$  and  $m_2$  are the coefficient of temperature and obvious pollution, respectively. Values of  $m_1$  and  $m_2$  are either 0.5 or 1.0 based on the temperature and presence of emission.

### 4.3.2 National Sanitation Foundation (NSF) WQI

Brown modified the Horton model and developed the NSF-WQI model (Abrahão et al., 2007; Lumb et al., 2011). Parameter selection was done using the Delphi technique (Ewaid, 2017; Rocha et al., 2015) and considered 11 different parameters of five different categories namely physical, chemical, microbiological, nutrient and toxic parameters (Dhany Sutadian et al., n.d.; Lumb et al., 2011). Linear scaling was used for sub-indexing and the values ranged from 0 to 1. 1 is considered to be in the recommended guideline (M. G. Uddin et al., 2021). The Delphi technique is also followed to evaluate the parameter weighting with an expert panel. Unequal weighting values are used, and the total sum of the weighting is 1. However, later applications of this model have used some modified weight values for some parameters (Noori et al., 2019; Tomas et al., 2017). And for the final step, the original model was developed with an additive aggregating function but later in 1973, Brown et al. suggested the multiplicative function to aggregate the sub-indices into an overall index (Brown et al., 1973; M. G. Uddin et al., 2021).

The proposed method for comparing the water quality of various water sources is based upon nine water quality parameters such as temperature, pH, turbidity, fecal coliform, dissolved oxygen, biochemical oxygen demand, total phosphates, nitrates and total solids. The water quality data are recorded and transferred to a weighting curve chart, where a numerical value of  $Q_i$  is obtained. The mathematical expression for NSF WQI is given by

$$WQI = \sum_{i=1}^n Q_i W_i$$

Where,  $Q_i$  = sub-index for  $i$ th water quality parameter and  $W_i$  = weight associated with  $i$ th parameter and  $n$  is the number of parameters. The model was first proposed with 9 water quality parameters. These are – DO, pH, BOD, temperature, total phosphate, nitrate, turbidity, total solids and fecal coliform. With that, the model proposed a specific weighting value for each parameters. Later, researchers modified the number of parameters and changed the parameters. They also defined their own parameter weight values. Table 4-1 shows a brief summary of the parameter weight values used by different researchers.

Table 4-1: Different parameter weight value used for NSF WQI model in different studies

WQ Parameter	Model Recommended Weight, $W_i$	Weight value defined in different studies					
		(Effendi et al., 2015)	(Babaei Semiromi et al., 2011)	(Shah & Joshi, 2017)	(Hoseinzadeh et al., 2015)	(Ewaid, 2017)	(Tomas et al., 2017)
DO	0.17	0.20	0.129	0.17	0.17	0.17	0.2857
pH	0.11	0.11	0.133	0.11	0.12	0.11	0.0714
BOD	0.11	0.13	-	0.11	0.10	0.11	0.2142
Temperature	0.10	0.12	-	0.10	0.10	0.10	-
Total Phosphate	0.10	0.12	-	0.10	0.10	0.10	0.1429
Nitrate	0.10	0.12	0.128	0.10	0.10	0.10	0.1429
Turbidity	0.08	0.10	0.155	0.08	0.08	0.08	-
Total Solids	0.07	0.08	0.100	0.07	0.08	0.07	-
Fecal Coliform	0.16	-	0.182	0.16	0.15	0.16	-
Total Odor Number	-	-	-	-	-	-	0.1429
Total Suspended Solids	-	-	0.173	-	-	-	-

### 4.3.3 The Scottish Research Development Department (SRDD) Index

The SRDD-WQI model was developed by the Scottish Research Development Department and was used to assess surface water quality from different tropical-sub-tropical countries such as – Iran, Romania, Portugal (Carvalho et al., 2011; Dadolahi-Sohrab et al., 2012; IONUŞ, 2010). In Eastern Thailand, a modified version SRDD model has been used for evaluating water quality (Bordalo et al., 2006). All three processes, parameter selection, sub-indexing, and parameter weighting, were conducted using the Delphi technique (M. G. Uddin et al., 2021). The SRDD model used the following modified additive aggregating function based on the NSF-WQI model (Lumb et al., 2011):

$$SRDDWQI = \frac{1}{100} \left( \sum_{i=1}^n S_i w_i \right)^2$$

#### 4.3.4 Canadian Council of Ministers of Environment (CCME) WQI

The CCME-WQI model is a modified version of the British Columbia Water Quality Index (BCWQI) model (Abbasi & Abbasi, 2012b). The CCME-WQI has been in application with a wide range of surface water bodies (M. G. Uddin et al., 2017). The CCME-WQI requires a minimum of four parameters to determine the water quality index; parameters are not specified and decided by the users (*CCME WATER QUALITY INDEX USER'S MANUAL 2017 UPDATE*, 2017). This model does not include parameter sub-indexing and weighting. After user-defined parameters are selected the model directly uses the aggregating function. In this model, 3 factors  $F_1$ ,  $F_2$  and  $F_3$  namely, 'scope,' 'frequency' and 'amplitude' respectively are calculated before aggregating the parameters. These factors are defined and expressed as follows –

$F_1$  = percent parameters that do not meet the specified objective

$$F_1 = \frac{\text{number of failed parameters}}{\text{number of total parameters}} \times 100$$

$F_2$  = percent individual test values that do not meet the specified objective value

$$F_2 = \frac{\text{number of failed tests}}{\text{number of total tests}} \times 100$$

$F_3$  = amount by which test values failed to meet their objective and this factor is calculated in three steps –

- a. The number of times by which any individual parameter deviates from the objective is called excursion and is calculated using the following expressions –

$$\text{excursion}_i = \left( \frac{\text{Failed Test Value}_i}{\text{Objective}_i} \right) - 1, \text{ when test values exceed the objective}$$

$$\text{excursion}_i = \left( \frac{\text{Objective}_i}{\text{Failed Test Value}_i} \right) - 1, \text{ when test values fall below the objective}$$

- b. The total extent of deviation is calculated by summing the individual excursions and dividing by the total number of tests

$$\text{the normalized sum of excursion, } nse = \frac{\sum_{i=1}^n \text{excursion}_i}{\text{number of total tests}}$$

- c. Finally, an asymptotic function is used to calculate the factor  $F_3$  to yield a range from 0 to 100 with the following expression –

$$F_3 = \frac{nse}{0.01 nse + 0.01}$$

Finally, the function used for aggregation is expressed as:

$$CCMEWQI = 100 - \left[ \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \right]$$

#### 4.3.5 Bascaron Index (BWQI)

Bascaron developed this model based on the Spanish water quality guidelines in 1979 (Sun et al., 2016). Many South American countries and some southern Asian regions adopted the Bascaron Index and some other countries developed a modified indexing model based on the Bascaron's WQI model (M. G. Uddin et al., 2021). BWQI model was developed with considering 26 parameters which is the highest number of parameters considered for developing any WQI model (Nong et al., 2020). Linear sub-indexing is applied based on the local water quality guidelines and a value ranging from 0 to 100 is developed against each of the parameters (Kannel et al., 2007; Pesce & Wunderlin, 2000). Unequal weight values ranging from 1 to 4 are used. Two modified additive functions were used to aggregate the sub-indices. One is to determine the objective index value and the other one is to determine the subjective index value. Expression for these two aggregating functions are as follows –

$$BWQI_{Obj} = \frac{\sum_{i=1}^n w_i S_i}{\sum w_i}$$

$$BWQI_{Sub} = k \times \frac{\sum_{i=1}^n w_i S_i}{\sum w_i}$$

The fundamental difference in these two equations is the constant,  $k$  which can be defined as the constant of visual assessment. values of  $k$  are achieved with a visual inspection of the river (Pesce & Wunderlin, 2000) and one of the following is taken as the value for  $k$  –

$k = 1.00$  ; clear water without apparent contamination of natural solids suspended.

$k = 0.75$  ; light contaminated water, indicated by light non-natural colour, foam, light turbidity for no natural reason.

$k = 0.50$  ; contaminated water, indicated by non-natural colour, light to moderate odour, high turbidity (non-natural), suspended organic solids, etc.

$k = 0.25$  ; highly contaminated water, indicated by blackish colour, hard odour, visible fermentation, etc.

#### 4.3.6 Fuzzy Interface System (FIS) Based Index

Fuzzy logic has been a popular tool for decision-making since the 1960s and many researchers have used fuzzy interface systems for environmental assessments (Peche & Rodríguez, 2012). Several studies included FIS to develop WQI models in recent years (Lermontov et al., 2009; F. Lu et al., 2014; Xia & Chen, 2015). The FIS-based WQI model also uses four steps, and these are analogous to the general steps of developing WQI models (Lermontov et al., 2009). These steps are a) fuzzy sets and membership function – related to the parameter selection and based on the correlation studies; b) fuzzy set operations – normalizing the water quality

parameters with FIS; c) fuzzy logic – weight values are determined with Fuzzy logic function; and d) interference rules – aggregating the water quality parameters using a range of fuzzy interference rules.

### 4.3.7 Malaysian Index (MWQI)

Department of Environment, Malaysia developed a WQI model in 1974 to evaluate the surface water quality based on the national water quality guidelines of Malaysia (Gazzaz et al., 2012). Six water quality parameters were selected for developing the model based on an expert panel. For sub-indexing, unique quality function curves were developed for each parameter. Parameter weighting was also done using the expert panel's opinion and unequal weighting was applied (Khuan et al., 2002). This model also uses the simple additive function for aggregating the sub-indices into the final WQI value. They classified the water quality based on the adopted WQI. Table 4-2 shows the water quality classification and Table 4-3 shows the WQI classification.

### 4.3.8 West Java Index (WJ-WQI)

This is the most recent WQI model developed in literature as per Sutadin et al. (2018) (Khuan et al., 2002). This model has all the four general steps of developing WQI models. This model was developed focusing on reducing the uncertainty of the previously existing WQI models. Parameter selection includes two sub-steps – a) using statistical assessment to evaluate parameter redundancy and b) identifying common parameters across the sampling stations. A linear scaling interpolation function is introduced to determine the sub-indices based on the maximum and minimum allowable guideline values of different water quality parameters. Parameter weighting was conducted with an expert panel and the panel's opinion was evaluated using the Analytical Hierarchy Process (AHP). Finally, the model uses the same multiplicative aggregating function used in the NSF-WQI model.

### 4.3.9 WQI Evaluation

The following table shows the evaluation classes recommended by different models.

Table 4-2: Different WQI Model Evaluation Classes

WQI Name	No. of Parameters Used and Evaluation Classes	WQI Range	Remarks on Water Quality
<b>Horton Index</b>	8 Parameters 5 Classes	91-100	Very Good
		71-90	Good
		51-70	Poor
		31-50	Bad
		0-30	Very Bad
<b>NSF-WQI</b>	11 Parameters 5 Classes	90-100	Excellent
		70-89	Good

		50-69	Medium
		25-49	Bad
		0-24	Very Bad
<b>SRDD Index</b>	10 Parameters 7 Classes	90-100	Clean
		80-89	Good
		70-79	Good Without Treatment
		40-69	Tolerable
		30-39	Polluted
		20-29	Several Polluted
		0-19	Piggery Water
<b>CCME-WQI</b>	4 Parameters 5 Classes	95-100	Excellent
		80-94	Good
		65-79	Fair
		45-64	Marginal
		0-44	Poor
<b>BWQI</b>	26 Parameters 5 Classes	90-100	Excellent
		70-90	Good
		50-70	Medium
		25-50	Bad
		0-25	Very Bad
<b>MWQI</b>	6 Parameters 3 Classes	81-100	Clean
		60-80	Slightly Polluted
		0-59	Polluted
<b>WJ-WQI</b>	13 Parameters 5 Classes	90-100	Excellent
		75-90	Good
		50-75	Fair
		25-50	Marginal
		5-25	Poor

Table 4-3: WQI parameters used in different WQI models

WQ Parameters	Existing WQI Models						
	Horton Index	NSF-WQI	SRDD Index	CCME-WQI	BWQI	MWQI	WJ-WQI
Temperature		√	√	Any Four Selected By User	√		√
pH	√	√	√		√	√	
Electrical Conductivity (EC)	√		√		√		
Total Solids (TS)		√					√
Total Dissolved Solids (TDS)	√						
Dissolved Oxygen (DO)	√	√	√		√	√	√
Biochemical Oxygen Demand (BOD)		√	√		√	√	
Total Suspended Solids (TSS)			√			√	
Chemical Oxygen Demand (COD)						√	√
Chloride (Cl)	√				√		√
Total Coliform (TC)					√		
Nitrate (NO <sub>3</sub> <sup>-</sup> )		√			√		√
Phosphate (PO <sub>4</sub> <sup>3-</sup> )		√	√				√
Ammonia (NH <sub>3</sub> )			√		√	√	
No. of Other Parameters	3	4	2		17	0	6

## Chapter 5: Approach and Methodology

### 5.1 Introduction

For establishing the water quality index of Dhaka city-based rivers, samples will be collected bi-monthly from four rivers: Buriganga, Turag, Balu and Shitalakhya. Samples will be collected from 6 monitoring sites located on the selected rivers and 14 water quality parameters will be analyzed for each sample. After data collection, spatial interpolation of the sampling sites and parameter concentrations will be studied using ArcGIS software.

To achieve the main objective, the principal component analysis will be conducted using the sampling data and based on the result, i.e., based on the principal components, the WQI model will be developed by using a modified expression of the NSF-WQI model.

### 5.2 Site Selection

Site selection was mainly based on the Department of Environment's (DoE) sample collection points as well as on-site field visits. The Department of Environment has established a comprehensive monitoring network to diligently track surface water quality and this network entails collecting monthly samples from specific sampling stations strategically placed along rivers that are part of the network. Site selection is done for all four rivers. Table 5-1 Represents the reasons behind site selection and the coordinates of the sites are given in Appendix B. Figure 5-1 represents the map location of the sites.

The general location of the sampling sites was expressed using the following naming convention:

First Letter of the river with a distinguished middle letter(e.g., BG, TR, BL or SL) \_Sampling station number (e.g., 1, 2, 3,...).

Table 5-1: Site selection reasons

Site No.	River Name	Site Code	Sampling Sites	Reasons Behind Selection
1	Buriganga	BG1	<i>Pagla Ghat, Dhaka</i>	Nearby brick resellers and food industries
2		BG2	<i>Postogola Bridge, Dhaka</i>	Nearby small and large iron recycling shops and factories and an unknown liquid discharge line in the river
3		BG3	<i>Sadarghat Launch Terminal</i>	One of the busiest and most important areas in Buriganga
4		BG4	<i>Chandir Ghat</i>	Nearby plastic recyclers, secondary recycling shops and factories
5		BG5	<i>Kholamora Ghat, Kamranghirchar</i>	Municipal discharge and transportation, recently sucker fish have been found and are in visible amounts on this site
6		BG6	<i>Gudaraghat, Hazaribag</i>	Nearby former tannery industries
7	Turag	TR1	<i>Near Fulpukuria Thread, Tongi, Gazipur</i>	Nearby textile and chemical industries
8		TR2	<i>Tongi Rail Bridge, Gazipur</i>	Nearby public bazar, construction sites and hospitals
9		TR3	<i>Ashulia Kacha Bazar</i>	Municipal waste disposal area beside the river
10		TR4	<i>Ashulia Landing Station</i>	The recent addition of the landing station and is the joining point of Turag and Tongi khal
11		TR5	<i>Rustampur Ghat, Mirpur road</i>	Household waste discharge
12		TR6	<i>Diabari Ghat, Mirpur road</i>	Nearby public bazar and the discharged municipal liquid waste into the river
13	Balu	BL1	<i>Near Balu Bridge, 300 feet, Dhaka</i>	Nearby bazar and future cricket stadium, a prime attraction for tourists
14		BL2	<i>Near LGED Bridge, Jolshiri</i>	Nearby recreational spot- jolshiri and industrial zone Pran-RFL city
15		BL3	<i>Beraid Boat Ghat, Beraid</i>	The recent addition of the nearby Balu bridge and public bazar
16		BL4	<i>Near Edarkandi-Fakirkhali Road</i>	Nearby hybrid poultry farms by local people
17		BL5	<i>Near Eastern Straw Board and Paper Mill</i>	Nearby paper mills and residential area
18		BL6	<i>Rajakhali Ghat, Demra</i>	Nearby demra where Shitalakhya and Balu meets together, there is a nearby residential area and public bazar

<b>19</b>	Sitalakhya	SL1	<i>Kholapara-Ghorashal Kheyaghat, Ghorashal</i>	Nearby several painting and chemical industries and industrial discharge
<b>20</b>		SL2	<i>Kaliganj Kazir Char Kheyar Ghat, Kaliganj</i>	Nearby newly developing AK Khan economic zone
<b>21</b>		SL3	<i>Beldi Bazar, Beldi</i>	Nearby brick manufacturing industry
<b>22</b>		SL4	<i>Near Hatabo Bazar, Rupganj</i>	Nearby paper mills
<b>23</b>		SL5	<i>Habib Ghat, Boralu Bazar</i>	Industrial zone, several paper mills, food industries
<b>24</b>		SL6	<i>South Rupshi Masjid Ghat, Rupshi</i>	Nearby public bazar with small waste dumps

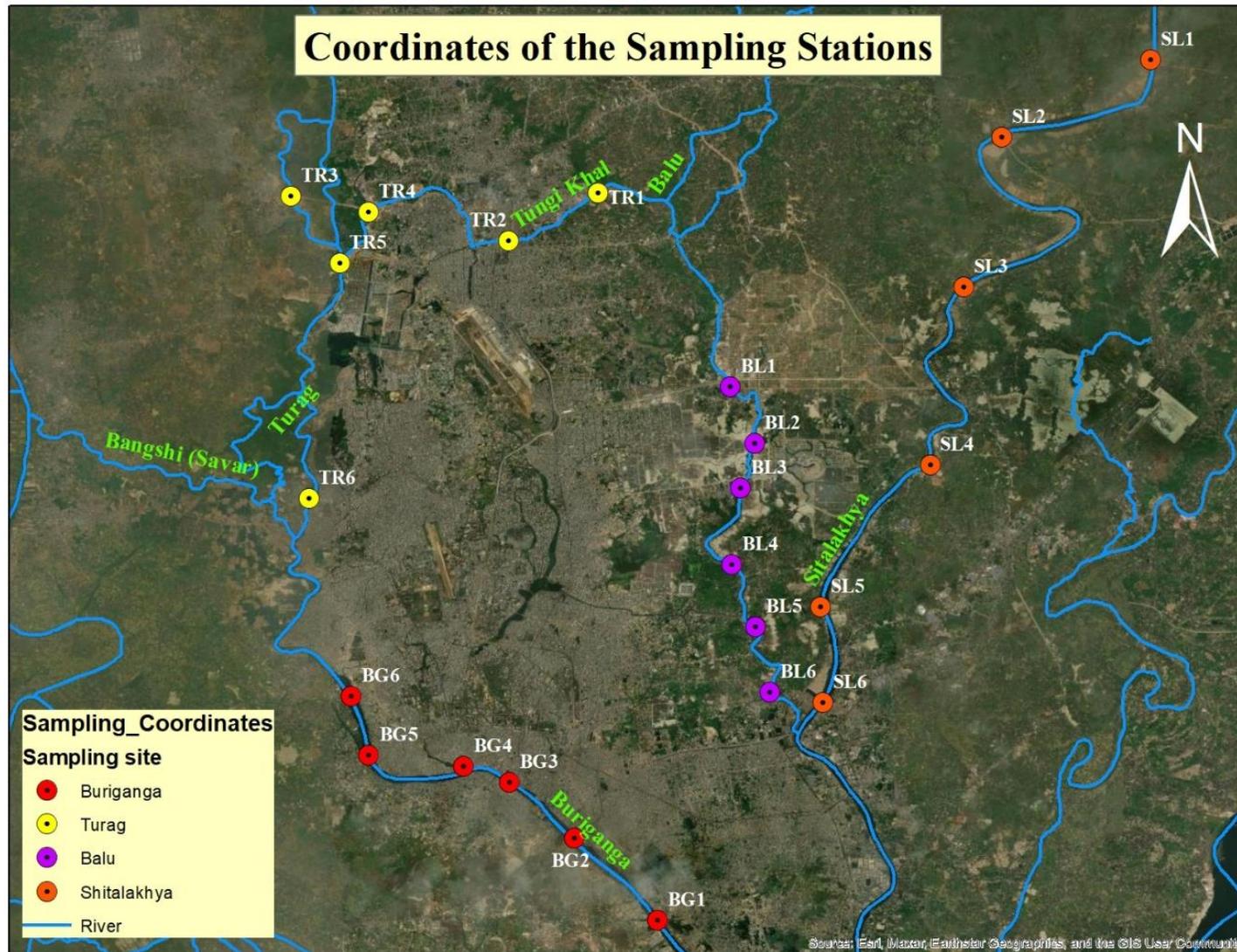


Figure 5-1: Sampling site locations for river

### 5.3 Data Collection

The methodology starts with a preliminary analysis of the surface water quality variables for the evaluation of temporal and spatial variations and the interpretation of the concentration of pollutants. This analysis helps to identify the most polluted segments of the four rivers and their adjacent streams, as well as the quality variables with the best and worst performances. This is a valuable step towards designing the more complex and rigorous procedure that we describe in the succeeding paragraphs. Initially, we would select 6 numbers of monitoring sites for each river based on the literature review of previous reports of WARPO, DoE and IWM, field surveys and analysis through the ArcGIS software. The sample collection would be done for each of the four seasons- pre-monsoon, monsoon season, post-monsoon and dry season. Table 5-2 shows the different parameters that will be considered for this study.

Table 5-2: Different considerable parameters and analyzing instruments/methods

Sl. No.	Parameter	Unit	Standard Protocol
1	Temperature	°C	APHA-2550
2	pH	-	ASTM-D-1293
3	Electrical Conductivity (EC)	mS/cm	Glass Electrode
4	Turbidity	NTU	-
5	Total Solids (TS)	mg/L	APHA-2540 B
6	Total Dissolved Solids (TDS)	mg/L	APHA-2540 C
7	Dissolved Oxygen (DO)	mg/L	Glass Electrode
8	Biochemical Oxygen Demand (BOD)	mg/L	APHA-5210 B
9	Total Suspended Solids (TSS)	mg/L	APHA-2540 D
10	Chemical Oxygen Demand (COD)	mg/L	APHA-5220 D
11	Chloride (Cl <sup>-</sup> )	mg/L	ASTM-E-201
12	Total Coliform (TC)	CFU	US-EPA
13	E. Coli	CFU	US-EPA
14	Nitrate (NO <sub>3</sub> <sup>-</sup> )	mg/L	APHA-4500- NO <sub>3</sub> <sup>-</sup> B
15	Phosphate (PO <sub>4</sub> <sup>3-</sup> )	mg/L	Spectrophotometric
16	Ammonia (NH <sub>3</sub> )	mg/L	APHA-4500- NH <sub>3</sub> B

### 5.4 ArcGIS Modeling

GIS is an indispensable tool for natural resource management, particularly in the areas of land use planning, animal habitat analysis, and natural hazard assessment to name a few of its many applications (Nandini et al., 2007). Although statistical surfaces, in the sense that they are defined by cartographers, do not exist in the same way that land does, it is possible to conceptualize them in the same way. For the preparation of statistical surfaces, the spatial interpolation process in GIS is commonly utilized (Wu et al., 2019). Given the impossibility of collecting field data at every location in the study area, spatial interpolation methods will be used to extrapolate information from sampled locations to locations where it could not be collected. Aiming toward the creation of maps depicting the spatial distribution of water quality

parameters, ArcMap 10 GIS software would be used to correlate the water quality to the sampling locations (Nath et al., 2018). There are various spatial interpolation methods available. The inverse distance weighting (IDW) method will be used in this case (G. Y. Lu & Wong, 2008). By using the spatial interpolation method, maps depicting the spatial distribution of the water quality parameters with specified sites will be created. Identifying variations in the accumulations of various parameters in the surface water of the research area was performed.

## 5.5 Principal Component Analysis

A statistical approach, PCA, will be used to analyze the water quality of Dhaka-based four rivers (Buriganga, Turag, Balu and Shitalakshya). The developed method can be easily applied to other river basins suffering from similar environmental problems. The developed method can rigorously reflect the trends and patterns in water quality and is easy to apply and interpret. Among various techniques, PCA has received interest in water quality analysis. PCA can be used to identify the critical parameters that severely affect the water quality of a specific water body, which would significantly decrease the load of data for further data processing.

PCA will be utilized to reduce the huge amount of data into some meaningful, simpler data that will be further used to find out the prominent factors that are responsible for the variation of the water quality characteristics. To achieve this, R language will be used for data analysis. And software MS Excel and RStudio 2023.9.1.0 will be used.

R is a programming language and environment specifically designed for statistical computing and graphics. It is an open-source software that provides a wide variety of statistical and graphical techniques, making it a powerful tool for data analysis, manipulation, and visualization. R has a large and active community of users and developers, contributing to its extensive library of packages for diverse statistical methods. Key features of R includes – data analysis, statistical modeling, data visualization, data manipulation and programming. RStudio is an integrated development environment (IDE) for R that enhances the R programming experience. It provides a user-friendly interface with features tailored to R development, making it easier to write, debug, and execute R code. Community version of RStudio will be utilized in this research project. RStudio have several advantages on the user-friendly interface such as – script editor, console, data viewer, plots and charts, package management etc. which can be handled in one window altogether making it highly suitable for data analysis.

PCA for water quality parameters using R would be achieved with the following steps. Required steps for performing PCA using R is demonstrated in the Figure 5-2.

1. Setting up the working directory in the computer which is selecting the folder where the data files and analysis outcomes will be stored.
2. Data collection and data importing which includes saving data into a .csv (comma separated value) file and importing these data into a data frame for further analysis.

3. Data exploration and processing, examining the structure of the dataset and imputing missing values or removing outliers.
4. Principal component analysis which can be done using different built-in functions such as `prcomp()` or `princomp()` with specified function variables for data standardization including scaling and centering.
5. Interpretation of the PCA includes the understanding of variance and selection of principal component based on the variance. Also, visualization of PCA using scree plot, biplot etc. would aid the interpretation process.
6. Finally, exporting the outcomes of PCA into a spreadsheet/excel (.xlsx) file for variable (parameter) selection will be done.

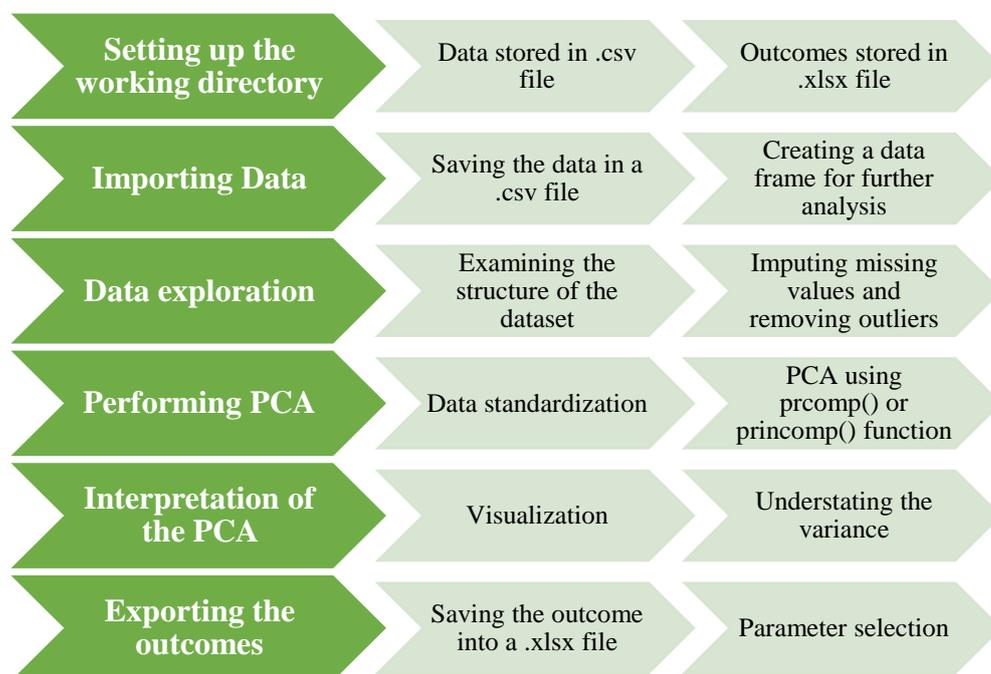


Figure 5-2: Steps involving in PCA using R

## **5.6 Development of the Water Quality Index Model**

The WQI will be established to analyze natural and artificial activities based on fundamental groundwater chemistry markers. The water quality index can also assess the quality of river water in general as well as in terms of its intended use for drinking, recreation, aquatic zones, agriculture, etc. There are many water quality index calculation methods available for evaluating the water. In this study, the WQI model will be developed using the critical parameters obtained from PCA. The obtained WQIs from different models will be compared. Then, based on the research and PCA, a modified expression of WQI based on the NSF-WQI model will be proposed for Bangladeshi rivers, which will make the WQI calculations quick, objective, and reproducible and enable the evaluation of changes in water quality in various regions.

## **5.7 Summary of Methodology**

The research methodology has begun with sampling site selection. After ensuring the reliable sites for sample collection various water quality parameters will be analyzed. Based on the water quality parameter data, spatial interpolation and principal component analysis will be conducted. After determining the critical components from the principal component analysis, the WQI model will be developed based on the critical components and a modified expression for the WQI model based on the NSF-WQI model will be proposed for Dhaka city-based rivers which can be useful for further study of surface water quality of Bangladesh. The overall flow chart of the methodology is shown in as below –

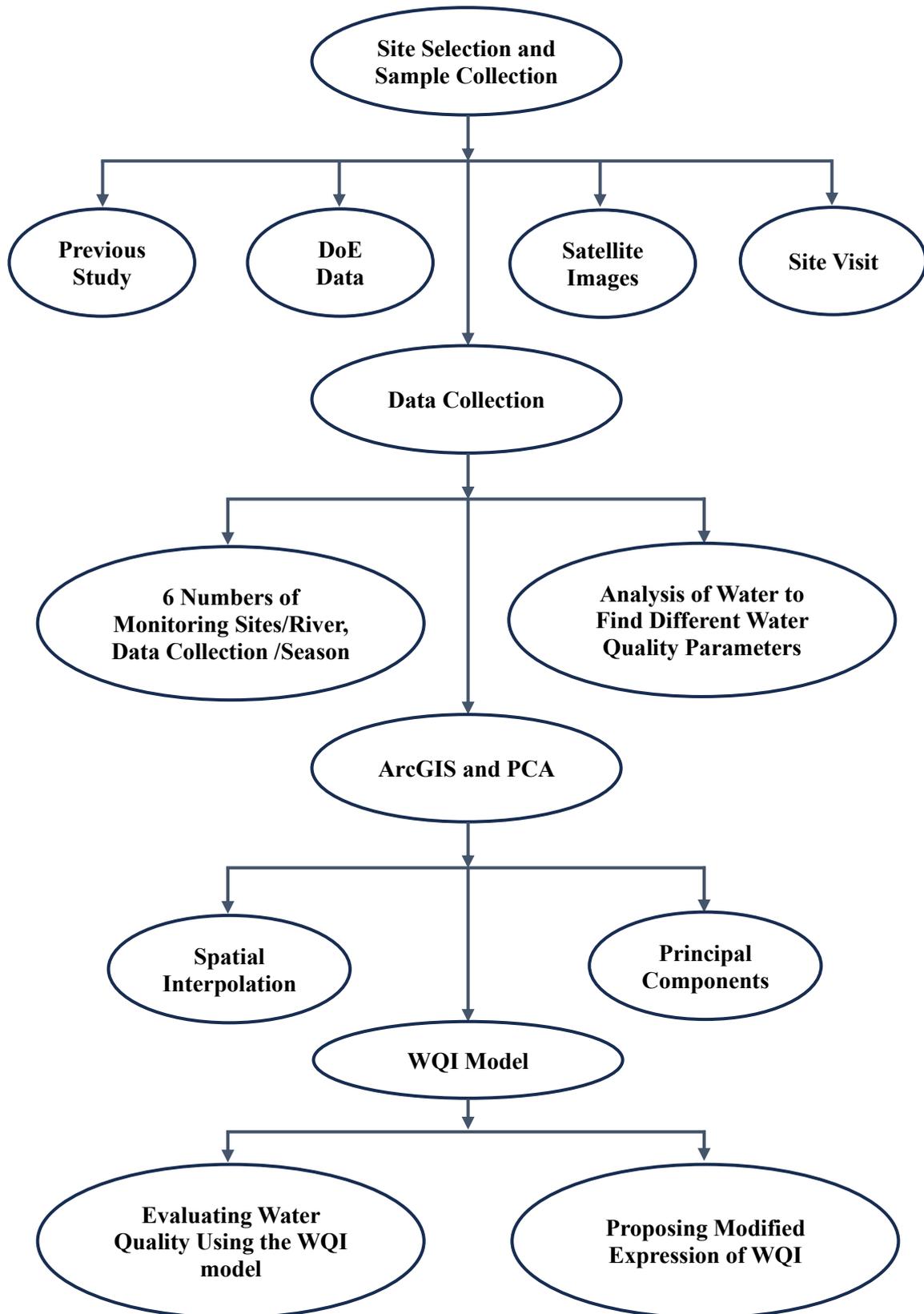


Figure 5-3: Flowchart for overall methodology

## Chapter 6: Work Plan and Progress

### 6.1 Agreement Signing

The agreement between Dept. of Chemical Engineering, BUET and the Water Resources Planning Organization (WARPO) to conduct the collaborative research project on “Establishment of Water Quality Index (WQI) through Principal Component Analysis for the Dhaka-based Rivers” was signed on 06 June 2023 (Figure 6-1). The Director General of WARPO, along with senior scientific officers and other officials, were present at this ceremony. On behalf of the Research team of the Dept. of ChE, the Principal Investigator, Professor Dr. Md. Shahinoor Islam was also present at this event. The ceremony was held in WARPO Bhaban, Green Road, Dhaka.



Figure 6-1: Contract signing between WARPO and Dept. of ChE, BUET

## 6.2 Field Visit and Sample Collection

The first sampling was conducted on Friday, 21<sup>st</sup> July, 2023. Members from the BUET team, including the principal investigator Dr. Md Shahinoor Islam and members from the WARPO team were present in the sample collection. The sample was collected from the BG4 station. A picture of the members present at the first sample collection is shown in Figure 6-2.



*Figure 6-2: Members present in sample collection at Buriganga (BG4)*

Another field visit was conducted on Friday, 04<sup>th</sup> August 2023, at the TR1 station in Turag river, near Fulpukuria Thread & Accessories Ltd. Images taken on the second sample collection are shown in Figure 6-3. A field visit and sample collection were conducted at the Balu river near Balu bridge, 300 feet, Dhaka. Figure 6-4 contains the images from the field visit at Balu Bridge on Wednesday, 06<sup>th</sup> September 2023. On 28<sup>th</sup> September another field visit was conducted in Ghorashal Gudara ghat at the SL1 station in Shitalakhya river. Figure 6-5 contains the images from the field visit at Ghorashal point.



*Figure 6-3: Members present in sample collection at Turag river (TR1)*



*Figure 6-4: Members present in sample collection at Balu river (BL1)*



*Figure 6-5: Members present in sample collection at at Shitalakhya river (SL1)*

### **6.3 Future Work Plan and Reporting**

After submitting the inception report, an inception workshop will be carried out within the 5<sup>th</sup> month (M5). In this workshop, a detailed project planning will be discussed to meet the final objectives of this research. An interim report regarding the progress of the research will be submitted at the end of 13<sup>th</sup> month (M13) following the submission of the draft final report at the end of 21<sup>th</sup> month (M21). A national-level final workshop will be conducted on 23<sup>th</sup> month (M23) following the final report submission at the end of the project (on month 24<sup>th</sup>). The detailed activity plan and schedule are illustrated in Table 6-1.

Table 6-1: Activity Schedule starting from June, 2023

SI	Description of Activities	Months																							
		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	M 22	M 23	M 24
1.	Mobilization																								
2.	Site Selection																								
3.	Site Visits and Secondary Data Collection and Analysis																								
4.	Sample Collection																								
5.	Lab Analysis																								
6.	Principal Component Analysis																								
7.	Determination of WQI																								
8.	Comparison of WQIs																								
9.	WQI Establishment																								
10.	Inception Report																								

Sl	Description of Activities	Months																							
		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17	M 18	M 19	M 20	M 21	M 22	M 23	M 24
11.	Inception Workshop																								
12.	Interim Report																								
13.	Draft Final Report																								
14.	Training																								
15.	Final Workshop at WARPO																								
16.	Final Report																								

**Continuous** 

**Intermittent** 

---

## Chapter 7: References

- 2010 ENVIRONMENTAL PERFORMANCE INDEX. (n.d.). Retrieved October 8, 2023, from <http://ciesin.columbia.edu>
- Abbasi, T., & Abbasi, S. A. (2012a). Water-Quality Indices: Looking Back, Looking Ahead. *Water Quality Indices*, 353–356. <https://doi.org/10.1016/B978-0-444-54304-2.00016-6>
- Abbasi, T., & Abbasi, S. A. (2012b). Water-Quality Indices of USA and Canada. *Water Quality Indices*, 175–186. <https://doi.org/10.1016/B978-0-444-54304-2.00010-5>
- Abbasi, T., & Abbasi, S. A. (2012c). Water Quality Indices. *Water Quality Indices*, 1–362. <https://doi.org/10.1016/C2010-0-69472-7>
- Abrahão, R., Carvalho, M., Da Silva, W. R., Machado, T. T. V., Gadelha, C. L. M., & Hernandez, M. I. M. (2007). Use of index analysis to evaluate the water quality of a stream receiving industrial effluents. *Water SA*, 33(4), 459–465. <https://doi.org/10.4314/WSA.V33I4.52940>
- Acquavita, A., Aleffi, I. F., Benci, C., Bettoso, N., Crevatin, E., Milani, L., Tamberlich, F., Toniatti, L., Barbieri, P., Licen, S., & Mattassi, G. (2015). Annual characterization of the nutrients and trophic state in a Mediterranean coastal lagoon: The Marano and Grado Lagoon (northern Adriatic Sea). *Regional Studies in Marine Science*, 2, 132–144. <https://doi.org/10.1016/J.RSMA.2015.08.017>
- Adams, S., Titus, R., Pietersen, K., Tredoux, G., & Harris, C. (2001). Hydrochemical characteristics of aquifers near Sutherland in the Western Karoo, South Africa. *Journal of Hydrology*, 241(1–2), 91–103. [https://doi.org/10.1016/S0022-1694\(00\)00370-X](https://doi.org/10.1016/S0022-1694(00)00370-X)
- Akay, E., & Dalkıran, N. (2020). Assessing biological water quality of Yalakdere stream (Yalova, Turkey) with benthic macroinvertebrate-based metrics. *Biologia*, 75(9), 1347–1363. <https://doi.org/10.2478/S11756-019-00387-9/METRICS>
- Akbor, M. A., Uddin, M., & Ahsan, M. A. (2017). Investigation of Water Quality Parameters at Different Points in the Buriganga River, Bangladesh. *Journal of Environmental Science and Natural Resources*, 10(1), 75–80. <https://doi.org/10.3329/JESNR.V10I1.34698>
- Al-Afify, A. D. G., Tahoun, U. M., & Abdo, M. H. (2019). Water Quality Index and Microbial Assessment of Lake Qarun, El-Batts and El-Wadi Drains, Fayoum Province, Egypt. *Egyptian Journal of Aquatic Biology and Fisheries*, 23(1), 341–357. <https://doi.org/10.21608/EJABF.2019.28270>
- Al-Shujairi, S. O. H. (2013). Develop and apply water quality index to evaluate water quality of Tigris and Euphrates Rivers in Iraq. *International Journal of Modern Engineering Research*, 3(4), 2119–2126.
- Ali, S. A., & Ali, U. (2013). Litho-Structural Mapping of Sind Catchment (Kashmir Basin), NW Himalaya, using Remote Sensing & GIS Techniques. *International Journal of Science and Research*, 4, 2319–7064. [www.ijsr.net](http://www.ijsr.net)
- Ali, S. A., & Ali, U. (2018). Hydrochemical characteristics and spatial analysis of groundwater quality in parts of Bundelkhand Massif, India. *Applied Water Science*, 8(1), 1–15. <https://doi.org/10.1007/S13201-018-0678-X/FIGURES/10>

- Ali, S. A., & Pirasteh, S. (2004). Geological applications of Landsat Enhanced Thematic Mapper (ETM) data and Geographic Information System (GIS): Mapping and structural interpretation in south-west Iran, Zagros Structural Belt. *International Journal of Remote Sensing*, 25(21), 4715–4727. <https://doi.org/10.1080/01431160410001688295>
- Almeida, C., González, S. O., Mallea, M., & González, P. (2012). A recreational water quality index using chemical, physical and microbiological parameters. *Environmental Science and Pollution Research*, 19(8), 3400–3411. <https://doi.org/10.1007/S11356-012-0865-5/FIGURES/8>
- Anderson, D. M., Fensin, E., Gobler, C. J., Hoeglund, A. E., Hubbard, K. A., Kulis, D. M., Landsberg, J. H., Lefebvre, K. A., Provoost, P., Richlen, M. L., Smith, J. L., Solow, A. R., & Trainer, V. L. (2021). Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae*, 102, 101975. <https://doi.org/10.1016/J.HAL.2021.101975>
- Babaei Semiromi, F., Hassani, A. H., Torabian, A., Karbassi, A. R., & Hosseinzadeh Lotfi, F. (2011). Evolution of a new surface water quality index for Karoon catchment in Iran. *Water Science and Technology*, 64(12), 2483–2491. <https://doi.org/10.2166/WST.2011.780>
- Batur, E., & Maktav, D. (2019). Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5), 2983–2989. <https://doi.org/10.1109/TGRS.2018.2879024>
- Belle-Oudry, D. (2008). Quantitative Analysis of Sulfate in Water by Indirect EDTA Titration. *Journal of Chemical Education*, 85(9), 1269–1270. <https://doi.org/10.1021/ED085P1269>
- Bengraïne, K., & Marhaba, T. F. (2003). Using principal component analysis to monitor spatial and temporal changes in water quality. *Journal of Hazardous Materials*, 100(1–3), 179–195. [https://doi.org/10.1016/S0304-3894\(03\)00104-3](https://doi.org/10.1016/S0304-3894(03)00104-3)
- Benkov, I., Varbanov, M., Venelinov, T., & Tsakovski, S. (2023). Principal Component Analysis and the Water Quality Index—A Powerful Tool for Surface Water Quality Assessment: A Case Study on Struma River Catchment, Bulgaria. *Water* 2023, Vol. 15, Page 1961, 15(10), 1961. <https://doi.org/10.3390/W15101961>
- Bigham, T., Dooley, J. S. G., Ternan, N. G., Snelling, W. J., Héctor Castelán, M. C., & Davis, J. (2019). Assessing microbial water quality: Electroanalytical approaches to the detection of coliforms. *TrAC Trends in Analytical Chemistry*, 121, 115670. <https://doi.org/10.1016/J.TRAC.2019.115670>
- Bonadonna, L., Briancesco, R., & La Rosa, G. (2019). Innovative analytical methods for monitoring microbiological and virological water quality. *Microchemical Journal*, 150, 104160. <https://doi.org/10.1016/J.MICROC.2019.104160>
- Bordalo, A. A., Teixeira, R., & Wiebe, W. J. (2006). A water quality index applied to an international shared river basin: The case of the Douro River. *Environmental Management*, 38(6), 910–920. <https://doi.org/10.1007/S00267-004-0037-6/FIGURES/7>
- Bozorg-Haddad, O., Delpasand, M., & Loáiciga, H. A. (2021). Water quality, hygiene, and health. *Economical, Political, and Social Issues in Water Resources*, 217–257.

- <https://doi.org/10.1016/B978-0-323-90567-1.00008-5>
- Brown, R. M., McClelland, N. I., Deininger, R. A., & Landwehr, J. M. (1973). Validating the WQI. *Proceedings of the National Meeting of American Society of Civil Engineers on Water Resources Engineering, Washington, DC, USA, 29*.
- Bu, J., Liu, W., Pan, Z., & Ling, K. (2020). Comparative Study of Hydrochemical Classification Based on Different Hierarchical Cluster Analysis Methods. *International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 9515, 17(24)*, 9515. <https://doi.org/10.3390/IJERPH17249515>
- BWDB. (2023). *Flood Forecasting and Warning Centre, BWDB, Bangladesh*. <http://www.ffwc.gov.bd/>
- Carvalho, L., Cortes, R., & Bordalo, A. A. (2011). Evaluation of the ecological status of an impaired watershed by using a multi-index approach. *Environmental Monitoring and Assessment, 174(1–4)*, 493–508. <https://doi.org/10.1007/S10661-010-1473-9/METRICS>
- CCME WATER QUALITY INDEX USER'S MANUAL 2017 UPDATE. (2017). <https://ccme.ca/en/res/wqimanualen.pdf>
- Chan, J. Y. Le, Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics 2022, Vol. 10, Page 1283, 10(8)*, 1283. <https://doi.org/10.3390/MATH10081283>
- Chauhan, V. S., & Sharma, A. (2003). Studies on organoleptic properties of food products from fresh egg and egg powder through principal component analysis. *Food / Nahrung, 47(2)*, 102–105. <https://doi.org/10.1002/FOOD.200390018>
- Chowdhary, P., Bharagava, R. N., Mishra, S., Khan, N., Chowdhary, P., Bharagava, R. N., Mishra, S., & Khan, N. (2020). Role of Industries in Water Scarcity and Its Adverse Effects on Environment and Human Health. *Environmental Concerns and Sustainable Development, 235–256*. [https://doi.org/10.1007/978-981-13-5889-0\\_12](https://doi.org/10.1007/978-981-13-5889-0_12)
- Cooper, C. M. (1993). Biological Effects of Agriculturally Derived Surface Water Pollutants on Aquatic Systems—A Review. *Journal of Environmental Quality, 22(3)*, 402–408. <https://doi.org/10.2134/JEQ1993.00472425002200030003X>
- Country profile-Bangladesh. (2014). [www.fao.org/](http://www.fao.org/)
- Country Profile Bangladesh COUNTRY OVERVIEW. (2011). [http://climatechange.worldbank.org/sites/default/files/documents/EACC\\_Bangladesh.pdf](http://climatechange.worldbank.org/sites/default/files/documents/EACC_Bangladesh.pdf)
- Cude, C. G. (2001). OREGON WATER QUALITY INDEX A TOOL FOR EVALUATING WATER QUALITY MANAGEMENT EFFECTIVENESS1. *JAWRA Journal of the American Water Resources Association, 37(1)*, 125–137. <https://doi.org/10.1111/J.1752-1688.2001.TB05480.X>
- Dadolahi-Sohrab, A., Arjomand, F., & Fadaei-Nasab, M. (2012). Water quality index as a simple indicator of watersheds pollution in southwestern part of Iran. *Water and Environment Journal, 26(4)*, 445–454. <https://doi.org/10.1111/J.1747-6593.2011.00303.X>

- Devendra Swaroop Bhargave, B., & Asce, F. (1985). Expression for Drinking Water Supply Standards. *Journal of Environmental Engineering*, 111(3), 304–316. [https://doi.org/10.1061/\(ASCE\)0733-9372\(1985\)111:3\(304\)](https://doi.org/10.1061/(ASCE)0733-9372(1985)111:3(304))
- Dhany Sutadian, A., Muttill, N., Yilmaz, A., & Perera, C. (n.d.). *Development of River Water Quality Indices-A Review*.
- Dojlido, J., Raniszewski, J., & Woyciechowska, J. (1994). Water quality index applied to rivers in the Vistula river basin in Poland. *Environmental Monitoring and Assessment*, 33(1), 33–42. <https://doi.org/10.1007/BF00546659/METRICS>
- Donald, A. E., Blessing, U. A., Anyanwu, C., & Donald, E. (2019). Index approach to water quality assessment of a south eastern Nigerian river. ~ 153 ~ *International Journal of Fisheries and Aquatic Studies*, 7(1). [www.fisheriesjournal.com](http://www.fisheriesjournal.com)
- Duclos Alege, J., & Gnauck, A. (2006). *Signal Analysis of Water Quality Indicators*.
- Effendi, H., Romanto, & Wardiatno, Y. (2015). Water Quality Status of Ciambulawung River, Banten Province, Based on Pollution Index and NSF-WQI. *Procedia Environmental Sciences*, 24, 228–237. <https://doi.org/10.1016/J.PROENV.2015.03.030>
- Ewaid, S. H. (2017). Water quality evaluation of Al-Gharraf river by two water quality indices. *Applied Water Science*, 7(7), 3759–3765. <https://doi.org/10.1007/S13201-016-0523-Z/TABLES/7>
- Fatema, K., Begum, M., Zahid, M. Al, & Hossain, M. (2018). Water quality assessment of the river Buriganga, Bangladesh. *Journal of Biodiversity Conservation and Bioresource Management*, 4(1), 47–54. <https://doi.org/10.3329/JBCBM.V4I1.37876>
- Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V., & Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environmental Monitoring and Assessment*, 144(1–3), 269–276. <https://doi.org/10.1007/S10661-007-9989-3/METRICS>
- Franco, H. D., Custodio, M., Peñaloza, R., & De La Cruz, H. (2021). Application of Multivariate Statistical Methods and Water Quality Index for the Evaluation of Surface Water Quality in the Cunas River Basin, Peru. *Asian Journal of Water, Environment and Pollution*, 18(4), 19–27. <https://doi.org/10.3233/AJW210039>
- Garg, S., Chowdhury, Z. Z., Faisal, A. N. M., Rumjit, N. P., & Thomas, P. (2022). Impact of Industrial Wastewater on Environment and Human Health. *Environmental Science and Engineering*, 197–209. [https://doi.org/10.1007/978-3-030-83811-9\\_10/TABLES/1](https://doi.org/10.1007/978-3-030-83811-9_10/TABLES/1)
- Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64(11), 2409–2420. <https://doi.org/10.1016/J.MARPOLBUL.2012.08.005>
- Girones, R., Ferrús, M. A., Alonso, J. L., Rodriguez-Manzano, J., Calgua, B., de Abreu Corrêa, A., Hundesa, A., Carratala, A., & Bofill-Mas, S. (2010). Molecular detection of pathogens in water – The pros and cons of molecular techniques. *Water Research*, 44(15), 4325–4339. <https://doi.org/10.1016/J.WATRES.2010.06.030>
- Goodman, C. M. (1987). The Delphi technique: a critique. *Journal of Advanced Nursing*, 12(6),

- 729–734. <https://doi.org/10.1111/J.1365-2648.1987.TB01376.X>
- Gupta, S., & Gupta, S. K. (2021). A critical review on water quality index tool: Genesis, evolution and future directions. *Ecological Informatics*, *63*, 101299. <https://doi.org/10.1016/J.ECOINF.2021.101299>
- Hasan, H. H., Jamil, N. R., & Aini, N. (2015). Water Quality Index and Sediment Loading Analysis in Pelus River, Perak, Malaysia. *Procedia Environmental Sciences*, *30*, 133–138. <https://doi.org/10.1016/J.PROENV.2015.10.024>
- Horton, R. K. (1965). An index number system for rating water quality. *J Water Pollut Control Fed*, *37*(3), 300–306.
- Hoseinzadeh, E., Khorsandi, H., Wei, C., & Alipour, M. (2015). Evaluation of Aydughmush River water quality using the National Sanitation Foundation Water Quality Index (NSFWQI), River Pollution Index (RPI), and Forestry Water Quality Index (FWQI). *Desalination and Water Treatment*, *54*(11), 2994–3002. <https://doi.org/10.1080/19443994.2014.913206>
- Hsu, C.-C., & Sandford, B. A. (2019). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research, and Evaluation*, *12*(1), 10. <https://doi.org/https://doi.org/10.7275/pdz9-th90>
- Huang, F., Wang, X., Lou, L., Zhou, Z., & Wu, J. (2010). Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water Research*, *44*(5), 1562–1572. <https://doi.org/10.1016/J.WATRES.2009.11.003>
- Hudson, N., Baker, A., Ward, D., Reynolds, D. M., Brunson, C., Carliell-Marquet, C., & Browning, S. (2008). Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water quality assessment? An example from South West England. *Science of The Total Environment*, *391*(1), 149–158. <https://doi.org/10.1016/J.SCITOTENV.2007.10.054>
- Huey, G. M., & Meyer, M. L. (2010). Turbidity as an Indicator of Water Quality in Diverse Watersheds of the Upper Pecos River Basin. *Water 2010, Vol. 2, Pages 273-284*, *2*(2), 273–284. <https://doi.org/10.3390/W2020273>
- IONUȘ, O. (2010). WATER QUALITY INDEX-ASSESSMENT METHOD OF THE MOTRU RIVER WATER QUALITY (OLTENIA, ROMANIA) INDICELE DE CALITATE A APEI-METODĂ DE DETERMINARE A CALITĂȚII APEI RÂULUI MOTRU (OLTENIA, ROMÂNIA). *OF THE UNIVERSITY OF CRAIOVA*, *74*.
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). <https://doi.org/10.1098/RSTA.2015.0202>
- Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental Monitoring and Assessment*, *173*(1–4), 625–641. <https://doi.org/10.1007/S10661-010-1411-X/METRICS>
- Kabir, A., Sraboni, H. J., Hasan, M. M., & Sorker, R. (2022). Eco-environmental assessment of the Turag River in the megacity of Bangladesh. *Environmental Challenges*, *6*, 100423.

<https://doi.org/10.1016/J.ENVC.2021.100423>

- Kamal, M. M., Malmgren-Hansen, A., & Badruzzaman, A. B. M. (1999). Assessment of pollution of the river Buriganga, Bangladesh, using a water quality model. *Water Science and Technology*, 40(2), 129–136. <https://doi.org/10.2166/WST.1999.0104>
- Kannel, P. R., Lee, S., Lee, Y. S., Kanel, S. R., & Khan, S. P. (2007). Application of water quality indices and dissolved oxygen as indicators for river water classification and urban impact assessment. *Environmental Monitoring and Assessment*, 132(1–3), 93–110. <https://doi.org/10.1007/S10661-006-9505-1/METRICS>
- Kasiske, D., Klinkmüller, K. D., & Sonneborn, M. (1978). Application of high-performance liquid chromatography to water pollution analysis. *Journal of Chromatography A*, 149(C), 703–710. [https://doi.org/10.1016/S0021-9673\(00\)81022-5](https://doi.org/10.1016/S0021-9673(00)81022-5)
- Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., Sarfraz, R. A., Baig, J. A., & Shah, A. Q. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, 72(2), 301–309. <https://doi.org/10.1016/J.ECOENV.2008.02.024>
- Khan, M. H., Akash, N. M., Akter, S., Rukh, M., Nzediegwu, C., & Islam, M. S. (2023). A comprehensive review of coconut-based porous materials for wastewater treatment and CO<sub>2</sub> capture. *Journal of Environmental Management*, 338, 117825. <https://doi.org/10.1016/J.JENVMAN.2023.117825>
- Khatun, M. A., Rashid, M. B., & Hygen, H. O. (n.d.). *Climate of Bangladesh*.
- Khuan, L. Y., Hamzah, N., & Jailani, R. (2002). Prediction of water quality index (WQI) based on artificial neural network (ANN). *2002 Student Conference on Research and Development: Globalizing Research and Development in Electrical and Electronics Engineering, SCOReD 2002 - Proceedings*, 157–161. <https://doi.org/10.1109/SCORED.2002.1033081>
- Koçer, M. A. T., & Sevgili, H. (2014). Parameters selection for water quality index in the assessment of the environmental impacts of land-based trout farms. *Ecological Indicators*, 36, 672–681. <https://doi.org/10.1016/J.ECOLIND.2013.09.034>
- Kumar, R., Dutt, V., Raina, A., & Sharma, N. (2022). Spatial water quality assessment of a mountain stream in northwestern India using multivariate statistical techniques. *Environmental Monitoring and Assessment*, 194(10), 1–24. <https://doi.org/10.1007/S10661-022-10386-0/TABLES/7>
- Lermontov, A., Yokoyama, L., Lermontov, M., & Machado, M. A. S. (2009). River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. *Ecological Indicators*, 9(6), 1188–1197. <https://doi.org/10.1016/J.ECOLIND.2009.02.006>
- Li, J., Abdulmohsin, H. A., Hasan, S. S., Kaiming, L., Al-Khateeb, B., Ghareb, M. I., & Mohammed, M. N. (2019). Hybrid soft computing approach for determining water quality indicator: Euphrates River. *Neural Computing and Applications*, 31(3), 827–837. <https://doi.org/10.1007/S00521-017-3112-7/TABLES/2>
- Linstone, H. A. (1985). The Delphi technique. *Environmental Impact Assessment, Technology*

- Assessment, and Risk Analysis. Proc., Les Arcs, 1983, 621–649.*  
[https://doi.org/10.1007/978-3-642-70634-9\\_22/COVER](https://doi.org/10.1007/978-3-642-70634-9_22/COVER)
- Liou, S. M., Lo, S. L., & Wang, S. H. (2004). A generalized water quality index for Taiwan. *Environmental Monitoring and Assessment, 96*(1–3), 35–52.  
<https://doi.org/10.1023/B:EMAS.0000031715.83752.A1/METRICS>
- Liu, C. W., Lin, K. H., & Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of The Total Environment, 313*(1–3), 77–89. [https://doi.org/10.1016/S0048-9697\(02\)00683-6](https://doi.org/10.1016/S0048-9697(02)00683-6)
- Lobato, T. C., Hauser-Davis, R. A., Oliveira, T. F., Silveira, A. M., Silva, H. A. N., Tavares, M. R. M., & Saraiva, A. C. F. (2015). Construction of a novel water quality index and quality indicator for reservoir water quality evaluation: A case study in the Amazon region. *Journal of Hydrology, 522*, 674–683.  
<https://doi.org/10.1016/J.JHYDROL.2015.01.021>
- Lu, F., Chen, Z., & Liu, W. (2014). A Gis-based system for assessing marine water quality around offshore platforms. *Ocean & Coastal Management, 102*(PA), 294–306.  
<https://doi.org/10.1016/J.OCECOAMAN.2014.10.003>
- Lu, G. Y., & Wong, D. W. (2008). An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences, 34*(9), 1044–1055.  
<https://doi.org/10.1016/J.CAGEO.2007.07.010>
- Lumb, A., Sharma, T C, Bibeault, J.-F., Sharma, T. C., & Bibeault, J.-F. (2011). A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. *Water Quality, Exposure and Health 2011 3:1, 3*(1), 11–24.  
<https://doi.org/10.1007/S12403-011-0040-0>
- Ma, Z., Li, H., Ye, Z., Wen, J., Hu, Y., & Liu, Y. (2020). Application of modified water quality index (WQI) in the assessment of coastal water quality in main aquaculture areas of Dalian, China. *Marine Pollution Bulletin, 157*, 111285.  
<https://doi.org/10.1016/J.MARPOLBUL.2020.111285>
- Mahapatra, S. S., Sahu, M., Patel, R K, Biranchi, , Panda, N., Sahu, M., Patel, R. K., & Panda, B. N. (2012). Prediction of Water Quality Using Principal Component Analysis. *Water Quality, Exposure and Health 2012 4:2, 4*(2), 93–104.  
<https://doi.org/10.1007/S12403-012-0068-9>
- Maskooni, E. K., Naseri-Rad, M., Berndtsson, R., & Nakagawa, K. (2020). Use of Heavy Metal Content and Modified Water Quality Index to Assess Groundwater Quality in a Semiarid Area. *Water 2020, Vol. 12, Page 1115, 12*(4), 1115. <https://doi.org/10.3390/W12041115>
- Matthew, C., Lawoko, C. R. O., Korte, C. J., & Smith, D. (1994). Application of canonical discriminant analysis, principal component analysis, and canonical correlation analysis as tools for evaluating differences in pasture botanical composition. *New Zealand Journal of Agricultural Research, 37*(4), 509–520. <https://doi.org/10.1080/00288233.1994.9513090>
- McGarigal, K., Stafford, S., & Cushman, S. (2000). Multivariate Statistics for Wildlife and Ecology Research. *Multivariate Statistics for Wildlife and Ecology Research.*  
<https://doi.org/10.1007/978-1-4612-1288-1>

- Ministry of Water Resource. (2001). *National Water Management Plan*. [https://warpo.portal.gov.bd/sites/default/files/files/warpo.portal.gov.bd/page/d921b920\\_da0c\\_4775\\_a5c2\\_5c33e6938232/nwmp\\_vol4.pdf](https://warpo.portal.gov.bd/sites/default/files/files/warpo.portal.gov.bd/page/d921b920_da0c_4775_a5c2_5c33e6938232/nwmp_vol4.pdf)
- Moog, O., & Scientific Conference Rivers in the Hindu Kush-Himalaya - Ecology & Environmental Assessment 2008 Kāthamāṇḍū; Dhulikhel. (n.d.). *ASSESS-HKH a product of ASSESS-HKH, development of an assessment system to evaluate the ecological status of rivers in the Hindu Kush-Himalayan region ; [proceedings of the Scientific Conference Rivers in the Hindu Kush-Himalaya - Ecology & Environmental Assessment]*.
- Mukate, S., Wagh, V., Panaskar, D., Jacobs, J. A., & Sawant, A. (2019). Development of new integrated water quality index (IWQI) model to evaluate the drinking suitability of water. *Ecological Indicators*, *101*, 348–354. <https://doi.org/10.1016/J.ECOLIND.2019.01.034>
- Nandini, N., Jumbe, A. S., Tandon, S., & N., S. (2007). Application of Remote Sensing, GIS, and GPS Technology on Environment and Earth's Natural Resources Management An Introduction. *Mapana Journal of Sciences*, *6*(1), 96–117. <https://doi.org/10.12723/MJS.10.8>
- Nath, B. K., Chaliha, C., Bhuyan, B., Kalita, E., Baruah, D. C., & Bhagabati, A. K. (2018). GIS mapping-based impact assessment of groundwater contamination by arsenic and other heavy metal contaminants in the Brahmaputra River valley: A water quality assessment study. *Journal of Cleaner Production*, *201*, 1001–1011. <https://doi.org/10.1016/J.JCLEPRO.2018.08.084>
- Naubi, I., Zardari, N. H., Shirazi, S. M., Ibrahim, N. F. B., & Baloo, L. (2016). Effectiveness of Water Quality Index for Monitoring Malaysian River Water Quality. *Polish Journal of Environmental Studies*, *25*(1), 231–239. <https://doi.org/10.15244/PJOES/60109>
- Naushad Alam, M., Fazle Elahi, P., & Md Didar-UI-Alam, P. (n.d.). *Risk and Water Quality Assessment over view of River Sitalakhya in Bangladesh*. Retrieved October 8, 2023, from [www.acadjournal.com](http://www.acadjournal.com)
- Navarro, A., Tauler, R., Lacorte, S., & Barceló, D. (2010). Occurrence and transport of pesticides and alkylphenols in water samples along the Ebro River Basin. *Journal of Hydrology*, *383*(1–2), 18–29. <https://doi.org/10.1016/J.JHYDROL.2009.06.039>
- Nazma Sultana, M., Sahadat Hossain, M., & Ara Latifa, G. (2019). WATER QUALITY ASSESSMENT OF BALU RIVER, DHAKA BANGLADESH. *Water Conservation & Management (WCM)*. <https://doi.org/10.26480/wcm.02.2019.08.10>
- Nong, X., Shao, D., Zhong, H., & Liang, J. (2020). Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Research*, *178*, 115781. <https://doi.org/10.1016/J.WATRES.2020.115781>
- Noori, R., Berndtsson, R., Hosseinzadeh, M., Adamowski, J. F., & Abyaneh, M. R. (2019). A critical review on the application of the National Sanitation Foundation Water Quality Index. *Environmental Pollution*, *244*, 575–587. <https://doi.org/10.1016/J.ENVPOL.2018.10.076>
- Oketola, A. A., Adekolurejo, S. M., Osibanjo, O., Oketola, A. A., Adekolurejo, S. M., & Osibanjo, O. (2013). Water Quality Assessment of River Ogun Using Multivariate Statistical Techniques. *Journal of Environmental Protection*, *4*(5), 466–479.

- <https://doi.org/10.4236/JEP.2013.45055>
- Olsen, R. L., Chappell, R. W., & Loftis, J. C. (2012). Water quality sample collection, data treatment and results presentation for principal components analysis – literature review and Illinois River watershed case study. *Water Research*, *46*(9), 3110–3122. <https://doi.org/10.1016/J.WATRES.2012.03.028>
- Omer, N. H., & Omer, N. H. (2019). Water Quality Parameters. *Water Quality - Science, Assessments and Policy*. <https://doi.org/10.5772/INTECHOPEN.89657>
- Oseke, F. I., Anornu, G. K., Adjei, K. A., & Eduvie, M. O. (2021). Assessment of water quality using GIS techniques and water quality index in reservoirs affected by water diversion. *Water-Energy Nexus*, *4*, 25–34. <https://doi.org/10.1016/J.WEN.2020.12.002>
- Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., & Huang, C. H. (2006). Assessment of seasonal variations in surface water quality. *Water Research*, *40*(20), 3800–3810. <https://doi.org/10.1016/J.WATRES.2006.08.030>
- Ouyang, Ying. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water Research*, *39*(12), 2621–2635. <https://doi.org/10.1016/J.WATRES.2005.04.024>
- Parmar, K. S., & Bhardwaj, R. (2014). Water quality management using statistical analysis and time-series prediction model. *Applied Water Science*, *4*(4), 425–434. <https://doi.org/10.1007/S13201-014-0159-9/FIGURES/3>
- Peche, R., & Rodríguez, E. (2012). Development of environmental quality indexes based on fuzzy logic. A case study. *Ecological Indicators*, *23*, 555–565. <https://doi.org/10.1016/J.ECOLIND.2012.04.029>
- Pesce, S. F., & Wunderlin, D. A. (2000). Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquía River. *Water Research*, *34*(11), 2915–2926. [https://doi.org/10.1016/S0043-1354\(00\)00036-1](https://doi.org/10.1016/S0043-1354(00)00036-1)
- Pinto, L. C., de Mello, C. R., & Ávila, L. F. (2013). Water quality indicators in the Mantiqueira Range region, Minas Gerais state. *CERNE*, *19*(4), 687–692. <https://doi.org/10.1590/S0104-77602013000400020>
- Platikanov, S., Rodriguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., & Tauler, R. (2014). Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. *Journal of Environmental Management*, *140*, 33–44. <https://doi.org/10.1016/J.JENVMAN.2014.03.006>
- Principal Component Analysis for Special Types of Data. (2002). *Principal Component Analysis*, 338–372. [https://doi.org/10.1007/0-387-22440-8\\_13](https://doi.org/10.1007/0-387-22440-8_13)
- Rahman, A., Islam, M., Hossain, M., & Ahsan, M. A. (2012). Study of the seasonal variations in Turag river water quality parameters. *African Journal of Pure and Applied Chemistry*, *6*(10). <https://doi.org/10.5897/AJPAC12.023>
- Rahman, M. A., & Al Bakri, D. (2010). A Study on Selected Water Quality Parameters along the River Buriganga, Bangladesh. *Official Peer Reviewed Journal of Babol Noshirvani University of Technology BUT*, *1*(2), 81–92.

- Rahman, S., & Hossain, F. (2008). Spatial assessment of water quality in Peripheral Rivers of Dhaka City for optimal relocation of water intake point. *Water Resources Management*, 22(3), 377–391. <https://doi.org/10.1007/S11269-007-9167-Y/METRICS>
- Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology* 2008 26:3, 26(3), 303–304. <https://doi.org/10.1038/nbt0308-303>
- Rocha, F. C., Andrade, E. M., & Lopes, F. B. (2015). Water quality index calculated from biological, physical and chemical attributes. *Environmental Monitoring and Assessment*, 187(1), 1–15. <https://doi.org/10.1007/S10661-014-4163-1/TABLES/9>
- Roy, R. (2018). An Introduction to water quality analysis. *ESSENCE – International Journal for Environmental Rehabilitation and Conservation*, 94–100. <https://doi.org/10.31786/09756272.18.9.2.214>
- Said, A., Stevens, D. K., & Sehlke, G. (2004). An innovative index for evaluating water quality in streams. *Environmental Management*, 34(3), 406–414. <https://doi.org/10.1007/S00267-004-0210-Y/FIGURES/3>
- Salman, M. A., Ahmed, S., Peas, M. H., & Khan, N. (2018). Water quality assessment of the Buriganga river, Dhaka, Bangladesh. *International Journal of Emerging Technology and Advanced Engineering*, 8(6), 17–23.
- Sánchez, E., Colmenarejo, M. F., Vicente, J., Rubio, A., García, M. G., Travieso, L., & Borja, R. (2007). Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators*, 7(2), 315–328. <https://doi.org/10.1016/J.ECOLIND.2006.02.005>
- Sarkar, C., & Abbasi, S. A. (2006). Qualidex - A new software for generating water quality indice. *Environmental Monitoring and Assessment*, 119(1–3), 201–231. <https://doi.org/10.1007/S10661-005-9023-6/METRICS>
- Sayed Ahammed, S., Tasfina, S., Ayaz Rabbani, K., & Adbul Khaleque, M. (2016). An Investigation Into The Water Quality Of Buriganga-A River Running Through Dhaka. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 5(03). [www.ijstr.org](http://www.ijstr.org)
- Shah, K. A., & Joshi, G. S. (2017). Evaluation of water quality index for River Sabarmati, Gujarat, India. *Applied Water Science*, 7(3), 1349–1358. <https://doi.org/10.1007/S13201-015-0318-7/TABLES/5>
- Sharma, A., Singh, A. K., & Kumar, K. (2012). Environmental geochemistry and quality assessment of surface and subsurface water of Mahi River basin, western India. *Environmental Earth Sciences*, 65(4), 1231–1250. <https://doi.org/10.1007/S12665-011-1371-7/FIGURES/11>
- Shibata, T., Solo-Gabriele, H. M., Fleming, L. E., & Elmir, S. (2004). Monitoring marine recreational water quality using multiple microbial indicators in an urban tropical environment. *Water Research*, 38(13), 3119–3131. <https://doi.org/10.1016/J.WATRES.2004.04.044>
- Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental*

- Modelling & Software*, 22(4), 464–475.  
<https://doi.org/10.1016/J.ENVSOF.2006.02.001>
- Singh, A. K., Mondal, G. C., Kumar, S., Singh, T. B., Tewary, B. K., & Sinha, A. (2008). Major ion chemistry, weathering processes and water quality assessment in upper catchment of Damodar River basin, India. *Environmental Geology*, 54(4), 745–758.  
<https://doi.org/10.1007/S00254-007-0860-1/FIGURES/13>
- Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research*, 38(18), 3980–3992.  
<https://doi.org/10.1016/J.WATRES.2004.06.011>
- Singh, S. K., Singh, P., & Gautam, S. K. (2016). Appraisal of urban lake water quality through numerical index, multivariate statistics and earth observation data sets. *International Journal of Environmental Science and Technology*, 13(2), 445–456.  
<https://doi.org/10.1007/S13762-015-0850-X/TABLES/6>
- Stoner, J. D., Andrus, C. D., & Menard, H. W. (n.d.). *Water-Quality Indices for Specific Water Uses United States Department of the Interior Geological Survey*.
- Subramanian, A., & Baskar, S. (2022). Water quality assessment of Noyyal river using water quality index (WQI) and multivariate statistical techniques. *Water Science*, 36(1), 85–98.  
<https://doi.org/10.1080/23570008.2022.2129146>
- Summers, K., & Summers, K. (2020). Water Quality - Science, Assessments and Policy. *Water Quality - Science, Assessments and Policy*. <https://doi.org/10.5772/INTECHOPEN.77531>
- Sun, W., Xia, C., Xu, M., Guo, J., & Sun, G. (2016). Application of modified water quality indices as indicators to assess the spatial and temporal trends of water quality in the Dongjiang River. *Ecological Indicators*, 66, 306–312.  
<https://doi.org/10.1016/J.ECOLIND.2016.01.054>
- Sutadian, A. D., Muttill, N., Yilmaz, A. G., & Perera, B. J. C. (2017). Using the Analytic Hierarchy Process to identify parameter weights for developing a water quality index. *Ecological Indicators*, 75, 220–233. <https://doi.org/10.1016/J.ECOLIND.2016.12.043>
- Sutadian, A. D., Muttill, N., Yilmaz, A. G., & Perera, B. J. C. (2018). Development of a water quality index for rivers in West Java Province, Indonesia. *Ecological Indicators*, 85, 966–982. <https://doi.org/10.1016/J.ECOLIND.2017.11.049>
- Swamee, P. K., & Tyagi, A. (2000). Describing Water Quality with Aggregate Index. *Journal of Environmental Engineering*, 126(5), 451–455. [https://doi.org/10.1061/\(ASCE\)0733-9372\(2000\)126:5\(451\)](https://doi.org/10.1061/(ASCE)0733-9372(2000)126:5(451))
- Swamee, P. K., & Tyagi, A. (2007). Improved Method for Aggregation of Water Quality Subindices. *Journal of Environmental Engineering*, 133(2), 220–225. [https://doi.org/10.1061/\(ASCE\)0733-9372\(2007\)133:2\(220\)/ASSET/D70772CB-9778-4BC0-AEE9-FED8EAA111BA/ASSETS/IMAGES/LARGE/6.JPG](https://doi.org/10.1061/(ASCE)0733-9372(2007)133:2(220)/ASSET/D70772CB-9778-4BC0-AEE9-FED8EAA111BA/ASSETS/IMAGES/LARGE/6.JPG)
- Tan, G., Yan, J., Gao, C., & Yang, S. (2012). Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering*, 31, 1194–1199.  
<https://doi.org/10.1016/J.PROENG.2012.01.1162>

- Tharmar, E., Abraham, M., Prakash, R., Sundaram, A., Flores, E. S., Canales, C., & Alam, M. A. (2022). Hydrogeochemistry and Water Quality Assessment in the Thamirabarani River Stretch by Applying GIS and PCA Techniques. *Sustainability* 2022, Vol. 14, Page 16368, 14(24), 16368. <https://doi.org/10.3390/SU142416368>
- Tirkey, P., Bhattacharya, T., Chakraborty, S., Poonam, T., Tanushree, B., & Sukalyan, C. (2013). Water quality indices-important tools for water quality assessment: a review. *International Journal of Advances in Chemistry (IJAC)*, 1(1). <https://www.researchgate.net/publication/262730848>
- Tomas, D., Čurlin, M., & Marić, A. S. (2017). Assessing the surface water status in Pannonian ecoregion by the water quality index model. *Ecological Indicators*, 79, 182–190. <https://doi.org/10.1016/J.ECOLIND.2017.04.033>
- Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., & Nõges, T. (2016). First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sensing* 2016, Vol. 8, Page 640, 8(8), 640. <https://doi.org/10.3390/RS8080640>
- Uddin, G., Nash, S., & Olbert, A. I. (2022). Optimization of parameters in a water quality index model using principal component analysis. 5739–5744. <https://doi.org/10.3850/IAHR-39WC2521711920221326>
- Uddin, M. G., Moniruzzaman, M., & Khan, M. (2017). Evaluation of Groundwater Quality Using CCME Water Quality Index in the Rooppur Nuclear Power Plant Area, Ishwardi, Pabna, Bangladesh. *American Journal of Environmental Protection*, Vol. 5, 2017, Pages 33-43, 5(2), 33–43. <https://doi.org/10.12691/ENV-5-2-2>
- Uddin, M. G., Nash, S., & Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218. <https://doi.org/10.1016/J.ECOLIND.2020.107218>
- Ustaoğlu, F., Tepe, Y., & Taş, B. (2020). Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. *Ecological Indicators*, 113, 105815. <https://doi.org/10.1016/J.ECOLIND.2019.105815>
- Uthicke, S., Vogel, N., Doyle, J., Schmidt, C., & Humphrey, C. (2012). Interactive effects of climate change and eutrophication on the dinoflagellate-bearing benthic foraminifer *Marginopora vertebralis*. *Coral Reefs*, 31(2), 401–414. <https://doi.org/10.1007/S00338-011-0851-2/FIGURES/5>
- Wang, L., Mei, Y., Yu, K., Li, Y., Meng, X., & Hu, F. (2019). Anthropogenic Effects on Hydrogeochemical Characterization of the Shallow Groundwater in an Arid Irrigated Plain in Northwestern China. *Water* 2019, Vol. 11, Page 2247, 11(11), 2247. <https://doi.org/10.3390/W11112247>
- Whitehead, P., Bussi, G., Hossain, M. A., Dolk, M., Das, P., Comber, S., Peters, R., Charles, K. J., Hope, R., & Hossain, S. (2018). Restoring water quality in the polluted Turag-Tongi-Balu river system, Dhaka: Modelling nutrient and total coliform intervention strategies. *Science of The Total Environment*, 631–632, 223–232. <https://doi.org/10.1016/J.SCITOTENV.2018.03.038>
- Wu, C. Y., Mossa, J., Mao, L., & Almulla, M. (2019). Comparison of different spatial

- interpolation methods for historical hydrographic data of the lowermost Mississippi River. *Annals of GIS*, 25(2), 133–151. <https://doi.org/10.1080/19475683.2019.1588781>
- Wunderlin, D. A., María Del Pilar, D., María Valeria, A., Fabiana, P. S., Cecilia, H. A., & María De Los Ángeles, B. (2001). Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study:: Suquía River Basin (Córdoba–Argentina). *Water Research*, 35(12), 2881–2894. [https://doi.org/10.1016/S0043-1354\(00\)00592-3](https://doi.org/10.1016/S0043-1354(00)00592-3)
- Xia, R., & Chen, Z. (2015). Integrated Water-Quality Assessment of the Huai River Basin in China. *Journal of Hydrologic Engineering*, 20(2), 05014018. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001030/ASSET/2D06B252-D463-4117-8088-D8E15E4CA10D/ASSETS/IMAGES/LARGE/FIGURE3.JPG](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001030/ASSET/2D06B252-D463-4117-8088-D8E15E4CA10D/ASSETS/IMAGES/LARGE/FIGURE3.JPG)
- Yang, Q., Hou, Z., Wang, Y., Zhao, Y., & Delgado, J. (2015). A comparative study of shallow groundwater level simulation with WA–ANN and ITS model in a coastal island of south China. *Arabian Journal of Geosciences*, 8(9), 6583–6593. <https://doi.org/10.1007/S12517-014-1706-2/TABLES/1>
- Yang, Qingchun, Li, Z., Ma, H., Wang, L., & Martín, J. D. (2016). Identification of the hydrogeochemical processes and assessment of groundwater quality using classic integrated geochemical methods in the Southeastern part of Ordos basin, China. *Environmental Pollution*, 218, 879–888. <https://doi.org/10.1016/J.ENVPOL.2016.08.017>
- Yang, Qingchun, Zhang, J., Wang, Y., Fang, Y., & Martín, J. D. (2015). Multivariate Statistical Analysis of Hydrochemical Data for Shallow Ground Water Quality Factor Identification in a Coastal Aquifer. *Polish Journal of Environmental Studies*, 24(2), 769–776. <https://doi.org/10.15244/PJOES/30263>
- Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2), 265–286. <https://doi.org/10.1198/106186006X113430>

## Appendix C: Site Selection for Sample Collection

Site Selection was done with a prior site visit and inspection. A total 24 sites were selected for sample collection with 6 sites for each river. Following is a summary table for selected sampling sites with their coordinates.

Table 8-1: Sampling sites with coordinates

River Name	Site Code	Sampling Sites	Coordinates	
			Latitude	Longitude
<b>Buriganga</b>	BG1	<i>Pagla Ghat, Dhaka</i>	23°39'40.3"N	90°27'15.1"E
	BG2	<i>Postogola Bridge, Dhaka</i>	23°41'16.0"N	90°25'37.8"E
	BG3	<i>Sadarghat Launch Terminal</i>	23°42'21.6"N	90°24'20.8"E
	BG4	<i>Chandir Ghat</i>	23°42'39.8"N	90°23'26.5"E
	BG5	<i>Kholamora Ghat, Kamranghirchar</i>	23°42'53.0"N	90°21'34.9"E
	BG6	<i>Gudaraghat, Hazaribag</i>	23°44'02.1"N	90°21'13.9"E
<b>Turag</b>	TR1	<i>Near Fulpukuria Thread, Tongi, Gazipur</i>	23°53'51.6"N	90°26'05.4"E
	TR2	<i>Tongi Rail Bridge, Gazipur</i>	23°52'55.4"N	90°24'20.3"E
	TR3	<i>Ashulia Kacha Bazar</i>	23°53'48.0"N	90°20'03.3"E
	TR4	<i>Ashulia Landing Station</i>	23°53'29.5"N	90°21'34.7"E
	TR5	<i>Rustompur Ghat, Mirpur road</i>	23°52'29.2"N	90°21'00.8"E
	TR6	<i>Diabari Ghat, Mirpur road</i>	23°47'53.9"N	90°20'24.2"E
<b>Balu</b>	BL1	<i>Near Balu Bridge, 300 feet, Dhaka</i>	23°50'04.6"N	90°28'41.7"E
	BL2	<i>Near LGED Bridge, Jolshiri</i>	23°48'58.5"N	90°29'09.9"E
	BL3	<i>Beraid Boat Ghat, Beraid</i>	23°48'05.9"N	90°28'53.9"E
	BL4	<i>Near Edarkandi-Fakirkhali Road</i>	23°46'36.9"N	90°28'43.4"E
	BL5	<i>Near Eastern Straw Board and Paper Mill</i>	23°45'23.8"N	90°29'11.1"E
	BL6	<i>Rajakhali Ghat, Demra</i>	23°44'07.3"N	90°29'28.3"E
<b>Sitalakhya</b>	SL1	<i>Kholapara-Ghorashal Kheyaghat, Ghorashal</i>	23°56'27.4"N	90°36'57.4"E
	SL2	<i>Kaliganj Kazir Char Kheyar Ghat, Kaliganj</i>	23°54'57.1"N	90°34'01.9"E
	SL3	<i>Beldi Bazar, Beldi</i>	23°52'01.6"N	90°33'17.2"E
	SL4	<i>Near Hatabo Bazar, Rupganj</i>	23°48'33.6"N	90°32'37.2"E
	SL5	<i>Habib Ghat, Boralu Bazar</i>	23°45'46.8"N	90°30'28.0"E
	SL6	<i>South Rupshi Masjid Ghat, Rupshi</i>	23°43'55.1"N	90°30'30.8"E

Following are the pictures taken in inspection visits for site selection -

### Surrounding in BG1 Site



### Surrounding in BG2 Site



### Surrounding in BG3 Site



### Surrounding in BG4 Site



### Surrounding in BG5 Site



### Surrounding in BG6 Site



### Surrounding in TR1 Site



### Surrounding in TR2 Site



### Surrounding in TR3 Site



### Surrounding in TR4 Site



### Surrounding in TR5 Site



### Surrounding in TR6 Site



### Surrounding in BL1 Site



### Surrounding in BL3 Site



### Surrounding in BL4 Site



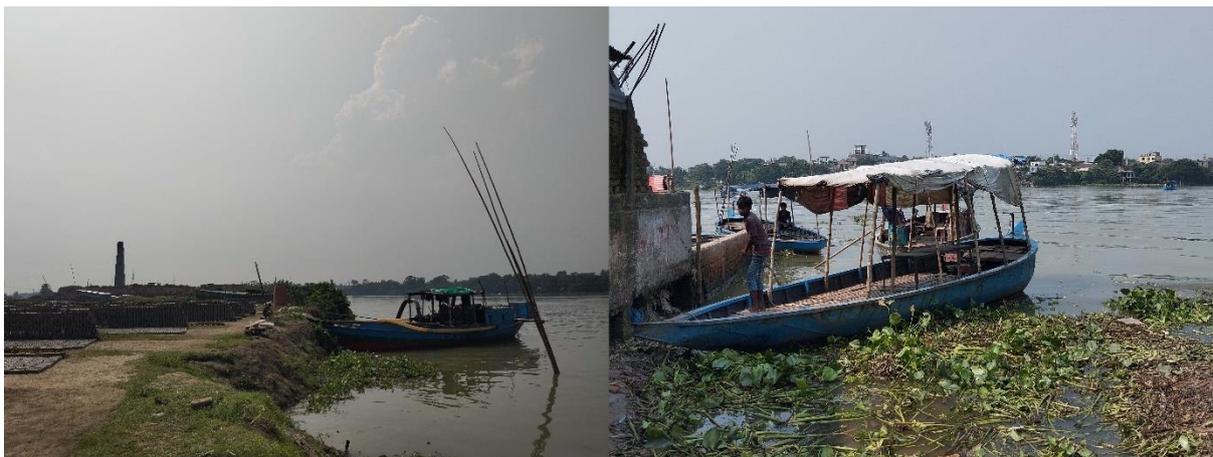
### Surrounding in SL1 Site



### Surrounding in SL2 Site



### Surrounding in SL3 Site



### Surrounding in SL4 Site



### Surrounding in SL5 Site



### Surrounding in SL6 Site



## Appendix D: Field Visit Images



*Figure 8-1: Field visit at Buriganga (BG4)*



Figure 8-2: Field visit at Turag (TR1)



*Figure 8-3: Field visit at Balu (BL1)*



Figure 8-4: Field visit at Shitalakhya (SL1)