Assessment and Modelling of River Water-Sediment Pollution Load based on Spatial and Temporal Variations

A thesis submitted

in partial fulfilment of the requirement

for the degree of

Doctor of Philosophy

by

Ankit Pratim Goswami

Department of Civil Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India
December 2022

Department of Civil Engineering
Indian Institute of Technology Guwahati
Assam-781039, India

CANDIDATE'S DECLARATION

I, **Ankit Pratim Goswami**, declare that this thesis titled "**Assessment and Modelling of River Water-Sediment Pollution Load based on Spatial and Temporal Variations**" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this Institute.
- In full or in portions, the contents of this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have clarified exactly what was done by others and what I have contributed myself.

Date:

Ankit Pratim Goswami
Registration No.: 176104105

Department of Civil Engineering
Indian Institute of Technology Guwahati
Assam-781039, India

CERTIFICATE

This is to certify that the thesis entitled **'Assessment and Modelling of River Water-Sediment Pollution Load based on Spatial and Temporal Variations** submitted by **Ankit Pratim Goswami (Roll No. 176104105),** a Research Scholar in the Department of Civil Engineering, Indian Institute of Technology Guwahati, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out by him under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and, in my opinion, has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for award of any degree or diploma.

Date:

Dr. Ajay Kalamdhad
Professor
Department of Civil Engineering
Indian Institute of Technology Guwahati

ACKNOWLEDGEMENT

As time draws near to bid adieu to the beautiful campus of the Indian Institute of Technology Guwahati, this is a righteous opportunity to mention the name of those wonderful and supportive persons who have shown me the right path and made conspicuous contributions to the completion of my research work. Without these supporters, I may not have gotten to where I am today, at least not sanely.

This Doctoral dissertation would not have been possible single-handedly without the help of many individuals over the past four years. First and foremost, I extend my sincere gratitude to my supervisor **Prof. Dr. Ajay Kalamdhad**, who has been my true mentor. This thesis is a result of his motivation and guidance throughout my research. His belief in my abilities has always inspired me to remain dedicated and focused. To the best of my abilities, I will try to carry forward the principles and work culture he instilled in me during these years.

I express my sincere gratitude to my Doctoral Committee, **Prof. Dr. Bimlesh Kumar (Chairman)**, **Prof. Dr. Vaibhav V. Goud**, and **Dr. Rishikesh Bharti**, for their valuable time and insightful suggestions provided at various stages of my study to improve the quality of my research work. I also extend my sincere thanks to the Director of IIT Guwahati, **Prof. Dr. T. G. Sitharam**, for providing the necessary facilities and a conducive academic environment. I also express my sincere thanks to the current head, **Prof. Dr. Sharad Gokhale** and former head, **Prof. Dr. Chandan Mahanta** of the Department of Civil Engineering, IIT Guwahati, for providing the necessary facilities. I am also grateful to the current Lab Incharge, **Prof. Dr. Saswati Chakraborty** and former Lab Incharge, **Prof. Dr. Pranab Kumar Ghosh** of Environmental Engineering Laboratory of Department of Civil Engineering, IIT Guwahati, for providing me with all the necessary facilities for the advancement of my research. I gratefully acknowledge the unstinted help provided by Ms. Jonali Saikia, Mr. Payodhar Pathak and Mr. Chitta Ranjan Medhi during all phases of my research work.

I am thankful to Central Instruments Facility, IIT Guwahati, for allowing me to carry out sample analysis to accomplish my PhD thesis objectives. I also wish to acknowledge the Ministry of Human Resource Development, India, for providing me with the fellowship to pursue my PhD.

I wish to express my gratitude to the Editors and Reviewers of all my manuscripts published in the respective Journals and Conference proceedings for their valuable comments on upgrading my research study.

I would like to extend my sincere thanks to Mr. Sanjib Das, Mr. Smitom Swapna Borah and Mr. Rahul Dutta for the unconditional support and respect you all gave me as an elder brother. I am very much indebted to my seniors, Dr. Kunwar Raghavendra Singh, Dr. Gaurav Goel, Dr. Sachin Kumar Tomar and Dr. Mayur Shirish Jain, for their help during my initial phase of this research work. My deepest thanks to

my friends, Ms. Kuwari Mahanta, Dr. Siddhant Dash, Dr. C. Venkatesh Reddy, Mr. Prakash Singh, Mr. Induchoodan TG, Mr. Arun Sathyan, and Mr. Vivekananda Haldar to whom I met during my PhD for their pleasant company, friendship, uncountable suggestions and unconditional help all through the study. I must thank all the cheerful group members of the Waste Management Research Group (WMRG) for their precious support and stimulating discussions. It was my great pleasure to be associated with the members of WMRG.

I am also indebted to my dear sister Ms Mridusmita Goswami for her care and affection. Finally, I must express my profound gratitude to my amazing family, my mother, Mrs. Taru Goswami and my father, Mr. Pranab Ch. Goswami, for providing me with unfailing support, continuous encouragement, unconditional love and blessing throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Above all, I would like to thank the Almighty for giving me the wisdom, health, strength and perseverance to complete this research.

Ankit Pratim Goswami

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 8

LIST OF ABBREVIATIONS

AAS	Atomic Absorption Spectrometer
AHP	Analytical Hierarchy Process
ANN	Artificial Neural Networks
APHA	American Public Health Association
BOD_5	5-day Biochemical Oxygen Demand
CCME	Canadian Council of Ministers of the Environment
CDI	Chronic Daily Intake
CF	Contamination Factor
CPCB	Central Pollution Control Board
CPHEEO	Central Public Health and Environmental Engineering Organisation
CR	Cancer Risk
DA	Discriminant Analysis
DO	Dissolved Oxygen
EC	Electrical Conductivity
EF	Enrichment Factor
E_r	Potential Ecological Risk of Individual Metal
FA	Factor Analysis
FAO	Food and Agriculture Organization
GCF	Global Contamination Factor
GIS	Geographic Information System
GPS	Global Positioning System
HCA	Hierarchical Cluster Analysis
HI	Hazard Index
HQ	Hazard Quotient
I_{geo}	Geoaccumulation Index
I_{POLL}	Pollution Index
HM	Heavy Metals
HRA	Health Risk Assessment
IC	Ion Chromatograph

ICF	Individual Contamination Factor
IS	Indian Standards
LULC	Land-Use-Land-Cover
MF	Mobility Factor
ML	Machine Learning
MLR	Multiple Linear Regression
MRI	Modified Risk Index
MST	Multivariate Statistical Techniques
NSF	National Sanitation Foundation
PCA	Principal Component Analysis
PCBA	Pollution Control Board Assam
PERI	Potential Ecological Risk Index
PLI	Pollution Load Index
PMF	Positive Matrix Factorization
RAC	Risk Assessment Code
RF	Random Forest
TA	Total Alkalinity
TCR	Total Cancer Risk
TDS	Total Dissolved Solids
TH	Total Hardness
Turb	Turbidity
WHO	World Health Organization
WQI	Water Quality Index

ABSTRACT

Rivers provide a significant portion of the water needed for human consumption and industrial processes; therefore, protecting their water quality is crucial. Numerous species of wildlife, as well as humans, depend on rivers for drinking. Rivers are also frequently used for recreation by people worldwide, but contaminants (biological, toxic, organic, and inorganic) pose a major environmental threat to these surface water resources. Water resources worldwide are being monitored and assessed in order to understand the activities causing their degradation. Water quality monitoring systems are now indispensable for providing decision-makers with the information they need to understand, interpret, and implement effective water conservation policies. However, a global issue has emerged in dealing with large amounts of random data generated by these monitoring systems and using them to extract relevant information about water quality. Methods based on environmetrics and artificial intelligence (AI) have shown to be efficient tools for understanding large and complex water quality datasets and assessing their spatiotemporal fluctuations in water quality of surface water bodies.

India is one of the few countries in the world endowed with a good quantity of water resources with a diverse range of climates and geographical settings. Similarly, northeast India, in particular Assam, is endowed with several perennial rivers which are used for drinking and irrigation. But, uncontrolled urbanisation and industrialisation negatively impact the water quality and aquatic fauna by introducing harmful pollutants into the rivers and also increase sediment pollution. Kolong river in Assam has been considered for this study, as this river is among the most polluted rivers in India as per the report of the Central Pollution Control Board, Govt. of India. Kolong River was formerly the only source of potable water, and settlements built up along its banks. Nagaon's portion of the Kolong River is currently degraded, posing a health and hygiene risk to the community. The Kolong River in Assam Nagaon's district illustrates how human activities have exacerbated environmental issues over the previous half-century.

This doctoral thesis attempts to understand the interrelationship between the two critical components of river ecosystem, i.e., its surface water and benthic sediments, as sediments are said to source and sink many contaminants in a river. In order to fully understand the aspects of the river ecosystem, the study has been divided into five objectives to get a better understanding of pollution load in surface water and benthic sediments, their sources and the need for monitoring and treatment. In the first part of **Objective I**, a survey of the probable catchment area of the river was carried out to select the sampling locations from which surface water and benthic sediment samples were collected. Sampling and analysis of surface water were carried out for 21 parameters

at 18 different sites, whereas sampling and analysis of benthic sediments were carried out for 9 parameters, particularly heavy metals, at 9 different sites. The analysis of surface water samples revealed that with respect to pollution load, heavy metals present a higher risk to the consumer of surface water of Kolong river. Following the first part of objective I, the aim of the second part of **Objective I** was to use the analysed surface water samples to evaluate the drinking water quality using fuzzy logic and information entropy. Fuzzy logic based index, i.e., fuzzy water quality index (FWQI) results are more stringent than information entropy based index, i.e., entropy weighted water quality index (EWQI), as it is more sensitive to parameter variation, because FWQI values will be higher even if only one or few parameters have exceeded the permissible drinking water limits.

After evaluation of the surface water pollution load, the aim of **Objective II** was to assess the benthic sediment quality or load in the river using total metal concentration and metal speciation fractions based indices. Total metal concentration based indices include geoaccumulation index (I_{geo}), pollution load index (PLI), enrichment factor (EF), potential ecological risk of individual metal (E^i_r), potential ecological risk index (PERI), and metal speciation fractions based indices, include pollution index (I_{POLL}), mobility factor (MF), individual (ICF) and global contamination factors (GCF), modified ecological risk index (MRI). All the total content indices indicate that compared to the background concentration, the metal pollution has increased in the benthic sediments of the river but does not pose any ecological risk at present. In comparison, the speciation based indices show high to medium risk for most of the river sites due to the presence of higher proportion of anthropogenic fractions of metals. After comparing both types of indices, it was found that the speciation-based index quantifies the risk associated with heavy metal contamination better as the indices developed using this approach gave consistent results and gel well with the variation of the dataset.

After establishing the pollution load in surface water and benthic sediment of the river, **Objective III** identified the latent sources of pollution using environmetrics tools such as cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA) and positive matrix factorisation (PMF) on the water quality variation and sediment quality variation. CA revealed that sites in the middle stretch were polluted due to anthropogenic metal pollution or runoff, and the rest were polluted due to domestic discharge or geogenic metal pollution. DA revealed that only 10 (turbidity, TDS, EC, hardness, alkalinity, Na^+, Ca^{+2}, Mg^{+2}, F^-, SO_4^{2-} and Cd) out of 21 parameters were responsible for the temporal variability of the water quality dataset of the studied area. PCA, Pearson correlation and PMF revealed that pollution in the winter and post-monsoon months attribute to domestic discharges and metal pollution. Moreover, another pollution source gets added in the monsoon and pre-monsoon months due to heavy precipitation in the region, which brings the agricultural or surface runoff quotient of pollution into the river. Furthermore, PMF was able to segregate the geogenic and the anthropogenic metal pollution. Geogenic metal

pollution was attributed to Ca^{+2}, Mg^{+2}, Fe, Mn, and Zn; anthropogenic metal pollution was attributed to Cd and Pb. These findings are also in-line with the benthic sediment quality study.

Till objective III, the assessment was carried out based on data collected during the study period, but to evaluate the effect of prolonged exposure to heavy metals on the human population, health risk assessment was carried out in **Objective IV** using deterministic and probabilistic approaches. Heath risk assessment revealed that the dermal pathway could be a root cause for carcinogenic and non-carcinogenic risk and children being at higher risk than adults. The sensitivity analysis carried out for surface water and benthic sediment samples revealed that Zn and Cd are responsible for non-carcinogenic risk, and Pb is responsible for the carcinogenic risk for the human population residing in the locality of the Kolong River.

Finally, in **Objective V**, linear regression and AI were utilised to develop two regression based predictive models to estimate the heavy metal concentration in the surface water of the river. These models were developed not only to act as an instrument of measurement but also to find the factors affecting its concentration in the river basin. Among the models developed using different techniques, the random forest models had better efficiency in predicting the concentration of heavy metals. These models will help in setting up remediation strategies for pollution abatement in the river.

Therefore, the current study demonstrates the necessity and effectiveness of environmetrics methods and AI for the evaluation and interpretation of large complex datasets in order to obtain more accurate information about the pollution load, which can aid in formulating policies for the protection and revitalisation of rivers worldwide.

Keywords: Surface water; Benthic sediment; Water quality index; Pollution indices; Environmetrics tools; Artificial intelligence; Health risk assessment; Random forest.

Chapter 1

INTRODUCTION

1.1 Overview

The presence of life on earth is largely attributable to the presence of water as one of the primary components. The eight billion people who live on earth consume almost 30 % of the total amount of renewable water that is easily accessible. Despite this, billions of people still do not have access to even the most basic water resources. India is one of the few countries in the world endowed with a good quantity of water resources with a diverse range of climates and geographical settings. The country receives roughly 4000 billion cubic metres (BCM) of precipitation annually, including snowfall. It is estimated that various river basins contain a total annual average water resource capacity of 1869 BCM, of which only 1086 BCM is usable. This usable volume comprises of 690 BCM surface water and 396 BCM groundwater (Jena et al., 2022).

Rivers supply water for various uses, including drinking, municipal, industrial, irrigation, hydroelectric power generation, waterway transport, and so on, making rivers an essential component in the hydrological cycle. Like many other natural systems, rivers are extremely complicated systems that are affected by a wide variety of factors. In order to have a complete grasp of the exogamic cycles of elements, it is necessary to have knowledge of the chemistry of the river ecosystem. River systems function as unified systems, any modifications to one section of a river system affect the other sections. A river is considered polluted when its water is unfit for its intended use. Domestic, industrial, and agricultural effluents are the three primary sources of wastewater that substantially impact the river system. During times of high precipitation, silt and other types of suspended solids, such as dirt, can be washed into rivers from agricultural fields that have been tilled, construction and logging sites, urban areas, and riverbanks that have been eroded, which also lead to pollution of the river. As a result of the sediments' introduction into various bodies of water, fish are unable to breathe properly, plant productivity and water depth are diminished, and aquatic species and the ecosystems in which they live become smothered. Also, due to biological activity, interactions between sediment and water, and shifts in chemical equilibrium of the sediment, pollutants can be discharged directly into the water from the sediments. It is essential to do concurrent research on both the water and the sediments of a river to achieve a holistic comprehension of the contamination or pollution load of the river (de Almeida et al., 2022).

1.2 Background Research

The quality of freshwater sources is impacted by increasing point and non-point sources of pollution. Discharges from these sources, whether treated or untreated, will always find their way into streams and pollute the streams. The discharges vary depending on the day, month or season, which is why there is a vast variation in pollution levels in different seasons. During winter, many large river basins dry up, leaving just a tiny amount of water to dilute their sewage, thus increasing river contamination. Changes in river pollution necessitate continuous monitoring of surface water bodies. The difficulty with continuous monitoring of surface-water bodies is that it produces enormous and complex data sets that are difficult to understand and interpret (Dixon and Chiswell, 1996; Iscen et al., 2008; Singh et al., 2020a). A long list of water quality parameters to explain surface water potability is neither logical nor entirely easily understandable (Bellman and Zadeh, 1970). The use of multivariate statistical techniques (MSTs) and water quality indices (WQIs) aids in dealing with the massive amount of data generated by continuous water resources monitoring. WQIs are an efficient way of explaining a source's water quality and have been used to characterise the acceptability of water sources for human consumption. WQIs can express large datasets in a condensed logical form. WQIs combine information from various sources to determine the overall state of a water system. They assist policymakers and promote public awareness about the issue of water quality. These strategies aid in detecting potential sources of water contamination and are an important tool in adequately managing water resources (Reghunath et al., 2002; Simeonov et al., 2003).

Initially, Horton in 1965, developed the first WQI in the world by choosing 10 common water quality parameters (Singh et al., 2019). In addition, in 1970, the Brown group also established a new WQI similar to Horton's index based on the weighting allocated to an individual parameter (Brown et al., 1970). Various scientists and experts have recently considered many modifications to the WQI principle (Dwivedi et al., 1997). The Delphi method, used for the last few decades, has led to weight aberration due to subjective factors that calculate or assign weight (Singh et al., 2019). Furthermore, these methods of weight distribution take time (Abbasi and Abbasi, 2012). Within these parameters, the calculation of fixed weights based on the parameter's intrinsic knowledge can remove subjective judgments (Singh et al., 2019). Information entropy can be used to characterise this useful information through entropy weights. Entropy-based weights have arisen as an effective methodology that utilises information entropy to apply weighting to parameters (Amiri et al., 2014). Information entropy addresses a random process to detect chaos or confusion (Cui et al., 2018). Water quality parameters are random variables, and their probability distribution impacts the importance and variance of the WQI (Landwehr, 1979). The weight allocation to a specific parameter depends on the uncertainty of its occurrence in a place. Lower weights are assigned or given to a parameter having higher uncertainty of occurrence at a site (Amiri et al., 2014). Entropy weighted water quality indices (EWQIs) are an upgrade over

traditional WQIs that often assign weightings to parameters focused on professional opinion and personal judgement. Such decisions and opinions may contribute to a lack of useful knowledge about the water quality in the region's rivers and lakes (Delgado and Romero, 2016; Fagbote et al., 2014).

Additionally, computational methods in the field of artificial intelligence (AI), such as neural networks, fuzzy logic, and genetic algorithms, have been increasingly applied to environmental issues (Chau, 2006) due to the widespread belief that they are effective in addressing the aforementioned issues. Zadeh (1965) introduced the topic of fuzzy logic, which is now one of the most widely used methods in AI. It is deemed suitable for developing environmental indices because it can incorporate human views and expertise into them, allowing it to deal with non-linear, uncertain, ambiguous, and subjective data. This allows us to incorporate qualitative and quantitative elements into the index, expanding its usefulness. Also, the conclusions of an assessment can be reported accurately and in a language that the general public, managers, decision-makers, and other non-specialists can understand (McKone and Deshpande, 2005; Silvert, 2000). As a result, the application of fuzzy logic to environmental indices has received significant attention, particularly in the context of water quality analysis (Chang et al., 2001; Gharibi et al., 2012a; Icaga, 2007; Lermontov et al., 2009; Liou et al., 2004; Mirabbasi et al., 2008; Ocampo-Duque et al., 2006).

In order to effectively prevent and manage surface water contamination, accurate water quality data must be gathered through regular monitoring (Singh et al., 2004; Varol et al., 2012). Managing the massive and complex data sets produced by monitoring stations due to the numerous water quality variables is a significant challenge (Dixon and Chiswell, 1996). Researchers studying surface water quality face a difficult challenge while attempting to interpret the data they collect. The observable data can be understood better with the use of the Environmetrics techniques, which enables quantitative environmental analysis. Principal component analysis (PCA), discriminant analysis (DA) and hierarchical cluster analysis (HCA) are three statistical methods that have seen widespread application in recent years for analysing large datasets of water quality measurements and identifying the sources of pollution (Batayneh et al., 2012; Bouguerne et al., 2017; Fan et al., 2010; Noori et al., 2010; Shrestha and Kazama, 2006). These tools are useful for manipulating, interpreting, and representing information about pollutants in surface water, as well as its geochemistry and sources of pollutants. But these statistical methods cannot be solely relied on to isolate the impact of every possible source of pollution. Receptor models, such as the positive matrix factorisation (PMF) model, are also utilised for this purpose. The PMF model was initially applied to an air pollution dataset to determine the contribution of various pollution sources. They have recently been applied to water quality datasets, together with HCA and PCA, to estimate the contributions of pollution sources (Gholizadeh et al., 2016; Zhang et al., 2020).

Similar to water contamination, the pollution of benthic sediments due to various anthropogenic activities has severely threatened the aquatic environment. As a result of excessive effluent

discharge (mostly from domestic, agricultural, and industrial wastewater), many heavy metals accumulate in the aquatic ecosystem, posing a severe threat to global sediment flux (Armagan et al., 2008; Azhar et al., 2015; Yao et al., 2021). These heavy metals are not only persistent, but they are also harmful when their concentrations exceed permissible limits since they do not degrade or decompose over time. In addition, these compounds are less mobile in water than other chemicals (Tchounwou et al., 2012). As a result, they continue to accumulate in natural aquatic systems, leading them to adhere to the top of the sediment column of the water body. As a result, sediment columns of water bodies are potential sources of heavy metals, which can be released into the water column or aquatic flora and fauna by natural or anthropogenic processes, and ultimately participate in the food chain (Yao et al., 2021; Yin et al., 2011). Heavy metals can be found in sediments in several chemical fractions, each related to their mobility and bioavailability (Liang et al., 2017; M. Wang et al., 2019). Since total metals do not fully reflect geochemical processes, chemical speciation, and metal bioavailability, sequential extraction is useful for investigating these topics (Hing et al., 2020; Sundaray et al., 2011; Zhang et al., 2017). It is worth noting that exchangeable metals are volatile and easily absorbed by humans and aquatic species (Liang et al., 2017). Heavy metals generated from anthropogenic activities are commonly found in an unstable state with greater mobility (Zhang et al., 2017). On the other hand, those that are lithogenic in origin are frequently associated with sediments that have taken on stable forms (Xiao et al., 2015). Therefore, the overall concentration of heavy metals in sediments is insufficient to determine the amount of sediment pollution. Its chemical form in sediments must be identified to determine how a metal will behave in the environment. In addition, the distribution of metals in different phases and their origins can be utilised to measure pollution levels (anthropogenic or natural)(Berrow, 1986; Dhanakumar et al., 2013; Kersten and Forstner, 1986). As a result, determining the heavy metal chemical fractions is crucial for source identification and risk assessment, particularly in areas with high levels of human activity.

In contrast to typical pollutants like nitrogen and phosphorus, toxic metals are extremely dangerous even in low quantities, such as a few micrograms per millilitre (µg/mL). As a result, residents of polluted areas, children and adults, are exposed to hazardous compounds via various sources and pathways, resulting in a wide range of toxic exposures for them (Chen et al., 2019; Wang et al., 2017; Xu et al., 2020). Chromium (Cr) and lead (Pb) are two heavy metals that have been identified as human carcinogens (Lyon, 1994). On the other hand, systemic non-cancerous metals (e.g. Cd, Co, Cu, and Hg, etc.) can also cause negative health consequences even at minute concentrations (Djadé et al., 2021). Groundwater, surface water, sediments, and soils can all be sources of exposure to toxic metals for people living in and around polluted areas (Adimalla et al., 2020; Gitter et al., 2020; Saha et al., 2017; Yang et al., 2019). Environmental pollution has been linked to several illnesses, including an increased risk of cancer in populations exposed to it and even deaths from poisoning (Cao et al., 2010; Dooyema et al., 2012). Studies in the clinical and

epidemiological fields are the most trustworthy tools for control and intervention in the public health system (ACHHRA, 2017; USEPA, 2005). However, the high monetary cost of such studies means that these studies are not always feasible for evaluators. Risk assessment is one easy-to-use tool that can provide a quantitative estimate of health problems caused by pollution. In other words, human exposure to harmful pollutants can be measured with the help of health risk assessment (HRA) (Huang et al., 2018; Singh et al., 2019). HRA can be calculated using deterministic or probabilistic methods. The output of health risk is represented as a single point value by the deterministic method. In probabilistic risk assessment (PRA), the output risk is a range of values resulting from the combination of the probability distributions of numerous input factors in the risk equation (Rajasekhar et al., 2020; Saha and Rahman, 2020). PRA is most useful when the deterministic risk outcome is close to the safe exposure limits (Barrio-Parra et al., 2019). Furthermore, probabilistic analysis, in combination with sensitivity analysis, may be used to investigate the impact of input parameter uncertainty and variability on risk calculation outcomes. HRA generally supplies information that can contribute to decision-making by quantitatively estimating risk. In addition, it can facilitate the allocation of resources to reduce exposure to environmental threats (Harris et al., 2017).

In times when heavy metals have become a prime and critical source of pollution in surface water and groundwater, the estimation of heavy metals is a cause of concern. Accurate measurement of heavy metals requires instruments with sufficient accuracy like atomic absorption spectrometer (AAS), inductively coupled plasma mass spectrometer (ICPMS) etc. The availability of such instruments at every research location or measuring agencies is difficult due to lack of funding or other reasons. In this regard, multivariate statistical techniques (MSTs) and AI can help develop models that estimate heavy metal concentrations with sufficient accuracy. MSTs and AI have the ability to model and predict non-linear and complex mathematical relationships between dependent and independent variables (Dehghanian et al., 2016; Kardam et al., 2012). The biological neurons process inspired the ANN concept, and the elements were connected similarly to the human brain and nerves, allowing it to learn the mapping structure from input data to output data without requiring explicit mathematical relationships (Uzun et al., 2017). Machine Learning (ML) is a subset of AI that enables software platforms or packages to become more accurate at predicting outcomes without being specifically designed to do so. ML algorithms estimate new output values by using historical data as input. Recently, a new data-mining AI algorithm has been developed to overcome regression problems and lessen the drawbacks of AI. New algorithms like random forest (RF) are being used to study issues in climatology, hydraulics, and hydrology. These algorithms can simulate solar radiation (Sharafati et al., 2019), estimate reference evaporation (Khosravi et al., 2020), measure suspended sediment transport (Khosravi et al., 2018), and predict shear stress in rivers (Khozani et al., 2016). The RF is a relatively new ensemble ML technique built on decision trees that make predictions by voting for categorisation and averaging regression

(Auret and Aldrich, 2012). Better modelling performance than ANN and MLR is made possible by the decision tree-based algorithms' lack of hidden layers and modelling transparency (Kisi et al., 2012). The proposed models could significantly minimise human and material resources in future experiments if they are accurate for new experimental data.

Therefore, extensive field research was carried out, considering the aforementioned issues related to surface water bodies and the need for continuous monitoring of these bodies. In order to accomplish this study's overarching goal, a variety of specific objectives were developed.

1.3 Research Objectives

In the present day, the need for potable drinking water has led humans to search for every possible source of water present on the earth. Rivers, lakes and groundwater are the primary drinking water sources available among the vast majority of water resources on the earth's surface. Increasing population and industrialisation have led to the overutilisation of groundwater. Due to this, many governments worldwide had to make strict policies on groundwater use and advocate using surface water such as rivers. Numerous rivers are present all over the world, which are perennial. Using rivers directly as a potable water source is far from the existing scenario. Improper drainage, disposal of solid and liquid waste directly into the river without any treatment, and numerous other reasons have deteriorated river water quality. Before utilising the river water for various purposes, proper monitoring and treatment of the water is required. The objective of this research work is based on the knowledge gap found from the literature review carried out in chapter 2 and also to provide possible sources of setup treatment, which are as follows–

- **Objective I:** Characterisation of the surface water of the river using 21 physicochemical parameters, including heavy metals. Also, to explore and develop water quality indices (WQIs) for the river basin based on the studied water quality parameters and compare the WQI with the existing water quality standard.
- **Objective II:** Assessment of sediment quality with respect to heavy metals. This involves the assessment of total metal concentration and speciation fractions of heavy metals to understand their contribution to water pollution load using the following indices:
  - Total metal concentration: Pollution load index (PLI), geoaccumulation index (I_{geo}), enrichment factor (EF), potential ecological risk index (PERI) and sediment quality guidelines (SQGs)
  - Speciation fractions of heavy metals: Pollution index (I_{POLL}), mobility factor (MF), individual contamination factor (ICF), global contamination factor (GCF), risk assessment code (RAC) and modified risk index (MRI).
- **Objective III:** Identification of latent sources of pollution in water and benthic sediment samples in the study area using discriminant analysis (DA), principal component analysis, positive matrix factorisation (PMF) and hierarchical cluster analysis (HCA).

- **Objective IV:** Assessment of health risk due to prolonged exposure to heavy metals through incidental ingestion and dermal contact using deterministic and probabilistic approaches.
- **Objective V:** Development of two regression models explaining heavy metal contamination in the water column of a river using the (i) physicochemical parameters and (ii) metal speciation fractions in the study area (MLR, ANN and RF).

1.4 Need of the Study

Assam is one of the eight states in the North-eastern region of India, having a diverse population, sociocultural background, and ethnic origins. Over 80% of the population resides in rural areas. The majority of the state's labour force (63%) is employed in agriculture and related fields. Assam is home to the world's largest tea-growing region. As the state is endowed with several perennial rivers, they are the primary source of irrigation. The rate of water use in agriculture has increased dramatically during the past few decades. The uncontrolled urbanisation and industrialisation negatively impact the water quality and aquatic fauna by introducing toxins into the rivers and also increase sediment pollution. The discharge of urban garbage, untreated effluents from various enterprises, and agrochemicals in certain rivers has reached a concerning level. So, there is an immediate need to monitor water resources and understand the various ways in which pollutants enter the water system. This is only conceivable if we can develop a thorough understanding of the surface water ecosystem through rigorous monitoring programmes, paving the way for innovative approaches for tackling critical concerns such as complex data processing. Furthermore, ways for measuring water quality using environmetrics tools such as multivariate statistics, information entropy and fuzzy logic will benefit the scientific community in a more comprehensive and time-efficient manner. This study will also help to understand the current contamination levels in the sediment column, allowing the various agencies involved in surface water protection and management to take the necessary actions for planning proper resource management. Furthermore, given the comprehensive monitoring and assessment carried out in this study, the findings will provide a comprehensive understanding of the heavy metals in an aquatic ecosystem.

1.5 Scope of the Research

The first and most important effort in achieving the broad objectives of the current research was collecting, transporting, and storing water and sediment samples continually. As a result, the primary scope of the research was the continuous sampling of various aquatic ecosystem components. Following that, laboratory analyses for various parameters were performed. From sample collection to analysis, the entire procedure demanded a thorough understanding of standard protocols, which remains one of the essential focuses of this research. Because the entire thesis is primarily concerned with substantial dataset processing and analysis, many computational activities were carried out, including the use of programmed software such as Microsoft Excel, SPSS

(v. 25), and EPA-PMF (v. 5.0). The study site's features were plotted using ArcGIS-ArcMap (v. 10.2). In addition to the computational software, statistical tools like MINITAB (v.20), MATLAB (v. 2022a), and Jupyter Notebook (v. 6.5.4) for python codes were used to develop the heavy metal regression model in MLR and AI.

1.6 Innovations Made in the Study

- A fuzzy logic based WQI is developed to express the water quality of rivers.
- Developed a Random Forest (machine learning) model to estimate the heavy metal concentrations in rivers.
- Developed another Random Forest (machine learning) model to understand the contribution of heavy metal pollution from benthic sediments to river water.

1.7 Thesis Organisation

The entire thesis has been divided into the following chapters:

- Chapter 1 introduces the importance of surface water resources, their pollution and ways to monitor and assess the pollution levels effectively. This chapter also discusses the objectives, needs, and scope of the study.
- Chapter 2 gives a detailed literature review of water quality and water quality index (WQI), development of WQIs, application of different environmetrics tools, information entropy and fuzzy logic in surface water quality monitoring and assessment, and finally, heavy metals prediction using regression models.
- Chapter 3 is associated with the flow chart of the various phases of the research, study area and sampling strategy. This chapter also discusses the methods and methodologies adopted for the present study.
- Chapter 4 is associated with the discussion of the observed water quality datasets and the application of information entropy and fuzzy logic in surface water quality assessment.
- Chapter 5 details benthic sediment quality and assesses sediment contamination through different approaches concerning heavy metals.
- Chapter 6 deals with the application of different environmetrics tools such as CA, DA, PCA, and PMF for the identification of various factors responsible for variability in water quality and benthic sediment quality.
- Chapter 7 deals with human health risk assessment concerning heavy metals.
- Chapter 8 is associated with the development of regression models for the prediction of heavy metals.
- Chapter 9 is associated with the overall conclusions of the present study and recommendations for future work.

Chapter 2

LITERATURE REVIEW

Although the literature covers a wide range of topics, this review will concentrate on those pertinent to this study. The chapter highlights previous research work that is relevant to the prospective study.

2.1 Water Quality and Water Quality Index (WQI)

Surface water pollution is a serious issue worldwide and is related to a wide range of biological and chemical contaminants (Al-Janabi et al., 2019). Surface water ecosystem pressures are classified as (i) hydromorphological pressures; (ii) non-point sources of pollution such as agricultural runoff and atmospheric precipitation; and (iii) point sources of pollution: particularly industry and sewage treatment plants (Dedić et al., 2020). The basic reason for investigating hydromorphological stresses is to understand better the effects induced by natural processes and diverse human activities (Wątor and Zdechlik, 2021). The hydromorphological characteristics of the aquatic system are critical for producing a safe environment. The evaluation of hydromorphological features is strongly related to the evaluation of physical, chemical and biological datasets for determining the ecological conditions of water (Poikane et al., 2020). To reduce the possible threats to public health, it is critical to identify pollution sources and design an effective management strategy (Vadde et al., 2018). The amount and type of contaminants contained in water considerably influence its fitness for any application. Physical, chemical, and biological pollutants in water are identified and expressed using the contamination characterisation (Soumaila et al., 2019). To establish the extent and status of pollution in any river, the water quality must be regularly monitored (Hussein et al., 2019). However, explaining each water quality measure to non-specialists and the general public is challenging and illogical. WQI has proven to provide information to citizens and legislators in an understandable and interpretable way (Stanly et al., 2021). The fundamental purpose of WQI is to offer a single value for water quality by converting data on the components and concentrations in a sample into a numerical score. WQIs allow for observation of changes in water over time or comparisons between different bodies of water based on the single unitless score. WQI is among the most effective methods for characterising the status and estimating the suitability of water for various purposes. WQI utilises data on water quality to help environmental authorities change their environmental policy (Nowicki et al., 2020).

2.1.1 Common steps for the development of any WQI

The four steps outlined below are used to develop WQI:

1. Selection of parameters.
2. Converting the parameters to a common scale.
3. Weights are assigned to all parameters.
4. Then, the sub-indices are aggregated into an index score (Ewaid et al., 2020; Vadde et al., 2018).

2.1.2 Justification for the use of WQI

WQIs are intended to be used first and foremost to analyse water quality monitoring data, allowing for extensive interpretation of outcomes, particularly when pollutant concentrations are below water quality standards (Uddin et al., 2021). WQIs assist professionals in interpreting monitoring data from a broader perspective, allowing administrative decision-makers to assess the performance of regulatory programmes and convey water quality information to the public in an understandable and easy-to-understand manner (Sutadian et al., 2016). Indices are used in monitoring activities, including water quality evaluation, treatment, and utilisation, public communication, R&D projects, and environmental planning (Abbasi and Abbasi, 2012).

2.1.3 History of WQI concept

In 1965, a panel of the President of the United States Science Consultative Committee on Environment proposed the establishment of a scientific technique for establishing a numerical index for identifying chemical water pollution (Eberhardt, 1967). According to the committee, the approach must be sensitive to various chemical contaminants. Its outcome is nearly proportionate to the negative consequences of water pollution on humans or aquatic life. This statement prompted Horton to release the first water quality index (WQI) the same year (Horton, 1965). Since that day, WQIs have become a popular and useful technique for measuring the water quality of water bodies across the world (Kulkarni, 2020). Following Horton, Brown et al., (1970) created a WQI with a structure comparable to Horton's index (Brown et al., 1970). The study of Brown et al. (1970) was sponsored by the National Sanitation Foundation (NSF), despite the rigidity with which parameters were selected. Because of this, Brown's index is also known as the National Sanitation Foundation Water Quality Index (NSFWQI).

Prati et al. (1971) introduced an index using a water quality categorization system. The degree of contamination is expressed numerically in the form of the index value. It takes into account all of the contaminants present at the same time. However, unlike the Horton index and the NSFWQI, the PRATI index, named after the person who developed it, measures each pollutant separately and has an increasing scale with pollution. The index scale went from 0 to 15 (or more) for high quality (or no pollution).

Dinius (1972) as part of the social counting method, assessed the water pollution control expenses by developing a WQI. In this WQI, sub-index functions are calculated using simple equations. The index score is then computed by adding the weighted sub-indices.

Stoner (1978) created an index by combining two large water-use categories: public supply and irrigation. The index can handle the two water uses by applying the weightings and sub-index functions to the final index aggregation. The Stoner index may also be used for different types of water.

Dinius (1987) suggested a multiplicate index to combine the pollutant into a single system. The Delphi approach involves a panel of seven water quality specialists, and the resulting index may be used to determine the level of freshwater pollution.

Smith (1990) focused on four types of water usage: (i) general, (ii) public bathing, (iii) water supply, and (iv) fish spawning. The Delphi approach was used to construct the index, which comprises selecting parameters for each water type, formulating sub-indices, and weights assignment to the selected parameters.

CCME (2017) presented a mathematical framework for analysing ambient water quality conditions in relation to water quality targets established by the Canadian Council of Ministers of Environment (CCMEWQI) in 2001 and revised in 2017. The index is made up of three essential components: (i) scope denotes the number of parameters not meeting the water quality objectives; (ii) frequency denotes the number of times the objectives were not met; and (iii) amplitude denotes the amount by which the objectives were not fulfilled. The index's output spans from 0 to 100, with 0 representing the poorest water quality and 100 representing the best.

Many scientists and specialists have recently considered numerous alterations to the WQI approach (Abtahi et al., 2015; Beamonte et al., 2005; Bhargava et al., 1990; Boyacioglu, 2007; Dwivedi et al., 1997; Ewaid and Abed, 2017; Landwehr et al., 1974; Liou et al., 2004; Medeiros et al., 2017; Nikoo et al., 2011; Sutadian et al., 2018, 2016; Thi Minh Hanh et al., 2011; Yaseen et al., 2018). These indices assign weightage to each water quality parameter before evaluating WQI. Despite their simplicity, these types WQIs cannot be relied on due to the inclusion of limited variables and crisp weight assignment (Dahiya et al., 2007; Lermontov et al., 2009). Applying crisp weights to the parameters in the WQI renders the evaluation of water quality conditions for changing seasons, temperatures, seasonal rainfalls, and surface runoffs impractical (Dahiya et al., 2007). The subjectivity of the WQI weights reduces the accuracy of its water quality evaluations in addition to ignoring toxic pollutants (Chang et al., 2001; Hou et al., 2016; Ocampo-Duque et al., 2006). Therefore, an advanced strategy is necessary to address the shortcomings of the WQI approach (Gharibi et al., 2012b).

Fuzzy logic and information entropy based indices can integrate significant parameters of water quality and classify random environmental factors into understandable and transparent subsets for the public and experts, and provide a much-appreciated formalism that accounts for the

inherent uncertainty of linear and non-linear information (Gharibi et al., 2012b; Mahapatra et al., 2011; Vadiati et al., 2016). Their ability to integrate quantitative and qualitative data offers a significant edge over other indices, as the portrayal of ecological complexity closely resembles reality (Mahmoudi et al., 2018). Information entropy is concerned with recognising uncertainty or disorder in a random process (Singh, 2013). Many random hydrological and meteorological processes in the environment bring into the limelight the effectiveness of information theory in their applications (Cheng et al., 2004; Fleming, 2007; Mishra et al., 2009; Singh, 2015, 2013). The assignment of weights to a specific parameter at a certain location depends on the occurrence uncertainty of that parameter. The greater occurrence uncertainty at a place results in lower parameter weights at that site (Amiri et al., 2014). In contrast, fuzzy models permit nonlinear relations of the input variables and link them to the output using if-then rules (Lermontov et al., 2009). They transform verbal concepts defining expert judgement into a mathematical framework based on fuzzy set theory (Marchini et al., 2009). Therefore, information entropy and fuzzy logic based indices are superior than conventional WQIs [based on the Delphi method, analytical hierarchy process (AHP), and expert survey technique], which assign weights to elements based on subjective evaluations and expert opinion (Amiri et al., 2014; Fagbote et al., 2014). Such judgements and opinions frequently result in the loss of important water quality data. Moreover, such procedures are arduous and time-consuming (Abbasi and Abbasi, 2012).

2.1.4 Current literature on WQIs

A qualitative and systematic approach was used to identify the current literature on WQIs. The procedures below were performed to retrieve the needed literature. First, a comprehensive literature review for research and review papers is conducted in the Scopus database using different keywords. After eliminating duplicates and doing an initial screening, documents are selected. The following keywords were entered: **TITLE-ABS-KEY** (("water quality index" OR "wqi") AND ("fuzzy logic" OR "information entropy") AND ("surface water" OR "river")) and a total of 14 relevant articles were found after a thorough screening out 49 articles.

Ocampo-Duque et al. (2006) investigated a water quality assessment approach based on fuzzy inference systems (FIS). A multi-attribute decision-aiding strategy was used to deal with the relative value of water quality indicators included in the fuzzy inference process. A case study in the Ebro River in Spain was used to assess the fuzzy index's possible use. The findings correspond with government reports and expert judgments about the pollution problems in the studied area. As a result, this technique emerges as an appropriate and alternative technique for developing effective water management strategies.

Lermontov et al. (2009) investigated the application of AI by developing the fuzzy water quality index (FWQI), which is based on fuzzy logic. The effectiveness of the proposed index is evaluated by comparing it to numerous water quality indices (WQIs) presented in the literature, using data from hydrographic surveys of the Ribeira de Iguape River in the southern portion of

the state of So Paulo, Brazil, from 2004 to 2006. The index was relatively similar to the other indices and correlated well with the WQI, normally constructed in Brazil. This new index might be utilised as a decision-making tool in environmental management.

Semirom et al. (2011) investigated the development of a new indicator known as the Fuzzy water quality index. It gives a simplified depiction of the numerous and complicated parameters (physical, biological, and chemical) that affect the overall quality of drinkable surface water. Six water quality parameters were utilised to develop the FWQI to assess the quality at a sampling station on the Karoon River in Iran, which includes DO, turbidity, pH, TDS, nitrate, and faecal coliform. The new index is expected to help decision-makers to report on the river water quality and investigate spatial and temporal variations. The authors believe that if applied logically, fuzzy logic might be a valuable tool for some environmental policy challenges.

Gharibi et al. (2012a) sought to create a unique water quality index based on a fuzzy logic technique for routine evaluation of surface water quality, specifically for human consumption. Twenty indicators were chosen for their vital relevance to overall water quality and their influence on human health. To evaluate the suggested index's performance under actual settings, a case study was done at Mamloo dam in Iran, using water quality data from four sampling sites from 2006 to 2009. The findings of this study suggested that the overall water quality in the sample sites was fairly poor (in the range of 45–55). Furthermore, a comparison of the suggested fuzzy-based index's outputs with those of the NSF water quality index (NSFWQI) and Canadian Water Quality Index (CWQI) revealed similar results but was sensitive to changes in water quality parameter levels. However, the current study's suggested index yielded more stringent outcomes than the NSFWQI and CWQI. The sensitivity study findings indicated that the index is resistant to rule modifications. In conclusion, the suggested index appears to provide accurate and dependable results and may thus be utilised as an instrument for monitoring water quality, particularly for drinking water analysis.

Singh et al. (2019) developed a fuzzy-based water quality evaluation method to analyse the overall state of water quality of a river. A case study of the Yamuna River was conducted to assess the usefulness of the fuzzy comprehensive water quality index (FCWQI). The FCWQI was determined exclusively for the usage of water for drinking purposes in this study, but this methodology may be extended to other applications as well. The FCWQI created here is based on an integrated system that allows a single number to indicate the overall water quality status.

Singh et al. (2019b) used shannon entropy (information entropy) to assess water quality based on end-use. Entropy weights were used to create an entropy-weighted water quality index (EWQI) for drinking water. The suggested method represented the Beki river's water quality as "excellent" (EWQI < 50) to "poor" (150 < EWQI < 200).

Chanapathi and Thatikonda (2019) created a Mamdani-type fuzzy-based regional water quality index (FRWQI) that includes dissolved oxygen (DO), faecal coliforms (FC), biological

oxygen demand (BOD), pH, nitrogen, suspended solids (SS), alkalinity, turbidity, chemical oxygen demand (COD), and electrical conductivity (EC). It may be used to assess the water quality index of various river basins worldwide. Despite the diverse geographic origins, the FRWQI is comparable to water quality models used in India, Malaysia, and the United States, and it can aid in the self-assessment of regional water quality on a worldwide scale.

Nayak et al. (2020) used fuzzy modelling to make predictions on the water quality of Indian rivers. The triangular and trapezoidal membership functions for fuzzification, along with the centroid, bisector, and mean of maxima (MOM) approaches for defuzzification, have been utilised in developing fuzzy models. It has been observed that the fuzzy model with a triangular membership function that utilises the bisector defuzzification method performs better than the triangular and trapezoidal membership functions that utilise the centroid and MOM methods of defuzzification. The values of the fuzzy logic based water quality index have been compared with those of the NSFWQI and Vietnam's official water quality index (VWQI). The values of the fuzzy logic based water quality index have been observed to be more indicative of the real river water quality state of Indian rivers compared to the NSFWQI and VWQI. This is because the fuzzy logic approach utilised is equally sensitive to all parameters and can precisely reflect even minute shifts in the value of any parameter, and is especially true for sections of contaminated rivers.

Hue & Thanh (2020) proposed a method based on fuzzy inference systems (FIS) to analyse water quality in the Nhue River, an inter-provincial irrigation network in Northern Vietnam. For the quality evaluation, nine fuzzy variables were used: dissolved oxygen (DO), biological oxygen demand (BOD_5), chemical oxygen demand (COD), pH, ammonium (NH_4^+), phosphates (PO_4^{3-}), total suspended solids (TSS), turbidity, and total coliform (Tco). The fuzzy inference system comprises 3250 rules with scores ranging from 0 to 100 and five verbal categories. By comparing it to Vietnam's official water quality index VWQI, the FWQI was confirmed. The study's findings demonstrate that the Fuzzy Inference System (FIS) technique was a realistic, straightforward, and helpful instrument for assessing surface water quality in rivers and irrigation networks.

Golshan et al. (2020) used the Mamdani fuzzy technique to assess the water quality of the Karun River and created a fuzzy water quality index (FWQI). A hierarchical method was used to overcome the curse of dimensionality in fuzzy rules, considering the correlations between 21 factors in the form of physicochemical, chemical, biological, and heavy metals. Fuzzy groups were processed using trapezoidal membership functions and procedures for evaluating parameter relationships. Based on data from 2010 to 2015, the FWQI of 36.78 ± 2.17 was estimated, whereas the NSFWQI and CCMEWQI indices obtained 41.8 ± 1.33 and 33.65 ± 2.87, respectively. According to the ANOVA test, changes in FWQI were not significant at sample sites. Furthermore, the data demonstrate a decline in water quality from upstream to downstream of the Karun watershed. Based on these findings, it is apparent that the water quality in the Karun River must be addressed to benefit the local ecology.

Siddique et al. (2021) used the entropy water quality index (EWQI) and irrigation water quality index (IWQI) to assess the surface water quality of the Dhaleshwari River in Bangladesh. During the dry and wet seasons, fifty surface water samples were collected from five sampling stations and analysed for sixteen parameters. The resultant IWQI followed a similar trend to the EWQI, indicating poor to good water quality. The entropy hypothesis identified $Mg^{2+,}$ Cr, TDS, and Cl^- as the primary pollutants impacting during the dry season and Cd, Cr, Cl, and SO_4^{2-} as the primary pollutants during the wet season affecting irrigation water quality. Overall, both IWQI and EWQI will aid in regulating water quality in a cost-effective and unbiased way.

Singh et al. (2021a) measured 20 physicochemical parameters in the Puthimari and Baralia Rivers in Assam between May' 2016 and June' 2017. The information entropy was used to assess the water quality of the river in terms of the entropy weighted water quality index (EWQI). EWQI ranged from 61.62 to 314.68. The study demonstrated the importance of information entropy in assessing surface water quality.

Pandey et al. (2022) suggested a new water quality index based on fuzzy logic to assess freshwater and saltwater intrusion in the Mandovi River. The parameters used to create this index were chloride, BOD, fluoride, nitrate, nitrite, and TDS. The index results not only categorize water quality but also offer a comparative assessment of water quality. The findings ensure that borderline scenarios are handled more user-friendly, delivering a more thorough status on water quality than in an earlier study.

2.2 Benthic Sediment Pollution Load

Sand, clay, and silt are all examples of sediment, which are the soil particles that eventually sink to the bottom of a body of water. Erosion of soil and the decay of plants and animals are two sources of sediment. These particles are carried to bodies of water by wind, water, and ice. Sediment is the most prevalent pollutant in aqueous bodies, according to the Environmental Protection Agency. Nearly 30% of all sediment is produced by natural erosion, whereas 70% is produced by accelerated erosion due to human usage of land (Walling, 2009). The following are some of the ways in which sediment degrades water quality and creates problems for humans and wildlife (i) sediment clogs storm drains and catch basins, which increases the risk of flooding; (ii) sediment-polluted water becomes cloudy, preventing animals from seeing food; (iii) sediment-polluted water inhibits natural vegetation growth in water; (iv) sediment in stream beds disrupts the natural food chain by destroying the habitat where the smallest stream organisms live and causing massive declines in fish populations; (v) sediment contributes to the increased price of purification of water because unpleasant odours and flavours; and (vi) accumulation of sediment in rivers can change the flow of water and diminish the depth, making it more challenging for both navigation and recreational use.

The contamination of benthic sediments by diverse anthropogenic activities has become a serious issue in recent years. Many heavy metals have accumulated in the aquatic ecosystem as a

result of excessive effluent discharge (mainly from domestic, agricultural, and industrial effluents), posing a major threat to global sediment influx (Armagan et al., 2008; Azhar et al., 2015; Yao et al., 2021). Because they do not degrade or decompose over time, these heavy metals are persistent and harmful when their quantities exceed permissible limits. Furthermore, these substances are less mobile in water than other chemicals. As a result, they continue to accumulate in natural aquatic systems, causing them to adhere to the top of the sediment column. Thus, sediment columns are potential sources of heavy metals, which can be released into the water column or aquatic flora and fauna by natural or anthropogenic processes and contribute to the food chain (Yao et al., 2021; Yin et al., 2011). Heavy metals can be found in sediments in various chemical fractions, each connected to mobility and bioavailability. Thus, measuring the total concentration of heavy metals is not sufficient in understanding the behaviour of metals in the aquatic environment because total metals do not fully reflect geochemical processes, chemical speciation, and metal bioavailability. To evaluate or measure the different fractions or species of heavy metals in the sediments, a sequential extraction technique can be used for the purpose (Liang et al., 2017; Wang et al., 2019; Hing et al., 2020; Sundaray et al., 2011; Zhang et al., 2017). Anthropogenic sources of heavy metals tend to be in a more mobile, unstable form (Liang et al., 2017; Zhang et al., 2017). On the other hand, those of lithogenic origin tend to be linked to sediments that have taken on stable forms (Xiao et al., 2015). Therefore, it is essential for source identification and risk assessment to know the chemical fractions of heavy metals, especially in highly contaminated areas.

2.2.1 Metal speciation

The purpose of a speciation analysis is to ascertain the presence of different chemical species or fractions of an element in sediment, soil or sludge. Determining the metal speciation fractions helps to assess the toxicity and mobility of metals, which inturn, allows for investigation of the extent to which benthic sediments have been contaminated. The metal speciation fractions are determined by the application of sequential extraction (Sahuquillo et al., 2003). Thus, there is a growing interest in using this method for investigating metals in aquatic sediments, and currently, there are more than a dozen methodologies for determining the percentages of these particular fractions. The European Community Bureau of Reference recommends using Tessier's five-stage sequential extraction method, which can separate heavy metal contamination in soil or sediments into five different fractions. The fractions or species in which metals occur are (Tessier et al., 1979):

1. **Exchangeable or F1 fraction**: This fraction is made up of metals that have been adsorbed on a solid surface. Changes in the ionic content of the water or a shift in the sorption-desorption equilibrium might cause this fraction to flow into the solution. This fraction is readily available and mobile.

2. **Associated with carbonates or F2 fraction:** Metals in the form of carbonates or co-precipitated with carbonates make up this fraction. Lowering the pH disrupts the carbonate equilibrium, allowing metals to enter the solution.
3. **Associated with the hydrated oxides of iron and manganese or F3 fraction:** This fraction is made up of metals that are adsorbed on the extended surfaces of hydrated oxides of iron and manganese that precipitate under anoxic (reducing) conditions. The sediment may dissolve, and the metals move into the solution when iron and manganese are reduced.
4. **Associated with organic matter or F4 fraction:** This fraction includes metals adsorbed on the surface of organic matter and metals associated with sulphides. Although temporarily immobilised, they may later move to one of the other fractions due to the natural mineralisation of the sediment.
5. **Stably associated with minerals or F5 fraction**: This fraction consists of metals permanently immobilised in the crystal lattices of both primary and secondary minerals; they do not harm the environment under normal conditions.

The first three are particularly risky because they can release metal loads due to decreased pH. The metal present in the inert phase, i.e. the F5 fraction, corresponds to detrital or lattice-bound metals that cannot be remobilised. The proportion of metal found in this category can be used as an indicator of the level of contamination in the water zone of a river or stream (Pardo et al., 1993). These fractions have different remobilisation behaviours under varying environmental circumstances (Förstner, 1993). Thus, to accurately assess their toxicity, it is required to determine the overall concentration of heavy metals in sediment and their presence in various chemical forms (speciation) in sediments.

2.2.2 Current literature on pollution in benthic sediments due to heavy metals

A qualitative and systematic approach was used to identify the current literature on pollution in sediments due to heavy metals. The procedures below were performed to retrieve the needed literature. First, a comprehensive literature review for research and review papers is conducted in the Scopus database using different keywords. After eliminating duplicates and doing an initial screening, documents are selected. The following keywords were entered: **TITLE-ABS-KEY** ("sediments" AND "heavy metals" AND "speciation" AND ("factor" OR "index") AND ("surface water" OR "river")) and a total of 46 relevant articles were found after a thorough screening out 151 articles.

Liu et al. (2011) investigated the concentrations of heavy metals and chemical speciation in sediment obtained from the Shuangtaizi River estuary's tidal flat. The reduction in Cd and Pb concentrations in the upper 66 cm of the core is striking. The distribution of chemical fractions supports the natural origin of the residual fractions of these metals. In contrast, due to pollution, only the non-residual components of Cd and Pb have increased the core in the past five decades.

The geoaccumulation index (I_{geo}) reveals that most heavy metals are unpolluted, even with increased concentration.

Wu et al. (2012) assessed the environmental quality, pollution level, bioavailability, and toxicity of heavy metals in the benthic sediments obtained from the Daliao River and Yingkou Bay estuary. The geoaccumulation index (Igeo) results reveal that, except for Pb, the pollution levels of four metals at 15 sampling sites fall into the "unpolluted" category. The pollutant load index (PLI) reveals that most sites in the study areas are not polluted, except for sites 1, 2, 3, 8, 12 and 13. Pb is the most polluting heavy metal, followed by Zn, Co. Cu. Co, Cu, Pb, and Zn pose a "low" potential ecological risk at all sampling sites. The order of potential ecological risk is as follows: Pb > Co > Cu > Zn. Sequential extraction of the metals reveals that Pb, Cu, Co, and Zn in the sediment are in a relatively stable state at majority of estuary locations in the Daliao River and Yingkou Bay, indicating that this region receives minimal pollution. Co, Pb, and Zn are also present in a labile state at a few sites and, as such, are deemed to be of anthropogenic origin.

Zhao et al. (2012) used two standard index methodologies to analyse the pollution and ecological risks of six metals (As, Cu, Mn, Pb, Sb, and Zn) in Yangtze Estuary sediments (i.e., total content and speciation indices). Each sampling site's ecological risk was uniformly low to moderate. According to the speciation index, the contamination levels at these sites were low to moderate, while the total content indices showed moderate to significant contamination. When assessing the ecological risks of each sampling site and each metal (Cu, Pb, or Zn), considerable correspondences and discrepancies were observed between the total content and the speciation indices when evaluating the contamination of specific locations in the estuary area.

Chen et al. (2013) used 5-step sequential extraction technique to evaluate the metal speciation and level of pollution in sediments in Kaohsiung Harbour, Taiwan. Individual contamination factors (ICF) and a global contamination factor (GCF) were utilised to quantify the level of sediment pollution. The results revealed that Kaohsiung Harbour was at high risk for Pb, Cd, Cr, and Cu. GCF data revealed that stations near the river mouth, fish port, and industrial locations contributed to Kaohsiung Harbour's high potential risk. The risk of heavy metals to the environment was analysed using the risk assessment code (RAC), and the results showed that Pb, Cd, Zn, and Mn posed a medium to high risk, while Hg, Cr, Cu, and Ni posed a low to medium risk.

Yuan et al. (2014) studied some geochemical properties, major elements and trace metals of sediments from a river, estuary, and lake in the Yangtze River Delta plain, a fluvial plain. Enrichment factors evaluated the heavy metal contamination levels. The comprehensive RACs of sediments followed the order of river sediments > estuary sediments > lake sediments. Among heavy metals, Cd and Pb revealed the highest risks for sediments.

Cui et al. (2014) used inductively coupled plasma mass spectrometry to assess heavy metals (Cr, Cu, Ni, Pb, Cd, As, Zn, and Hg) spatial distribution, sources, and potential ecological risk in sediments of the Wuyuer River and its tributaries. Cd was found to be mobile, whereas Cu, Pb,

Zn, Cr, and Ni were found to be bioavailable. Hg and As were mostly found in the residual fraction. According to Hakanson's potential risk index, the overall potential ecological risk of these heavy metals in the Wuyuer River was moderate. Among the metals tested, Hg and Cd were the two most potentially toxic, and they were mostly found around the cities of Keshan, Yi'an, and Fuyü. Cd concentrations should be regularly monitored since the speciation of Cd in river sediments revealed Cd to be extremely bioavailable.

Liu and Shen (2014) used HCl-HNO_3-HF-$HClO_4$ digestion and an optimised BCR sequential extraction method to assess the total and extractable quantities of Cu, Pb, and Zn in the surface sediments of west Chaohu Lake (China). The enrichment factor was used to measure metal pollution; sediment quality guidelines (SQG) and risk assessment code (RAC) were used to estimate possible eco-risk. The metal enrichment factors based on total and extractable concentrations all revealed greater values in the northern lake region and decreasing values in the south. Overall, sediments Cu, Pb and Zn constitute a low-to-high risk to the environment.

Pandey et al. (2014) investigated the geochemical fractions of nine heavy metals (Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Pb) present in Ganga river sediments in Varanasi using the sequential extraction method (SEP) and acid digestion. The Pollution Load Index (PLI) ranged between 1.2 and 3. Cd and Pb had higher geo-accumulation index (Igeo) concerns, whereas Mn, Fe, and Ni had negative Igeo at all sampling stations. Significant amounts of Pb, Cd, Cu, and Ni were found in the available fractions.

Bo et al. (2015) collected water and sediment samples from residential, mixed (commercial and industrial), and agricultural sectors in the Taihu Lake region of China. They analysed the levels of six heavy metals in these samples: Cd, Cr, Cu, Ni, Pb, and Zn. Among the six metals found in water and sediments, the concentration of Zn was the highest, while the concentration of Cd was the lowest. Most of the Cd, Cr and Pb were found in the acid-soluble fraction, residual fraction, and reducible fraction in the sediments, respectively. The ecological risk assessment results reveal that Cd constituted the greatest threat to the ecosystem.

Bo et al. (2015b) investigated the ecological risks in sediments from residential, mixed (industrial and residential), and agricultural sectors of rural rivers in southern Jiangsu Province, China, using a European community bureau of reference sequential extraction method. In all three locations, there were larger amounts of Cd in the acid-soluble fraction, Cr in the residual fraction, and Pb in the reducible fraction. As a result, Cd and Pb might pose extremely significant environmental risks. Furthermore, risk assessment code (RAC) study revealed that most sediment samples in residential and mixed areas were categorized as having a 'high' to 'extremely high' risk due to their Cd, Ni, and Zn contents.

Islam et al. (2015) used a sequential extraction approach to examine surface sediment from a downstream river (Paira) related to the marine ecosystems of the Bay of Bengal. This study aimed to assess the mobility and dynamics of heavy metals and their potential ecological consequences.

The findings demonstrated a high environmental risk of Cd due to its increased availability in the exchangeable fraction (21%) and a significant amount in the carbonate bound fraction. According to the risk assessment code (RAC), the highest mobility of Cd provides a greater environmental risk and hazard to aquatic biota.

Pandey et al. (2015) used physicochemical and metal analysis to examine the influence of urban drains on river water and sediments. Metal speciation and elemental composition analysis (SEM-EDS) were utilised to estimate metal pollution load in river sediments. Metal speciation study revealed that accessible and labile fractions dominated in Cr, Ni, Cu, Zn, Cd, and Pb, except for Mn and Fe, which were prominent in residual forms. The pollution Load Index (PLI) ranged from 1.8 (S1) to 3.9. (S15). Cd (6.88-8.97) and Pb (2.41–3.24) had the maximum Geo-accumulation index (GAI).

Fernandes and Nayak (2015) collected two sediment cores from the Swarna estuary and the Gurpur estuary in Karnataka, India, to examine metal bioavailability and toxicity. Metal speciation investigation on selected core samples for Fe, Mn, Ni, Zn, Cu, Co, and Cr revealed that Fe was predominantly connected with the residual fraction, accounting for 93±0.5% in Gurpur and 84±6% in the Swarna estuary. When total metals were compared to sediment quality values (SQV), the values of Co in both cores exceeded the apparent effect threshold (AET). Except for Co in the Swarna estuary, where it posed a significant risk of toxicity associated with the sediments, metal speciation studies revealed little impact on aquatic life.

Lei et al. (2016) investigated the spatial variation and ecological risks of six heavy metals (Cd, Cr, Cu, Ni, Pb, and Zn) in estuarine surface sediments from 10 estuaries in the Hai River Basin (HRB), China. Metal speciation revealed that Cd is linked with the exchangeable and carbonate fractions (averaging 21.3%) and was classified as medium to high risk by the Risk Assessment Code (RAC). Moreover, half of the Cr, Cu, Ni, and Zn were connected with the residual fraction. The mobility order of these heavy metals was Cd > Pb > Zn > Cu > Ni > Cr based on the total of the first three fractions. Compared to the cinnamon soil background values, the potential ecological risk index (RI) values varied from 25.6 to 168, with an average of 91.2, suggesting a low ecological risk in HRB estuarine areas. The toxic-response factor was dominated by Cd and Pb (45.8 and 25.5%, respectively).

Tatone et al. (2016) assessed the potential risks of toxic metals in settling material collected with sediment traps in the Uruguay River using enrichment factors (EF), sediment quality guidelines (SQG), and speciation using a four-step sequential extraction procedure. The average EF of Zn, Cr, Ni, and Pb is less than 1.5, suggesting natural sources, although Cu and Mn are consistently higher (EF > 2), implying some anthropogenic effects. Cu concentrations consistently duplicated the SQG (35.7 g/g), implying that unfavourable biochemical consequences may occur. However, the speciation findings indicated that most metals are connected with the residual

fraction, closely related to the mineral matrix, and unavailable to aquatic species. The only exception is Mn, which is mostly found in non-residual fractions (∑F1 − F3 = 82 %).

Helali et al. (2016) collected three core samples from the Mejerda River Delta in Tunisia and tested them for major and trace elements to determine pollution and bioavailability. The enrichment factor (EF), the geo-accumulation index (Igeo), and chemical speciation were all employed. Pb, and to a lesser extent Zn, are the most polluting elements off the Mejerda outlet, according to estimates of trace element accumulation using the EF and the I_{geo} index. These metals are also the most hazardous in terms of bioavailability.

Duodu et al. (2016) examined the percentages of heavy metals in sediment cores collected from a wetland in the Huaihe River Basin, China, utilising BCR sequential extraction procedures. Enrichment factor (EF) and geoaccumulation index (I_{geo}) values increased significantly from a depth of 4 cm to a maximum of 5 and 1 near the surface, respectively, showing the significant increase of pollutant discharge in recent years, particularly for Cd, Cu, and Zn. Moreover, the proportions of the various fractions of Cd, Cr, Cu, Pb, and Zn had substantially changed. Significant increases in the amounts of mobile fractions of heavy metals from a depth layer of 4 cm to the surface layer revealed that the bioavailability of Cd, Cu, Cr and Pb increased significantly. The sharp rise suggested that Cd and Cu have very high mobility and bioavailability. In addition, significant proportions of Zn's reactive fractions suggested that Zn was highly mobile and presented a considerable ecological risk to aquatic creatures in the riverine wetland.

Liu et al. (2017) collected and analysed sediment samples from 22 metal-polluted sections in Xiangjiang River (XJR), China. Various indices evaluated ecological risk and enrichment of heavy metals. Pollution levels initially increased during industrialisation but decreased after hundreds of polluting industries were prohibited. Cd, As, Pb, and Hg contaminated the majority of the sections with severe pollution and elevated risk. Both the enrichment factor and the geoaccumulation index followed the order Cd > Hg > Zn > Mn > Pb > Cu > As > Cr. Cd (66.93%), Zn (33.80%), Pb (30.81%), Mn (18.38%), Hg (17.58%), Cu (10.20%), As (9.81%), and Cr (7.65%) were the most bioavailable metals. The assessment of their bioavailability, biotoxicity, and abundance revealed that Cd contamination was the most prominent, but other metals' contamination should not be overlooked.

Xu et al. (2017) studied the speciation of heavy metals in nine coastal rivers draining into Laizhou Bay. Throughout the river stretch, Cr Zn, Cr, Ni, and Pb were bound to the residual fraction; however, Cd was preferentially linked to the exchangeable phase. The combined potential ecological risk indices were utilised to assess possible threats. The majority of sampling locations in the Laizhou Bay watershed present a moderate ecological risk from metals.

Ke et al. (2017) collected sediment samples from 19 sites inside the protected area of the Liaohe River and analysed them for heavy metals to determine the potential ecological risk posed by their presence. The results revealed that the pollution caused by seven heavy metals falls in the

following order: Cd, As, Cu, Ni, Pb, Cr, and Zn. The investigation of metal speciation revealed that Cd, Pb, and Zn were dominated by non-residual fractions and exhibited high mobility and bioavailability, indicating substantial anthropogenic origins. Based on the potential ecological risk index (PERI), the geo-accumulation index (I_{geo}), and the risk assessment code (RAC), it was determined that Cd posed a high to extremely high potential ecological threat in the area.

Zhao et al. (2017) assessed the ecological risks and existing forms of heavy metals in Jialu River, China using sediment samples collected from 22 locations. The results revealed that the exchangeable fraction was mostly connected with Cr, with a mean percentage of 54.23%. Mn and Zn were primarily found in the reducible fraction, constituting 42.72% and 37.71% of total contents, respectively. Co, Ni, Cu, and Cd had the highest mean percentages of the residual fraction, with 58.75%, 59.02%, and 48.21%, respectively. The sediment samples from the lower reach exhibited higher heavy metal levels, and Cd posed a significant risk compared to other metals.

Jun et al. (2017) employed chemical speciation to evaluate heavy metal contamination (Cu, Fe, Mn, Ni, Zn, Cd, Cr, and Pb) in surface sediments in the Yellow River. The immobile fraction (residual) quantity of Fe, Cu, Cd, and Cr was higher than the mobile fractions. The significant enrichment of Ni and Cr is a serious environmental concern. Based on the geoaccumulation index, enrichment factor, and pollution load index, the degree of contamination of heavy metals in sediment was Cr > Cd > Pb > Cu > Mn > Ni > Fe > Zn.

Tang et al. (2017) thoroughly examined heavy metal contamination, toxicity, and ecological risk assessment for surface river sediments in China's Haihe Basin. The geoaccumulation and pollution load indices found that most surface river sediments in the Haihe Basin were contaminated with the analysed metals, particularly at the junction zone of the Zi Ya He and Hei Long Gang watersheds. The five heavy metals in the sediments were all anthropogenic, and the enrichment degrees were in the order Cu > Pb > Zn > Cr > Ni. The analysed sediments of the Haihe Basin have a low potential ecological risk, according to the potential ecological risk index (38.9). The sediments, however, were possibly ecologically detrimental to health based on the mean likely effect concentration quotient (0.547), which might be attributed to the speciation of the five metals in the sediments.

Zhang et al. (2017a) investigated heavy metal (Cr, Cu, Ni, Pb, and Zn) pollution and the risks resulting from heavy metals in riverine sediments in mountainous (MB), mountain–plain (MPB), and plain urban-belt (PB) areas in the Haihe Basin, China. The enrichment factors suggested that sediments in the MPB were more contaminated with Cu and Zn. The risk assessment code and individual contamination factor demonstrated that Zn was mobile and posed ecological concerns, with the exchangeable fractions comprising 21.1%, 21.2%, and 19.2% of the total Zn contents in samples from the MB, MPB, and PB, respectively. The non-residual portions of Cr, Cu, and Zn in the MPB were 56.2%, 54.9%, and 56.6%, respectively, of the total values, indicating that these metals may be highly bioavailable.

Mna et al. (2017) assessed the level of pollution in the Bizerte Lagoon, which is subjected to high anthropogenic pressure, using sediment cores from two locations. The enrichment factor and geoaccumulation index were evaluated along with their chemical speciation fractions. Enrichment factor and geoaccumulation index indicate a build-up of Cd, Zn, Cr, and Pb; however, chemical speciation indicates that only Cd and Mn pose a concern.

Kang et al. (2017) investigated the speciation of heavy metals (Cr, Mn, Ni, Cu, Zn, As, Cd, and Pb) in the sediments of Jiaozhou Bay. Based on the summation of the first three phases, the mobility sequence was in the order: Mn > Pb > Cd > Zn > Cu > Ni > Cr > As. Enrichment factors (EF) suggested that heavy metals in Jiaozhou Bay exhibited negligible enrichment to a small degree of enrichment. The potential ecological risk index (RI) suggested that Jiaozhou Bay exhibited a low ecological risk and a rising trend since the 1940s due to the growth in anthropogenic activities.

Sebei et al. (2018) investigated 18 surface sediments and collected 5 water samples from the Tessa River near two mining sites. There is no spatial variation in the chemical speciation of these metals. Except for Cd, which is bound to the residual fraction and the carbonates, all other heavy metals are bound to the five sediment chemical fractions: the residual fraction (>52%), the oxyhydroxides fraction (21%), the carbonates (16%), and finally the organic matter and the exchangeable fraction (10%). The bioavailable component of the heavy metals tested surpasses 45%, posing a toxicity risk.

Ji et al. (2018) investigated the distribution and risk assessment of heavy metals, specifically Cr, Cu, Mn, Ni, Pb, and Zn, in the Yongding River's overlaying water, porewater, and sediments. To assess the ecological risk of heavy metals in sediments, sediment quality guidelines (SQGs) and pollutant load index (PLI) were used. Heavy metal concentrations increase in the order shown below: porewater < sediment < overlaying water. Pb and Zn concentrations in porewater exceed the criteria maximum concentrations (CMCs). According to principal component analysis, Mn in sediments is predominantly connected with the exchangeable fraction, Cu and Zn with the reducible fraction, and Cr, Ni, and Pb with the residual fractions. Overall, PLIs of sediments in the Yongding River at sites 1, 4, and 10 demonstrate moderate contamination from Cr, Cu, Mn, Ni, Pb, and Zn.

Islam et al. (2018) investigated the ecological risk of heavy metals (Cr, Ni, Cu, As, Cd, and Pb) in the surface sediment of Bangladesh's Buriganga River. In sediments, the mean concentrations of Cr, Ni, Cu, As, Cd, and Pb were 297, 240, 280, 21, 7.7, and 731 mg/kg, respectively. The pollution load index (PLI) and contamination factor (CF) were utilised to assess the ecological risk. PLI ranged from 2.9 to 13 in the winter to 2.4 to 11 in the summer, demonstrating increased sediment deterioration due to metal pollution. CF values of Cd ranged from 6.7 to 61, demonstrating that Cd had a significant impact on the investigated sediments. The decreasing order of pollutants was Cd > Pb > Cu > Ni > As > Cr, based on the severity of potential ecological risk

for a particular metal (E_r^i). The potential ecological risk of Buriganga River sediments ranged from significant to extremely high.

Pal and Maiti (2018) investigated the speciation of metals (Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn) in water and sediment in the Damodar River during the pre and post-monsoon seasons. The calculated values of I_{geo} in river sediment during both seasons revealed that the majority of metals were found in the I_{geo} class 0-1, indicating that the sediment was unpolluted to moderately polluted. The partition coefficient results showed that Cr, Pb, Co, and Mn possess high retention capacity, whereas Cd, Zn, Cu, and Ni are highly mobile.

Amor et al. (2019) determined the total and speciation concentrations of heavy metals (Pb, Cd, Cu, Zn, Ni, and Cr) in the Rades-Hamam Lif coast surface sediments of the Mediterranean Sea. Several geochemical indices were used to estimate the risk of contamination and environmental concerns associated with heavy metals on surface sediments. The sequential extraction revealed high potential Cd bioavailability due to a higher proportion of the exchangeable fraction. On the other hand, Cr and Cu were largely attached to the residual fraction, indicating minimal toxicity and bioavailability. The mobility order was Cd > Pb > Ni > Zn > Cr > Cu, while the level of pollution was Cd > Pb > Ni > Zn > Cr > Cu.

He et al. (2019) investigated the potential sources and environmental concerns of heavy metals by studying the quantities and chemical speciation of 8 heavy metals (Zn, Cu, Ni, Pb, Cr, Hg, As, Cd) in 34 sediment samples taken from the Changjiang (Yangtze River) estuarine area. The sequential extraction results indicate that, except for Cd, most heavy metals are bonded to the inert residual fractions. The geoaccumulation index (Igeo) and enrichment factor (EF) reveal that there is no noticeable enrichment or pollution of these heavy metals in this area. The Cd adsorbed to river sediments is bonded to exchangeable, carbonate, and Fe-Mn oxides associated phases, which may be chemically reactive in estuarine processes and thus presents significant environmental threats as human activities increase.

Sun et al. (2019) investigated the pollution status and ecological risks of heavy metals (Cu, Zn, Pb, Ni, and Cr) in Songhua River sediments in an urban region having petrochemical industries. The freezing period had the highest mean concentration of total heavy metals near the petrochemical industrial area. According to the geo-accumulation index, Cu was the most polluting heavy metal. The potential ecological risk index indicated that the overall and specific heavy metal concentrations were linked with low ecological risk. Based on the overall concentrations, the speciation and risk assessment code results suggested that these heavy metals in the sediment possibly had poor mobility and bioavailability, with non-residual fractions of 35.2% for Cu, 37.46% for Zn, 33.83% for Pb, 24.59% for Ni, and 36.04% for Cr. Thus, heavy metals in the sediment posed no to moderate ecological threats.

Verma and Pandey (2019) evaluated the distribution of eight heavy metal fractions (Zn, Cr, Cu, Pb, Cd, Ni, Fe, and Mn) in the Ganga River bed sediment. The study was conducted during

low summer flow (March to June 2017) and concentrated on a 285 km middle stretch of the Ganga River between Lucknow upstream and Banaras downstream. The majority of the metals had a significant residual proportion. However, Zn, Pb, and Cd accounted for approximately 20-30% of the exchangeable form. The contamination factor (CF), enrichment factor (EF), pollution load index (PLI), geoaccumulation index (I_{geo}), and risk assessment code (RAC) revealed moderate to severe contamination, indicating considerable anthropogenic influence. According to the United States Environmental Protection Agency, concentrations of metals were extremely toxic in numerous regions.

Gadkar et al. (2019) collected and analysed Mangrove and mudflat sediment cores at the Cumbarjua Canal and Zuari River confluence site for metal enrichment and bioavailability within the Zuari Estuary. The contamination factor revealed a moderate level of contamination in mangroves and mudflats for Fe, Mn, Cu, and Co. The pollution load index indicated metal contamination in the Zuari Estuary sediments. The metal speciation analysis showed the lithogenic source of Fe (in mudflats), Cu, Zn, and Co. Bioavailable Fe (in mangroves) and Mn (in mangroves and mudflats), on the other hand, suggest probable mobilisation, preferential accumulation, and bioavailability. Furthermore, sediment quality data suggested bioavailable Mn and Co toxicity, and the risk assessment code demonstrated that Mn posed a medium risk.

Park et al. (2020) studied the spatial and temporal distribution and the sources of particulate heavy metals in Masan Bay. This area has been polluted by waste from nearby massive manufacturing complexes and urban areas for a long time. The primary sources of several heavy metals would be river water and the release mechanism of river debris. During June and August, metal-enriched particles from river water and WWTP discharge were the primary contributor to particulate metal concentrations; however, in October and November, particulate metal concentrations were provided by detrital particles comparable to sediments. Metal concentrations in offshore samples were used as the background level to generate the contamination index. Mn, Co, and Zn contamination levels in zone I ranged from 3 to 5, indicating moderate to severe contamination.

Ayyanar and Thatikonda (2020) investigated the presence of toxic heavy metals at locations associated with industrial effluents, and idol immersion practises in India's contaminated Lake Hussain Sagar. The high concentrations of As, Cd, Pb, Zn, Cu, and Ni found in surface water were caused by the industrial effluent discharge. Surface sediments have an increasing sorption capacity as Pb > Cd Ni Cr Zn Cu, and desorption demonstrated that Pb and Cu were retained to a larger extent due to high clay and organic content. Chemical speciation revealed that 20-50% of Zn and 50-80% of Cd were connected with exchangeable and carbonate fractions. Zn and Cd are rated as 'high risk' to 'very high risk' by the risk assessment code, while Cr, Pb, Ni, and Cu are rated as 'medium risk'. The enrichment factor value of sediments for Cd (20.42-119.48), Zn (2.19-4.85), Cu (2.02-3.19), and Pb (2.85-7.72) indicates that anthropogenic activities have caused significant contamination.

Cai et al. (2021) determined the chemical speciation fractions of As, Cd, Cu, Pb, Mn and Zn in the surface sediments of Wujiang River. The results showed that, except for As, the mean total concentrations of all heavy metals were greater than their background levels. Cu, Zn, Pb, As, and Hg are typically found in trace amounts. The reducible fraction and exchangeable fraction dominated the Mn and Cd, respectively. Cu, Zn, Pb, As, and Hg showed low risk in this study area, according to the risk assessment code (RAC); however, Mn and Cd posed a significant risk at most sample sites. According to the I_{geo} categorization, Cd poses the greatest environmental concern, and heavy metal contamination levels in surface sediments were frequently in the order: Cd > Zn > Cu > Pb > Hg > Mn > As. The potential ecological risk index (PERI) categorized 2 sample locations in the significant risk category and the rest in the moderate risk category.

Shin et al. (2021) researched the levels of contamination and the distribution, speciation, and sources of heavy metals in the sediment of the Taewha River. The heavy metals in the sediment displayed the following order: Zn > Pb > Cr > Cu > Ni > As > Cd. The total content of heavy metals rose throughout the sampling period in the downstream region. Anthropogenic activities contaminated the sediment in the sample stations to varying degrees, as determined by EPA guidelines for sediment pollution, which include the pollution intensity index (I_{POLL}) and potential ecological risk index (RI). A sudden increase in Pb, Zn, and Cd concentrations was seen at Station 3 over the summer and fall, which was connected to increasing clay mineral content induced by seasonal and lithological changes. This rise can be ascribed to a neighbouring industrial complex or the oxidation of sulphide minerals, which could be tied to an abandoned amethyst mine. Sequential extraction tests reveal that the potential toxicity of each element varies. Metals with larger percentages of exchangeable fractions and fractions bonded to carbonates, on the other hand, can be extremely hazardous.

Liu et al. (2021) investigated the distribution, level of pollution, and potential ecological risk of potentially toxic elements (PTEs) from Mn mining in Changyang, Western Hubei, China. River water and sediments were collected to measure 7 PTEs (As, Cd, Cr, Cu, Mn, Pb, and Zn), as well as pH and redox potential (Eh). The findings revealed that Mn was the most predominant contaminant, and a combination of anthropogenic and natural processes dominated the pollution concentration of Mn in the studied region. The contamination factor (CF) and pollution load index (PLI) values also found Mn as the predominant contaminant in river water. There was minimal As and Pb pollution downstream, but there was considerable Cu pollution upstream. Upstream and downstream were the primary polluted river sections. The geo-accumulation index (I_{geo}) and potential ecological risk index (PERI) data indicated negligible Mn pollution in river sediments. The PERI of PTEs from river sediments was tolerable, with only Mn posing a moderate ecological risk upstream.

Abdallah (2021) employed sequential extraction and ecological risk assessment methodologies to analyse sediment samples from six Nile River sectors for heavy metal concentrations of As,

Cd, Mo, and Sn. The findings revealed that: (1) total heavy metal content was affected by industrial effluents, runoff, and black water, which were all greater in the coastal area; and (2) sequential extractions revealed that As and Mo were primarily composed of residual, Cd was bound to exchangeable, and Sn was dominated by iron and manganese oxide fractions. P Pollutant load index (PLI) value > 1 indicates impairment in sediment quality. The average Igeo values for the metals studied indicated that they were heavily polluted. Cd and Sn had the highest bioavailability levels across all regions studied, whereas As had the highest in the El-Tina area, indicating a potential threat to aquatic life. According to the ICF data, Cd and Sn had the most ability to be released from the sediments into the overlying water.

Hao et al. (2021) investigated the bioavailability and speciation of heavy metals in sediments using seven heavy metals. The results indicated that all HMs in the sediments were within their permissible exposure limit (PEL); however, Cd and Zn were substantially higher than the soil baseline. The majority of HMs were discovered to be in a residual fraction, with an exceedingly low exchangeable percentage. The lack of correlations between pH, electrical conductivity (EC), total dissolved solids (TDS), and HM bioavailability suggested that HMs in the carbonate-bound phase are stable and resistant to environmental fluctuations. In contrast, the significant correlations between redox potential (Eh), turbidity, organic matter (OM), main grain size (Mz), and HM bioavailability suggested that HMs in the reducible and oxidizable forms are susceptible. As a result, rather than their carbonate-bound form, the variance in HM bioavailability in karstic rivers is predominantly regulated by their reducible and oxidizable forms.

Anandkumar et al. (2022) utilised a modified BCR sequential extraction methodology to assess the speciation of trace metals in the coastal sediments of Miri and decipher the seasonal geochemical processes responsible for observed phenomena, and identify potential sources of these trace metals. The findings of the granulometric analysis revealed that coastal currents helped by monsoon winds have altered the grain size distribution of the sediments, allowing us to separate the research region into north-eastern and southwestern segments with differing geochemical compositions. Cu (> 84 %) and Zn (82 %) concentrations are largely related to the easily accessible exchangeable fraction. Pb and Cd dominate the non-residual fractions, while Fe, Mn, Co, Ni, and Cr dominate the residual fraction. Using Pearson's correlation and factor analysis, the major mechanisms controlling the chemistry of the sediments are determined to be the association of Cu and Zn with fine fraction sediments, sulphide oxidation in the SW portion of the study area, atmospheric fallout of Pb and Cd in the river basins, precipitation of dissolved Fe and Mn supplied by the rivers, and remobilization of Mn from the coastal sediments. Various pollution indices suggest that the coastal sediments of NW Borneo are polluted with Cu and Zn and are highly bioavailable, posing a hazard to the native aquatic species, coral reefs, and mangroves.

Miao et al. (2022) analysed seven heavy metals (HMs) in the surface water and sediments of the Liujiang River Basin in order to comprehend the HMs' speciation and the environmental

variables influencing their accumulation and conversion. The results indicated that the amounts of HM in water are well below the primary water quality standard; however, Cd and Zn concentrations in sediments are much higher than their respective soil baselines. Only Cd and Pb are mostly present in non-residue forms (carbonate-bound and reducible fractions). Non-significant correlations revealed that pH and redox potential (Eh) might have little effect on HM concentrations in water, whereas strong connections emphasised the effects of Eh, organic matter, and mean particle size on metal accumulation in sediments. Opposing associations between EC, TDS, pH, and Cd revealed that acidic wastewater discharge contributed to the buildup of Cd in sediment. Only certain forms of Cd, As, Cu, Zn, and Pb were observed to convert significantly between water and sediments, suggesting that their unique forms should substantially govern the conversion of HMs in sediments. The extremely high risk revealed by the elevated levels of monomial ecological risk factor (E_r^i) and potential ecological risk index (RI) is only present upstream, but the elevated level of Cd should be viewed as a severe environmental concern.

Dan et al. (2022) employed chemical fractionation of heavy metals (Zn, Pb, Cu, and Cd) to assess the pollution status and related risks in the Cross-River Estuary (CRE) surface sediments and nearshore areas surrounded by a deteriorating mangrove ecosystem. Cd and Zn were the most contaminated heavy metals according to the contamination factor (CF) and geo-accumulation (I_{geo}). The presence of high proportions of Zn (63.78%), Pb (64.48%), Cd (76.72%), and Cu (48.57%) in non-residual fractions suggests that these heavy metals are bioavailable. Cd was shown to have a moderate to high ecological and bioavailability risk using the ecological risk (Er) and risk assessment code (RAC).

2.3 Statistical and Multivariate Analysis

The data on water quality encompasses information about the surface water body and the surrounding environment. It is crucial for water quality management to interpret specific qualities and their accompanying information (Li et al., 2015). In this regard, multivariate statistical techniques (MSTs) have the ability to appreciate and understand the complicated interaction between the acquired data and uncover latent factors. These statistical tools assist decision-makers in identifying the pollution sources that affect water quality and spatiotemporal variations. Also, these tools help interpret complex water quality data sets, identify information in high variance data matrices, and provide a correct understanding of the monitoring data (Bhat et al., 2014; Gholizadeh et al., 2016; Li et al., 2015; Njuguna et al., 2020; Xia et al., 2020). Frequently applied MSTs include principal component analysis (PCA), positive matrix factorization (PMF), cluster analysis (CA), principal component analysis/multiple linear regression (PCA-MLR or APCS-MLR), chemical mass balance (CMB), and edge analysis (UNMIX) (Li et al., 2003; Shi et al., 2011). It is necessary to select the best appropriate statistical technique to accurately reveal the pollutant source profile of the studied area due to the diverse physical and mathematical background of these MSTs (Bilgin, 2015; Olsen et al., 2012; Sergeant et al., 2016). PCA has been acknowledged

as an efficient technique for identifying probable pollution sources (Jiang et al., 2018; Zhang et al., 2018). In addition, the PMF proposed by the United States Environmental Protection Agency (USEPA) has been widely used to allocate pollution sources (Jiang et al., 2017). Thus, PMF and PCA are the most preferred MSTs because they can handle complicated data sets generated through exhaustive field surveys and do not require backward analytic techniques (Lee et al., 2016; Salim et al., 2019).

However, distinguishing the types of pollution sources and the lack of understanding regarding the transport pathways of pollutants from the source to a water body are obstacles to these procedures. In addition, these methods generate preliminary results that are subject to substantial uncertainty due to inadequate data quality and measurement mistakes in samples (Liu et al., 2018; Zhang et al., 2009). Thus, no single method can precisely and uniquely identify the sources associated with various land uses. Therefore, merging the methodologies is considered beneficial for detecting the origins and transport pathways of pollutants (Hu et al., 2018; Wang et al., 2019). The combination of PCA and PMF models increases the precision of analytical information regarding probable pollution sources, offers correct estimates of the proportion of a given pollutant that originates from each source, and eliminates biased judgement in source allocation (Ran et al., 2021; Xu et al., 2021; Zhang et al., 2018). In addition, integrating PCA with PMF can improve understanding and precisely identify and allocate the probable sources and transport routes of pollutants in a body of water (Ahmed et al., 2016; Guan et al., 2018).

2.3.1 Principal Component Analysis (PCA)

PCA is a dimensionality-reduction method frequently used to reduce the dimensionality of large datasets by decreasing the number of variables into a smaller set that retains the majority of the information of the original set. Reducing the number of variables in a dataset reduces accuracy, but the idea of dimensionality reduction is to trade off a little accuracy for simplicity. Because smaller datasets are simpler to analyse, comprehend, and process without redundant variables. To summarise, the goal behind PCA is straightforward: decrease the number of variables in a dataset while retaining as much information as feasible.

PCA reduces the dimensionality of high-dimensional data while preserving trends and patterns. It accomplishes this by reducing the data to fewer dimensions that serve as feature summaries. High-dimensional data are highly prevalent in biology and develop when numerous features, such as gene expression levels, are recorded for each sample. This type of data offers various issues that PCA mitigates, including computational expense and a higher error rate due to repeated test corrections when testing each feature for correlation with an outcome. PCA is an unsupervised learning method (Altman and Krzywinski, 2017) that detects patterns without knowing whether the samples are from different treatment groups or have phenotypic differences.

PCA reduces data by geometrically projecting it onto smaller dimensions known as principal components (PCs), with the goal of determining the best summary of the data using a small

number of PCs. The first PC is chosen to reduce the overall distance between the data and its projection onto the PC. Minimizing this distance maximises the variance (σ^2) of the projected points. The second (and subsequent) PCs are chosen similarly, with the added criterion that they be uncorrelated with all preceding PCs. In other words, projection onto PC1 is unrelated to projection onto PC2; PCs are orthogonal geometrically. Because no correlation is required, the maximum number of PCs feasible is either the number of samples or the number of features, whichever is less. The PC selection procedure maximises the correlation (R^2) (Altman and Krzywinski, 2015a) between data and projection and is similar to doing multiple linear regression on the projected data against each variable of the original data (Altman and Krzywinski, 2015b; Krzywinski and Altman, 2015) For example, when utilised in multiple regression with PC1, the projection onto PC2 has the highest R^2.

2.3.2 Positive matrix factorization (PMF)

Paatero and Tapper (1994) and Paatero (1997) provided detailed descriptions of the positive matrix factorization (PMF) model. It has been used as an alternative to PCA in a variety of research initiatives, including wet deposition (Anttila et al., 1995; Juvela et al., 1996), harbour bottom sediments (Poulton et al., 1995), and sources of aerosols (Huang et al., 1999; Polissar et al., 1996). The major distinction between PCA and PMF is that the non-negativity of components (loadings and scores) is integrated into the PMF model. Furthermore, instead of relying on information from the correlation matrix, PMF employs a point-by-point least-squares minimization approach. As a result, the generated profiles can be compared directly to the input matrix without processing. This is PMF's distinct advantage over PCA.

2.3.3 Current literature on statistical and multivariate analysis

A qualitative and systematic approach was used to identify the current literature on pollution source apportionment in surface water bodies using PCA and PMF combined. The procedures below were performed to retrieve the needed literature. First, a comprehensive literature review for research and review papers is conducted in the Scopus database using different keywords. After eliminating duplicates and doing an initial screening, documents are selected. The following keywords were entered: **TITLE-ABS-KEY** (("surface water" OR "river" OR "sediments") AND ("principal component analysis" AND "positive matrix factorization")) and a total of 21 relevant articles were found after a thorough screening out 74 articles.

Comero et al. (2011) employed Positive Matrix Factorization (PMF) to analyse a geochemical dataset acquired from XRF analysis on sediments from Italian alpine lakes. PMF discovered four factors related to the chemical properties of lake sediments in catchment areas: heavy metal-bearing minerals, carbonates, silicates, and phosphate and sulphur supply. Also, a new PMF factor was identified to modify individual uncertainty estimates properly, explaining a possible Pb contamination source.

Wang et al. (2015) applied principal components analysis (PCA) and positive matrix factorization (PMF) to identify and apportion pollution sources of toxic metals in the Yangtze River estuary's surface sediments. PC1 can be classified as a sewage component, PC2 as an atmospheric deposition component, and PC3 as an agricultural nonpoint component based on PCA. Eight sources were identified with PMF to identify better and quantify the concentrations: agricultural/industrial sewage mixed (18.6%), mining wastewater (15.9%), agricultural fertiliser (14.5%), atmospheric deposition (12.8%), agricultural non-point (10.6%), industrial wastewater (9.8%), marine activity (9.0%), and nickel plating industry (8.8%). The presence of toxic elements appears to be linked to anthropogenic activity rather than natural sources. The PCA results provided a generic classification of sources, which served as the foundation for the PMF study. PMF resolves more variables with greater explained variation than PCA. The combination of the two strategies may yield more realistic and consistent results.

Gholizadeh et al. (2016) assessed the water quality and identified probable pollution sources in South Florida's three major rivers using absolute principal component score-multiple linear regression (APCS-MLR), principal component analysis (PCA), and receptor modelling (PMF) technique. The PCA technique revealed five and four possible pollution sources in the rainy and dry seasons, respectively. The APCS-MLR revealed that the principal sources of river water contamination were point source pollution discharges from agricultural fields and industrial and domestic wastewater. The data matrix was also subjected to the PMF receptor model and compared PCA and APCS-MLR models, and there were considerable disparities in the projected contribution for each probable pollution source, notably during the wet season.

Chen et al. (2016) used several GIS-based multivariate statistical approaches to analyse datasets on water quality in the Liao River system in China. Positive matrix factorization (PMF) and principal component analysis (PCA) discovered eight probable pollution factors for each data structure element, accounting for more than 61% of the overall variance. The largest latent pollution factor for group A was oxygen-consuming organics from farmland and woodland runoff. The principal contaminants in Group B were oxygen-consuming organics, oil, nutrients, and faeces. The identified contaminants for group C included oxygen-consuming organics, oil, and hazardous organics.

Alves et al. (2018) used positive matrix factorization (PMF) and principal component analysis (PCA) as well as Spearman's correlation analysis and Kruskal-Wallis test to interpret a water quality data set resulting from a nearly two-year monitoring programme (May 2013 to April 2015) for water samples collected from the municipal water treatment plant (WTP). PCA was useful in determining the most critical criteria for changes in water quality. For summer-autumn: total coliforms (TCOLI), winter: water level (WL), water temperature (WT), and electrical conductivity (EC), and spring: colour (COLOR) and turbidity (TURB) were the critical criteria. PMF was applied to the entire data set, allowing for the source apportionment of water pollution via three

anthropogenic sources: the discharge of household sewage (primarily expressed by Escherichia coli (ECOLI)), industrial effluent, and agricultural runoff.

Salim et al. (2019) investigated the effectiveness of PCA-MLR and PMF models using stormwater runoff data gathered from a small catchment (Site 1) with urban development projects and a sub-watershed outlet (Site 2). The PCA-MLR model found three pollution sources at both sites, but PMF detected five with a comprehensive source mechanism that included two extra sources. According to the modelling results, domestic wastewater and soil erosion were the dominant sources of contamination at Sites 1 and 2. The PMF model outperformed the PCA-MLR model in terms of performance assessment statistics such as the Nash coefficient (0.86-0.99), percent error (14 to 2), and regression coefficients ($R^2 \leq 0.99$). Overall, the PMF receptor modelling approach proved more robust for the current study sites with varying land-use types.

Chen et al. (2019) employed principal component analysis (PCA), modified grey relational analysis (MGRA), absolute principal component score-multiple linear regression (APCS-MLR), and positive matrix factorization (PMF) receptor modelling technologies to evaluate the groundwater quality and apportion the potential contamination sources in the Lalin river basin, China. The contamination assessment with PCA and MGRA suggested that human activities polluted the Lalin river basin groundwater. The PCA method identified five and four potential contamination sources in wet and dry seasons, respectively, and the main sources were the same. The APCS-MLR and PMF methods apportioned the source contributions to each groundwater quality variable. The final results showed that agricultural sources, including wastewater, agrochemicals and fertilisers, were identified as the main sources of groundwater contamination in wet and dry seasons.

Dash et al. (2020) used four environmetrics tools to examine the water quality and geochemistry of Deepor Beel, Assam, India. The 23 sampling locations were categorised as high, low, or moderate contamination sites based on hierarchical clustering (HCA). The discriminant analysis (DA) of the water quality dataset resulted in 9 parameters, which were predominantly responsible for cluster distinction. The standardized dataset was then subjected to principal component analysis (PCA) to identify potential contamination sources. PCA produced two important main components that defined the anthropogenic and natural elements that define water contamination. Finally, PMF with four pre-defined factors was applied to the dataset matrix. The major contributors are surface water runoff, Leaching from the Boragaon landfill site, discharge from the Basistha River, and effluent discharge from industries in the wetland.

Sun et al. (2021) identified the sources of heavy metals (HMs) in river sediments at a Pb-Zn mine in Danzhai, Guizhou, China. The concentrations and coefficient of variation of the HMs demonstrated that the river sediments surrounding this Pb-Zn mine were polluted with HMs. Based on the results of the positive matrix factorization (PMF) model, principal component

analysis (PCA), and correlation analysis, the primary sources of selected HMs are agricultural activities (livestock and poultry rearing) and Pb-Zn mining and smelting activities, coal mining activities.

Jafarabadi et al. (2021) identified possible sources of potentially harmful metals (PTMs) in Persian Gulf coastal sediments (Iran). Total and fraction analyses revealed significant pollution levels of metals. PTMs were largely linked with the oxidizable and residual fractions in most cases. The combined PCA-PMF modelling technique found four major metal sources in the research region (anthropogenic, vehicle-related, agricultural, and lithogenic).

Wu et al. (2021) quantified contamination sources of eight heavy metals (Mn, Ni, Pb, Zn, As, Cr, Cd, and Cu) in the Beiyun River. Positive matrix factorization (PMF) and principal component analysis-multiple linear regressions (PCA-MLR) were applied to identify pollution sources. The PMF determination coefficient (R^2) is closer to one, implying that the pollution source investigation is highly precise. The Beiyun River is estimated to receive the most heavy metals via industrial operations (23.0%), transportation (17%), agriculture (16%), and atmospheric deposition (16%).

Magesh et al. (2021) quantified the sources of heavy metals in the Schirmacher Hills lacustrine systems. The positive matrix factorization (PMF) model was recommended to understand metal relationships and designate likely sources. The results show that natural weathering of source rocks (78.53%), followed by human-induced activities coupled with atmospheric deposition (21.47%), are the primary sources of heavy metals.

Mamun and An (2021) employed multivariate statistical techniques (MSTs) to examine spatial and temporal fluctuations in water quality in order to identify and quantify potential pollution sources influencing the Yeongsan River. A 15-year dataset containing 11 water quality parameters from 16 monitoring locations was used. The results of the principal component analysis (PCA), factor analysis (FA), and positive matrix factorization (PMF) revealed that two sewage treatment plants, agricultural operations, and animal farming all had a negative influence on river water quality.

Luo et al. (2021) investigated and evaluated heavy metal concentrations in the Fenghe River Basin (FRB) in Shaanxi Province, revealing their sources. The Water Quality Index (WQI), the Numero Index (Pn), the Geological Accumulation Index (I-geo), and the Potential Ecological Risk Index are used to assess heavy metals in water and sediments (RI). The positive matrix factorization (PMF) and principal component analysis (PCA) models are used to investigate the connection and origin of heavy metals. The results show that the majority of heavy metals in the water are within the environmental quality criteria for surface water but exceed the background level concentration in the soil. In considerable detail, PMF models were able to identify the factors or sources of heavy metals in water and sediment. The primary sources of pollution in the region are urban construction and traffic, the electronics sector, machinery manufacturing, and tourism.

Niu et al. (2021) utilized two receptor models, PCA-MLR and PMF, to estimate the source types and contributions of heavy metals in the Pearl River Delta's Modaomen Estuary. This estuary's many sources included industrial products, imports from the Pearl River Estuary, traffic pollution, fuel/oil combustion, and agricultural uses. Among these heavy metal pollution sources, industrial activity was the most significant contributor, accounting for 29 % in 2003, 28 % in 2015, and 38 % in 2018.

Celen et al. (2022) assessed the spatiotemporal variation in 13 water quality parameters (TOC, TN, NO_2^-, NO_3^-, TP, SO_4^{2-}, Cl-, TSS, colour, pH, temperature, DO, EC) at 8 monitoring stations in the Ergene basin for a year using multivariate statistical techniques such as positive matrix factorization (PMF) and principal component analysis (PCA). Based on pollution levels and point/non-point sources determined by field observations, the eight monitoring stations were divided into five groups (five for Gr-A and three for Gr-B). The principal component analysis yielded four and three latent variables that explained 87% and 89% of the overall variation in the Gr-A and Gr-B datasets, respectively. The PMF model-derived parameters revealed that the principal pollution sources for Gr-A areas include textile and leather sector discharges, agricultural activities, household discharges, and seasonal influences. Gr-B sites include household garbage, agricultural fertilisers, and industrial contaminants. Thus, PMF analysis for standard water quality measures provides a reliable statistical method for identifying complex contamination sources.

El-Ouaty et al. (2022) examined the concentrations of heavy metals (HM) and sulfurs in the bottom sediments of Nador lagoon in North-East Morocco. Enrichment Factor (EF), QGis Inverse Distances Weighted (IDW) interpolation models, multivariate statistical techniques (Principal Component Analysis - Agglomerative Hierarchical Clustering - Pearson correlation matrix), Positive matrix factorization (PMF model), and grain size/organic matter analysis were used to assess the spatial distribution and source identification. The following are the average heavy metal concentrations in the sediments: Sr > Ba > V > Zr > Zn > Cr > Rb > La > Cu > Pb > Ni > Ce > Nd > Co > Sc > Nb > Ga > Th > Y > Hf, and the Sulfurs. The findings revealed that the average amounts of S, Sr, Pb, and Nd in bottom sediments were greater than the background levels. The EF results revealed that anthropogenic activities enhanced HM contents, particularly for Sr and Sc and then for Th, V, Ce, Nd, and Pb. The potential HM and sulfurs identified sources revealed that: 1) Ga, Y, Hf, Nb, Rb, Sc, Zr, Nb, Co, Zn, Th, Ni, V, Ba, Cr, and Cu were derived from industrial activities related to stealing and mining; 2) Wastewater treatment station and lagoon watershed were the main sources of Sr in the lagoon; 3) Zn, Cu, and Pb were linked to sulfurs, mostly originating from agricultural inputs and mining activities;

Shi et al. (2022) investigated the source apportionment of accessible occurrence forms of heavy metals (AHMs) in surface sediments using absolute principal component scores-multivariate linear regression (APCS-MLR) and positive matrix factorization (PMF) receptor models in Dianchi Lake, southwest China. Both the APCS-MLR and PMF models revealed three kinds of

probable sources that were comparable: (1) agricultural fertilizer/insecticide, atmospheric deposition, and traffic emissions; (2) natural transitions; and (3) industrial and sewage wastes. Furthermore, the comparative findings indicated that the PMF model was more viable for assessing AHMs sources in wetland sediments because it can investigate one additional source, namely plant maintenance and waterfowl feeding, and has greater accuracy in forecasting AHM concentrations.

Yang et al. (2022) characterised the spatiotemporal variation of the water pollution and the pollution sources during rainfall and no-rainfall events in an urban-rural marginal catchment in the Qingshan District, Wuhan City, China. The PMF model was more compelling and supplied more sources than the PCA-MLR model. Although non-point sources were the dominant pollution source (56.7~71.9 %), receptor models confirmed that the dominating sources shifted across urban, urban-rural, and rural regions. Due to extensively constructed activities and industrial emissions, runoff and land surface dusk were priority pollution sources for urban-rural and urban regions, with contributions of 39.1~41.7 %. However, industrial factors and land surface darkness (42.4%) dominated rural pollution causes. The runoff and residential factors (39.3~45.9 %) governed the primary pollution sources during rainfall events in urban and urban-rural settings. The study's findings provide a scientific foundation for addressing specific pollution causes and restoring water quality in the urban-rural marginal watershed.

Jiang et al. (2022a) investigated the concentrations of seven potentially hazardous metals (PHMs) in the surface sediments of a typical bay in southern China that has been subjected to long-term human impacts to elaborate on environmental pollution characteristics, risks and sources of PHMs. According to the findings, the mean concentrations of each element were as follows: 7.00 (As), 0.164 (Cd), 79.1 (Cr), 29.3 (Cu), 25.4 (Pb), 0.042 (Hg), and 107.4 (Zn). PHMs in the surface sediments displayed decreasing patterns from the nearshore to the offshore sea regions, displaying high, moderate, and low concentrations in the inner, middle, and outer bay, respectively. These trends were seen as one moved from the nearshore to the offshore sea areas. According to the geo-accumulation index, the most contaminated metal was Cd, followed by Cu, Hg, Zn, Cr, and Pb. The inner bay was found to have the highest pollution levels according to the modified contamination of degree (mCd) and the nemerow integrated pollution index (NIPI). Both the potential ecological risk index (PERI) and the toxic risk index (TRI) demonstrated that the presence of PHMs has been an unfavourable factor in the overall health of the bay, particularly the inner bay. However, PERI indicated that Cd and Hg were the primary contributors to the negative impacts, but TRI demonstrated that Cu and Cr were the significant contributors. Based on the principle component analysis results, it was determined that As originated from natural sources, whilst other PHMs originated from anthropogenic sources. Positive matrix factorization model further explained that human activities were the major sources, including industrial and marine transport activities (40.5%) and discharge of domestic sewages (29.3%), while natural

sources contributed 30.2%. Based on these findings, prevention and control of PHMs pollution, as well as the management and remediation measures in coastal bays, can be implemented.

Wen et al. (2022) collected samples of periphytic biofilm and surface water from 16 different locations along the Lancang River in China in order to evaluate the spatial distribution, enrichment factor (EF), and potential ecological risk index (RI) of heavy metal elements found in the samples. According to the findings, the amounts of heavy metals found in the samples taken from the biofilm were statistically substantially higher than those found in the samples taken from the surface water (one-way analysis of variance, $p<0.001$). In addition, 37.50% of the biofilm samples had considerable heavy metal pollution, as indicated by a mean EF>5. As and V were the most polluting anthropogenic metals, while natural sources were considered responsible for Ni and Co enrichment. The findings of the RI evaluation demonstrated that As poses a consistent ecological concern. The results of the principal component analysis with multiple linear regression (PCA-MLR) and positive matrix factorization (PMF) model indicated the presence of heavy metal ions in the biofilm samples could largely be attributed to industrial activities (PCA–MLR: 68.89%; PMF: 76.39%). This is followed by a mixed source of natural and agricultural activities (PCA–MLR: 18.12%; PMF: 13.56%), as well as traffic emissions (PCA–MLR: 12.99%; PMF: 10.05%).

2.4 Health Risk Assessment

Heavy or trace metals are necessary for the regular functioning and metabolism of the human body in small amounts, but excessive amounts can be hazardous to human health. However, because heavy metals are not easily biodegradable or chemically, they accumulate over prolonged periods, creating significant health problems. Heavy metals can enter the body through direct ingestion, inhalation, and skin absorption from contaminated water (USEPA and EPA 1991). Many researchers accomplished the significance of trace elements to human health and plant growth and understood the mechanisms of metal transport (Bhuiyan et al., 2010; Gupta et al., 2008; Kale et al., 2010; Pawar and Nikumbh, 1999). Heavy metal contamination is one of the world's most important health issues due to its persistence and accumulation, which endangers living creatures and the ecosystem (Belkhiri et al., 2017). The pollutants bioaccumulate in water and sediments and continue to biomagnify in aquatic animals, posing a cancer risk once they enter the food chain (Ali and Khan, 2018). Consumers using water and fish from rivers contaminated with heavy metals face serious health consequences. These heavy metals, under favourable conditions, pose a risk to the entire aquatic ecosystem (Sundaray et al., 2011). These heavy metals are considered conservative pollutants because they are inert in all environmental samples (Olivares-Rieumont et al., 2005).

Although metals such as Cr, Cu, and Co are necessary for regulating human metabolism, excessive amounts induce metabolic problems (Jolly et al., 2013). Cr (VI) can enhance the risk of stomach cancer in humans (Welling et al., 2015). Toxic metals such as Pb and Cr are known carcinogens and have been linked to problems of key organs such as the kidneys and the

neurological system, which become apparent after many years of exposure (Singh et al., 2010; Zhuang et al., 2009). In addition to the impacts of individual heavy metals and in combination provide both carcinogenic and non-carcinogenic health concerns to humans (Bermudez et al., 2011). Endocrine disruptors in humans and animals include heavy metals such as Cd, As, Ni, Hg, Pb, and Zn. Pb can impair adrenal and ovarian steroidogenesis, resulting in compromised female fertility. The systemic non-cancerous metals Cd, Co, Cu, and Hg, etc., on the other hand, can also have health consequences even at minute concentrations (Djadé et al., 2021). As a result, residents of polluted areas, children and adults, are exposed to hazardous compounds via various sources and pathways, resulting in a wide range of toxic exposures for them (Chen et al., 2019; Wang et al., 2017; Xu et al., 2020). The health risk levels posed by various contaminants can be assessed by evaluating the human health risk factors (Wu et al., 2010). Three major pathways, direct ingestion, inhalation and dermal absorption, govern human health risk factors, but the inhalation pathway largely does not account for contaminated water (Kim et al., 2011).

2.4.1 Current studies on health risk assessment

A qualitative and systematic approach was used to identify the current literature on human health risk assessment in India. The procedures below were performed to retrieve the needed literature. First, a comprehensive literature review for research and review papers is conducted in the Scopus database using different keywords. After eliminating duplicates and doing an initial screening, documents are selected. The following keywords were entered: **TITLE-ABS-KEY** (("surface water" OR "river" OR "sediments") AND "heavy metals" AND "health risk assessment" AND "India")) and a total of 17 relevant articles were found after a thorough screening out 36 articles.

Krishna and Mohan (2014) studied the seasonal variation in concentrations of potentially harmful heavy metals (As, Cd, Cr, Cu, Ni, Pb, and Zn) in surface and groundwater samples taken from an industrially contaminated site in Hyderabad, India, during the pre-monsoon and post-monsoon. Health risk assessments, such as chronic daily intake (CDI) and hazard quotient (HQ), were also computed. Surface and groundwater had elevated amounts of hazardous components measured by the CDI and HQ indices, which are safe for human consumption only with appropriate treatment.

Magesh et al. (2017) sought to determine the amount of trace elements present in the groundwater in the Tamiraparani river basin. There is a declining trend of trace element content: Fe > Zn > Mn > Cu > Cr > Ni > Pb. Except for Fe, Mn and Pb, all trace elements were below the global drinking water limits. The exposure dosage index (CDI) and the hazard quotient (HQ) were utilised to carry out a health risk assessment. The order of the CDI values of trace elements for the oral and dermal pathways was: Fe > Zn > Mn > Cu > Cr > Ni > Pb and Fe > Zn > Mn > Cu > Cr > Ni > Pb, respectively. The Hazard quotient of all the elements was less than one, implying a low probability of contamination occurring via the oral and dermal channels.

Mitra et al. (2018) investigated the spatial and temporal variation of 13 trace elements in the surface water along the north-south gradient of the Hooghly River Estuary, India, and subsequently assessed the risk to human health using USEPA guidelines. Using the hazard quotient and hazard index values, it was determined that the elements under study posed non-carcinogenic risk. However, As, Cd, Pb, and Cr exceeded the upper limit of cancer risk (10^{-4}), resulting in carcinogenic risk concerns for both children and adults, with children being more prone.

Wagh et al. (2018) procured 40 groundwater samples from dug and bore wells in rural habitations during the 2011 pre-monsoon season and assessed them for toxic metals (Pb, Cd, Zn, Fe, Mn, Cr, Cu, Co and Ni). The human exposure risk based on HI reveals that groundwater is not safe to drink for people of all age groups, as the value of HI ≥1.

Shil and Singh (2019) conducted studies in the Mahananda River to identify dissolved heavy metals in the water and to assess the non-carcinogenic and carcinogenic human health hazards using the hazard index (HI) and incremental lifetime cancer risk (ILCR). For this, 11 heavy metals (Mn, Cr, Fe, Cu, Zn, Ni, Co, As, Cd, Pb, and Hg) were analysed and found to be safe for drinking. HI_{Adult} analysis showed that 7 sampling locations in pre-monsoon and two in post-monsoon exceeded the hazard limit (HI > 1). HI_{Child} values were greater than one (>1) at 6 sampling locations in the pre-monsoon season. The ILCR analysis showed that 15 sample locations (55.55%) in pre-monsoon and 6 sampling locations in post-monsoon had exceeded the cancer risk limitation value (ILCR > 10^{-4}). Because of the reduced amount of water available and the increased concentrations of metals in river water, pre-monsoon values were higher than post-monsoon values.

Singh et al. (2020) evaluated the surface water quality of the Kameng River (Assam, India). The contamination index calculated the degree of contamination of each sampling site due to heavy metals. Its associated human health risk assessment was done by computing average daily intake and Hazard quotient (HQ). The HQ of all sampling sites varied from 0.14 to 1.21.

Mukherjee et al. (2020) investigated heavy metal toxicity in groundwater and the associated human and ecological risks in Birbhum district, India. The investigation found that the target heavy metals in groundwater has the following order: Fe > Zn > Sr > Mn > Cr > Pb > Ni > Cu > Cd, with mean concentrations of the carcinogens, Pb and Fe are above their maximum allowable limits. The study also indicated that residents might be under very significant cancer risks as a result of ingestion and dermal contact with carcinogenic heavy metals in groundwater. Children in the study area were more vulnerable to carcinogenic and non-carcinogenic risks than adults, primarily through oral exposure. Thus, the study recommends that residents drink treated groundwater.

Setia et al. (2020) conducted a systematic study to assess the spatial and temporal variability of metal contamination in water from the entry point of the Sutlej River in Indian Punjab to its tail end when it leaves the country. The likelihood of cancer risk was also estimated through

human health risk assessment. The likelihood of cancer risk due to ingestion of metals through water was in the order: Cd > Ni > Cr > As, and the risk is higher in the areas along the trans-boundary.

Arisekar et al. (2020) collected samples from four locations along the Thamirabarani River. The concentrations of heavy metals in fish and sediments ranged from 0.001 to 9.505 mg/kg and 0.294 to 106.25 mg/kg, respectively. Lifetime cancer risk (LCR) levels were higher than the acceptable threshold value (ATV), and children were more exposed to health concerns. Except for site 4, the LCR of hazardous metals except Cd were within the ATV (10^{-6}~10^{-4}); Cd was over the ATV and posed a substantial cancer risk to downstream residents.

Maity et al. (2020) investigated the amounts of geogenic arsenic and other heavy metals influencing As release in aquifers in the Bhojpur area of Bihar, India. The assessment of human health risks in two demographic groups in the region revealed that the general water quality is slightly contaminated, although the threat linked with it is low.

Prasad et al. (2020) evaluated the human health risk from heavy metal contamination in surface water in the Upper Ganga river, India. Health risk assessment was carried out using chronic daily intake (CDI) and hazard quotient (HQ), which demonstrates that exposure via ingestion and dermal routes currently poses no substantial threat to human health (HQ < 1) for any metal in particular. $HQ_{Ingestion}$ levels for Cr (adults 0.51, child 0.55) and Pb (adult 0.31, child 0.34) were substantially higher than for other heavy metals in the two population groups studied. The total effect of all the heavy metals is expressed in terms of hazard index (HI). $HI_{Ingestion}$ ranged from 0.85 to 1.64 for adults and 0.92 to 1.77 for children, showing that both groups face health risks, with children being more vulnerable to either exposure pathway. Furthermore, during the post-monsoon season, HI levels suggested an increased risk to health for both groups. Higher hazard index (HI) values (>1) in the Upper Ganga river imply an ever-increasing non-carcinogenic risk to the riverine population. The study underlines the role of heavy metals in compromising the water quality of the Upper Ganga River and calls for immediate action to reduce human health hazards.

Kumar et al. (2020) investigated the groundwater quality of the Sutlej River Basin (SRB), Punjab (India), to be used as drinking and irrigation water and further accessed their impacts on human health. The quality parameters investigated are pH, conductivity, cations, anions, and heavy metals. According to the health risk assessment, there is a possibility that Cr and As in soil and groundwater may cause cancer. The cumulative exposure of U (HQ-1.21), NO_3^- (HQ-0.37), and F^- (HQ-0.34) through groundwater consumption may pose long-term adverse repercussions, according to a non-carcinogenic risk assessment (hazard index-1.98). Constant long-term monitoring is necessary to keep a check on changes in the quality of the groundwater and soil system, even if the contaminants currently do not pose any threats to human health.

Dixit and Siddaiah (2021) investigated the seasonal variation in concentration by collecting 50 samples from four wetlands in Gurugram, Haryana. In addition, the research area's overall water quality and risk were assessed using a heavy metal pollution index (HPI) and a health risk index (HRI). According to BIS guidelines, Fe and Cr exceed the permissible limits in summer, but all metals are within limits in winter. Metals Fe and Mn had the highest concentrations, followed by those in the following order: Zn, Cu, Ni and Cr. A critical index limit of 100 is set for HPI calculations, and all the values are less than that. But most of its values are >30, which falls into the high-class HPI category in W1, W3 and W4 wetlands signifying high deterioration. HI being > 1 in W1 and W4 may pose a non-carcinogenic risk to children and adults in summer and winter through ingestion and dermal exposure, while the risk is limited to summer only in W3 and W4.

Khan et al. (2021) analysed the level of heavy metal contamination and health concerns connected with millions of people drinking water from the Gomti River. The degree of contamination (Cd) was determined to be '11.93,' indicating 'high' risk levels due to heavy metal contamination in River Gomti, spanning an approximate distance of 61 km in Lucknow city areas. Due to the 'near neutral' pH of river water, heavy metals were found to have little mobility. The human health risk assessment results show that the hazard index linked with non-carcinogenic risks exceeded the allowable limits at all sample locations. The largest health risk was observed downstream of Lucknow city at the Bharwara sewage treatment plant, indicating higher levels of heavy metal in the river water following treatment from the Bharwara STP. The results of the carcinogenic risk assessment indicated that children were more vulnerable to health concerns.

Parween et al. (2021) evaluated the environmental quality of the most contaminated stretch of the Yamuna River in Delhi. The study was carried out to investigate the toxicity and health risks related with persistent contaminants found in the fluvial ecosystem. In order to evaluate heavy metals (Cd, Co, Cr, Cu, Fe, Mn, Ni, Pb and Zn), 84 sediment samples were obtained from the Delhi part of the Yamuna flood plain. Sediment quality criteria, enrichment factor, geo-accumulation index, degree of contamination, and Pollution Load Index were used to conduct the eco-toxicological assessment. Cd, Pb, and Zn had maximum amounts matching level 3 of concern, while Cr and Ni had the greatest concentrations, corresponding to level 4 of concern. Sediment samples were highly enriched and moderately contaminated with Cd and Pb, as represented by enrichment and contamination factors (CF). Zn > Cd > Cr~Ni > Pb~Cu had the highest CF for metals. The pollutant load index was highest in Delhi, near the Yamuna's outflow point. The non-carcinogenic and carcinogenic potentials of vegetables were evaluated in terms of human health risks.

Bhat et al. (2022) collected stream water and surface soil samples from the Tungabhadra River watershed in Karnataka to analyse heavy metal contamination and its impacts on human health. In soils, the sequence of non-carcinogenic (NCR) and carcinogenic risk (CR) is ingestion > dermal > inhalation, and in water samples, it is ingestion > dermal > inhalation. In soil and

water, the hazard index (HI) is mostly < 1, suggesting that there are no NCRs for adults or children. Two water samples exhibited HI > 1 for As in adults, while 50% of water samples in children showed HI > 1, owing to which children were more exposed to arsenic contamination than adults. The total carcinogenic risk (TCR) of Pb in adults was less than the recommended limit of 10^{-5}. TCRs for As and Cr in adults and Pb in children were found to be within the range of 10^{-6}~10^{-4}, indicating a lesser carcinogenic risk. On the other hand, the TCR of As and Cr in children is above the threshold values of 10^{-5}, indicating serious carcinogenic effects. TCR of As and Cd in water samples exceeded the 10^{-5} threshold limit, indicating a considerable carcinogenic risk for adults and children.

Selvam et al. (2022) evaluated the quality of the Thamirabarani River system in south India using physicochemical variables (pH, EC, TDS, DO, BOD, turbidity and NO3), microbiological parameters (total coliform bacteria, faecal coliform bacteria, faecal streptococci and escherichia coli), and toxic metals (As, Cr, Fe, Cu, Zn, Cd, and P). The findings revealed a 20% reduction in the contamination ratio during the lockout period compared to the period before the lockdown was implemented. The health risk assessment models (HQ, HI, and TCR) identified carcinogenic and non-carcinogenic concerns for children and adults due to dermal adsorption and ingesting exposures. During the lockdown period, the HI values for both As and Cr were higher than the recommended level (>1), but the potential risk for children and adults was still modest compared to the time before the lockdown.

2.5 Prediction of Heavy Metals using Regression Techniques

With global economic development, more and more heavy metals are being discharged into aquatic habitats. Heavy metals can enter aquatic environments through anthropogenic or natural processes such as rainfall, surface runoff, mineral mining, industrial effluents, and urban sewage (Başak and Alagha, 2010; Concas et al., 2006; Ghosh and Maiti, 2018; Palanques and Diaz, 1994; Sylaios et al., 2012). According to studies, heavy metals in aquatic habitats harm the ecosystem and degrade the quality of drinking water (Chowdhury et al., 2016; Martin and Holdich, 1986; Utgikar et al., 2004). Monitoring concentrations of heavy metals and their behaviour is critical for environmental management, particularly when safeguarding drinking water sources.

Many studies have investigated the risks of dissolved heavy metal concentrations in rivers and lakes (Deng et al., 2018; Jiang et al., 2012; Liang et al., 2018; Müller et al., 2008; Rajeshkumar et al., 2018). Several physicochemical and environmental variables, including pH, organic matter, suspended matter, and channel morphology, have been found in studies to influence heavy metal behaviour (Ahmed et al., 2013; La Colla et al., 2015; Y. Wang et al., 2016; Yang et al., 2016). Therefore, continuous monitoring of heavy metal concentrations and other environmental variables is thus required to fully explain the behaviour and associated risks of heavy metals in aquatic systems. Heavy metal (HM) detection techniques include inductively coupled plasma mass spectrometry (ICP-MS), inductively coupled plasma optical emission spectrometry (ICP-OES), inductively coupled plasma atomic emission spectrometry (ICP-AES), flameless atomic absorption spectrophotometry (FAAS) and atomic absorption spectroscopy (AAS) etc. (Koelmel

and Amarasiriwardena, 2012; Massadeh et al., 2016; Sreenivasa Rao et al., 2002; Daşbaşi et al., 2016; Siraj and Kitte, 2013). These apparatuses are highly sensitive, selective, relatively expensive and involve complex operational procedures. Therefore, detecting heavy metal concentration statistically has become a research focus for relevant government departments, academia, and environmentalists (Muyessar et al., 2013). Modelling is an alternate solution for lowering the cost of water quality monitoring while measuring the concentration of heavy metals.

Some previous studies have developed models based on geochemical processes to predict heavy metal concentrations in aquatic environments (Braga et al., 2010; de Blois et al., 2003; Fall and Fall, 2001; Garneau et al., 2017). However, there are still several challenges in simulating heavy metal concentrations in the real world using process-based models. Process-based models typically require detailed descriptions of geochemical processes with numerous input variables, which exceed the limited available monitoring data (Palani et al., 2008; Thorslund et al., 2017). Furthermore, the internal mathematical representations of geochemical processes can be of a significantly higher order than what can be observed externally, resulting in poor performance due to parameter uncertainties and imprecise mathematical descriptions of geochemical processes (Cho et al., 2016; Rode and Suhr, 2007). The model outputs in the case of process based models are sensitive to critical factors, and improper validation with little data may happen, which reduces the model accuracy (Ciffroy, 2020). Thus, there is a clear need for a relatively rapid and effective model that can predict heavy metal concentrations using minimal monitoring data. Regression techniques such as multiple linear regression (MLR) and machine learning (ML) techniques can aid in estimating the concentration of heavy metals in surface water.

2.5.1 Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. MLR aims to model the linear relationship between the explanatory (independent) and response (dependent) variables.

Formula and calculation of multiple linear regression:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \in \quad (2.4)$$

where, for i = n observations:

y_i = dependent variable

x_i = explanatory variable

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

$\in$ = the model's error term (also known as the residuals)

The ordinary least squares method is the most widely used procedure for developing estimates of the model parameters. The multiple regression model is based on the following assumptions:

- There is a linear relationship between the dependent variables and the independent variables
- The independent variables are not too highly correlated with each other
- y_i observations are selected independently and randomly from the population
- Residuals should be normally distributed with a mean of 0 and variance σ

The coefficient of determination (R^2) is a statistical metric used to determine how much variation in results can be explained by variance in the independent variables. Even if the predictors are unrelated to the outcome variable, R^2 always increases when additional predictors are added to the MLR model. R^2 cannot thus be used to determine which predictors should be included and which should be omitted from a model. R^2 can only be between 0 and 1, with 0 indicating that none of the independent variables can predict the outcome and 1 indicating that all independent variables can predict the outcome without errors.

2.5.2 Machine Learning (ML) Algorithms

ML algorithms are increasingly involved in a number of facets of daily life, from what people can read and watch to how people can shop, meet, and travel. Consider the detection of fraud. When someone uses a credit card to make a purchase, ML algorithms immediately check the transaction to see whether it is fraudulent. They anticipate whether the transaction is fraudulent or not depending on the traits of the previous purchases. ML has a wide range of applications in search and recommendation systems. These ML algorithms are at the foundation of how commercial search engines work, beginning with the first query entered. Furthermore, search engines often use data about how people interact with the search site to improve their future efficacy, such as which pages are clicked, how long someone reads the pages, and so on. Similarly, movie recommendation websites employ ML algorithms to predict what customers enjoy based on previous ratings. Ready-to-use ML algorithms for speech recognition, language translation, text classification, and many other tasks are now being offered as web-based services on cloud computing platforms, significantly expanding the audience of developers who can use them and making it easier than ever to put together solutions that apply ML at a high level. In general, there are two types of ML algorithms. The first kind is supervised learning, in which the prediction of some output variable is associated with input items. So, for a credit card transaction, if the output is a projected category from a finite number of alternatives, such as fraudulent or not. This is a classification problem within supervised learning, and the function used to complete the classification task is known as the classifier. If the output variable we wish to forecast is a real-valued number rather than a category, such as the time it would take a car to accelerate from 0 to 100 kmph, it is a regression problem and must apply a regression function. To make predictions, supervised learning requires a training set of labelled objects. Labelling some problems can be simple or challenging, depending on the amount of labelled data required, the level of human experience or expert knowledge required to deliver an accurate label, and the complexity of the labelling process,

among other things. Crowdsourcing, Amazon's Mechanical Turk, and Crowd Flower are among the most popular platforms. A continuous member function can be used to solve any regression problem. Any classification problem can be mapped to distinct categories in the same way. Unsupervised learning is the second major class of ML methods, in which the input data lacks labels. In a method known as clustering, such problems are tackled by identifying some relevant structure in the incoming data. So, once this structure is discovered in the form of clusters, groups, or other fascinating subsets, the result will be provided in the form of a meaningful summary of the input data and possibly visualising the structure. Unsupervised learning enables us to approach issues with little or no knowledge of the end outcome. In this study, the ML algorithm used is supervised learning, more specifically, a regression problem.

2.5.2.1 Artificial Neural Network (ANN)

The inventor of the first neurocomputer, Dr. Robert Hecht-Nielsen, defines a neural network as – "*...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs*". The ANN is a mathematical algorithm for non-linear data processing that can connect several input variables to produce one or more output variables (Mustafa et al., 2012; Tao et al., 2019). It is based on the biological operation of neural connections in the human brain. This method is usually made up of three layers: input, hidden, and output. The hidden layer, which can be one or more layers, is responsible for making sense of the input layer's multidimensional expansion (Ehteram et al., 2021). An ANN's architecture comprises units known as neurons, sometimes nodes (Mustafa et al., 2012; Tao et al., 2019). One disadvantage of ANN is that it is computationally expensive and heavily dependent on hardware capability (Mijwil, 2018; Poblete et al., 2017; Zor et al., 2017). The ANN requires processing power parallel with its structure for the models to be trained with realistic and efficient training durations.

Furthermore, no clear set of guidelines appears to exist to determine the ANN structure throughout model building or coding. As a result, an appropriate ANN architecture must be developed through model development experience and trial-and-error techniques. The ANN is a black-box model, which limits its capacity to identify causal linkages between factors and specific output, and it may overfit during training due to model interaction or non-linearity (Poblete et al., 2017; Zor et al., 2017). According to several studies (Ehteram et al., 2021; Mustafa et al., 2012; Tao et al., 2019), the mathematical model of an ANN may be represented by:

$$y_i = f\left(\sum_{i=1}^{N} w_{ij} x_i + b_j\right) \tag{2.5}$$

where y_i is the output variable, N is the number of neurons, ω_{ij} is the weight connecting the j^{th} neuron and the i^{th} neuron, x_i is the input vector, b_j is the bias of the j^{th} neuron, and f is the activation function. As mentioned by Mustafa et al. (2012), there is no certain rule for selecting

the number of neurons in the hidden layer and other hyperparameters. Therefore, the hyperparameters must be selected through trial and error.

2.5.2.2 Random Forests

Random forests, also known as 'random decision forests,' are an ensemble learning approach that combines many algorithms to improve classification, regression, and other tasks. The performance of each classifier or regressor is poor on its own, but when coupled with others, it can generate exceptional results. This is the basic idea behind the random forest, wherein a number of decision trees are coupled together to enhance the performance of the model. The algorithm begins with a 'decision tree' (a tree-like graph or model of decisions), with an input at the top. The data is then split into smaller and smaller groupings based on specific variables as it moves down the tree.

RF is a popular ML approach for making predictions from a given collection of predictors using an ensemble of decision trees (Breiman, 2001). Prediction outcomes, such as mode (in a classification framework) or mean prediction (in a regression framework), are obtained by constructing decision trees on a training dataset (Rasaei and Bogaert, 2019). In RF, many decision trees are built, each one using its own bootstrapped subset of the training data (Khanal et al., 2018). The dataset is randomly partitioned into similar subsets using the bootstrap sample algorithm. Each tree is grown and trained using a random subset of the data, while the remaining samples are utilised for validation and to quantify the model's accuracy (Rasaei and Bogaert, 2019; Were et al., 2015).

2.5.3 Current literature on prediction of heavy metals

A qualitative and systematic approach was used to identify the current literature on the estimation or prediction of heavy metals in surface water bodies. The procedures below were performed to retrieve the needed literature. First, a comprehensive literature review for research and review papers is conducted in the Scopus database using different keywords. After eliminating duplicates and doing an initial screening, documents are selected. The following keywords were entered: **TITLE-ABS-KEY** ("heavy metals" AND ("regression" OR "MLR" OR "ANN" OR "RF") AND ("surface water" OR "river" OR "sediments")) and a total of 26 relevant articles were found after a thorough screening out 533 articles.

Rate et al. (2000) measured total Cu, Pb, and Cd contents and analytically defined fractions in near-shore sediments of the upper Swan River estuary. No single parameter could explain the variance in metal concentrations for all metals based on linear regressions between sediment metal concentrations and physicochemical features of the sediments (pH, organic carbon, and particle size distribution).

Chou et al. (2004) devised a method for modelling sediment chemistry changes associated with the deposition of aquaculture waste. Al, Cu, Fe, Li, Mn, and Zn; organic carbon and <63 µm particles were utilised to evaluate the extent of noticeable effects in the vicinity of the cage. This

study revealed significant differences in sediment chemistry between aquaculture sites and the natural background: (1) negative correlations between sediment Cu and Zn with Al; (2) poor correlations between metals and Li; and (3) decreased concentrations of Fe and Mn with increased accumulation of organic carbon. Metals, organic carbon, and particles were subjected to principal components analysis (PCA) to establish the projected or modified environmental monitoring program (EMP) ratings based on their similarities. Utilising the EMP rating based on sediment chemistry, this method gives a regression model with R^2=0.945 compared to R^2=0.653 for the model using uncorrected EMP for evaluating environmental conditions.

Roach (2005) assessed the potential environmental risk by analysing metal concentrations in surface sediments from Lake Macquarie using enrichment factor and sediment quality guidelines (SQGs). Cd, Pb, Hg, Se, Ag, and Zn were enriched in the surface sediments of the entire lake out of the 12 analysed metals. Comparisons using SQGs and effects range median quotients revealed that sediments from a location in Cockle Bay had metal concentrations with the greatest likelihood of generating deleterious effects on sediment-associated biota and that this likelihood declined with distance from Cockle Bay. Compared to previous sediment quality data, surface metal concentrations in the lake have decreased significantly during the past 15 years. Due to low background concentrations, it was unable to build models for all metals. Simple linear regression models were appropriate for the majority of metals; however, for Se and As, a multiple regression model produced a more accurate estimate of background values.

Buszewski and Kowalkowski (2006) investigated the chemometric treatment of data generated from an ex-situ column leaching experiment and described the development of a model based on an artificial neural network (ANN) topology. The column leaching experiments have been used to find dependencies between different physicochemical parameters of soil and heavy metals concentrations in leachate. Investigations of three different soil types are sufficient to create the local model of heavy metal transport within the soil profile. The application of ANNs seems to be the future tool for modelling the transport of inorganic substances in real soil profiles. However, ANNs are not able to extrapolate results over the experimental range. The adaptation of new datasets obtained for different soil profiles can improve modelling precision and provide the possibility of modelling more complicated soil compositions, like organic ones. The proposed structure of Generalized regression neural networks (GRNNs) is complicated and can be adopted only as the "black box": where data is processed properly. The output values agree with the experimental data.

Parizanganeh et al. (2007) examined metal concentrations and their spatial variation in nearshore sediments throughout the Iranian Caspian Sea coast. 14 sample locations were selected along the shore, and 400 g of surficial sediments were collected. All acquired data were subjected to statistical techniques. Linear regression analysis revealed that sediment particle size did not play

a significant role in determining heavy metal contents and spatial variations. The Box and Whisker plots revealed that metal concentrations were not uniformly distributed.

DeLaune et al. (2008) measured background metal concentrations in 220 surface soil samples collected in southwest coastal Louisiana. This dataset calculates regression relationships between Al, Fe, and different metals. At the 1% significance level, Al content in sediments was favourably linked with Cu (r = 0.577), Pb (r = 0.936), Cr (r = 0.969), Ni (r = 0.830), Cd (r = 0.617), and Zn (r = 0.506), but only at the 5% significance level with Mn (r = 0.148). At the 1% significance level, Fe content in the sediments was positively linked with Cu (r = 0.586), Pb (r = 0.847), Cr (r = 0.875), Ni (r = 0.932), Cd (r = 0.803), Zn (r = 0.551), and Mn (r = 0.479). These correlations were used to assess metal pollution at various sites. Data from two known polluted sites, Capitol Lake (Baton Rouge, LA) and Bayou Trepagnier (LA) were far greater than the prediction values established using the Chenier Plain Al and Fe metal regression lines for Cr, Cd, Pb and Zn. Thus, metal/Al and metal/Fe regression correlations can be used to identify regions of probable metal pollution in the coastal zone, but they must be regionally correlated.

Nouh and Al-Noman (2009) developed regression models for predicting mean concentrations of selected heavy metals (Cu, Pb, Ni, Zn, and Fe) in stormwater runoff using data from five residential urban desert catchments. Two groups of equations were used to forecast the concentrations of the specified metals. The first set of equations links suspended sediment concentrations to stormwater parameters, whereas the second set relates suspended sediment concentrations to heavy metals in stormwater runoff. The results of the predictions encouraged recommendations on the use of the equations in the investigated catchments and identified the relative importance of the stormwater runoff and dust storms on the accumulation and transportation of heavy metals in the stormwater runoff. Based on the findings, recommendations for water quality regulation in arid areas are proposed.

Kazemi and Hosseini (2011) assessed the spatial variation patterns of six heavy metals in Caspian Sea sediments: As, Cd, Cu, Hg, Pb, and Zn. For spatial variability modelling, ordinary kriging (OK), genetic algorithm based on artificial neural network (GA-ANN), adaptive network fuzzy inference system (ANFIS), and conditional simulation (CS) have been employed. The results show that the CS realisations produce interpolation values, indicating that the parsimony criterion cannot be followed. The simulated maximum and minimum values are less and greater than the corresponding observed values, respectively. The OK realisation smoothed out spatial variability and extreme measured values between the observed minimum and maximum values for all contaminants. The GA-ANN model is also capable of simulating minimum contaminant values. Besides Cd and Hg, ANFIS, GA-ANN, and OK can also mimic average contaminant values. Finally, when comparing the four interpolation approaches, the GA-ANN model performs best in terms of retaining the statistical properties of the observed data for all pollutants; nevertheless, the ANFIS model performs best in terms of simulation errors.

Rooki et al. (2011) predicted heavy metal (Cu, Fe, Mn, Zn) concentrations from acid mine drainage (AMD) in Shur River, southeast Iran, using backpropagation neural network (BPNN), general regression neural network (GRNN), and multiple linear regression (MLR), while accounting for pH, SO_4^{2-}, and Mg concentrations in the AMD. The findings suggest that the ANN might be a promising tool for promptly and cost-effectively predicting heavy metals in the AMD.

Sadeghi et al. (2012) investigated the feasibility of estimating suspended sediment concentration (SSC) utilising heavy metal concentration as predictor variables (Fe, Cr, Zn, and Ni). From November 2007 to June 2008, suspended sediment samples were collected twice a week from the left bank of the Kojour River during runoff periods. The bivariate and multivariate regression models correlated the SSC and particle size distribution (PSD) with HMC. The results revealed a strong association between dissolved and particulate Cr levels and efficiency coefficients greater than 77% ($P < 0.001$). However, the link between SSC and nickel content was weak. These findings demonstrate that SSC can effectively estimate the HMC in watersheds with varying degrees of accuracy and vice versa. It is also believed that regulating SSC can readily regulate heavy metal pollution.

Su et al. (2013) gathered data from 41 monitoring sites between 1996 and 2003 and pre-treated it for seven variables, including petroleum, Cr (VI), Cd, Pb, Hg, CN^-, and volatile phenol. The outcomes demonstrated that primary predictors and the predictive power of spatial regression varied with respect to variables and scales. Topology, river source distance, land use/land cover (LULC), population density, and gross domestic product (GDP) can be used to predict hazardous chemical factors between 1996 and 2003. Between 1996 and 2003, LULC types accurately predicted increases in cyanide and heavy metals, whereas GDP and population density contributed to petroleum dynamics. This research indicated that spatial regression is a potential method for developing indicators to combat toxic chemical contamination. The study suggests utilising multiscale techniques to determine the dynamics of toxic substances.

Lai et al. (2013) investigated the effect of sediment particle size and land use on the variation of Cr, Ni, Zn, Pb, Cd, Cu, As, and Hg in sediments collected from 19 sub-basins of the Han River Basin in Korea. Except for Zn and Al, all examined metals exhibited significant associations with the fine silt fraction of the sediments, indicating that grain size distribution may contribute to metal enrichment. Only Cd, Cu, Hg, and Pb revealed a close association between metal concentrations and a percentage of urban area in the sub-basin. There was no correlation between agricultural land usage and metal enrichment. The amounts of Cd, Cu, Hg, and Pb were accurately estimated by multiple regression analyses based on fine silt fraction and urban area % ($r > 0.70$, $p < 0.005$).

Fulton and Meyer (2014) provide a regression model to generate a water effect ratio (WER) based on site-specific criteria for a network of ephemeral and intermittent streams in the southwestern United States. A multiple-regression model developed from DOC and alkalinity explained

85% of the toxicity variance in site samples collected, offering a good prediction tool that may be included in the WER framework when site-specific criteria levels are determined.

Abdallaoui and Badaoui (2015) compared the predictive ability of MLR and ANN for the estimation of four heavy metals (Cd, Cr, Cu, and Pb) concentration from eight physicochemical variables (organic matter, moisture content, fine Fraction, pH, $CaCO_3$, carbon and phosphorus in the sediment, and suspended solids in the water column) in the Beht River Basin in Morocco. Using the MLR approach, the determination coefficients varied from $R^2 = 0.26$ for Cd to 0.83 for Cr, with intermediate values for Cu ($R^2 = 0.55$) and Pb ($R^2 = 0.67$). The determination coefficients (R^2) for Cd of 0.88, Cr of 0.93, Cu of 0.96, and Pb of 0.80 was obtained using the back-propagation technique of ANN. These outcomes of ANN were much superior to those obtained using the MLR approach, indicating the superior predictive ability of ANN.

Ipeaiyeda and Onianwa (2016) developed a regression model for each metal (Ni, Zn, Cr, Co, Cu, Cd and Pb) and investigated expressing the metal concentration as a function of distance from the discharge point. The model related the observed metal concentrations with predicted metal concentrations and described the dispersion pattern of the metals in the Olosun river sediment with R^2 varying from 0.9360 (Cu) to 0.9999 (Cr). The proposed regression equations find relevance in the quest for sediment quality planning and pollutant prediction.

Lu et al. (2019) simulated dissolved, particulate, and total heavy metal concentrations in the Taihu Lake region, China, utilising artificial neural network (ANN) and support vector machine (SVM) models with water temperature, pH, suspended solids, turbidity, and total nitrogen, nitrate nitrogen, ammonia nitrogen, total phosphorous, orthophosphate, and permanganate. With the Nash-Sutcliffe efficiency coefficients >0.8, ANN and SVM provided rapid simulation models which successfully simulated particulate heavy metal concentrations. When modelling dissolved and total heavy metal concentrations, models fared poorly. In addition, sensitivity analysis aided in identifying significant factors influencing the behaviour of heavy metals and enhancing environmental monitoring campaigns and management plans.

Bouragba et al. (2020) developed a K_d model that considers four physicochemical parameters of stream water to simulate heavy metal concentrations in the sediment of heavily polluted urban streams. The amounts of lead (Pb) in the sediment of the Harrach River, Algeria, were simulated using a one-dimensional distributed hydrological model that included the K_d model. A multivariable equation of the K_d model with physicochemical parameters (pH, suspended solid concentration (SS), chemical oxygen demand (COD), and biological oxygen demand (BOD)) was derived through multivariate regression analysis of observational data in diverse environmental situations. The Kd model was compared to a constant Kd value in hydrological simulations. The numerical findings were more comparable with the K_d model than with constant K_d, with R^2 increasing from 0.05 to 0.67 for the K_d model.

Stoichev et al. (2020) assessed metal contamination in the sediment samples of a coastal lagoon by employing a novel multiple regression methodology. The metal concentrations were expressed as a function of the geochemical properties of the sediments (fine fraction, organic carbon, Ca, Al, and Mn) and the distance between sample stations. For Li, Co, Ni, Ba, V, and Cr, the influence of distance on concentrations was insignificant, and only geochemical characteristics particular to each element characterized its spatial variation. As, Cu, Zn, and Pb concentrations were affected by geochemical and spatial distance factors, with the latter signifying anthropogenic effect and the extent of transfer of pollutants away from the source.

Ozel et al. (2020) investigated Cu, Fe, Zn, Mn, Ni, and Pb and used radial basis neural network (RBANN), multilayer perceptron (MLP) neural network, and adaptive neuro-fuzzy inference system (ANFIS) to model the results using temperature, pH, EC, COD, BOD, and SS as independent variables. During the testing phase, the RMSE and MAE values of all heavy metal models were determined to have very low error levels. During the testing phase, the models generated with MLP had R^2 values greater than 0.77. The R^2 values of the models utilising the RBN procedure in the test phase varied from 0.773 to 0.989. The ANFIS model's test phase R^2 value was more than 0.80. RBN was successful for Cu, Zn, and Mn when the MAE and RMSE values were ranked from best to worst based on the test assessment findings. In contrast, the MLP model was successful for Ni, while the ANFIS model was successful for Fe and Pb. As a result, it can be concluded that the heavy metal content may be roughly predicted using AI in a timely and cost-effective manner.

Bhagat et al. (2021) developed AI models for sediment Pb prediction in two Australian Bays (Bramble (BB) and Deception (DB) stations). A feature selection (FS) technique termed extreme gradient boosting (XGBoost) is utilised and verified against principal component analysis (PCA), recursive feature elimination (RFE), and the genetic algorithm (GA) to abstract the associated input parameters for the Pb prediction. For predicting Pb, the XGBoost model is deployed using a grid search technique (Grid-XGBoost) and assessed against the generally used AI models, artificial neural network (ANN) and support vector machine (SVM). The input parameter selection procedures re-dimensioned the 21 into 9-5 parameters without losing their learnt information during the training phase of the models. The mean absolute percentage error (MAPE) values (0.06, 0.32, 0.34, and 0.33) for the XGBoost-SVM, XGBoost-ANN, XGBoost-Grid-XGBoost, and Grid-XGBoost models were obtained at the BB station, respectively. The XGBoost-Grid-XGBoost and Grid-XGBoost models achieved the lowest MAPE values at the DB station, 0.25 and 0.24, respectively.

Liao et al. (2021) utilized stepwise multiple linear regression to predict heavy metal migration. The bioavailability of heavy metals recovered from sandy sediments by diethylene triamine pentaacetic acid was substantially lower compared to non-sandy sediments. Stepwise multiple linear regression research revealed that heavy metal prediction equations comprise of multiple

physicochemical elements. With R^2 values ranging from 0.82 to 0.97, all predicted and tested values were of the same order of magnitude.

Zhang et al. (2021) predicted the spatial distributions of heavy metals (HMs) in reclaimed coastal lands. A total of 241 surface (0-20 cm) soil and sediment samples were collected from an eastern Chinese reclamation zone. The random forest (RF) model accurately forecasted HM distributions based on soil parameters. When independent variables were considered, the RF model demonstrated a slightly decreased but acceptable ability in HMs prediction. The temporal rise and close association between soil Cd and phosphorus in reclaimed soils indicated the possible risks of Cd pollution in coastal reclaimed soils.

Gad et al. (2021) evaluated the surface water quality of Qaroun Lake using the weighted arithmetic water quality index (WAWQI), heavy metal pollution index (HPI), metal index (MI), contamination index (Cd), and pollution index (PI) and multivariate analyses such as cluster analysis (CA), principal component analysis (PCA), and support vector machine regression (SVMR). According to the findings of the WQIs, the lake's surface water is unfit for use because it is adversely impacted by pollution. The findings of the PI showed that surface water samples taken from Qaroun Lake were significantly impacted by Al, moderately impacted by Cd and Cu, and only slightly impacted by Zn as a result of uncontrolled releases of domestic and industrial wastewater. As an alternative approach to predicting the WQIs, the SVMR models based on physiochemical factors demonstrated the highest overall performance.

Farahat et al. (2021) developed new prediction models for evaluating the probable uptake of heavy metals (HMs) by the invasive grass Vossia cuspidata. These models contain sediment parameters such as pH, organic matter, and silt and clay concentrations. In Nile islands, samples of plants and sediment were taken from the microsites that best represent the species' natural distribution. According to the findings, the root was the primary organ responsible for accumulating the tested HMs (Fe, Mn, Zn, Cu, Ni, and Pb). No significant discrepancies were found between HM concentrations measured and predicted in any of the organs of the species. This demonstrates that the regression models that were constructed have strong robustness.

Mokarram et al. (2022) collected water samples from 25 different locations and measured 21 different water quality parameters across the industrial centres of the Kor River basin. The interpolation maps of each parameter were computed using the Kriging method, and the water quality was evaluated using the Water Quality Index (WQI) method. WQI values ranged from 28 to 73, indicating a higher concentration of pollution in the areas downstream of the industries than upstream. The findings of the principal component analysis (PCA) suggested that the biological oxygen demand (BOD), chemical oxygen demand (COD), nitrous oxide concentration (N_2O), and coliform count were the most critical variables among the 21 parameters that affect water quality. The findings of linear regression revealed that the parameters with the highest R^2 values

for determining coliform levels and WQI values, respectively, were biological oxygen demand (BOD) and levels of Cd, Pb, Hg, and Zn.

Pan et al. (2022) investigated the spatial and temporal variation in the water quality of the Yellow River's main channel during the spring and fall of 2019. Water Quality Index (WQI) was derived from 15 different water parameters measured at 44 sampling stations in 26 river stretches and six reservoirs. The amount of pollution caused by eight heavy metals was measured using a heavy metal evaluation index (HEI). The WQI values dropped from the Yellow River's source region to its estuary, and greater values were found in the majority of reservoirs in comparison to the nearby natural river sections. Stepwise regression was used to generate eight different WQI models, including four that were weighted and four that were unweighted. The best modelling performance was achieved by a model that included six parameters: total suspended solids, ammonia-nitrogen, permanganate index, electrical conductivity, and dissolved oxygen. The second highest modelling performance was achieved by a model that included nitrate-nitrogen. After introducing parameter weights, the WQI model improved both the accuracy of its predictions and the degree of fit.

2.6 Knowledge Gap

The physical, chemical and biological characteristics of the water system determine the state of the water, which is referred to as the water quality. Water has the unique ability to dissolve and carry a wide range of substances inside itself; as a result, it is susceptible to becoming contaminated even though it plays an essential role in maintaining life. Because of rapid population increase, urbanisation, and industrialisation, the aquatic ecosystem has been subjected to enormous strain, resulting in a decline in the water quality and its biodiversity. The contamination of surface water has developed into a problem that affects the ecosystem on a global scale. Continuous monitoring is of the utmost importance to keep the water quality at a particular level.

It is important to note that sediments are the storage units of aquatic ecosystems. To gauge the human effect on our environment, sediment quality is a good indicator. More often than not, sediments are the primary source of information about long-term changes. In lakes, rivers, and coastal waters, sediments are believed to be the primary source of toxicity. It is more likely that large contaminated sediments will be found near where they came from, so they may be found in more active waters like rivers and creeks. The deposition of small sediments happens when the water is relatively still, so it is most likely in lakes, reservoirs, estuaries, bays, and harbours. If sediments are not disturbed, pollutants can accumulate in the pore fluids around them, but this is less common. The environmental conditions influence the properties of sediment grains in the basin, and they are further modified by the different natural processes of digenesis that transform loose sand and gravel into hard rock.

Different WQIs and pollution indices have been widely utilised to express the status of contamination in the surface water body and the benthic sediments. Nevertheless, the application of

information entropy and fuzzy logic, which at present is widely acknowledged as a better option for understanding the random processes occurring inside a river basin, is found to be rare. In the case of benthic sediments, the pollution indices developed and utilised to express the contamination level or ecological risk level are mostly based on the total heavy metal concentrations. The availability of heavy metals in benthic sediments is directly linked to metal speciation that divides the metal contamination in sediments into 5 fractions. Various literatures have inferred from their studies that pollution indices based on total metal concentration produce a misleading level of risk. The application of pollution indices based on metal speciation or bioavailable fractions is very limited.

Source apportionment of pollution in surface water and benthic sediments is primarily done using PCA. Nevertheless, the literature suggests that combining PCA and PMF would give a better understanding of the pollution sources in both surface water and benthic sediments. This type of study is also limited worldwide and especially in India in the case of surface water.

The most problematic contamination in most surface water bodies is due to heavy metals. Risk assessment of heavy metals is carried out through health risk assessment studies. Health risk assessment assesses cancer and non-cancer risk of prolonged exposure to heavy metals via two different approaches, deterministic and probabilistic approaches. These studies are very important in utilising surface water for various purposes, but literature on such studies is very limited in India and non-existent in non-eastern India.

Heavy metal pollution is a critical source of contamination in surface water bodies. Heavy metals assessment necessitates using extremely expensive and accurate tools, which are not always available at all study locations or research institutes. Developing a metal assessment model is required to estimate heavy metal concentrations using basic and easily measured physicochemical characteristics with sufficient accuracy. Such studies are limited in India, as well as in the rest of the globe. In addition, the relationship between metal pollution in the water column and sediment column based on metal speciation has to be investigated. These two models will aid in estimating metal pollution and also formulating remediation plans for a surface water body.

Based on the research gaps from the literature review, the objectives of the present study were decided.

Chapter 3

MATERIALS AND METHODS

This chapter details the experimental and analytical approaches utilised to complete the research objectives formulated in chapter 3. The research work was carried out in different phases, and the detailed methodology is given in this chapter.

3.1 Design of Research

In order to fully understand the limnology aspect of the river ecosystem, the research work demanded a systematic investigation and, therefore, needed an experimental design, as depicted in 3. 1. To complete the research work systematically, the study was divided into 5 objectives.

In Objective I, a detailed reconnaissance survey was carried out using LULC maps and field visits to determine the approachable and important sampling sites. After selecting the sampling sites, benthic sediment and surface water samples were collected and analysed over two years (from May' 2018 to October' 2019) to understand spatiotemporal variation. Also, surface water samples were characterised for their possible locations as potable water using information entropy and fuzzy logic.

In Objective II, benthic sediment samples were then analysed for their contamination status using two techniques. Firstly, it was analysed with respect to the total metal content and, secondly, with respect to the bioavailable fractions of the metal concentration in the sediments.

In Objective III, pollution source apportionment was done using different statistical techniques or environmetrics tools for surface water and benthic sediments. The methodologies adopted were Discriminant analysis, Principal component analysis, Positive matrix factorization, Pearson correlation matrix and Cluster analysis.

In Objective IV, the health risk associated with the river for the inhabitants was evaluated for exposure to heavy metals for a prolonged duration in the form of carcinogenic and non-carcinogenic health effects.

Lastly, in Objective V, two models were formulated to assess heavy metal contamination in surface water. The first was formulated with respect to the physicochemical properties of surface water, and the second was formulated with respect to metal speciation fractions in the sediments. These two models will help to assess the heavy metal contamination in the surface water and also help formulate remediation strategies.

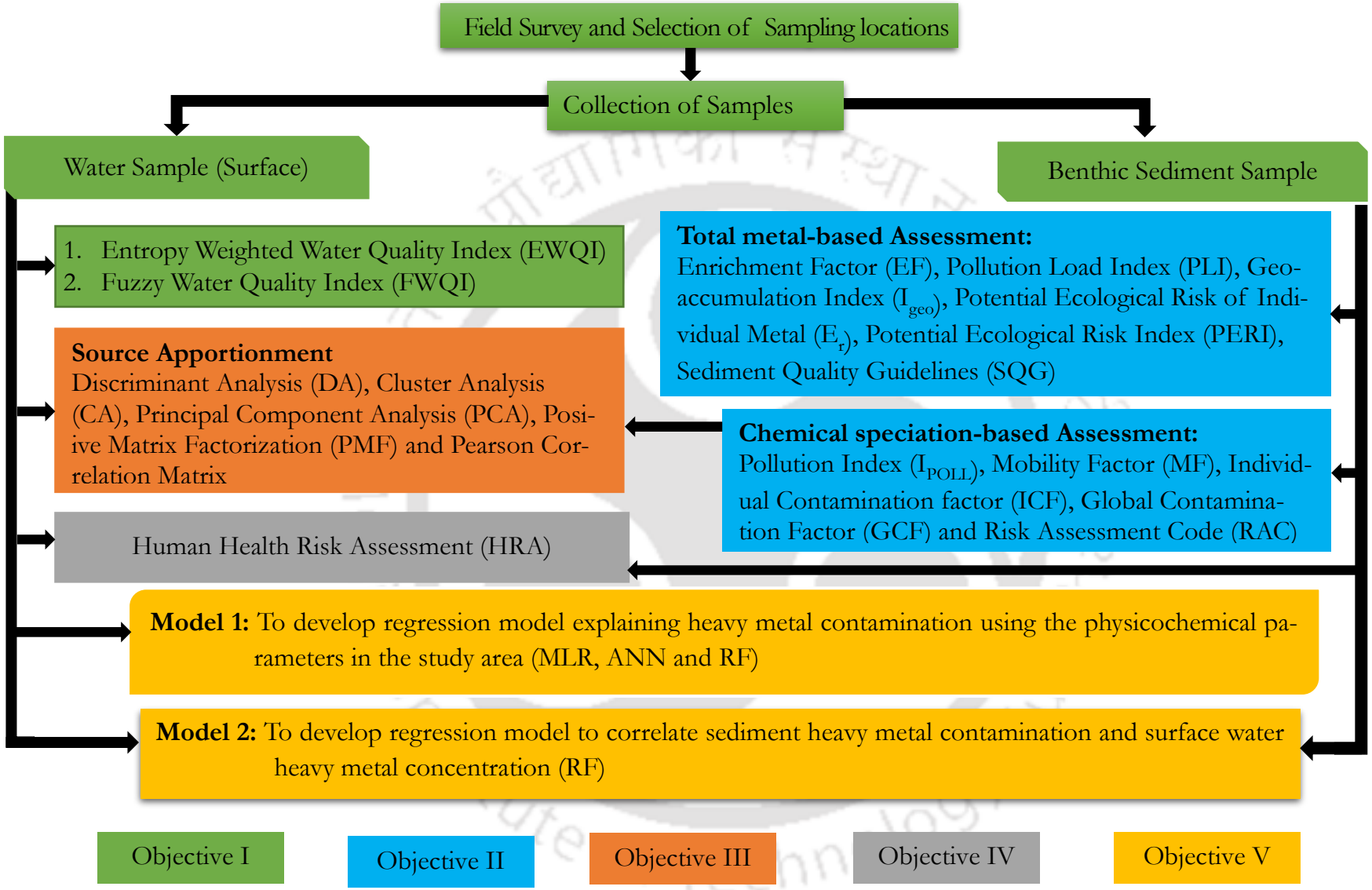


Fig. 3.1. Design of research

3.2 General Description of Study Area

The Kolong River is a tributary of the Brahmaputra, which originated as a spill channel or anabranch (PCBA, 2013). The river has its source in the Kukurakata and Hatimura hills. The river's total length is roughly 218.62 kilometres, and it passes through the metropolises of Nagaon, Morigaon, and Kamrup. The river has an average depth of 1-2m with a maximum height of 4-6m during the highest rainfall period (Bora and Goswami, 2017). The Kolong River combines with the Kopili River, a vast southern river of the Brahmaputra near Jagibhakatgaon, Morigaon district, before returning to its source, the Brahmaputra River near Guwahati. Nagaon District is home to the Kolong River, which flows for more than 100 kilometres. The river divides Nagaon township into two east-west halves, Haibargaon and Nagaon, in the middle of the township. The Kolong River receives water from several tributaries, including the Diju, Misha, Diphalu, Haria-Nanoi, and Titaimari or Rahasuti. After receiving water from the aforementioned rivulets, the Kolong River swells. The Kopili River joins the Kolong River course in the Morigaon district at Dukhutimukh in Jagibhakatgaon town (Singh et al., 2020; Bora and Goswami, 2015). Kolong River was formerly the only source of potable water, and settlements built up along its banks. Nagaon's portion of the Kolong River is currently degraded, posing a health and hygiene risk to the community. The Kolong River in Assam Nagaon's district illustrates how human activities have exacerbated environmental issues over the previous half-century. The Kolong River turned into a succession of intermittent dry stretches and sluggish pools due to this massive human intervention over the years that followed. Because of its limited capacity for self-purification, the river is currently under a considerable degree of anthropogenic stress and acts as a major recipient of pollutants. As a result, the Central Pollution Control Board lists the Kolong River as one of India's most contaminated rivers. The Kolong River's quality can only be improved over the long run with a comprehensive restoration scheme (Bora and Goswami, 2014; CPCB, 2015). Industry does not dominate in the Nagaon district, which still depends mostly on agriculture. Existing industries which exist in the region and contribute to the pollution load of the river include the following:

- Tea industry (largest industry)
- Handloom and handicrafts industries
- Forests and wood industry
- Assam Cooperative Jute Mill Ltd
- Kampur Cooperative Sugar Mill
- Katimari Weaving Project
- Sack Craft paper project

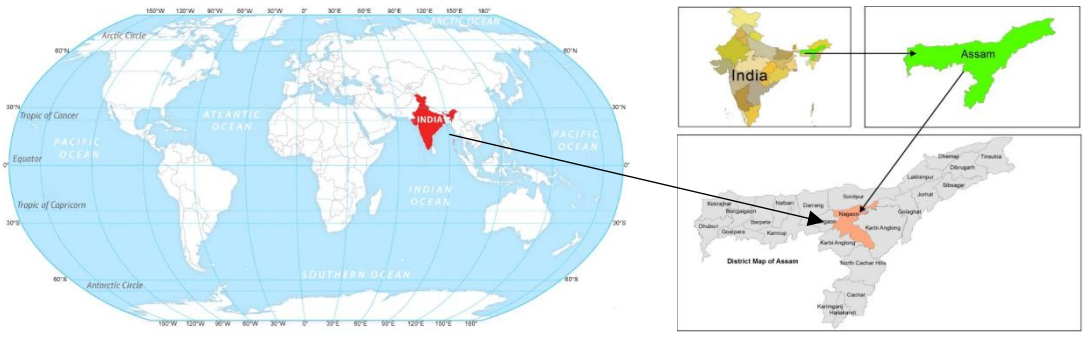


Fig. 3.2. Map of Assam

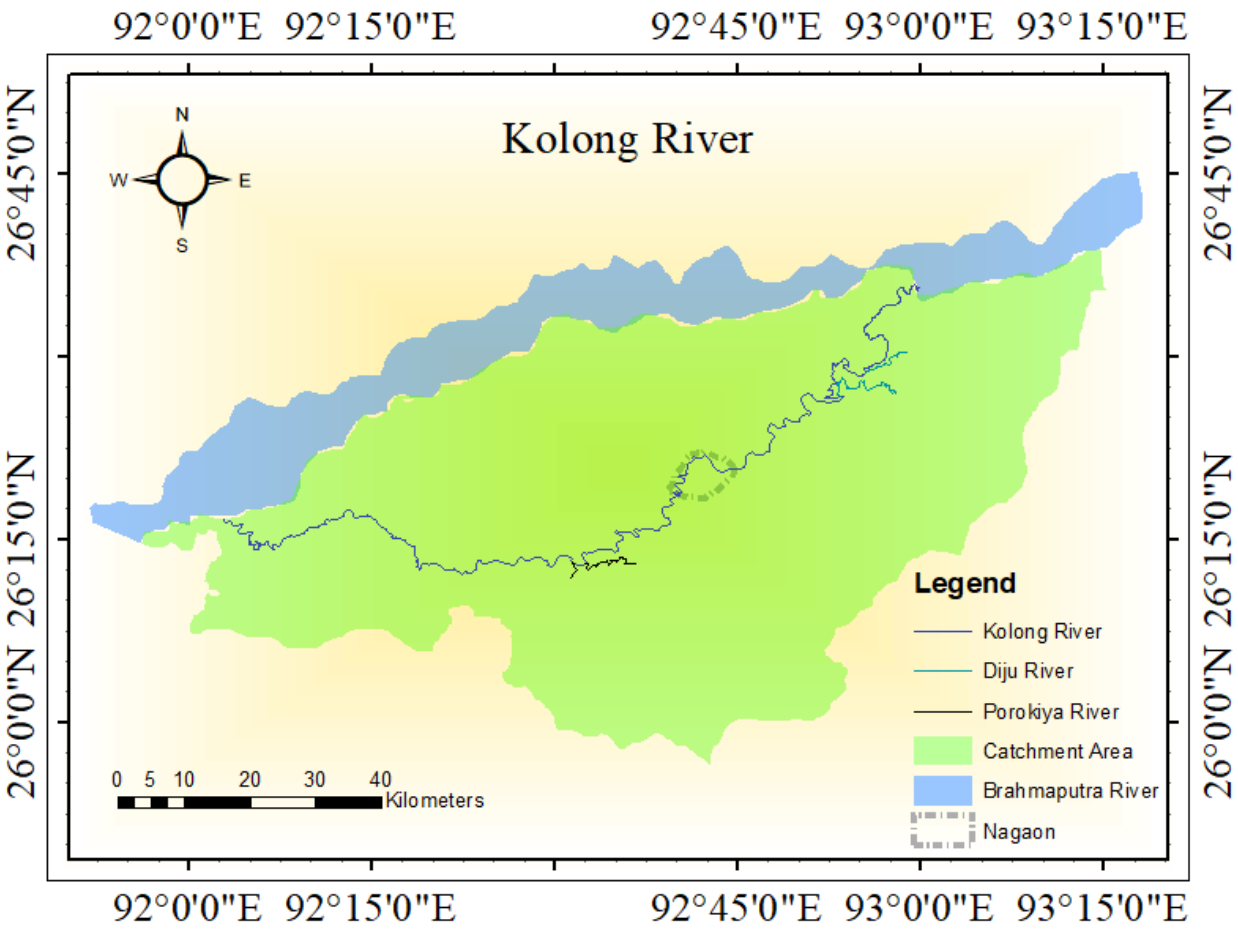


Fig. 3.3. Study Area showing Kolong river

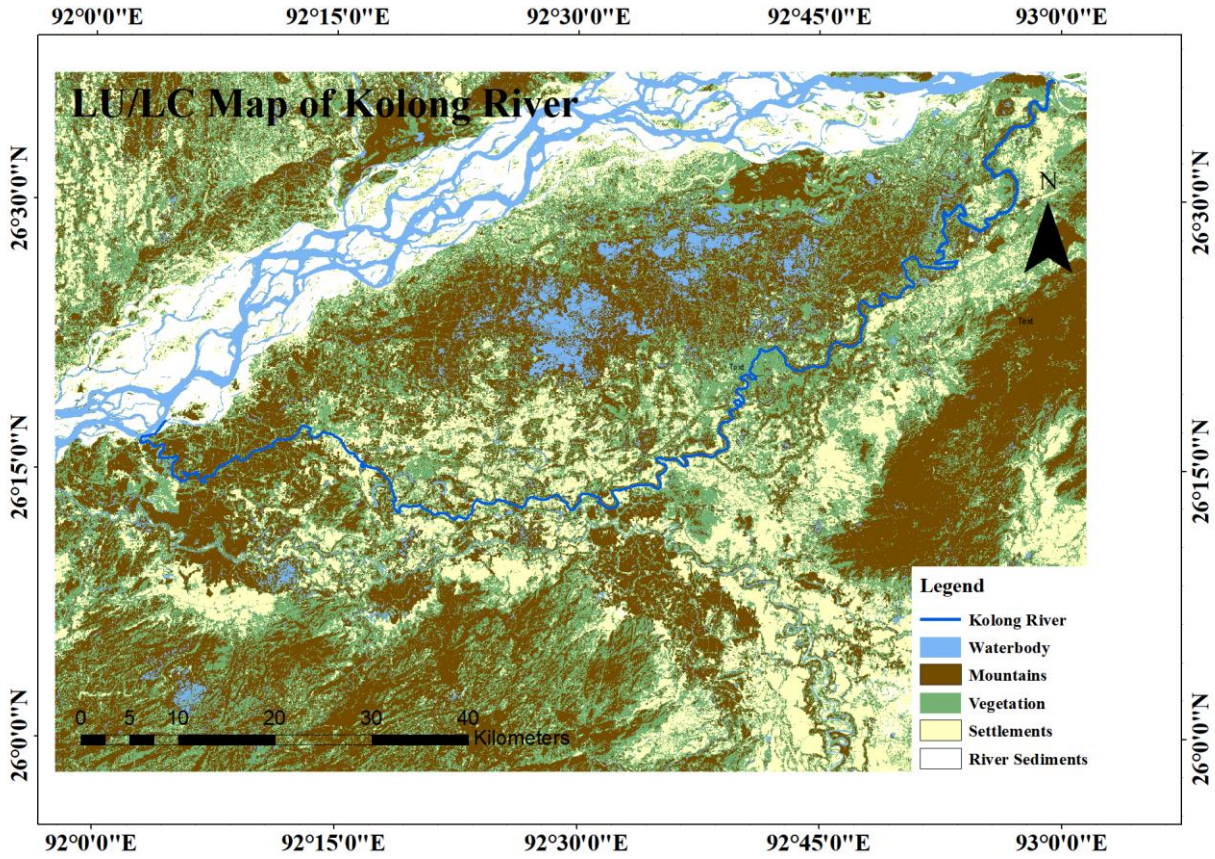


Fig. 3.4. LULC Map around Kolong river

3.3 Objective I (Part I): Sampling Strategy and Data Acquisition

In part I of objective I, the main focus of work will be to collect samples in the form of water samples and benthic sediment samples from the river for 12 months, from May' 2018 to October' 2019, to characterise the water quality of the river as well as some natural contributing factors of pollution. To understand the variation in data, the 12 months of data are grouped into four seasons pre-monsoon, monsoon, post-monsoon and winter. Water samples were analysed for physicochemical parameters and heavy metals, whereas benthic sediment samples were analysed only for their heavy metal contamination. After the reconnaissance survey, 18 surface water samples were collected throughout the flow of the river from its origin to its fallout. However, only 9 sediment sampling sites (almost every alternate surface water sampling location) were found suitable and accessible for collecting sediment samples based on the location of the probable pollution sources. More sampling locations were centred around Nagaon town, a densely populated area located on the flow path of the river. The sampling sites for collecting surface water and benthic sediment samples are shown in Fig. 3.5 and Fig. 3.6, respectively, and their GPS locations are shown in Table 3.1.

Table 3.1. Sampling sites

Surface Water	Latitude	Longitude	Benthic Sediments	Latitude	Longitude
SSKR1	N 92° 31' 21.60"	E 26° 13' 22.72"	**SS1**	N 92° 31' 21.60"	E 26° 13' 22.72"
SSKR2	N 92° 34' 55.88"	E 26° 15' 41.54"	**SS2**	N 92° 38' 48.14"	E 26° 17' 26.14"
SSKR3	N 92° 38' 48.14"	E 26° 17' 26.14"	**SS3**	N 92° 40' 41.18"	E 26° 20' 29.49"
SSKR4	N 92° 40' 18.48"	E 26° 18' 51.55"	**SS4**	N 92° 41' 12.90"	E 26° 21' 26.04"
SSKR5	N 92° 40' 41.18"	E 26° 20' 29.49"	**SS5**	N 92° 42' 53.13"	E 26° 21' 6.28"
SSKR6	N 92° 40' 49.52"	E 26° 20' 56.71"	**SS6**	N 92° 45' 37.12"	E 26° 21' 41.64"
SSKR7	N 92° 41' 12.90"	E 26° 21' 26.04"	**SS7**	N 92° 47' 51.67"	E 26° 23' 28.49"
SSKR8	N 92° 42' 17.25"	E 26° 21' 54.73"	**SS8**	N 92° 50' 43.78"	E 26° 25' 34.05"
SSKR9	N 92° 42' 53.13"	E 26° 21' 6.28"	**SS9**	N 92° 53' 17.80"	E 26° 26' 25.53"
SSKR10	N 92° 44' 29.19"	E 26° 20' 48.94"			
SSKR11	N 92° 45' 37.12"	E 26° 21' 41.64"			
SSKR12	N 92° 47' 35.18"	E 26° 22' 27.73"			
SSKR13	N 92° 47' 51.67"	E 26° 23' 28.49"			
SSKR14	N 92° 49' 10.72"	E 26° 24' 43.49"			
SSKR15	N 92° 50' 43.78"	E 26° 25' 34.05"			
SSKR16	N 92° 51' 7.75"	E 26° 26' 34.45"			
SSKR17	N 92° 53' 17.80"	E 26° 26' 25.53"			
SSKR18	N 92° 56' 3.22"	E 26° 29' 19.86"			

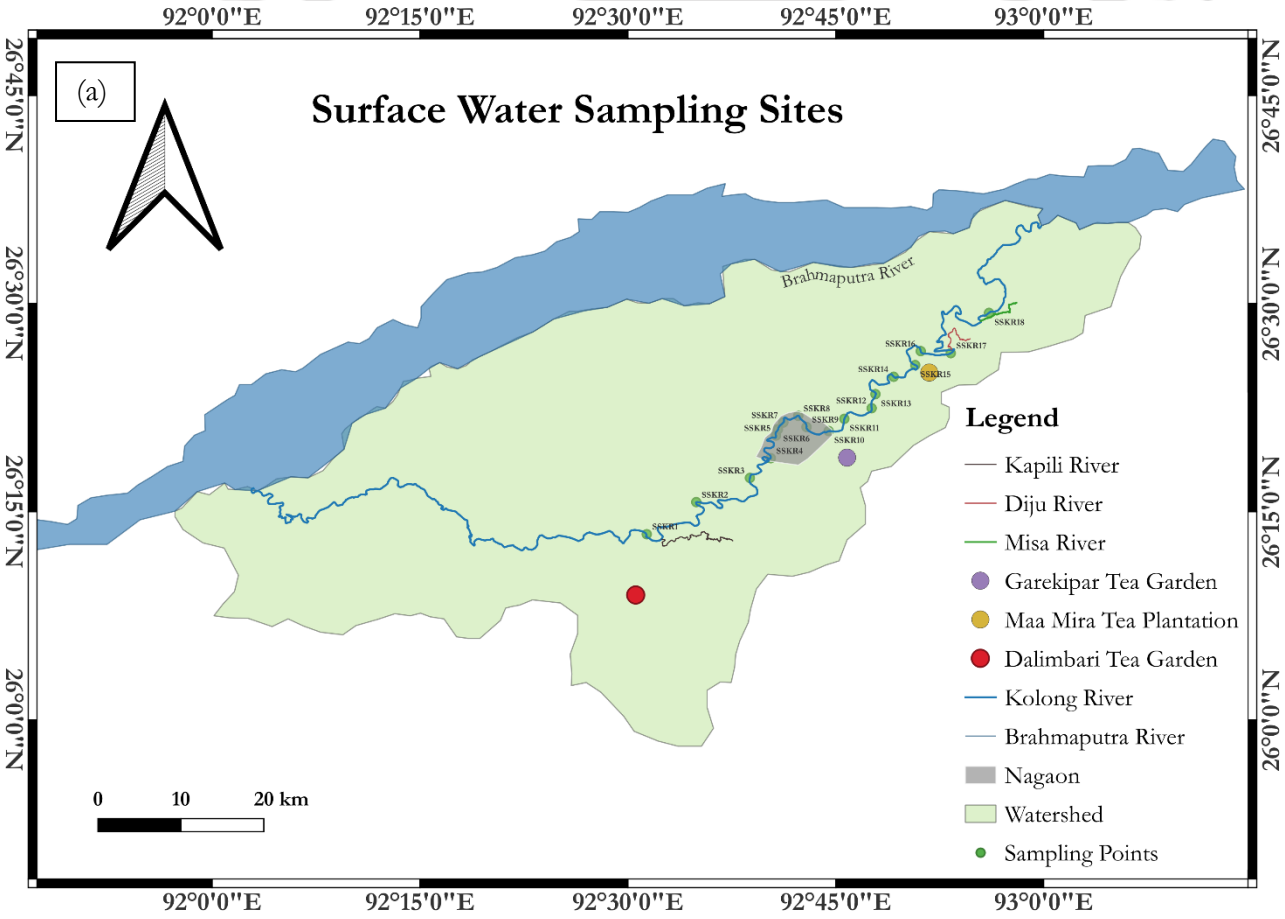
3.3.1 Sample and data collection

3.3.1.1 Surface water samples

The sampling technique for surface water entails collecting information on various parameters affecting water quality. All water samples were collected in triplicate at 0.5 m below the surface of the river. A 1 L sampler was used to collect water samples, which were then placed in polycarbonate bottles for laboratory analysis. Separate 0.5 L samples of water were collected for heavy metal analysis, in which 0.5 mL acid was added to make the pH around 2 in order to avoid precipitation of heavy metals. Before sampling, bottles were rinsed three times with the sample collected.

3.3.1.2 Benthic sediment samples

Sediments are chemically and physically heterogeneous, with pollutant distribution highly dependent on grain size; therefore, care must be taken while selecting sites and designing a sampling programme. Sediments are typically taken from the top 10 cm, where the majority of biological activity occurs, although some organisms can burrow to higher depths. Sediment samples were collected using a grab sampler from the river bed. Collected samples were kept in laboratory grade zipped polythene bags. After collecting the samples, samples were oven-dried and removed the unwanted particles by handpicking, such as small pebbles, stone, wooden and grass particles. Samples were first sun-dried and smashed by hand and later oven dried to remove the water particles from the soil matrices and sieved through a 200µ sieve for further experiment.



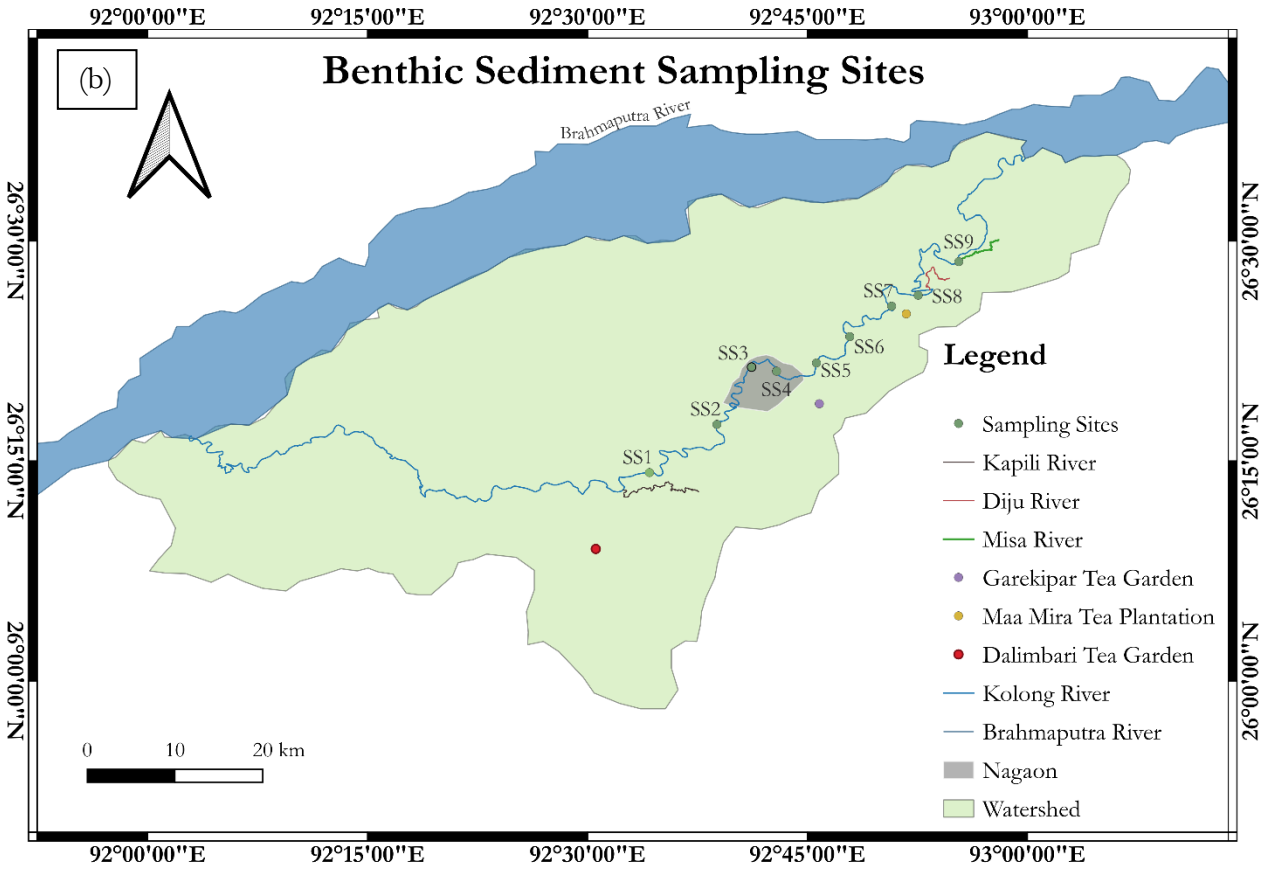


Fig. 3.5. Sampling sites for (a) surface water samples and (b) benthic sediment samples

3.3.2 Analytical procedures

3.3.2.1 Surface water samples

Analytical procedures of Standard Methods for the Examinations of Water and Wastewaters 20th edition, published by APHA (2012), have been followed throughout the analysis. Certain parameters, such as pH, DO, and EC, were analysed at the site. These non-conservative parameters change over time. Afterwards, samples were placed in coolers and kept at 4°C until the next analysis was scheduled to be carried out. A quality control procedure was maintained throughout, including the recalibration of instruments. Reagents were prepared as recommended by APHA Standard Methods (20th edition, 1998). Unless otherwise stated, all chemicals and reagents used in the analyses were of analytical grade. Deionized water was used for all dilutions. Standard solutions were prepared by diluting the stock solutions.

Dissolved oxygen (DO) was measured on-site with the aid of a portable DO meter. pH and electrical conductivity (EC) in mS/cm were measured using a multiparameter pH/ORP /EC/TDS/Salinity/Pressure/Temperature Waterproof Meter (HI98194). Alkalinity and hardness were determined titrimetrically and expressed in mg/L as $CaCO_3$. BOD_5 (mg/L) was determined using modified Winkler's method in 300 ml BOD bottles. Cations, namely sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}) in mg/L, were analysed using Flame photometer. Anions, namely fluoride (F^-), chloride (Cl^-), sulphate (SO_4^{2-}), and nitrate (NO_3^-), were determined using Ion Chromatography. The samples preserved at a pH of 2 were analysed for Iron (Fe), Manganese

(Mn), Chromium (Cr), Lead (Pb), Cadmium (Cd), Copper (Cu) and Zinc (Zn) in mg/L using AAS.

3.3.2.2 Benthic Sediment samples

Analytical procedures were followed in accordance with EPA protocol 3050B acid digestion for sediments, sludges, and soils for total metal digestion with nitric acid and hydrogen peroxide and Tessier sequential extraction procedures for the metal speciation (Tessier et al., 1979; USEPA, 1996). Samples were analysed in AAS after the preparation of the sample. Quality control procedures were followed throughout the experiment by calibrating the instrument every time before operating the instrument using freshly prepared standards (APHA, 2012). All the chemicals used throughout the experiments were laboratory grades, and deionized water was used for all dilutions. Quality control was done as per Standard Methods. The following experimental procedures were undertaken following the Tessier 5-step, sequential extraction techniques to prepare samples for various heavy metal fractions (Tessier et al., 1979):

Step I: 1g of sample (sediment) is treated at pH 7 with 1 M magnesium chloride; stirred with a shaker for 24 h at room temperature; centrifuged @5000 rpm for 10 minutes, and then the suspension is diluted with deionised water and analysed.

Step II: The previous step's residual solid is treated with sodium acetate. After 5 h in the shaker at room temperature, centrifugation results in solid-liquid separation, and the metal-bearing solution is diluted and analysed.

Step III: The residual solids in the first stage are processed in a 6 h water bath shaker at a temperature of 85°C with Hydroxyl ammonium chloride. As in previous phases, there is a solid-liquid separation by centrifugation, and the solution is then diluted and analysed.

Step IV: The remaining solid of the previous step is treated in the hot water bath shaker using nitric acid and hydrogen peroxide and agitated for 2 h at 85°C. After 2 h, 3 mL of H_2O_2 is added and then further heated by adding NH_4OAc and HNO_3 and then diluted to 20 mL solution. Remove the supernatant after agitating it for 30 mints.

Step V: The previous step's residual solid is combined with a 5:1 acid mixture of H_2SO_4 and $HClO_4$ and digested for 2 h at 300°C before being filtered through Whatman 42 filter paper.

3.3.3 List of instruments/methods used and make

Instruments used in carrying out the experiment are listed below with the make and model number, as listed below in Table 3.2.

Table 3.2. List of instruments/methods

Sl. No.	Name	Make	Model	Parameters
1	Atomic Absorption Spectrophotometer (AAS)	THERMO SCI-ENTIFIC	iCE 3000 SE-RIES	Mg^{+2}, Fe, Mn, Cr, Hg, Pb, Cu & Zn in mg/L
2	Multiparameter pH/ORP /EC/TDS/Salinity/DO/Pressure/ Temperature Waterproof Meter	HANNA	HI98194	pH, EC (µS/cm), DO (mg/L)
3	Water bath Shaker	ICT		
4	Laser particle Size Analyser	MALVERN	HYDRO 2000 MU	
5	Flame Photometer 128	SYSTRONICS		Na^{+}, K^{+} & Ca^{+2} in mg/L
6	Ion Chromatograph (IC)	792 Basic IC	METROHM	SO_4^{2-}, NO_3^{-}, Cl^{-}, F^{-} in mg/L
7	Graviometric	Oven Drying at 103-105˚C		TDS in mg/L
8	Titration			Alkalinity & Hardness in mg/L
9	5-day BOD			BOD_5 in mg/L

3.4 Objective I (Part II): Indexing Approach to establish Drinking Water Applicability

Conventional WQIs developed in the last decade and before exhibit several weak points, allowing the assignment of quality values using limited parameters. In addition, most indices exclude toxic compounds like heavy metal pollution. However, the biggest flaw in these indices is that they do not account for the inherent uncertainty and subjectivity of this complex environmental issue. (Chang et al., 2001; Mpimpas et al., 2001; Silvert, 2000).

In this section, two indexing approaches were utilised to solve the above-mentioned complexity related to assessing the drinking water quality of surface water samples. From the literature review conducted in chapter 2, the two assessment tools for this purpose are information entropy and fuzzy logic. The following describes the steps for the calculation of entropy weighted water quality index (EWQI) developed by Singh et al. (2019) and the fuzzy logic-based water quality index (FWQI) developed in this study.

3.4.1 Entropy weighted Water Quality Index (EWQI)

For the calculation of EWQI, weights are allocated to each parameter based on their entropy value. When it comes to the long-term behaviour of random processes, entropy and associated information metrics can provide useful information about it. The calculation steps for EWQI are (Singh et al., 2019):

- Step 1. A matrix was constructed for all 'm' water samples (m = 1, 2...m) and 'n' calculated parameters (n = 1, 2...n)

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & \cdots & x_{mn} \end{bmatrix} \quad (3.1)$$

- Step 2. To eliminate the error caused by various units and sizes, matrix A was transformed into a normalised matrix B. After transformation, the standard grade matrix B was obtained

$$b_{ij} = \frac{x_{mn} - (x_{mn})_{min}}{(x_{mn})_{max} - (x_{mn})_{min}} \quad (3.2)$$

where b_{ij} for the measured parameter (n) in a specific water sample is the construction function of normalisation (m).

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & \cdots & b_{mn} \end{bmatrix} \quad (3.3)$$

where B = standard grade matrix

- Step 3. The following formula calculated the information entropy:

$$P_{ij} = \frac{b_{ij}}{\sum b_{ij}} \quad (3.4)$$

$$E_j = \frac{1}{\ln(m)} \sum_{i=1}^{m} P_{ij} \ln(P_{ij}) \quad (3.5)$$

- Step 3. The parameter (j)'s entropy weight was calculated by:

$$W_n = \frac{(1 - E_j)}{\sum_{j=1}^{n} (1 - E_j)} \quad (3.6)$$

- Step 5. For each parameter quality rating scale was assigned by:

$$Q_j = \left(\frac{C_j}{S_j} \right) \times 100 \quad (3.7)$$

where,

Q_j = quality rating scale,

C_j = measured concentration of the parameter, and

- Step 6. WQI was calculated using the following formula:

$$EWQI = \sum_{j=1}^{n} W_j \times Q_j \quad (3.8)$$

Once the overall EWQI results have been established, it is possible to quantify how safe the water is over a given time according to the scale recommended by Jianhua et al. (2011). Table 3.3 shows the EWQI ranges for various pollution levels.

Table 3.3. Water quality scale for EWQI

EWQI	Water Quality
< 50	Excellent
50-100	Good
100-150	Average
150-200	Poor
>200	Extremely poor

Water sources with excellent or good quality may sustain a high diversity of aquatic life. Water will also be ideal for all modes of recreation, including direct water communication.

3.4.2 Fuzzy logic-based Water Quality Index (FWQI)

The necessity for more appropriate approaches to regulate the water quality variables and interpretation of an acceptable range for each parameter needs integration of different metrics involved in the evaluation process is well acknowledged. In this regard, some alternative approaches have emerged from AI. These approaches, primarily fuzzy logic and fuzzy sets, can be evaluated on real-world environmental problems with the ultimate goal of eliminating the ambiguity and imprecision in the criteria used in decision-making tools (Chang et al., 2001; McKone and Deshpande, 2005).

Fuzzy sets, which are conceptually simple and based on natural language, have been effectively utilised to describe nonlinear functions, develop inference systems on top of expert expertise, and deal with imprecise data (Pham and Pham, 1999; Romano et al., 2004; Ross, 2010; Zadeh, 1999). These benefits have been used to address difficult water-related environmental issues (Ghosh and Mujumdar, 2006; McKone and Deshpande, 2005; Sadiq and Rodriguez, 2004; Vemula et al., 2004). In this study, fuzzy logic was also utilised to analyse water quality by constructing a fuzzy logic based water quality index.

3.4.2.1 Fuzzy Interference System

Fuzzy set theory was developed to model complex systems in uncertain and imprecise contexts (Ross, 2010). Sets with unclear boundaries are used in fuzzy logic. Fuzzy logic can be used to map inputs to appropriate outputs. Figure 3.6 depicts an input-output map for the water quality categorization problem: "Given a comprehensive collection of indicators of water quality, what is the water quality in a stream?". Water quality indicators and stream conditions are imprecise definitions since they lack clearly defined boundaries.

In general, fuzzy inference systems are composed of membership functions, inference rules, and defuzzification. In a fuzzy inference system, a variable or input has a domain known as its universe of discourse, and each set in this space is expressed by a language phrase such as "low," "medium," or "high." If-then rules define the relationships between the variables. Finally, a

defuzzification algorithm is utilised to determine the optimal crisp value. In other words, a fuzzy output variable must be defuzzified into a single value (Ross, 2010).

3.4.2.2 Membership Functions

A membership function is a curve that defines how each point in the input space is assigned a membership value ranging from 0 to 1. The universe of discourse refers to the input space. The output axis is known as the degree of membership μ. If X denotes the universe of discourse and its elements, then a fuzzy set A is defined as a collection of ordered pairs:

$$A = \{x, \mu_A(x) \mid x \in X\} \tag{3.9}$$

where, $\mu_A(x)$ is the membership function of x in A. A membership function is an arbitrary curve whose shape is determined by practical considerations. Membership functions come in various shapes, including triangular, trapezoidal, and Gaussian. Membership functions have been constructed in various ways, such as using the expert's knowledge (Norwich and Turksen, 1984; Turksen, 1991), fuzzy clustering (Bezdek and Pal, 1992; Gustafson and Kessel, 1978; Jang and Sun, 1995; Keller and Hunt, 1986; Peizhuang, 1983; Takagi and Hayashi, 1991) neural networks (de Campos Souza, 2020; Shihabudheen and Pillai, 2018).

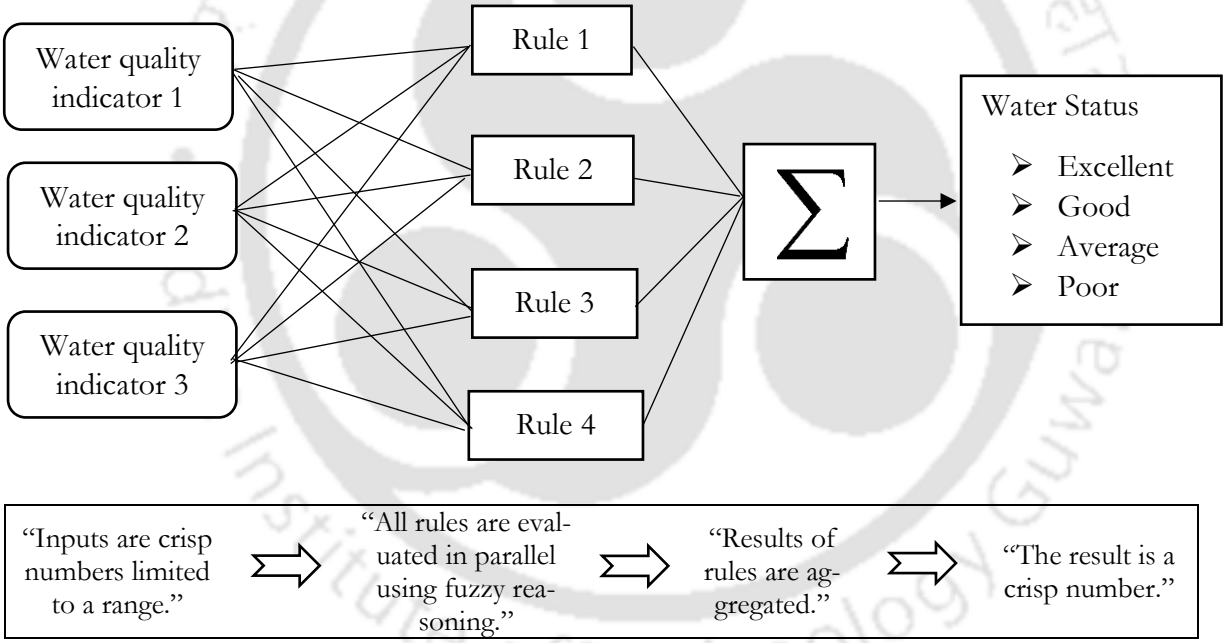


Fig. 3.6. Flow chart for FWQI development

3.4.2.3 Interference Rules

The if-then rules are used to formulate the inference rules. The if-then rule comprises two parts: the if component, known as the antecedent, and the then part, known as the consequent. The output variable with one or more consequents is derived from one or more premise antecedents in the inference rules. The rules are as follows:

1. If (parameter 1 is Low) and (parameter 2 is Low), then (Set is Good)
2. If (parameter 1 is Low) and (parameter 2 is Medium), then (Set is Bad)
3. If (parameter 1 is Low) and (parameter 2 is High), then (Set is Worse)
4. If (parameter 1 is Medium) and (parameter 2 is Low), then (Set is Bad)

5. If (parameter 1 is Medium) and (parameter 2 is Medium), then (Set is Bad)
6. If (parameter 1 is Medium) and (parameter 2 is High), then (Set is Worse)
7. If (parameter 1 is High) and (parameter 2 is Low), then (Set is Worse)
8. If (parameter 1 is High) and (parameter 2 is Medium), then (Set is Worse)
9. If (parameter 1 is High) and (parameter 2 is High), then (Set is Worse)
10. If (parameter 1 is Low), then (Set is Good)
11. If (parameter 1 is Medium), then (Set is Bad)
12. If (parameter 1 is High), then (Set is Worse)
13. If (parameter 2 is Low), then (Set is Good)
14. If (parameter 2 is Medium), then (Set is Bad)
15. If (parameter 2 is High), then (Set is Worse)

The if-then rules might change depending on the type of parameter (e.g. pH, DO).

Fuzzy set operators express the relationships between the fuzzy sets of the inputs. There are three standard fuzzy set operations: intersection (AND), union (OR), and complement (negation or NOT). If A and B are defined as sets in the universe X, then for a given element x belonging to X, the following operations can be carried out:

$$\text{AND: } \mu_{A\cap B}(x) = \mu_A(x) \cap \mu_B(x) = \min \mu_A(x), \mu_B(x) \quad (3.10)$$

This AND operation creates the intersection of sets A and B.

$$\text{OR: } \mu_{A\cap B}(x) = \mu_A(x) \cup \mu_B(x) = \max \mu_A(x), \mu_B(x) \quad (3.11)$$

This OR operation creates the union of the two sets.

$$\text{NOT: } \mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (3.12)$$

The negation operator NOT creates the complement set of the set operated below.

Aggregation operators can take several forms, but the functions "min", "max", and "not" are the most common forms and are sufficient most of the time (Ross, 2010; Yager and Filev, 1994).

3.4.2.4 Defuzzification

A fuzzy-based system generates a final fuzzy variable that is not in a form that allows non-experts to comprehend it. Therefore, the fuzzy variable is transformed into a clear value via de-fuzzification. Several approaches have been utilised to determine the optimal crisp value, including the centre of gravity method (COG), the mean of maxima method, the weighted average method, and the max method. The most used defuzzification techniques are the mean of maxima and COG. The former calculates the mean of the output membership function maxima, whereas the latter finds the fuzzy variable's geometric centroid (Lee, 2005; Zimmermann, 1996).

3.4.2.5 Development of FWQI

For drinking water quality, twenty-two parameters were selected for developing a fuzzy logic-based water quality index (FWQI). The parameters selected were: pH, DO, EC, BOD_5, turbidity, TDS, hardness, alkalinity, Na^+, K^+, Ca^{+2}, Mg^{+2+}, F^-, Cl^-, SO_4^{-2}, NO_3^-, Fe, Mn, Cd, Pb, Cu, and Zn. 11 fuzzy sets in terms of membership functions were defined for each input indicator. Each

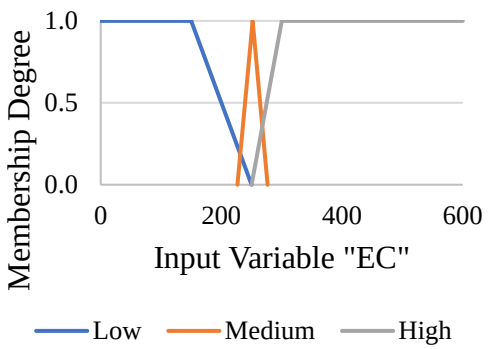
membership function links the elements in the universe of discourse to the interval [0,1]. Membership functions used in this study have trapezoidal for the twenty-one parameters. Each set in the universe of discourse X is introduced by a linguistic term. In this study, low, medium, and high were used as linguistic terms for the fuzzy sets. Fig. 3.7. shows the membership functions that fuzzyify the input data according to eq. (3.13).

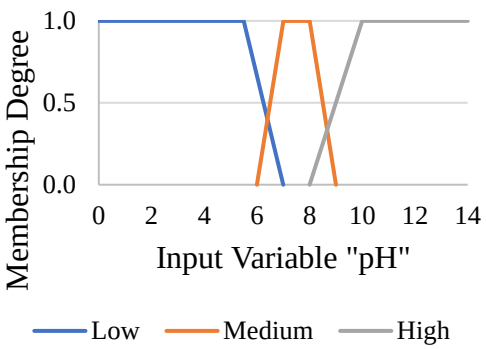
$$\text{Trapezoidal: } f(x; a, b, c, d) = \begin{cases} 0 & x < a \text{ or } d < x \\ \frac{(a-x)}{(a-b)} & a \le x \le b \\ 1 & b \le x \le c \\ \frac{(d-x)}{(d-c)} & c \le x \le d \end{cases} \quad (3.13)$$

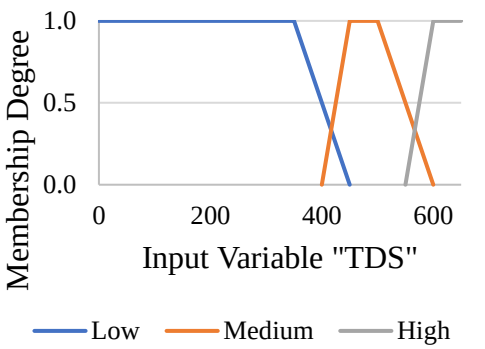
where, a, b, c, and d are membership function parameters, and x represents every single point on the x-axis. The parameters were normalised using membership functions (Fig. 3.7) and aggregated to generate a WQI (Fig. 3.8) using the Mamdani fuzzy inference system (Gharibi et al., 2012a). The fuzzy inference system helps imitate the knowledge of experts and also the restrictions placed on the drinkable water quality, simulating human behaviour and environmental processes (Ross, 2010). The performance and reliability of a system are determined by the quantity and quality of its rules. The rules cover all potential FWQI modes for the 21 input parameters. Based on the knowledge of specialists and IS 10500: 2012, about 300 rules were developed for this study (Chapman, 2021; Gharibi et al., 2012a; Ocampo-Duque et al., 2013).

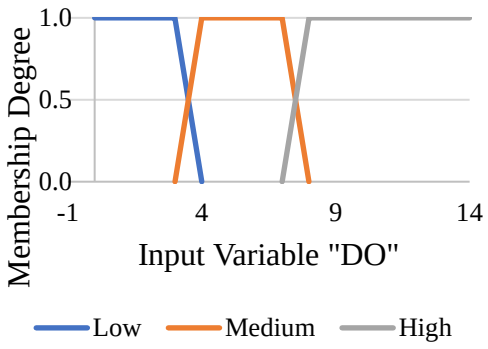
The following are examples of the rules set for the EC-pH group, complied with the experts' knowledge obtained from different literatures:

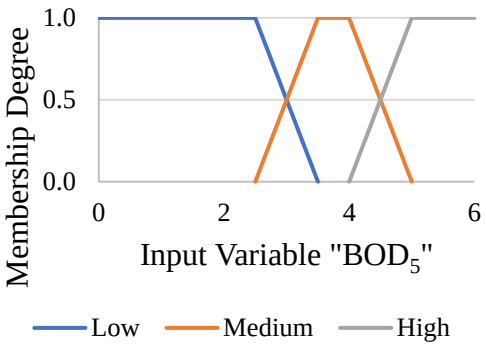
1. If (EC is Low) and (pH is Low), then (FG1Sc is Worse)
2. If (EC is Low) and (pH is Medium), then (FG1Sc is Good)
3. If (EC is Low) and (pH is High), then (FG1Sc is Bad)
4. If (EC is Medium) and (pH is Low), then (FG1Sc is Worse)
5. If (EC is Medium) and (pH is Medium) then (FG1Sc is Bad)
6. If (EC is Medium) and (pH is High) then (FG1Sc is Bad)
7. If (EC is High) and (pH is Low) then (FG1Sc is Worse)
8. If (EC is High) and (pH is Medium) then (FG1Sc is Worse)
9. If (EC is High) and (pH is High), then (FG1Sc is Worse)
10. If (EC is Low) then (FG1Sc is Good)
11. If (EC is Medium) then (FG1Sc is Bad)
12. If (EC is High) then (FG1Sc is Worse)
13. If (pH is Low) then (FG1Sc is Worse)
14. If (pH is Medium) then (FG1Sc is Good)
15. If (pH is High) then (FG1Sc is Bad)

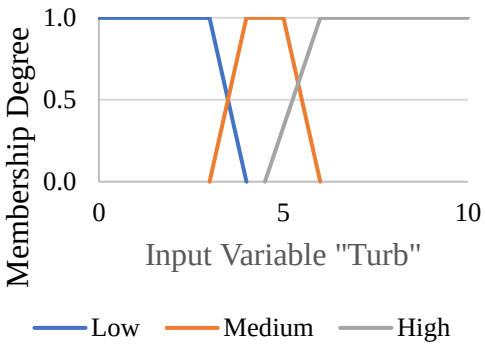


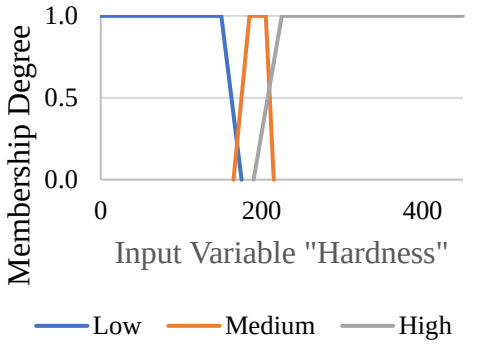


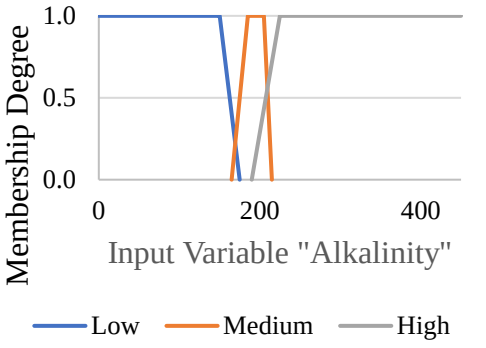


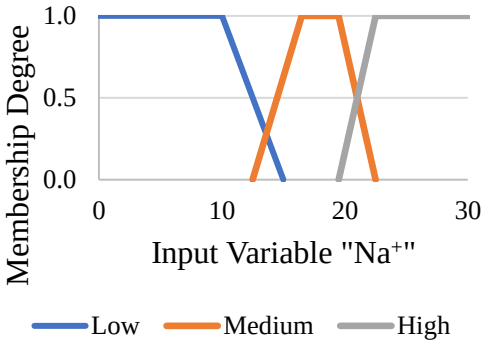


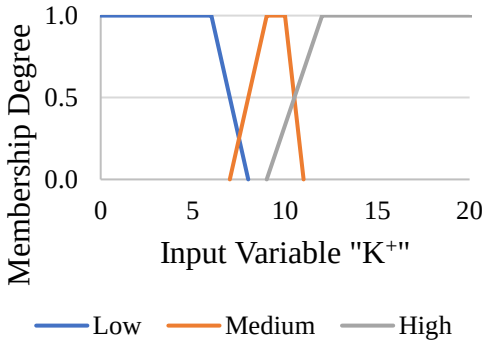


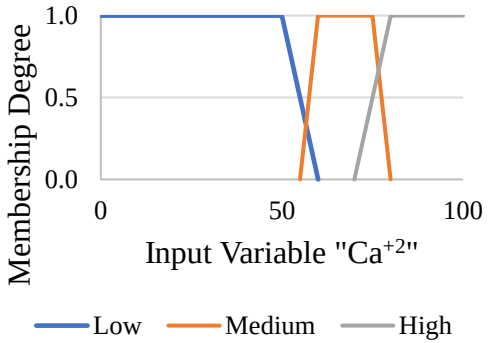


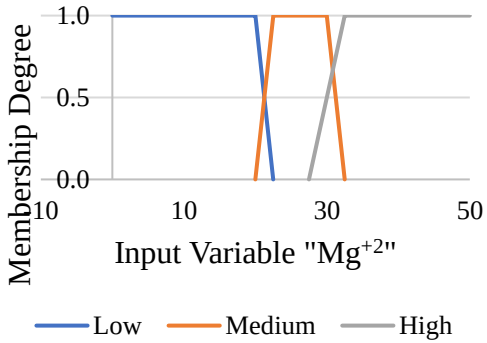


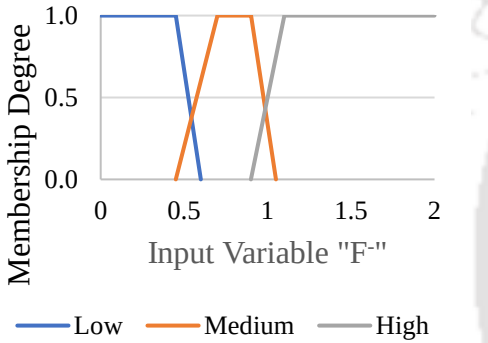


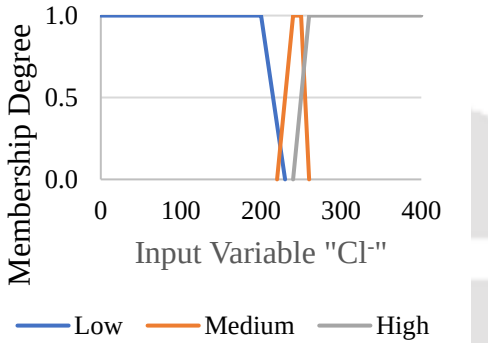


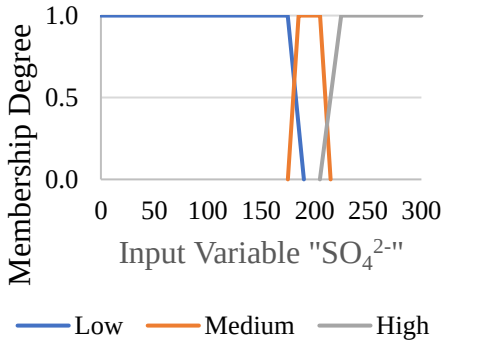


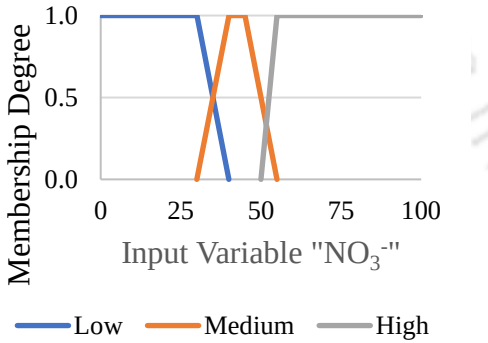


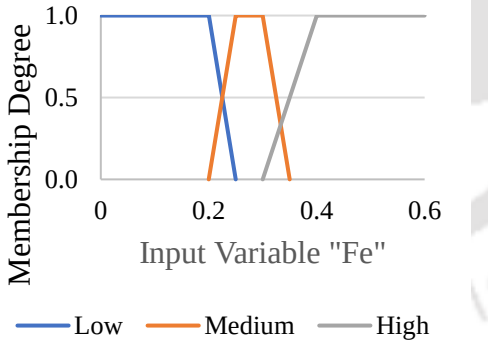


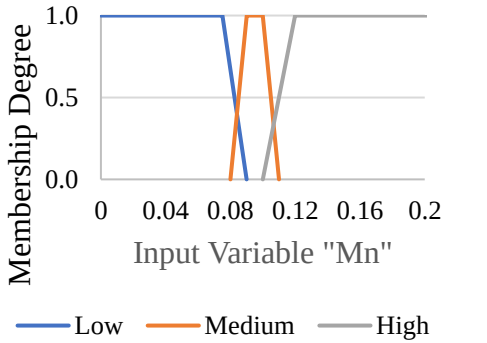












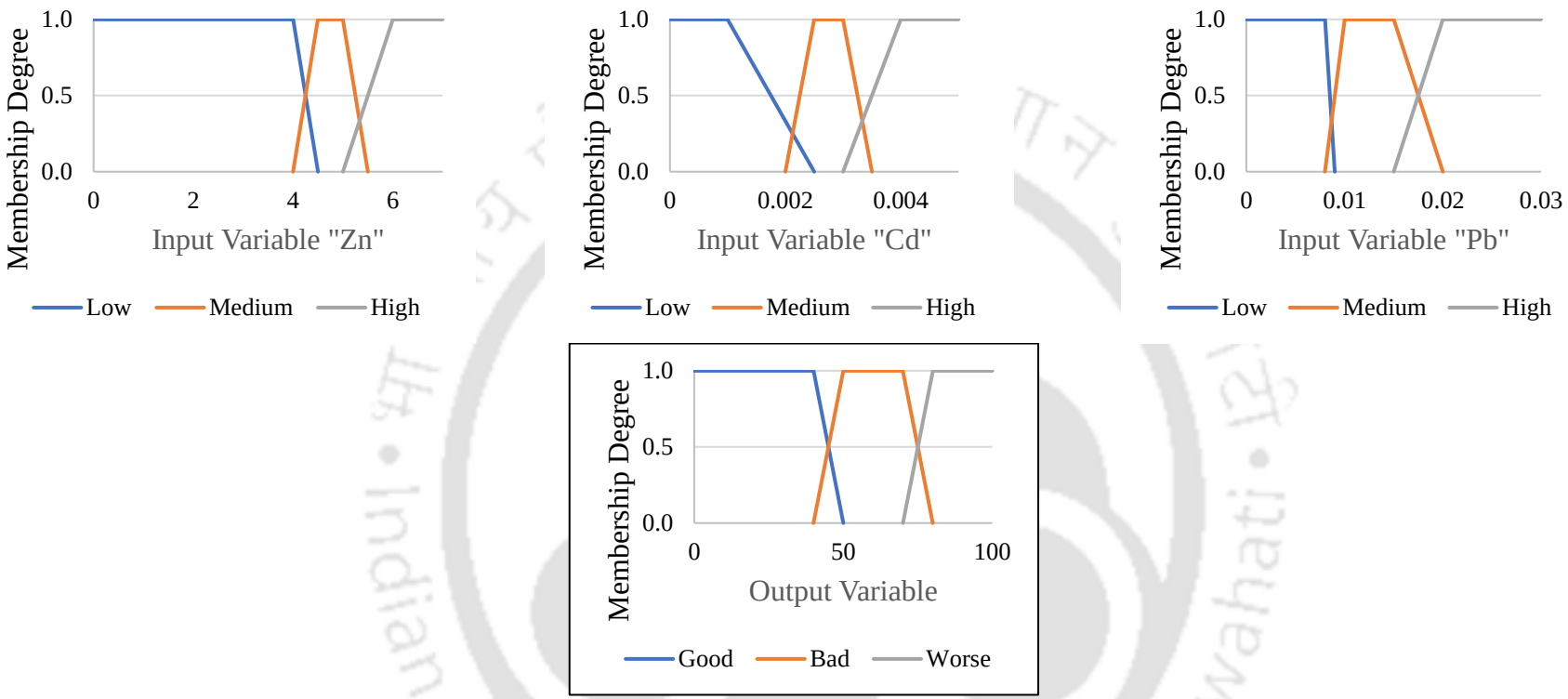
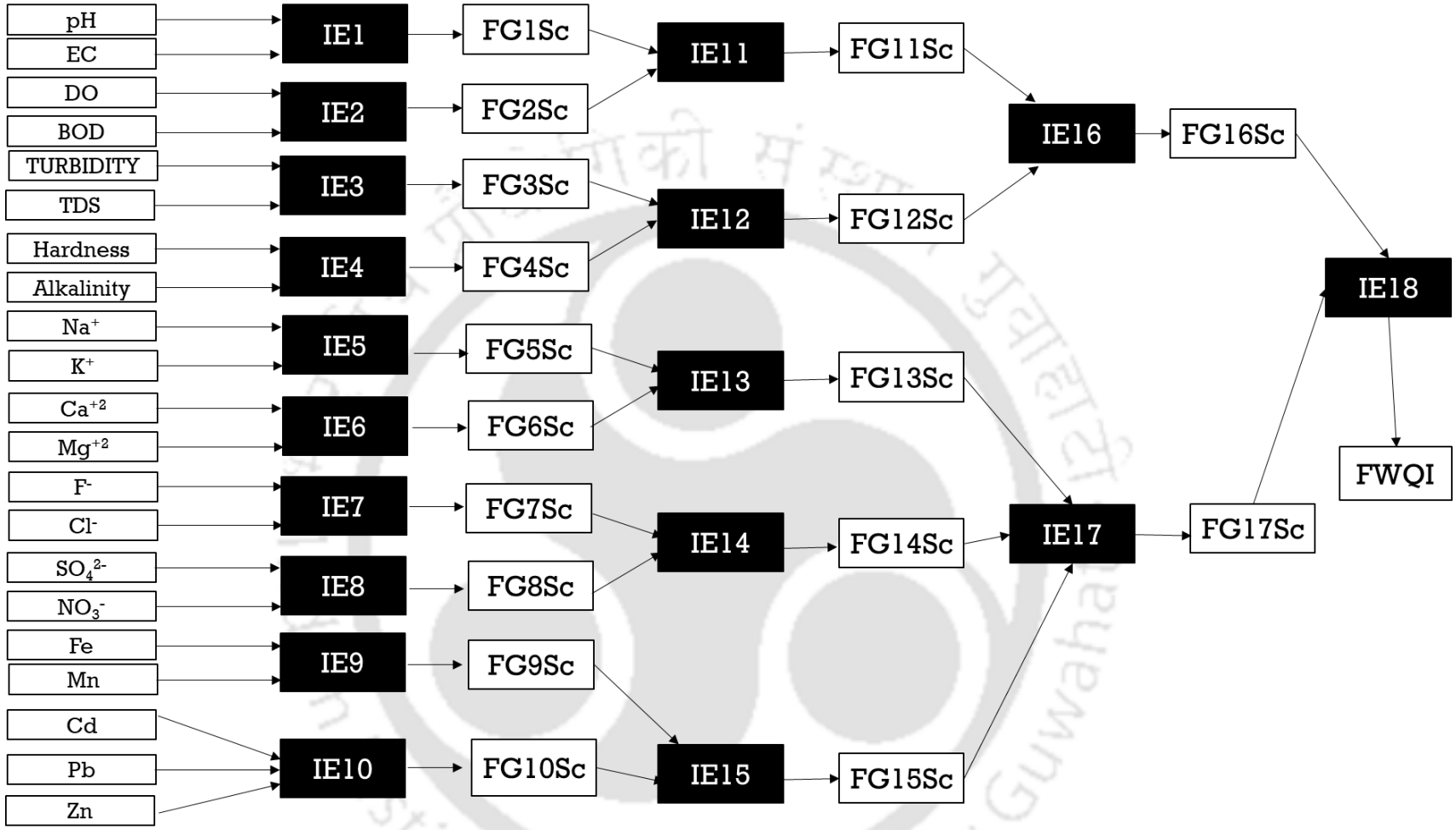


Fig. 3.7. Trapezoidal memberships functions to fuzzify and de-fuzzify input-output



- Index Calculation flow diagram (Made in Simulink)
- IE: Interference Engine; FG1Sc: Fuzzy Group 1 Score
- FWQI: Fuzzy Water Quality Index

Fig. 3.8. Aggregation of 21 input parameters in FWQI

Furthermore, the COG method, the most conventional and physically applicable defuzzification method (Jha et al., 2020), was used to de-fuzzify the outputs. The de-fuzzified values were obtained using Eq. (3.14) (Ross, 2010):

$$z = \frac{\int \mu(z) z \, dz}{\int \mu(z) \, dz} \tag{3.14}$$

where, z is FWQI.

All rules and calculations were made using the fuzzy logic toolbox in MATLAB (2022b), and FWQI aggregation was done using SIMULINK.

3.5 Objective II: Benthic Sediment Pollution Assessment

Sediments are the sink of metals in freshwater and marine environments (Arjonilla et al., 1994; Luoma, 1989; Weimin et al., 1994). Total concentrations of most metals in sediments are several orders of magnitude higher than aqueous concentrations (Luoma, 1989). Many researchers have feared that high and excessive amounts of heavy metals in the sediments may eventually contaminate the human and animal food chain (Iwegbue, 2007; Singh and Kalamdhad, 2013). This is because metals remain in the environment for a long time, unlike some organic pollutants. They are not biodegradable, and their residence time in the sediments can be thousands of years (Ekwutosi et al., 2020). The determination of total heavy metal concentrations can only be used to assess heavy metal pollution but cannot be used to assess the mobility and bioavailability of heavy metals (Nimyel et al., 2015; Zhang et al., 2017). Speciation can be an efficient tool for environmentalists to determine or assess the toxicity, mobility, reactivity, and bioavailability of heavy metals in the environment (Wan et al., 2017; Zhang et al., 2017). Therefore, heavy metal speciation in sediments plays an important role in environmental risk assessment of sediment heavy metal pollution. In this study, heavy metal sediment pollution assessment is carried out in two folds:

- Total metal concentration based assessment
- Metal speciation fractions-based assessment

3.5.1 Total metal concentration based assessment

3.5.1.1 Pollution Load Index (PLI)

PLI refers to the geometric mean of heavy metals measured in the same sediment in river bed sediments (Jorfi et al., 2017). This index makes it possible to standardise the amounts of heavy metals in river bed sediment with precision. Unskilled staff can easily understand this index effectively and rapidly to determine the pollution situation in various places (Elias et al., 2012). This index will allow environmental management policy decisions to be made easily. Tomlinson et al., (1980) have established a PLI, typically as follows:

$$CF_i = \frac{C_{m(i)}}{C_{bv(i)}} \tag{3.15}$$

$$PLI = \sqrt[nm]{\prod_{i=1}^{nm} CF_i} \quad (3.16)$$

where,

CF_i = contamination factor of i^{th} metal,

nm = number of metals,

$C_{m(i)}$ = heavy metal concentration measured (i^{th} metal),

$C_{bv(i)}$ = background value of i^{th} metal.

Anthropogenic contamination is inferred if the PLI value is >1, while <1 does not suggest contamination (Harikumar et al., 2009). There is a linear correlation between the population in urban areas and the impact of urbanisation on the river system, usually represented by PLI values (Singh et al., 2002). Geometric mean data for metal concentrations in sediments are typically smaller than the equivalent arithmetic mean and should be used as background sediment metal concentrations for assessing environmental pollution (Chen et al., 1999; Hseu et al., 2002).

3.5.1.2 Geoaccumulation Index (I_{geo})

The following equation was used to calculate the I_{geo} of the sediment (Müller, 1969) :

$$I_{geo} = \log_2\left(\frac{C_{m(i)}}{1.5 \times C_{bv(i)}}\right) \quad (3.17)$$

where,

$C_{m(i)}$ = heavy metal concentration measured (i^{th} metal),

$C_{bv(i)}$ = Geochemical concentration (background) of element i.

A factor of 1.5 is considered for potential variations in background data owing to lithological changes. The I_{geo} obtained is graded as Table 3.4 in seven grades (Boszke et al., 2008)

3.5.1.3 Enrichment Factor (EF)

Geological and human actions on sediment quality are assessed by EF, which is shown in Eq.(16) (Omwene et al., 2019):

$$EF = \frac{\left(\frac{C_{m(i)}}{C_r}\right)_{sample}}{\left(\frac{C_{m(i)}}{C_r}\right)_{backgound}} \quad (3.18)$$

where,

$C_{m(i)}$ = concentration of heavy metal,

C_r = reference heavy metal concentration,

Five types of contamination are commonly correlated with the enrichment factor shown in Table 3.4 (Qingjie et al., 2008). Here, iron (Fe) is used as a tracer to distinguish natural and anthropogenic contamination.

3.5.1.4 Potential Risk of Individual Metal (E_r^i) and Potential Ecological Risk Index (PERI)

Hakanson (1980) proposed the PERI and the possible risk of i^{th} metals (E_r^i), representing the overall effects of heavy metals and the toxicity of heavy metals (Cheng and Yap, 2015). The equation to determine E_r^i and PERI are as follows:

$$C_f^i = \frac{C_{m(i)}}{C_{bv(i)}} \quad (3.19)$$

$$C_d = \sum_{i=1}^{m} C_f^i \quad (3.20)$$

$$E_r^i = t_r^i \times C_f^i \quad (3.21)$$

$$PERI = \sum_{i=1}^{nm} E_r^i \quad (3.22)$$

where C_f^i = single element pollution factor, $C_{m(i)}$= concentration of heavy metal,, $C_{bv(i)}$= geochemical concentration (background) of element i, C_d = integrated pollution level (sum of all C_f^i of all the metals examined), E_r^i= potential ecological risk of i^{th} metal, t_r^i= Toxicity of i^{th} metal (Zn = 1, Pb = 5, Cu = 5, and Cd = 30). Different degrees of ecological risk per PERI have been listed in Table 3.4.

Table 3.4. Classification of I_{geo}, EF, E_r^i and PERI

I_{geo}, I_{POLL}	**Remarks**
(I_{geo}, I_{POLL}) ≤ 0	Uncontaminated
0 < (I_{geo}, I_{POLL}) ≤ 1	Uncontaminated to moderately contaminated
1 < (I_{geo}, I_{POLL}) ≤ 2	Moderately contaminated
2 < (I_{geo}, I_{POLL}) ≤ 3	Moderately to heavily contaminated
3 < (I_{geo}, I_{POLL}) ≤ 4	Heavily contaminated
4 < (I_{geo}, I_{POLL}) ≤ 5	Heavily to extremely contaminated
5 < (I_{geo}, I_{POLL})	Extremely contaminated
Enrichment Factor (EF)	**Remarks**
EF ≤ 2	Depletion to minimum enrichment
2 < EF ≤ 5	Moderate enrichment
5 < EF ≤ 20	Significant enrichment
20 < EF ≤ 40	Very high enrichment
EF >40	Extremely high enrichment
Potential Risk Individual Metal (E_r^i)	**Remarks**
E_r^i<40	Low ecological risk
40 ≤ E_r^i 80	Moderate ecological risk
80 ≤ E_r^i < 160	Considerable ecological risk

Potential Risk Individual Metal (E^i_r)	Remarks
$160 \leq E^i_r < 320$	High ecological risk
$E^i_r \geq 320$	Very high ecological risk
PERI and MRI	**Remarks**
PERI, MRI ≥150	Low ecological risk
150 < PERI, MRI ≤ 300	Moderate ecological risk
300 < PERI, MRI < 600	Severe ecological risk
PERI, MRI ≥ 600	Serious ecological risk

3.5.2 Metal speciation fractions based assessment

3.5.2.1 Pollution Index (I_{POLL})

Karbassi et al., 2008 updated Muller's I_{geo} formula as pollution index as follows:

$$I_{POLL} = \log_2\left(\frac{B_c}{L_p}\right) \tag{3.23}$$

where, B_c denotes bulk fraction and L_p denotes lithogenous fraction, respectively. The I_{POLL} obtained is graded as Table 3.4 in seven grades (Boszke et al., 2008)

3.5.2.2 Mobility Factor (MF)

The proportions of the metal fractions are used to determine the mobility of heavy metals in the environment. The chemical fractions of metals bound to sediments were split into two categories based on the degree to which they were associated with distinct phases. The mobility factor was then calculated using the information from these two groups. The F1 and F2 fractions comprise the mobile group, whereas the stationary group comprises of the F3, F4, and F5 fractions. The mobility factor (MF) was used to calculate the heavy metal mobility index was derived by Kabala and Singh (2001).

$$MF = \frac{F1 + F2}{F1 + F2 + F3 + F4 + F5} \times 100(\%) \tag{3.24}$$

3.5.2.3 Individual and Global Contamination Factors (ICF & GCF)

Evaluating heavy metal contamination is critical because it indicates the level of risk that heavy metals pose to the environment. The individual contamination factor (ICF) for each metal was computed by dividing the total of the first three extractions (F1, F2, and F3) by the residual (F5) fraction for each metal. The Global Contamination Factor (GCF) for each site was calculated by summing the ICF for each heavy metal detected (Ikem et al., 2003; Naji and Sohrabi, 2015). The ICF and GCF were determined with the help of the following equation:

$$ICF_i = \frac{F1 + F2 + F3}{F5} \tag{3.25}$$

$$GCF = \sum_{i=1}^{n} ICF_i \tag{3.26}$$

According to Zhao et al. (2012), the ICF and GCF categories were as follows: GCF<6 & ICF<1– indicates low contamination, 6<GCF<12 & 1<ICF<3–moderate contamination, 12<GCF< 24 & 3<ICF<6- considerable contamination, and GCF>24 & ICF>6– high contamination.

3.5.2.4 Risk Assessment Code (RAC)

Sediment contamination in the form of heavy metals are dispersed in a range of fractions in variable proportions, demonstrating a clear link between heavy metals in sediments and freshwater systems (Gao and Chen, 2012). RAC has the potential to express and regulate varying levels of risk associated with distinct fractions (Hing et al., 2020; Wang et al., 2010). The risk associated with contaminated sediments due to heavy metals is determined using the RAC method, which uses a rating to the proportion of heavy metals (%) found in the carbonate and exchangeable fractions of the sedimentary association to sediments (Sun et al., 2019; Sundaray et al., 2011). If the overall ratio of the carbonate and exchangeable fractions is < 1%, the ecosystem is deemed safe; and if the overall ratio is > 50%, the risk is considerable, and heavy metals may well infiltrate the food supply, according to the RAC (Ayyanar and Thatikonda, 2020). RAC has been used to categorise the risk posed by bioavailable fractions, as shown in Table 3.5.

Table 3.5. Risk Assessment Code (RAC) and values of ∂

Category	Risk	Bioavailable Fraction (%)	Toxic index (∂)
1	No Risk	<1	1.0
2	Low Risk	1-10	1.0
3	Medium Risk	11-30	1.2
4	High Risk	31-50	1.4
5	Very High Risk	>50	1.6

3.5.2.5 Modified risk index (MRI)

A toxic index is introduced for the exchangeable and carbonate fractions (%) to estimate the risk, in addition to the procedure followed for PERI to calculate MRI, which is as follows:

$$\varphi = \partial A + B \tag{3.27}$$

$$\sim C_{m(i)} = \sim C_{m(i)} \varphi \tag{3.28}$$

$$\sim C_f^i = \frac{\sim C_{m(i)}}{C_{bv(i)}} \tag{3.29}$$

$$\sim E_r^i = t_r^i \times \sim C_f^i \tag{3.30}$$

$$MRI = \sum_{i=1}^{nm} \sim E_r^i \tag{3.31}$$

Where $\sim C^i_f$, $\sim C_{m(i)}$, $\sim E^i_r$, and MRI are the revised forms of C^i_f, $C_{m(i)}$, E^i_r, and PERI, respectively;

φ is the modification factor for the concentration of heavy metals;

A is the cumulative percentage of carbonate and exchangeable fractions;

B = 1-A;

∂ is the toxic factor that corresponds to various carbonate and exchangeable fraction ratios; its values are provided in Table 3.5.

As per MRI, different degrees of ecological risk have been listed in Table 3.4.

3.5.3 Sediment Quality Guidelines (SQGs)

SQGs are scientific methods or procedures that indicate the potential adverse environmental effects of heavy metal concentration in sediments (CCME, 2001). They were developed to address many environmental concerns and regulatory programmes. SQGs from the literature are found to use two or three assessment scales of values for assessing pollution. Some commonly used SQGs are TEL-PEL (MacDonald et al., 2000) and ERL-ERM (MacDonald et al., 2000). The threshold effect level (TEL) and probable effect level (PEL) approaches use geometric means of percentiles of effects and no-effects distributions for slightly differing threshold levels. Another system for assessing sediment quality is the ERL-ERM guidelines. It considers the concentration of substances in the aquatic environment based on their likelihood of causing adverse effects. Concentrations corresponding to the 10th and 50th percentiles of the data distribution of incidence of adverse effects were named the effects range low (ERL), and effects range median (ERM), respectively. These correspond to concentration values wherein the incidence of adverse effects was unlikely or frequent. The SQG may be regarded as a screening approach for providing an overall perspective of the environmental sensitivity of a water body. As there are no established guidelines in India for assessing metal enrichment in sediments, the SQG proposed by researchers elsewhere are used in this study to assess sediment quality (Swarnalatha and Nair, 2017).

Table 3.6. Sediment Quality Guidelines (SQGs) classification

Sediment Quality Guidelines (SQGs)		**Effect**
TEL and PEL guidelines	<TEL	Not associated with adverse biological effects
	Between TEL-PEL	May occasionally be associated with adverse biological effects
	>PEL	Frequently associated with adverse biological effects
LEL and SEL guidelines	<LEL	Dredged sediments may have no contamination
	Between LEL-SEL	The impact is moderate
	>SEL	Severely impacted
ERL and ERM guidelines	<ERL	Minimal effects range
	Between ERL-ERM	Effects would occasionally occur
	>ERM	Effects would frequently occur

3.6 Objective III: Source Apportionment Using Environmetrics Tools

The present study uses four environmetrics tools such as hierarchical cluster analysis (HCA), discriminant analysis (DA), principal component analysis (PCA), and positive matrix factorization (PMF) for the assessment of water quality and sediment pollution in Kolong River, Assam, India.

3.6.1 Cluster Analysis (CA)

According to Ward (1963), agglomerative hierarchical CA, a squared Euclidean distance, was used to identify multivariate commonalities in coastal water quality. The observations with the most similarities are initially grouped, and then these initial groups are combined based on their commonalities. As similarity declines, eventually, all subgroups are consolidated into a single cluster (Alkarkhi et al., 2009). To calculate the distances between clusters, the Ward technique employs an analysis of variance. At each phase, it seeks to minimise the sum of squares of any two (hypothetical) clusters that can be generated. The spatial variability of water quality is expressed as Dlink/Dmax, the ratio of the linkage distances for a certain case to the maximum linkage distance. The quotient is then multiplied by 100 to standardize the linkage distance represented on the y-axis (Simeonov et al., 2003; Singh et al., 2005; Singh et al., 2004; Wunderlin et al., 2001). The initial stage in CA should be data normalisation (Liu et al., 2003), which tends to boost the influence of variables with small variances and decrease the influence of variables with large variance (Gupta et al., 2009). This analysis produces a dendrogram, which is a highly interpretable, comprehensive graphical explanation of hierarchical clustering and is one of the primary reasons for the popularity of hierarchical clustering approaches (Ruppert, 2004).

In this study, the software package IBM SPSS Statistics 25 was used to carry out CA using all the water quality parameters. Spatial CA was carried out for post-monsoon, winter, pre-monsoon and monsoon seasons. The squared Euclidean distance (Kukreja et al., 2016) was calculated between two clusters (a and b) from the standardized values given below:

$$D^2 = \sum_{i=1}^{n} (a_i - b_i)^2 \tag{3.32}$$

where D^2 squared Euclidean distance

a_i: standardized values of the i^{th} parameter for cluster a

b_i: standardized values of the i^{th} parameter for cluster b

n: number of parameters

i: water-quality parameters

3.6.2 Discriminant analysis (DA)

Discriminant analysis (DA) is a technique for identifying and classifying cases into categorical dependent values, typically a dichotomy. Multiple quantitative features are employed in DA to distinguish two or more naturally occurring groupings. DA provides a statistical classification of samples performed with previous knowledge of an object's membership in a specific group or cluster. Furthermore, DA aids in the clustering of samples with similar features. As shown in the

equation below, the DA approach creates a discriminant function for each group that acts on raw data (Härdle and Simar, 2013; Singh et al., 2005; Singh et al., 2004; Wunderlin et al., 2001):

$$f(G)i = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \quad (3.33)$$

where i is the number of groups (G), k_i is the constant inherent to each group, n is the number of parameters used to classify a set of data into a given group, and w_j is the weight coefficient assigned by DA to a given selected parameter (p_j). The weight coefficient maximises the distance between the means of the dependent variable (criterion). The classification table, known as a confusion, assignment, or prediction matrix or table, is utilised to evaluate the performance of DA. This is merely a table with rows representing the observed categories of the dependents and columns representing the expected categories. When the forecast is accurate, all instances will be located on the diagonal. The proportion of cases on the diagonal represents the proportion of accurate classifications.

In this study, DA was applied to the raw data matrix using standard, forward stepwise, and backward stepwise modes to create discriminant functions to assess the temporal differences in water quality within the river basin. The temporal (season) variables were the main (dependent) variables, while all observed parameters were independent.

3.6.3 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a holistic method for identifying the dynamics of all variables in a system (Gorgoglione et al., 2019). It seeks to reduce the dimensionality of a multivariate dataset by extracting information in the form of a fewer number of principal components that represent typical environmental characteristics. Thus, by curtailing the dimensions of the multivariate dataset, the main principal factors comprising all the necessary information may be extracted with minimum information loss (Abdi and Williams, 2010). The following sequence of five principal operations usually takes place while extracting the principal components in PCA (Yang et al., 2020).

(a) Firstly, the original data matrix was obtained as Eq. 3.34:

$$X = (x_{ij})_{n\times p} = \begin{vmatrix} x_{11} & . & . & x_{1p} \\ . & . & . & . \\ . & . & . & . \\ x_{n1} & . & . & x_{np} \end{vmatrix} \quad (3.34)$$

(b) The original data was standardized to reduce the dimensionality using Eq. 3.35:

$$x_{ij}^{*} = \frac{(x_{ij} - \overline{x}_j)}{s_i} \quad (3.35)$$

(c) The correlation coefficient matrix was obtained based on Eq. 3.36:

$$R = (r_{ij})_{p\times p} = \frac{1}{n-1}\sum_{t=1}^{n} x_{ti}^{*} \times x_{ij}^{*} \quad (3.36)$$

(d) Eigenvalues and eigenvectors of the correlation coefficient matrix were obtained as Eq. 3.37:

$$F_i = u_{i1}x_1^{*} + u_{i2}x_2^{*} + ... + u_{in}x_n^{*} (i = 1,2,.....n) \quad (3.37)$$

(e) Principal components were obtained using Eq. 3.38:

$$F = \frac{\lambda_1}{\lambda_1 + \lambda_2 + ... + \lambda_n} F_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2 + ... + \lambda_n} F_2 + ... + \frac{\lambda_n}{\lambda_1 + \lambda_2 + ... + \lambda_n} F_n \quad (3.38)$$

where, n and p are the number of sampling sites and water quality parameters, respectively; x_{ij} and x_{ij}^{*} are the originally measured data and standardized variable, respectively; x_j is the average value for j^{th} indicator, s_j is the standard deviation of j^{th} indicator, λ_i and u_i are the eigenvalues and eigenvectors, respectively; and x_i^{*} is the standardized indicator variable.

3.6.4 Positive Matrix Factorization (PMF)

The PMF receptor model is one of the most important and widely used methods in source allocation (Yan et al., 2019). In the PMF model, n (samples) × m (species concentration) matrix (X) is decomposed into two-factor matrices: profiles (F) and contributions (G); and residual (E) (Eqn. 3.39). Two files were used as input to the PMF model: one comprising concentrations of the assessed parameters and the other including uncertainty values derived using Eqn. (3.40) or (3.41). The ideal number of factors is obtained by repeatedly running the model for several iterations and then selecting the best run or solution with the minimum value of Q (robust), which indicates the model's ability to fit data (Jiang et al., 2019). Eqn. (3.43) has been used to determine this parameter. The primary goal of PMF factor analysis is to minimise Q for G and F while ensuring that all or some of G and F's components have non-negative values (Paatero, 1997).

$$X = FG + E \quad (3.39)$$

If the species concentration (i.e. concentration of heavy metal (C_s)) was more than the standard deviation (SD), the uncertainty (Unc.) was calculated by Eqn. (3.40); otherwise, the C_s were substituted by SD/2, and the uncertainty was presumed to be supplied by Eqn. (3.41) (G. Wang et al., 2016).

$$Unc. = 0.1C_s + \frac{SD}{3} \quad (3.40)$$

$$Unc. = 5 \times \frac{SD}{6} \quad (3.41)$$

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{e_{ij}}{s_{ij}}\right)^2 \quad (3.42)$$

where e_{ij} is the sum of squared differences between the original matrix data (X) and output of PMF (GF), and s_{ij} is the calculated uncertainties.

The investigation was conducted using the EPA PMF 5.0. The optimum number of factors was estimated by looking at the value of Q, which indicates how well the model fits the data. A global minimum was obtained for each model run by varying the seed value from 1 to 20 times (Bzdusek et al., 2006).

3.7 Objective IV: Human Health Risk Assessment

The human health risk models developed by the USEPA, which include carcinogenic and non-carcinogenic models, have been effective and accepted globally. There is currently no approved limit for allowable maximum carcinogenic and non-carcinogenic risk levels in India. In order to analyse the possible risks to human health posed by heavy metal pollution, the USEPA model and its associated threshold values were used. The health risk evaluation consisted of four steps: (1) hazard identification, (2) dose-response analysis, (3) exposure analysis, and (4) risk characterisation (USEPA, 2004, 1992, 1989). Two sources, namely sediments and surface water, were utilised to compute the heavy metal exposure dosages. Humans may be exposed to heavy metals from a river basin by four primary routes: (1) oral ingestion of sediment particles, (2) dermal contact with sediment particles, (3) oral intake of surface water, and (4) dermal contact with surface water from the river basin.

3.7.1 Exposure Parameters

Incidental ingestion and dermal contact (during recreational activities) with surface water and sediments were used to estimate health risks to humans due to exposure to heavy metals in children and adults. The chronic daily intake (CDI) was calculated using the USEPA's procedures (Eqns. (31)-(34)) in mg/kg (USEPA, 2001, 2004).

$$CDI_{ingestion-water} = \frac{C_w \times EF \times ET \times IR_w \times ED}{AT \times BW} \times CF \quad (3.43)$$

$$CDI_{ingestion-sediments} = \frac{C_s \times EF \times ET \times IR_s \times ED}{AT \times BW} \times CF \quad (3.44)$$

$$CDI_{dermal-contact-water} = \frac{C_w \times EF \times ET \times ED \times SA \times k_p}{AT \times BW} \times CF \quad (3.45)$$

$$CDI_{dermal-sediments} = \frac{C_s \times EF \times ET \times ED \times SA \times AF \times ABS}{AT \times BW} \times CF \quad (3.46)$$

where, C is the concentration of heavy metals in water (C_w: mg/L) and sediments (C_s: mg/kg). The rest of the details and values of the terms in Eqn. (3.43) -(3.46) are given in Table 3.7.

3.7.2 Risk Characterization

A risk assessment of heavy metal exposure, both non-carcinogenic and carcinogenic, was conducted to determine the health consequences following eqns. (3.47) and (3.48). Hazard quotients (HQs) were used to compute the non-carcinogenic risk for all heavy metals through the exposure pathways. The hazard index (HI) was computed by adding all HQs. If HQ and HI are more than 1, the suggested acceptable limits have been exceeded (Phillips and Moya, 2014). Potential carcinogenic health impact (CR) from inadvertent intake of sediments was estimated using Eqn (3.48). The cancer risk was evaluated using reported slope factors (SF_{oral}). The total cancer risk (TCR) was determined by adding the CR values for each route of exposure and comparing them to permissible standard values (USEPA, 2001; Yang et al., 2018). The reference dosage (R_fD) and slope factors (SF) were provided on the RAIS website (RAIS, 1992).

$$HQ_{ingestion} = \frac{CDI_{ingestion}}{R_f D_{oral}} \quad (3.47)$$

$$HQ_{dermal} = \frac{CDI_{dermal-contact}}{R_f D_{dermal}} \quad (3.48)$$

$$CR_{ingestion} = CDI_{ingestion} \times SF_{oral} \quad (3.48)$$

Table 3.7. Deterministic and probabilistic assessment parameters and values

Parameters		Deterministic Approach Point Estimate (RME)	Probabilistic Approach Distribution	Values	Reference
EF	Exposure Frequency (day/year)	120	Triangular	120 (26 -260)	
ED	Exposure Duration (Year)				
	Adults	30	Lognormal	11.36 ± 13.72	(Israeli and Nelson, 1992)
	Child	6	Uniform	01-06	(Anon, 2019)
ET	Exposure Time: adults and child (hour/event)	2.6	Triangular	2.6 (0.5-6)	(Anon, 2019)
SA	Skin surface area (swimming) (cm^2)				
	Adults	23000	Normal	18400 ± 2300	(Phillips and Moya, 2013)
	Child	7280	Normal	6800 ± 600	(Anon, 2019; Carr, 1994)
BW	Body Weight (kg)				
	Adults	70	Normal	72 ± 15.9	(Carr, 1994)
	Child	15	Normal	15.6 ± 3.7	(Phillips and Moya, 2013)
IR_w	Ingestion Rate of water (mg/event)				
	Adult	0.053	-	0.053	(USEPA, 2011)
	Child	0.090	-	0.090	(Goldblum et al., 2006)

Parameters		Deterministic Approach Point Estimate (RME)	Probabilistic Approach Distribu-tion	Values	Reference
IR_s	Ingestion Rate of Sedi-ments (mg/event)				
	Adult	12.5	-	12.5	(Goldblum et al., 2006)
	Child	50	-	50	(Goldblum et al., 2006)
AT	Averaging Time (day)				
	non-carcinogen	365xED	-	365xED	(USEPA, 2001)
	carcinogen	365x70	-	365x70	(USEPA, 2001)
AF	Adherence Factor (mg/cm^2)				
	Adult	0.07	-	0.07	(USEPA, 2001)
	Child	0.2	-	0.2	(USEPA, 2001)
ABS	Dermal Adsorption Factor (Unitless)	0.001	-	0.001	(USEPA, 2011; Wang et al., 2017)
K_p	Permeability Constant (cm/hour)	Cd, Co, Cu, Fe and Mn=10^{-3}, Pb = 10^{-4}, Zn = $6^{x}10^{-4}$,	-	Cd, Co, Cu, Fe and Mn=10^{-3}, Pb = 10^{-4}, Zn = $6^{x}10^{-4}$,	(RAIS, 1992)

3.7.2.1 Deterministic Method

Models for risk assessment generally have only one value assigned to each model parameter, with the outcome being a single value for the risk associated with that parameter. (Jiménez-Oyola et al., 2021). This technique is simple and easy to grasp, but input variables do not account for variability (Bazeli et al., 2020; Sheng et al., 2021). The deterministic approach's exposure parameters and generic exposure factors are shown in Table 3.7.

3.7.2.2 Probabilistic Method

Monte Carlo simulation (MCS) is a frequently used approach in probabilistic risk assessment, in which variables are described by their distribution (Jiménez-Oyola et al., 2021; Saha and Rahman, 2020). A wide variety of probable outcomes is generated using statistical sampling procedures in MCS in the form of probability distributions, which considers the data's inherent unpredictability and uncertainty (Low et al., 2015; Tong et al., 2018). MCS and other probability-based approaches are frequently used in human health risk assessment because they examine the sensitivity of the model input variables (Bazeli et al., 2020; Saha et al., 2017; Saha and Rahman, 2020; US EPA, 2001). The risk assessment for probabilistic modelling was done using MCS in Oracle Crystal Ball (Gitter et al., 2020; Xu et al., 2020). The number of iterations (for each run) was set at 10,000 to construct the probabilistic risk distributions (Yang et al., 2019). Heavy metal concentrations prior to MCS were fitted with the triangular distribution. For the probabilistic evaluation, the standard distributions and exposure parameter values are shown in Table 3.7.

3.7.2.3 Sensitivity Analysis

The impact of the input factors on cancer and non-cancer risk estimations is assessed using a sensitivity analysis based on the contribution to variance or rank coefficient correlation (Chen et al., 2019). This approach is commonly used in decision-making and risk management to discover the most critical factors influencing risk outcomes (Harris et al., 2017). The sensitivity analysis in this study was performed using the Spearman rank order correlation coefficient, which assesses how strongly and in which direction quantitative variables are connected to outcomes rather than the values themselves. (USEPA, 2001). The sensitivity analysis was performed using a 95% confidence level and 10,000 iterations.

3.8 Objective V: Regression Modelling

Quantification of heavy metals present in anything around us is very important in the present day because of its toxic characteristics. But, accurate assessment requires precise instruments that are expensive and require skilled persons to operate. These instruments are not available at every research institute, making it difficult to conduct research relating to heavy metals. In this context, predictive models can help estimate the concentration of heavy metals in surface water bodies and other water sources. Parametric statistical and deterministic models have been the traditional approaches for modelling water quality, these models require vast information on various hydrological subprocesses to arrive at the end results (Singh et al., 2009). Moreover, these models require precisely determined rate constants/coefficients pertaining to various hydrological, chemical, physical and biological processes, which are largely time and space specific. Additionally, such models have analytical solutions but have boundary conditions as limitations (Basant et al., 2010). Also, since many factors affect the water quality, it has a complicated nonlinear relation with the variables; therefore, traditional data processing methods are no longer good enough for solving the problem (Ranković et al., 2012). In recent years, several researches have been conducted on water quality forecast models (Chen et al., 2020; Najah et al., 2013; Ömer Faruk, 2010; Palani et al., 2008; Wu et al., 2014). Over the past several years, many nonlinear models such as ANN (Gandhimathi and Meenambal, 2012; Hanoon et al., 2021; Y. Jiang et al., 2022; Lu et al., 2019; Zhou et al., 2015), decision trees, random forest, support vector machines (Emamgholizadeh et al., 2014; Lu et al., 2022; Mohammed et al., 2022; Paes et al., 2022; Xia et al., 2022; Xiao et al., 2022) which belong to ML techniques among others have been used for the prediction and forecasting of water resource variables. Also, ML models have been widely accepted as a potentially useful way of modelling hydrological processes and have been applied to various areas, including water quality, rainfall runoff, sedimentation and rainfall forecasting (Tiyasha et al., 2020). In this study, quantification of heavy metals in water column is done in two-fold ways:

(1) Firstly, an attempt has been made to correlate the physicochemical parameters with heavy metals. This has been done keeping in view the estimation of heavy metals and remediation

options it will provide from the mathematical or physical relations with physicochemical parameters generated from the model.

(2) Secondly, the relation between the metal speciation fractions in the sediment column and metal concentrations in the water column will be explored. This model, after development with help in understanding the contribution of metal pollution load from the sediment column to the water column and thus help in the formulation of sediment remediation strategies.

These two models will be developed using multiple linear regression (MLR) and machine learning (ML) techniques such as artificial neural network (ANN) and random forest (RF) methods. Their efficacies will be determined using coefficient of determination (R^2).

Chapter 4

CHEMISTRY AND QUALITY OF RIVER WATER

4.1 Seasonal and Spatial Variation of Water Quality

4.1.1 Dissolved Oxygen (DO)

In surface water, DO varies from 8 ppm at 25°C to 15 ppm at 0°C. DO determination is an integral part of assessing water quality because almost every biological and chemical process in water bodies involves oxygen or its effects. Less than 4 ppm concentration harms aquatic communities and leads to death at less than 2 ppm in most fish communities. The estimation of DO can be used as an index of the presence of organic matter in a contaminated environment. The presence of DO in the Kolong River was noticeable throughout the year, except in the post-monsoon season. The minimum DO was 3.99 and 3.86 ppm at sites 'SSKR9', and 'SSKR11' in the post-monsoon, indicating a significant amount of organic pollution was discharged (Fig. 4.1). It never went below 4.5 ppm for the rest of the seasons. All the seasons have a dip in the DO variation curve, which indicates the discharge of organic pollution and the river's dilution capacity.

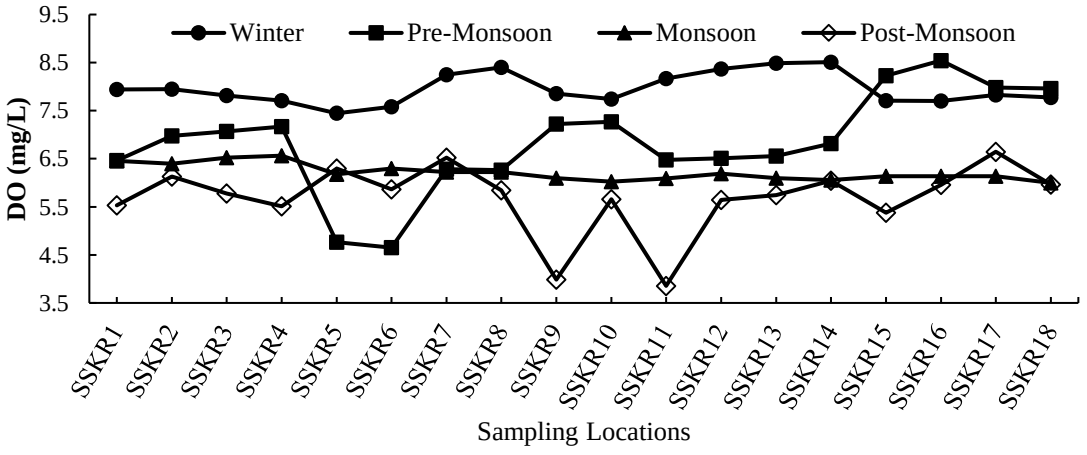


Fig. 4.1. Seasonal variation of DO

4.1.2 pH

The values measured during one year of sampling (Pre-monsoon, Monsoon, Post-Monsoon, and Winter) do not indicate any significant seasonal fluctuation of pH (Fig. 4.2) or any abnormality. The average pH values ranged from 6.71 to 7.44, which is the near-neutral nature of the Kolong river.

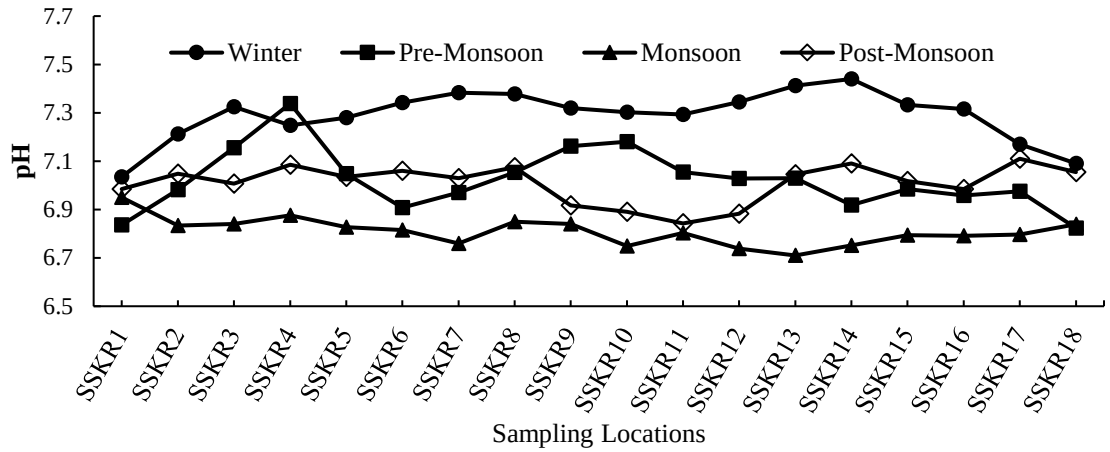


Fig. 4.2. Seasonal variation of pH

4.1.3 Turbidity, electrical conductivity (EC), hardness (TH), alkalinity (TA) and total dissolved solids (TDS)

Turbidity is a measure of a liquid's relative clarity. It is an optical property of water that measures the quantity of light scattered by materials in water when a light is shone through a sample of water. The higher the turbidity, the greater the intensity of scattered light. IS 10500: 200 recommends a permissible limit of 1 NTU for drinking purposes. The monsoon had the highest turbidity values, followed by post-monsoon, winter and pre-monsoon. With the onset of monsoon, i.e., in the pre-monsoon season, an increase in precipitation dilutes the turbidity value, but at the height of monsoon, various sources bring in high volume of solid particles, thereby increasing the turbidity values. With the receding monsoon, the turbidity values start to fall in the post-monsoon season (Fig. 4.3).

EC represents dissolved solid concentrations in water predominantly caused by inorganic ions. A high value of EC indicates a higher level of dissolved salts and vice versa. Conductivity values of the Kolong river for winter and post-monsoon seasons at upstream locations are higher than those prescribed by IS 10500: 2012 (Fig. 4.3). There is a dilution of EC for all seasons at the downstream section of the river. Hardness due to calcium is usually more prevalent (max: 70%), though hardness due to magnesium can reach up to 50-60% in a few cases. Hardness varies in river water due to seasonal fluctuations in flow, with the lowest values in high flow (flood) conditions and the highest values during lean flow conditions. Hardness has no adverse effects on health; however, some evidence has been given to indicate its role in heart disease. Hardness can be helpful in some situations, as it prohibits rusting in pipes by developing a thin layer of scale. Water hardness is an important parameter determining its utility for domestic and commercial activities. According to IS10500 2012, up to 200 ppm of hardness is acceptable. The river's hardness value has never exceeded the permitted limit throughout the sampling period (Fig. 4.3). The variation of alkalinity is similar to hardness values indicating the absence of non-carbonate hardness.

TDS in water or wastewater apply to the dissolved material. In several ways, solids can adversely influence water or effluent quality. Water with a large number of dissolved solids is not

ideal for consumption and can lead to physiological reactions in consumers. The maximum acceptable dissolved solid level is 500 ppm for drinking water (IS 10500 2012). TDS values of river water are influenced by environmental factors such as agricultural practices, flow barriers, etc. The TDS values are within the acceptable limit for all the seasons, with 475 ppm being the highest TDS value in the winter (Fig. 4.3) (Bora and Goswami, 2015). In the Kolong River, EC, TDS and hardness were the highest in the winter season and the lowest in the monsoon season.

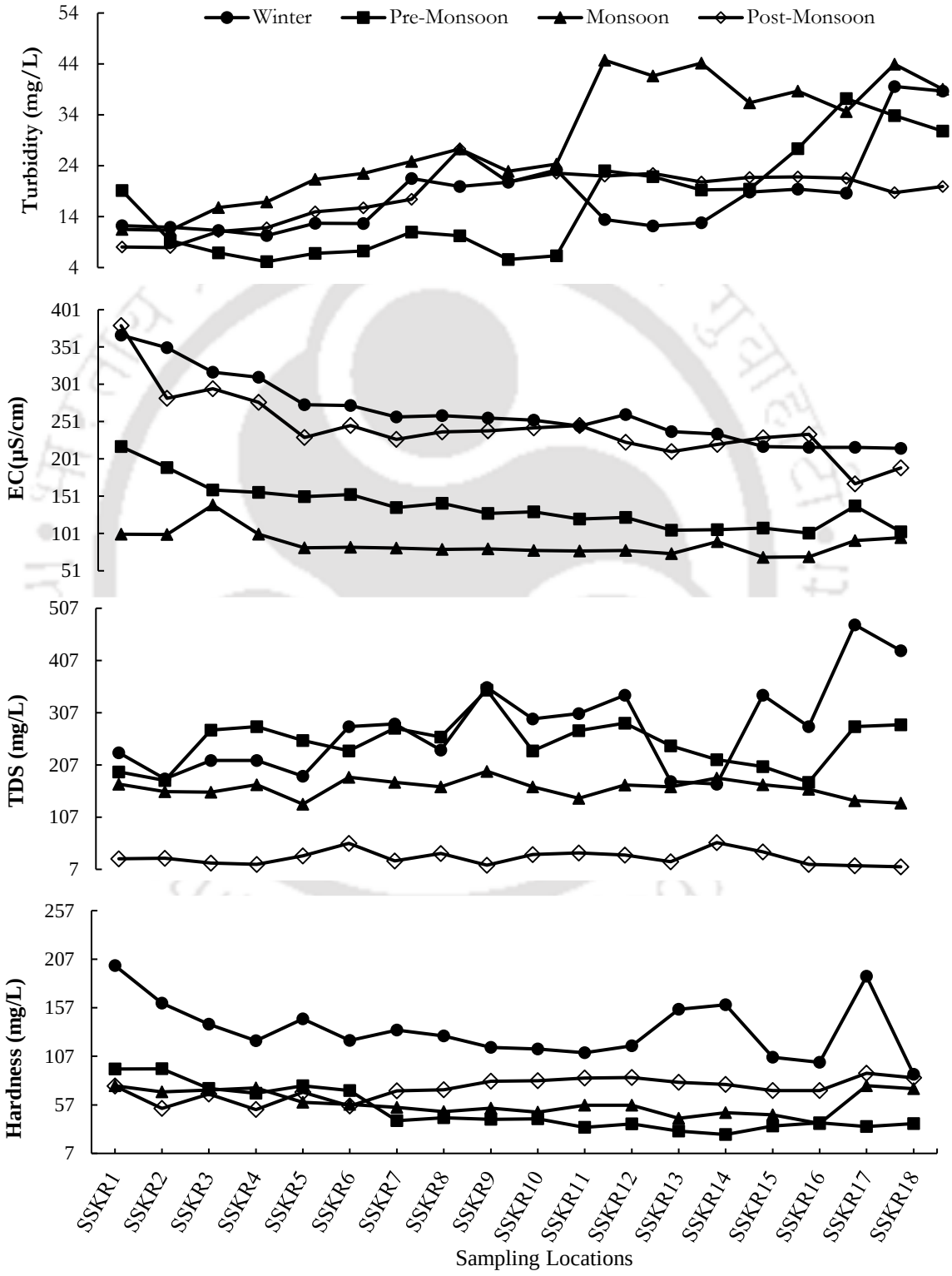


Fig. 4.3. Seasonal variation of turbidity, EC, TDS and hardness

4.1.4 Major ion chemistry

4.1.4.1 Sodium (Na^+), potassium (K^+), calcium (Ca^{+2}) and magnesium (Mg^{+2})

Highly water-soluble sodium salt is one of the most common elements. The concentration of Na^+ in natural waters differs significantly based on local geological factors and discharges of wastewater (Michalopoulos et al., 2014). Augmented amounts in surface waters can arise from residential and commercial wastewater. IS 10500:2012 recommends an acceptable limit for Na^+ in drinking water as 200 ppm. Na^+ concentrations in Kolong River water ranged from 25.30 ppm (pre-monsoon season) to 5.77 ppm (post-monsoon season) (Fig. 4.4). Potassium ions are found in lesser amounts in natural waters, as rocks containing K^+ are relatively weather-resistant (Ralph and Diane, 1980). No significant variation of K^+ was observed during the year of study (Fig. 4.4). Higher values were observed in the winter or pre-monsoon seasons, i.e. lean periods. For Na^+ and K^+, the lowest values were found in monsoon or post-monsoon season, i.e. high flow periods, which might be due to the high discharge of an enormous volume of water during this period. Ca^{+2} in river water mainly comes from weathering carbonates, sulphates and silicate minerals. The concentration of Ca^{+2} ions has been found to vary with the nature of the basin. It is a critical micronutrient in the aquatic system and plays a significant role in the hardness of water in the environment (Hunsaker and Pratt, 1971). In the Kolong river, concentration of Ca^{+2} (Fig. 4.4) fluctuated between the following range; 62.5 to 22.1 ppm in the winter season, 27.8 to 4.1 ppm in pre-monsoon, 22.0 to 6.45 ppm in monsoon season; 45.6 to 28.5 in the post-monsoon season. Higher values were observed in the winter or post-monsoon, i.e. lean period. The lowest values were observed in monsoon or pre-monsoon season, i.e. high flow periods, which might be due to the high discharge of an enormous volume of water during this period, which is similar to sodium variation. Like Ca^{+2}, Mg^{+2} also occurs in natural water and is derived from the same sources as calcium. The contribution of Mg^{+2} is always somewhat less than Ca^{+2}. It is to be noted that high Mg content makes them unfit for domestic use and reduces potability. The concentration of Mg in the Kolong river water with season-wise variation in a year is shown in Fig.4.4. Concentration of calcium varied from 15.4-8.0 ppm in the winter season, 9.9-2.0 ppm in the pre-monsoon season, 7.4-5.0 ppm in monsoon season and 24.4-12.8 ppm in the post-monsoon season. Concentration of Na^+, K^+, $Ca^{+2}$and Mg^{+2}, however, has fluctuations, but their values are within acceptable or permissible limits. Variations of Na^+ and K^+ are similar, whereas the variation of Ca^{+2} and Mg^{+2+} are similar. No abnormal increase or decrease in values was observed during the one-year study.

4.1.4.2 Fluoride (F^-), chloride (Cl^-), nitrate (NO_3^-) and sulphate (SO_4^{2-})

Fluoride (F^-) is a crucial element of human metabolism and can be beneficial as well as detrimental to its presence in drinking water according to concentration. The two extremes of F^- concentration, deficiency and excess, are associated with dental disorders and fluorosis (Seppä et al., 1998). F^- occurrence has been documented in drinking water in nearly all parts of the world

(Mariappan et al., 2000). F^- in potable water depends on the source, environment and geology. According to IS 10500:2012, the acceptable limit is 1.0 ppm in potable water. In the present study, the F^- concentration in the Kolong river is 0.07-0.65 ppm, with winter and pre-monsoon having the highest F^-. The seasonal F^- variation in the river is depicted in Fig. 4.5.

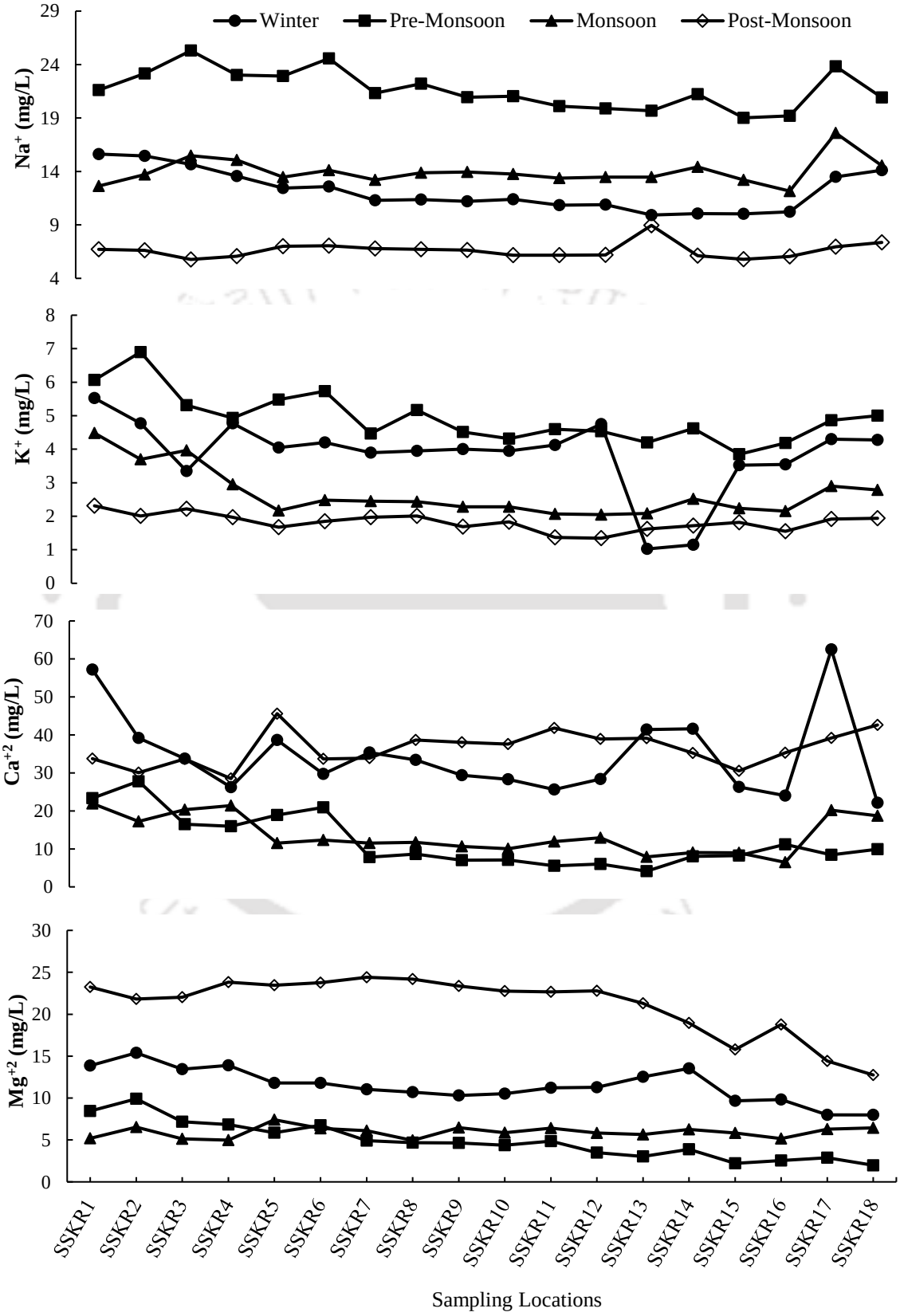
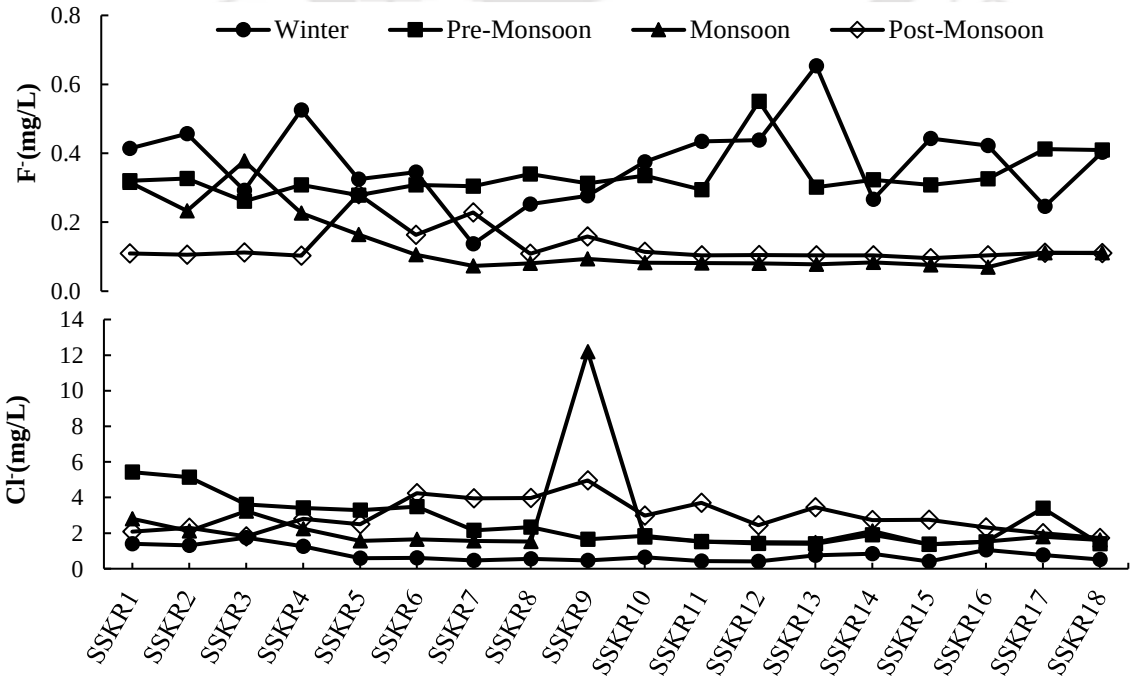


Fig. 4.4. Seasonal variation of Na^+, K^+, Ca^{+2} and Mg^{+2}

Chloride (Cl^-) occurs naturally in ground and surface water. The primary sources of chloride are atmospheric precipitation, sedimentary rocks, and industrial and domestic sewage. Without other sources, the concentration of Cl^- in natural water by atmospheric precipitation should not exceed 20 ppm (Sibley, 1979). A higher level of Cl^- concentration in natural water bodies indicates pollution due to industrial and domestic sewage. In the Kolong river, chloride concentration variation varied from 0.41 to 12.2 ppm over the sampling period, for which the monsoon has the highest industrial and domestic discharge (Fig. 4.5).

NO_3^- is one of the most common pollutants in surface water. It is controlled primarily in potable water; surplus amounts may induce "blue baby" or methemoglobinemia disease in newborns. While NO_3^- is not an immediate threat to teenagers and adults, its presence indicates one or more severe contaminants like bacteria or pesticides (Winton et al., 1971). The nitrate comes primarily from manure storage, fertilisers, and septic systems. NO_3^- content in the Kolong river varies between 20.77 ppm to 0.36 ppm for the entire sampling period (Fig. 4.5). Not much fluctuation is observed in nitrate concentration, with sampling locations 'SSKR8' and'SSKR10' showing an unusual rise in concentration, indicating possible non-point sources of pollution existing.

Due to biological activity, sulphate (SO_4^{2-}) may come from the organically bound sulphur. This phenomenon is observed mainly in the marine environment due to the ample supply of sulphur in seawater (Golterman et al., 1982). Sulphate is also an essential component of calcium and magnesium hardness. Sulphate produces an unpleasant taste at 300-400 ppm. The suggested upper limit for water sulphate is 250 ppm for human use (Sawyer et al., 1978). During the sampling period, the sulphate concentration in the Kolong river ranged between 13.54 and 0.21 ppm. The maximum sulphate concentration was found during the monsoon season, indicating that agricultural runoff entering the river occurred during the monsoon season. The seasonal variation of sulphate is shown in Fig. 4.5.



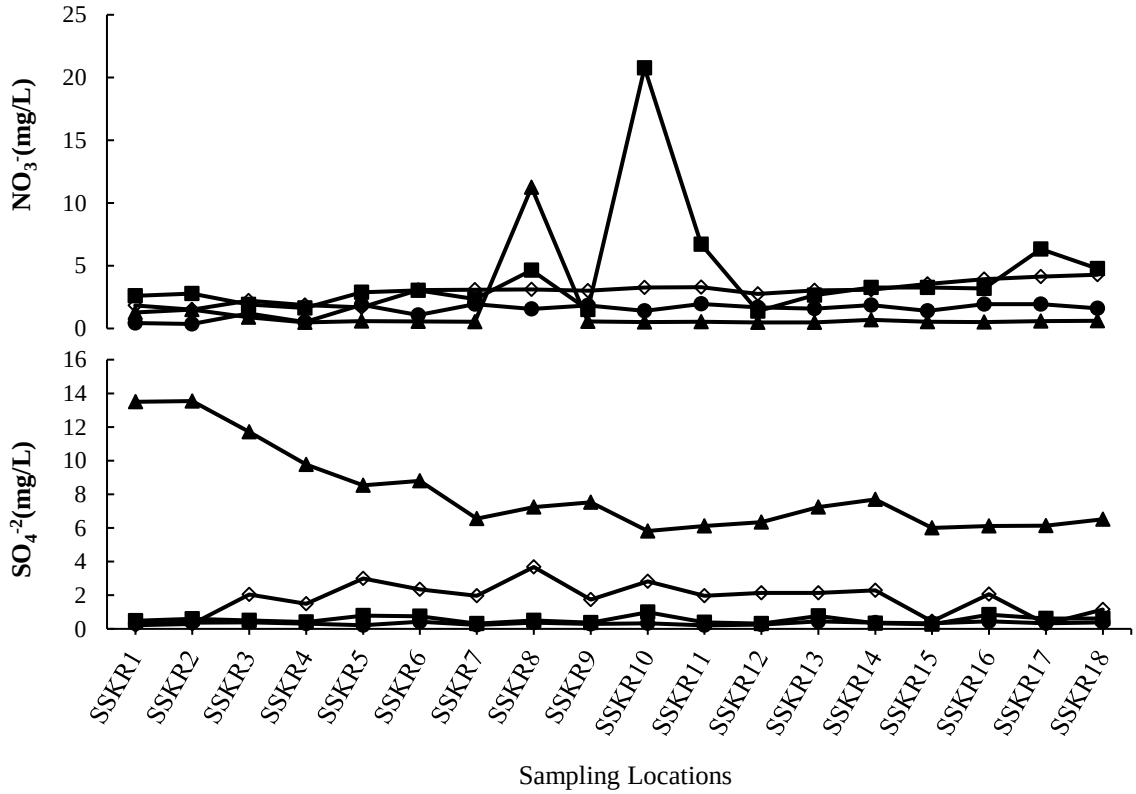


Fig. 4.5. Seasonal variation of F^-, Cl^-, NO_3^-, and SO_4^{2-}

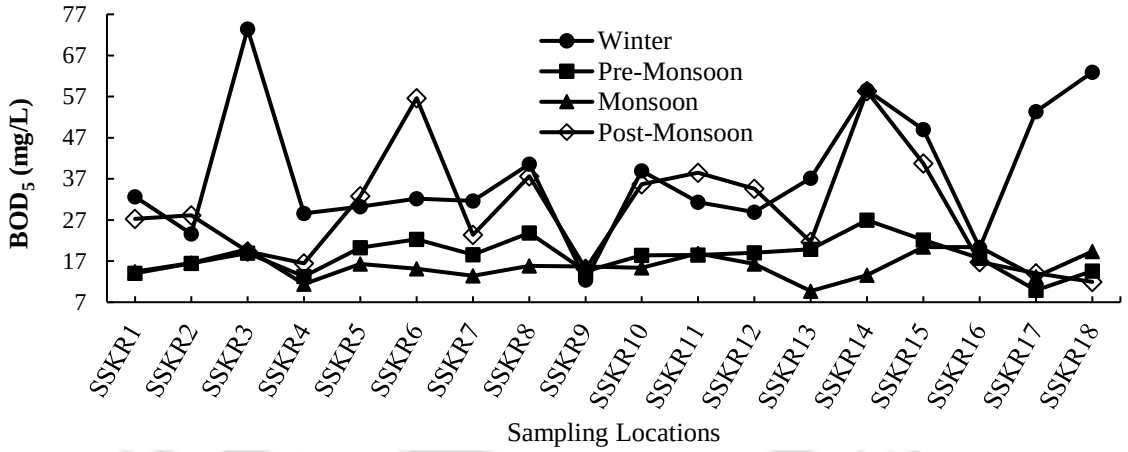


Fig. 4.6. Seasonal variation of BOD

4.1.5 Biochemical Oxygen Demand (BOD)

The quantity of DO needed for organic matter to be decomposed by "aerobic microorganisms" is designated as BOD. Therefore, BOD is an indicator of environmental pollution, with a higher value implying more significant natural pollution (Ranjith et al., 2019). Values greater than 5 ppm are objectionable and suggest organic contamination of some form. BOD ranged from 9.30 to 73.1 ppm in this study, with the winter season having the highest and the monsoon season having the lowest BOD values showing the dilution in the river system (Fig. 4.6) (Bora and Goswami, 2015).

4.1.6 Heavy metals

Heavy metals are persistent contaminants as they are neither removed nor degraded. Therefore, in sediments and soils, they appear to accumulate. Higher levels of heavy metals can impact aquatic biota in the marine or aquatic environment and pose a risk to human consumption. Due

to urbanisation and industrialisation, trace metals have grown more in various water bodies, particularly heavy metals. More than 50 elements are listed as heavy metals, of which 17 are extremely toxic (Tchounwou et al., 2012).

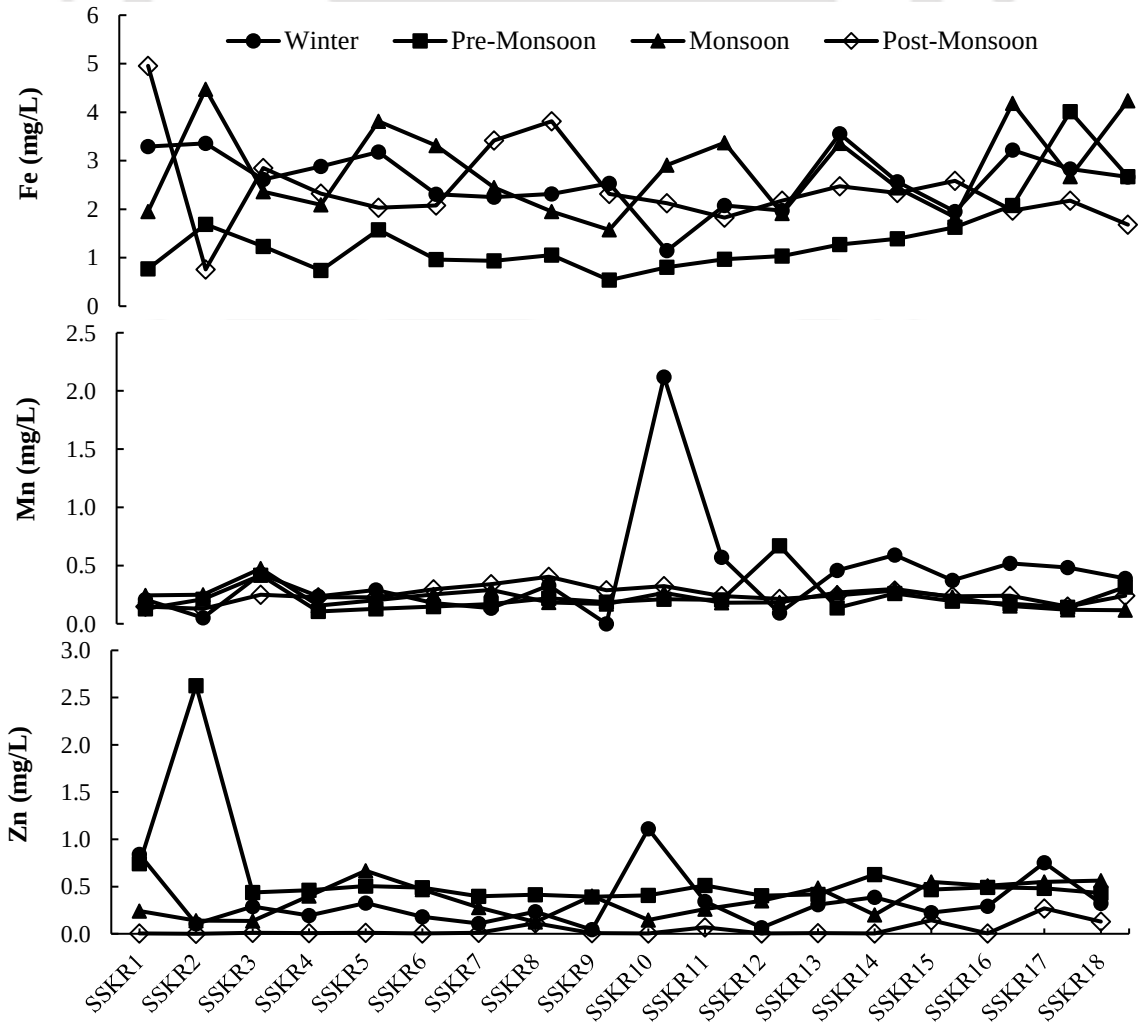
The presence of iron in the riverine system depends on the region's regional geology and the other chemicals present in the river system. Iron is known to be one of man's most critical metal elements. The human body has approximately 3000-5000 mg of iron (Fernández-Luqueño et al., 2013; Landis et al., 1995). As long as iron levels in the environment are not very high (0.3 ppm), they do not affect the human body. But, the iron in industrial processes or environments can cause undesirable problems without adequate wastewater management, e.g. iron hydroxide precipitation can block valves, causing turbidity issues, and iron deposits can cause a foul smell or taste in the water. IS 10500:2012 recommends 0.3 ppm as the acceptable limit in potable water. Fe concentration in the Kolong river varied from 0.54 to 5 mg/L for the entire sampling period (Fig. 4.7). Fe concentration has consistently exceeded the permissible limit for all the seasons for all sampling locations, which indicates Fe is a critical parameter for river water quality. Iron sources in the river are anthropogenic and geogenic, which might be the reason for the high concentration.

Mn is typically found as suspended or dissolved in surface waters at concentrations less than 0.05 mg/L. Mn is objectionable in drinking water supplies for several reasons. Mn stains plumbing fixtures and clothing at concentrations over 0.15 mg / L, creating unpleasant drinks. As with Fe, microbial growth in the distribution system can occur by the presence of Mn in the water. Mn is an essential element for most organisms, but large amounts result in pneumonitis, which affects the central nervous system (Manivasakam, 2016; Mena et al., 1969). IS 10500:2012 recommended a value of 0.1 ppm as the allowable drinking water limit. The Kolong river was found to have the following Mn variations in the entire period of sampling; Winter: 2.12 to 0.05 ppm, Pre-Monsoon: 0.67 to 0.11 ppm, Monsoon: 0.47 to 0.12 ppm, Post-Monsoon: 0.40 to 0.13 ppm. Mn concentrations, for the most part, sampling has exceeded the permissible value. The Winter season has seen the highest Mn concentration, and with the advancement of monsoons, values have diluted to some extent (Fig. 4.7).

Zn is a crucial trace element. Its shortcomings and its abundance are both harmful. Recent studies have shown that zinc is particularly critical for fetal development and infant nutrition (Hotz and Brown, 2004). Excess amounts of zinc can induce cramps, nausea, vomiting, and central nervous system disorder, on the other hand. IS 10500:2012 acceptable value of Zn for potable water is 5 ppm. In the Kolong river, Zn concentration varies from 2.6 ppm to 0.0 ppm (Fig. 4.7), lower than the acceptable value. Winter and pre-monsoon seasons show higher Zn concentration fluctuation than monsoon and post-monsoon seasons. Also, there are higher dilutions in the values in the monsoon and post-monsoon seasons.

Cd is a non-beneficial element and is recognised as highly poisonous. Minute amounts of cadmium are thought to be responsible for harmful changes in the arteries of human kidneys. Cd can cause kidney damage, anaemia, bone marrow disorders, and high blood disorders (Perry et al., 1961). Owing to Cd's toxicity, an acceptable value of 0.005 ppm in potable water was recommended by IS 10500:2012. The concentration of Cd was observed to be higher in the winter season when the concentration exceeded the acceptable value. Still, its value has decreased with the onset of monsoon, which can be observed with the lowered values in the pre-monsoon season (Fig. 4.7). But with its entire rainy season, the Cd level has gone below the detection limit.

Of all the environmental neurotoxins, Pb is considered one of the most dreadful because of its widespread subtle and insidious impact on man. Therefore, it is the most extensively studied toxic metal (Bryce-Smith, 1989). It reaches the human body primarily by ingestion or inhalation of contaminated water and food. An acceptable value of 0.05 ppm of lead in potable water was suggested by IS 10500:2012. There were not many variations in Pb concentrations in the Kolong river during the pre-monsoon and monsoon months. However, after the monsoon, the concentration was higher; due to the declining rainy season, which leads to a decrease in the dilution capacity of the river (Fig. 4.7). The Kolong river was found to have the following Pb concentrations; Pre-Monsoon: 0.09 to 0.04 ppm, Monsoon: 0.05 to 0.01 ppm; Post-Monsoon: 0.092 to 0.07 ppm.



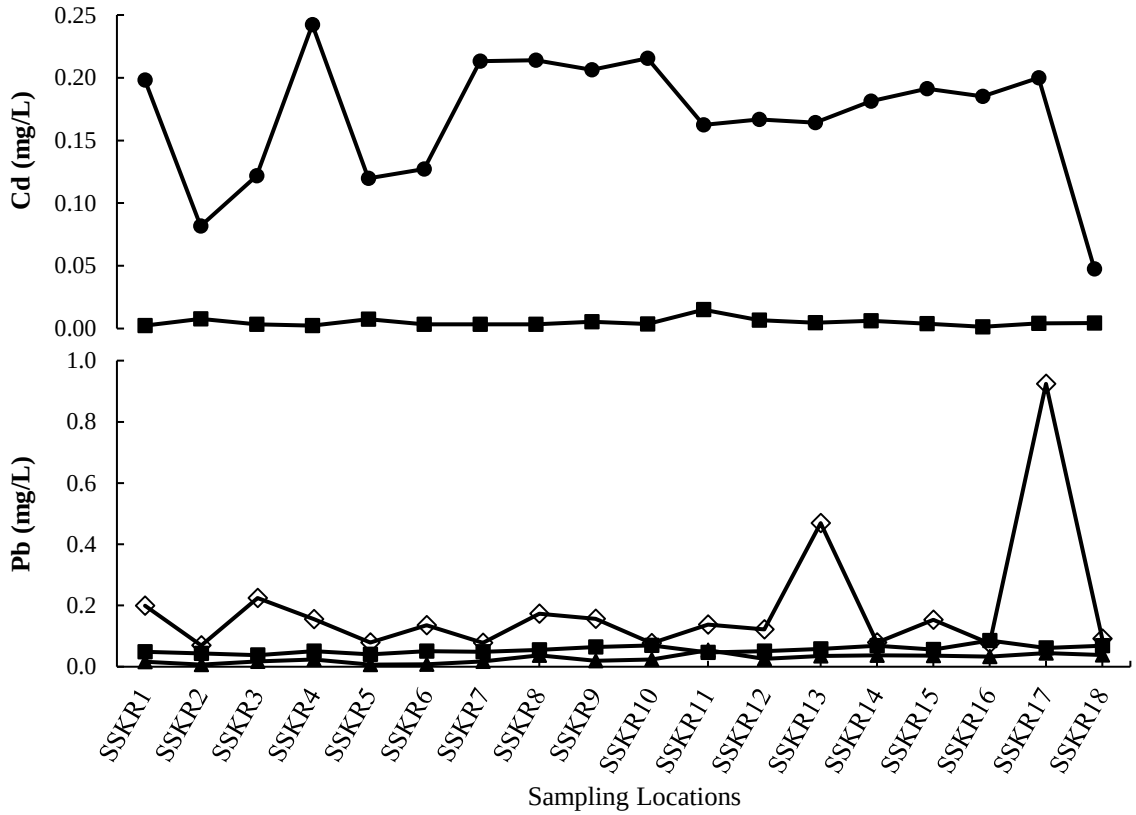


Fig. 4.7. Seasonal variation of Fe, Mn, Zn, Cd and Pb

4.2 Spatial and Temporal Variability of EWQI

The EWQI of all sampling locations in the Kolong river with seasonal variation is shown in Fig. 4.8, and the classification basis is shown in Table 3.3. The classification is based on the fact that the 'excellent' water quality indicates that water quality is directly available for consumption. With the addition of contaminants into the river, water quality deteriorates, and the degree of treatment required for consumption increases. The 'good' water quality is the least treatment required, and 'extremely' poor water quality is the highest treatment for water consumption at that location (Singh et al., 2019). In the winter season, EWQI varies from 109.84 to 241.32, depicting 'average' to 'extremely poor' quality water. During the Pre-Monsoon season, EWQI varies from 68.93 to 165.75, representing 'Good' to 'Poor' water quality. In the Monsoon season, EWQI varies from 67.28 to 114.19, depicting 'Good' to 'Average' water quality. In the Post-Monsoon season, EWQI ranges from 102.36 to 259.91, defining 'Average' to 'Extremely Poor' quality of water. The values obtained after calculating EWQI showed no value falls in the 'Excellent' water quality classification indicating the extent of pollution in the river during sampling (Bora and Goswami, 2015). The sampling sites 'SSKR3', 'SSKR6', 'SSKR8', SSKR14', 'SSKR15', SSKR17', and 'SSKR18' are showing 'Extremely Poor' quality of water during the various seasons of the sampling period. Thus, indicating a few of the probable non-point pollution sources in the river. Also, values observed in the winter decrease in the pre-monsoon and monsoon seasons with an increase in the region's precipitation. During the Post-monsoon season, the EWQI numbers rise as the monsoon season ends since less water is available to dilute the sources of pollutants.

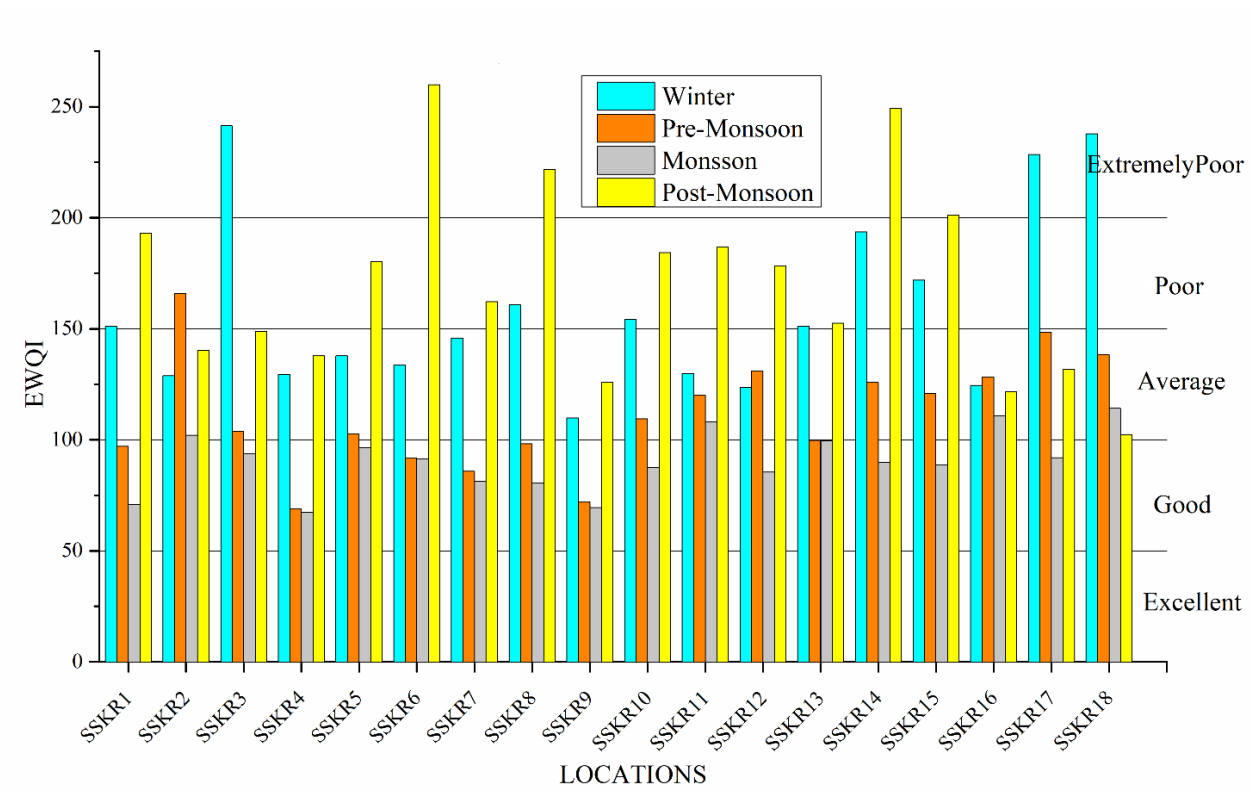


Fig. 4.8. Seasonal variation of EWQI

4.3 Spatial and Temporal Variability Of FWQI

The FWQI of all sampling locations in the Kolong river with seasonal variation is shown in Fig. 4.9 and classified using the output membership function in Fig. 3.7. The classification is based on the fact that higher FWQI values have worse quality and lower FWQI values have better water quality. The results obtained from FWQI were similar to that obtained from EWQI. Post-monsoon had the worst water quality, with most of the values falling in the bad quality and site 'SSKR10' having the worst quality. For most sites, post-monsoon has higher FWQI values followed by winter, pre-monsoon and monsoon, indicating the dilution of FWQI values, i.e. improvement in quality due to high precipitation in the monsoon and pre-monsoon seasons. The results indicate that FWQI is at par with EWQI but can be said to be easy to apply as knowledge of the permissible limits are enough to predict the water quality of a source. No, difficult statistical know-how is required to formulate this index. The basic difference between EWQI and FWQI is that the FWQI results are more stringent than EWQI, which is obvious from Fig. 4.8 and 4.9. The results of FWQI for all the sites vary from bad to worse, whereas EWQI results vary from good to extremely poor. In FWQI, the final rank of any sample is very much affected by toxic parameters. It means that a sample with an acceptable range of all parameters (except one toxic parameter) falls under the unacceptable rank. Therefore, some samples, which have a good quality based on EWQI, show worse quality (higher rank) in FWQI. As a result, in areas where water chemistry shows some toxic elements in the surface water resources, the FWQI classification of water for drinking gives more reliable results.

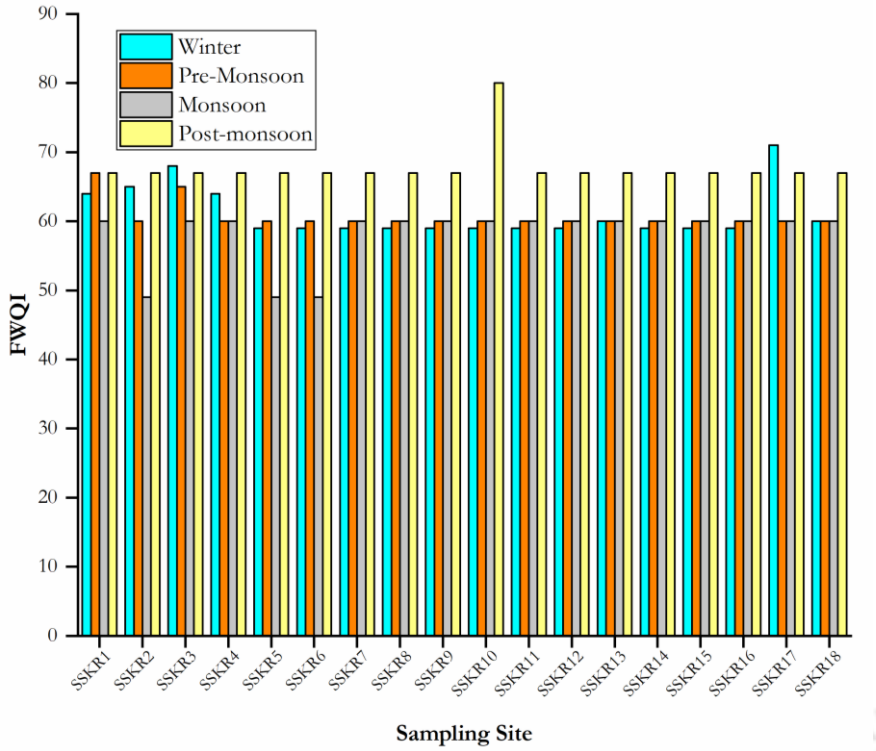


Fig. 4.9. Seasonal variation of FWQI

4.4 Summary

In this chapter, water quality in the Kolong river has been evaluated to understand better the seasonal variance of contaminants and how they shift from dry to wet and back again. After laboratory analysis, water quality parameters were utilised to determine the water quality status in terms of EWQI and FWQI. The following concluding remarks can be made from this study:

1. DO at a few sites have gone below the permissible limit of 4mg/L, EC at upstream sites in the winter and post-monsoon season are higher than the permissible limit of 250 mg/L, BOD_5, i.e., the organic load of the river is higher throughout the seasons. Considering the case of heavy metals, Fe, Mn, Cd, and Pb are also higher than the permissible limit throughout the seasons. The rest of the physicochemical parameters are within their limit throughout the study period. With respect to pollution load, heavy metals present a higher risk to the consumer of the surface water of Kolong river.
2. Rivers that encounter pollution sources (point or non-point sources) in their paths have deteriorating water quality, as observed from EWQI values. The water quality of the Kolong River ranged from 'Good' to 'Extremely poor'. The poorer water quality is attributed to less dilution available in the river, i.e. the dry season, while good water quality was found in the heavy rainfall season. The sampling sites' SSKR3', 'SSKR6', 'SSKR8', SSKR14', 'SSKR15', SSKR17', and 'SSKR18' are showing 'Extremely Poor' quality of water during the various seasons of the sampling period, indicating few of the probable non-point pollution sources in the river. FWQI showed similar but stricter results even though many parameters are within the limit except a few parameters outside it. This implies that the FWQI is much more sensitive to each parameter variation.

Chapter 5

BENTHIC SEDIMENT LOAD ASSESSMENT

In this chapter, sediment load assessment is carried out in terms of heavy metals in two folds. Firstly, pollution load assessment is carried out in terms of total metal content in sediments and secondly, it is carried out in terms of bioavailable and non-bioavailable fractions i.e.in terms of metal speciation fractions measured using a 5-step sequential extraction technique.

5.1 Variation of Total Metals in Benthic Sediments

Total metal concentration at each of the sediment locations is shown in Table 5.1, and their order of variation in different seasons is shown in Fig. 5.1. It is observed that metals in Kolong river sediment followed the trend: Fe > Zn> Mn > Pb > Cd > Cu > Co. According to Bodek et al., 1988, the normal range of Fe is 20,000-550,000 mg/kg in soils. Native Fe concentrations are regionally diverse and can differ significantly in localised regions due to soil types and other sources (USEPA, 2005). In this study, Fe concentration is observed in the range of 1633-23551 mg/kg of sediments, which is in the range provided by the USEPA for iron concentrations in soil (US EPA, 2005). The Mn range in soil ranged between 35 and 1483 mg/kg, far exceeding the FAO/WHO limits for 2 mg/kg (Sakan et al., 2009; WHO, 2013). Pb values varied between 4.94 and 186.09 mg/kg in sediment. These sediment values surpassed both the US (15mg/kg) and the German (70mg/kg) limits (Kittrick, 1971; USEPA, 2002). Cu levels in sediments ranged from 8-46 mg/kg. The requirements were within the European Commission General Director of the Environment (ECDGE) limits (Mortvedt, 1995). The concentration of Zn in the sediment was detected in the range of 0-23550 mg/kg, which extends beyond the acceptable limits of the (EC-DGE) 2010 norm of 100-150 mg/kg (Mortvedt, 1995). Mn, Pb, Cu, and Zn exceeded the limits of presence in soil or sediment prescribed by different agencies, which is a cause of concern keeping in their harmful effects (de Groot, 2018; Mahurpawar, 2015). The maximum values of these metals can cause problems for the aquatic ecosystem and need to be further analysed to understand the extent of anthropogenic or natural contamination in the sediments.

5.2 Speciation of Heavy Metals in Benthic Sediments

The maximum, minimum, and average concentration (mg/kg) of different fractions of heavy metals in the sediment along with standard deviation is shown in Table 5.2, and the mean per-centages of the metals associated with the different phases (Fig. 5.2) are as follows:

- F1: Cu, Mn (19%) > Cd (17%) > Co (13%) >Pb (12%) > Zn (10%) > Fe (3%)

Table 5.1. Maximum and minimum of total metal concentration at sampling locations in benthic sediments

Sampling Locations	Cadmium (mg/kg)		Cobalt (mg/kg)		Copper (mg/kg)		Iron (mg/kg)		Lead (mg/kg)		Manganese (mg/kg)		Zinc (mg/kg)	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
SS1	63.13	0.35	33.54	9.77	39.96	9.13	23550.64	2531.20	164.01	6.74	1138.91	38.44	23550.64	0.35
SS2	61.13	0.87	49.18	9.41	36.19	8.56	21223.41	2248.24	89.64	5.25	999.83	39.58	21223.41	0.87
SS3	63.13	0.65	39.65	0.00	36.19	8.74	22691.48	2298.26	176.36	5.01	1243.22	36.27	22691.48	0.00
SS4	61.53	1.25	44.37	9.42	37.39	9.11	22481.40	2362.16	185.67	5.30	1483.57	35.07	22481.40	1.25
SS5	59.73	0.18	39.99	9.34	45.80	8.64	21534.04	2387.09	136.30	4.94	1006.99	41.31	21534.04	0.18
SS6	60.93	0.46	37.47	9.35	37.99	8.70	22459.58	2456.26	159.76	4.99	1165.71	37.96	22459.58	0.46
SS7	52.63	0.54	39.57	9.88	41.59	9.11	18611.41	2360.17	156.92	7.16	886.93	43.75	18611.41	0.54
SS8	54.63	0.00	32.27	8.51	38.61	8.13	17581.51	1938.57	107.58	5.72	698.78	39.58	17581.51	0.00
SS9	54.63	0.29	45.72	0.00	36.79	8.70	20427.84	1632.62	186.09	6.80	905.22	40.18	20427.84	0.00

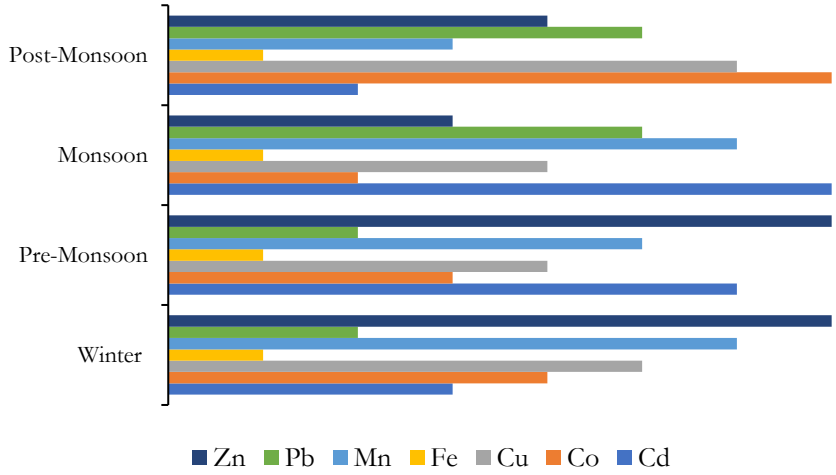


Fig. 5.1. Order of variation of metals in different seasons

- F2: Zn (24%) > Cd (15%) > Pb (12%) > Co (10%) > Mn: Cu (9%) > Fe (4%)
- F3: Fe (38%) > Mn (35%) > Cu (21%) >Zn (16%) > Cd (13%) > Pb (12%) > Co (11%)
- F4: Cd (25%) > Pb (15%)> Cu, Zn (12%)> Co, Mn (7%)> Fe (4%)
- F5: Co (59%) > Fe (51%) > Pb (49%)> Cu (39%)> Zn (38%) > Cu, Mn (30%)

Metals are observed to have a greater association with F1, F3, and F5 phases of sediments, which are well represented with higher percentages. Fe and Mn have a greater association with F1, F3, and F5, indicating anthropogenic and natural sources of pollution, whereas Cu, Pb, and Zn have greater F1 and F2 phases, thus indicating an anthropogenic source. This contribution of pollution sources can further be evaluated using lithogeneous and anthropogenic shares of the metals, which are calculated by adding the exchangeable, carbonate bound, and Fe/Mn-oxide bound phases (Fig. 5.3) (Förstner, 1982; Karbassi et al., 2011). The order of the mean anthropogenic portions of the metals, expressed as a percentage of their mean concentration: Mn (63%) > Zn (50%) > Cu (49%) >Cd (46%) > Fe (45%) > Pb (36%) > Co (34%).

5.2.1 Seasonal variation

In the case of Pb, the organic matter fraction is slightly increased during pre-monsoon and post-monsoon seasons. This may be associated with the acidic pH of the sediment affecting Pb's organic matter fraction (F4) sorption. The speciation analysis (Table 5.2) showed that the major part of Pb is bound to the residual fraction. The prominent residual fraction (F5) presence of the sediment with a small exchangeable (F1) and carbonate fraction (F2) indicates its lower bioavailability (Jordão and Nickless, 1989; Kumar and Thakur, 2017). The Zn speciation (Table 5.2) demonstrates its strongest association with the (Fe-Mn) oxide component of the sediment (F3). From winter to the other seasons, there is also a rise in the concentration of the organic fraction (F4). A small fraction of Zn is associated with the carbonate (F2) and exchangeable fraction (F1) of the sediments. A significant quantity is also related to the sediment's residual fraction (F5). These findings are consistent with (Fe-Mn) oxides demonstrated capacity to scavenge Zn from solution (Gadh et al., 1993; Nriagu and Coker, 1980; Yuan et al., 2018).

The chemical partitioning of Co (Table 5.2) demonstrates its relationship with exchangeable fraction (F1) and a carbonate fraction (F2). A minor fraction (5%) is also related to the sediment's (Fe-Mn) oxide concentration (Baruah et al., 1996). Cu is readily absorbed by sediments, resulting in a high residual level. Sorption rate varies with clay/sediment type, pH, competing cations, and the presence of ligands and Fe/Mn oxides. The Cu speciation reveals that it is mostly connected with the Fe-Mn oxide fractions of the sediments. However, it is also associated with a significant fraction of the organic matter content. This is possible because of the complex formation with organic matter, particularly during the monsoon period. A significant amount of Cu is affiliated with the residual fraction (Jha et al., 1990; Maharana et al., 2018). High percentages of Cd are found in exchangeable (F1) and carbonate (F2) fractions in different seasons.

In contrast, lower percentages are found in organic matter (F4) and Fe-Mn oxide (F3) fractions, but the dominant fraction in the case of Cd is also the residual fraction. These findings suggest that controlling metal concentration in the soluble phase relies heavily on Cd adsorption. Cd is effectively absorbed by Fe-Mn oxides (Chester, 2018; Gadde and Laitinen, 1974; van der Weijden, 1976) and organic materials of natural origin (Gardiner, 1974; Mikhailenko et al., 2020). A considerable percentage of Cd from exchangeable and carbonate fractions suggests that after salinity increases or slight pH decreases, a substantial percentage of Cd particles is solubilised and easily accessible (Fernandes et al., 2019; van der Weijden et al., 1977). According to speciation studies, Mn compounds are found primarily in the first three fractions (>50%) in all seasons, with higher percentage fractions found in the pre-monsoon and monsoon seasons. Furthermore, similar to Fe, Mn is weakly absorbed from organic materials. It is possible that Mn bioavailability in sedimentary organic materials is lower. In addition, it may be established that exchangeable Mn species are plentiful in Kolong river sediments due to industrial and agricultural activity (Akcay et al., 2003; Rajeshkumar et al., 2018). Speciation studies of Fe show that the highest form is found in residual fraction (F5) and oxide fraction (F3). Mn, Cu, and Cd have the highest exchangeable fractions, while Fe and Co have the highest in the residual (Fig. 5.2). The appreciable quantity of metals in the exchangeable fractions is a cause of concern for the aquatic environment as the fractions indicate anthropogenic contamination (Gadh et al., 1993; Nriagu and Coker, 1980).

From Tessier's 5-step sequential extraction, residual fraction or F5 fraction dominates the metal distribution in sediments in all the different seasons except for Zn (winter), Mn (Pre-monsoon) and Pd (post-monsoon) (Table 5.2). This indicates that the heavy metals are also locked up in the lithogeneous phase (Sundaray et al., 2011). After F5 fraction, F3 dominates the metal distribution in the sediments and F4, F2 and F1 have equitable distribution. This metal fraction (F5) is inert and contained within the crystalline structure of some primary and secondary minerals (Ramos-Gómez et al., 2011). Under normal environmental conditions, the metals associated with this fraction are stable and insoluble. As a result, they have little effect on surface water quality and are not biologically accessible (Idriss and Ahmad, 2013). All the metals also have an appreciable proportion in F2 and F3 fractions due to metals affinity toward carbonates and Fe-Mn oxides (Sundaray et al., 2011). For all the metals, F5 and F4 combined, i.e. non-bioavailable fractions, are dominant species in the sediments in the winter season. While with the onset of monsoon, the proportion of bioavailable fractions, i.e. the F1 and F2 fractions, increases to an appreciable quantity due to the turbulence in the river stream (Aguilar-Hinojosa et al., 2016). This could pose a higher risk for the aquatic environment as these fractions easily into the river stream and the aquatic biota (Al-Mur, 2020; Liu et al., 2017; Sun et al., 2019). This also indicates that metals from the stream might have adsorbed onto the sediments, leading to a decrease in metal concentration in the stream (Akindele et al., 2020; Akindele and Olutona, 2014).

Table 5.2. Distribution of heavy metal (%) in various metal speciation fractions in the Kolong river sediments

		Winter					Pre-monsoon					Monsoon					Post-monsoon				
		F1	**F2**	**F3**	**F4**	**F5**	**F1**	**F2**	**F3**	**F4**	**F5**	**F1**	**F2**	**F3**	**F4**	**F5**	**F1**	**F2**	**F3**	**F4**	**F5**
Zn	**Min**	6.46	26.66	2.30	2.16	39.73	4.06	10.92	14.57	9.32	13.32	3.67	6.80	16.71	16.28	23.45	0.50	12.06	22.59	18.21	9.94
(%)	**Max**	11.60	38.89	12.71	8.79	58.54	22.37	22.71	24.05	43.93	35.58	24.14	13.99	36.50	24.43	40.48	14.79	24.14	43.42	33.72	32.07
	Average	9.03	32.54	6.67	4.03	47.73	15.26	16.63	21.62	23.25	23.24	11.73	9.83	26.35	19.36	32.72	8.35	17.92	30.53	23.34	19.86
	Std. Dev*	1.95	4.30	3.57	2.07	6.09	5.85	4.14	2.81	9.41	6.75	6.23	2.22	6.96	2.52	5.78	4.19	4.67	6.93	4.67	7.07
Mn	**Min**	19.07	8.03	12.18	15.59	22.39	16.42	2.45	36.49	1.23	22.73	12.82	8.92	25.45	6.86	20.24	9.43	4.57	16.08	12.25	36.15
(%)	**Max**	27.87	16.86	16.56	30.20	34.68	26.43	8.51	48.13	2.54	32.79	23.69	15.03	41.22	18.99	39.43	17.27	11.50	32.45	18.05	48.64
	Average	23.90	11.98	14.21	21.05	28.86	20.95	6.09	43.20	1.77	27.99	15.99	11.69	32.29	10.14	29.89	13.25	6.79	25.29	15.13	39.54
	Std. Dev*	3.17	3.03	1.54	4.29	4.42	3.50	1.85	4.27	0.38	3.31	3.44	2.00	5.79	3.58	6.87	2.64	2.19	4.70	2.05	4.03
Pb	**Min**	9.09	7.87	9.76	21.91	24.27	5.69	5.85	8.95	12.93	52.32	12.91	12.67	11.87	11.46	45.57	9.68	12.36	15.06	2.71	44.24
(%)	**Max**	21.50	14.10	14.55	42.09	36.56	10.89	12.61	11.85	17.17	58.73	14.55	14.55	14.32	13.43	49.14	16.31	15.33	23.43	4.67	56.89
	Average	16.42	10.36	11.63	30.90	30.69	8.36	10.36	10.39	14.80	56.10	13.90	13.46	12.91	12.66	47.07	13.48	13.52	19.56	3.78	49.66
	Std. Dev*	3.41	2.39	1.68	6.49	4.19	1.77	1.93	1.03	1.47	2.09	0.61	0.56	0.85	0.63	1.43	2.13	1.21	2.27	0.66	4.02
Fe	**Min**	1.10	3.17	38.02	6.86	26.87	0.80	1.84	34.50	1.19	43.13	3.20	2.43	20.37	2.12	41.06	5.38	8.87	16.13	9.53	46.40
(%)	**Max**	1.95	8.44	53.28	15.32	47.97	1.50	2.39	51.50	2.87	61.31	6.76	5.69	42.83	6.03	70.12	7.56	12.45	26.11	11.63	54.49
	Average	1.55	4.85	43.27	9.66	40.67	1.11	2.08	41.50	1.68	53.63	5.09	4.43	32.86	3.08	54.53	6.71	11.05	21.14	10.58	50.53
	Std. Dev*	0.30	1.86	5.39	2.86	7.31	0.26	0.22	5.58	0.51	5.76	1.25	1.12	8.90	1.20	11.13	0.61	1.32	3.41	0.87	2.99
Cu	**Min**	15.11	6.26	9.18	14.82	29.40	13.49	3.00	17.67	4.65	30.48	9.39	8.49	11.40	8.77	36.10	15.96	13.06	19.75	12.71	19.24
(%)	**Max**	27.96	18.69	12.98	27.96	45.75	25.73	8.27	33.07	9.14	51.87	17.38	14.01	18.87	20.22	53.34	27.10	15.97	23.79	15.71	34.54
	Average	21.27	11.00	11.29	18.72	37.72	19.05	5.36	26.51	7.23	41.86	12.27	11.12	16.26	15.85	44.50	20.77	14.40	21.66	14.03	29.15
	Std. Dev*	4.45	3.83	1.56	4.09	5.11	3.53	1.66	4.71	1.54	6.69	2.35	1.71	2.57	3.57	5.59	3.77	1.06	1.31	1.07	4.48
Co	**Min**	10.98	10.83	10.62	20.41	22.94	9.55	9.33	5.68	1.96	64.34	8.27	7.71	13.09	5.10	53.00	28.18	6.63	9.84	6.28	31.46
(%)	**Max**	19.17	20.99	13.87	37.70	36.29	16.05	10.94	8.76	3.80	70.07	9.88	10.40	19.62	11.13	63.16	42.83	10.05	13.18	8.63	49.07
	Average	14.72	14.87	12.46	27.53	30.42	12.32	10.21	6.96	2.89	67.63	9.15	9.25	14.99	8.06	58.55	31.87	8.29	11.53	7.39	40.92
	Std. Dev*	3.05	3.76	1.20	5.36	4.92	2.52	0.54	0.95	0.50	2.25	0.57	0.81	1.97	1.67	2.92	4.29	1.10	1.12	0.70	5.05
Cd	**Min**	11.55	8.21	11.71	19.58	23.52	13.06	10.23	11.96	5.51	38.85	20.97	17.10	3.42	6.84	29.39	3.62	10.12	6.55	5.76	55.15
(%)	**Max**	20.09	19.16	14.87	38.02	38.11	17.59	18.57	32.36	19.28	47.09	30.91	27.07	13.86	9.06	45.37	12.38	14.19	14.62	13.19	63.13
	Average	15.57	13.08	12.98	26.73	31.64	16.03	15.46	15.83	9.09	43.59	26.62	21.64	6.89	8.16	36.69	6.78	12.21	12.32	10.30	58.39
	Std. Dev*	3.20	4.19	1.27	5.74	5.48	1.82	2.49	6.28	3.95	2.48	3.11	3.50	3.44	0.64	4.98	2.36	1.64	2.33	2.06	2.85

*Standard Deviation

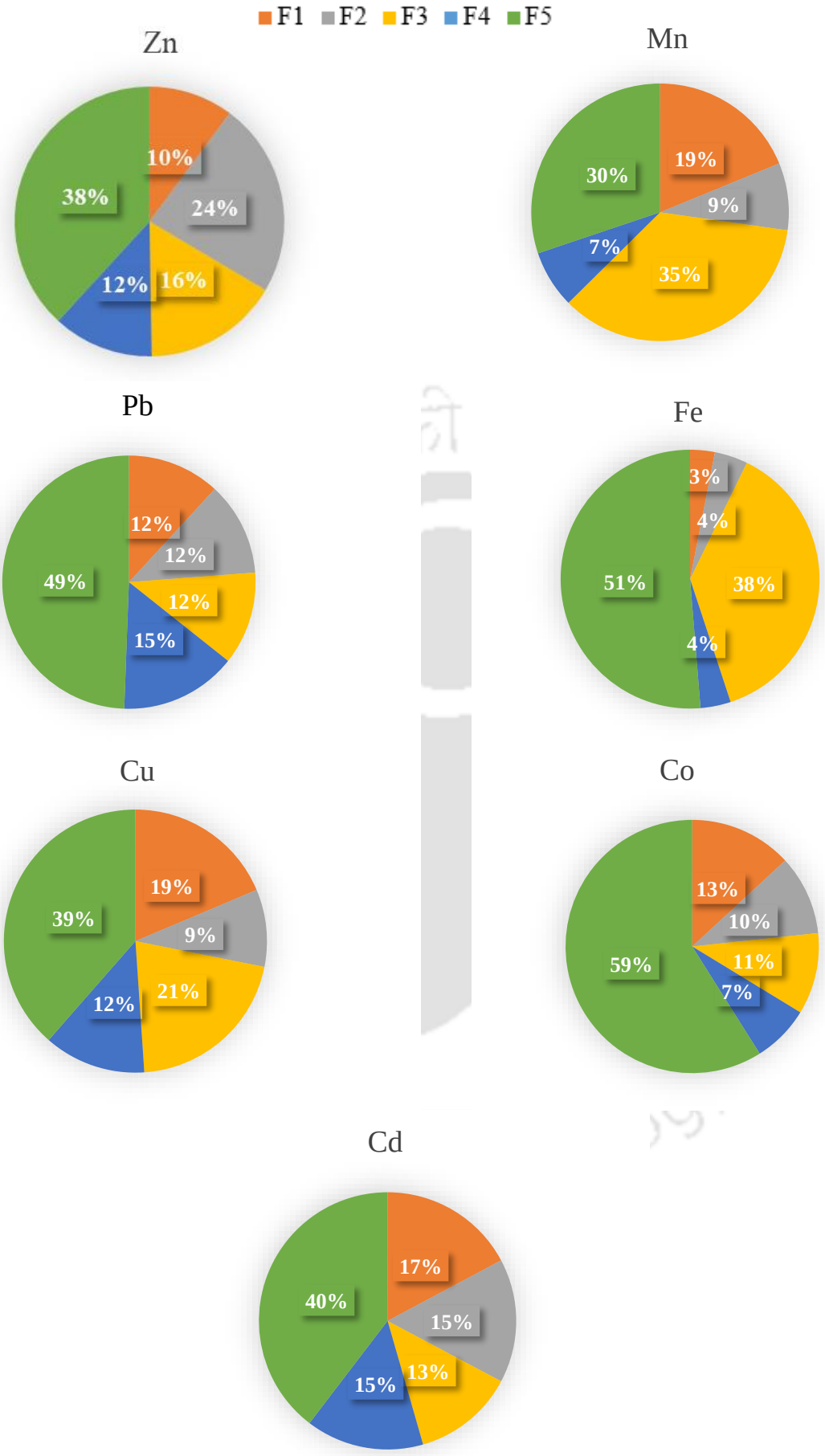


Fig. 5.2. Presence of metals in different sedimentary phases or chemical fractions in benthic sediments

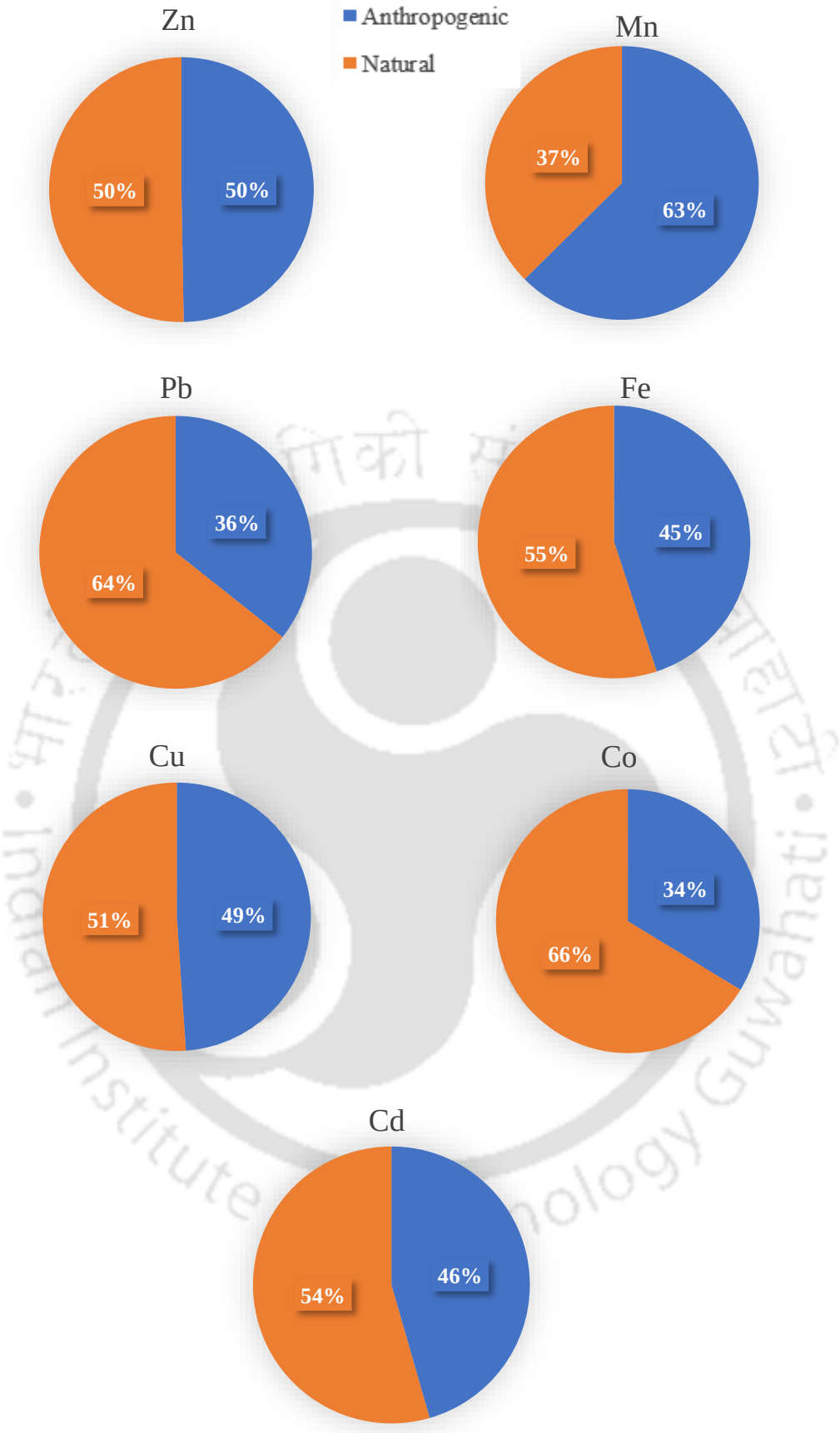


Fig. 5.3. Pie chart showing anthropogenic and natural fractions of metals in benthic sediments of the river

5.3 Particle size and Metal Speciation Factions

The particle size of sediments was analysed using a laser particle size analyser (Make: MALVERN, model: HYDRO 2000 MU). Mean particle size ($d_{0.5}$) sediments are plotted against metal speciation fractions, and the results are shown in fig. 5.4. For Cd, it is observed that smaller sediment particles have a higher F5 fraction indicating historical presence of Cd. For Co, it can be observed that there is a high proportion of F1 and F5 for the smaller size of sediment particles. In the case of Cu and Fe, F3 and F5 are higher in proportion in the smaller size of sediments. However, in the case of Pb and Zn, there is an equitable proportion for all the fractions, still higher proportions of F3, F4 and F5 in smaller sediment particles. From this, it can be said that F5 is found to be high in the case of the smaller size of sediments except for Mn, wherein the F3 fraction is high. Also, along with F5, a high concentration was observed in the F3 fraction in cases of Cu and Fe. So, it can be inferred from the particle size study that in the case of Cu, Fe and Mn, there is some form of anthropogenic and natural pollution during the studied period, whereas in the rest, mostly the rest of the metal pollution is due to natural sources. However, principal component analysis and positive matrix factorization need to be carried out to confirm the pollution sources.

5.4 Sediment Load Assessment

Sediment load assessment is carried out in terms of heavy metals in two folds. Firstly, pollution load assessment is carried out in terms of total metal content in sediments and secondly, it is carried out in terms of bioavailable and non-bioavailable fractions i.e.in terms of metal speciation fractions measured using 5-step sequential extraction technique.

5.4.1 Via total heavy metal concentrations in the benthic sediments

5.4.1.1 Pollution Load Index (PLI)

PLI values have been calculated and observed in the range of 0.48-1.89 and are shown in Fig. 5.5. It has been observed that in August '18, September '18, December '18, March '18, May '19, June '19, and October '19, the PIL value is >1, with the highest value being found in June '19 and October '19. This is due to the rivers' high flow and turbulent condition, which allows all the metal concentrations to get adsorbed on the surface of sediment particles. Additionally, ample amounts of chemical pesticides and herbicides are used during the peak period for harvesting tea leaves (Golian et al., 2020; Hazarika and Bhuyan, 2013; Ostad-Ali-Askari and Shayannejad, 2021). PLI values for the samples collected in 2019 were higher than those collected in 2018. This can be associated with the anthropogenic addition of pollutants in the river stream, which ultimately get deposited in the sediments over a period of time. High values of PLI are also observed in the drier months like Dec'18 and Mar'19, which indicates that during the low flow, metals tend to precipitate on the surface of sediments and get attached, thereby increasing the metal concentration (Kaushik et al., 2009). The values are PLI are observed to vary similarly in the year 2018 and 2019, with monsoon and pre-monsoon months having lower pollution load than post-monsoon

and winter months due to lower turbulence in the post-monsoon and winter months. This is because heavy metals tend to settle on the river bed due to lower turbulence in these months, giving higher PLI. The values of PLI in the year 2018 are less than those in 2019, indicating continuous anthropogenic addition of pollutants in the river giving rise to higher PLI values.

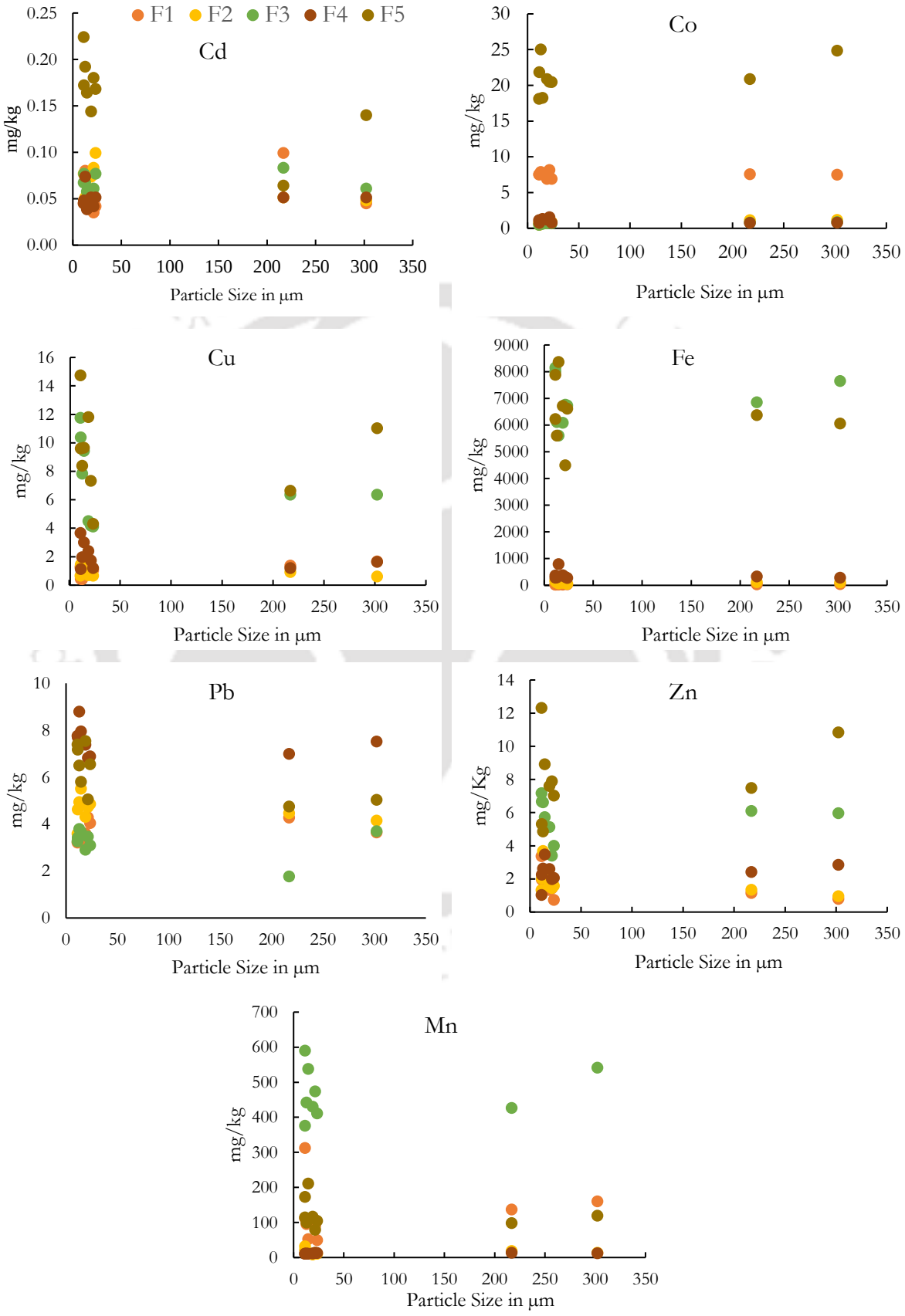


Fig. 5.4. Particle size and metal speciation

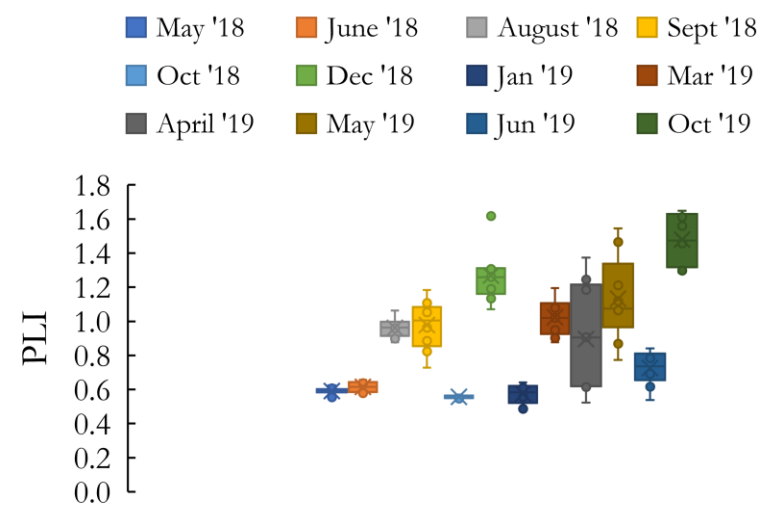


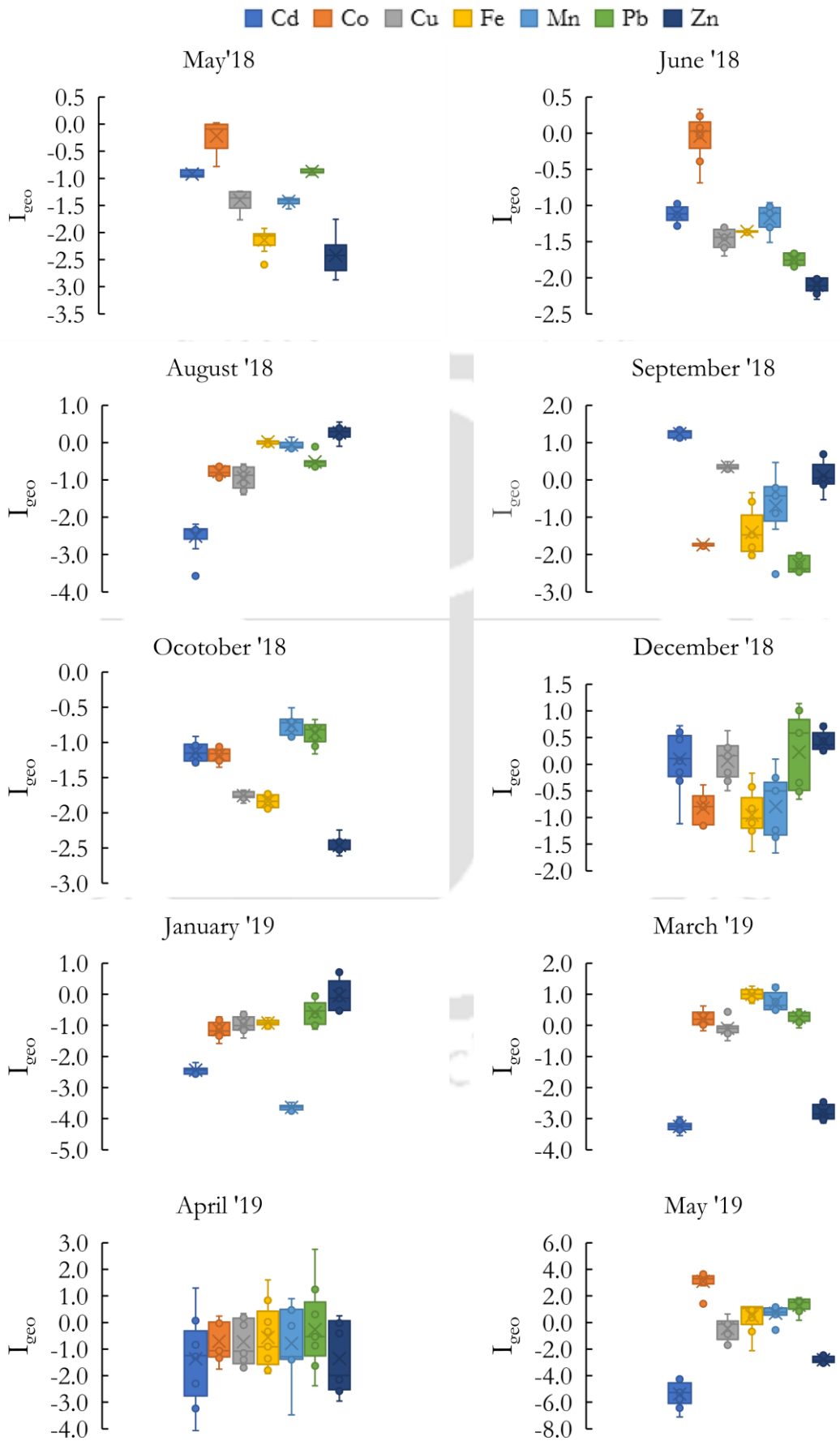
Fig. 5.5. Variation of PLI values

5.4.1.2 Geoaccumulation Index (I_{geo})

The I_{geo} values calculated are shown in Fig 5.6. Sediments are mostly found in the range of moderately polluted due to cobalt in May and June '18. In August '18, Zn was a significant contributor to pollution varying in the range of 0.29-0.56, which is moderately polluted, which can also be observed from a high anthropogenic fraction of the metal. In September '18, Cd was the main contributor to pollution, and in December '18, Pb was the major contributor of pollutants. In January '19, sediments were found to be moderately polluted. In March '19, Fe and Mn were the highest contributors to pollution according to the I_{geo} values, followed by Pb and Co. The variation in the concentrations of the metals is likely to be due to the runoff water released in the rainy season from the agricultural field and tea garden, which uses chemical fertilisers, pesticides, and herbicides during the rush cropping periods for tea production in April to June and October to December (Bora and Goswami, 2014). In May'19, all the sites were uncontaminated to moderately contaminated with each metal analysed except for Pb, which is moderate to heavily contaminated. Again with onset of the monsoon, i.e. June' 19, I_{geo} values decreased below the previous months and increased in October after the high precipitation period ended, except for Co, which has higher values than the rest of the metals. Co and Pb have significant load in the sediments of the river due to anthropogenic pollution, as the metals have high anthropogenic fractions in the sediments as compared to lithogenic fractions (Karbassi et al., 2011).

Also, it is observed during the monsoon months, i.e. May'18 and May'19, June'18 and June'19, I_{geo} values for most metals are less than 1, whereas in the rest of the months are greater than 1. This variation can be attributed to the fact that in the monsoon months, due to high flow in the river, i.e. turbulent flow, the metals tend to release the sediment particles and move into the water, thereby increasing the pollution load of the river. After the monsoon season, the exact opposite happens wherein the stillness of the river, metals tend to deposit onto the sediment particles and get adsorbed. For this reason, the values in the driest months, i.e. December and January, are

greater than 1. However, for Pb and Co, there is a constant increase in I_{geo} values irrespective of the season. This might be because of the anthropogenic pollution sources existing in the region, continuously adding pollutants to the river, leading to an increase in I_{geo} values.



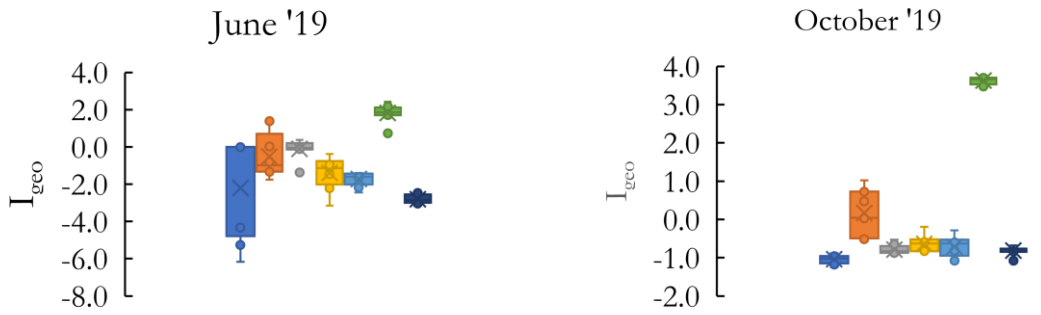


Fig. 5.6. Variation of I_{geo} values

5.4.1.3 Enrichment Factor (EF)

EF predicts the ecological degradation of the Kolong River due to the impact of anthropogenic activities. EF estimation is based on elementary concentration and reference in the earth's crust of the sediment to the Fe concentration and is shown in Table 5.3 (Lim et al., 2013; Salomons and Forstner, 2012). Among the metals found in Kolong River sediments (Cd, Co, Cu, Fe, Mn, Pb, Zn), Co (42.51: May'19) has the highest EF value observed during the period of study, which falls in the category of very high enrichment. Co (23.24: June'19, and 9.36: September'18), Pb (29.15: June'19, 12.43: May'19, and 5.26: May'19), Cu (9.82: June'19, and 5.32: September'18), Cd (9.36: September'18) and Mn (7.45: May'19) have significantly enriched values in the sediments of Kolong river suggesting an anthropogenic source of the metal which can be directly correlated with the presence of higher fractions of metals relating to anthropogenic contamination. Mn 63%, Cu 49%, Cd 46%, Pb 36%, and Co 34%, respectively, have higher contamination fractions, which is anthropogenic, leading to higher enrichment factor (Karbassi et al., 2011). In the monsoon months, i.e. in May'19 and June'19, Co and Pb have the highest EF values in the river. This indicates that the monsoon brings anthropogenic pollutants into the river, leading to river ecosystem degradation. EF values observed over two years indicate that the pollution level with respect to benthic sediments is increasing and can cause serious environmental degradation. Thus, serious attention should be given to sediment pollution in the Kolong river.

5.4.1.4 Potential Risk of Individual Metal (E_r^i) and Potential Ecological Risk (PERI)

Table 5.4 shows E_r and PERI values for all months of sampling. The maximum values for Co (113.66, 74.30) in September and December and Pb (42.65) in June are rated as 'moderate ecological risk', which can also be directly related to high anthropogenic fractions. The remaining values of E_r for Cd, Cu, Pb and Zn were found below 40, indicating low ecological risk. PERI values were less than 150, categorising the sites as low ecological risk in different sampling months (Hakanson, 1980; Teodorof et al., 2020; Zhang et al., 2016). Although PLI, I_{geo}, EF and E_r suggested some form of anthropogenic contamination of the Kolong River sediments, PERI levels showed a low ecological risk. This indicates that even though the metals have higher anthropogenic contamination in the sediments, they have not exceeded the permissible sets of different agencies.

Table 5.3. Maximum values of EF in different months

Sampling Months	Enrichment Factor (Maximum Value)					
	Cd	Co	Cu	Mn	Pb	Zn
May'18	3.36	5.71	2.16	2.20	3.23	1.78
June'18	1.31	3.21	1.05	1.31	0.82	0.65
August'18	0.21	0.66	0.66	1.14	0.95	1.43
September'18	9.36	1.28	5.32	3.14	1.00	4.47
October'18	1.76	1.86	1.21	2.74	2.26	0.82
December'18	3.54	1.54	4.05	1.69	4.67	4.00
January'19	0.41	1.02	1.15	0.17	2.08	3.55
March'19	0.06	0.72	0.56	1.23	0.69	0.09
April'19	3.11	3.85	2.52	1.75	5.26	2.64
May'19	0.05	42.51	3.98	7.45	12.43	0.78
June'19	0.53	23.24	9.82	3.39	29.15	1.61
October'19	1.33	1.19	1.39	1.57	1.70	1.31

Table 5.4. Maximum values of E_r and PERI in different months

Sampling Month	Er				PERI
	Cd	Cu	Pb	Zn	
May'18	25.19	3.18	4.26	0.44	32.75
June'18	23.00	3.04	2.42	0.38	28.24
August'18	9.87	5.03	6.95	2.20	22.56
September'18	113.66	10.53	1.95	2.46	126.92
October'18	23.85	2.34	4.69	0.32	30.03
December'18	74.30	11.60	16.48	2.54	94.02
January'19	9.81	5.11	7.64	2.46	21.78
March'19	5.87	10.12	10.77	0.28	27.00
April'19	6.50	7.72	5.27	0.97	18.12
May'19	23.48	8.12	30.88	0.25	59.47
June'19	2.45	2.18	42.65	1.86	48.23
October'19	24.45	5.19	9.77	0.94	39.33

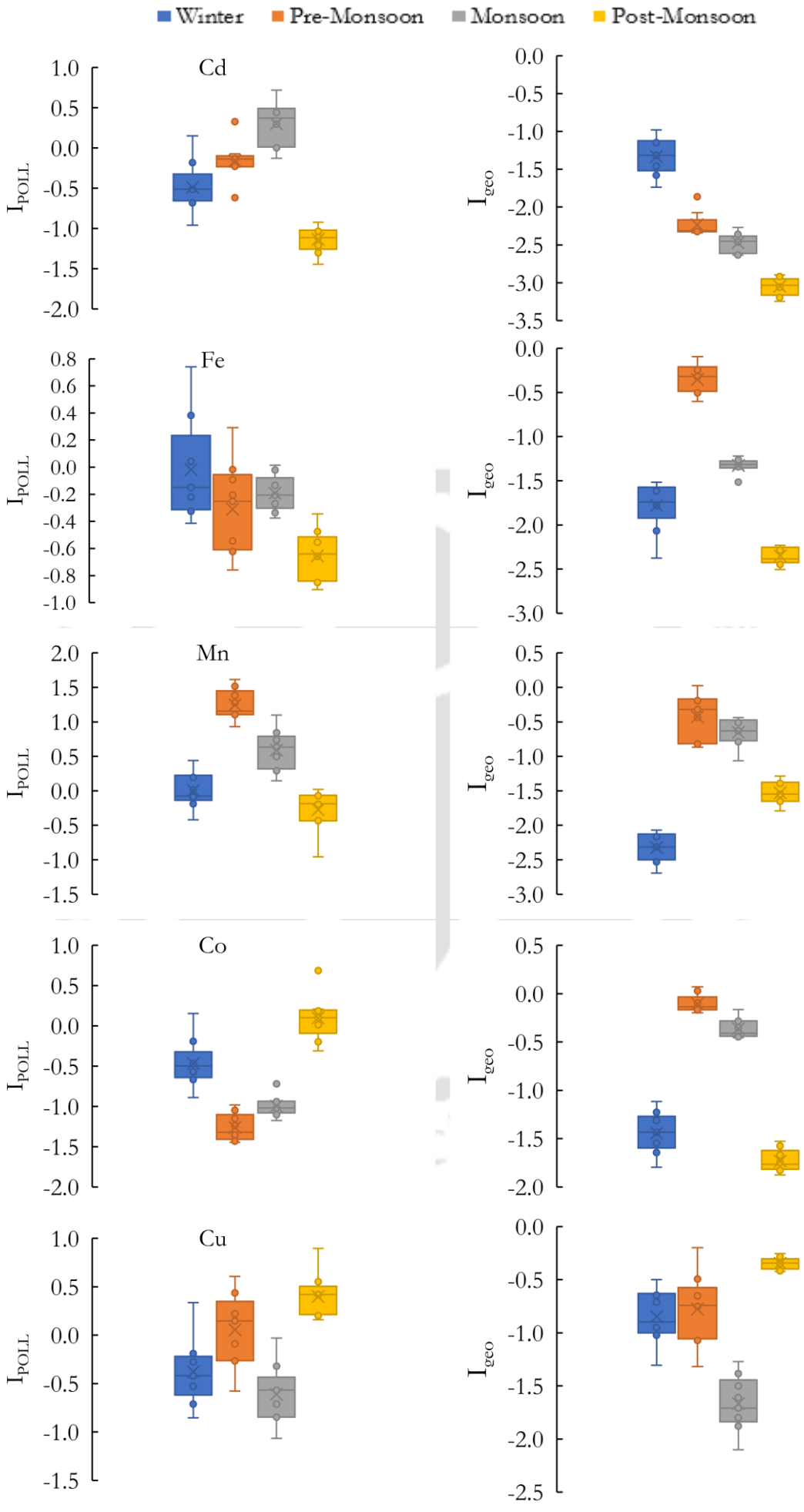
5.4.2 Via metal speciation fractions in the benthic sediments

5.4.2.1 Pollution Index (I_{POLL})

Seasonal variations of I_{POLL} and I_{geo} are shown in Fig. 5.7. The I_{POLL} index values show a wide range of contamination (from 'uncontaminated' to 'moderately uncontaminated'). As a result, a considerably clearer picture of anthropogenic contamination is presented (Karbassi et al., 2008; Roudposhti et al., 2016). Staring with Cd, I_{POLL} initially increases with the monsoon and then recedes with the receding monsoon. Similar variation in being is also shown in the case of Fe and

Mn. So, for these metals, increased turbulence leads to metals getting attached to sediment particles from the water column of the river. Whereas, in the case of Co, Cu, Pb and Zn, the pollution decreases in the monsoon season and increases in the pre-monsoon season. This can be attributed to the fact that in the monsoon season when there is high turbulence in the river, attached metals get desorbed from the sediment particles and thus have lower value in monsoon season.

I_{POLL} is a modification of I_{geo} index by considering bioavailable fractions with respect to non-bioavailable fractions, as non-bioavailable fractions are hardly released into any ecosystem due to bond with the sediment particles and causing very less harm to the aquatic environment. As it is well established in the literature that an increase in bioavailable fractions poses a rather than just increase in metal concentration. The effectiveness of I_{POLL} needs to be understood by comparing its values with the I_{geo}. For Cd, I_{POLL} values increase after winter with advancing monsoon rainfall and then decrease with receding monsoon, whereas in the case of I_{geo}, there is a constant decrease from winter to post-monsoon season due to a decrease in overall concentration of Cd, but the proportion of bioavailable fraction increase initially (Table 5.2) than decrease which I_{POLL} first increase than decrease. For Fe, I_{geo} values are less than 0, which indicates the sediment pollution due to metals is on an 'uncontaminated' level. However, when I_{POLL} values are looked upon, it is observed that it indicates 'uncontaminated' to 'moderate' pollution level for winter amd pre-monsoon season, and then pollution level decreases for the rest. This is due to higher bioavailable fractions in winter and pre-monsoon season, and then the release of bioavailable fractions takes place (Table 5.2), due to which metal pollution decreases. Similar variations were observed in Mn, wherein I_{geo} show an uncontaminated pollution level, but I_{POLL} values show 'uncontaminated to moderate' contamination and 'moderately contaminated' pollution level. The higher risk shown by I_{POLL} is due to higher bioavailable fractions (Table 5.2). The seasonal variation depicted in cases Cd, Fe, and Mn shows that in monsoon months, due to turbulence in the river, metal pollution increases and then decreases in the rest of the season. In the case of Co and Cu, the variation shown by I_{geo} and I_{POLL} are exactly the opposite. I_{geo} shows an increase in pollution trend during monsoon and then decrease in the rest, but I_{POLL} values decreasing trend in monsoon and an increasing trend in the rest. This can be directly linked to the proportion of bioavailable fractions in monsoon compared to rest of the season. Thus, it can be said that for these metals during monsoon, i.e. during high turbulence period, the proportion of bioavailable fractions decreases, indicating release of metals from sediment particles, thereby decreasing the pollution level. Till now, all metals considered have higher pollution levels with I_{POLL} than with I_{geo}. But for Pb and Zn exact opposite has been observed, i.e. lower I_{POLL}, which is due to lower bioavailable fractions in the case of these two metals. This is because I_{geo} is based on shale concentrations rather than their physicochemical fractions, which might vary from one place to another.



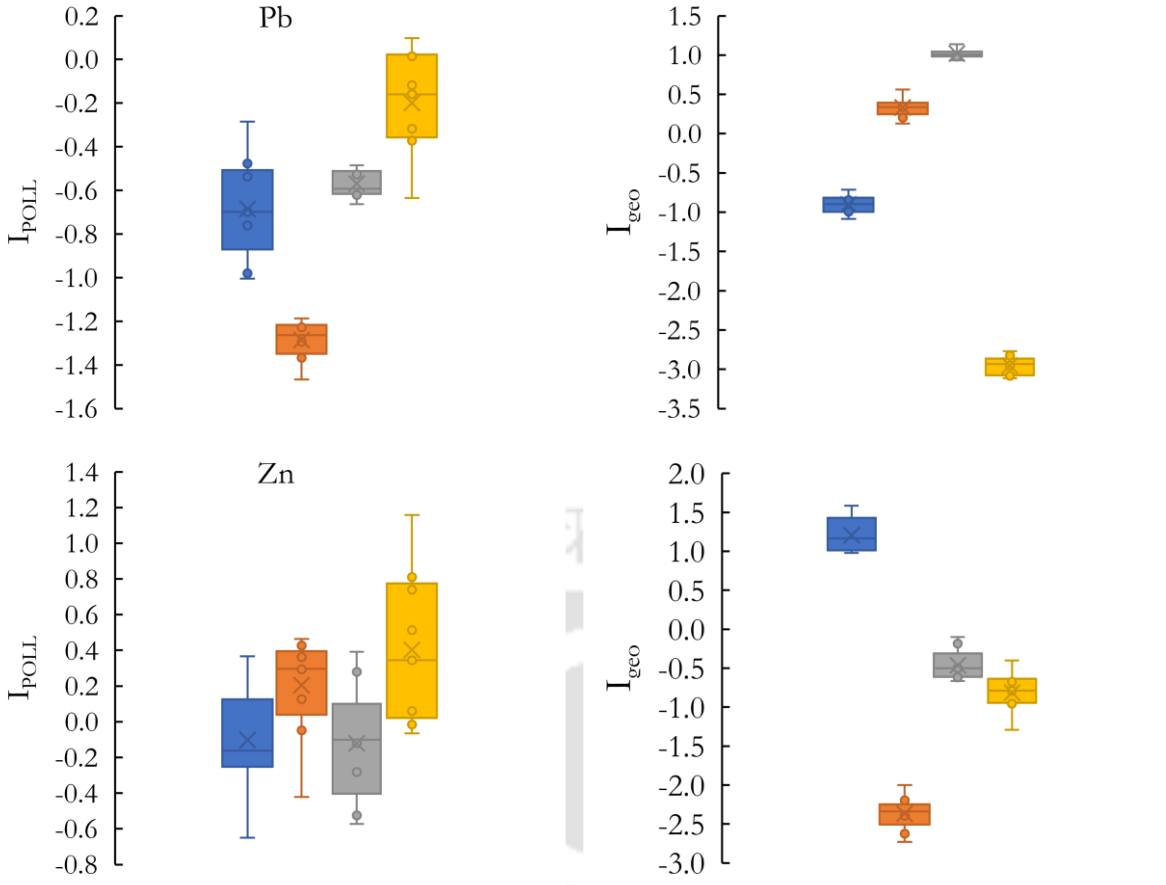
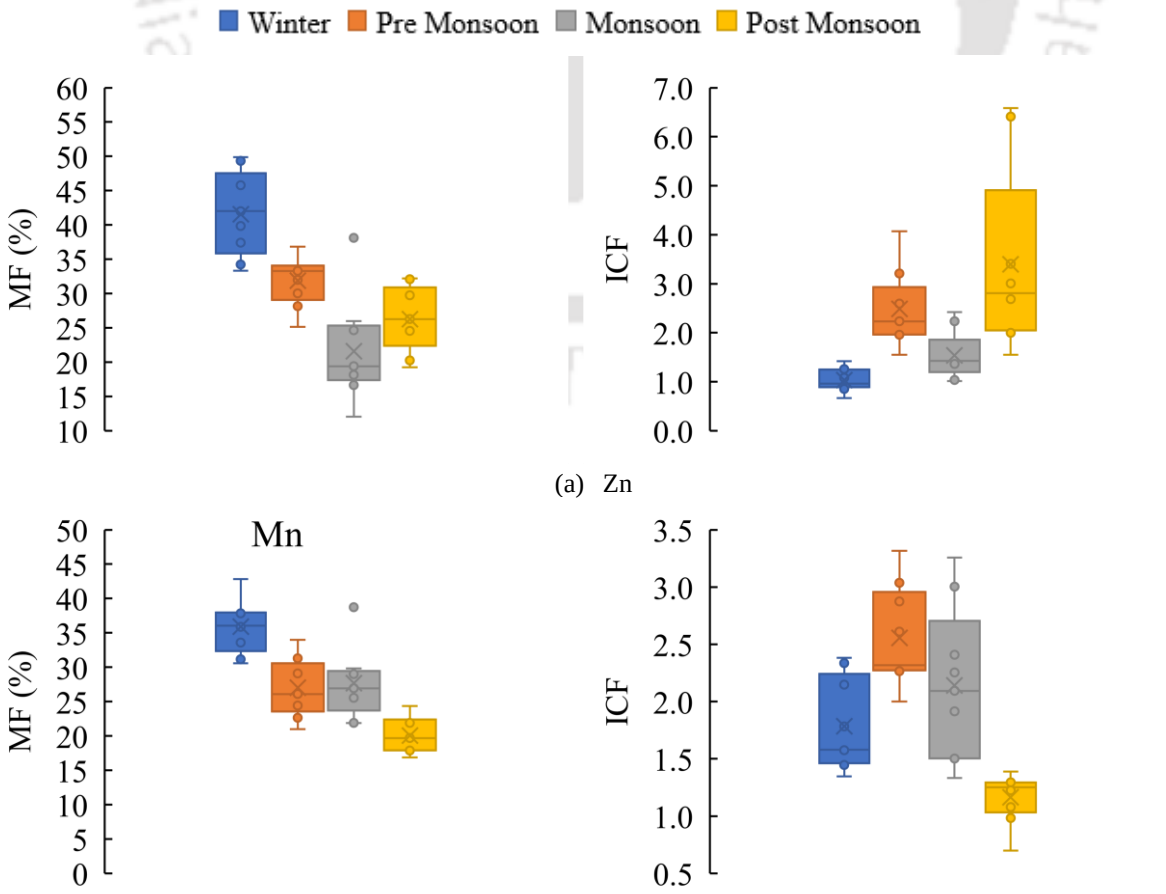


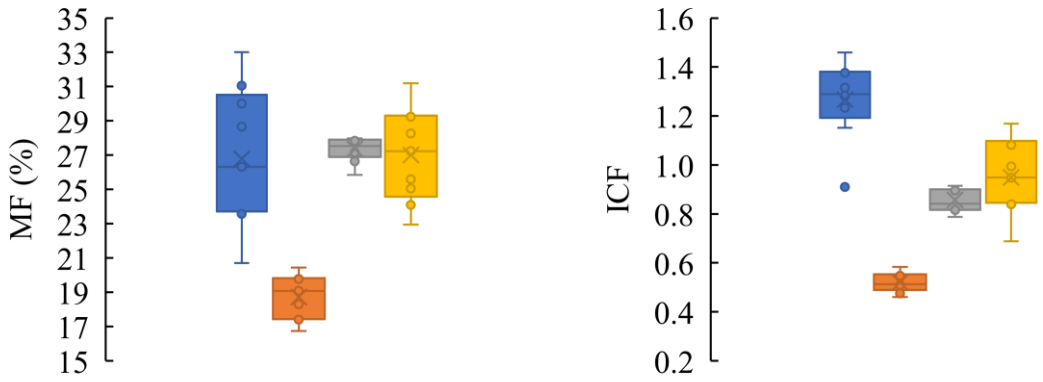
Fig. 5.7. Comparison of seasonal variation of I_{POLL} and I_{geo} values

5.4.2.2 Mobility Factor (MF), Individual Contamination Factor (ICF) and Global Contamination Factors (GCF)

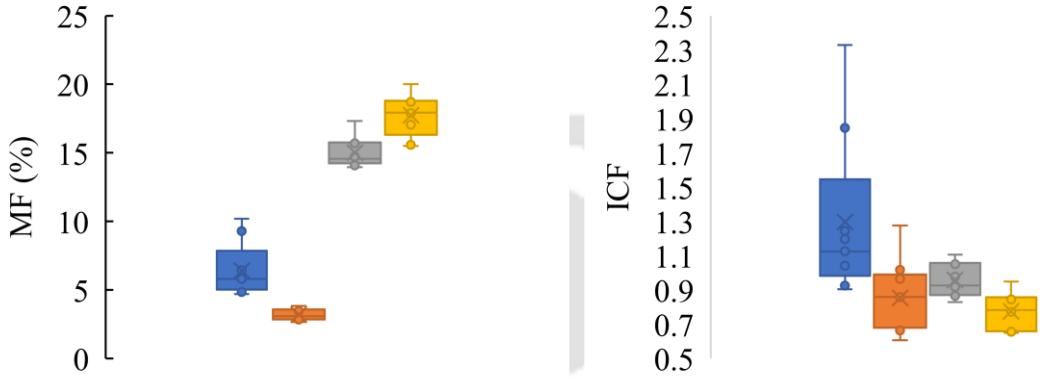


(a) Zn

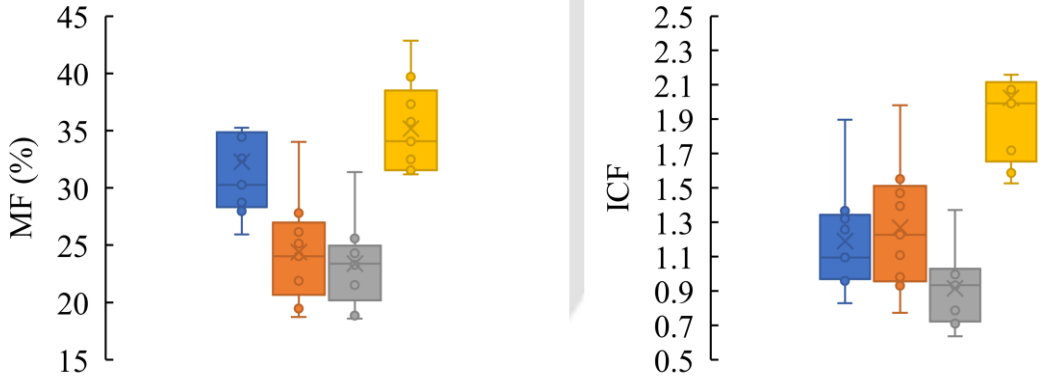
(b) Mn



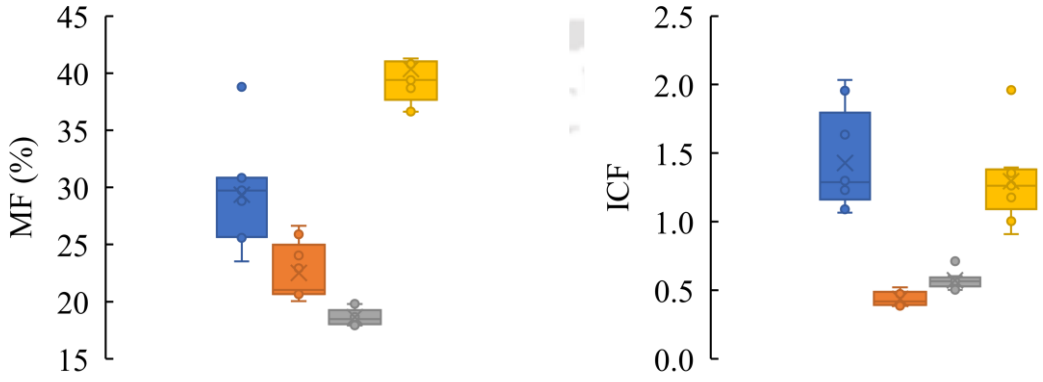
(c) Pb



(d) Fe



(e) Cu



(f) Co

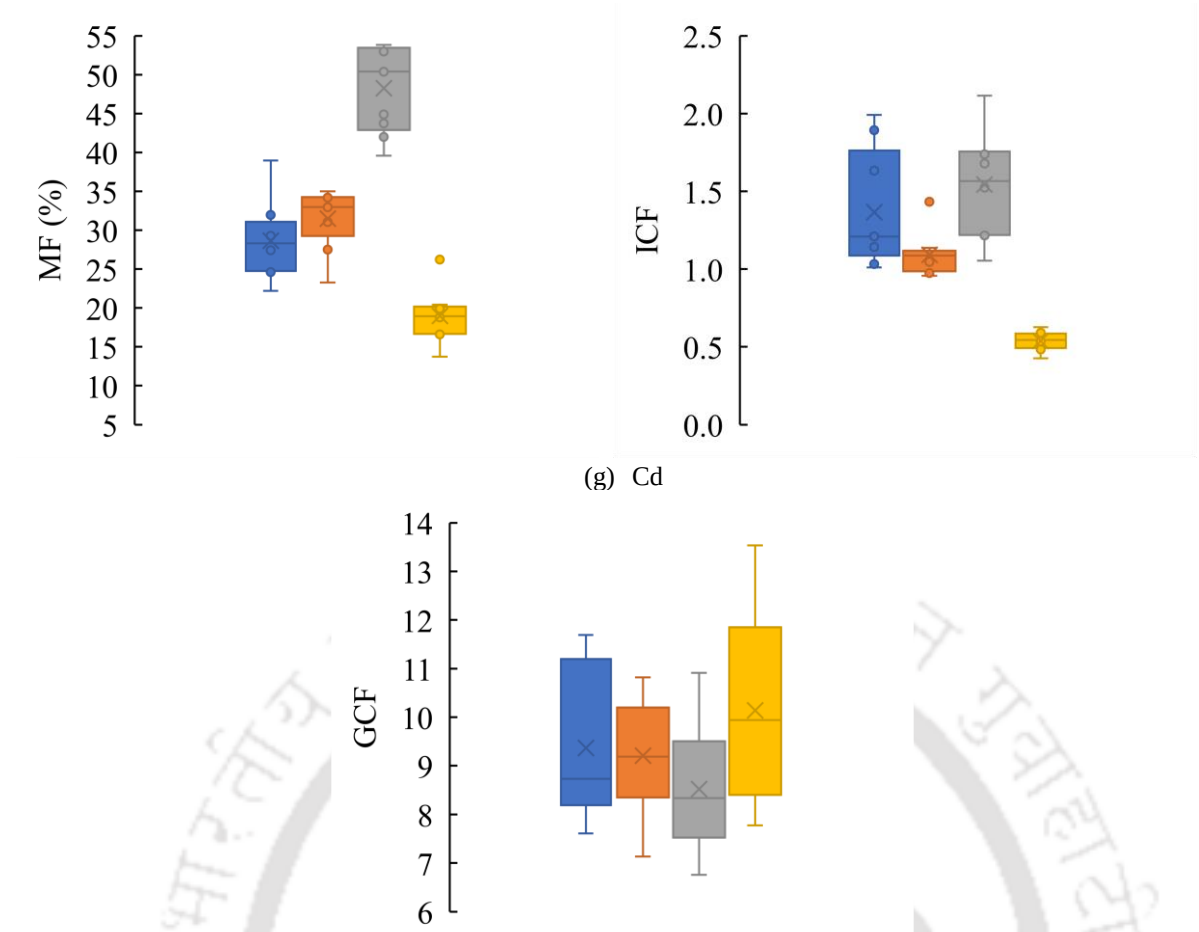


Fig. 5.8. Mobility Factor (MF), and Individual Contamination Factor (ICF) of (a) Zn, (b) Mn, (c) Pb, (d) Fe, (e) Cu, (f) Co, and (g) Cd, and their Global Contamination Factors (GCF)

The box plot in Fig. 5.8 depicts the changes in MF, ICF, and GCF temporally and spatially. The MF factor represents the mobility of bioavailable fractions in the sediments. The average order of variations for MF and ICF are as follows:

MF:	**ICF:**
Mn: W > PRM > M > POM	Mn: PRM> M > W > PTM
Zn: W > PRM > PTM> M	Zn: PTM > PRM > M > W
Pb: W > PTM > M > PRM	Pb: W > PTM > M > PRM
Fe: PTM > M > W > PRM	Fe: W > M > PRM > PTM
Cu: PTM > W > PRM > M	Cu: PTM > PRM > W > W
Co: PTM > W > PRM > M	Co: W > PTM > M > PRM
Cd: M > PRM > W > PTM	Cd: M > W > PRM > PTM

The greater amount of mobile fractions present in the sediments, the more is the risk for the ecosystem, as the fractions can be released from the sediments into the stream. Zn, Mn and Pb have the highest MF in winter, whereas Cu and Co have the highest MF in the post-monsoon season, i.e. in the dry season. In the wet season, i.e. the pre-monsoon and monsoon, the MF has decreased, which might be due to the desorption of mobile metal fractions from the sediments due to the increase in turbulence in the river (Akindele et al., 2020). For Cd, the exact opposite

happened, wherein MF is higher in the wet season than in the dry season. This indicates that turbulence has an opposite effect on Cd, which leads to the adsorption of Cd onto the sediment particles. For Fe, there is a constant increase of bioavailable metals from winter to post-monsoon season, which MF represents, indicating the anthropogenic addition of metals in the river and, in turn, setting down onto the sediment particle.

The ICF value indicates the anthropogenic contamination with respect to each metal, whereas the GCF value gives an idea about overall contamination at a site. The ICF value for Zn and Cu is highest in the post-monsoon, whereas Co, Fe, and Pb have the highest values in the winter. For most metals, dry seasons pose a higher risk than wet seasons. This indicates that metal gets adsorbed onto the sediment particles during the lean flow period, i.e. during the winter and post-monsoon seasons and gets desorbed during the high turbulent period, i.e. during the pre-monsoon and monsoon seasons. The ICF values of Zn constantly rose from low contamination in the winter to moderate contamination, indicating that Zn is continuously added to the river anthropogenically or externally. Mn, Pb, Fe, Cu, Co, and Cd have moderate to low contamination, and Zn has the highest contamination value. The cumulative effect of these metals at a site is represented by GCF, which has increased after the monsoon season (in post-monsoon) to considerable contamination, compared to the pre-monsoon and monsoon seasons, where the majority of GCF values fall into the moderate contamination category. This indicates that with the onset of monsoon or with increase in turbulence in the river, desorption of heavy metals from sediment can be observed with decrease in the labile or bioavailable fraction of metals. In other words, with monsoon receding, i.e. in post-monsoon season, the bioavailable fractions of metals in sediments rise, indicating the metal getting adsorbed onto the sediment particles giving an increased risk (Akindele and Olutona, 2014; Mna et al., 2021). From this study, it is observed that the tendency for metals to accumulate in sediments happens in dry seasons or the lean flow period of the river. Thus, the anthropogenic addition of metals in the Kolong River might be happening during the dry season, which can be the reason for high values during this period (Akindele et al., 2020). When metals are added to the sediments, they are an unstable bond which makes them bioavailable, and in such cases, the MF, ICF and GCF will show high values as in the study area.

5.4.2.3 Risk Assessment Code (RAC)

The bioavailability (Förstner and Salomons, 1980a; Lim et al., 2013; Morillo et al., 2002) of heavy metals in Kolong river samples of sediments are shown in Table 5.5, determined by summation of F_1 and F_2 chemical fractions, and non-bioavailability fractions (%). The bioavailable fraction is a portion of metal that can be absorbed or ingested by aquatic animals or plants under favourable conditions, thus causing toxicity to the environment. The bioavailable metal fraction followed the order during different seasons, shown in Fig. 5.9.

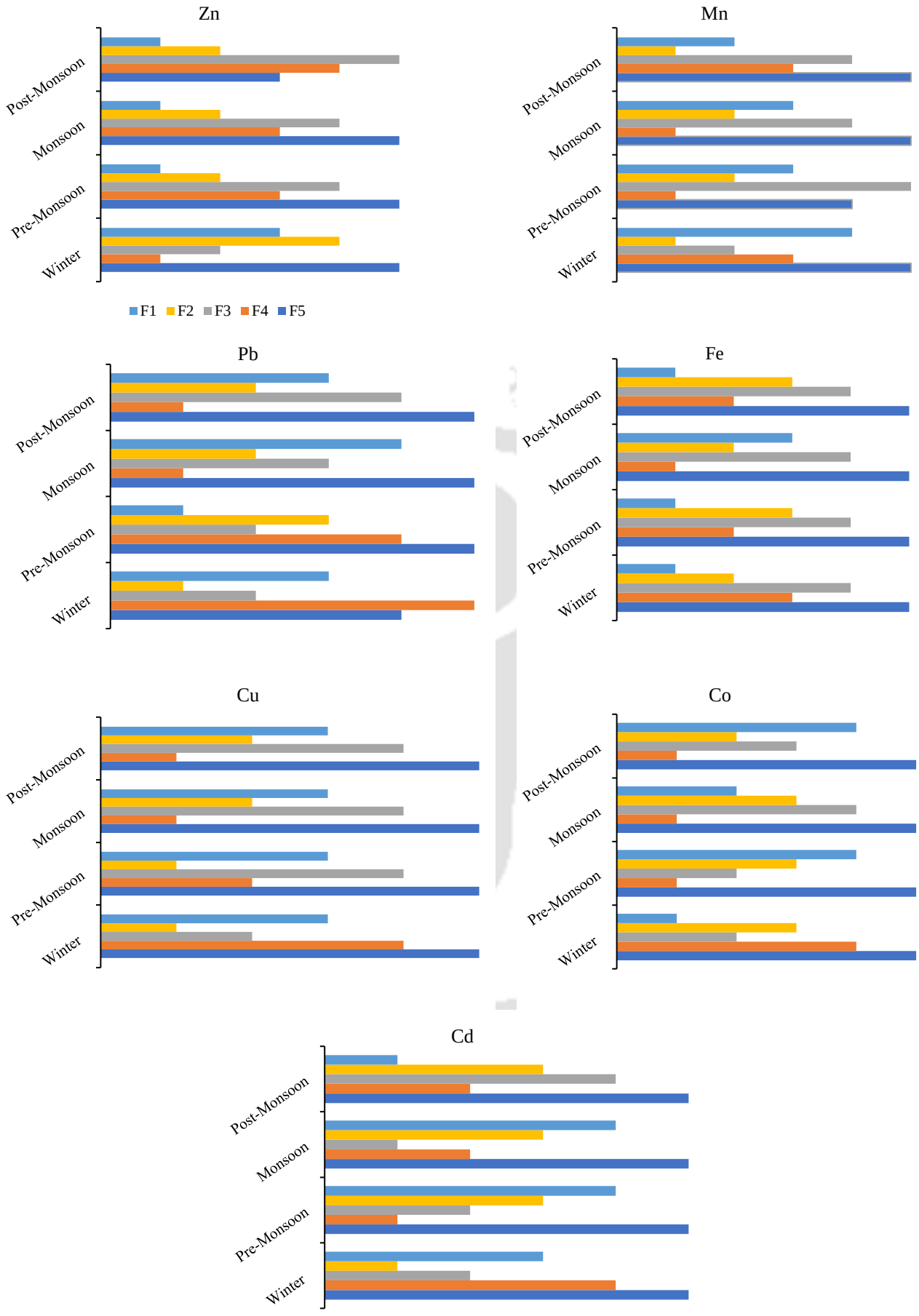


Fig. 5.9. Order of variation of different fractions of metals in different seasons

In this study, Zn, Mn, Cd, Co, and Cu have the most bio-available fraction of the benthic sediment of the Kolong river and pose a high degree of environmental risk. Risk assessment code (RAC) can be used to evaluate the threat posed by highly mobile heavy metal fractions using Table 4.4, and the results of the comparison are presented in Table 5.6 (G et al., 1985; Jain et al., 2007; Karbassi and Shankar, 2005; Singh et al., 2005). Using RAC, it was observed that for Zn, there was a decrease in risk from 'High Risk' to 'Medium' with the onset of monsoon, which indicates the release of bioavailable fraction of metals into the river stream during the time that is harmful to the environment. Similar is the case for Mn and Pb. The exact opposite is observed in the case of Cd, Co, and Cu, i.e. with the onset of monsoon, more and more metals were getting absorbed into the sediment particle giving rise to 'High Risk' or 'Very High Risk' assessment. In other words, it can be said that metals from the river stream were getting transferred onto the sediment particles due to turbulence in the river stream during the monsoon months. In the case of Fe, most of the metal is bound to the non-bioavailable phases of surficial sediment, thus posing a less severe risk to the aquatic ecosystem (Al-Mur, 2020; Liu et al., 2017; Sun et al., 2019; Zhang et al., 2017). Similar observations were made from the use of ICF. After estimating the potential risk using RAC, trace metal concentrations should be compared with the SQGs, if the total metal concentration poses any risk.

5.4.2.4 Modified Risk Index (MRI)

MRI is a modification of PERI using toxic index (∂) and modification factor (φ) based on the bioavailable fractions of metals in the sediments. The results of the comparison between MRI and PERI are shown in Fig. 5.10. As can be observed from the figure, MRI results show a greater risk than PERI. This is because specific importance has been given to samples with higher proportion of bioavailable fractions of metals present. The seasonal variation of PERI shows a decrease in ecological risk from winter to pre-monsoon and then increases in the monsoon season, and then again decreases in the post-monsoon season. This can be due to the fact that during monsoon season, increased precipitation increases the surface runoff, thus, bringing in more and more pollutants into the river ecosystem and thus increasing the ecological risk level. However, there is not much change in the proportion of bioavailable fractions present in the sediments, and thus risk associated is also the same, which MRI beautifully represents. The values obtained in the study show that metal contamination in the sediments presents a 'Low ecological risk' in the aquatic ecosystem even though from RAC, metal contamination assessment shows 'High Risk' at some during the sampling period for each of the metals. This might be because PERI and MRI classification were developed with keeping metal contamination threshold limits in sediments, and none of the sites during the sampling crossed the threshold limits. Therefore, from this study, it can be said that using MRI in place of PERI would give the correct picture of risk in the study area.

Table 5.5. Percentage wise distribution of different metal speciation fractions in different seasons

Heavy metals		Zn		Mn		Pb		Fe		Cu		Co		Cd	
Seasons	Sites	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*
		%	%	%	%	%	%	%	%	%	%	%	%	%	%
Winter	SS1	42.24	57.76	36.90	63.10	31.05	68.95	6.43	93.57	34.45	65.55	30.29	69.71	29.28	70.72
	SS2	39.84	60.16	42.82	57.18	28.67	71.33	10.20	89.80	46.58	53.42	38.80	61.20	38.94	61.06
	SS3	45.78	54.22	35.90	64.10	23.85	76.15	4.69	95.31	35.25	64.75	29.73	70.27	28.29	71.71
	SS4	49.33	50.67	30.56	69.44	23.57	76.43	5.18	94.82	25.93	74.07	23.52	76.48	22.22	77.78
	SS5	49.90	50.10	36.06	63.94	23.85	76.15	4.84	95.16	28.73	71.27	28.82	71.18	27.43	72.57
	SS6	33.33	66.67	31.18	68.82	33.00	67.00	5.80	94.20	30.27	69.73	25.69	74.31	24.88	75.12
	SS7	34.26	65.74	33.57	66.43	26.31	73.69	5.30	94.70	28.68	71.32	25.60	74.40	24.63	75.37
	SS8	37.42	62.58	38.11	61.89	20.71	79.29	5.91	94.09	32.57	67.43	30.85	69.15	30.20	69.80
	SS9	42.02	57.98	37.86	62.14	30.00	70.00	9.29	90.71	27.98	72.02	32.96	67.04	31.95	68.05
Pre-Monsoon	SS1	33.92	66.08	29.11	70.89	20.45	79.55	2.86	97.14	18.72	81.28	24.04	75.96	27.51	72.49
	SS2	28.17	71.83	34.00	66.00	16.73	83.27	2.86	97.14	19.48	80.52	25.91	74.09	31.04	68.96
	SS3	34.19	65.81	31.32	68.68	19.08	80.92	2.84	97.16	24.04	75.96	20.05	79.95	23.28	76.72
	SS4	25.12	74.88	26.11	73.89	19.78	80.22	3.08	96.92	25.14	74.86	26.64	73.36	31.55	68.45
	SS5	36.83	63.17	24.93	75.07	17.45	82.55	2.66	97.34	22.46	77.54	22.94	77.06	33.58	66.42
	SS6	30.01	69.99	24.43	75.57	17.40	82.60	3.53	96.47	26.14	73.86	20.65	79.35	32.96	67.04
	SS7	31.98	68.02	20.98	79.02	19.42	80.58	3.48	96.52	34.00	66.00	21.03	78.97	34.28	65.72
	SS8	33.29	66.71	22.66	77.34	19.85	80.15	3.83	96.17	27.81	72.19	20.69	79.31	35.01	64.99
	SS9	33.56	66.44	29.82	70.18	18.29	81.71	3.62	96.38	21.87	78.13	20.77	79.23	34.21	65.79

Heavy metals		Zn		Mn		Pb		Fe		Cu		Co		Cd	
Seasons	Sites	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*	B*	NB*
		%	%	%	%	%	%	%	%	%	%	%	%	%	%
Monsoon	SS1	19.40	80.60	21.90	78.10	26.63	73.37	7.38	92.62	18.59	81.41	18.14	81.86	39.59	60.41
	SS2	12.04	87.96	21.88	78.12	27.98	72.02	8.26	91.74	23.58	76.42	18.36	81.64	44.90	55.10
	SS3	18.89	81.11	26.30	73.70	25.85	74.15	5.63	94.37	23.29	76.71	17.92	82.08	42.01	57.99
	SS4	16.64	83.36	25.53	74.47	27.46	72.54	10.90	89.10	21.53	78.47	18.48	81.52	43.73	56.27
	SS5	24.64	75.36	29.07	70.93	27.93	72.07	8.02	91.98	18.85	81.15	17.92	82.08	53.82	46.18
	SS6	20.24	79.76	26.95	73.05	27.53	72.47	12.02	87.98	23.40	76.60	18.73	81.27	53.83	46.17
	SS7	26.00	74.00	29.81	70.19	27.85	72.15	12.27	87.73	31.39	68.61	18.65	81.35	52.96	47.04
	SS8	18.13	81.87	29.01	70.99	27.14	72.86	9.80	90.20	25.60	74.40	19.80	80.20	53.09	46.91
	SS9	38.13	61.87	38.72	61.28	27.86	72.14	11.43	88.57	24.31	75.69	17.59	82.41	50.39	49.61
Post-Mon-	SS1	24.56	75.44	20.32	79.68	25.59	74.41	17.06	82.94	42.85	57.15	39.09	60.91	20.42	79.58
soon	SS2	32.07	67.93	25.08	74.92	29.24	70.76	17.93	82.07	32.50	67.50	41.30	58.70	19.61	80.39
	SS3	32.21	67.79	23.14	76.86	27.23	72.77	18.87	81.13	37.32	62.68	40.00	60.00	18.80	81.20
	SS4	24.64	75.36	29.74	70.26	28.27	71.73	18.20	81.80	39.71	60.29	38.69	61.31	16.71	83.29
	SS5	27.42	72.58	21.77	78.23	29.37	70.63	20.01	79.99	35.76	64.24	40.78	59.22	18.92	81.08
	SS6	26.29	73.71	32.22	67.78	24.10	75.90	17.89	82.11	31.19	68.81	50.74	49.26	19.91	80.09
	SS7	20.25	79.75	24.53	75.47	25.05	74.95	18.72	81.28	34.07	65.93	39.40	60.60	13.75	86.25
	SS8	29.74	70.26	21.80	78.20	31.20	68.80	15.58	84.42	31.56	68.44	36.63	63.37	26.22	73.78
	SS9	19.24	80.76	28.01	71.99	22.95	77.05	15.49	84.51	31.53	68.47	34.81	65.19	16.59	83.41

B*: bioavailable fraction; NB*: nonbioavailable fraction.

Table 5.6. Results from comparison with RAC

Heavy metals		Zn	Mn	Pb	Fe	Cu	Co	Cd
Seasons	**Sites**	**Risk Category**						
Winter	**SS1**	High Risk	High Risk	High Risk	Low Risk	High Risk	High Risk	Medium
	SS2	High Risk	High Risk	Medium	Medium	High Risk	High Risk	High Risk
	SS3	High Risk	High Risk	Medium	Low Risk	High Risk	Medium	Medium
	SS4	High Risk	High Risk	Medium	Low Risk	Medium	Medium	Medium
	SS5	High Risk	High Risk	Medium	Low Risk	Medium	Medium	Medium
	SS6	High Risk	High Risk	High Risk	Low Risk	High Risk	Medium	Medium
	SS7	High Risk	High Risk	Medium	Low Risk	Medium	Medium	Medium
	SS8	High Risk	High Risk	Medium	Low Risk	High Risk	High Risk	High Risk
	SS9	High Risk	High Risk	High Risk	Low Risk	Medium	High Risk	High Risk
Pre-	**SS1**	High Risk	Medium	Medium	Low Risk	Medium	Medium	Medium
monsoon	**SS2**	Medium	High Risk	Medium	Low Risk	Medium	Medium	High Risk
	SS3	High Risk	High Risk	Medium	Low Risk	Medium	Medium	Medium
	SS4	Medium	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS5	High Risk	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS6	High Risk	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS7	High Risk	Medium	Medium	Low Risk	High Risk	Medium	High Risk
	SS8	High Risk	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS9	High Risk	Medium	Medium	Low Risk	Medium	Medium	High Risk

Heavy metals		Zn	Mn	Pb	Fe	Cu	Co	Cd
Seasons	**Sites**	**Risk Category**						
Monsoon	**SS1**	Medium	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS2	Medium	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS3	Medium	Medium	Medium	Low Risk	Medium	Medium	High Risk
	SS4	Medium	Medium	Medium	Medium	Medium	Medium	High Risk
	SS5	Medium	Medium	Medium	Low Risk	Medium	Medium	Very High Risk
	SS6	Medium	Medium	Medium	Medium	Medium	Medium	Very High Risk
	SS7	Medium	Medium	Medium	Medium	High Risk	Medium	Very High Risk
	SS8	Medium	Medium	Medium	Low Risk	Medium	Medium	Very High Risk
	SS9	High Risk	High Risk	Medium	Medium	Medium	Medium	Very High Risk
Post-monsoon	**SS1**	Medium	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS2	High Risk	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS3	High Risk	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS4	Medium	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS5	Medium	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS6	Medium	High Risk	Medium	Medium	High Risk	Very High Risk	Medium
	SS7	Medium	Medium	Medium	Medium	High Risk	High Risk	Medium
	SS8	Medium	Medium	High Risk	Medium	High Risk	High Risk	Medium
	SS9	Medium	Medium	Medium	Medium	High Risk	High Risk	Medium

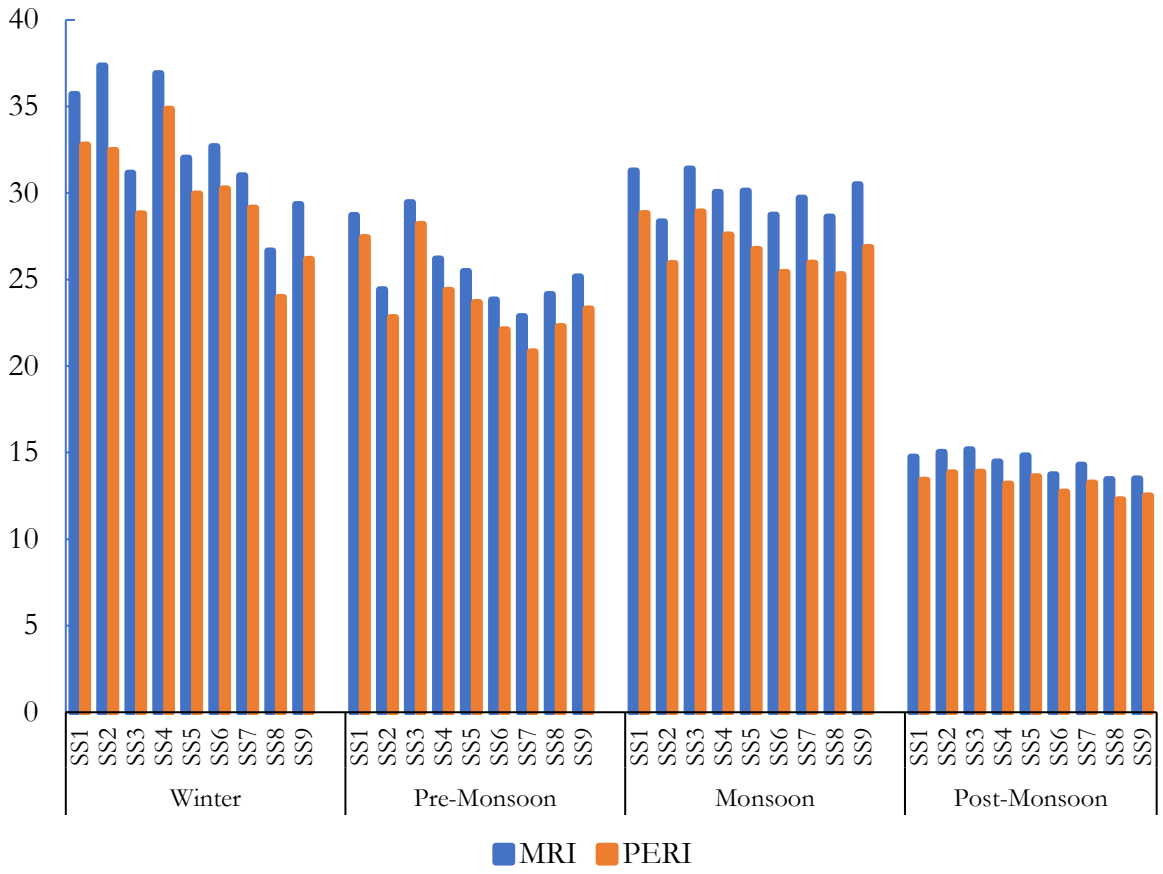


Fig. 5.10. Comparison between MRI and PERI

5.5 Sediment Quality Guidelines (SQGs)

It is known that the different metals pose different threats to the aquatic environment based on their interaction with sediment particles. In this section, three sets of sediment quality standards: ERL (effect range low) and ERM (effects range medium), TEL (Threshold Effect Level) and PEL (probable effects level) guidelines, and LEL (lowest effect level) and SEL (several effect levels) guidelines are utilised evaluate the question of whether the total metal sediment concentration is threatening the aquatic life (effects range medium) (Long et al., 1995; Macdonald et al., 1996; MacDonald and Ingersoll, 2003). Table 5.7 displays the specifics of these guidelines. The summation of the 5 fractions determined by the Tessier 5-step sequential extraction technique is total metal concentration, compared with the different guidelines (Table 5.7). However, risk assessment results concluded that the potential threat posed by the heavy metal contamination is in the 'medium' or 'high risk' or 'very high risk' category, but none of the sites has surpassed various sediment quality guidelines. This high proportion of bioavailable metals is a problem for the aquatic environment, but it is well within limits at the time of the study. Therefore, appropriate measures should be taken not to increase trace metal sediment contamination as the proportion of bioavailable metals in the surficial sediment of the Kolong river is high, which is a cause of concern for the future.

Table 5.7. Sediment Quality Guidelines (SQGs) classification

Sediment Quality Guidelines (SQGs)		**Effect**					
TEL and PEL guidelines	<TEL	Not associated with adverse biological effects					
	Between TEL-PEL	May occasionally be associated with adverse biological effects					
	>PEL	Frequently associated with adverse biological effects					
LEL and SEL guidelines	<LEL	Dredged sediments may have no contamination					
	Between LEL-SEL	The impact is moderate					
	>SEL	Severely impacted					
ERL and ERM guidelines	<ERL	Minimal effects range					
	Between ERL-ERM	Effects would occasionally occur					
	>ERM	Effects would frequently occur					
	Metal Concentration (mg/kg of sediments)						
	Mn	Zn	Cu	Pb	Cd	Co	Fe
TEL	-	124	18.7	30.2	0.68	-	-
PEL	-	271	108	112	4.21	-	-
LEL	460	120	16	31	0.6	-	-
SEL	1100	270	110	110	9	-	-
ERL	-	150	34	46.7	1.2	-	-
ERM	-	410	270	218	9.6	-	-
Measured Values in this study							
Range	0.075-0.496	0.023-0.457	0.007-0.026	0.003-0.061	0.003-0.013	0.009-0.033	1.736-10.575
Average	0.238	0.146	0.017	0.027	0.006	0.019	4.75

5.6 Summary

In this chapter, sediment pollution assessment due to heavy metals is carried out using two methodologies. The first methodology involved carrying out the assessment based on the total contamination of metals, and the second one carried out the assessment based on metal speciation fraction, i.e. the various bioavailable and non-bioavailable fractions in sediments. Various critical observations were made, and the following conclusions, listed below, are deduced from the study.

- The trace metals concentration was abundant and in the order of Fe > Mn > Zn > Cd > Pb > Co > Cu.
- PLI values show that the river bed sediments are moderately polluted. PLI value is higher in the dry season, which shows that the contaminants are added into the streams anthropogenically. This might be due to low water flow where the contaminants are settled into the river bed sediment. Possible pollutant sources may be the tea garden runoff, which uses different chemical fertilisers, herbicides, and acid forming pesticides. Additionally, agricultural and domestic waste also contribute to the river at various locations. From the Geoaccumulation index data, it has been seen that the pollution is highest due to the presence of Cadmium (1.34), Iron (1.26), Manganese (1.23), followed by Zinc (0.76), Cobalt (3.80), Copper (0.63), and Lead (5.75). In Kolong river sediments, Zn (7.98), Cd (6.40) and Pb (5.08) are significantly enriched in the river sediments indicating the source of metal as anthropogenic.
- The E_r of Cd (113.66) and Pb (42.65) were the highest. It falls in the moderate and considerable ecological risk category, which can also be observed from the higher anthropogenic fraction of the metals. The PERI values found at all sampling sites in different months were below 150 at low ecological risk.
- Sequential extraction results showed that F5 is the dominant metal speciation in surficial sediment samples collected during different seasons. However, fractions F1, F2, and F3 also substantially contribute to the metal contamination, posing a risk to the aquatic environment because of the liable bond with the sediment particles.
- The following pattern shows anthropogenic fractions in the sediments: Mn (63%) > Zn (50%) > Cu (49%) >Cd (46%) > Fe (45%) > Pb (36%) > Co (34%).
- The pollution index (I_{POLL}) is a modification of the geoaccumulation index (I_{geo}) and gives higher contamination levels for Cd, Fe, Mn, Cu, and Co and lower for Pb and Zn. This is because Cd, Fe, Mn, Cu, and Co have higher bioavailable fractions, whereas Pb and Zn have higher non-bioavailable fractions.
- The mobility factor (MF) and the individual contamination factor (ICF) for most metals analysed were higher in the winter or post-monsoon months, i.e. during the dry season. However, the variation of Cd was the exact opposite to the other metals, i.e. higher values

were found in pre-monsoon and monsoon season (wet period) and identified the increase in the Fe and Zn contamination due to anthropogenic addition of pollutants. Global contamination factor (GCF) showed that the post-monsoon season had the most considerable contamination, followed by the winter, pre-monsoon, and monsoon seasons. This contamination in benthic sediments leads to a very distinct result that during high rainfall seasons wherein flow in the river is turbulent, the contamination falls from higher form contamination to lower form contamination due to the desorption of metals from sediments

- RAC yielded the varying nature risk posed by different heavy metals in different seasons. Zn, Mn, and Pb posed a higher risk in the drier season, indicating these metals adsorb onto sediments during the lean flow period when there is low turbulence in the river leading to precipitation and adsorption onto sediment particles. Whereas with onset of monsoon or an increase in turbulence, these particles get released into the river, decreasing the risk posed by sediments to the aquatic environment. However, the exact opposite case happens with Cu, Co, and Cd, i.e. with the onset of monsoon, wherein turbulence and mixing increase in the river; metals get adsorbed onto sediment particles and released into the river during the lean period. In Fe's case, the dominant fraction is the fraction F5, which is the metal fraction that is non-reversibly bonded to the sediment particles' crystal structure, due to which the risk posed by Fe is less, even though the concentration of Fe is appreciably high.
- PERI and MRI show that the bioavailable fraction-based index (MRI) was able to quantify the risk or hazard of metal contamination. This is because when the bioavailable fractions were lower in the sediments, MRI calculated the risk very close to the PERI values, but when the bioavailable fractions were higher, the values were higher than the PERI values, thus giving them higher risk.
- Comparison with the SQGs revealed that the total metal concentration never exceeded SQGs, suggesting that sediment contamination does not threaten the aquatic environment at present. However, a high proportion of fractions F1, F2, and F3 indicates a severe threat to the aquatic environment with an increase in future contaminations.

Thus, this study calls for the use of bioavailable fractions-based evaluation parameters in sediment pollution assessment for better quantification of the risk of metal contamination.

Chapter 6

APPLICATION OF ENVIRONMENTRICS TOOLS FOR WATER QUALITY MONITORING AND ASSESSMENT

In this chapter, hierarchical cluster analysis (HCA), discriminant analysis (DA), principal component analysis (PCA) and positive matrix factorisation (PMF) have been applied to water quality and sediment quality datasets to understand the nature of the variations in the datasets in the river ecosystem. In first half of this chapter, the results from analysis of the water quality dataset are discussed, and in the 2nd half, the results from the sediment quality data analysis are discussed.

6.1 Analysis of Water Quality Dataset

6.1.1 Discriminant Analysis (DA)

The aim of carrying out the discriminant analysis (DA) is to find out parameters affecting the seasonal variation in the river. This will help formulate remediation strategies based on the parameters causing the river's water quality degradation, as observed in chapter 5. Standard and stepwise modes were used for constructing classification functions (CFs). It was observed that only 10 (turbidity, TDS, EC, hardness, alkalinity, Na^{+}, Ca^{+2}, Mg^{+2}, F^{-}, SO_4^{2-} and Cd) out of 21 parameters (based on the Fisher's Linear Discriminating Functions) were responsible for the temporal variability of the water quality dataset of the studied area (Table 6.1). Turbidity, TDS, EC, hardness, and alkalinity explain the domestic discharge quotient, Na^{+}, Ca^{+2}, Mg^{+2}, and F^{-} explains soil pollution contribution to river pollution, SO_4^{-2} explains mostly the agricultural runoff quotient, and Cd explains the anthropogenic or geogenic metal pollution quotient of the seasonal variation in the water quality dataset. Therefore, any remediation strategies formulated can be based on these quotients of pollution in the Kolong river. Thus, DA will help formulate targeted pollution abatement policies for this study area and any polluted water body.

6.1.2 Most probable source identification

6.1.2.1 Using Principal Component Analysis (PCA)

Samples from the Kolong river were collected for 12 months. After analysing the collected samples for the water quality parameters, the data has been clubbed into four seasons: pre-monsoon, monsoon, post-monsoon, and winter. PCA was performed on a standardised log-transformed dataset for each season to identify the latent factors, which were then used for source

identification. This assessment was conducted primarily to establish an entirely new collection of variables that were significantly fewer in number than the initial dataset used in the subsequent analysis. According to the eigenvalue-one criterion and the factor loadings of the components were classified as 'weak', 'moderate,' and 'strong,' with absolute loading values ranging from 0.30 to 0.50, 0.50 to 0.75, and > 0.75, respectively (Liu et al., 2003; Singh et al., 2021b).

Table 6.1. Classification functions for discriminant analysis of the temporal variation in Kolong river

Classification Function Coefficients								
	Standard Mode				Stepwise Mode			
	Season				Season			
	Winter	Pre-Monsoon	Mon-soon	Post-Monsoon	Win-ter	Pre-Mon-soon	Mon-soon	Post-Mon-soon
pH	1345.70	1322.56	1308.30	1335.44				
DO	-49.91	-52.52	-56.77	-58.74				
Turbidity	8.85	8.43	9.76	7.66	1.56	1.21	2.60	0.25
BOD_5	-0.01	-0.50	-0.46	-0.59				
TDS	0.09	-0.07	-0.06	-0.09	0.21	0.09	0.10	0.09
EC	0.72	0.60	0.42	1.02	-0.01	-0.17	-0.34	0.15
Hardness	1.09	0.21	0.77	-2.12	2.16	1.25	2.01	-0.42
Alkalinity	1.14	0.92	1.97	0.58	0.13	0.16	1.09	-0.29
Na+	4.72	14.55	7.42	-1.74	5.29	14.07	7.76	0.44
K^+	-0.73	-1.58	-1.82	-1.07				
Ca^{+2}	-0.55	0.25	-1.03	8.08	-5.38	-4.22	-6.11	1.64
Mg^{+2}	-11.60	-12.38	-15.27	-7.18	0.98	0.09	-2.45	4.91
F^-	272.12	204.94	166.51	121.88	111.96	84.80	41.11	13.87
Cl^-	1.48	1.64	2.07	2.29				
SO_4^{-2}	-1.28	1.91	9.85	-2.88	2.55	3.62	12.03	-0.28
NO_3^-	-0.62	-0.63	-0.54	-0.62				
Fe	15.33	12.66	16.95	14.69				
Mn	84.20	136.56	124.72	196.44				
Zn	31.58	48.35	37.82	32.59				
Pb	-82.25	-98.53	-77.91	7.86				
Cd	-307.48	-491.67	-625.56	-769.57	397.79	279.85	128.82	23.02
(Constant)	-5032.00	-4801.03	-4726.62	-4772.08	-195.30	-192.62	-247.29	-88.16
Fisher's linear discriminant functions								

PCA of winter season

For the winter season dataset, across 5 PCs (Table 6.2), PC1 describes 31.91% of the total variance and has strong positive loading on Cd. This factor explains metal pollution's contribution to deteriorating river quality (Schlosser and Karr, 1981; Wang et al., 2017). PC2 explains 20.43% of the overall variance and has strong loadings on turbidity, alkalinity, and Mg^{2+}, which can be related to domestic discharges (Miller et al., 2007). PC3, PC4, and PC5, respectively, explain 12.87%, 7.72%, and 6.80% of the total variance. This can be directly related to the various densely

populated settlements on the bank of the Kolong river (Fig. 3.4), and since the town does not have a sewage treatment plant, all the settlements discharge their household waste directly into the Kolong river (Bora and Goswami, 2015).

Table 6.2. Results of the PCA for water quality parameters for winter

	Component				
	1	2	3	4	5
pH	-.075	.026	-.919	.022	.043
DO	.717	.113	-.135	-.115	-.092
Turbidity	-.343	-.807	.288	.079	.044
BOD_5	-.231	-.282	-.004	.100	.817
TDS	.559	-.451	.517	-.092	-.096
EC	.441	.709	.514	.066	-.038
Hardness	-.221	.155	.036	.946	.095
Alkalinity	-.125	.883	.161	-.033	-.027
Na^+	.520	.302	.695	.058	.040
K^+	-.068	.221	-.039	.062	.895
Ca^{+2}	-.218	-.026	.047	.958	.090
Mg^{+2}	-.063	.920	-.043	.149	.048
F^-	.647	.215	.263	-.333	-.131
Cl^-	-.529	.451	.290	.179	.451
SO_4^{2-}	-.075	-.088	-.206	.145	.127
NO_3^-	.082	-.351	-.197	-.007	-.112
Fe	-.686	.105	-.223	.353	.080
Zn	-.280	-.082	.434	.290	.331
Cd	.849	.164	.228	-.137	-.131
Eigenvalue	4.794	3.667	2.518	2.277	1.894
% total variance	25.233	19.301	13.254	11.987	9.971
Cumulative % variance	25.233	44.534	57.789	69.775	79.746

PCA of pre-monsoon season

PC1 describes 32.44% of the total variance across 5 PCs (Table 6.3), with strong loads on EC, K^+, Mg^{+2}, F^- and Cl^- for the pre-monsoon season dataset. PC2 and PC3, respectively, account for 14.31% and 10.84% of the overall variation and have strong loads on turbidity, hardness, alkalinity and Ca^{+2}. These three main components illustrate domestic discharge's relationship to river water quality (Miller et al., 2007). The overall variance described by PC4 and PC5, respectively, were 8.92% and 7.58% and SO_4^{2-} and NO_3^- have strong loads with the PCs. These loadings are caused by overland flow due to the start of rainfall in these months, which leads to surface run-off and agricultural run-off (CPCB, 2015; Goonetilleke et al., 2005; Schlosser and Karr, 1981). Also, from

Fig. 3.4, it can be observed that the Kolong river is surrounded by agricultural fields, forests, or tea gardens at some of the locations on the bank of the river, and this factor can be related to that (Mendivil-Garcia et al., 2020; Zhou et al., 2012).

Table 6.3. Results of the PCA for water quality parameters for pre-monsoon

	Component				
	1	**2**	**3**	**4**	**5**
pH	.616	-.085	-.098	.310	.103
DO	-.385	-.209	.587	.263	-.094
Turbidity	.061	-.228	.812	-.013	.019
BOD_5	.265	.182	-.170	.107	.722
TDS	-.316	-.235	.017	-.346	-.614
EC	.899	-.117	-.134	.074	-.269
Hardness	.350	.814	-.186	.134	.298
Alkalinity	-.023	.824	-.366	.093	-.214
Na^+	-.144	.355	.174	-.600	-.476
K^+	.830	.422	.011	-.002	.080
Ca^{+2}	.004	.936	-.031	-.016	.263
Mg^{+2}	.853	.415	-.170	.091	.123
F^-	.762	-.060	.152	.248	.232
Cl^-	.888	.322	.029	.004	.133
SO_4^{2-}	.122	.092	.018	.807	-.004
NO_3^-	.129	.050	-.034	.752	-.020
Fe	.047	.090	.778	-.283	-.103
Mn	-.158	.139	.020	-.367	.677
Zn	.282	.629	.362	-.180	.216
Cd	.665	-.029	-.194	.152	.442
Pb	-.429	.028	.564	.102	-.163
Eigenvalue	5.202	3.442	2.408	2.256	2.252
% total variance	24.771	16.390	11.467	10.744	10.724
Cumulative % variance	24.771	41.161	52.628	63.373	74.097

PCA of monsoon season

PC1 describes 42.11% of the total variance across 5 PCs in the monsoon season dataset (Table 6.4) and has high loads on pH, DO, TDS, EC, alkalinity, Na^+, Mg^{+2}, SO_4^{2-}, and Zn. This component mainly determines residential discharge's relationship to water quality, with a small impact from wastewater and metal contamination (Wang et al., 2017). Of the overall variance, 16.41% and 10.66% are respectively explained by PC2 and PC3, which are strongly loaded by Ca^{+2}, hardness, and turbidity, respectively. These two main components describe domestic discharge's

relationship to the river's water quality (Miller et al., 2007). PC4 and PC5 describe 6.19% and 5.01% of the overall variation, respectively, and have strong positive loads on Cl^- and NO_3^- which are related to agricultural or surface runoff (Juahir et al., 2011; Kayhanian et al., 2012; Vieira et al., 2013), as well as industrial discharges (Edokpayi et al., 2017; Zhao et al., 2011). Household sewage is dumped directly into the river by those who live along its banks. Nagaon and Morigaon do not have a sewage treatment facility at present. Kitchen and bathroom wastewater are flushed into the river through an exposed drain. Local industries such as vehicle repair shops, gas stations, and others contribute to industrial pollution. Also, from Fig. 3.4, there is a lot of vegetation or agricultural fields form which runoff can take place and pollute the river. These pollution sources are common in any river as population settlement mainly occurs on the bank of the river (Zhao et al., 2011). Also, there are a lot of major and minor roads near the river, from which runoff can take place and pollute the river in the process (Fig. 3.4).

PCA of post-monsoon season

For the post-monsoon season dataset, PC1 describes 43.52% of total variation across 6 PCs (Table 6.5) and has heavy loads on pH, turbidity, hardness, Cl^-, and Cd. This component mainly describes domestic discharge's relationship to water quality, with a small contribution of metal contamination (Ha and Bae, 2001; Singh et al., 2021a). PC2, PC3, and PC4, respectively, explain 11.55%, 9.39%, and 7.57% of the overall variance and have strong loadings on Zn, Pb, Mn, and Cu. These principal components illustrate the presence of metal contaminants in the river, which can be of geogenic or anthropogenic origin depending on the local region's geology (Chowdhury et al., 2016). In this study, Zn and Mn can be traced to be geogenic as these metals can be found in appreciable quantity in the area (Brahma and Misra, 2014; Krishan et al., 2016), and Cu and Pb can be traced to industrial sources because of the industries located in the area. PC5 and PC6 account for 5.75% and 5.31% of the total variance and have significant component loadings with TDS and DO, which can be linked to domestic discharges (Schlosser and Karr, 1981). Since pollution sources change over time, it may be concluded that the Kolong River pollution source is non-point. The identification of these sources will significantly aid the formulation of mitigation plans.

6.1.2.2 Using Correlation Matrix

Winter season correlation matrix

In the correlation analysis for the winter season, which is shown in Table 6.6, TDS is correlated with turbidity (R^2=0.78); alkalinity (R^2=0.88) and Na^+ (R^2=0.73) is correlated with EC; Ca^{+2} ((R^2=0.96) and Fe (R^2=0.50) with hardness; Mg^{+2} is correlated with EC (R^2=0.81) and alkalinity (R^2=0.75); Cl^- is correlated with alkalinity (R^2=.52) and Mg^{+2} (R^2=0.64). The high correlations in the dataset of the winter season can be deduced to be due to domestic discharges (Osei et al., 2010). The highly correlated compounds like alkalinity, Na^+, EC, Ca^{+2}, hardness, Mg^{+2}, and Cl^- are mostly components of domestic discharges. In this analysis, since metals like Ca^{+2} and Fe are

highly correlated with hardness, these metals can be said to be from domestic discharges or from the soil condition of the region rather than the industrial origin. So, in the correlation matrix analysis of the winter season dataset, the source of pollution can be said to be domestic or geo-genic (soil condition) (Ahmed et al., 2019).

Table 6.4. Results of the PCA for water quality parameters for monsoon

	Component				
	1	2	3	4	5
pH	.837	.131	.339	.002	-.004
DO	-.974	.043	-.013	-.079	-.070
Turbidity	-.134	-.119	-.795	.136	.036
BOD5	.331	-.534	.473	.252	.043
TDS	.824	.273	-.103	-.019	-.014
EC	-.951	.129	-.046	-.018	-.046
Hardness	.229	.937	-.049	.118	.066
Alkalinity	.965	.019	.129	.054	.085
Na^+	.952	-.019	.147	.081	.079
K^+	-.259	-.344	.736	-.091	-.020
Ca^{+2}	-.020	.971	-.083	.018	.057
Mg^{+2}	.828	.107	.094	.341	.042
F^-	.172	.585	.651	.083	-.002
Cl^-	-.164	-.231	.199	.164	-.831
SO_4^{2-}	.861	-.041	.289	-.126	.109
NO_3^-	-.100	-.139	.135	-.831	.097
Fe	.036	-.341	.178	.517	.573
Mn	-.725	-.096	.026	.085	.057
Pb	-.319	.228	-.723	-.066	.092
Zn	.772	.028	-.009	.201	.071
Eigenvalue	8.032	2.954	2.700	1.297	1.091
% total variance	40.162	14.771	13.502	6.485	5.454
Cumulative % variance	40.162	54.933	68.434	74.919	80.973

Pre-monsoon correlation matrix

In the correlation analysis of the pre-monsoon dataset, which is shown in Table 6.7: DO is correlated with turbidity (R^2=0.69); EC is highly correlated with hardness (R^2=0.91), alkalinity (R^2=0.90), Na^+ (R^2=0.60), K^+(R^2=0.83), Ca^{+2} (R^2=0.82), Mg^{+2} (R^2=0.90) and Cl^- (R^2=0.95); hardness is correlated with alkalinity (R^2=0.99), Na^+ (R^2=0.62), K^+(R^2=0.86), Ca^{+2} (R^2=0.95), Mg^{+2} (R^2=0.91), Cl^- (R^2=0.90) and Zn (R^2=0.54); alkalinity is correlated with Na^+ (R^2=0.58), K^+(R^2=0.86), Ca^{+2} (R^2=0.94), Mg^{+2} (R^2=0.92) and Cl^- (R^2=0.88); Na^+ is correlated with

K^+(R^2=0.66), Ca^{+2} (R^2=0.61), Mg^{+2} (R^2=0.62) and Cl^- (R^2=0.70); K^+ is correlated with Ca^{+2} (R^2=0.89), Mg^{+2} (R^2=0.84) and Cl^- (R^2=0.82); Ca^{+2} is correlated with Mg^{+2} (R^2=0.84) and Cl^- (R^2=0.89); Mg^{+2} is correlated with Cl^- (R^2=0.89). These correlations directly indicate the domestic discharge component of pollution (Osei et al., 2010). Apart from these correlations, turbidity is correlated with iron (R^2=0.60); F^- is correlated with Mn (R^2=0.68); Zn is correlated with hardness (R^2=0.54), alkalinity (R^2=0.57), K^+ (R^2=0.70), Ca^{+2} (R^2=0.64), Mg^{+2} (R^2=0.61) and Cl^- (R^2=0.57); Pb is correlated with DO (R^2=0.60) and turbidity (R^2=0.53). This correlation indicates metal pollution component, which may be geogenic or anthropogenic (Ahmed et al., 2019). The correlation between turbidity and chlorine can suggest an anthropogenic source of pollution, and the rest suggest a geogenic source. Lastly, NO_3^- and SO_4^{-2} are highly correlated (R^2=0.52), indicating or suggesting the agricultural runoff portion of the pollution (Park et al., 2014).

Table 6.5. Results of the PCA for water quality parameters for post-monsoon

	Component					
	1	**2**	**3**	**4**	**5**	**6**
pH	-.795	-.225	-.303	-.115	-.249	.281
DO	.044	.014	-.037	-.010	-.002	.908
Turbidity	-.881	-.142	-.045	-.032	-.013	.207
BOD_5	.078	-.308	.688	.059	-.300	.264
TDS	.159	.011	-.121	.127	.840	-.008
EC	-.542	-.442	-.396	-.051	-.035	.084
Hardness	.762	.463	.292	.112	.124	.036
Alkalinity	.479	-.109	.465	.283	.601	-.070
Na^+	.229	.667	.136	.207	.447	-.092
K^+	.494	.199	.193	.614	.277	-.047
Ca^{+2}	.588	.536	.381	-.077	.190	.165
Mg^{+2}	.659	-.056	.410	.289	.438	-.099
F^-	.531	.313	.434	.066	.435	.168
Cl^-	-.891	-.145	-.056	-.017	-.183	-.264
SO_4^{2-}	-.090	-.022	.288	-.003	-.033	-.091
NO_3^-	-.612	-.150	-.169	.104	-.458	-.280
Fe	-.278	.043	.111	.896	.050	-.033
Mn	.142	.188	.808	.008	.130	-.207
Zn	.100	.895	-.123	.087	.089	-.001
Pb	.214	.809	.033	-.103	-.285	.049
Cu	.186	-.017	-.094	.933	.085	.030
Cd	.820	.189	.338	-.030	.169	.107
Eigenvalue	6.745	3.033	2.518	2.347	2.318	1.320
% total variance	30.657	13.786	11.444	10.667	10.537	5.999
Cumulative % variance	30.657	44.444	55.887	66.555	77.091	83.090

Monsoon correlation matrix

In the correlation matrix for the monsoon dataset, which is shown in Table 6.8: pH is correlated with DO (R^2=0.61), hardness (R^2=0.64), K^+ (R^2=0.70), Ca^{+2} (R^2=0.68), and F^- (R^2=0.70); DO is correlated with EC (R^2=0.67), hardness (R^2=0.56), alkalinity (R^2=0.51), K^+ (R^2=0.72), Ca^{+2} (R^2=0.65), and F^- (R^2=0.83); EC is correlated with hardness (R^2=0.73), K^+ (R^2=0.80), Ca^{+2} (R^2=0.77), and F^- (R^2=0.88); hardness is correlated with Na^+ (R^2=0.57), K^+ (R^2=0.75), Ca^{+2} (R^2=0.98), and F^- (R^2=0.71); alkalinity is correlated with F^- (R^2=0.50); Na^+ is correlated with Ca^{+2} (R^2=0.54), K^+ is correlated with Ca^{+2} (R^2=0.80), and F^- (R^2=0.89); Ca^{+2} is correlated with F^- (R^2=0.75). These high correlations indicate or suggest the domestic discharge component of the pollution in the river (Osei et al., 2010). Also, SO_4^{-2} is correlated with pH (R^2=0.66), DO (R^2=0.80), EC (R^2=0.68), hardness (R^2=0.58), alkalinity (R^2=0.66), K^+ (R^2=0.87), Ca^{+2} (R^2=0.61) and F^- (R^2=0.88). These correlations imply that the river during monsoon season is polluted by agricultural or surface runoff (Park et al., 2014). Besides the above correlations, Mn is correlated with EC (R^2=0.55), and Pb is correlated with Zn (R^2=0.82). These two correlations indicate that river is contaminated and polluted with metal to some extent which may be either geogenic or anthropogenic (Ahmed et al., 2019).

Post-monsoon correlation matrix

In the correlation matrix dataset of the post-monsoon season, shown in Table 6.9: pH is correlated with DO (R^2=0.54) and K^+ (R^2=0.54); turbidity is correlated with alkalinity (R^2=0.56) and Mg^{+2} (R^2=0.53); EC is correlated with K^+ (R^2=0.53) and Mg^{+2} (R^2=0.52); hardness with Ca^{+2} (R^2=0.68) and NO_3^- (R^2=0.59); alkalinity is correlated with Mg^{+2} (R^2=0.90); and Mg^{+2} is correlated with Cl^- (R^2=0.54). These correlations suggest that river contamination is due to domestic discharges from the vicinity (Osei et al., 2010). Apart from the above: Mn is correlated with Cl^- (R^2=0.63) and SO_4^{-2} (R^2=0.79); Zn is correlated with NO_3^- (R^2=0.59) and Pb (R^2=0.68). These correlations suggest that the metal pollution content of the river can again be anthropogenic or geogenic (Ahmed et al., 2019). The results from the correlation matrix are in line with PCA conducted in this study. Pollution in the months of winter and post-monsoon were attributed to domestic discharges and metal pollution. Whereas another pollution source gets added in pre-monsoon and monsoon months due to heavy precipitation in the region during this period which brings in agricultural or surface runoff quotient of pollution into the river.

6.1.2.3 Using Positive Matrix Factorisation (PMF)

For analysis of sources using PMF, the number of factors is set as the number of PCs generated during PCA.

Table 6.6. Winter season correlation matrix

	pH	*DO*	*Tur-bidity*	*BOD_5*	*TDS*	*EC*	*Hard-ness*	*Alka-linity*	*Na^+*	*K+*	*Ca^{+2}*	*Mg^{+2}*	*F^-*	*Cl^-*	*SO_4^{2-}*	*NO_3^-*	*Fe*	*Zn*	*Cd*
pH	1.00																		
DO	0.44	1.00																	
Turbid-ity	-0.32	-0.09	1.00																
BOD_5	-0.06	0.06	0.39	1.00															
TDS	-0.35	-0.22	0.78	0.13	1.00														
EC	-0.37	-0.11	-0.60	-0.19	-0.54	1.00													
Hard-ness	-0.27	0.20	-0.08	0.10	-0.28	0.48	1.00												
Alkalin-ity	-0.11	-0.20	-0.71	-0.38	-0.59	0.88	0.19	1.00											
Na^+	-0.77	-0.39	0.02	0.18	-0.02	0.73	0.43	0.49	1.00										
K+	-0.04	-0.19	-0.20	0.50	-0.12	0.38	0.05	0.31	0.44	1.00									
Ca^{+2}	-0.32	0.16	0.15	0.15	-0.05	0.28	0.96	-0.02	0.37	-0.02	1.00								
Mg^{+2}	0.09	0.19	-0.80	-0.15	-0.85	0.81	0.44	0.75	0.34	0.23	0.18	1.00							
F^-	-0.12	0.01	-0.37	-0.20	-0.24	0.11	-0.10	0.04	-0.03	-0.19	-0.19	0.27	1.00						
Cl^-	-0.31	-0.17	-0.35	0.21	-0.47	0.69	0.46	0.52	0.65	0.62	0.31	0.64	0.16	1.00					
SO_4^{2-}	0.10	-0.02	0.09	0.23	-0.13	-0.12	-0.04	-0.23	0.02	0.09	-0.04	0.01	0.27	0.33	1.00				
NO_3^-	0.44	0.22	0.44	0.08	0.39	-0.82	-0.30	-0.69	-0.68	-0.19	-0.12	-0.67	-0.40	-0.64	-0.20	1.00			
Fe	-0.29	-0.03	-0.17	-0.14	-0.37	0.32	0.50	0.14	0.36	-0.01	0.43	0.36	0.31	0.51	0.32	-0.25	1.00		
Zn	-0.41	-0.16	0.32	0.26	0.14	0.03	0.37	-0.13	0.19	-0.03	0.45	-0.14	-0.02	0.17	-0.01	-0.07	-0.18	1.00	
Cd	0.28	0.21	-0.07	-0.29	0.01	-0.08	0.11	0.12	-0.40	-0.23	0.14	-0.05	-0.15	-0.10	-0.21	0.10	-0.28	0.23	1.00

Table 6.7. Pre-monsoon season correlation matrix

	pH	*DO*	*Tur-bidity*	*BOD_5*	*TDS*	*EC*	*Hard-ness*	*Alka-linity*	*Na^+*	*K^+*	*Ca^{+2}*	*Mg^{+2}*	*F^-*	*Cl^-*	*SO_4^{2-}*	*NO_3^-*	*Fe*	*Mn*	*Zn*	*Pb*	*Cd*
pH	1.00																				
DO	0.03	1.00																			
Tur-bidity	-0.56	0.60	1.00																		
BOD_5	-0.14	-0.39	-0.18	1.00																	
TDS	0.42	-0.09	-0.26	-0.29	1.00																
EC	0.01	-0.37	-0.47	-0.31	-0.21	1.00															
Hard-ness	0.03	-0.40	-0.50	-0.22	-0.30	0.91	1.00														
Alka-linity	0.05	-0.42	-0.55	-0.16	-0.31	0.90	0.99	1.00													
Na^+	0.17	-0.39	-0.51	-0.16	0.14	0.60	0.62	0.58	1.00												
K+	-0.21	-0.42	-0.37	-0.17	-0.26	0.83	0.86	0.86	0.66	1.00											
Ca^{+2}	-0.15	-0.33	-0.35	-0.16	-0.45	0.82	0.95	0.94	0.61	0.89	1.00										
Mg^{+2}	0.16	-0.47	-0.65	-0.11	-0.21	0.90	0.91	0.92	0.62	0.84	0.84	1.00									
F^-	-0.22	0.20	0.42	-0.21	0.21	-0.20	-0.31	-0.33	-0.24	-0.12	-0.28	-0.37	1.00								
Cl^-	-0.08	-0.30	-0.35	-0.32	-0.32	0.95	0.90	0.88	0.70	0.88	0.89	0.87	-0.23	1.00							
SO_4^{2-}	-0.02	-0.08	-0.01	-0.08	-0.32	-0.01	0.14	0.13	0.13	0.11	0.18	-0.02	-0.14	0.08	1.00						
NO_3^-	0.28	0.12	-0.20	0.00	-0.09	-0.07	-0.12	-0.10	-0.09	-0.21	-0.20	-0.09	0.01	-0.15	0.52	1.00					
Fe	-0.40	0.46	0.69	-0.37	-0.09	-0.23	-0.23	-0.30	0.11	0.00	-0.06	-0.41	0.32	0.01	0.22	-0.13	1.00				
Mn	-0.02	0.08	0.10	0.16	0.24	-0.19	-0.18	-0.18	-0.10	-0.09	-0.22	-0.16	0.68	-0.26	-0.29	-0.04	-0.05	1.00			
Zn	-0.16	0.01	-0.14	-0.10	-0.49	0.49	0.54	0.57	0.21	0.70	0.64	0.61	-0.06	0.57	0.04	-0.09	0.08	-0.07	1.00		
Pb	-0.18	0.60	0.53	-0.03	-0.15	-0.61	-0.59	-0.60	-0.52	-0.53	-0.46	-0.66	0.21	-0.54	0.31	0.31	0.30	-0.11	-0.26	1.00	
Cd	0.01	-0.24	0.02	0.13	0.16	-0.12	-0.13	-0.11	-0.16	0.06	-0.15	0.06	0.02	-0.16	-0.21	-0.07	-0.07	0.16	0.23	-0.33	1.00

Table 6.8. Monsoon season correlation matrix

	pH	*DO*	*Tur-bidity*	*BOD_5*	*TDS*	*EC*	*Hard-ness*	*Alka-linity*	*Na^+*	*K^+*	*Ca^{+2}*	*Mg^{+2}*	*F^-*	*Cl^-*	*SO_4^{2-}*	*NO_3^-*	*Fe*	*Mn*	*Zn*	*Pb*
pH	1.00																			
DO	0.61	1.00																		
Tur-bidity	-0.64	-0.73	1.00																	
BOD_5	0.15	-0.12	0.06	1.00																
TDS	-0.07	0.12	-0.24	-0.30	1.00															
EC	0.47	0.67	-0.52	0.07	-0.17	1.00														
Hard-ness	0.64	0.56	-0.40	-0.08	-0.38	0.73	1.00													
Alka-linity	0.47	0.51	-0.74	-0.25	0.14	0.33	0.31	1.00												
Na^+	0.04	0.12	0.10	-0.23	-0.27	0.49	0.57	-0.15	1.00											
K^+	0.70	0.72	-0.65	0.00	-0.06	0.80	0.75	0.48	0.22	1.00										
Ca^{+2}	0.68	0.65	-0.42	-0.09	-0.31	0.77	0.98	0.31	0.54	0.80	1.00									
Mg^{+2}	-0.21	-0.45	0.13	0.05	-0.28	-0.22	0.04	-0.04	0.07	-0.29	-0.18	1.00								
F^-	0.67	0.83	-0.71	0.06	-0.14	0.88	0.71	0.50	0.20	0.89	0.75	-0.27	1.00							
Cl^-	0.19	-0.10	-0.20	-0.02	0.48	-0.01	-0.04	0.01	0.03	-0.04	-0.08	0.16	-0.02	1.00						
SO_4^{2-}	0.66	0.80	-0.80	-0.09	0.07	0.68	0.58	0.66	-0.01	0.87	0.61	-0.15	0.88	0.03	1.00					
NO_3^-	0.24	0.07	-0.11	0.00	0.04	-0.05	-0.12	0.34	-0.03	0.01	-0.04	-0.38	-0.07	-0.08	0.02	1.00				
Fe	-0.14	-0.21	0.07	0.22	-0.58	-0.06	-0.01	-0.09	-0.14	-0.06	-0.10	0.41	-0.07	-0.36	0.04	-0.22	1.00			
Mn	-0.11	0.37	-0.38	0.03	0.23	0.55	-0.03	0.35	-0.02	0.38	0.01	-0.18	0.49	-0.10	0.42	-0.09	-0.15	1.00		
Zn	-0.09	-0.31	0.39	0.04	-0.31	-0.38	-0.05	-0.62	0.07	-0.40	-0.13	0.38	-0.31	0.01	-0.38	-0.39	0.25	-0.52	1.00	
Pb	-0.29	-0.53	0.82	0.11	-0.26	-0.29	-0.21	-0.52	0.20	-0.37	-0.17	-0.17	-0.47	-0.17	-0.62	0.14	-0.07	-0.38	0.08	1.00

Table 6.9. Post-Monsoon season correlation matrix

	pH	*DO*	*Turbidity*	*BOD_5*	*TDS*	*EC*	*Hardness*	*Alkalinity*	*Na^+*	*K^+*	*Ca^{+2}*	*Mg^{+2}*	*F^-*	*Cl^-*	*SO_4^{2-}*	*NO_3^-*	*Fe*	*Mn*	*Zn*	*Pb*
pH	1.00																			
DO	0.69	1.00																		
Turbidity	0.24	0.06	1.00																	
BOD_5	-0.03	-0.03	0.41	1.00																
TDS	0.22	0.01	0.47	-0.34	1.00															
EC	-0.18	-0.21	0.38	0.04	0.30	1.00														
Hardness	-0.40	-0.20	-0.74	-0.15	-0.61	-0.43	1.00													
Alkalinity	-0.20	-0.13	0.56	0.22	0.38	0.49	-0.57	1.00												
Na^+	0.31	0.20	-0.07	-0.19	-0.03	-0.28	0.16	-0.10	1.00											
K^+	0.50	0.37	0.20	-0.21	0.26	0.53	-0.37	0.19	-0.04	1.00										
Ca^{+2}	-0.26	-0.09	-0.24	-0.07	-0.36	-0.48	0.68	-0.26	0.39	-0.46	1.00									
Mg^{+2}	-0.31	-0.24	0.53	0.23	0.28	0.52	-0.40	0.90	-0.06	0.01	-0.13	1.00								
F^-	0.08	0.23	0.24	0.03	0.18	-0.11	-0.15	0.42	0.19	-0.03	0.34	0.36	1.00							
Cl^-	-0.21	-0.48	0.19	0.28	-0.09	-0.11	-0.07	0.49	0.18	-0.27	0.03	0.54	0.28	1.00						
SO_4^{2-}	-0.15	-0.02	0.31	0.34	-0.04	-0.22	0.02	0.44	0.05	-0.33	0.41	0.49	0.33	0.43	1.00					
NO_3^-	-0.02	-0.01	-0.53	-0.08	-0.52	-0.70	0.59	-0.67	0.09	-0.31	0.31	-0.66	-0.27	0.07	0.01	1.00				
Fe	0.06	0.03	0.04	0.01	-0.04	0.49	0.12	0.20	0.01	0.49	-0.12	0.30	0.04	0.11	0.05	-0.13	1.00			
Mn	-0.04	-0.08	0.14	0.31	-0.17	-0.25	0.02	0.40	-0.02	-0.08	0.13	0.33	0.21	0.63	0.79	0.28	0.24	1.00		
Zn	0.31	0.16	-0.34	-0.19	-0.30	-0.51	0.42	-0.66	0.06	0.10	0.23	-0.70	-0.22	-0.22	-0.35	0.59	-0.02	-0.15	1.00	
Pb	0.33	0.23	-0.29	-0.33	0.00	-0.33	0.37	-0.50	0.38	0.10	0.15	-0.38	-0.20	-0.16	-0.34	0.30	0.07	-0.35	0.68	1.00

PMF of Winter Season

In the case of winter, the number of factors is set to 6 and analysis was carried out to generate parameter contribution profiles into different factors, which would help in identifying the sources of pollution existing in that season (Fig. 6.1). Only parameters contributing more than 30% were considered, gave a fruitful insight into the pollution sources. EC, F^-, and Na^+ have higher contributions to factor 1; BOD_5, hardness, Ca^{+2}, Cl^-, and SO_4^{-2} have higher contributions to factor 2; Ca^{+2}, NO_3^-, and Cd have higher contributions to factor 3; turbidity, BOD_5, and TDS have higher contributions to factor 4; alkalinity, K^+, and Fe have higher contributions in factor 5; and lastly Zn have higher contributions into factor 6. The compositions of the different factors allow us to identify or speculate on the sources of pollution. Factors 1 and 5 relate to soil or geogenic metal pollution. Factors 3 and 6 are related to the anthropogenic fraction of metal pollution, and factors 2 and 4 are related to domestic pollution. For PMF analysis in the winter season, domestic and metal pollution are the sources of pollution that deteriorate the river water quality.

PMF of Pre-monsoon Season

Here, the number of factors is set to 5, and the analysis generates parameter contribution profiles (Fig. 6.2). pH, DO, TDS, EC, alkalinity, Na^+, K^+, SO_4^{-2}, and Pb have high contributions to factor 1; BOD_5 and Mn have high contributions to factor 2; hardness, alkalinity, Ca^{+2}, SO_4^{-2}, Zn, and Cl^- have high contributions factor 3; turbidity, NO_3^-, and Fe have high contributions to factor 4; lastly Mg^{+2}, Cd, F^-, and Cl^- have high contributions to factor 5. These compositions have helped to identify the sources of pollution and represent them as different factors. Factors 1, 2, and 3 represent domestic pollution sources; factor 4 represents runoff sources and factor 5 represent soil or geogenic metal pollution source. The PMF analysis carried out for the pre-monsoon season also reveals domestic, runoff and metal pollution as the sources of pollution.

PMF of Monsoon Season

Here, the number of factors is set to 5, and the analysis generates parameter contribution profiles (Fig. 6.3). DO, BOD_5, EC, K^+, Cl^-, NO_3^- and Mn have high contributions to factor 1; hardness, Ca^{+2}, F^-, and NO_3^- have high contributions to factor 2; Pb has high contributions to factor 3; TDS, alkalinity, Mg^{+2}, SO_4^{-2}, and F^- have high contributions to factor 4; and lastly Fe and Mn have high contributions to factor 5. The compositions of the factors give away sources of pollution in the river. Factor 1 is related to domestic pollution; factors 2 and 4 are related to surface or agricultural runoff; factor 3 is related to anthropogenic metal pollution; and factor 5 is related to geogenic metal pollution. The PMF analysis for the monsoon season reveals three pollution sources: domestic, runoff and metal pollution.

PMF of Post-Monsoon Season

Here, the number of factors is set to 6 and the analysis is carried out to generate parameter contribution profiles (Fig. 6.4). BOD_5, hardness, alkalinity, Ca^{+2}, Mg^{+2}, and F^- have high contributions to factor 1; TDS and Zn have contributions to factor 2; pH, DO, turbidity, TDS, EC, Cl^-,

and NO_3^- have high contributions to factor 3; hardness, Ca^{+2}, and Pb have high contributions to factor 4; and lastly K^+ and NO_3^- have high contributions to factor 6. The compositions of the different factors reveal the sources of pollution existing in this season. Factors 1 and 3 represent the domestic pollution source; factor 2 represents soil or geogenic metal pollution source; factor 4 represents the anthropogenic metal pollution source; and lastly, factor 6 also represents the soil or geogenic metal pollution source. Two pollution sources exist in the post-monsoon season: domestic and metal pollution sources.

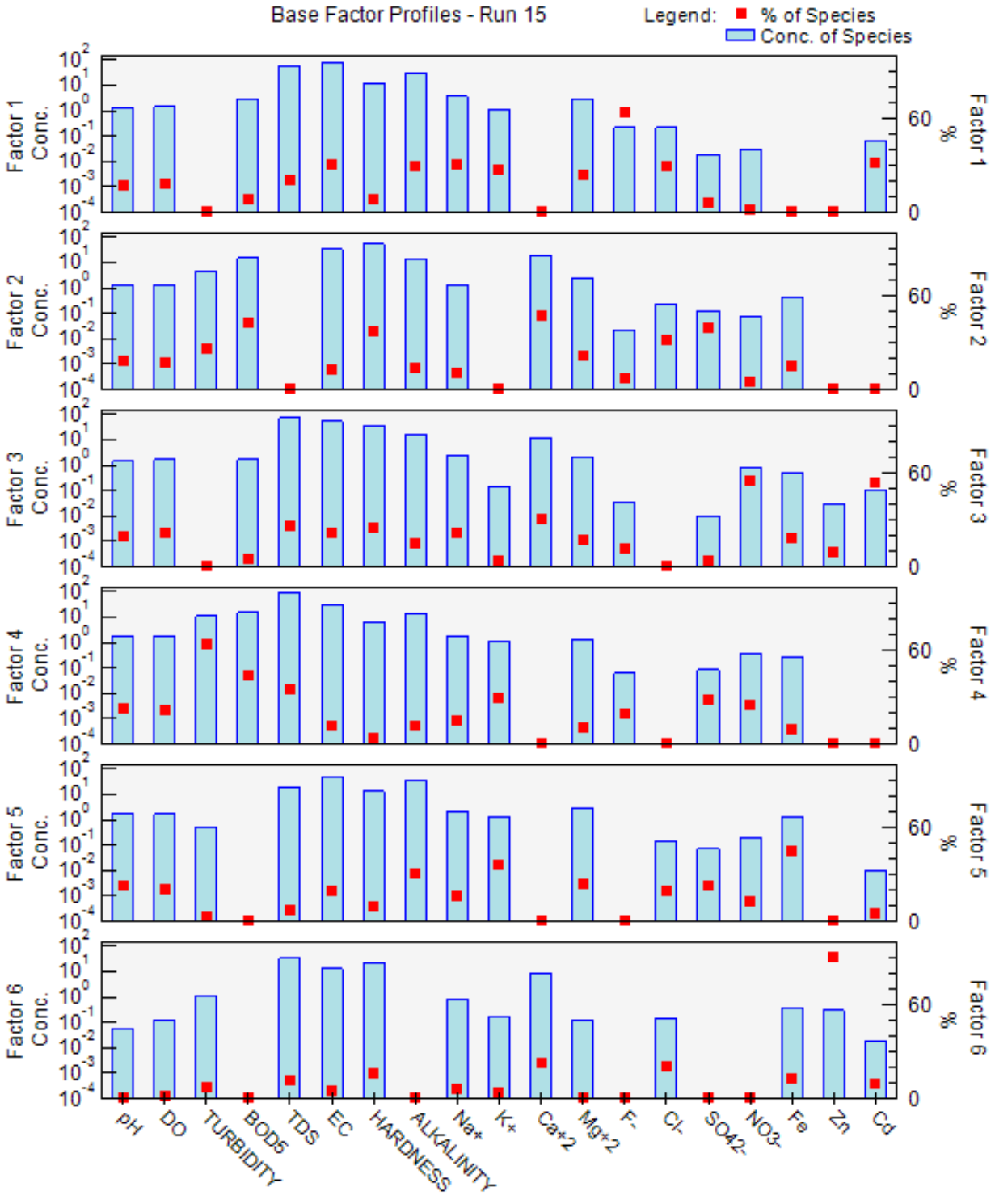


Fig. 6.1. Parameter contribution (%) to the factors for the winter season

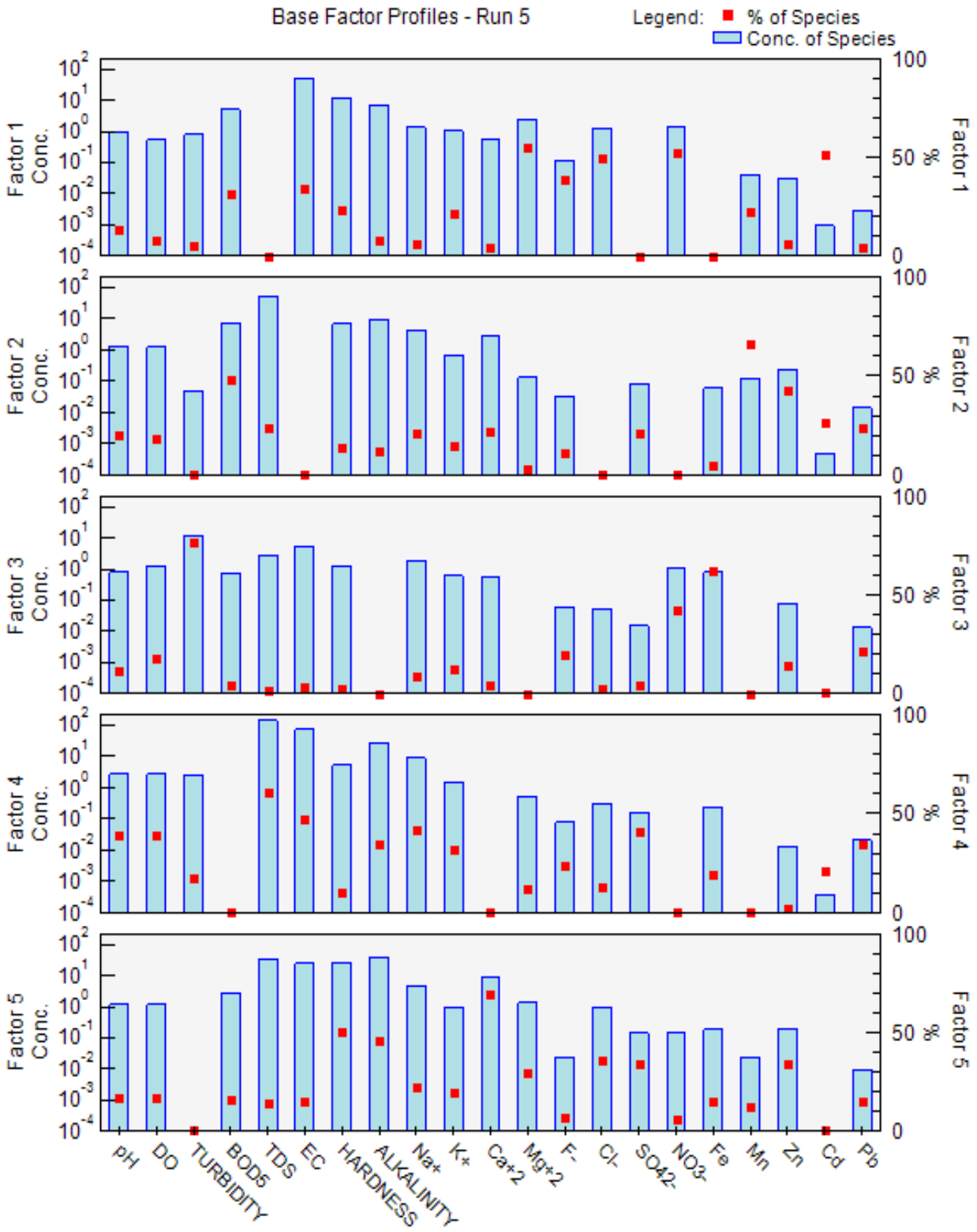


Fig. 6.2. Parameter contribution (%) to the factors for pre-monsoon season

The PMF results align with the PCA and Pearson correlation matrix carried out in this study. Pollution in the months of winter and post-monsoon were attributed to domestic discharges and metal pollution. In comparison, another pollution source gets added in pre-monsoon and monsoon months due to heavy precipitation in the region, which brings in agricultural or surface runoff pollution into the river.

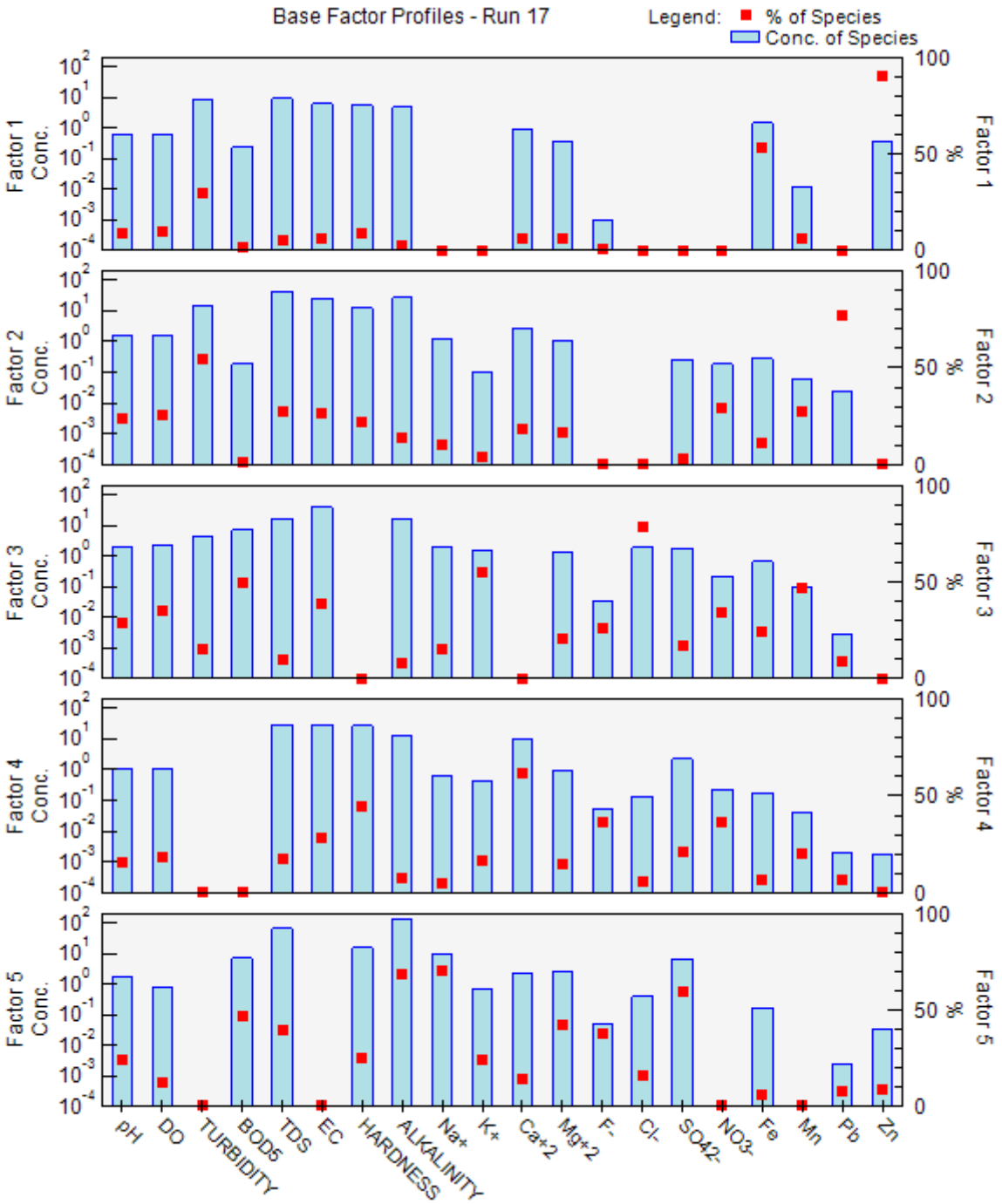


Fig. 6.3. Parameter contribution (%) to the factors for monsoon season

6.1.3 Hierarchical Cluster Analysis (HCA)

In the sections, hierarchical cluster analysis (HCA) is carried out on the sampling sites in different seasons to group them into different clusters and understand their classification based on WQI variation. This analysis will help to find out the most probable locations of point and non-point sources of pollution existing in the vicinity of the river. In the winter season (Fig. 6.5), HCA classified the sampling sites into three clusters. The 1st cluster contained the sampling sites from 1 to 4, the 2nd cluster contained the sampling sites from 5 to 16, and the 3rd cluster contained the final two sampling sites (17 and 18). From, Fig. 5.8 and 5.9, clusters 1 and 3 had EWQI values in the range of 'poor' or 'extremely poor', whereas the sites in cluster 2 had EWQI values in the

'average' range. HCA classified the sampling sites into three clusters in the pre-monsoon season (Fig. 6.6).

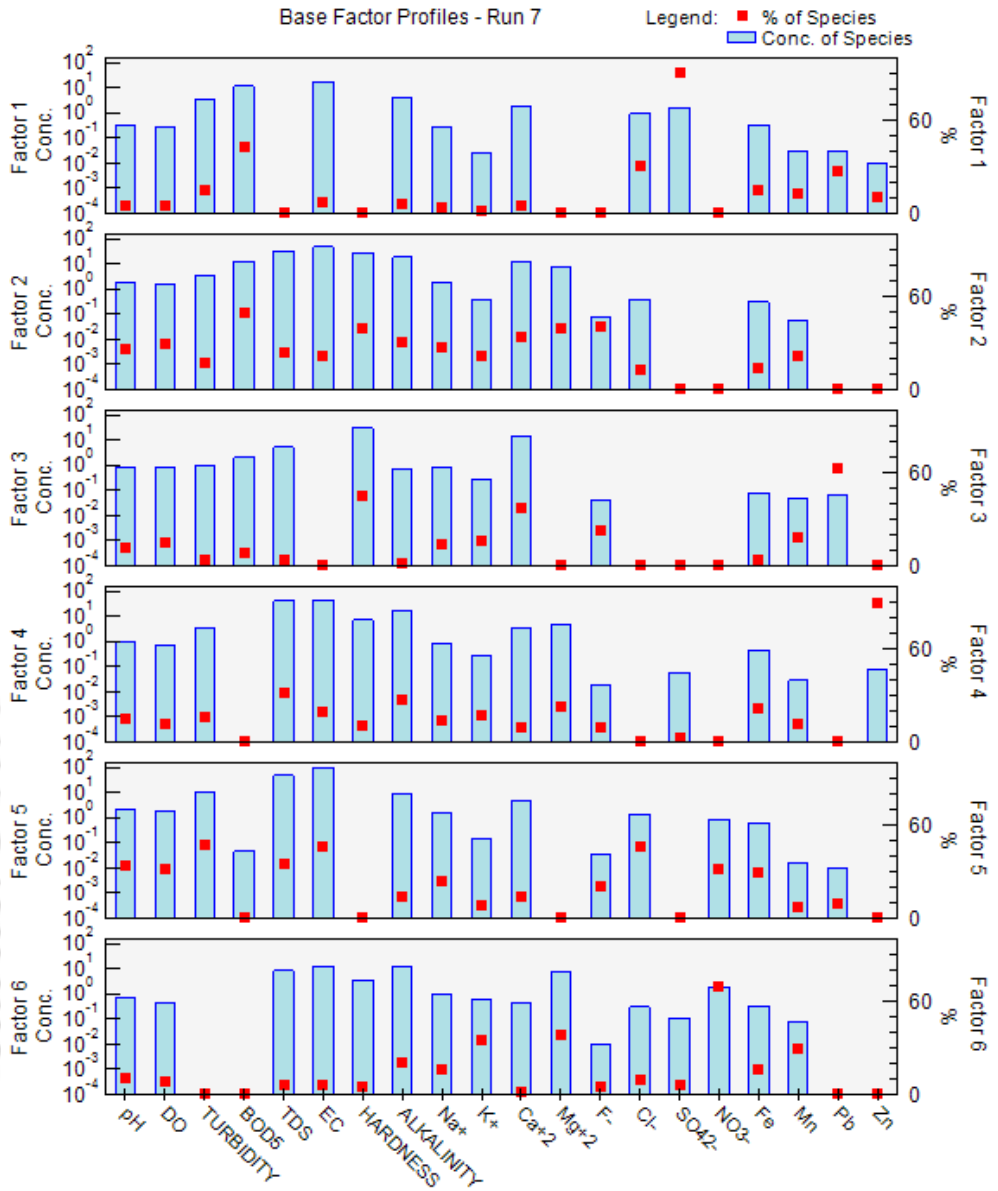


Fig. 6.4. Parameter contribution (%) to the factors for post-monsoon season

The 1st cluster contained sampling sites from 1 to 6, 2nd cluster contained sites located on the middle stretch of the river (from 7 to 12), and the 3rd cluster contained sites from 13 to 18. From Fig. 4.7, the 1st and the 3rd cluster contained the sites whose EWQI values ranged from 'poor to 'average', and the 2nd cluster contained sites whose EWQI values ranged from 'average' to 'good'. From winter to the pre-monsoon season, the EWQI values diluted, possibly due to increased precipitation over the seasons. In the monsoon season (Fig. 6.7), the clusters are exactly the same as in the winter, with the difference, EWQI values are even lower (i.e. the quality is improving) due highest precipitation values in the season, which have a dilution effect on the values of the different pollutant sources. The highest variation of the dataset is observed in the post-monsoon season, giving rise to four clusters (Fig. 6.8). From analysis, the identification of the non-point sources can be made from post-monsoon season data. 1st cluster contains the site from 1 to 4, 2nd

cluster contains the site from 5 to 8, 3rd cluster contains the site from 9 to 16, and 4th cluster contains sites 17 and 18. The 1st and 4th cluster contain the sites whose WQI value ranged from 'poor' to 'average', whereas the 2nd and 3rd cluster contains the site whose WQI values ranged from 'poor' to extremely poor'. The variation over winter, pre-monsoon and monsoon are similar, with the middle stretch showing better water quality status than the end stretches. However, the opposite happened in post-monsoon season, where the middle stretch showed degraded water quality status indicating probable locations of non-point sources, especially from increased precipitation over earlier months. Thus, this middle stretch, i.e. clusters 2 and 3 in the post-monsoon season, can be said to be the locations of metal pollution. This can be correlated with the analysis of PCA, Pearson correlation matrix, PMF and WQIs, wherein the post-monsoon season has a greater contribution or strong correlation from different heavy metals. Also, in the pre-monsoon season, cluster 3 (final cluster) has more number sampling sites. Thus, these sites can be inferred to be sites of agricultural or surface runoff in the pre-monsoon season, as can be observed from PCA, Pearson correlation matrix, and PMF. Thus, the sites from 9 to 16 can be said to be sites of anthropogenic pollution or non-point sources of pollution in the river, and the rest of the sites can be said to be sites of domestic discharge or geogenic metal pollution.

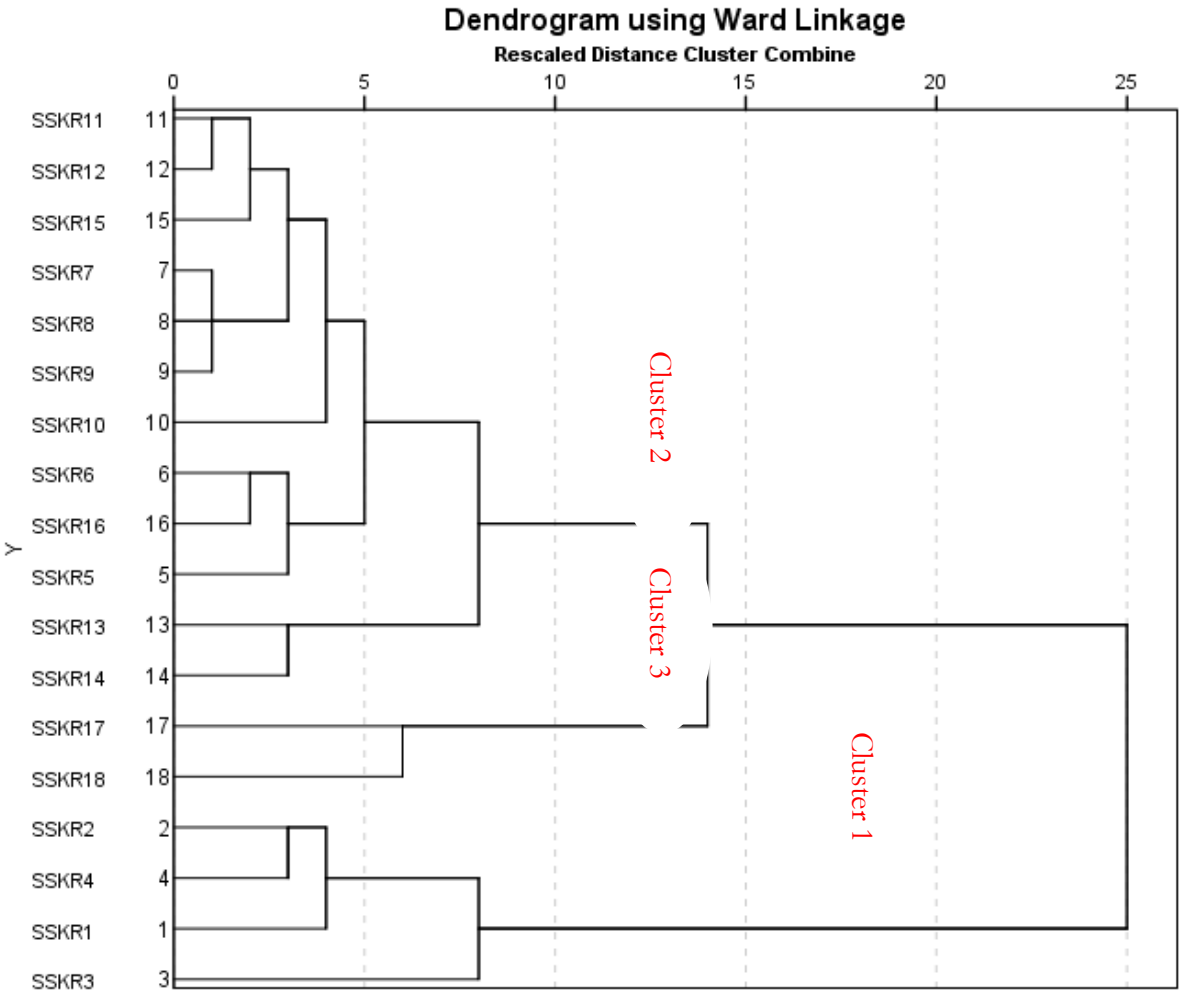


Fig. 6.5. HCA results of the winter season

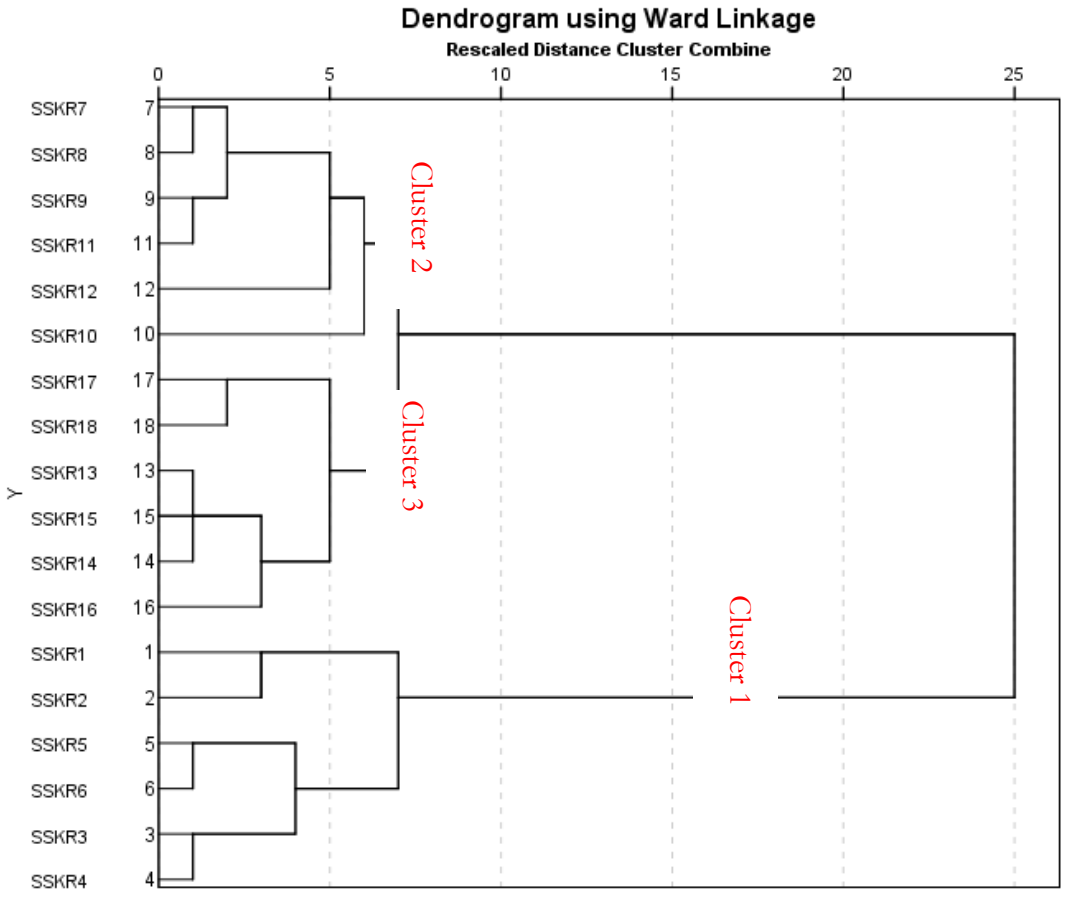


Fig. 6.6. HCA results of the pre-monsoon season

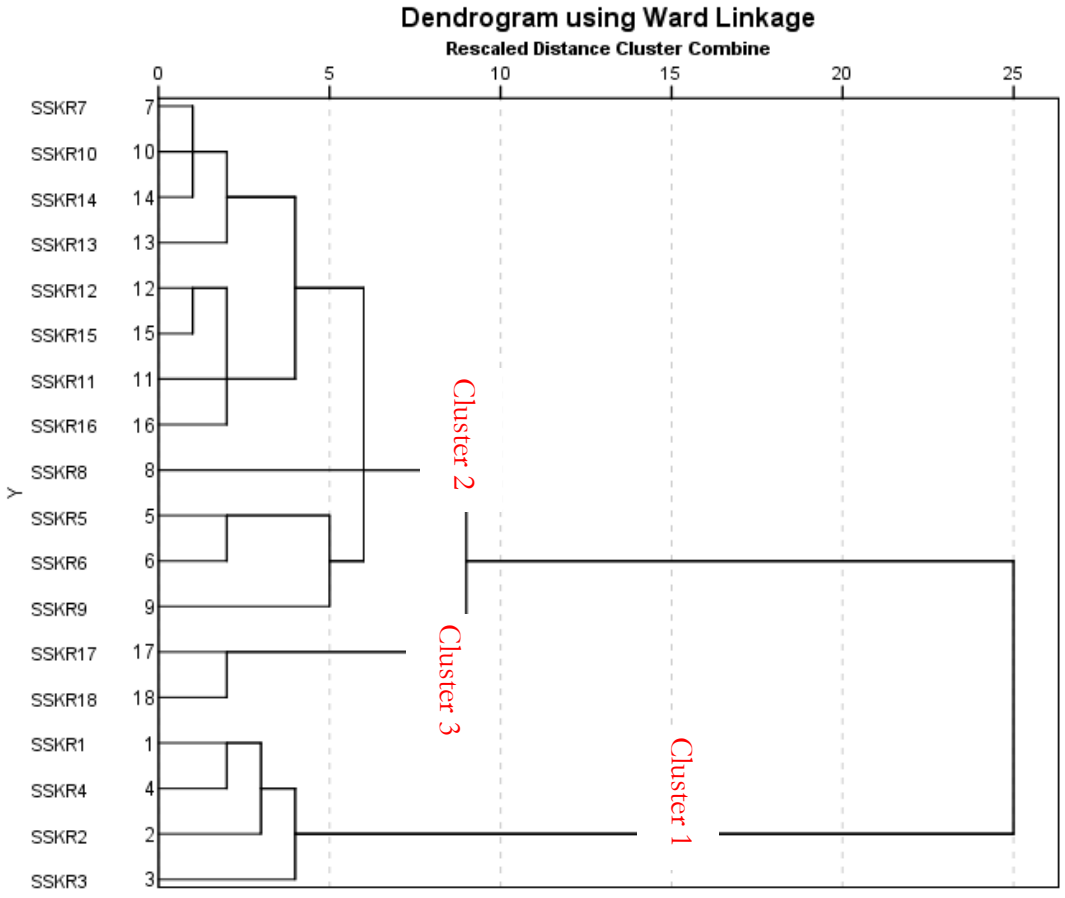


Fig. 6.7. HCA results of the monsoon season

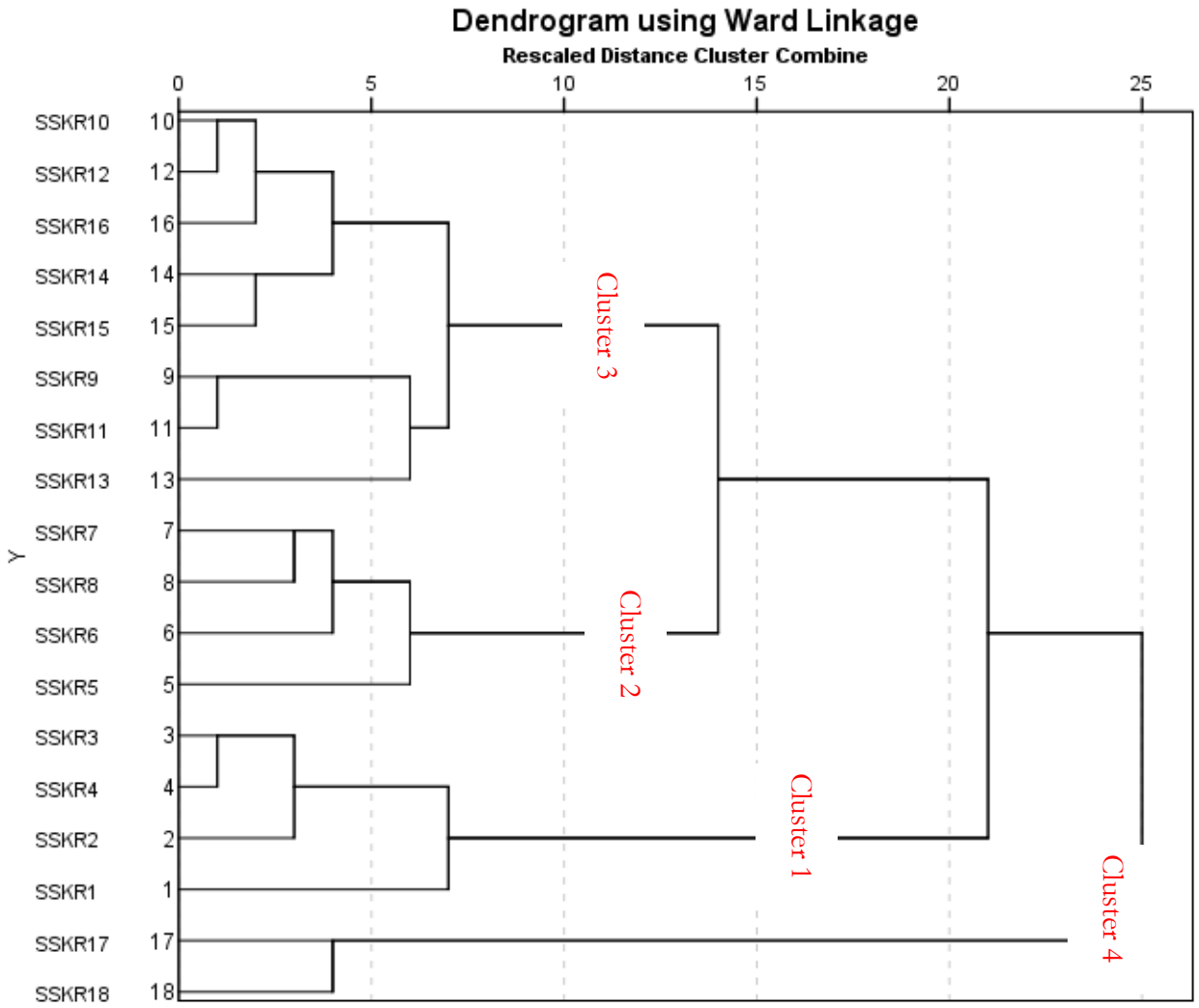


Fig. 6.8. HCA results of the post-monsoon season

6.2 Analysis of Sediment Quality Dataset

6.2.1 Most probable source identification

6.2.1.1 Using Principal Component Analysis (PCA)

PCA was conducted with SPSS 25 to classify bioavailable heavy metal sources in the river's surficial sediment. The PCA loadings in the current dataset have been further eliminated by contributions of major parameters (<0.75 factor score) (Singh et al., 2020a). The resulting values are shown in Table 6.10, along with their eigenvalues and the percentage of variances for sorted rotated factor loading values. 2 factors explain approximately 80% of the total variance. Among the 2PCs, PC1 explains 45.11 %, and PC2 explains 34.80 % of the total variance, respectively. PC1 is largely correlated to F1 and F2 fractions of Cu, Fe, and Mn, which is exchangeable fraction and carbonate bound species of the metal's contamination. In the case of Pb, all the fractions have a high correlation to PC1. PC2 is correlated to F1 and F2 fractions of Cd and Pb, which accounts for exchangeable fraction of the metal's contamination; F3 fraction of Mn, which again accounts for the Fe/Mn oxides bound species of the metal contamination; and F3, F4, F5 fraction of Zn which accounts for the species bound to Fe/Mn oxides, species bound to sulphides and organic matter, and residual fraction of the metal's contamination respectively (Förstner and Salomons,

1980b; Gibbs, 1973; Salomons and Förstner, 1980). After carrying out PCA, most of the metal contamination in the surficial sediment is largely contributed to F1, F2, and F3 fractions, which accounts for the anthropogenic contamination and, to a small extent, by F5 fraction, which is described as the residual or the inert fraction, i.e. they unlikely to be released by the sediments into the stream (Morillo et al., 2002; Pardo et al., 1993). Therefore, from PCA, it can be deduced that the metal contamination in the Kolong river surficial sediment is largely from anthropogenic sources.

6.2.1.2 Pearson correlation among metal fractions

The association of metal fractions can further be evaluated using lithogeneous (F4 and F5 metal fractions) and anthropogenic (F1, F2, and F3 metal fractions) metals' shares, which are calculated by adding the exchangeable, carbonate bound, and Fe/Mn-oxide bound phases (Förstner, 1982; Karbassi et al., 2011). The disparities between heavy metal fractions highlighted the transit of heavy metals and the sedimentary processes. Correlation analysis has therefore been performed to determine the effects of different factors. Table 6.11 shows the Pearson correlation coefficients between the heavy metal contents in the sediments. Fe, Mn, Cu and Cd concentrations were substantially correlated, with correlation coefficients above 0.68. These significant correlations suggest that the heavy metals came from similar sources and accumulated excessively in river sediments. All metals (Zn, Mn, Fe, Cu, Co, Pb, Cd) are highly correlated (R^2>0.83) to their respective concentrations in anthropogenic and lithogenic fractions. These significant relationships suggest that increasing heavy metal loadings will encourage heavy metal accumulation and that heavy metals will be transformed into more easily extractable components in the sediments. (Gujre et al., 2021; Liu et al., 2020; Qin et al., 2021).

6.2.1.3 Positive Matrix Factorization (PMF)

Q values were computed after iterations with several factors ranging from 2 to 7. Four factors were considered ideal because Q_{True}/Q_{exp} for iteration, in this case, is the least (Table 6.12). The model's fitness was determined by comparing observed and predicted heavy metal values, yielding the correlation coefficients (R^2). R^2 values greater than 0.95 were found in all heavy metals, indicating a strong relationship between predicted and observed values and, therefore, a high level of model dependability (Table 6.13).

Heavy metal toxicity to human health and the environment involves utilisation of source allocation models, viz. PMF model to decrease the likelihood of future contamination and, as a result, to more efficiently manage resources. The current investigation considers four factors that influence the source of pollution in the Kolong River. From Fig. 6.9, it can be observed that the contribution to factor 1 is shown only by Mn, Fe, Cu, Co, and Cd. The presence of Cu, Co and Cd can mainly be attributed to anthropogenic sources of contamination from the nearby region. So, factor 1 can be said to be related to anthropogenic contamination. In the rest of the factors, it is observed that the contribution is only from Fe, Mn, and Zn. As reported in various works of

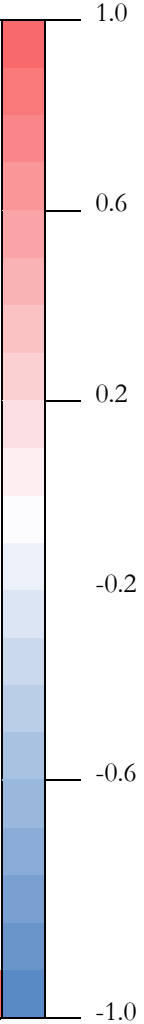
literature, the source of these metals in northeast India is natural or lithogenic (Borah et al., 2009; Haloi and Sarma, 2012). Thus, from source apportionment analysis using PMF, it can be concluded that among the four factors, factor 1 can be related to anthropogenic contamination and factor 2, 3 and 4 can be said to be associated with natural or lithogenic contamination.

Table 6.10. Rotated component matrix loading score for the metal speciation fractions in Kolong river sediments

Variables	Component	
	1	2
CdF1	-	-0.76
CdF2	-	-0.82
CdF3	-	-
CdF4	-	-
CdF5	-	-
CoF1	-	-0.85
CoF2	-	-0.80
CoF3	-	-
CoF4	-	-
CoF5	-	-
CuF1	0.93	-
CuF2	0.93	-
CuF3	-	-
CuF4	-	-
CuF5	-	-
FeF1	0.87	-
FeF2	0.79	-
FeF3	-	-
FeF4	-	-
FeF5	-	-
PbF1	-0.97	-
PbF2	-0.96	-
PbF3	-0.96	-
PbF4	-0.96	-
PbF5	-0.95	-
MnF1	-	-
MnF2	0.74	-
MnF3	-	0.89
MnF4	-	-
MnF5	-	-
ZnF1	-	-
ZnF2	-	-
ZnF3	-	0.84
ZnF4	-	0.88
ZnF5	-	0.94
Eigenvalues	15.78	12.17
% of Variance	45.10	34.79
Cumulative %	45.10	79.90
Remarks	Fe/Mn Bound Species & Carbonate Bound Species (Except for Pb)	Fe/Mn Bound Species, Fe/Mn oxides Bound Species & Carbonate Bound Species (Except for Zn)

Table 6.11. Correlation analysis of total concentration and geochemical fractions of heavy metals

	ZnA	ZnN	ZnT	MnA	MnN	MnT	PbA	PbN	PbT	FeA	FeN	FeT	CuA	CuN	CuT	CoA	CoN	CoT	CdA	CdN	CdT
ZnA	1.00																				
ZnN	0.43	1.00																			
ZnT	0.85	0.84	1.00																		
MnA	0.07	-0.40	-0.19	1.00																	
MnN	-0.10	-0.43	-0.31	0.75	1.00																
MnT	0.00	-0.44	-0.26	0.95	0.92	1.00															
PbA	-0.49	-0.54	-0.61	0.50	0.69	0.62	1.00														
PbN	-0.48	-0.54	-0.60	0.36	0.56	0.48	0.85	1.00													
PbT	-0.50	-0.56	-0.62	0.41	0.61	0.53	0.92	0.99	1.00												
FeA	0.29	0.40	0.41	0.31	-0.13	0.12	-0.18	-0.29	-0.27	1.00											
FeN	0.01	-0.58	-0.34	0.71	0.93	0.86	0.56	0.50	0.54	-0.11	1.00										
FeT	0.12	-0.41	-0.17	0.80	0.85	0.88	0.47	0.38	0.42	0.28	0.93	1.00									
CuA	0.03	-0.32	-0.17	0.78	0.81	0.85	0.41	0.26	0.31	0.28	0.82	0.90	1.00								
CuN	-0.22	-0.59	-0.48	0.85	0.74	0.85	0.52	0.31	0.38	0.01	0.69	0.67	0.61	1.00							
CuT	-0.13	-0.54	-0.39	0.91	0.85	0.94	0.53	0.32	0.39	0.13	0.82	0.84	0.85	0.93	1.00						
CoA	0.51	0.10	0.36	0.71	0.39	0.61	0.07	-0.17	-0.11	0.76	0.41	0.69	0.62	0.46	0.59	1.00					
CoN	-0.07	-0.27	-0.20	0.19	0.00	0.11	0.60	0.47	0.52	0.05	-0.04	-0.02	-0.18	0.11	-0.01	0.10	1.00				
CoT	0.13	-0.20	-0.04	0.44	0.15	0.33	0.55	0.35	0.42	0.34	0.12	0.24	0.08	0.28	0.22	0.47	0.92	1.00			
CdA	0.32	-0.04	0.17	0.68	0.55	0.67	0.45	0.50	0.51	0.47	0.56	0.72	0.58	0.29	0.45	0.67	0.34	0.56	1.00		
CdN	0.02	-0.05	-0.01	0.71	0.89	0.84	0.49	0.37	0.41	0.12	0.74	0.76	0.85	0.55	0.74	0.49	-0.19	0.02	0.60	1.00	
CdT	0.15	-0.05	0.06	0.78	0.85	0.86	0.53	0.46	0.50	0.28	0.74	0.82	0.83	0.50	0.70	0.61	0.01	0.24	0.83	0.94	1.00



A: Anthropogenic, N: Natural/Lithogenic, T: Total Metal Concentration

Table 6.12. Summary of PMF for different runs for heavy metals in surficial sediments in Kolong River

Factor	Qexp	Q(Robust)	Q(True)	Q(true)/Qexp
2	31	8.8501	8.8500	0.2855
3	15	1.7390	1.7390	0.1159
4	-1	0.8216	0.8216	-0.8216
5	-17	0.3065	0.3062	-0.0180
6	-33	0.0063	0.0058	-0.0002

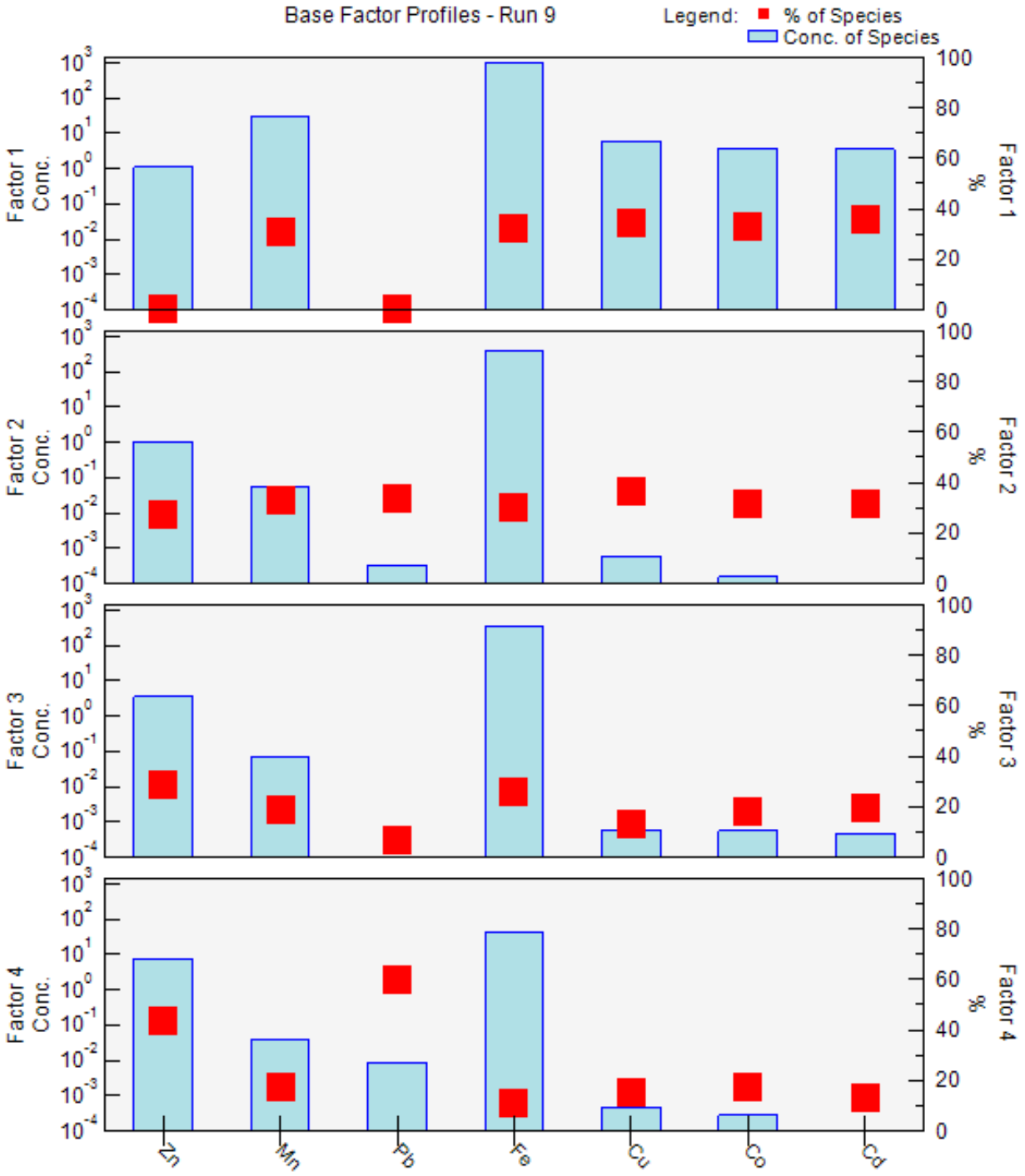


Fig. 6.9. Heavy metals (%) contribution to the 4 factors

Table 6.13. Correlation Coefficients (R^2) of heavy metals between observed and predicted values by the PMF model

Species	Intercept	Slope	SE	R^2
Zn	0.85	1.00	4.18	1.00
Mn	2.18	0.98	3.03	0.96
Pb	0.30	0.98	0.11	0.99
Fe	55.47	0.98	97.61	0.97
Cu	0.26	0.98	0.22	0.99
Co	0.13	0.99	0.13	0.99
Cd	0.10	0.99	0.10	1.00

6.3 Summary

This chapter gives a detailed insight into sources of the variation in the river water and benthic sediments of the Kolong River using environmetrics tools. Critical concluding remarks from the study:

Water Quality

(i) Hierarchical cluster analysis (HCA) clusters the 18 sampling sites into different clusters in different seasons. HCA clustered the sampling sites into three clusters in winter, pre-monsoon and monsoon season and four clusters in post-monsoon season. Sites in the middle stretch, i.e. from 9 to 16, were polluted due to anthropogenic metal pollution or runoff, and the rest were polluted due to domestic discharge or geogenic metal pollution.

(ii) From Discriminant analysis (DA), only 10 (turbidity, TDS, EC, hardness, alkalinity, Na^+, Ca^{+2}, Mg^{+2}, F^-, SO_4^{2-} and Cd) out of 21 parameters were responsible for the temporal variability of the water quality dataset of the studied area. Turbidity, TDS, EC, hardness, and alkalinity explain the domestic discharge quotient, Na^+, Ca^{+2}, Mg^{+2}, and F^- explains soil pollution contribution to river pollution, SO_4^{-2} explains mostly the agricultural runoff quotient, and Cd explains the anthropogenic or geogenic metal pollution quotient of the seasonal variation.

(iii) From PCA, Pearson correlation, and PMF, that pollution in the months of winter and post-monsoon were attributed to domestic discharges and metal pollution in the winter and post-monsoon months. Whereas, another pollution source gets added in the monsoon and pre-monsoon months due to heavy precipitation in the region, which brings the agricultural or surface runoff quotient of pollution into the river. Nevertheless, PMF was able to segregate the geogenic and the anthropogenic metal pollution. Geogenic metal pollution was attributed to Ca^{+2}, Mg^{+2}, Fe, Mn, and Zn; anthropogenic metal pollution was attributed to Cd and Pb. These findings are also inline benthic sediment quality study.

Benthic Sediment Quality

(i) From PCA, Pearson correlation matrix and PMF, benthic sediments were mainly contaminated with heavy metals from anthropogenic sources. Nevertheless, PMF helped to segregate these sources. Geogenic contamination is due to Fe, Mn, and Zn because of the higher presence of F4 and F5 fractions, whereas anthropogenic contamination is due to Co, Cu and Cd because of the higher presence of F1, F2, and F3 fractions. These results helped confirm the metal pollution sources in Kolong River water samples.

Chapter 7

HUMAN HEALTH RISK ASSESSMENT

Human health is jeopardised if hazardous chemicals are found in surface water and benthic sediment samples. The dangers of prolonged exposure to heavy metals on the health of the human population (both adults and children) were investigated in this study.

7.1 Surface Water

7.1.1 Deterministic Method

The results for deterministic approach calculation for health risk assessment through different exposure routes on the human population are shown in Fig. 7.1. For non-carcinogenic risk through ingestion pathway, the HQ values are less than 1 for both adults and children, with the highest risk being present by Zn. For non-carcinogenic risk associated with dermal contact, i.e. their HQ values are also less than 1 for Fe, Mn, Pb and Cd but for Zn is quite high with a maximum of 10.02 for and 14.80 for children. The non-carcinogenic risk values were higher for children than adults putting the children at higher risk than the adults. The cumulative effect for the pathways is represented by HI, shown in Fig. 7.2. For the high value of Zn, the overall effect of non-carcinogenic risk assessment values shows unacceptable values, i.e. greater than 1. There is rise in HI and HQ values in pre-monsoon season, indicating that with rise in precipitation, anthropogenically pollutants are added into the stream, which gives higher risk values. But, the height of monsoon dilutes these values.

The carcinogenic risk assessment (TCR) results are shown in Fig. 7.2. Though there is variation in the values of TCR throughout the different seasons, TCR values never went beyond the seasons of analysis limit of 10^{-5}. The maximum TCR is 1.91×10^{-6} for adults and 4.93×10^{-6} for children. Thus, the surface water of Kolong River does not present any carcinogenic risk for the inhabitant population but has a higher degree of non-carcinogenic risk, with the dermal route being the point of exposure for the population. Values present a fixed risk through the deterministic approach, but probabilistic approach calculation needs to be carried out to understand the full extent of the risk.

7.1.2 Probabilistic Method

The values of any parameter vary temporally and spatially during the sampling period; calculating risk using deterministic point values would give us an idea about the hazard posed by that

parameter but would present a full picture of the risk posed for the analysis. Also, quantifying every point at different points in time would be tedious and may result in calculation errors. In such cases, the probabilistic method of risk calculation would give a more descriptive result of the parameter variation and help in quantifying the spatial and temporal risk posed by the parameter. In this method, the data of heavy metals collected for different seasons is first fitted to a triangular distribution, and different parameters are set to different distributions, as shown in Table 3.7. This is then run through a Monte-Carlo simulation (10000 iterations) to quantify the cancerous and non-cancerous risk posed through the different routes of exposure, i.e. ingestion and dermal. The 5th and 95th percentile, i.e. maximum and minimum values of probabilistic analysis for HI and TCR, are shown in Fig 7.3, and their frequency-probability distributions graphs are shown in Fig. 7.4-7.7. Both the values show similar variations, which is rise in risk in the pre-monsoon season and decrease after that, thus indicating the dilution power of the river. The HI values for all seasons were greater, indicating non-cancerous risk for adults and children. In cancerous risk assessment (TCR), the values are mostly less than 1×10^{-5} for most of the year, but post-monsoon season possesses considerable risk as the 5th percentile value is greater than the limit. Cancer causing metals are entering the river ecosystem is a cause of concern. From this assessment, children are at higher risk compared to adults. Therefore, the use of Kolong River water should be treated for heavy metals before consumption for any particular purpose.

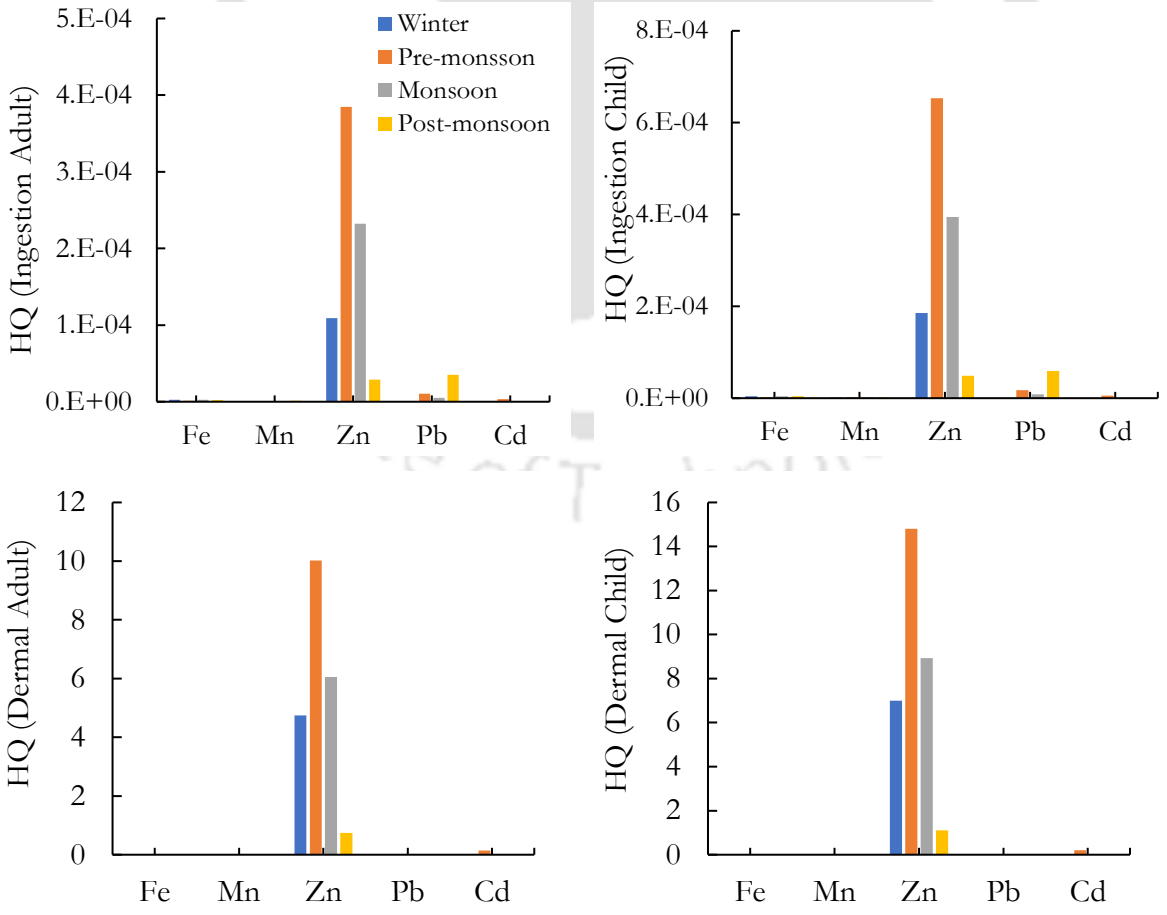


Fig. 7.1. Hazard Quotient (HQ) for ingestion and dermal contact in surface water of Kolong river

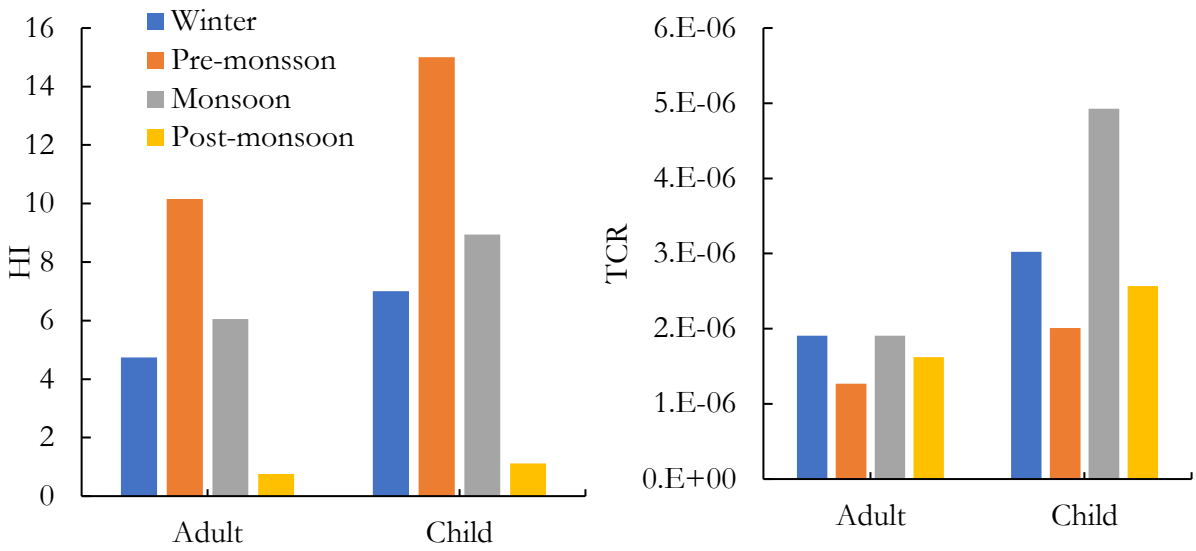


Fig. 7.2. Hazard index (HI) and Total carcinogenic risk (TCR) for the adults and children for surface water samples through the deterministic method

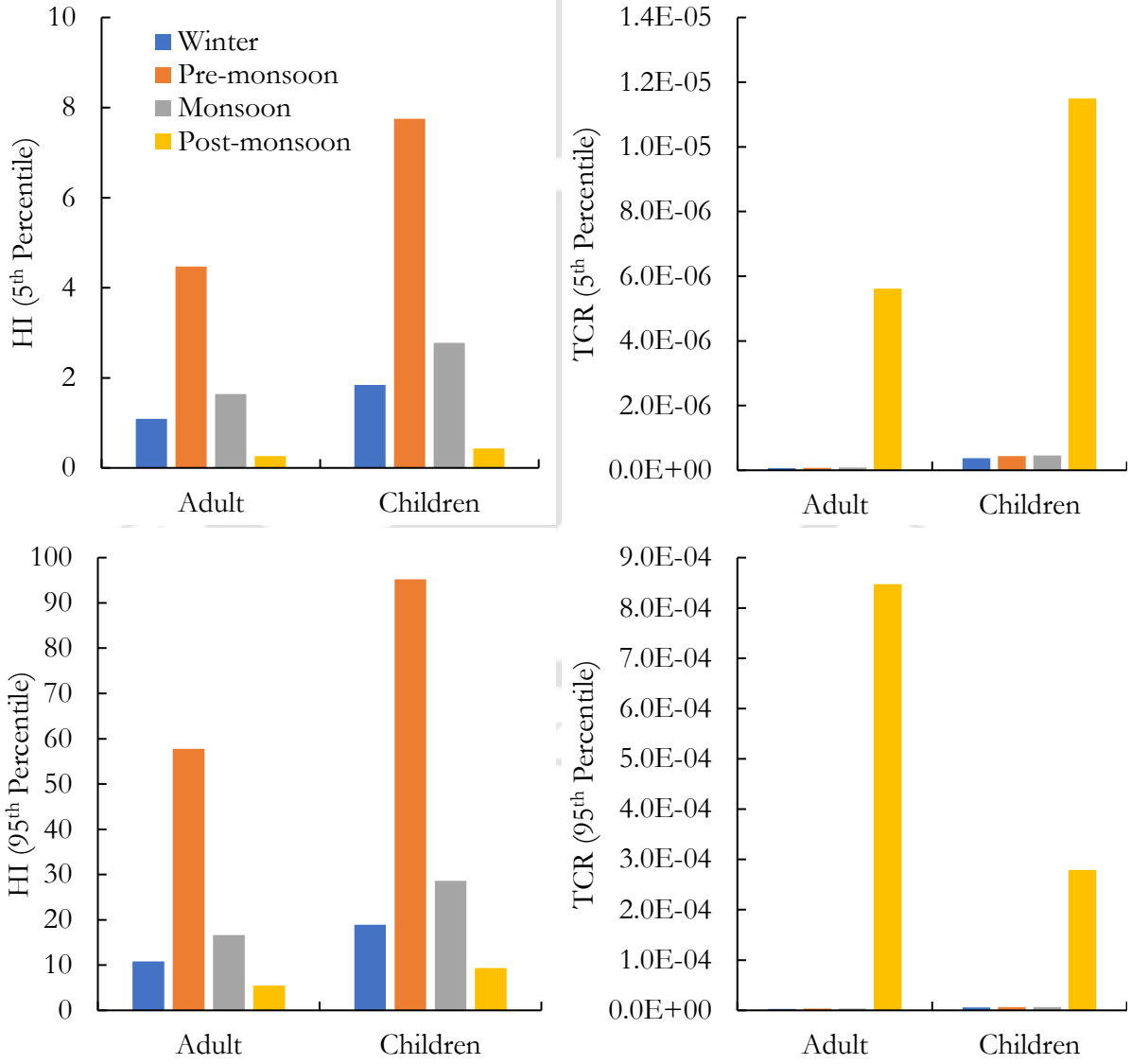


Fig. 7.3. HI and TCR from exposure to Heavy Metals in surface water through probabilistic method (5th and 95th percentile)

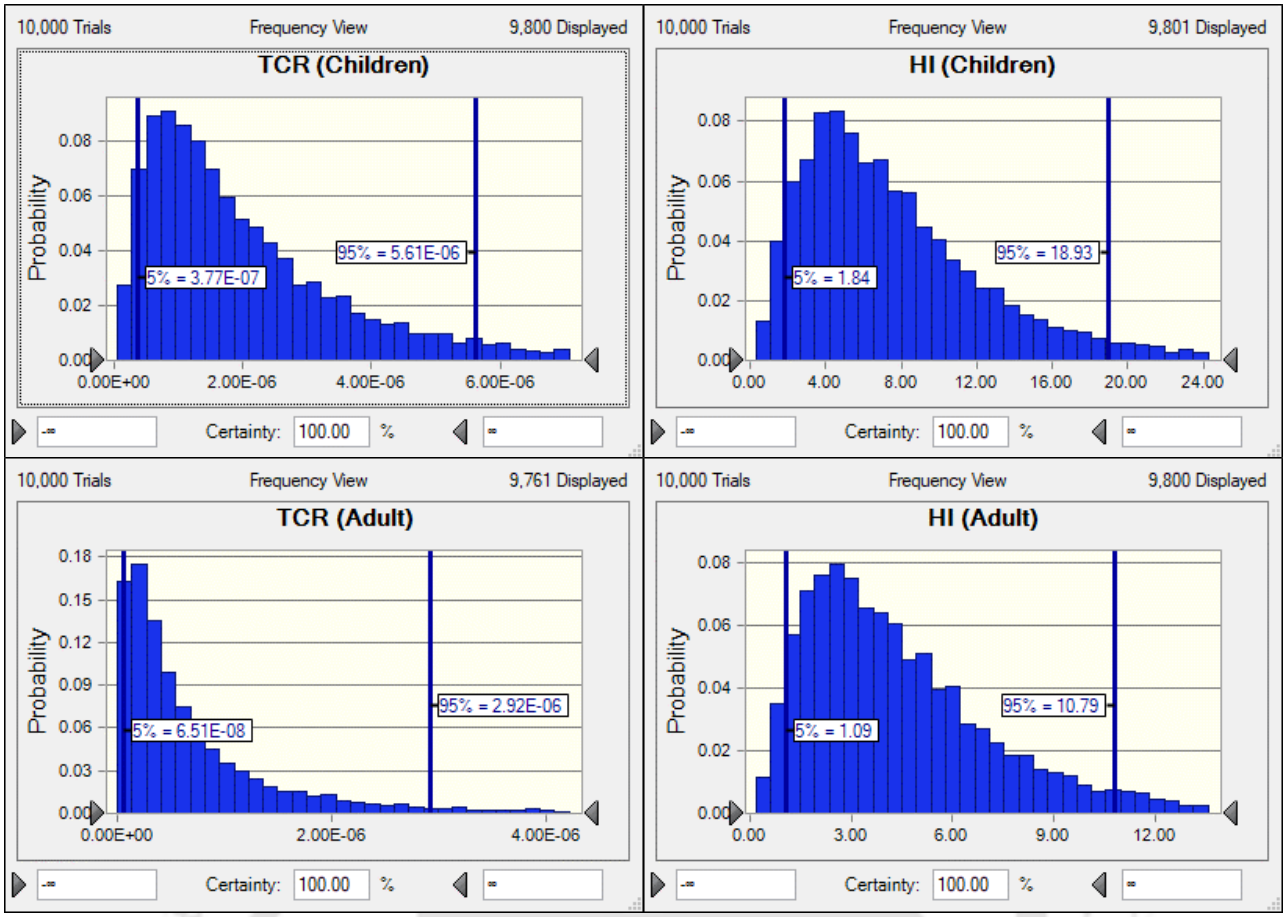


Fig. 7.4. Frequency-probability distribution for winter for surface water

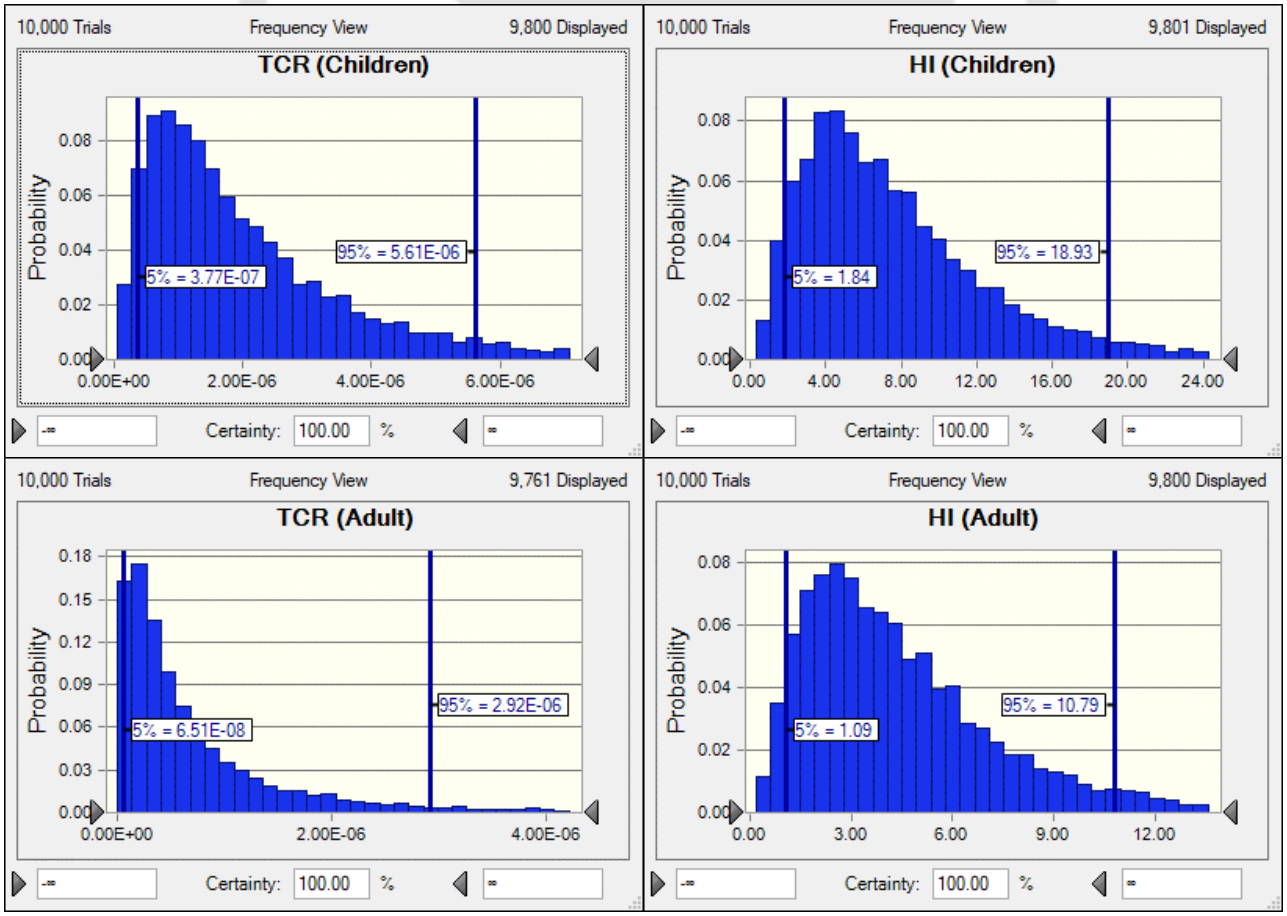


Fig. 7.5. Frequency-probability distribution for pre-monsoon for surface water

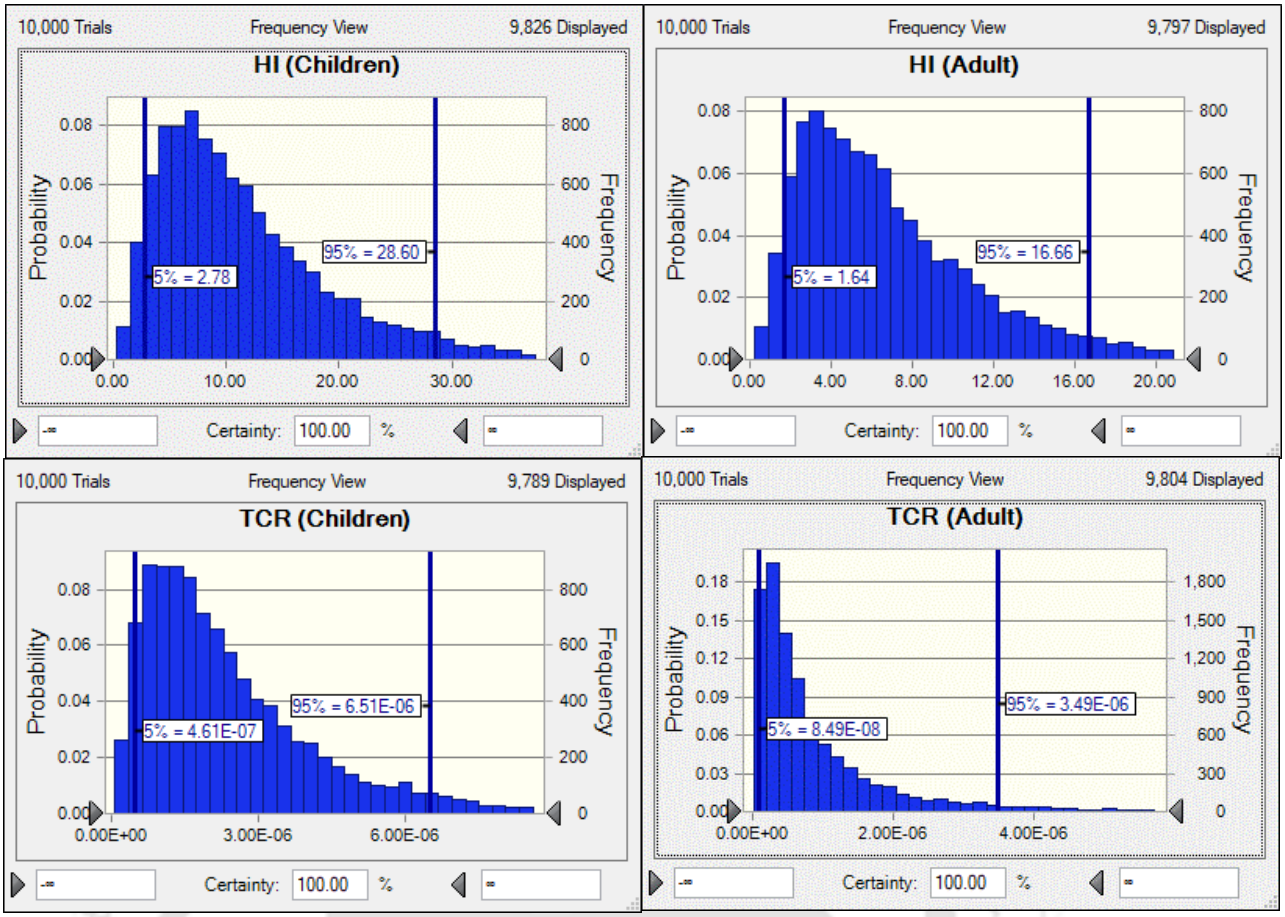


Fig. 7.6. Frequency-probability distribution for monsoon for surface water

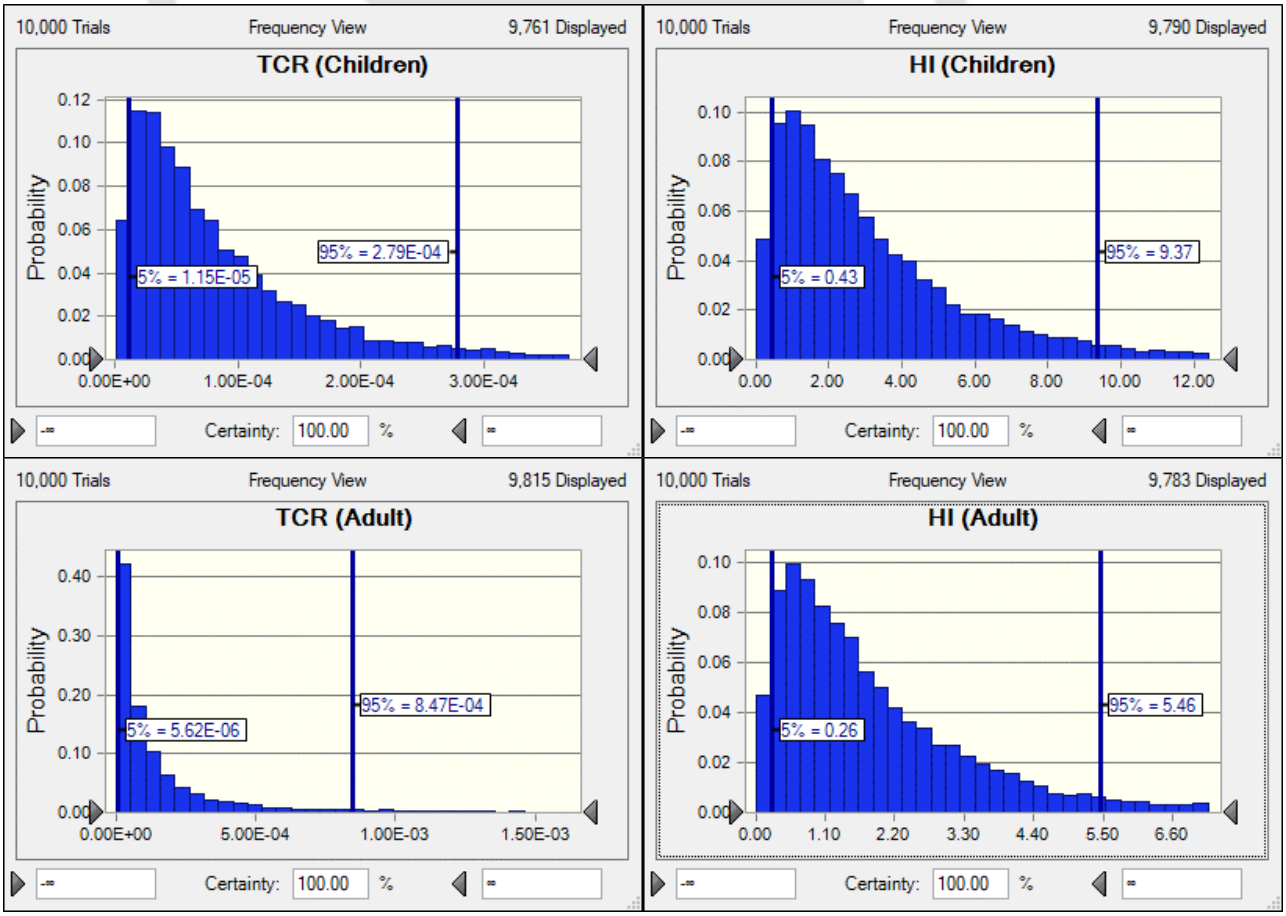
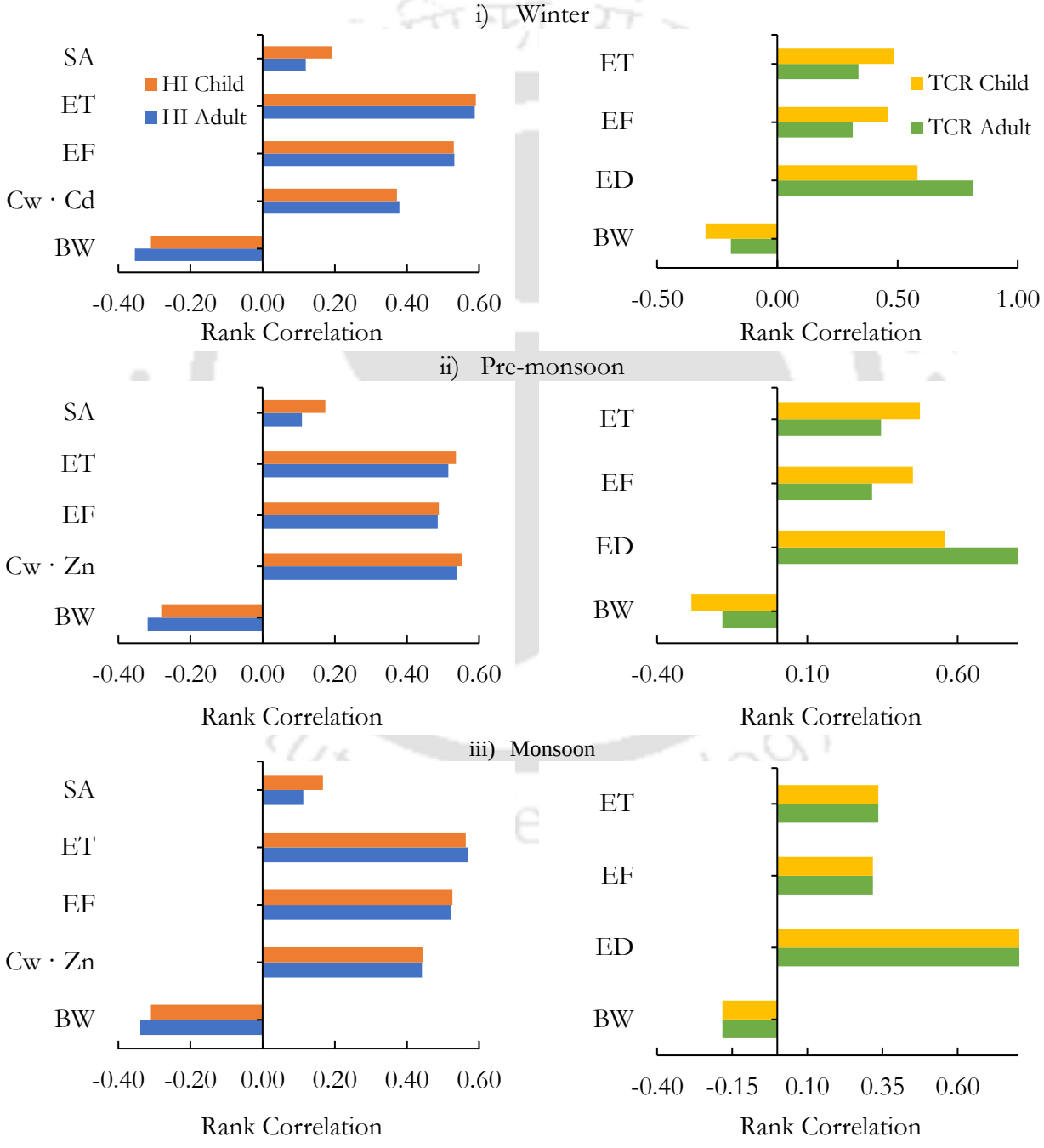


Fig. 7.7. Frequency-probability distribution for post-monsoon for surface water

7.1.2.1 Sensitivity Analysis

Fig. 7.8 shows the results of the sensitivity analysis on the factors responsible for carcinogenic and non-carcinogenic risk for the inhabitants or users of the surface water of the river. Surface area (SA), exposure frequency (EF), exposure time (ET), and body weight (BW) were the parameters irrespective of the season leading to the high non-carcinogenic risk. In winter, the high values were due to Cd (Cw · Cd), and for the rest of the season, the higher risk values were attributed to Zn (Cw · Zn). In the case of carcinogenic risk, the basic parameters, exposure frequency (EF), exposure time (ET), exposure duration (ED), and body weight (BW), were the factors responsible. The high carcinogenic risk in the post-monsoon can be attributed to Pb (Cw · Pb). From the health risk assessment of surface water, Cd, Zn, and Pb are heavy metals responsible for causing risks to the inhabitants.



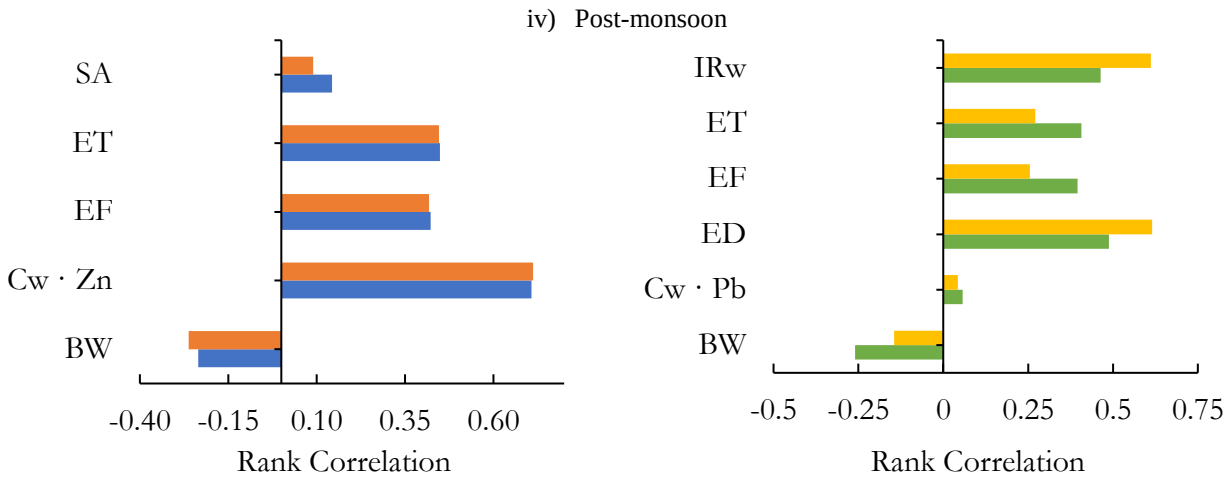
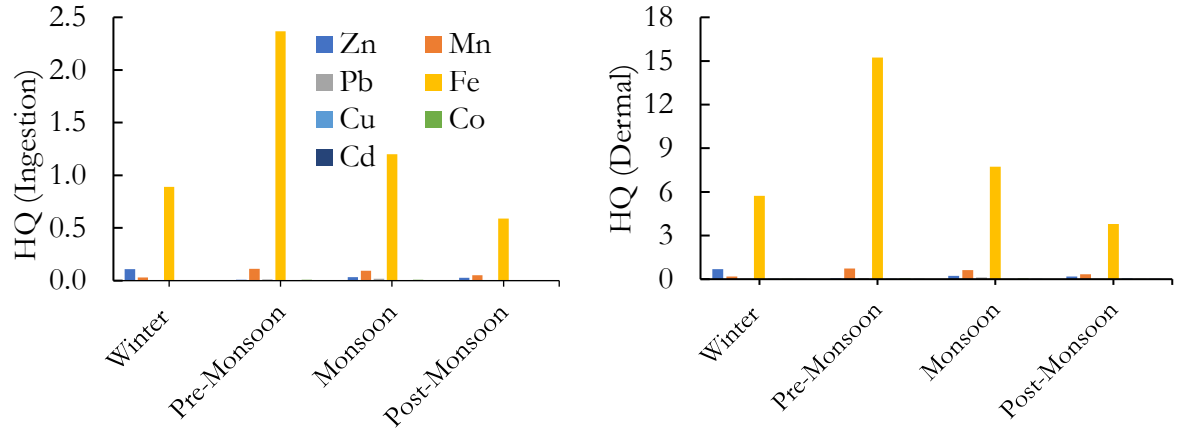


Fig. 7.8. Rank correlation charts of inputs of probability analysis (sensitivity analysis) for surface water

7.2 Benthic Sediments

7.2.1 Deterministic Method

Fig. 7.9 depicts carcinogenic and non-carcinogenic risk assessment findings using the deterministic method for exposure to heavy metals in sediments. During the sampling period, which included several seasons, the results of the deterministic analysis revealed unsatisfactory results, namely, HI>1 and TCR>10^{-5}. Non-carcinogenic risk for adults was highest in the pre-monsoon, with a value of 27.17 compared to a reference value of 1. They tend to decrease with increased precipitation, which is directly correlated with MF, which is also seen to decrease with onset of monsoon. The non-carcinogenic effect of these metals is observed to pose a serious health hazard for persons in the lower age group (i.e. Child), where the HI value varies from 30 to 130 in comparison to persons in the higher age group (i.e. adult) where variation is in between 5 to 20. Fe poses serious health concerns among the metals analysed in the study area, wherein HI value for the rest of the metals is mostly less than 1 (Fig. 7.9). Among the exposure routes, dermal absorption poses more concern for the inhabitants of the nearby area, thereby mainly restricting any recreational usage of the river. Considering the carcinogenic impact of heavy metals on the population of humans, the direct usage of the river should be banned as the TCR values are more than 1x10^{-5} for all seasons.



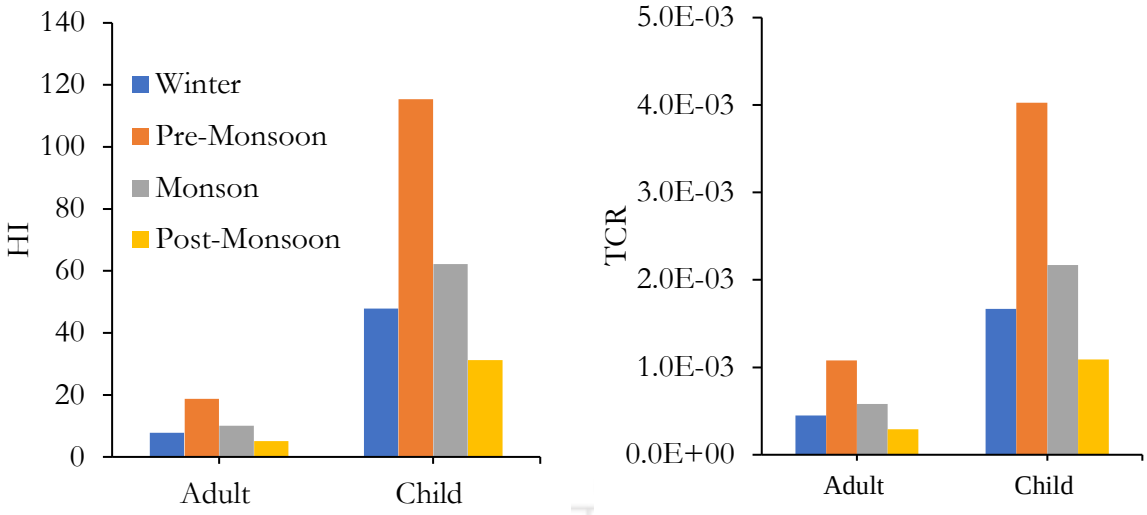
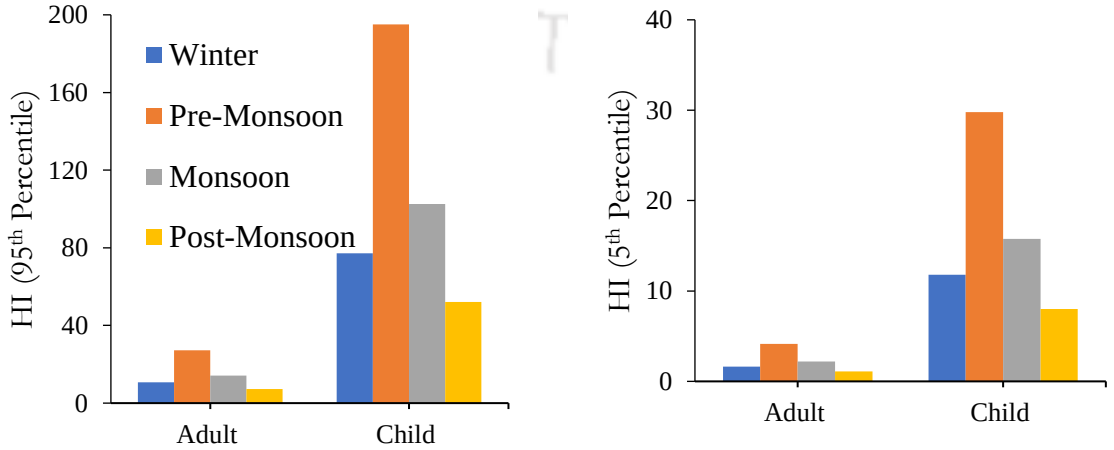


Fig. 7.9. HQ, HI and TCR outcomes of deterministic method for heavy metals in benthic sediments through the different exposure routes

7.2.2 Probabilistic Method

The 5th and 95th percentile values of HI and TCR are given in Fig. 7.10, and their probability-frequency distribution curves for the complete dataset of different seasons are provided in Fig. 7.11-7.14. Non-carcinogenic risk assessment evaluated using HI shows a variation of 1.63-10.69 in winter, 4.15-27.17 in pre-monsoon, 2.20-14.23 in monsoon, and 1.11-7.21in post-monsoon in the case of adults. Coming to the carcinogenic effect, TCR values also show a high variation. In winter, the variation is 1.70E-05 to 4.48E-04, in pre-monsoon, the variation is 3.01E-05 to 1.14E-03, in monsoon, the variation is 1.59E-05 to 5.96E-04, in post-monsoon, the variation is 1.16E-04 to 3.04E-04. Results show that neither in wettest nor driest months do the HI and TCR values fall below permissible values of 1 and 10^{-5}, respectively. The situation is even more harsher for the child population, where the values are much higher. The HI values are in order $10^{2,}$ and the TCR values are in order 10^{-3}, which is way higher than the permissible limits. Health risk assessment study thus calls for the treatment and stringent policies to be framed for pollution abatement.



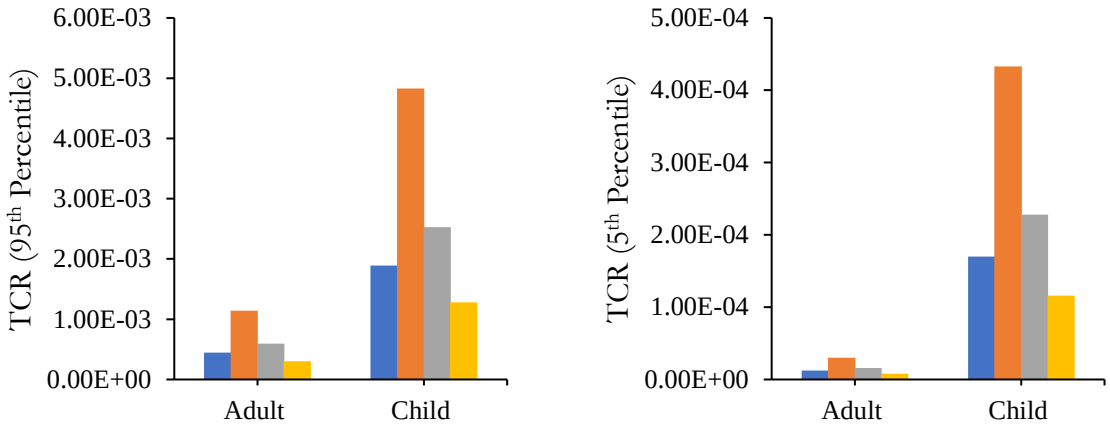


Fig. 7.10. HI and TCR from exposure to heavy metals in benthic sediments through probabilistic method (5th and 95th percentile)

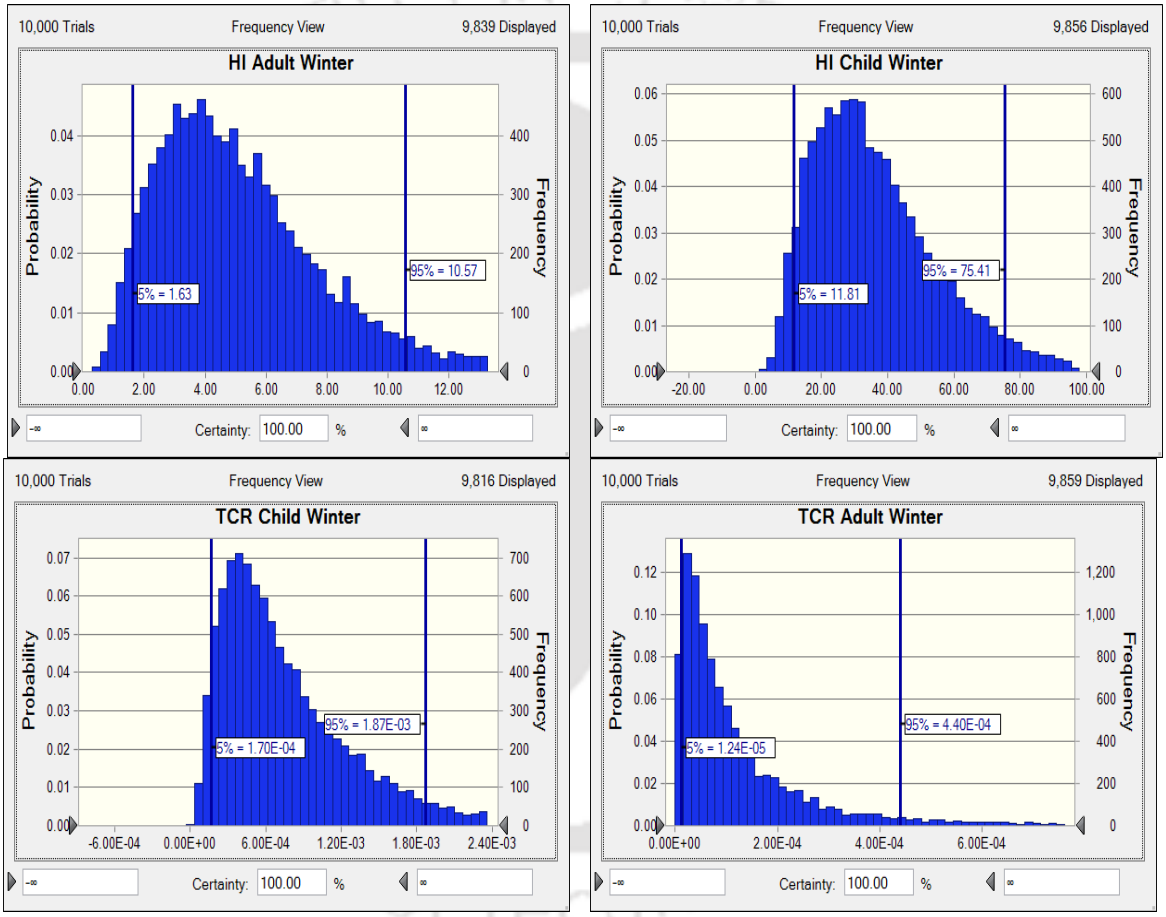
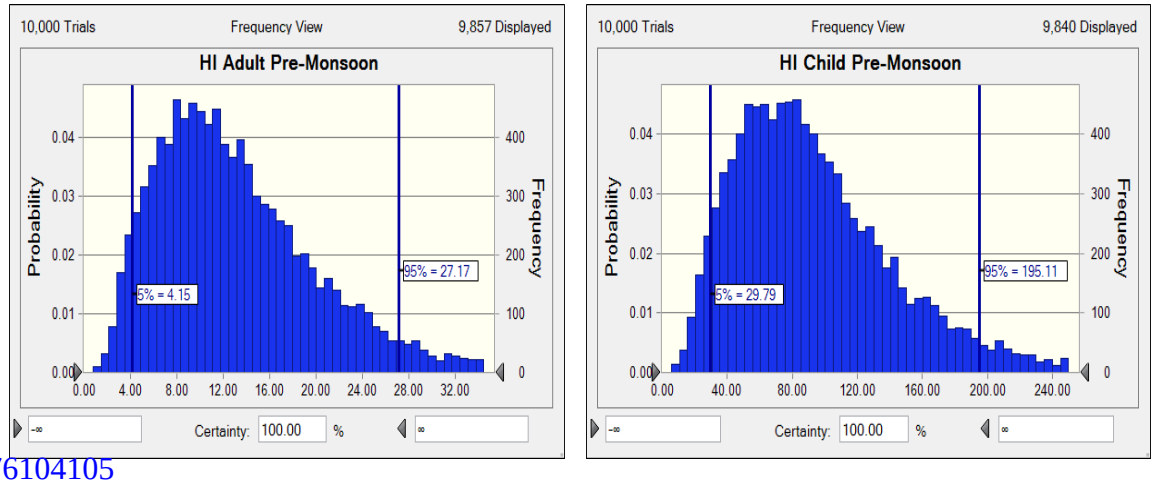


Fig. 7.11. Frequency-probability distribution chart for winter season (benthic sediments)



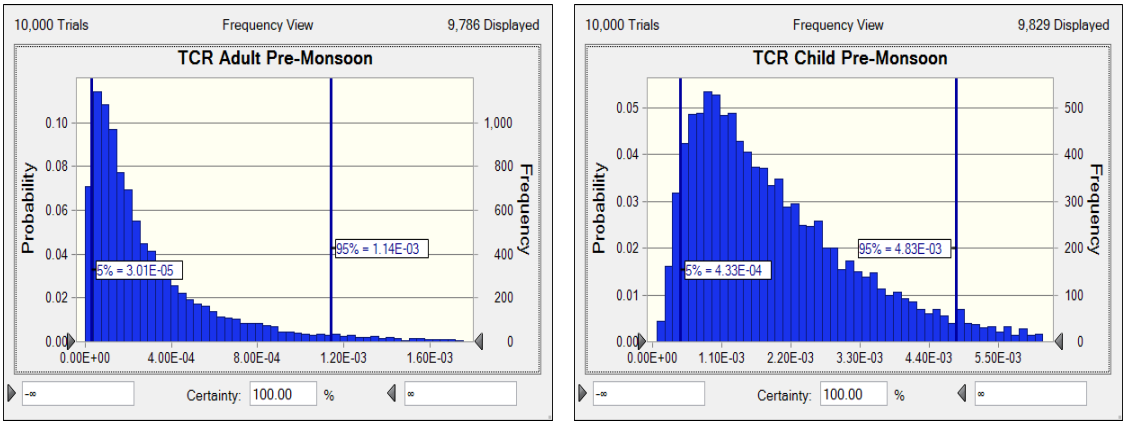


Fig. 7.12. Frequency-probability distribution chart for pre-monsoon (benthic sediments)

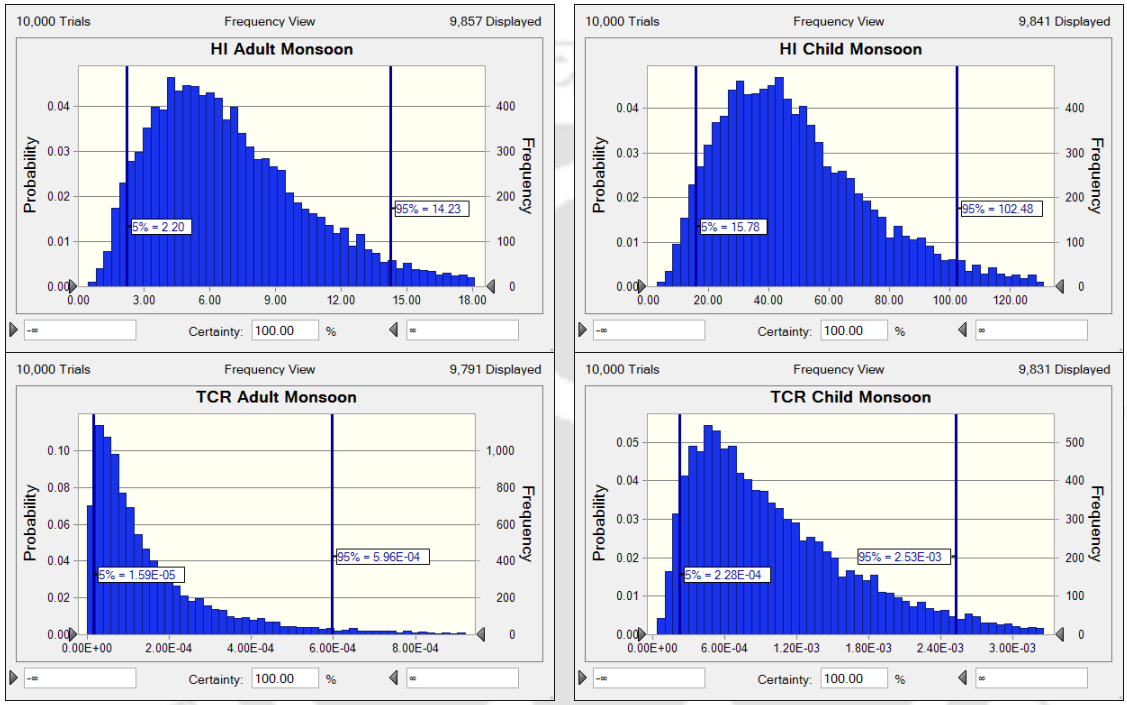


Fig. 7.13. Frequency-probability distribution chart for monsoon (benthic sediments)

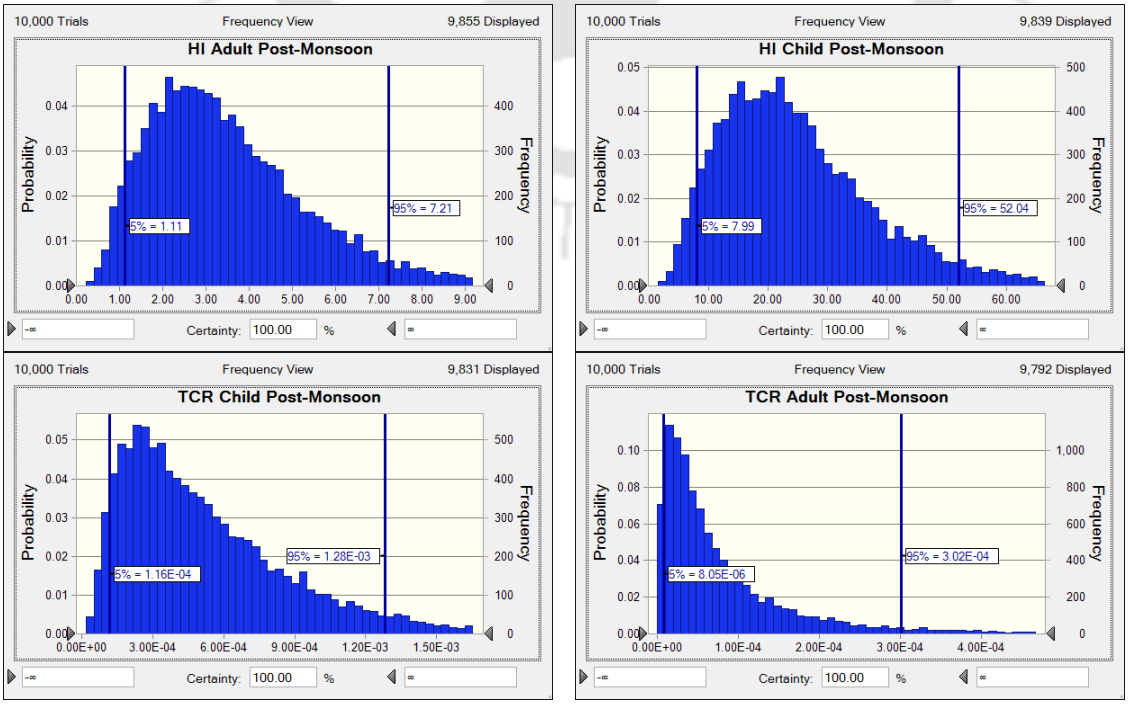
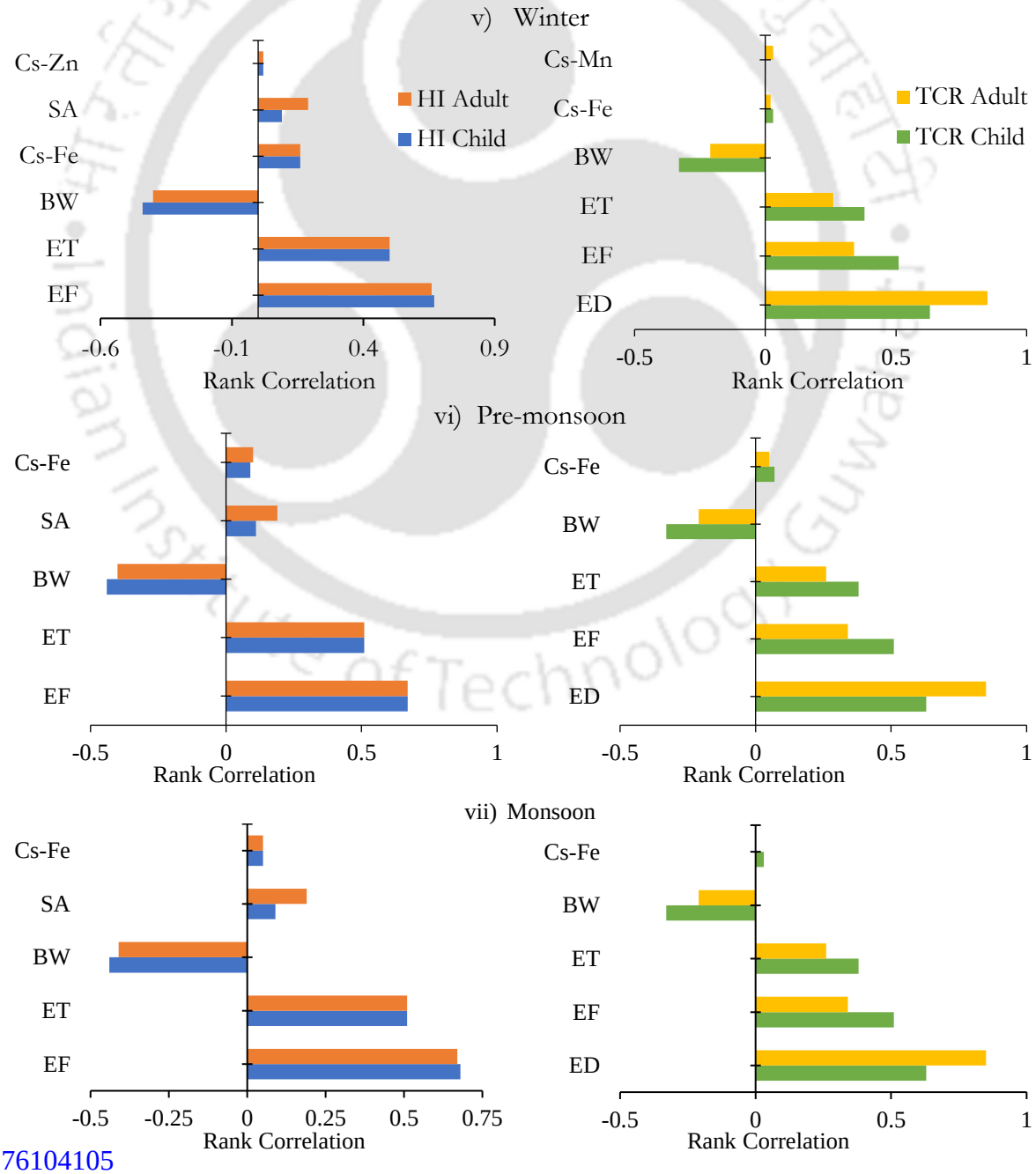


Fig. 7.14. Frequency-probability distribution chart for post-monsoon (benthic sediments)

7.2.2.1 Sensitivity Analysis

Fig. 7.15 shows the sensitivity analysis results to assess the prime parameters concerning the non-carcinogenic and carcinogenic risks for benthic sediments. Exposure frequency (EF), exposure time (ET), body weight (BW) of the population exposed to the risk, contact surface area (SA) in case of dermal exposure, and Fe concentration (Cs-Fe) are important risk factors for non-cancer risk. The contributing or differentiating factor causing the high risk in the winter months is the concentration of Zn (Cs-Zn). In the case of carcinogenic risk, the significant factors contributing to risk are the body weight (BW) of the population exposed to the risk, the concentration of Fe (Cs-Fe), exposure duration (ED), and exposure frequency (EF), and exposure time (ET). The contributing or differentiating factor causing the high risk in the winter months is the concentration of Mn (Cs-Mn). Thus, from sensitivity analysis, it can be concluded that the focus of remediation study in the region should primarily focus on the concentration of Fe, Mn and Zn. Then if required, the focus can also be shifted towards the rest of the heavy metals.



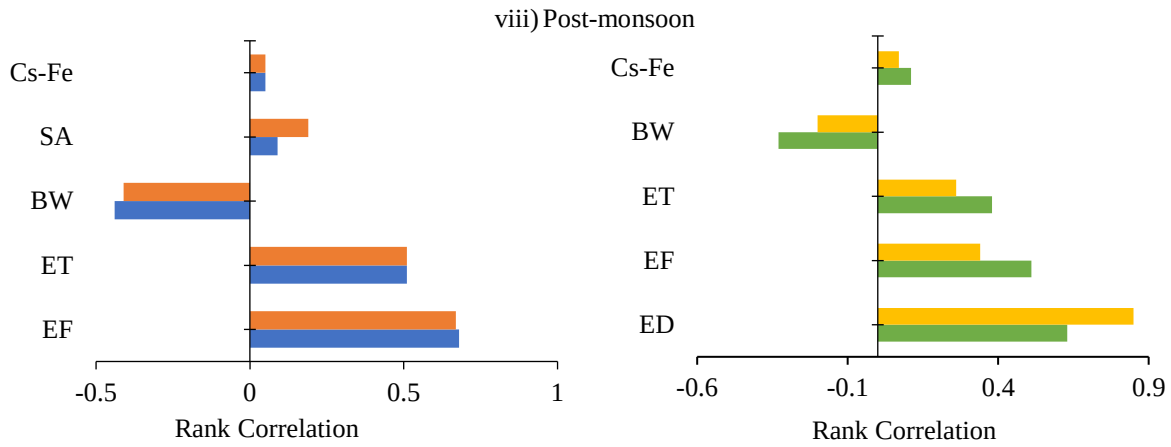


Fig. 7.15. Rank correlation charts of inputs of probability analysis (sensitivity analysis) for benthic sediments

7.3 Summary

The current study used environmental monitoring to assess the health risk of heavy metals in the Kolong River ecosystem, considering two essential components of a river ecosystem: surface water and benthic sediment. Deterministic and probabilistic techniques were used to calculate the non-carcinogenic and carcinogenic risks of heavy metals. Several critical observations were made, and the investigation yielded the following results, which are stated below:

i) The dermal pathway of exposure presents a considerable risk for the human population than the ingestion pathway of exposure in the case of both water column (surface water) and benthic sediments.

ii) The non-carcinogenic health risk assessment revealed Zn and Cd to pose a considerable hazard for the human population, and the carcinogenic health risk assessment revealed Pb to pose a considerable hazard for the human population in the case of surface water.

iii) The non-carcinogenic health risk assessment revealed Zn to pose a considerable hazard to the human population, and the carcinogenic health risk assessment revealed Mn to pose a considerable hazard to the human population in the case of benthic sediments.

Chapter 8

REGRESSION MODELLING

In this chapter, two regression models were generated to predict heavy metal concentration in the water column using multiple linear regression (MLR), artificial neural networks (ANN) and random forest (RF). In the first regression model, heavy metals in the water column were predicted using the physicochemical parameters, as metals in the water column vary greatly with the changes in the physicochemical parameters in the water column. The second regression model predicted heavy metals in the water column from the metal speciation fractions in the benthic sediments as the transfer of heavy metals from water to the sediment column or vice versa occurs continuously in a river under changing environmental conditions.

8.1 First Regression model

In this model, 16 physiochemical parameters such as electrical conductivity (EC), pH, turbidity (TURB), dissolved oxygen (DO), 5-day biochemical oxygen demand (BOD_5), hardness (TH), alkalinity (TA), total dissolved solids (TDS), sodium (Na^+), potassium (K^+), calcium (Ca^{+2}), magnesium (Mg^{+2}), fluoride (F^-), chloride (Cl^-), sulphate (SO_4^{-2}), and nitrate (NO_3^-) were utilised to evaluate heavy metal (Cd, Fe, Mn, Pb, and Zn) concentrations in the surface water Kolong river using MLR, ANN, and RF. In any river system, heavy metals are present along with the physicochemical parameters. Heavy metal concentrations in a river vary with changes in the physicochemical parameters. Keeping this condition in mind, this model is developed so that the parameters affecting the concentration of heavy metals in water can be found mathematically, and pollution abatement strategies can be formulated. Also, the measurement of heavy metals requires high-end instruments like AAS and ICPMS, which are very expensive. The availability of such instruments at every research institute or location is improbable due to their high price or lack of funding. This model is also developed to formulate a heavy metal prediction model so that heavy metals can be measured in any surface water samples with the help of physicochemical parameters, which can be measured relatively easily in any lab. To develop this model, the dataset utilised is shown in Table 8.1. In this table, mean, standard error (SE), standard deviation (Std Dev), minimum (Min), 1st quartile value (Q1), median, 3rd quartile (Q3) and maximum (Max) are shown to understand the variation in the dataset. The dataset includes the monthly sampling values of water samples collected and analysed from May ’2018 to October ’2019. In this period, sampling was done alm-

ost every alternate month. For the regression analysis, 75% of this dataset is used as training data and the remainder as testing data.

Table 8.1. Dataset utilised to develop the first regression model

Variables	N	Mean	Std Dev	Min	Q1	Median	Q3	Max
EC	213	205.27	67.96	88.00	158.15	193.50	244.75	509.00
pH	213	7.06	0.29	6.48	6.80	7.07	7.31	7.71
BOD5	213	22.15	16.24	0.60	10.50	19.20	27.30	117.00
DO	213	6.72	1.23	1.87	6.13	6.90	7.36	9.28
TDS	213	221.29	123.63	30.00	130.00	200.00	276.00	600.00
TURB	213	21.35	13.16	1.40	11.68	17.09	28.27	72.85
TH	213	70.08	46.87	4.54	31.35	64.57	100.67	236.83
TA	213	109.07	82.59	34.00	66.00	89.25	115.96	419.33
Na^+	213	14.10	7.05	4.56	7.90	12.65	20.56	32.08
K^+	213	3.81	4.02	0.95	2.19	3.60	4.43	56.85
Ca^{+2}	213	20.32	16.93	0.65	5.18	18.11	28.95	72.55
Mg^{+2}	213	9.33	7.40	0.34	4.40	8.02	11.80	36.98
F^-	213	0.25	0.20	0.00	0.14	0.21	0.30	1.17
Cl^-	213	2.31	4.99	0.19	0.50	0.92	2.94	66.07
SO_4^{2-}	213	2.55	5.56	0.04	0.35	0.58	2.28	38.86
NO_3^-	213	2.24	3.56	0.14	0.75	1.22	2.22	32.72
Fe	213	1.92	0.93	0.03	1.21	1.91	2.52	5.28
Mn	213	0.18	0.14	0.00	0.03	0.19	0.26	0.77
Pb	213	0.04	0.04	0.00	0.00	0.03	0.06	0.17
Cd	213	0.07	0.14	0.00	0.00	0.00	0.05	0.75
Zn	213	0.28	0.30	0.00	0.06	0.18	0.37	1.60

8.1.1 Multiple Linear Regression (MLR)

Multiple linear regression modelling was performed in two steps. First, all sixteen independent variables were entered into the model to check which parameters were statistically affecting the heavy metal. After observing the effect of all the parameters, a stepwise regression was carried out to determine the significant variables selected by the models. Stepwise regression results, i.e., the coefficient of determination (R^2) value, will always be less than the overall results as in MLR as R^2 increases with an increase in the number of variables. The increase may or may not be significant. However, the contribution of such variables in any model is misleading or not constant. So, such variables should not be included in any linear regression models. The ordinary least squares method is used to develop model parameter regression equations, and the assumptions for developing the equations are mentioned in section 2.5.1. MLR estimation of all the heavy metals considered in this study are as follows:

(a) Cadmium (Cd)

All variables

$$Cd = 0.761 + 0.000046\ EC - 0.1439\ pH + 0.001169\ BOD_5 + 0.02588\ DO + 0.000238\ TDS - 0.001683\ TURB - 0.000387\ TH + 0.000322\ TA - 0.00348\ Na^{+} - 0.00438\ K^{+} + 0.00074\ Ca^{+2} + 0.01199\ Mg^{+2} + 0.1594\ F^{-} - 0.00190\ Cl^{-} - 0.00259\ SO_4^{2-} - 0.00088\ NO_3^{-}$$

Stepwise regression

$$Cd = 0.574 - 0.1208\ pH + 0.03118\ DO + 0.000265\ TDS - 0.001526\ TURB - 0.01596\ K^{+} + 0.013147\ Mg^{+2} + 0.1670\ F^{-}$$

All variables included model for Cd have a training R^2 of 0.64 and a testing R^2 of 0.60. But, many of the variables or parameters were statistically not significant, i.e., $p>0.05$. While from the stepwise regression, only 7 parameters were required to estimate the Cd concentration, and all 7 parameters were statistically significant ($p<0.05$). This stepwise regression model for Cd has a training R^2 of 0.69 and a testing R^2 of 0.40. The coded coefficients for the stepwise regression model of Cd are shown in Table 8.2.

Table 8.2. Coded coefficients for stepwise regression model of Cd

Term	Coef	95% CI	T-Value	P-Value	VIF
Constant	0.061	(0.04950, 0.07245)	10.490	0.000	
pH	-0.035	(-0.04701, -0.02250)	-5.600	0.000	1.210
DO	0.038	(0.02542, 0.05144)	5.830	0.000	1.250
TDS	0.033	(0.02015, 0.04530)	5.140	0.000	1.280
TURB	-0.020	(-0.03298, -0.00720)	-3.080	0.002	1.380
K^{+}	-0.064	(-0.0960, -0.0323)	-3.980	0.000	1.410
Mg^{+2}	0.097	(0.08422, 0.11026)	14.750	0.000	1.280
F^{-}	0.033	(0.01909, 0.04631)	4.740	0.000	1.590

(b) Iron (Fe)

All Variables

$$Fe = 0.06 + 0.001447\ EC - 0.057\ pH + 0.00458\ BOD_5 + 0.1122\ DO - 0.000973\ TDS + 0.03468\ TURB + 0.01022\ TH - 0.00015\ TA + 0.00248\ Na^{+} + 0.03692\ K^{+} - 0.00551\ Ca^{+2} + 0.01547\ Mg^{+2} - 1.634\ F^{-} + 0.03516\ Cl^{-} + 0.0430\ SO_4^{2-} - 0.0393\ NO_3^{-}$$

Stepwise regression

$$Fe = -2.073 + 0.393\ pH - 0.001493\ TDS + 0.03314\ TURB + 0.007809\ TH + 0.002615\ TA + 0.04492\ K^{+} + 0.01654\ Mg^{+2} - 1.330\ F^{-} + 0.03473\ Cl^{-} - 0.0373\ NO_3^{-}$$

All variables included model for Fe have a training R^2 of 0.75 and a testing R^2 of 0.70. But, many of the variables or parameters were statistically not significant, i.e., $p>0.05$. While from the stepwise regression, only 10 parameters were required to estimate the Fe concentration, and all 10 parameters were statistically significant ($p<0.05$). This stepwise regression model for Fe has a

training R^2 of 0.73 and a testing R^2 of 0.67. The coded coefficients from the stepwise regression model for Fe are shown in Table 8.3.

Table 8.3. Coded coefficients for stepwise regression model of Fe

Term	Coef	95% CI	T-Value	P-Value	VIF
Constant	1.902	(1.8301, 1.9734)	52.390	0.000	
pH	0.113	(0.0311, 0.1951)	2.720	0.007	1.300
TDS	-0.185	(-0.2803, -0.0888)	-3.810	0.000	1.570
TURB	0.436	(0.3568, 0.5154)	10.860	0.000	1.240
TH	0.366	(0.2778, 0.4543)	8.190	0.000	1.500
TA	0.216	(0.1399, 0.2921)	5.610	0.000	1.070
K+	0.181	(0.1117, 0.2497)	5.170	0.000	1.130
Mg^{+2}	0.122	(0.0375, 0.2072)	2.850	0.005	1.390
F-	-0.260	(-0.3466, -0.1739)	-5.950	0.000	1.510
Cl^-	0.173	(0.1054, 0.2413)	5.040	0.000	1.100
NO_3^-	-0.133	(-0.2061, -0.0594)	-3.570	0.000	1.250

(c) Manganese (Mn)

All variables

$$Mn = 0.083 - 0.000747\ EC + 0.0655\ pH + 0.001769\ BOD_5 - 0.01177\ DO - 0.000081\ TDS - 0.000326\ TURB - 0.000295\ TH - 0.000072\ TA - 0.01058\ Na^+ + 0.01331\ K^+ + 0.002571\ Ca^{+2} - 0.00258\ Mg^{+2} - 0.1396\ F^- + 0.003402\ Cl^- - 0.00076\ SO_4^{2-} - 0.00913\ NO_3^-$$

Stepwise regression

$$Mn = 0.5335 - 0.000637\ EC + 0.002233\ BOD_5 - 0.00995\ DO - 0.000139\ TDS - 0.000217\ TA - 0.010156\ Na^+ + 0.00744\ K^+ - 0.0885\ F^- + 0.002961\ Cl^- - 0.00996\ NO_3^-$$

All variables included model for Mn have a training R^2 of 0.74 and a testing R^2 of 0.72. But, many of the variables or parameters were statistically not significant, i.e., $p>0.05$. While from the stepwise regression, only 10 parameters were required to estimate the Mn concentration, and all 10 parameters were statistically significant ($p<0.05$). This stepwise regression model for Mn had a training R^2 of 0.78 and a testing R^2 of 0.63. The coded coefficients from the stepwise regression model for Mn are shown in Table 8.4.

(d) Lead (Pb)

All Variables

$$Pb = -0.1149 + 0.000043\ EC + 0.01857\ pH + 0.000133\ BOD_5 - 0.00183\ DO + 0.000006\ TDS + 0.000964\ TURB - 0.000638\ TH - 0.000286\ TA + 0.002669\ Na^+ - 0.000676\ K^+ + 0.002430\ Ca^{+2} - 0.001419\ Mg^{+2} - 0.0300\ F^- + 0.000229\ Cl^- + 0.002202\ SO_4^{2-} + 0.001632\ NO_3^-$$

Stepwise regression

Pb = -0.1351 + 0.02097 pH + 0.000416 BOD5 + 0.000894 TURB - 0.000754 TH
- 0.000262 TA + 0.002832 Na^{+} - 0.001076 K^{+} + 0.002663 Ca^{+2} - 0.001371 Mg^{+2}
- 0.0297 F^{-} + 0.001785 SO_4^{2-} + 0.001707 NO_3^{-}

All variables included model for Pb have a training R^2 of 0.67 and a testing R^2 of 0.65. But, many of the variables or parameters were statistically not significant, i.e., $p>0.05$. While from the stepwise regression, only 12 parameters were required to estimate the Pb concentration, and all 12 parameters were statistically significant ($p<0.05$). This stepwise regression model for Pb has a training R^2 of 0.69 and a testing R^2 of 0.59. The coded coefficients from the stepwise regression model for Pb are shown in Table 8.5.

Table 8.4. Coded coefficients for stepwise regression model of Mn

Term	Coef	95% CI	T-Value	P-Value	VIF
Constant	0.179	(0.16875, 0.18839)	35.910	0.000	
EC	-0.043	(-0.05375, -0.03278)	-8.150	0.000	1.200
BOD_5	0.036	(0.02476, 0.04778)	6.220	0.000	1.380
DO	-0.012	(-0.02429, -0.00024)	-2.010	0.046	1.380
TDS	-0.017	(-0.02873, -0.00567)	-2.950	0.004	1.410
TA	-0.018	(-0.03078, -0.00509)	-2.760	0.007	1.640
Na^{+}	-0.072	(-0.08472, -0.05844)	-10.760	0.000	1.870
K^{+}	0.030	(0.01951, 0.04032)	5.680	0.000	1.380
F^{-}	-0.017	(-0.03107, -0.00356)	-2.490	0.014	1.470
Cl^{-}	0.015	(0.00553, 0.02402)	3.160	0.002	1.080
NO_3^{-}	-0.035	(-0.04538, -0.02556)	-7.070	0.000	1.230

Table 8.5. Coded coefficients for stepwise regression model of Pb

Term	Coef	95% CI	T-Value	P-Value	VIF
Constant	0.038	(0.03466, 0.04097)	23.690	0.000	
pH	0.006	(0.00255, 0.00952)	3.420	0.001	1.300
BOD_5	0.007	(0.00271, 0.01081)	3.290	0.001	1.790
TURB	0.012	(0.00815, 0.01539)	6.420	0.000	1.320
TH	-0.035	(-0.04453, -0.02613)	-7.580	0.000	9.180
TA	-0.022	(-0.03014, -0.01319)	-5.050	0.000	6.360
Na^{+}	0.020	(0.01549, 0.02443)	8.820	0.000	1.960
K^{+}	-0.004	(-0.00773, -0.00093)	-2.510	0.013	1.410
Ca^{+2}	0.045	(0.03556, 0.05459)	9.360	0.000	9.400
Mg^{+2}	-0.010	(-0.01477, -0.00551)	-4.320	0.000	2.190
F^{-}	-0.006	(-0.00979, -0.00183)	-2.880	0.004	1.750
SO_4^{2-}	0.010	(0.00233, 0.01753)	2.580	0.011	4.790
NO_3^{-}	0.006	(0.00290, 0.00926)	3.770	0.000	1.190

(e) Zinc (Zn)

All variables

$$Zn = -0.682 - 0.000893\ EC + 0.0488\ pH + 0.002391\ BOD_5 + 0.0018\ DO - 0.000326\ TDS + 0.00335\ TURB - 0.000587\ TH + 0.002464\ TA + 0.02508\ Na^{+} + 0.00971\ K^{+} + 0.00662\ Ca^{+2} - 0.00255\ Mg^{+2} + 0.0177\ F^{-} + 0.01009\ Cl^{-} - 0.01930\ SO_4^{2-} + 0.01475\ NO_3^{-}$$

Stepwise regression

$$Zn = -0.3579 - 0.000683\ EC + 0.002461\ BOD_5 - 0.000325\ TDS + 0.00338\ TURB + 0.002445\ TA + 0.02142\ Na^{+} + 0.01027\ K^{+} + 0.004831\ Ca^{+2} + 0.01075\ Cl^{-} - 0.01653\ SO_4^{2-} + 0.01364\ NO_3^{-}$$

All variables included model for Zn have a training R^2 of 0.72 and a testing R^2 of 0.71. But, many of the variables or parameters were statistically not significant, i.e., $p>0.05$. While from the stepwise regression, only 11 parameters were required to estimate the Pb concentration, and all 11 parameters were statistically significant ($p<0.05$). This stepwise regression model for Zn had a training R^2 of 0.71 and a testing R^2 of 0.68. The coded coefficients from the stepwise regression model for Zn are shown in Table 8.6.

Table 8.6. Coded coefficients for stepwise regression model of Zn

Term	Coef	95% CI	T-Value	P-Value	VIF
Constant	0.276	(0.2518, 0.3001)	22.560	0.000	
EC	-0.046	(-0.0762, -0.0167)	-3.080	0.002	1.290
BOD_5	0.040	(0.0109, 0.0691)	2.720	0.007	1.530
TDS	-0.040	(-0.0702, -0.0102)	-2.640	0.009	1.550
TURB	0.044	(0.0165, 0.0723)	3.140	0.002	1.240
TA	0.202	(0.1478, 0.2561)	7.370	0.000	4.850
Na^{+}	0.151	(0.1177, 0.1843)	8.940	0.000	1.830
K^{+}	0.041	(0.0159, 0.0667)	3.210	0.002	1.360
Ca^{+2}	0.082	(0.0514, 0.1122)	5.310	0.000	1.650
Cl^{-}	0.054	(0.0301, 0.0772)	4.510	0.000	1.120
SO_4^{2-}	-0.092	(-0.1459, -0.0379)	-3.360	0.001	4.370
NO_3^{-}	0.049	(0.0242, 0.0729)	3.940	0.000	1.210

8.1.2 Artificial Neural Network (ANN)

The regression learner app of MATLAB is used to estimate heavy metals using ANN. In the regression learner app, different parameters of ANN can be optimised, known as hyperparameter optimisation. In hyper-parameter optimisation, number of layers (1-3), number of neurons in each layer (1-300), activation function (ReLu, Tanh, Sigmoid, None), regularisation strength (Lambda), data standardisation required/not required can all be optimised to get the best results for the neural network. This optimisation process is performed either by Bayesian optimisation or

random search method. The optimised parameters were validated using the k-fold (5-fold) validation technique on the data within the training set. After this, the predicted performance of the training process is examined on the testing data. The typical ANN model for the study will consist of an input layer of 16 neurons, a hidden layer consisting of a maximum of 3 hidden layers, each consisting of 300 hidden neurons (maximum) and an output layer of a single neuron. The results of regression for the estimation of heavy metals are as follows:

(a) Cadmium (Cd)

The best neural network (NN) model for the prediction for estimation of Cd consists of 3 hidden layers, 1st layer of 1 neuron, 2nd layer of 2 neurons, and 3rd layer of again 2 neurons. The optimised hyperparameters are activation function: Tanh, Regularisation strength (Lambda): 1.04×10^{-3}, and standardisation required: Yes, and the iteration limit: 1000. Using this parameter, the regression ANN model is being developed, shown in Fig. 8.1, and the training/testing results are shown in Fig. 8.2. The neural network developed for Cd had a training R^2 of 0.76 and a testing R^2 of 0.72. The relative contribution of the various parameters in the ANN model are shown in Fig. 8.3. From the figure, it is observed that the most influencing parameters on Cd concentration are Mg^{+2}, F^-, DO, turbidity, hardness, Cl^-, Ca^{+2}, and SO_4^{-2}, having a relative contribution of more than 10 in the model.

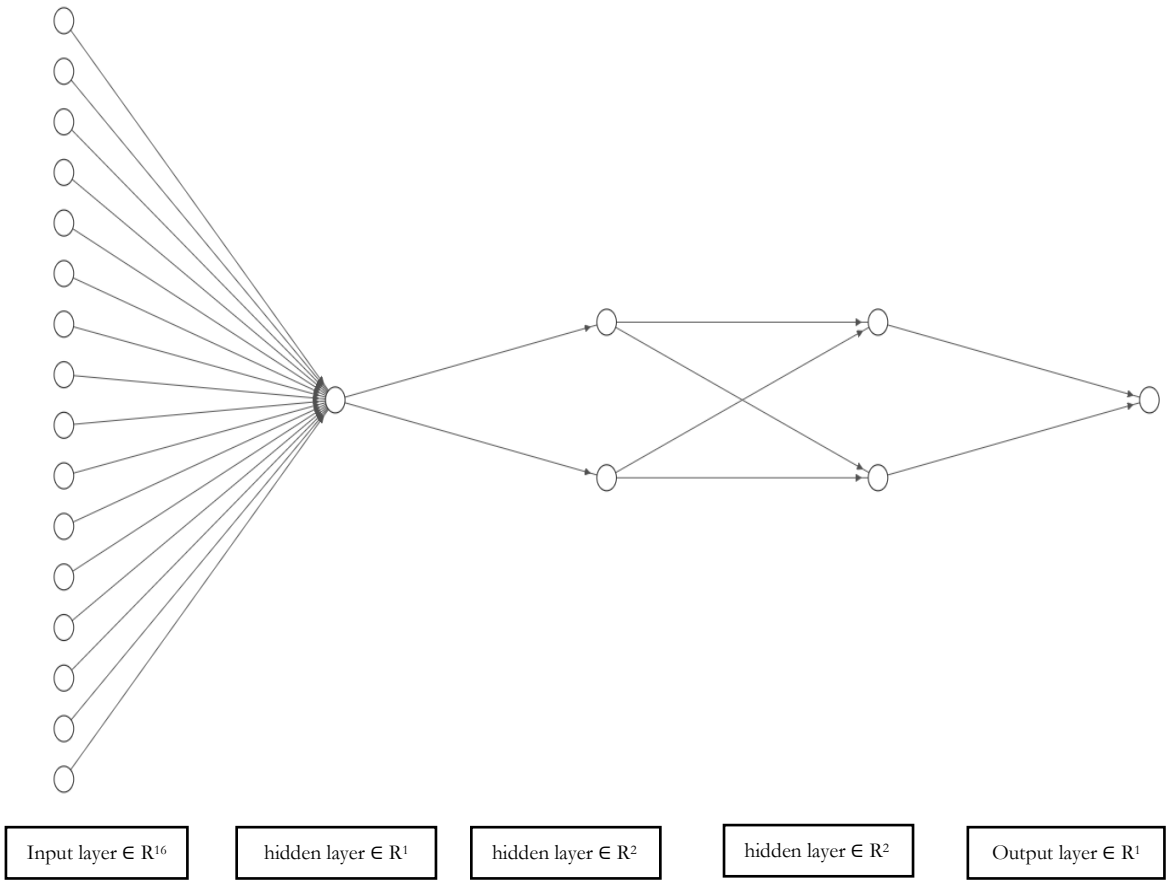


Fig. 8.1. ANN model for the prediction of Cd

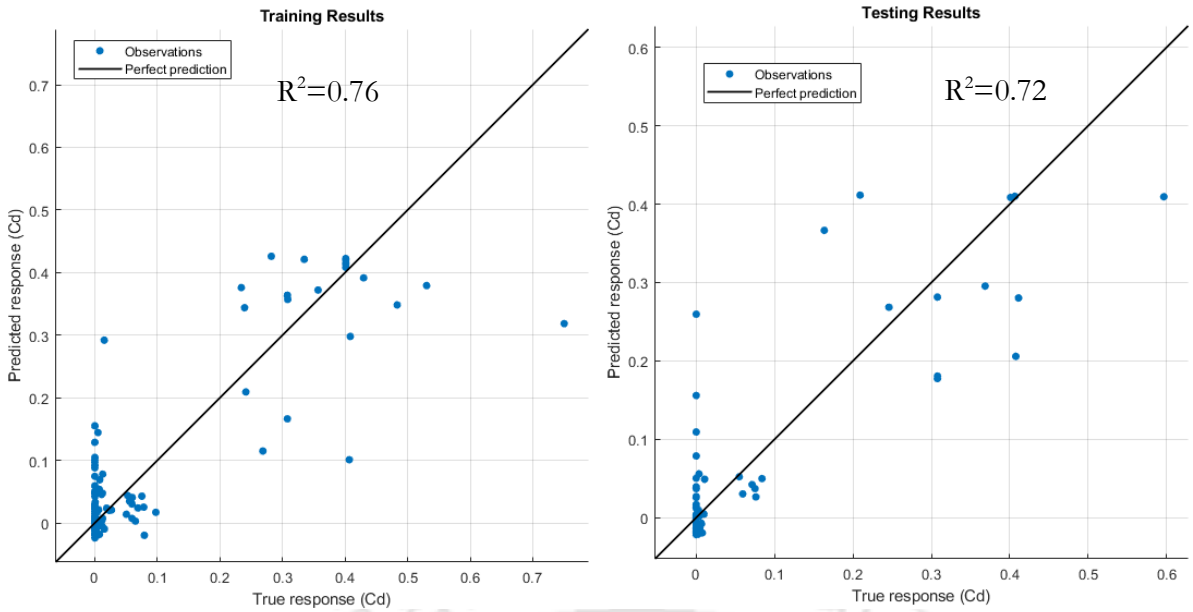


Fig. 8.2. Training/testing R^2 for ANN model for Cd

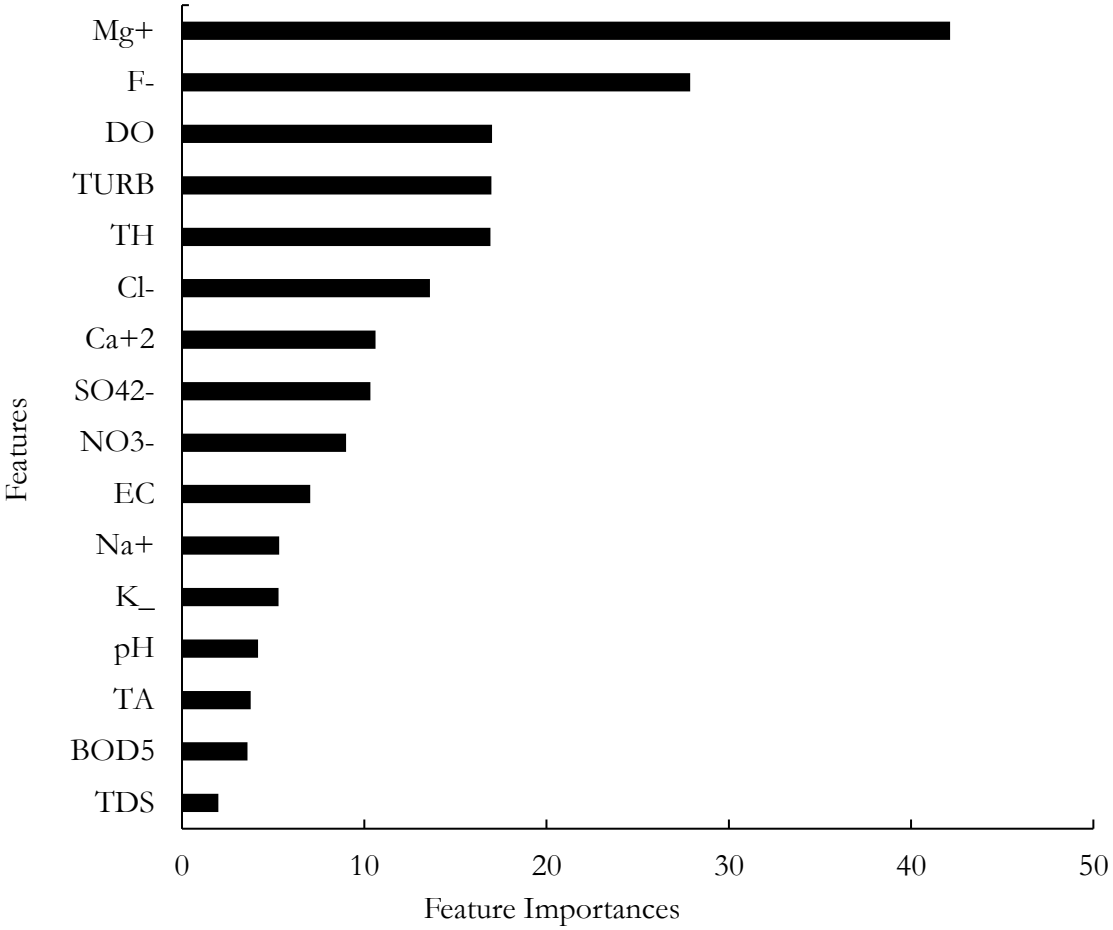


Fig. 8.3. Feature importances of the ANN model for Cd

(b) Iron (Fe)

The best NN model for the prediction for estimation of Fe consists of 2 hidden layers, 1st layer of 4 neurons and 2nd layer of 142 neurons. The optimised hyperparameters are activation function: None, Regularisation strength (Lambda): 9.24×10^{-3}, and standardisation required: Yes, and the iteration limit: 1000. Using this parameter, the regression NN model is being developed, shown in Fig. 8.4, and the training/testing results are shown in Fig. 8.5. The neural network developed

for Fe had a training R^2 of 0.73 and a testing R^2 of 0.72. The relative contribution of the various parameters in the ANN model are shown in Fig. 8.6. The figure shows that the most influencing parameters on the concentration of Fe are Ca^{+2}, F^-, hardness, pH, turbidity, and Na^+, having a relative contribution of more than 10 in the model.

(c) Manganese (Mn)

The best NN model for the prediction for estimation of Mn consists of 2 hidden layers, 1st layer of 16 neurons and 2nd layer of 4 neurons. The optimised hyperparameters are activation function: Sigmoid, Regularisation strength (Lambda): 1.40×10^{-3}, and standardisation required: Yes, and the iteration limit: 1000. Using this parameter, the regression ANN model is being developed, shown in Fig. 8.7, and the training/testing results are shown in Fig. 8.8. The neural network developed for Mn had a training R^2 of 0.75 and a testing R^2 of 0.70. The relative contribution of the various parameters in the ANN model are shown in Fig. 8.9. The figure shows that the most influencing parameters on the concentration of Mn are Na^+, BOD_5, F^-, TDS, pH, and Ca^{+2}, having a relative contribution of more than 10 in the model.

(d) Lead (Pb)

The best NN model for the prediction for estimation of Pb consists of 2 hidden layers, 1st layer of 98 neurons and 2nd layer of 125 neurons. The optimised hyperparameters are activation function: ReLu, Regularisation strength (Lambda): 4.90×10^{-4}, and standardisation required: Yes, and the iteration limit: 1000. Using this parameter, the regression NN model is being developed, shown in Fig. 8.10, and the training/testing results are shown in Fig. 8.11. The neural network developed for Pb had a training R^2 of 0.79 and a testing R^2 of 0.68. The relative contribution of the various parameters in the ANN model are shown in Fig. 8.12. The figure shows that the most influencing parameters on the concentration of Pb are alkalinity, turbidity, F^-, Mg^{+2}, and hardness, having a relative contribution of more than 10 in the model.

(e) Zinc (Zn)

The best NN model for the prediction for estimation of Zn consists of 3 hidden layers, 1st layer of 61 neurons and 2nd layer of 2 neurons, and 3rd layer of 2 layers. The optimised hyperparameters are activation function: None, Regularisation strength (Lambda): 1.33×10^{-2}, and standardisation required: Yes, and the iteration limit: 1000. Using this parameter, the regression NN model is being developed, shown in Fig. 8.13, and the training/testing results are shown in Fig. 8.14. The neural network developed for Zn had a training R^2 of 0.67 and a testing R^2 of 0.61. The relative contribution of the various parameters in the ANN model are shown in Fig. 8.15. The figure shows that the most influencing parameters on the concentration of Pb are alkalinity, Na^+, hardness, DO, SO_4^{-2}, NO_3^-, Mg^{+2}, and F^- , having a relative contribution of more than 10 in the model.

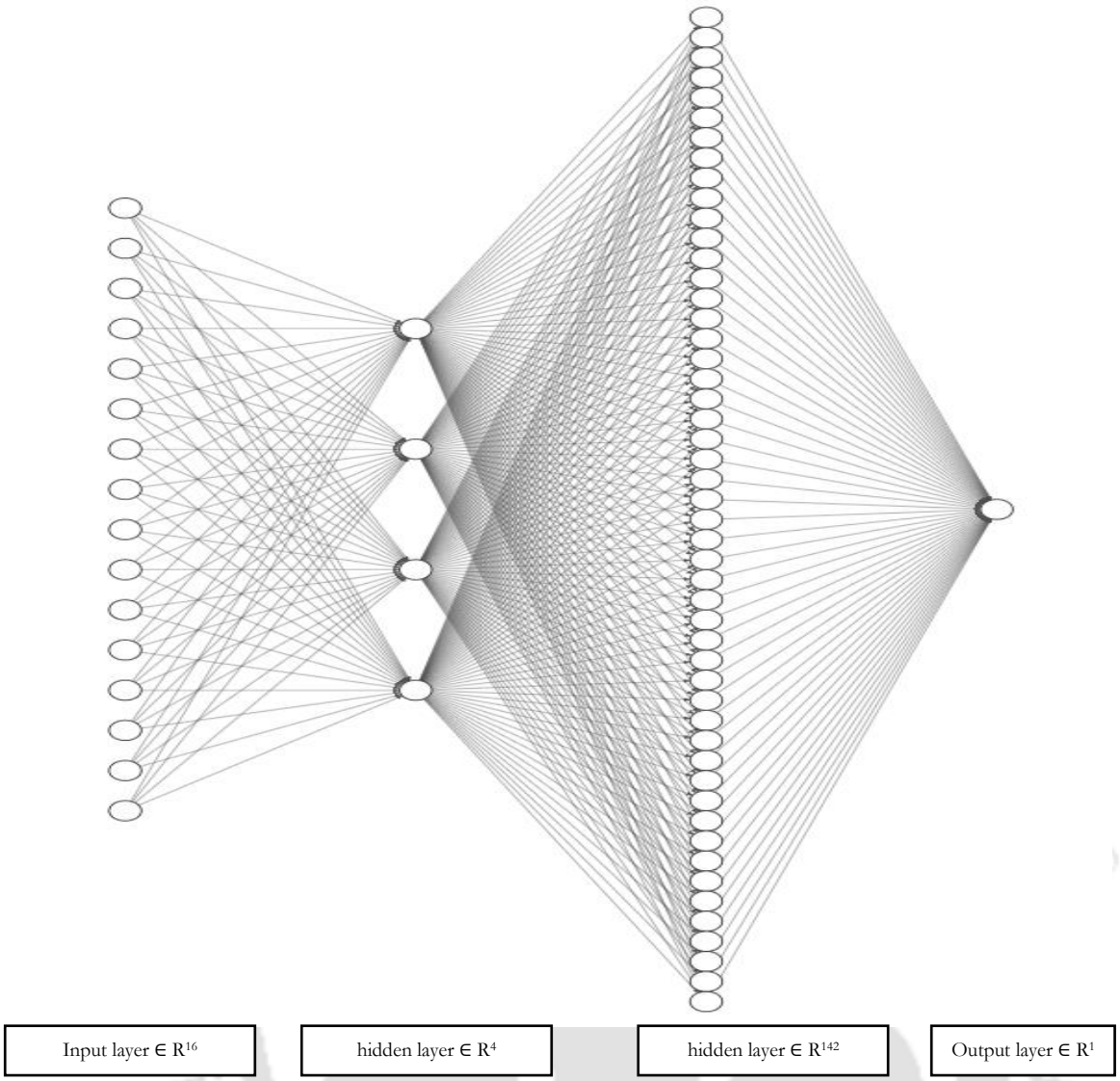


Fig. 8.4. ANN model for the prediction of Fe

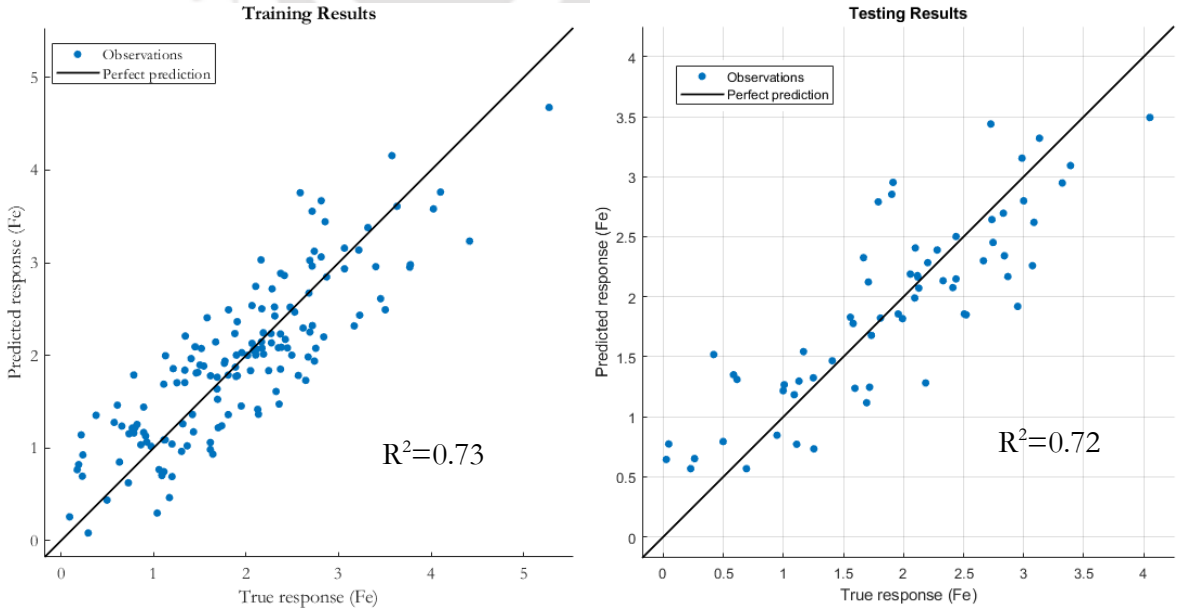


Fig. 8.5. Training/testing R^2 for ANN model for Fe

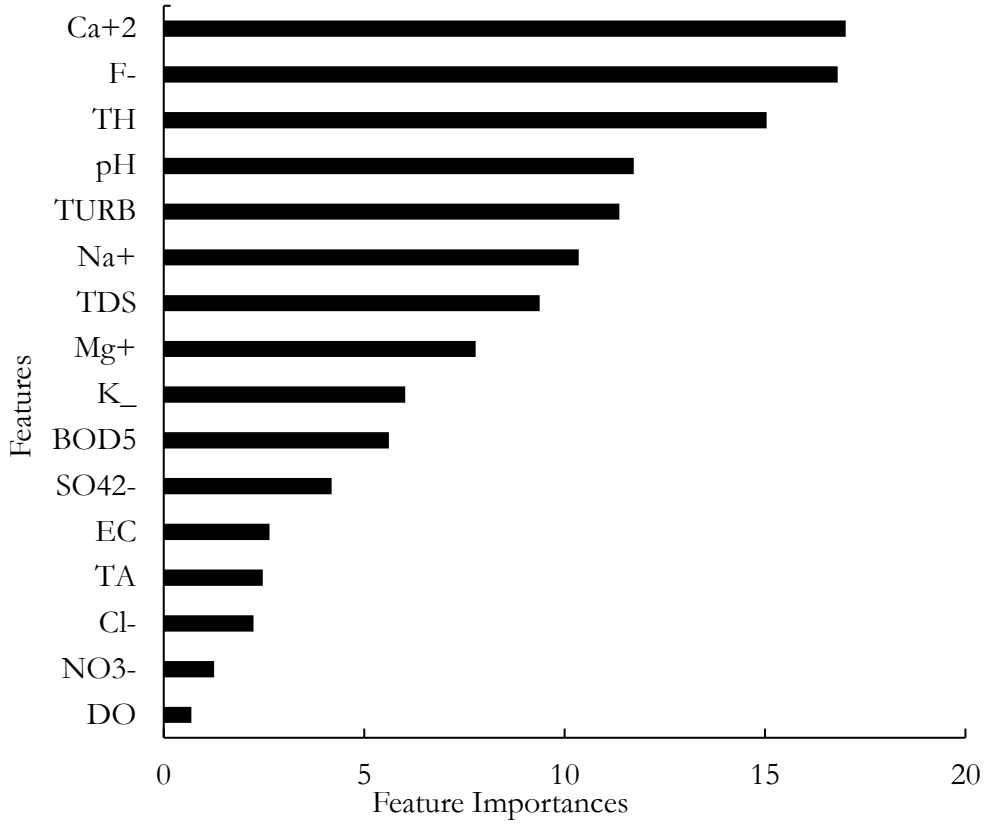


Fig. 8.6. Feature importances of the ANN model for Fe

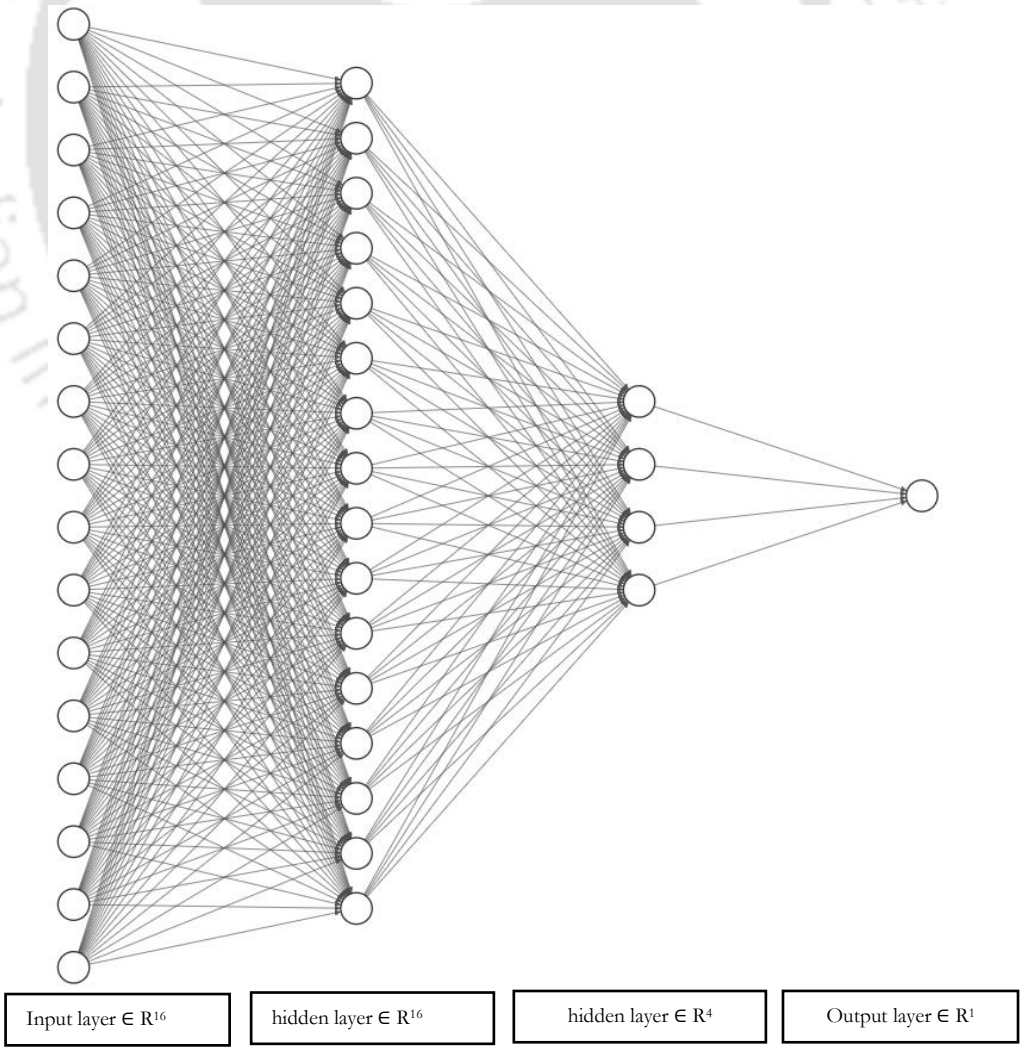


Fig. 8.7. ANN model for the prediction of Mn

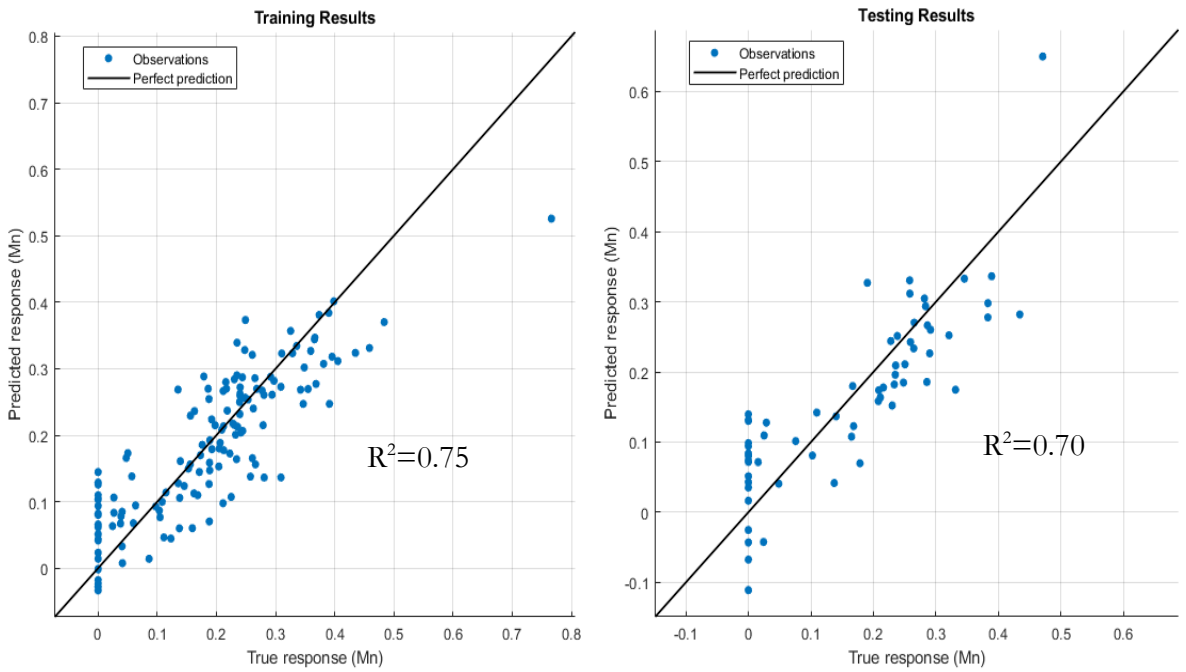


Fig. 8.8. Training/testing R^2 for ANN model for Mn

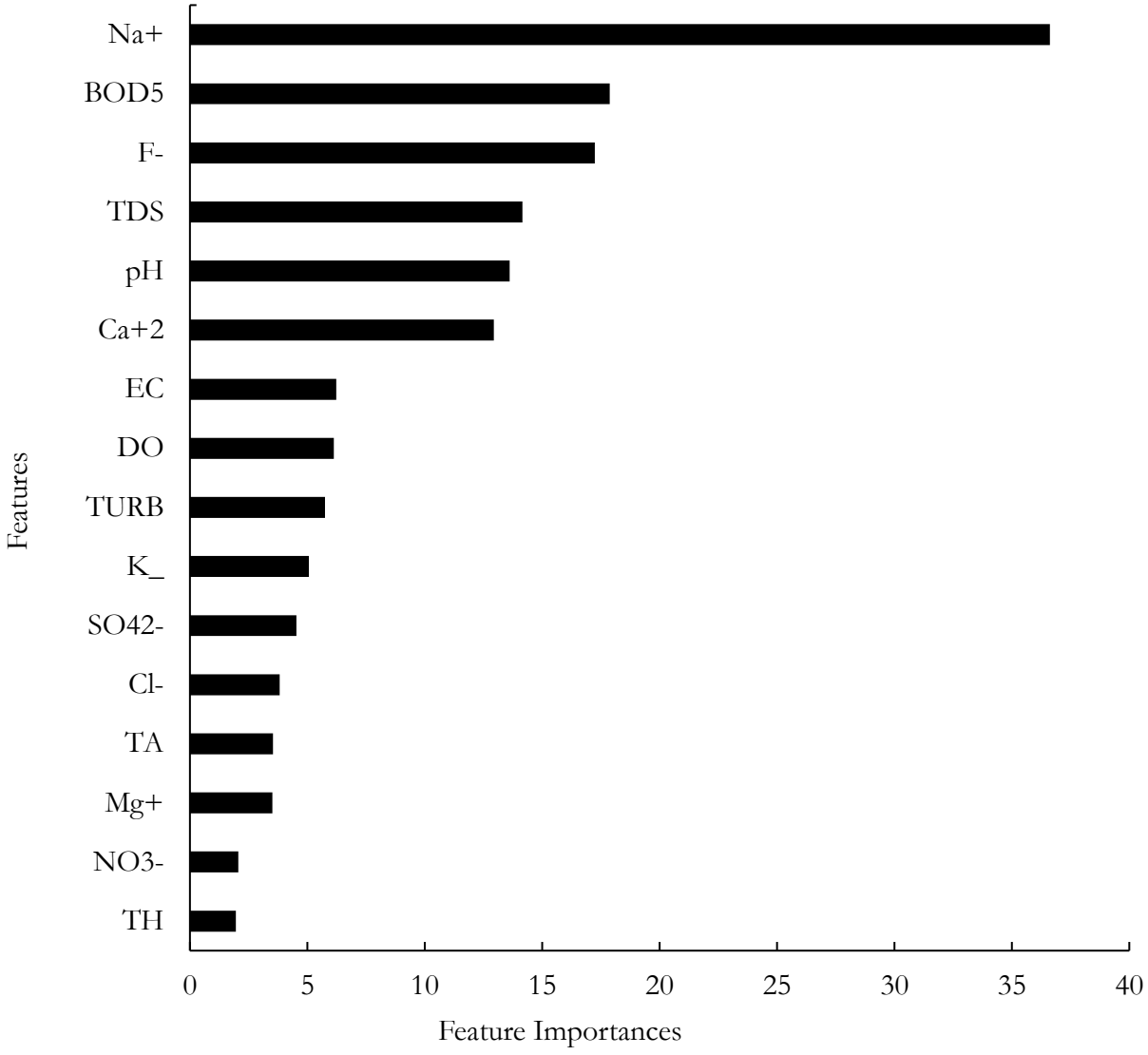


Fig. 8.9. Feature importances of the ANN model for Mn

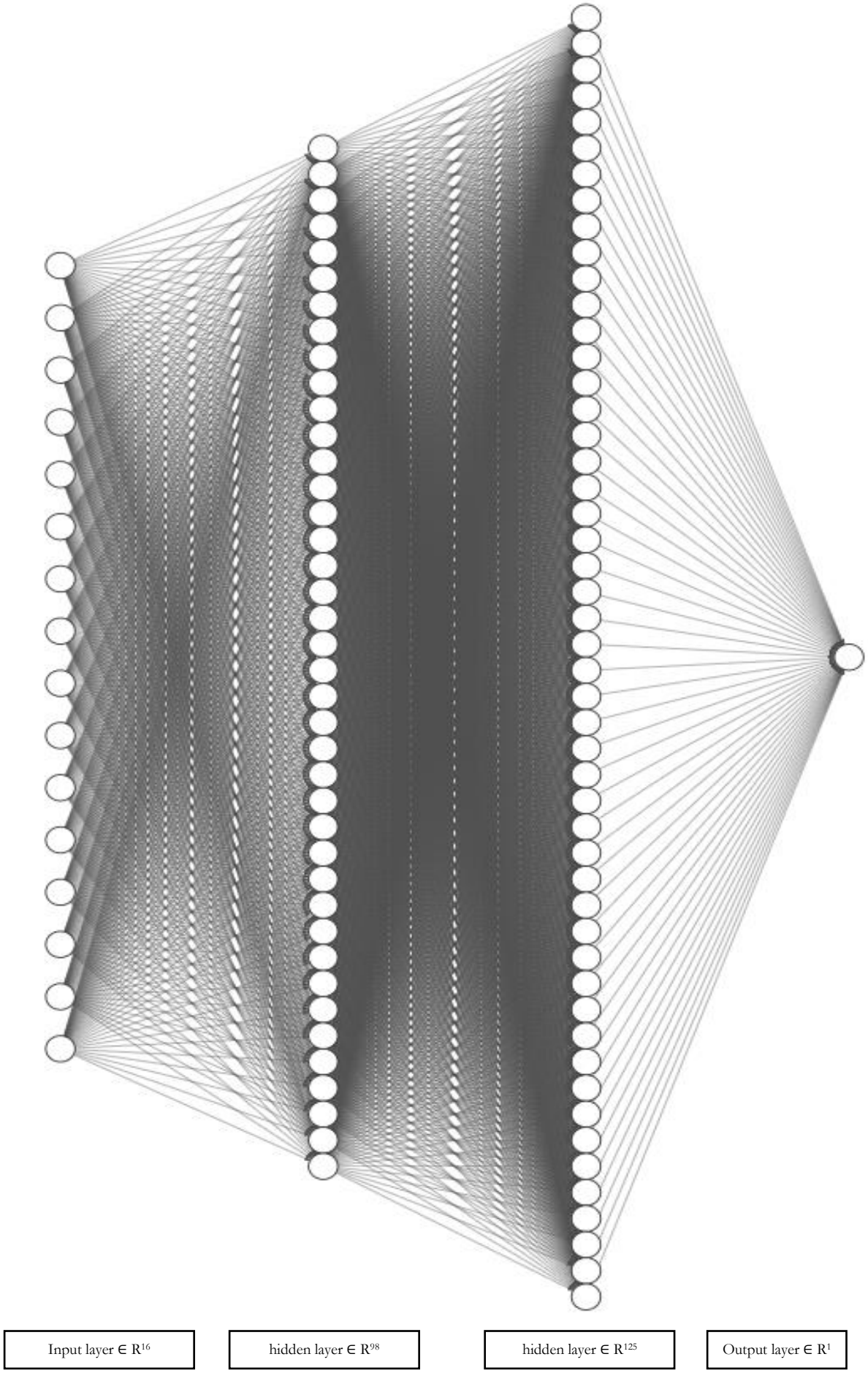


Fig. 8.10. ANN model for the prediction of Pb

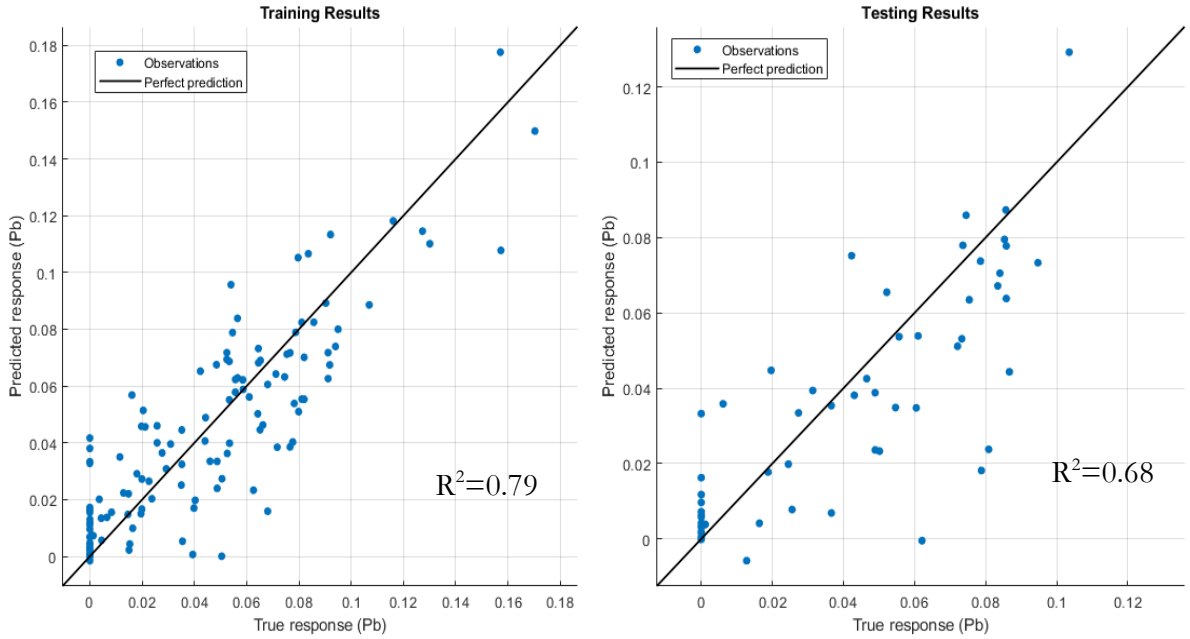


Fig. 8.11. Training/testing R^2 for ANN model for Pb

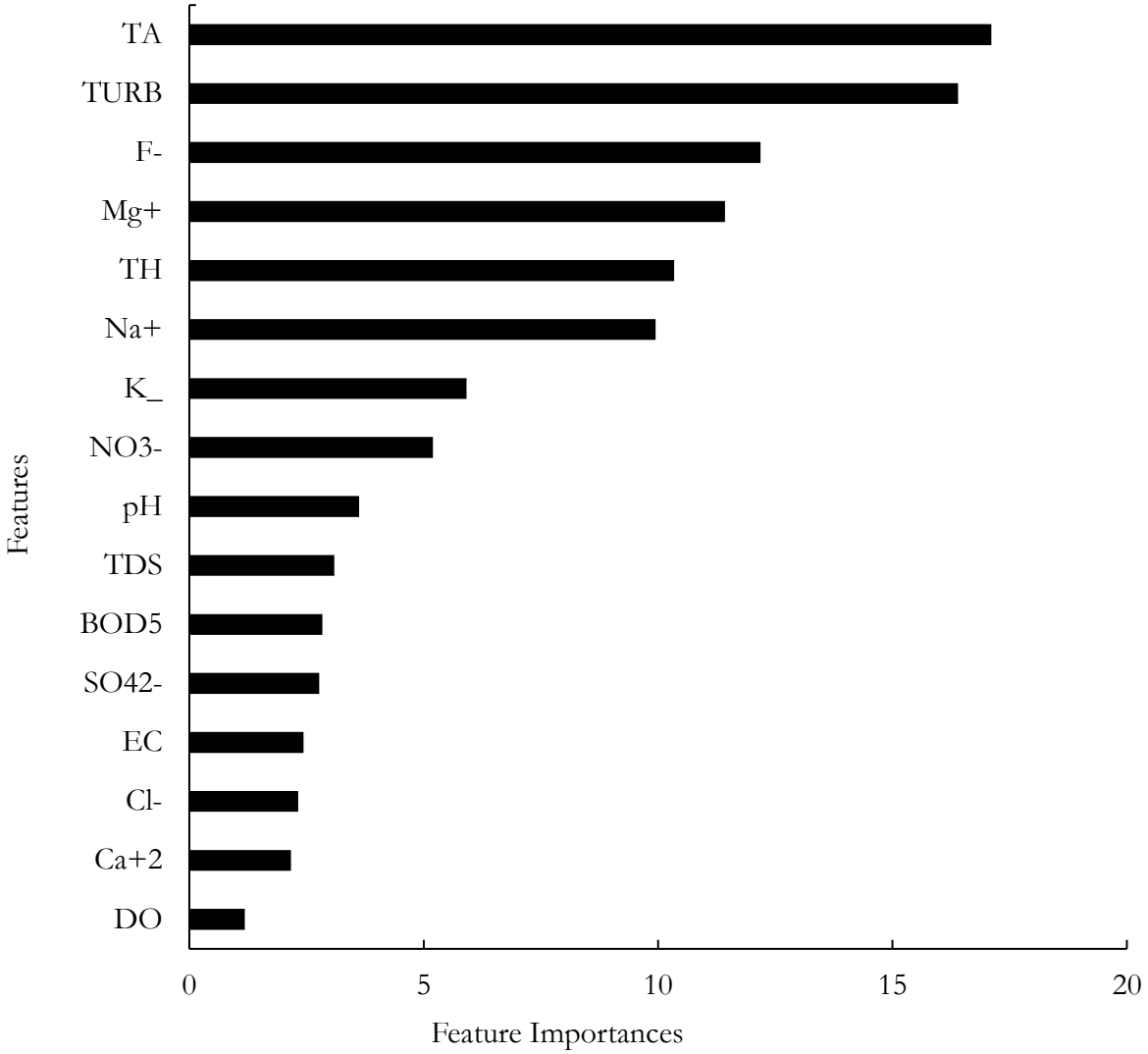


Fig. 8.12. Feature importances of the ANN model for Pb

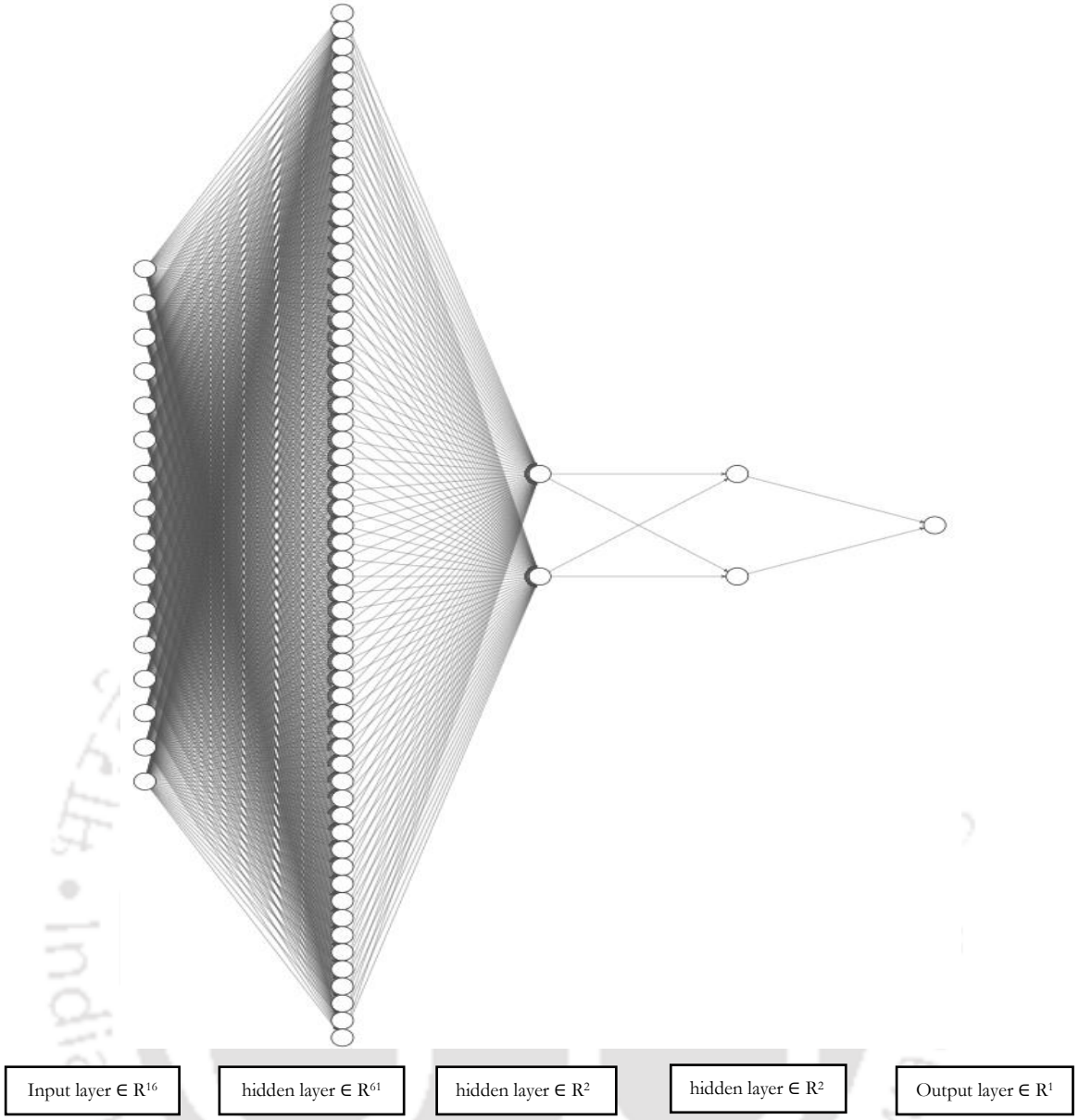


Fig. 8.13. ANN model for the prediction of Zn

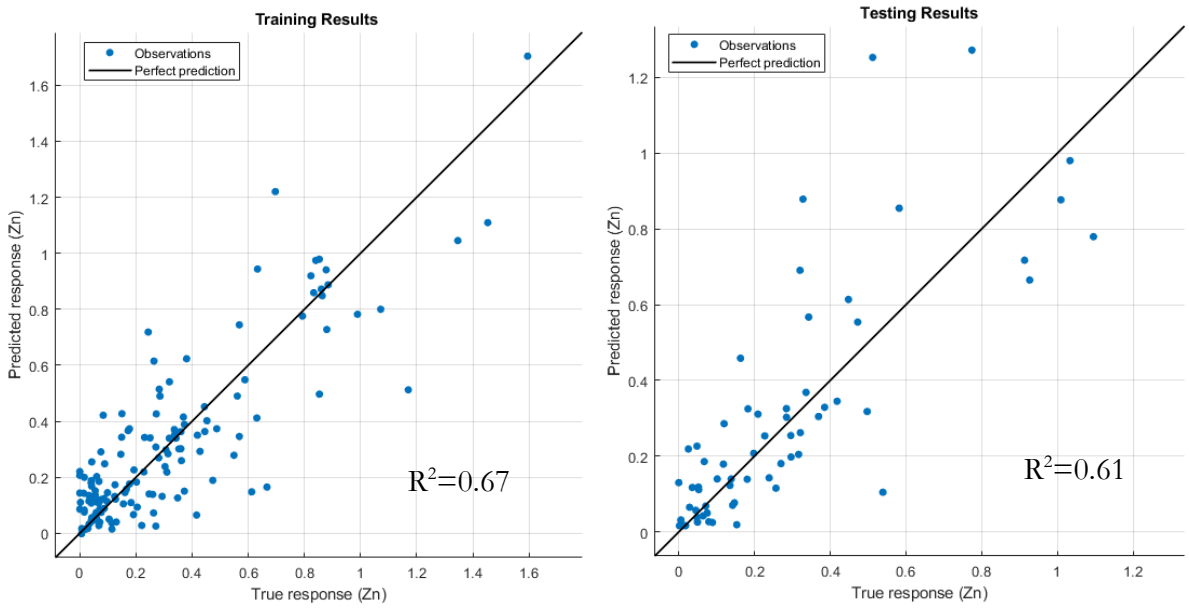


Fig. 8.14. Training/testing R^2 for ANN model for Zn

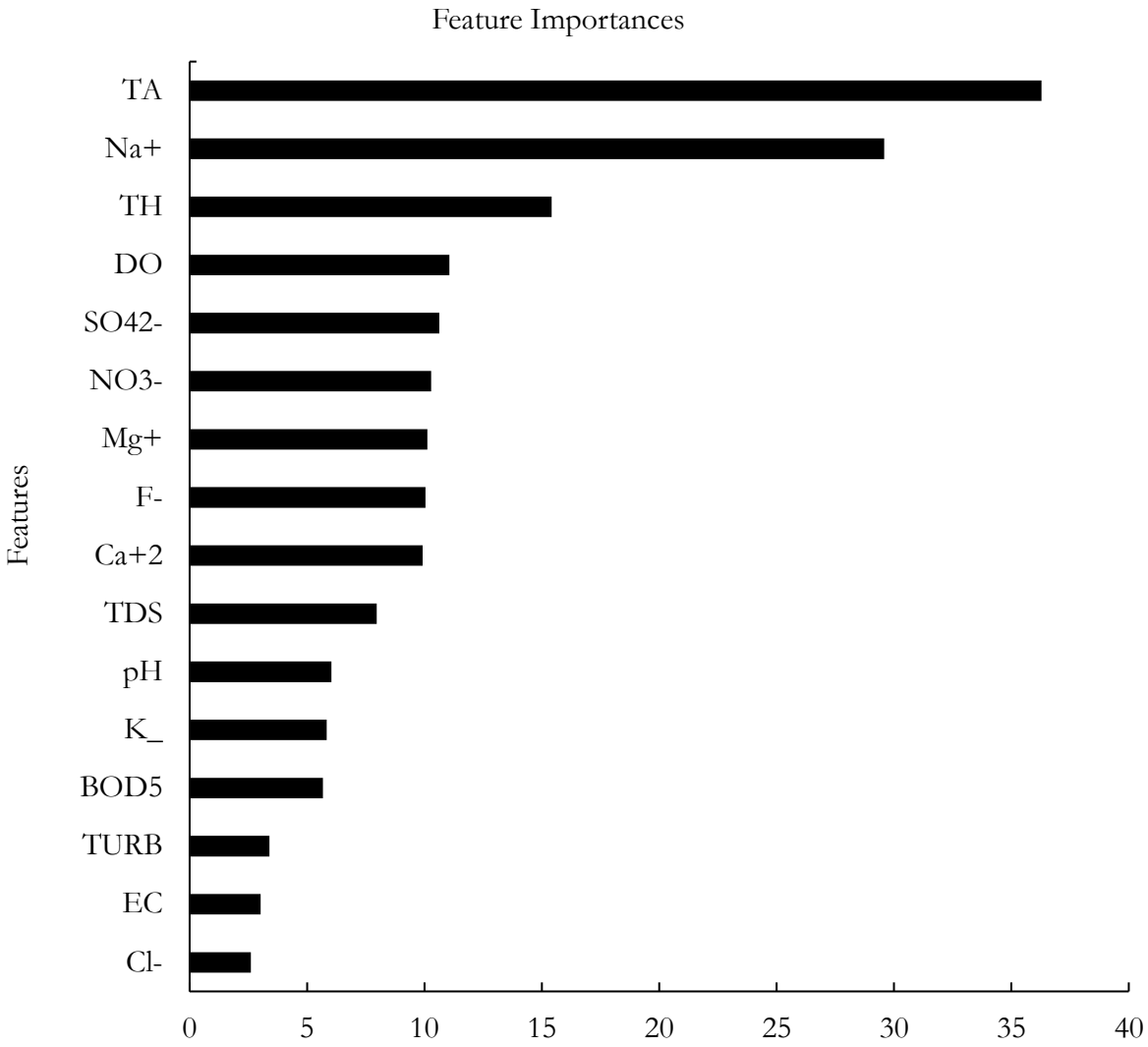


Fig. 8.15. Feature importances of the ANN model for Zn

8.1.3 Random Forest (RF)

The Python' sklearn' library is used to develop an RF model for heavy metal prediction. The dataset is divided into training and testing sets for this purpose. In the case of the RF Regression model, the hyperparameters to be optimised are 'n_estimators', 'max_depth', 'min_samples_split, 'min_samples_leaf,' and 'max_features' using 'RandomSearchCV' and 'GridSearchCV'. The k-fold (5-fold) validation technique is used to optimise these parameters on the training dataset, and the results are validated on the testing set. The following are the Random Forest Regression results for heavy metal estimation:

(a) Cadmium

The optimised hyperparameters for the Cd RF model are 'max_depth': 8, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 338. The random forest algorithm considered 338 (n estimators) different decision trees to develop this model and returned the best decision tree based on the R^2 value. A small part of one of the decision trees (up to depth =3) is shown in Fig. 8.16. The training and testing results for the Cd RF model are shown in Fig. 8.17, and the parameters affecting the concentration of Cd are shown in Fig. 8.18. The parameters that

have the greatest influence on Cd concentration are Mg^{+2}, Ca^{+2}, hardness and Cl^-, all of which have feature importances greater than 0.10.

(b) Iron (Fe)

The optimised hyperparameters for the Fe RF model are 'max_depth': 12, 'max_features': '0.5', 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 371. The random forest algorithm was used to develop this model, which considered 371 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. A small part of one of the decision trees (max_depth =3) is shown in Fig. 8.19. The training and testing results for the Fe RF model are shown in Fig. 8.20, and the parameters affecting the concentration of Fe are shown in Fig. 8.21. The parameters that have the greatest influence on Fe concentration are F^-, turbidity, Ca^{+2} and hardness, all of which have feature importances greater than 0.10.

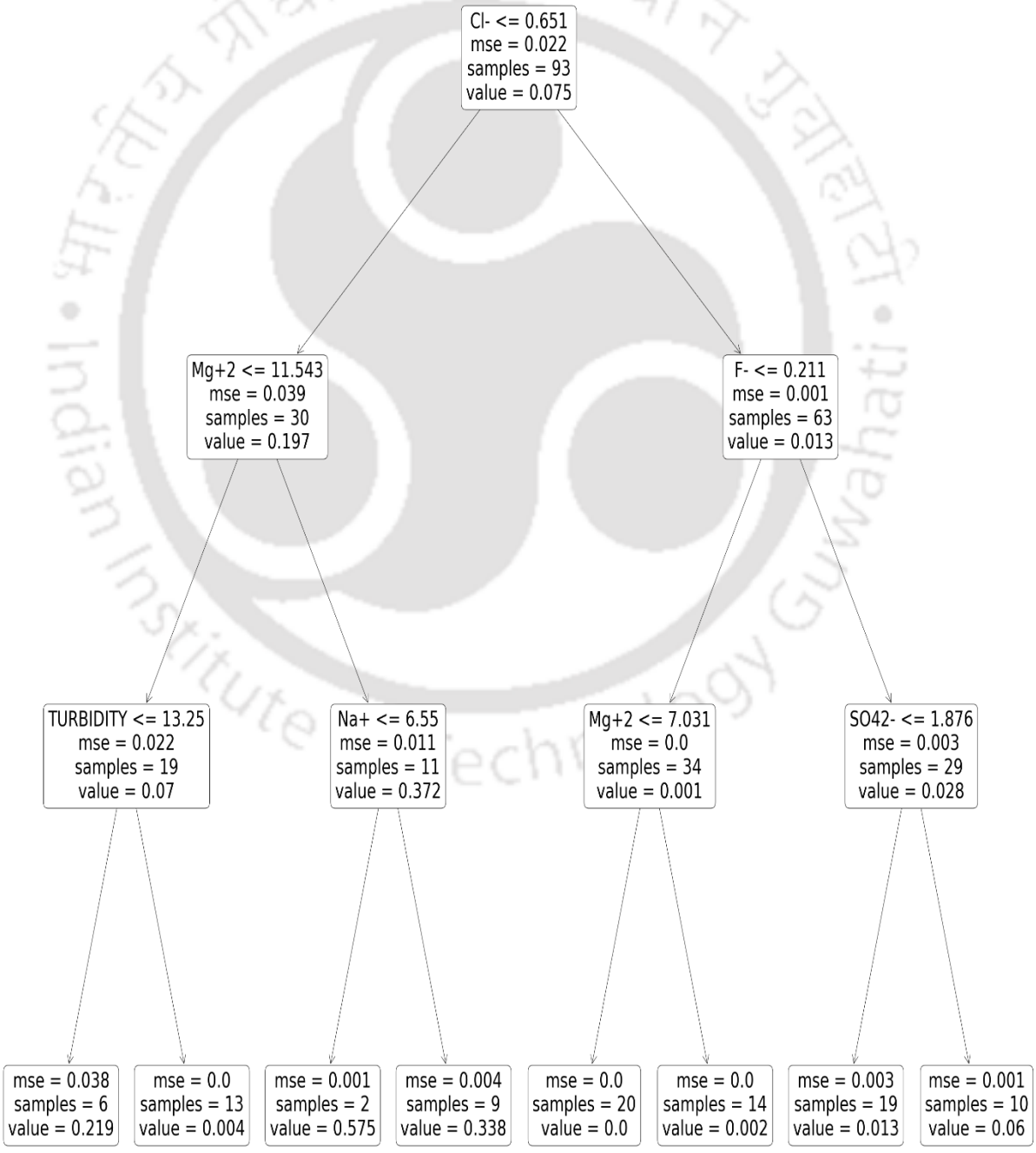


Fig. 8.16. A decision tree (max_depth=3) for the Cd RF model

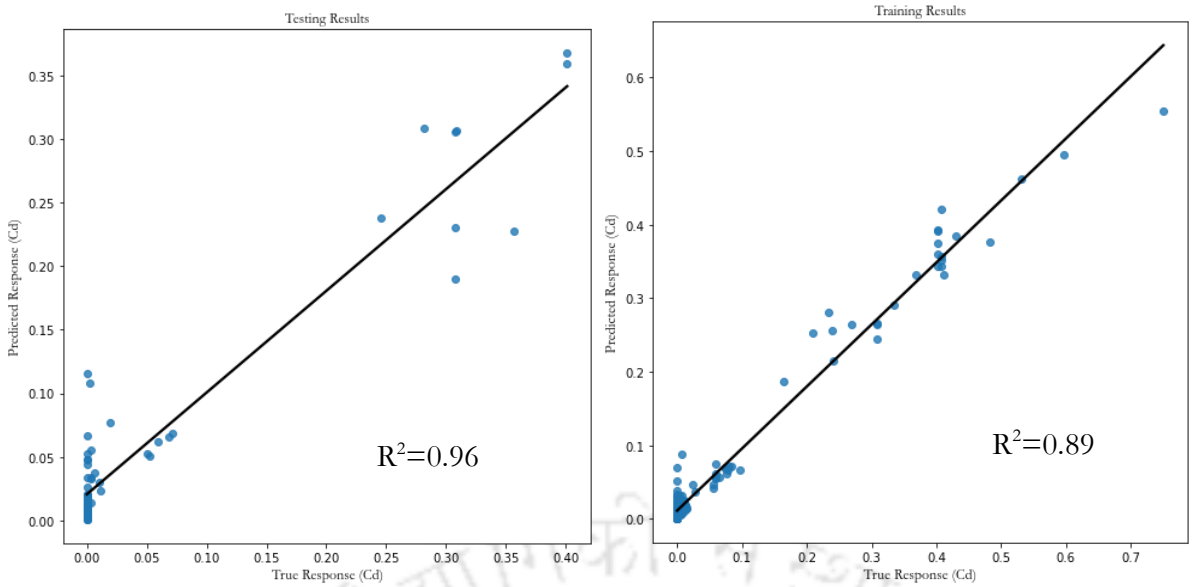


Fig. 8.17. Training and testing R^2 for the Cd RF model

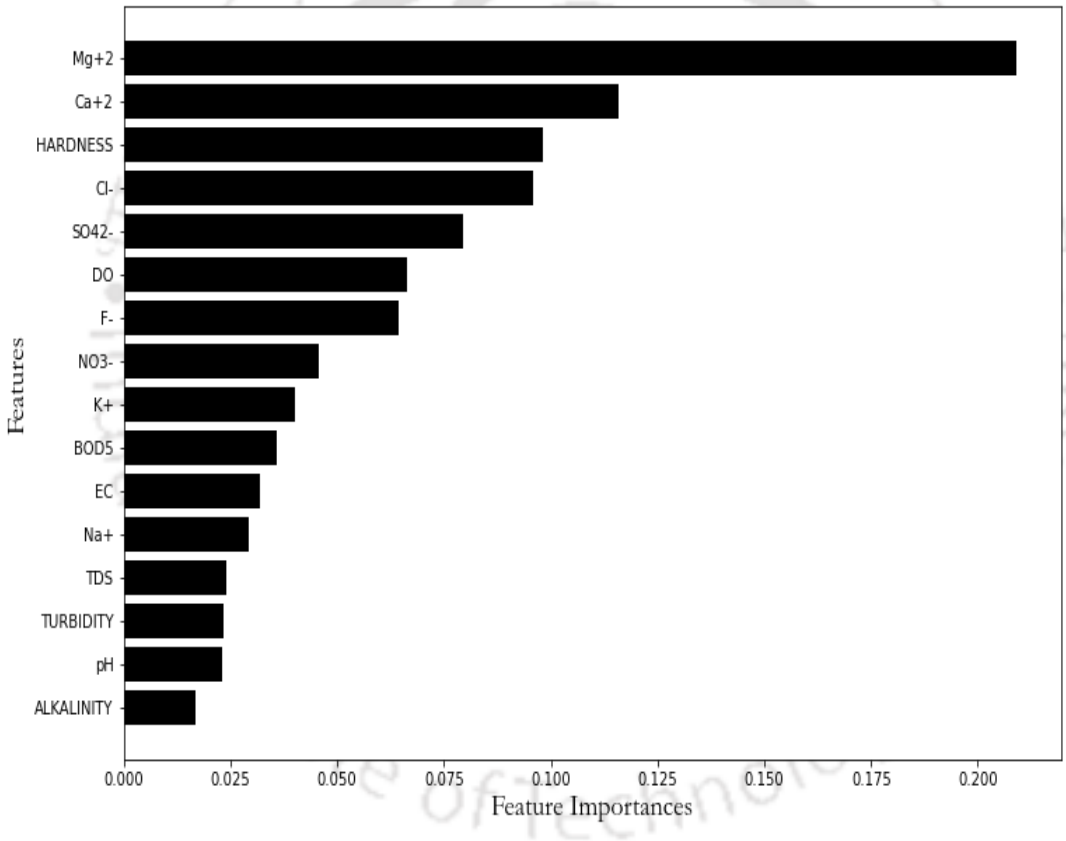


Fig. 8.18. Feature importance of the Cd RF model

(c) Manganese (Mn)

The optimised hyperparameters for the Mn RF model are 'max_depth': 8, 'max_features': sqrt, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 270. The random forest algorithm was used to develop this model, which considered 270 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. A small part of one of the decision trees (up to depth =3) is shown in Fig. 8.22. The training and testing results for the Mn RF model are shown in Fig. 8.23, and the parameters affecting the concentration of Mn are shown in Fig. 8.24.

The parameters that have the greatest influence on Mn concentration are Na^+ and BOD_5, all of which have feature importances greater than 0.10.

(d) Lead (Pb)

The optimised hyperparameters for the Pb RF model are 'max_depth': 10, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 4, 'n_estimators': 47. The random forest algorithm was used to develop this model, which considered 47 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. A small part of one of the decision trees (max_depth =3) is shown in Fig. 8.26. The training and testing results for the Pb RF model are shown in Fig. 8.27, and the parameters affecting the concentration of Pb are shown in Fig. 8.28. The parameters that have the greatest influence on Pb concentration are alkalinity, turbidity and NO_3^-, all of which have feature importances greater than 0.10.

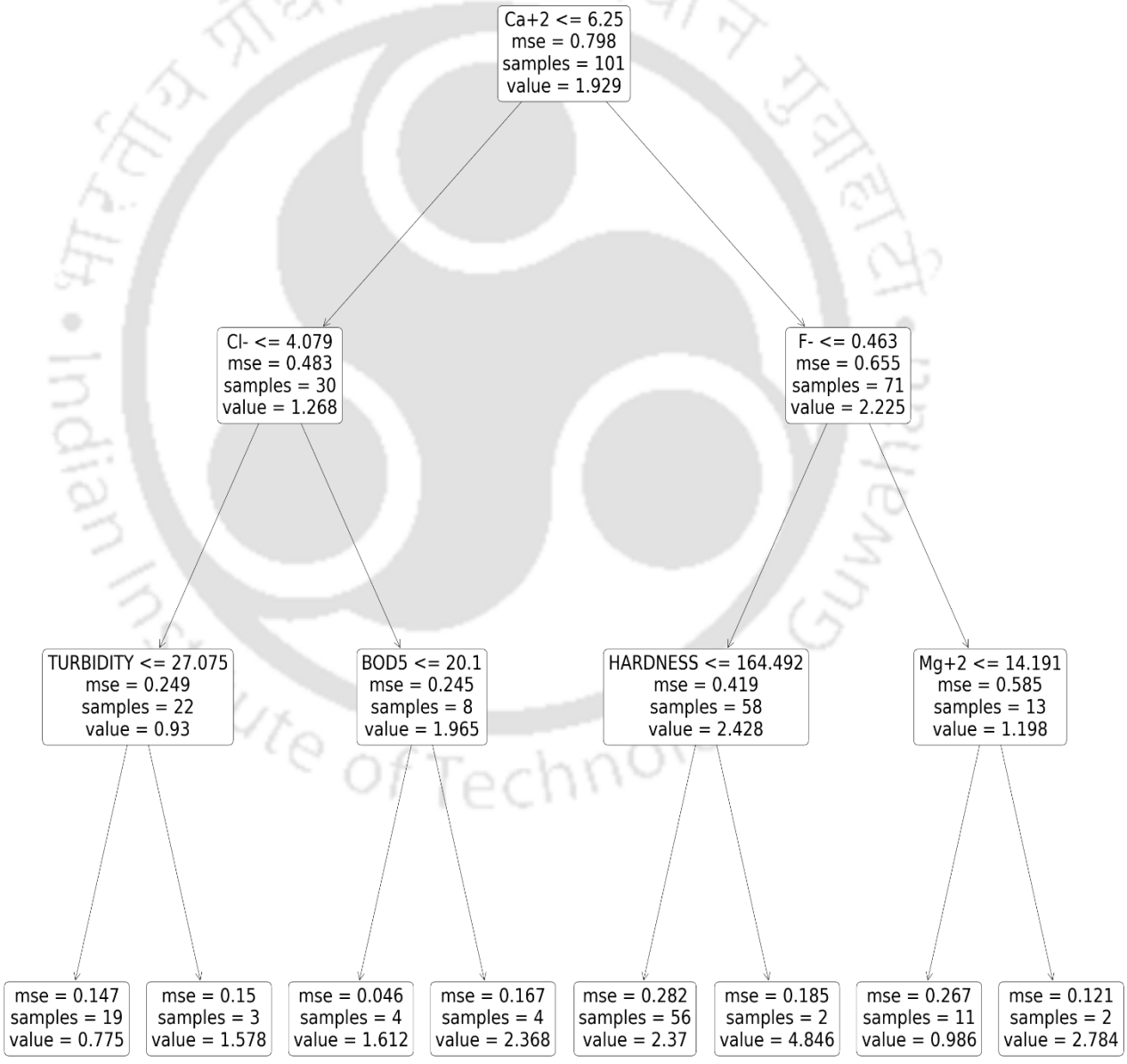


Fig. 8.19. A decision tree (max_depth=3) for Fe RF model

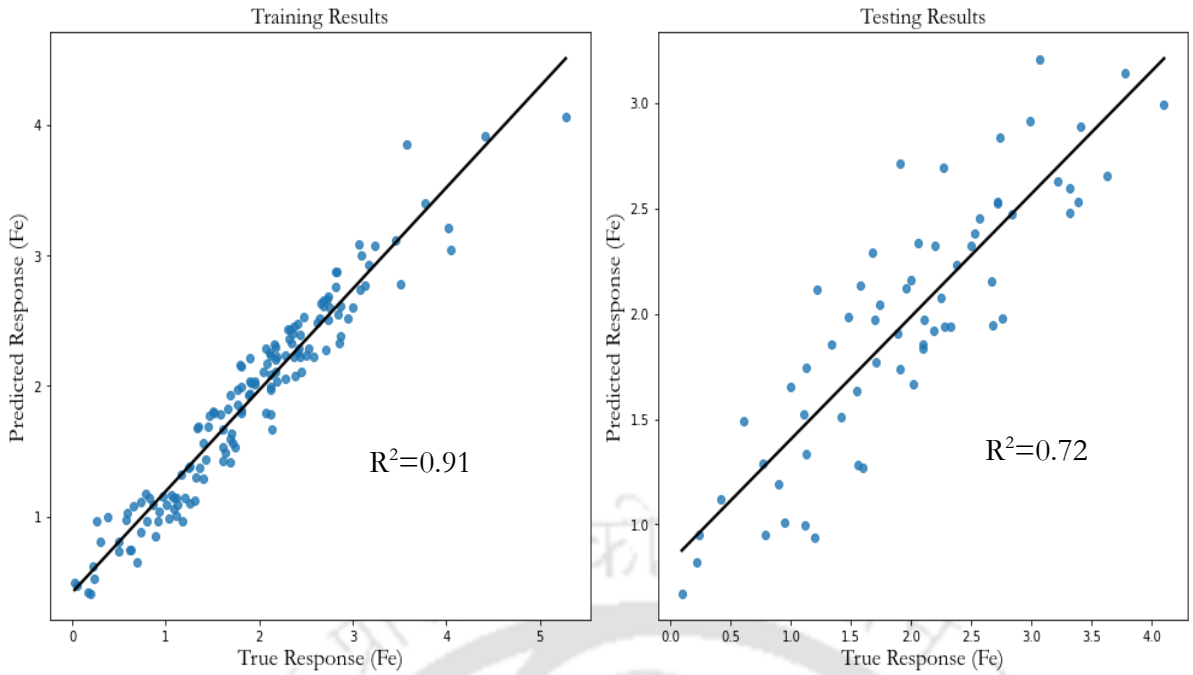


Fig. 8.20. Training and testing R^2 for the Fe RF model

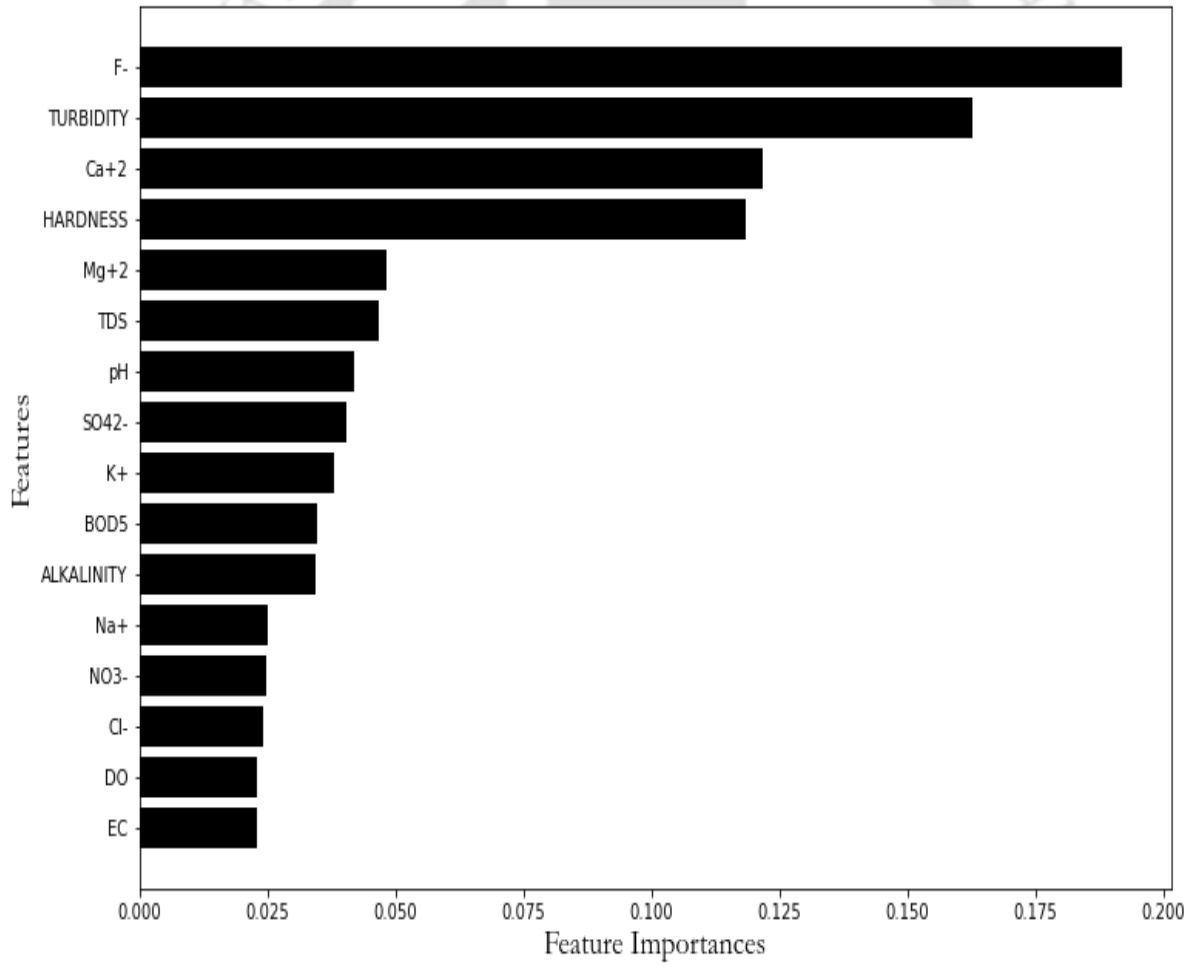


Fig. 8.21. Feature importance of Fe RF model

(e) Zinc (Zn)

The optimised hyperparameters for the Zn RF model are 'max_depth': 12, 'max_features': sqrt, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 363. The random forest algorithm was used to develop this model, which considered 363 (n estimators) different decision trees, and

the best decision tree was estimated using the R^2 value. A small part of one of the decision trees (max_depth =3) is shown in Fig. 8.28. The training and testing results for the Zn RF model are shown in Fig. 8.29, and the parameters affecting the concentration of Zn are shown in Fig. 8.30. The parameters that have the greatest influence on Cu concentration are Na^+ and alkalinity, all of which have feature importances greater than 0.10.

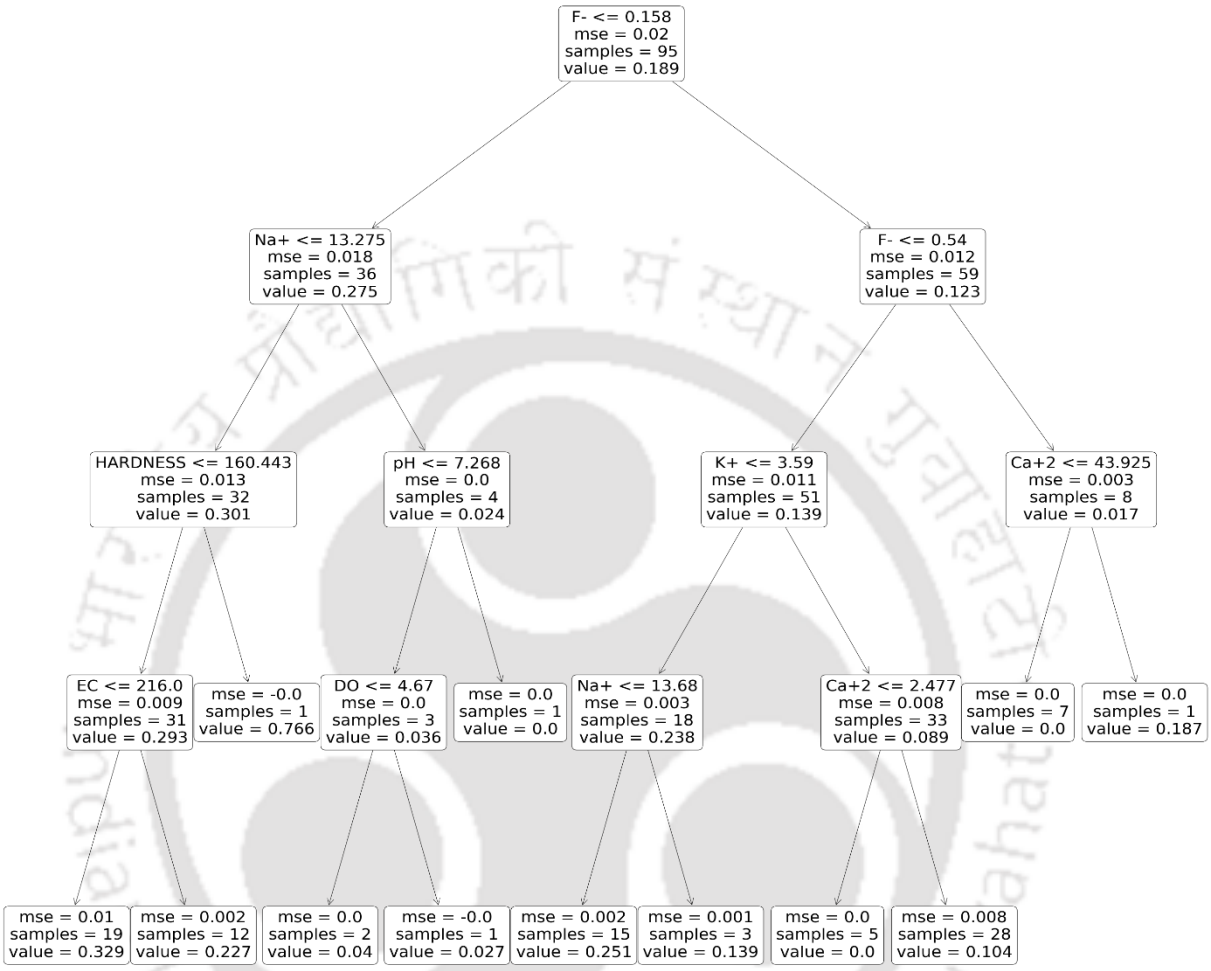


Fig. 8.22. A decision tree (max_depth=3) for Mn RF model

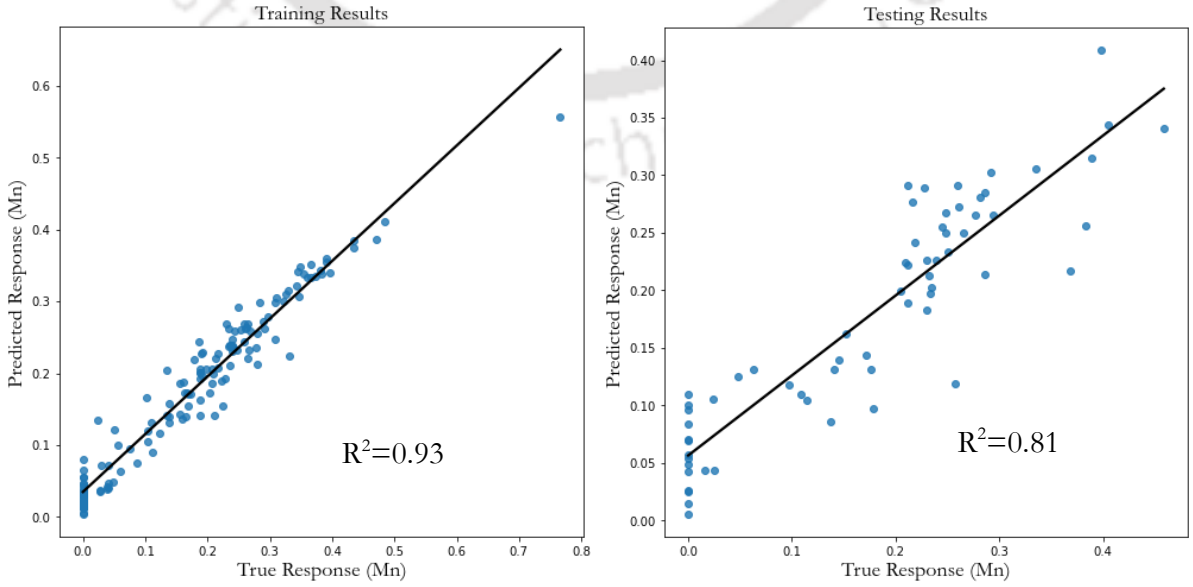


Fig. 8.23. Training and testing R^2 for Mn RF model

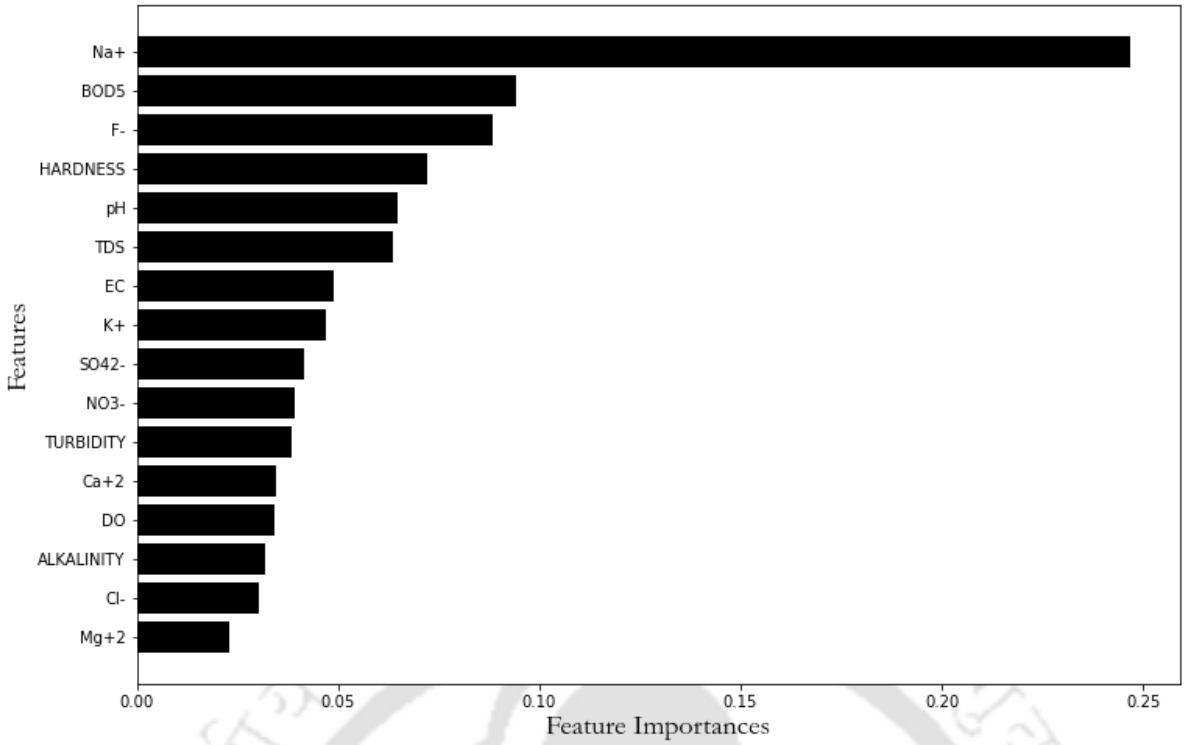


Fig. 8.24. Feature importance of Mn RF model

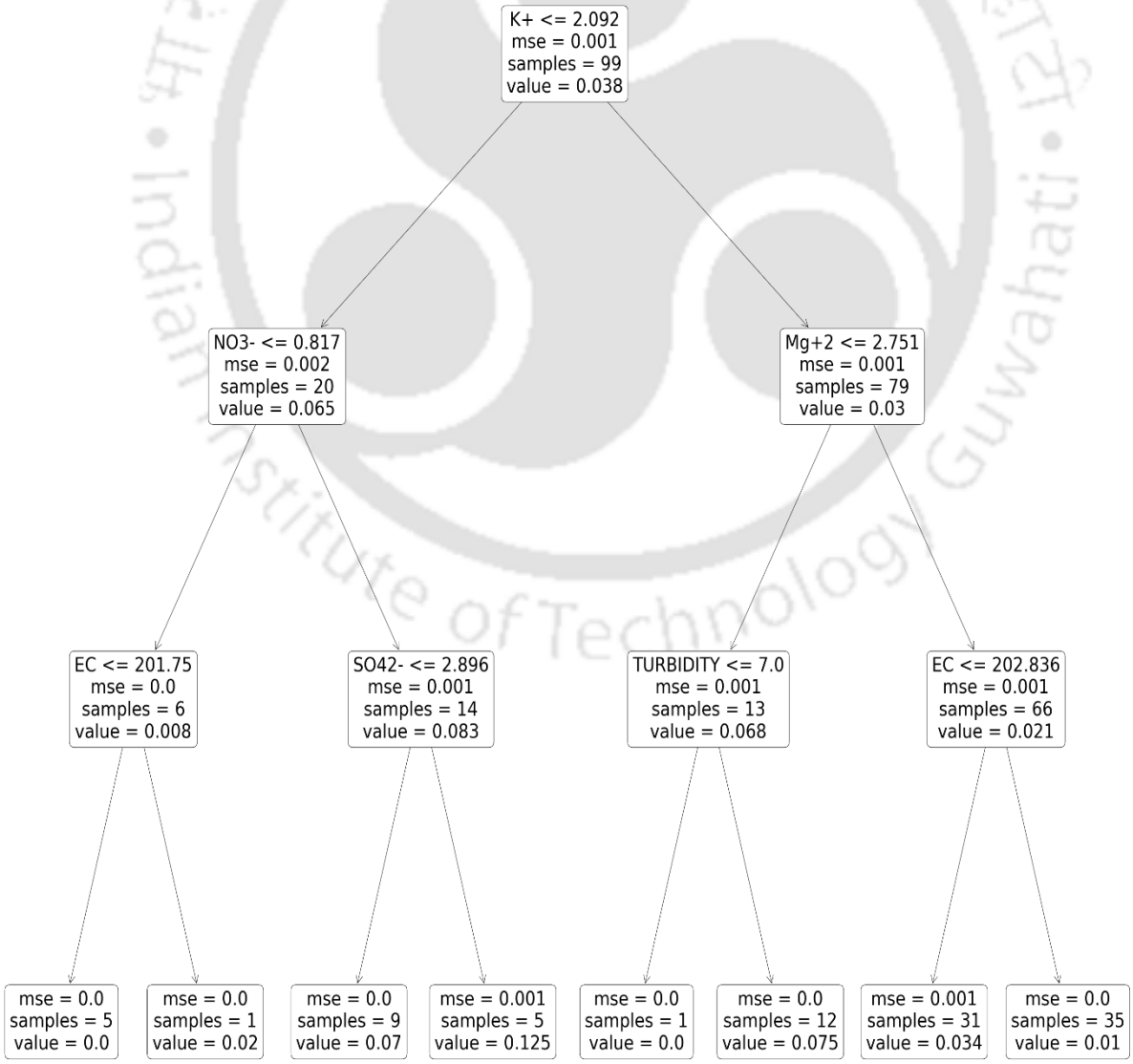


Fig. 8.25. A decision tree (max_depth=3) for Pb RF model

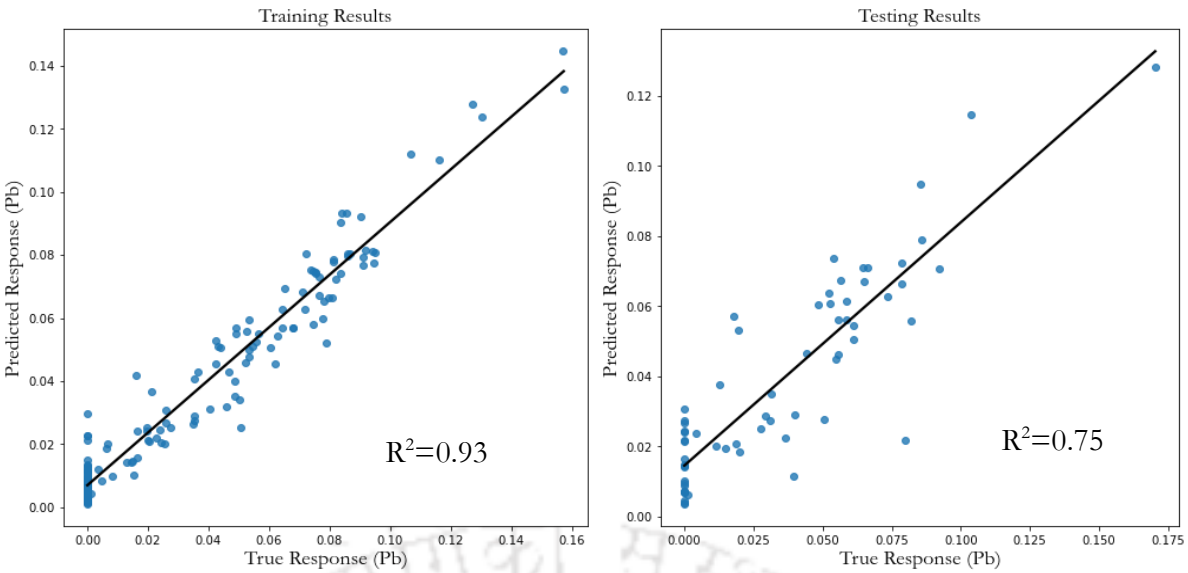


Fig. 8.26. Training and testing R^2 for the Pb RF model

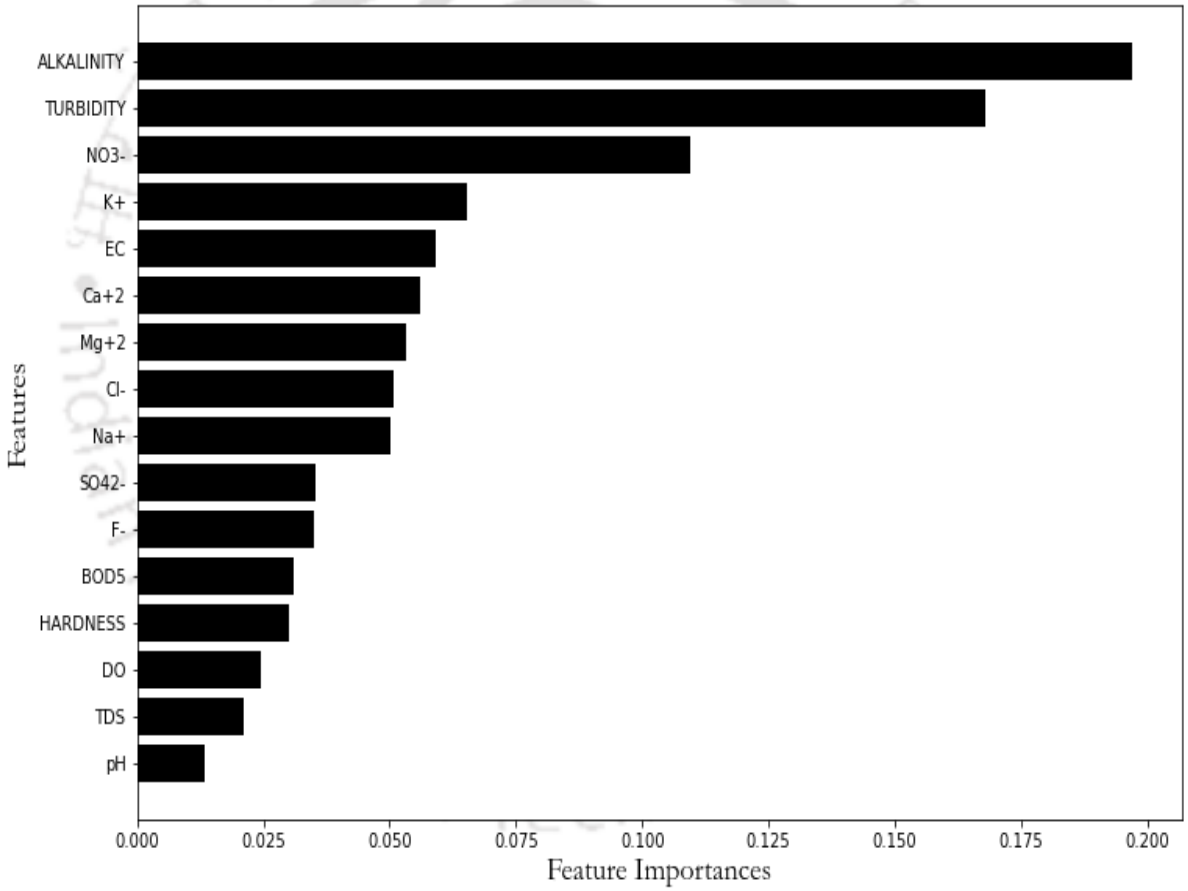


Fig. 8.27. Feature importance of the Pb RF model

8.1.4 Comparison between the models

The prediction capability of new models can only be evaluated using unknown data. For this purpose, the entire dataset of 213 observations is randomly divided into a training set and a testing set for each model. The training set represents 75% of the dataset or 160 observations, and the test set represents 25% of the dataset or 53 observations. It was observed that the RF-based models outperformed ANN and MLR models (Table 8.7). Table 8.8 shows the parameters influencing heavy metal concentrations identified from the RF models. These parameters differ from

heavy metal to heavy metal; no constant parameters were discovered to enhance or decrease heavy metal concentration. According to the RF models, the parameters that need to be addressed for heavy metal pollution abatement include Mg^{+2}, Ca^{+2}, Na^{+}, Cl^{-}, F^{-}, NO_3^{-}, hardness, alkalinity, turbidity, and BOD_5. These parameters should be considered while developing pollution control and management remediation strategies.

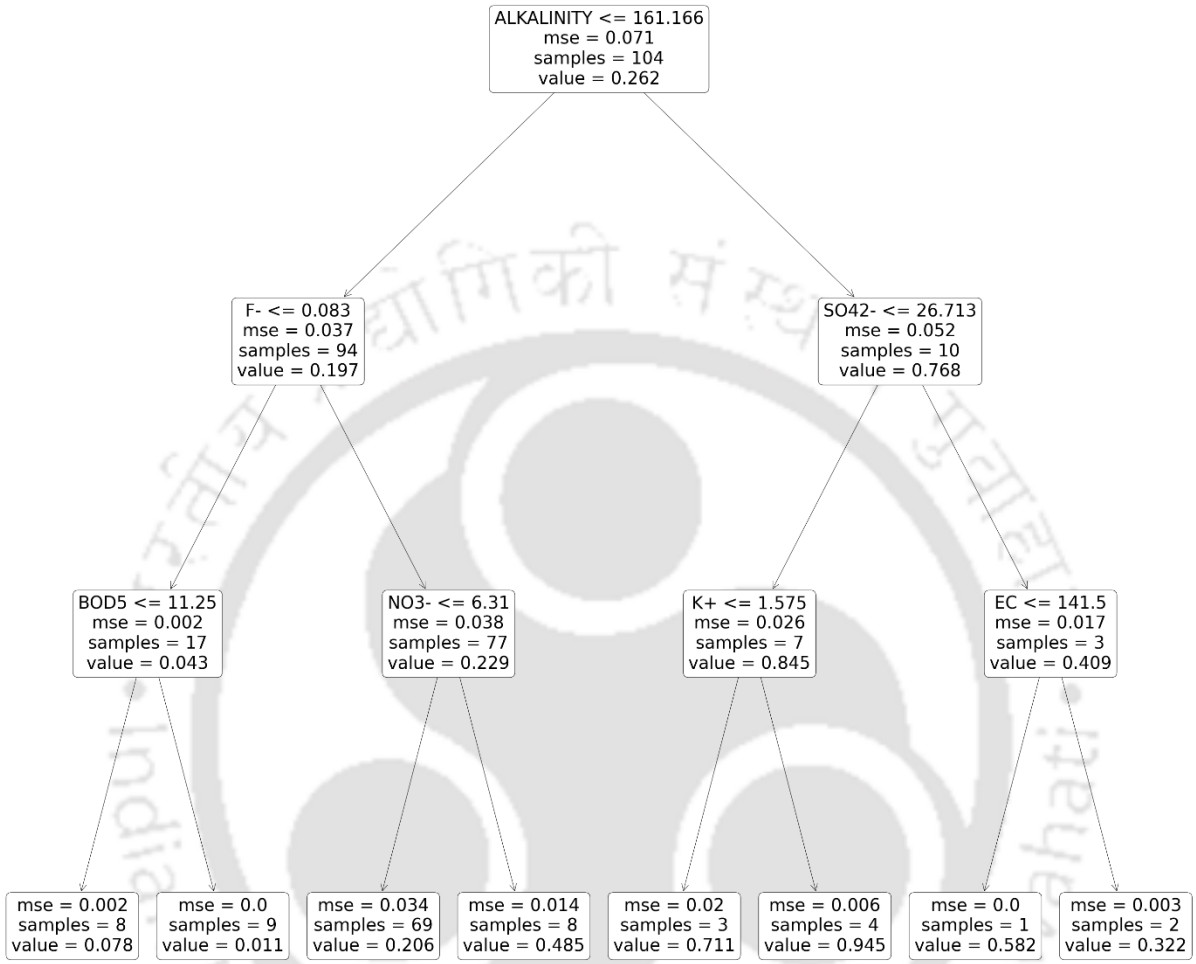


Fig. 8.28. A decision tree (max_depth=3) for Zn RF model

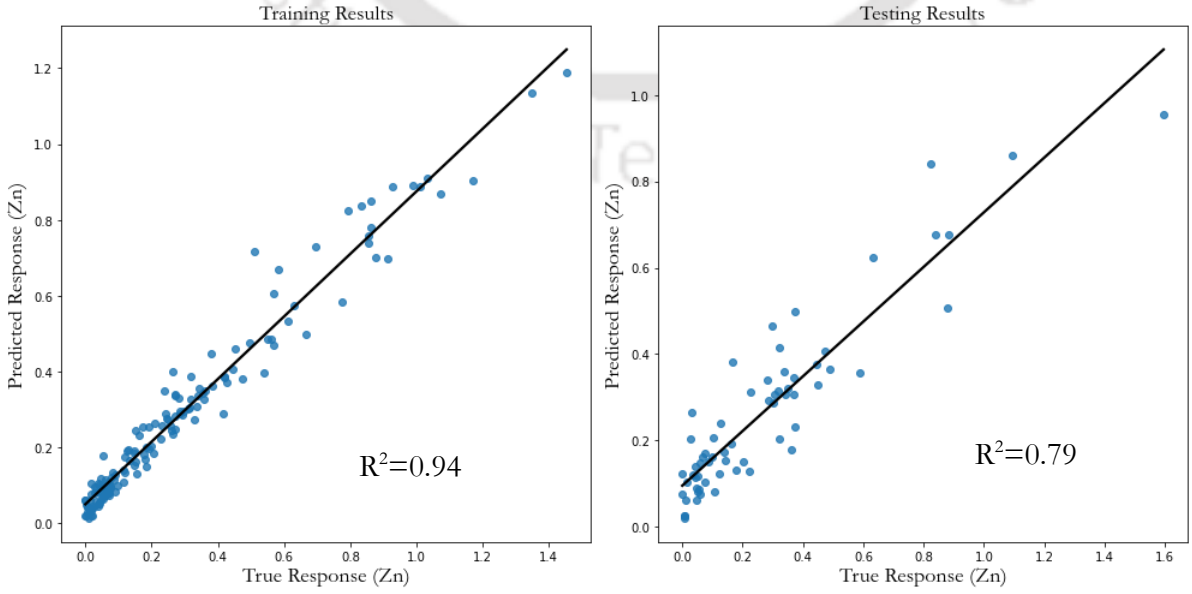


Fig. 8.29. Training and testing R^2 for the Zn RF model

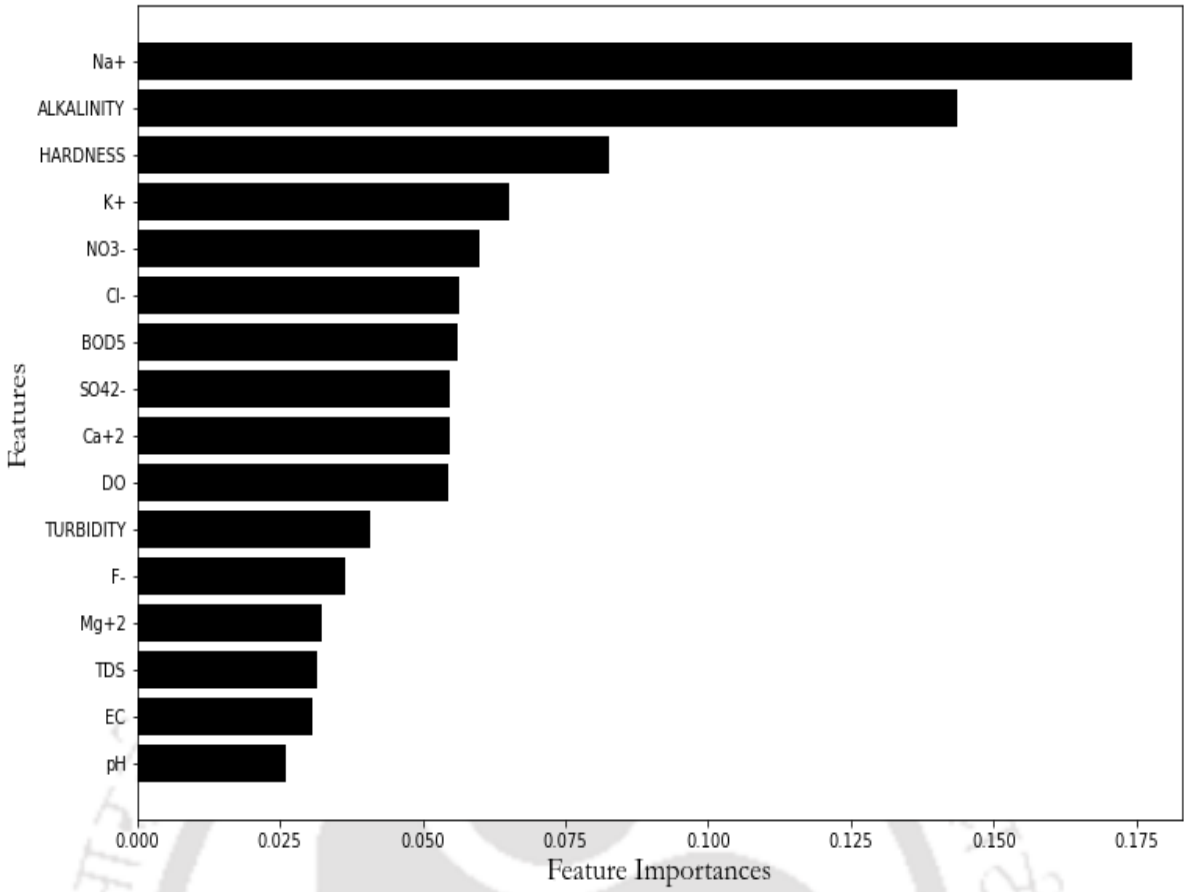


Fig. 8.30. Feature importance of Zn RF model

Table 8.7. The MLR, ANN, and RF (determination coefficients) result on the training (train) set and test set for the studied metals.

Heavy Metals / **Models**	**Cd**		**Fe**		**Mn**		**Pb**		**Zn**	
	Train Set	Test Set	Train Set	Test Set	Train Set	Test Set	Train Set	Test Set	Train Set	Test Set
MLR	0.69	0.4	0.73	0.67	0.78	0.63	0.69	0.59	0.71	0.68
ANN	0.76	0.72	0.73	0.72	0.75	0.7	0.79	0.68	0.67	0.61
RF	0.96	0.89	0.91	0.72	0.93	0.81	0.93	0.75	0.94	0.79

Table 8.8. Factors affecting heavy metal concentrations

Heavy Metals	**Affecting parameters**
Cd	Mg^{+2}, Ca^{+2}, hardness, Cl^-
Fe	F^-, turbidity, Ca^{+2}, hardness
Mn	Na^+, BOD_5
Pb	Alkalinity, turbidity, NO_3^-
Zn	Na^+, alkalinity

8.2 Second Regression Model

In sediments, heavy metals exist in five fractions, i.e., F1, F2, F3, F4, and F5. The bonding of these fractions with sediment particles is different for different fractions. The F1, F2, and F3 fractions can release their metal loads by lowering the pH and are the most dangerous. The F4 fraction can be released by the sediment to the water column under oxidising conditions such as dredging, current, flooding, tides, etc. The F5 fraction is unlikely to be released by the sediments into the water column. All these fractions have different behaviours with respect to remobilisation under changing environmental conditions (Förstner and Salomons, 1980a). An attempt has been made to develop a model to explain the relationship between heavy metals in the water column and heavy metals in sediment. From section 8.1, it is observed that the performance of the random forest (RF) regression models were better than multiple linear regression (MLR) and artificial neural network in dealing with water quality dataset. Therefore, the second regression model is developed only using random forest (RF) regression. The models for different models are as follows:

(a) Cadmium (Cd)

The dataset utilised for model development for Cd is shown in Table 8.9. The optimised hyperparameters for the Cd speciation RF model are 'max_depth': 8, 'max_features': auto, 'min_samples_leaf': 1, 'min_samples_split': 10, 'n_estimators': 29. The random forest algorithm was used to develop this model, which considered 29 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. Fig. 8.31 shows one of the decision trees employed by the random forest. The training and testing results for the Cd speciation RF model are shown in Fig. 8.32, and the fractions of metals in benthic sediments affecting the concentration of Cd in the water column are shown in Fig. 8.33. The contribution order of the different Cd fractions in benthic sediments is F1, F3, and F2. F1 has the highest 67% contribution, followed by F3 with 17% and F2 with 14% to the concentration of Cd in water.

Table 8.9. Dataset for model development of Cd

Variable		N	Mean	Std Dev	Min	Q1	Median	Q3	Max
Cd in Sediments (mg/kg)	F1	108	0.96	1.17	0.00	0.05	0.47	1.47	3.99
	F2	108	0.87	1.03	0.00	0.07	0.47	0.95	3.87
	F3	108	0.71	0.87	0.00	0.09	0.48	0.79	6.47
	F4	108	0.82	1.32	0.00	0.05	0.59	0.82	8.89
	F5	108	2.20	1.92	0.00	0.19	1.81	3.47	6.14
Cd in water (mg/L)		108	0.05	0.10	0.00	0.00	0.00	0.07	0.50

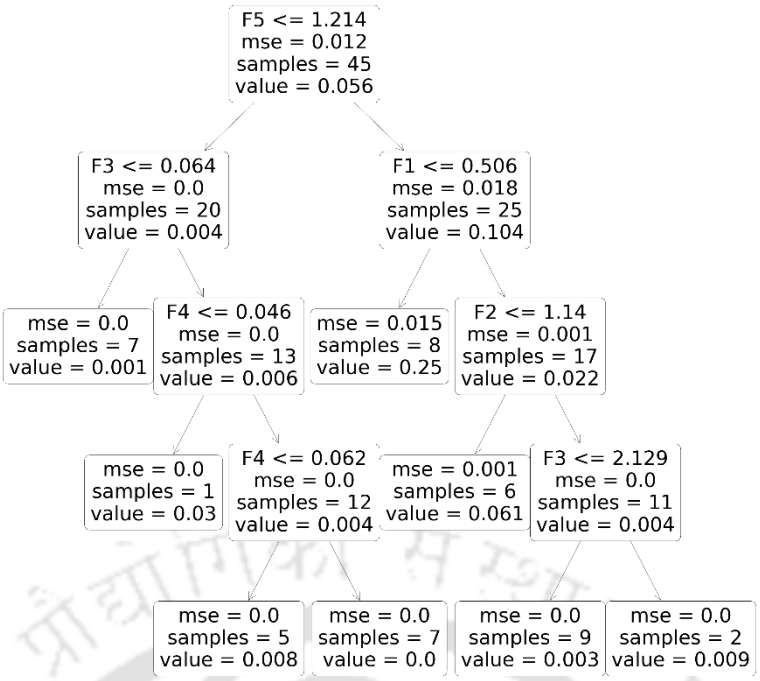


Fig. 8.31. A decision tree for the Cd speciation RF model

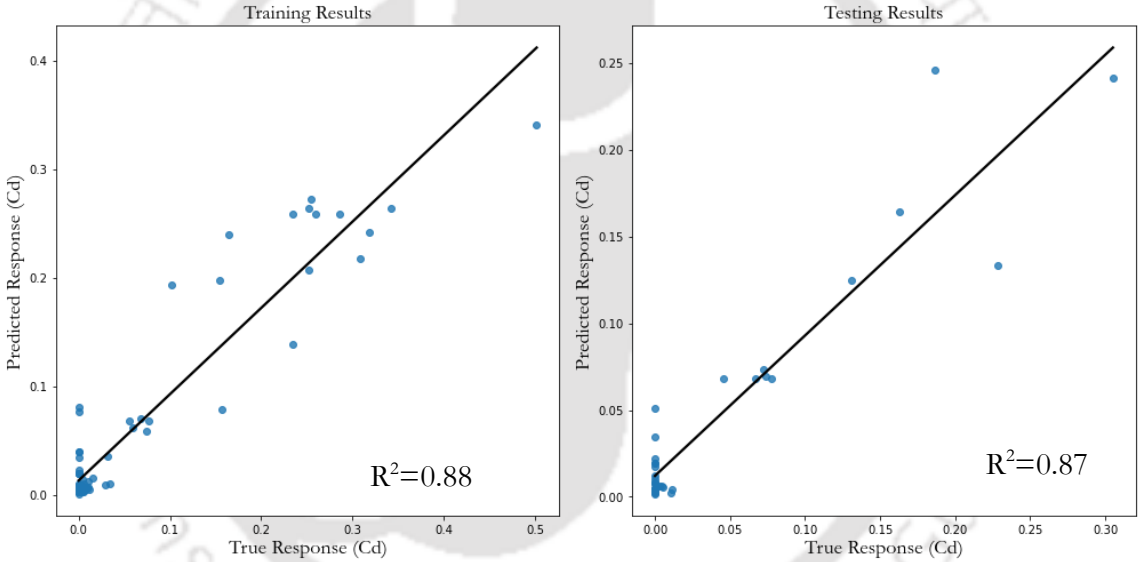


Fig. 8.32. Training and testing R^2 for Cd speciation RF model

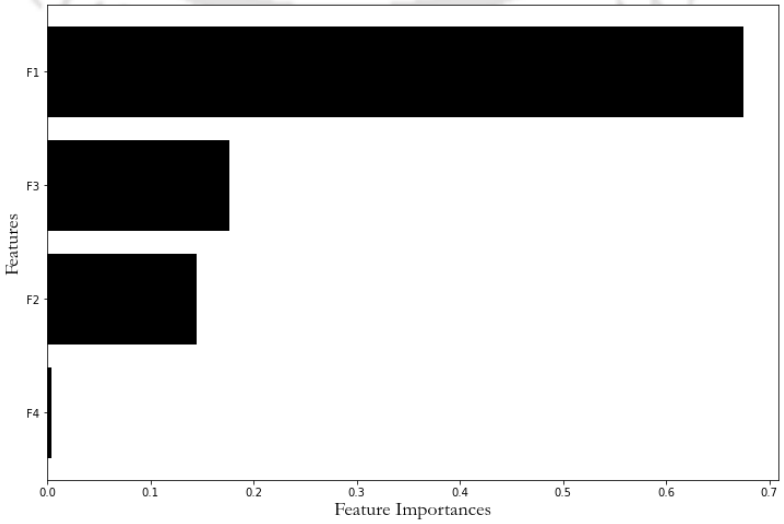


Fig. 8.33. Feature importances of Cd speciation RF model

(b) Iron (Fe)

The dataset utilised for model development for Fe is shown in Table 8.10. The optimised hyperparameters for the Fe speciation RF model are 'max_depth': 12, 'max_features': 0.5, 'min_samples_leaf': 1, 'min_samples_split': 6, 'n_estimators': 51. The random forest algorithm was used to develop this model, which considered 51 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. Fig. 8.34 shows one of the decision trees employed by the random forest. The training and testing results for the Fe speciation RF model are shown in Fig. 8.35, and the fractions of metals in benthic sediments affecting the concentration of Fe in the water column are shown in Fig. 8.36. The contribution order of the different fractions of Fe in benthic sediments is F3, F2, F4, and F1. F3 has the highest 39% contribution, followed by F2 with 22%, F4 with 20% and F1 with 19% to the concentration of Fe in water.

Table 8.10. Dataset for model development of Fe

Variable		N	Mean	Std Dev	Min	Q1	Median	Q3	Max
Fe in sediments (mg/kg)	F1	108	135.300	183.700	3.200	16.200	26.600	238.400	683.700
	F2	108	181.900	165.800	3.200	48.100	165.300	286.600	679.200
	F3	108	1649.000	2011.000	3.000	34.000	541.000	2668.000	8148.000
	F4	108	165.900	223.300	2.900	4.900	12.500	357.800	1044.600
	F5	108	2244.000	2485.000	10.000	507.000	821.000	4493.000	8360.000
Fe in Water (mg/L)		108	1.875	0.640	0.132	1.502	1.996	2.287	3.157

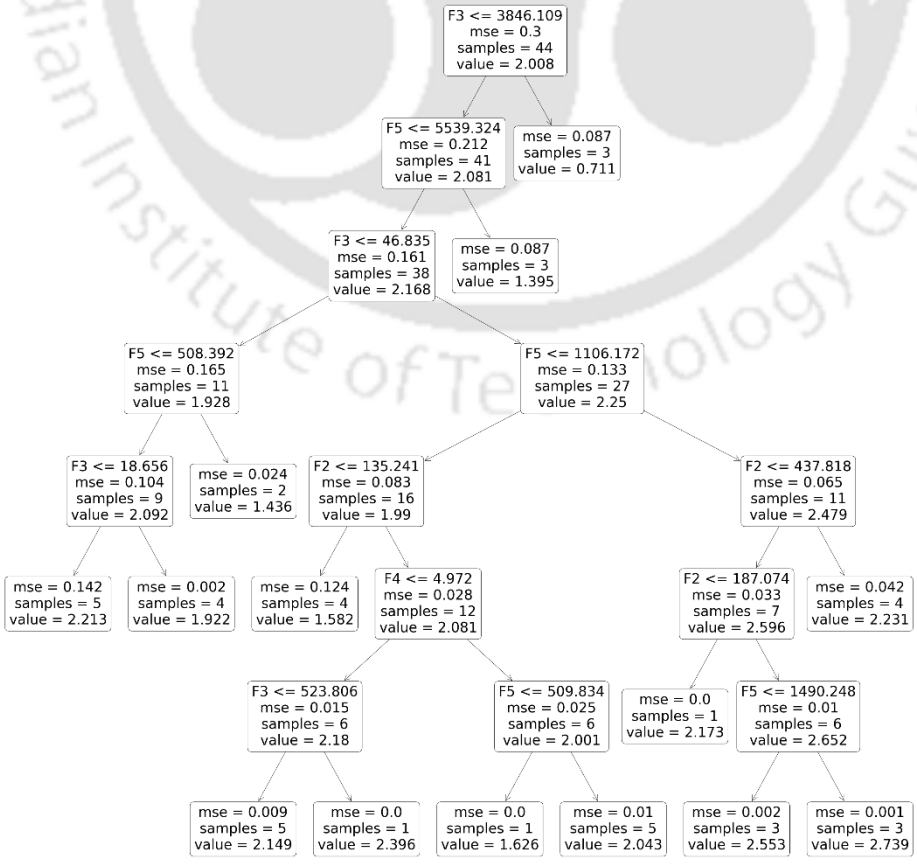


Fig. 8.34. A decision tree for the Fe speciation RF model

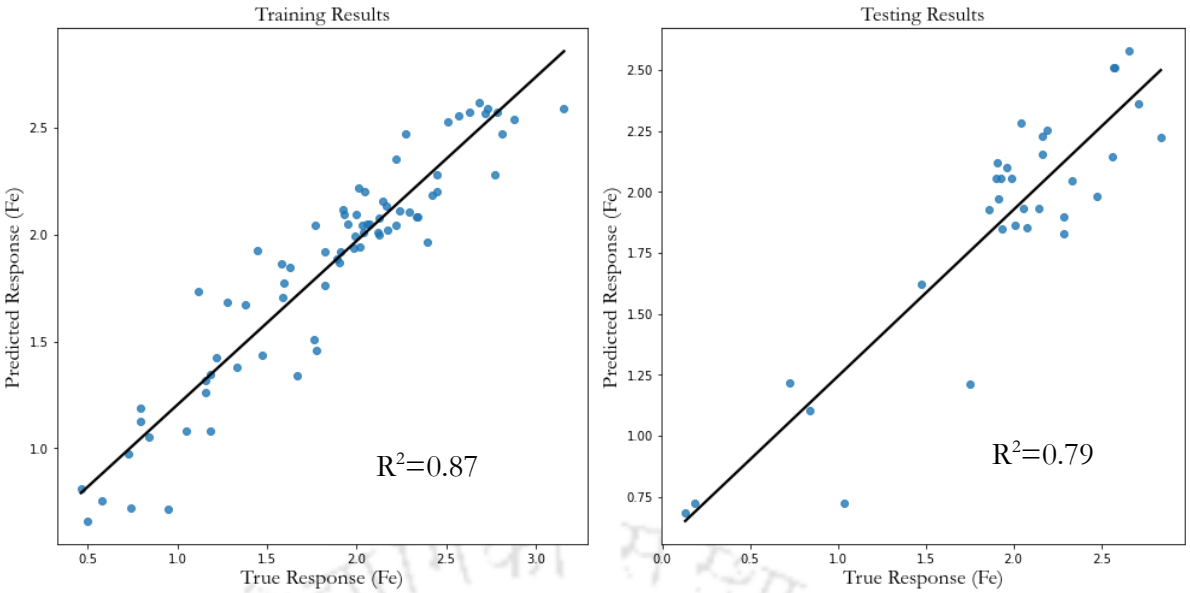


Fig. 8.35. Training and testing R^2 for Fe speciation RF model

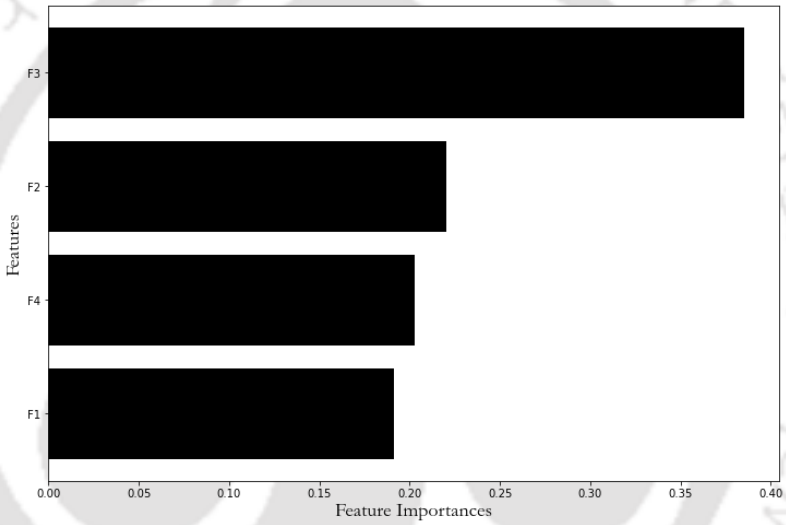


Fig. 8.36. Feature importance of Fe speciation RF model

(c) Manganese (Mn)

The dataset utilised for model development for Mn is shown in Table 8.11. The optimised hyperparameters for the Mn speciation RF model are 'max_depth': 10, 'max_features': sqrt, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 252. The random forest algorithm was used to develop this model, which considered 252 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. Fig. 8.37 shows a small part (max_depth=5) of one of the decision trees employed by the random forest. The training and testing results for the Mn speciation RF model are shown in Fig. 8.38, and the fractions of metals in benthic sediments affecting the concentration of Mn in the water column are shown in Fig. 8.39. The contribution order of the different fractions of Mn in benthic sediments is F4, F2, F3, and F1. F4 has the highest 59% contribution, followed by F2 with 15%, F3 with 13% and F1 with 12% to the concentration of Mn in water.

Table 8.11. Dataset for model development of Mn

Variable		N	Mean	Std Dev	Min	Q1	Median	Q3	Max
Mn in sedi-ments (mg/kg)	F1	108	49.580	56.160	3.530	20.220	32.040	56.330	312.200
	F2	108	22.320	17.830	0.000	10.440	17.990	31.740	84.590
	F3	108	93.500	130.400	4.600	17.600	34.300	100.600	590.100
	F4	108	18.320	17.320	0.000	2.760	15.960	28.690	84.990
	F5	108	80.260	64.440	4.080	24.490	76.410	113.410	299.590
Mn in Water (mg/L)		108	0.176	0.118	0.000	0.057	0.212	0.267	0.446

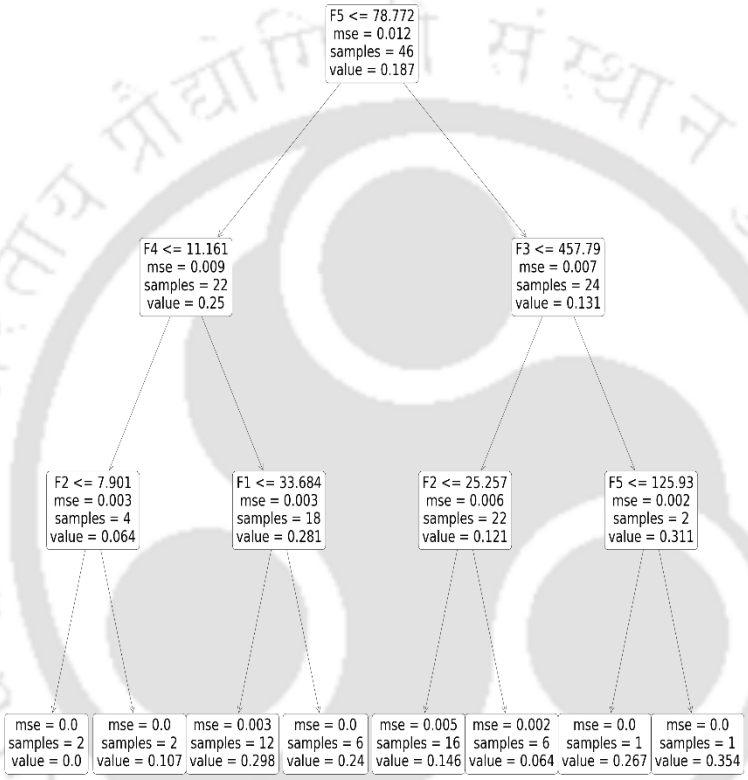


Fig. 8.37. A decision tree for Mn speciation RF model

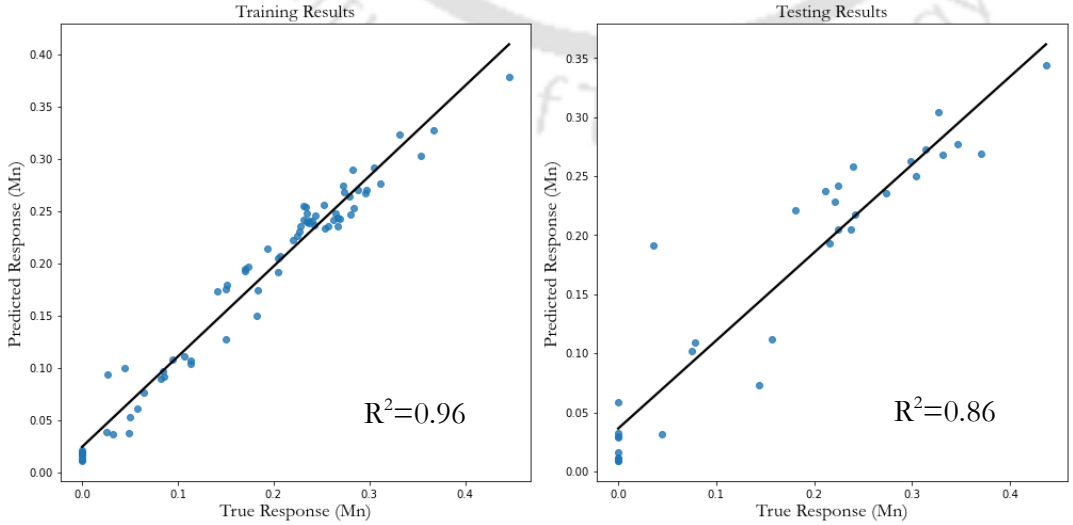


Fig. 8.38. Training and testing R^2 for Mn speciation RF model

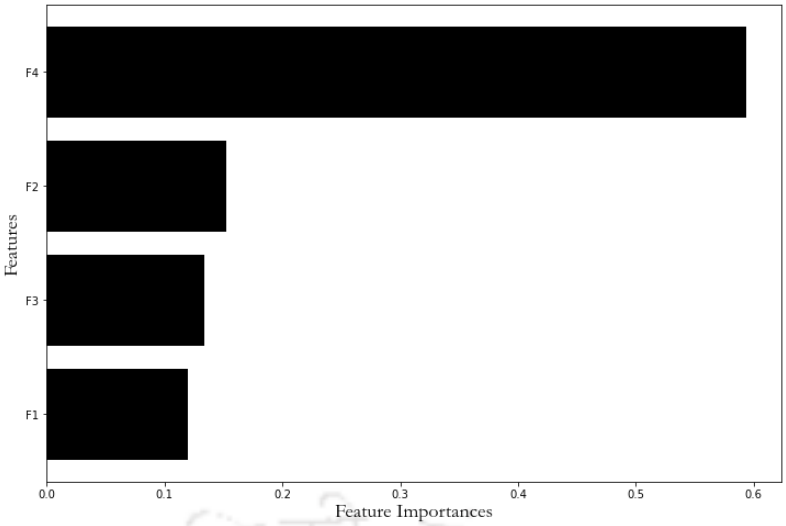


Fig. 8.39. Feature importance of Mn speciation RF model

(d) Lead (Pb)

The dataset utilised for model development for Pb is shown in Table 8.12. The optimised hyperparameters for the Pb speciation RF model are 'max_depth': 10, 'max_features': auto, 'min_samples_leaf': 2, 'min_samples_split': 3, 'n_estimators': 390. The random forest algorithm was used to develop this model, which considered 390 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. Fig. 8.40 shows one of the decision trees employed by the random forest. The training and testing results for the Pb speciation RF model are shown in Fig. 8.41, and the fractions of metals in benthic sediments affecting the concentration of Pb in the water column are shown in Fig. 8.42. The contribution order of the different fractions of Pb in benthic sediments is F1, F3, F4, and F2. F1 has the highest 34% contribution, followed by F3 with 33%, F4 with 23% and F2 with 10% to the concentration of Pb in water.

Table 8.12. Dataset for model development of Pb

Variable		N	Mean	Std Dev	Min	Q1	Median	Q3	Max
	F1	108	3.681	5.415	0.191	0.822	1.878	3.461	21.533
Pb in	**F2**	108	3.693	5.151	0.206	0.602	1.681	3.479	20.874
sediments	**F3**	108	3.685	4.699	0.358	1.006	2.171	3.480	18.714
(mg/kg)	**F4**	108	4.643	4.857	0.019	0.607	3.803	6.932	18.445
	F5	108	15.020	24.150	0.210	1.230	5.260	9.610	80.320
Pb in Water (mg/L)		108	0.031	0.028	0.000	0.000	0.031	0.052	0.122

(e) Zinc (Zn)

The dataset utilised for model development for Zn is shown in Table 8.13. The optimised hyperparameters for the Zn speciation RF model are 'max_depth': 8, 'max_features': auto, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 310. The random forest algorithm was used to develop this model, which considered 310 (n estimators) different decision trees, and the best decision tree was estimated using the R^2 value. Fig. 8.43 shows one of the random forest's

decision trees (max_depth=3). The training and testing results for the Zn speciation RF model are shown in Fig. 8.44, and the fractions of metals in benthic sediments affecting the concentration of Zn in the water column are shown in Fig. 8.45. The contribution order of the different fractions of Zn in benthic sediments is F1, F2, F4, and F3. F1 has the highest 64% contribution, followed by F2 with 16%, F4 with 14% and F3 with 5% to the concentration of Zn in water.

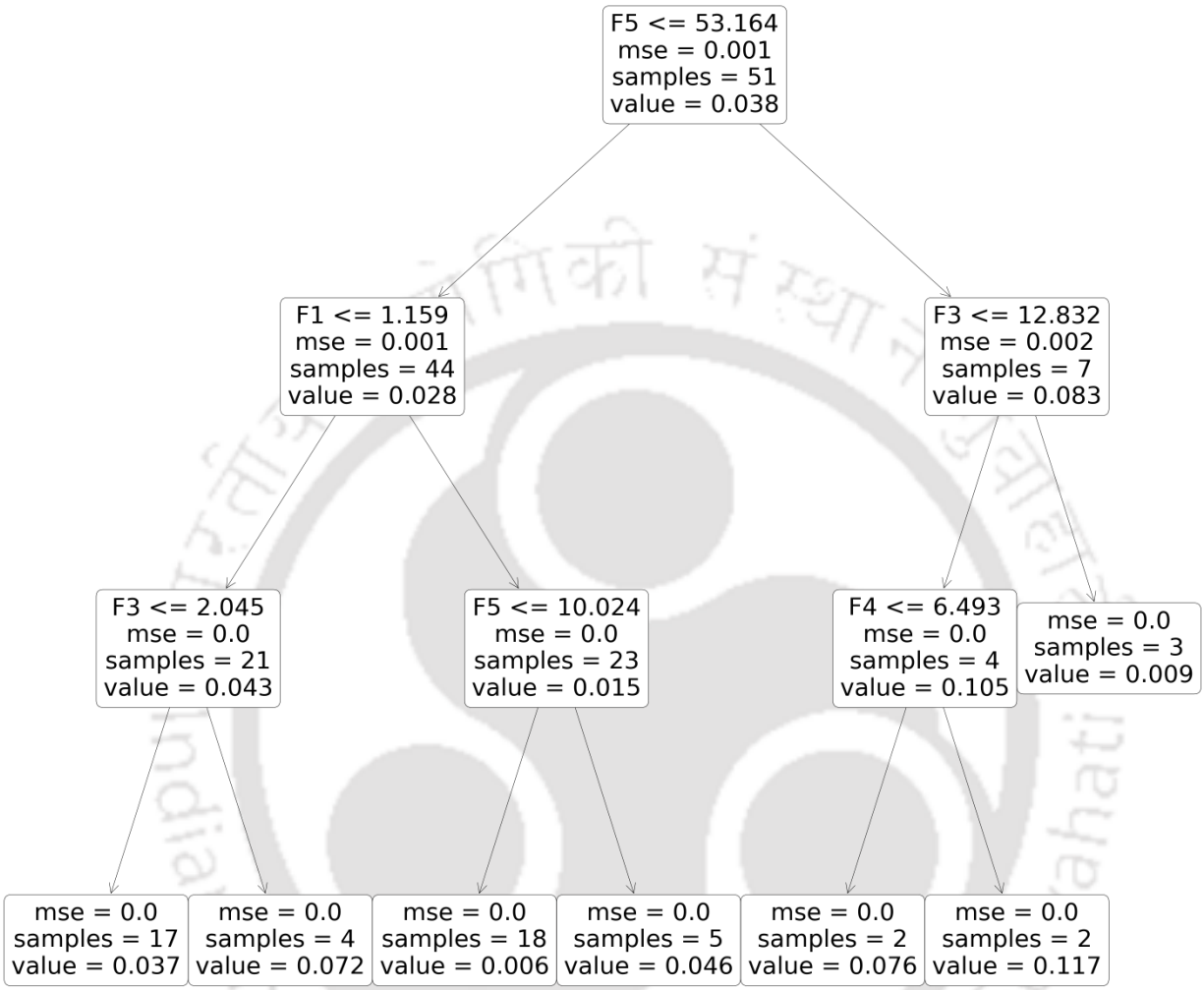


Fig. 8.40. A decision tree for the Pb speciation RF model

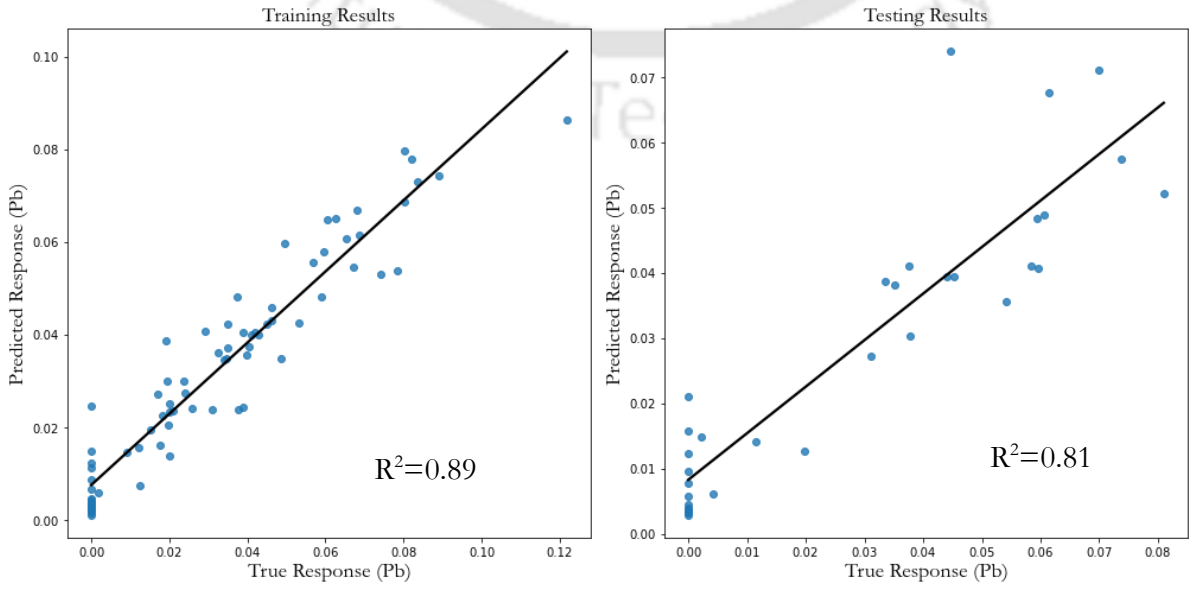


Fig. 8.41. Training and testing R^2 for Pb speciation RF model

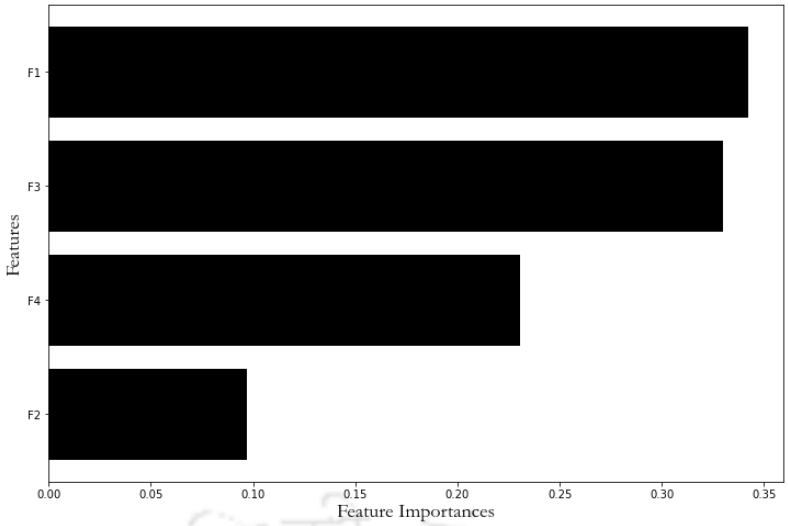


Fig. 8.42. Feature importance of Pb speciation RF model

Table 8.13. Dataset for model development of Zn

Variable		N	Mean	Std Dev	Min	Q1	Median	Q3	Max
Zn in sediments (mg/kg)	F1	107	11.960	17.020	0.000	1.840	5.190	14.670	88.980
	F2	107	26.700	56.260	0.950	4.520	7.410	19.960	247.160
	F3	107	19.310	20.220	3.400	6.540	8.220	27.420	108.350
	F4	107	14.420	14.170	1.030	3.090	11.020	19.120	66.400
	F5	107	44.310	88.980	1.600	4.860	10.280	45.220	489.800
Zn in Water (mg/L)		107	0.186	0.141	0.004	0.059	0.151	0.275	0.656

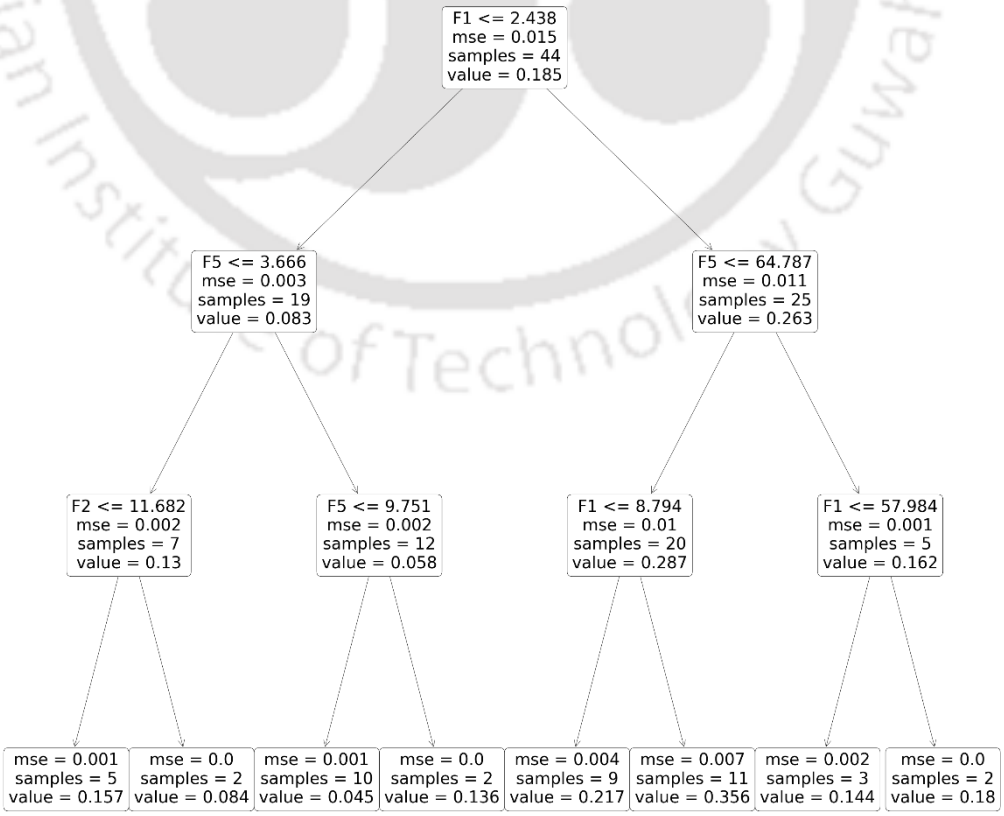


Fig. 8.43. A decision tree (max-depth=3) for Zn speciation RF model

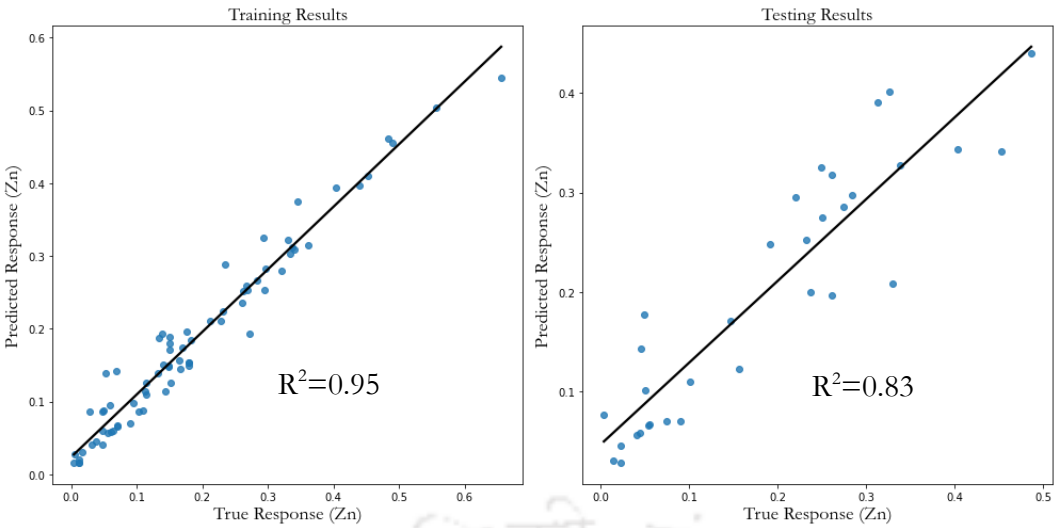


Fig. 8.44. Training and testing R^2 for Zn speciation RF model

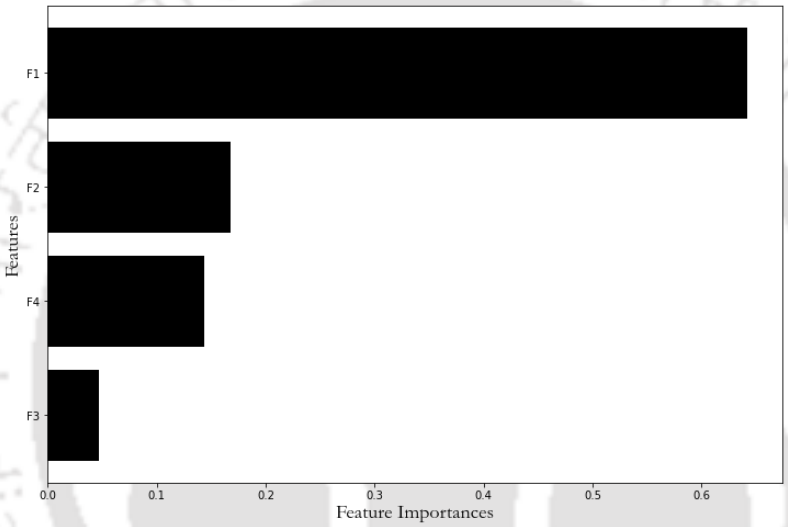


Fig. 8.45. Feature importance of Zn speciation RF model

8.2.1 Comparison of Different RF Models

The performance of models can only be judged based on their performance in predicting unknown datasets. For this purpose, the dataset of this model (108 observations) is randomly divided into a training set (81 observations) and a test set (27 observations). The best performing model for predicting metal concentration in the water column from metal fractions in benthic sediments are Cd (R^2 =0.87) and Mn (R^2 =0.86) models, and the worst performing model is Fe (R^2 =0.79) model (Table 8.14). Thus, RF was able to predict metal concentrations with sufficient accuracy.

Table 8.14. Performance evaluation (R^2) of RF models

Heavy Metals	Training Set	Test Set
Cd	0.88	0.87
Fe	0.87	0.79
Mn	0.96	0.86
Pb	0.89	0.81
Zn	0.95	0.83

8.3 Summary

In this chapter, two regression models were developed for the Kolong river. One was for depicting the relationship between physicochemical parameters and heavy metals using MLR, ANN and RF; the other was for depicting the relationship between heavy metals in the water column and heavy metal fractions in the sediment column. A few critical conclusions can be made, which are as follows:

a. RF performed better on the test and training set than the MLR and ANN models for the first regression model. This model found that heavy metal concentrations in the surface water were affected by the presence of Mg^{+2}, Ca^{+2}, Na^{+}, Cl^{-}, F^{-}, NO_3^{-}, hardness, alkalinity, turbidity and BOD_5. The surface water remediation strategies can now be formulated with the help of this regression model.

b. The second regression model was developed only using RF because of its performance in the first regression model. The RF technique performed best in the case of Cd and Mn and worst in the case of Fe. Metal speciation fractions F1 for Cd, F2 for Cu, F3 for Fe, F4 for Mn, F1 for Pb, and F1 for Zn made the most considerable contributions to the concentrations of heavy metals in the water column from benthic sediments. Therefore, this model has been used to estimate the contribution of benthic sediments metal fractions to heavy metal pollution in the surface water of the Kolong River.

Chapter 9

OVERALL CONCLUSIONS AND FUTURE SCOPE OF RESEARCH

9.1 Overall Conclusions

A comprehensive assessment was carried out to characterise the pollution load in surface water and benthic sediments of Kolong River to understand their risk and relationship using multivariate statistical techniques (MSTs) and artificial intelligence (AI). Important conclusions are summarised below:

1. The assessment of surface water pollution load revealed that DO, EC, BOD_5, Fe, Mn, Cd and Pb values exceeded the permissible limit during the sampling period. Although a few physicochemical were present in higher concentrations, heavy metals, because of their persistent or non-degradable nature, present a higher risk to the consumer of the surface water of Kolong river. The quality of surface water was expressed in terms of EWQI and FWQI, which identified few of the probable non-point pollution sources in the river. The FWQI results are more stringent than EWQI as they are more sensitive to each parameter variation.
2. The assessment of benthic sediment pollution load in terms of heavy metals was carried out using two approaches; total metal concentration and metal speciation fractions. Speciation based approaches were better at quantifying the risk or hazard posed by heavy metals, as the indices developed using this approach gave consistent results and gel well with the variation of the dataset.
3. Environmetric tools like hierarchical cluster analysis, discriminant analysis, principal component analysis, and positive matrix factorization model were employed to give insights into the sources of pollution. Hierarchical cluster analysis (HCA) clusters the 18 sampling sites into different clusters in different seasons. HCA clustered the sampling sites into three clusters in winter, pre-monsoon and monsoon season and four clusters in post-monsoon season. Sites in the middle stretch, i.e. from 9 to 16, were polluted due to anthropogenic metal pollution or runoff, and the rest were polluted due to domestic discharge or geogenic metal pollution. From DA, only 10 (turbidity, TDS, EC, hardness, alkalinity, Na^{+}, Ca^{+2}, Mg^{+2}, F., SO_4^{2-} and Cd) out of 21 parameters were responsible for the temporal variability

of the water quality dataset of the studied area. From PCA, Pearson correlation, and PMF, pollution in the months of winter and post-monsoon were attributed to domestic discharge and metal pollution. Whereas, another pollution gets added during pre-monsoon and monsoon, bringing agricultural or surface runoff pollution into the river. These findings are also inline with the benthic sediment quality study.

4. Human health risk assessment was conducted to evaluate the carcinogenic and non-carcinogenic risks posed by prolonged exposure to heavy metals through ingestion and dermal contact with surface water and benthic sediments of Kolong River. The dermal contact of exposure for both water column (surface water) and benthic sediments presents a considerable risk for the human population than the ingestion pathway of exposure. The non-carcinogenic health risk assessment revealed Zn (for both surface water and sediments) and Cd (only in the case of surface water) posing a considerable hazard for the human population, and the carcinogenic health risk assessment revealed Pb (for surface water) and Mn (for sediments) to be a considerable hazard for the human population. In the study area, children were at higher risk than adults for carcinogenic and non-carcinogenic risks.
5. The application of artificial intelligence (AI) and multiple linear regression (MLR) to develop two models for the prediction of heavy metals in the surface water of the Kolong River was carried out. One model was developed using physicochemical parameters of surface water, and the second model was developed using the metal speciation fractions to identify the factors affecting the metal concentrations in surface water. Random forest (RF) technique performed better on the test and training set than the multiple linear regression (MLR) and artificial neural network (ANN) based models. The first regression model found that heavy metal concentrations in the surface water were affected by the presence of Mg^{+2}, Ca^{+2}, Na^{+}, K^{+}, Cl^{-}, F^{-}, NO_3^{-}, hardness, alkalinity, turbidity and BOD_5. The second regression model found the F1 fraction for Cd, F2 fraction for Cu, F3 fraction for Fe, F4 fraction for Mn, F1 fraction for Pb, and F1 fraction for Zn as influencing factors. Therefore, river remediation strategies, particularly to curb heavy metal pollution, can now also be formulated with the help of the findings of these regression models.

The formulation and application of the current study can thus be of great assistance to various agencies responsible for water supply and pollution control, as these form a powerful tool for easy understanding, making their applicability uncomplicated. Indeed, these methodologies make the use of water quality datasets extremely simple and straightforward. This would help in properly managing the funds allotted for restoring and revitalising the water bodies. This is due to the fact that it will assist in identifying the critical pollution parameters that are essential in the process of restoring the water bodies, which will help in the process of updating the regulatory norms for future policies.

9.2 Future Scope of Research

- The present study on a river can be made applicable for other various other sources of water, such as groundwater, lakes, reservoirs, and estuaries, which can give a clearer picture of the water quality of a source in any geographical location.
- Indices developed in this study and till date has proved to be highly effective; still, newer methods such as artificial neural networks, artificial neuro-fuzzy interference system, and geographical interference system may be tried and tested to improve the interpretability of the indices. Also, water quality indices can be developed for many other purposes, such as irrigation, industrial, commercial etc., which can utilise their applicability. Furthermore, research on improving the proposed models are also highly welcome.
- One promising area of research would be the optimisation of water quality monitoring stations.
- Another area would be the fate of pollutants using information entropy and other methods.
- In the case of MLR, the ordinary least squares (OLS) technique has been used to estimate the regression equations; advanced techniques such as the generalised least squares (GLS) technique can also be used to enhance the accuracy of the predicted equations.
- Spatial modelling to model the pollution dissemination in a river could be another study area.
- As industries and factories are build up in areas near rivers, the environmental carrying capacity of rivers can also be investigated or evaluated.

REFERENCES

Abbasi, T., Abbasi, S., 2012. Water Quality Indices. Elsevier. https://doi.org/10.1016/C2010-0-69472-7

Abdallah, M.A.M., 2021. Chemo speciation of metals in surficial sediments of Nile delta coastal zone ecosystem, Egypt. Mar. Syst. Ocean Technol. 16, 43–53. https://doi.org/10.1007/s40868-021-00093-9

Abdallaoui, A., Badaoui, H. El, 2015. Comparative study of two stochastic models using the physicochemical characteristics of river sediment to predict the concentration of toxic metals. J. Mater. Environ. Sci. 6, 445–454.

Abdi, H., Williams, L.J., 2010. Principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. https://doi.org/10.1002/wics.101

Abtahi, M., Golchinpour, N., Yaghmaeian, K., Rafiee, M., Jahangiri-Rad, M., Keyani, A., Saeedi, R., 2015. A modified drinking water quality index (DWQI) for assessing drinking source water quality in rural communities of Khuzestan Province, Iran. Ecol. Indic. 53, 283–291. https://doi.org/10.1016/j.ecolind.2015.02.009

ACHHRA, 2017. Environmental Health Risk Assessment. Guidelines for assessing human health risks from environmental hazards.

Adimalla, N., Chen, J., Qian, H., 2020. Spatial characteristics of heavy metal contamination and potential human health risk assessment of urban soils: A case study from an urban region of South India. Ecotoxicol. Environ. Saf. 194. https://doi.org/10.1016/j.ecoenv.2020.110406

Aguilar-Hinojosa, Y., Meza-Figueroa, D., Villalba-Atondo, A.I., Encinas-Romero, M.A., Valenzuela-García, J.L., Gómez-Álvarez, A., 2016. Mobility and bioavailability of metals in stream sediments impacted by mining activities: the Jaralito and the Mexicana in Sonora, Mexico. Water. Air. Soil Pollut. 227, 51–124. https://doi.org/10.1007/s11270-016-3046-1

Ahmed, F., Fakhruddin, A.N.M., Imam, M.D.T., Khan, N., Khan, T.A., Rahman, M.M., Abdullah, A.T.M., 2016. Spatial distribution and source identification of heavy metal pollution in roadside surface soil: a study of Dhaka Aricha highway, Bangladesh. Ecol. Process. 5. https://doi.org/10.1186/s13717-016-0045-5

Ahmed, I.A.M., Hamilton-Taylor, J., Lofts, S., Meeussen, J.C.L., Lin, C., Zhang, H., Davison, W., 2013. Testing copper-speciation predictions in freshwaters over a wide range of metal-organic matter ratios. Environ. Sci. Technol. 47, 1487–1495. https://doi.org/10.1021/es304150n

Ahmed, N., Bodrud-Doza, M., Islam, S.M.D.U., Choudhry, M.A., Muhib, M.I., Zahid, A., Hossain, S., Moniruzzaman, M., Deb, N., Bhuiyan, M.A.Q., 2019. Hydrogeochemical evaluation and statistical analysis of groundwater of Sylhet, north-eastern Bangladesh. Acta Geochim. 38, 440–455. https://doi.org/10.1007/s11631-018-0303-6

Akcay, H., Oguz, A., Karapire, C., 2003. Study of heavy metal pollution and speciation in Buyak Menderes and Gediz river sediments. Water Res. 37, 813–822. https://doi.org/10.1016/S0043-1354(02)00392-5

Akindele, E.O., Olutona, G.O., 2014. Water physicochemistry and zooplankton fauna of Aiba Reservoir headwater streams, Iwo, Nigeria. J. Ecosyst. 2014, 1–11. https://doi.org/10.1155/2014/105405

Akindele, E.O., Omisakin, O.D., Oni, O.A., Aliu, O.O., Omoniyi, G.E., Akinpelu, O.T., 2020. Heavy metal toxicity in the water column and benthic sediments of a degraded tropical stream. Ecotoxicol. Environ. Saf. 190. https://doi.org/10.1016/j.ecoenv.2019.110153

Al-Janabi, Z., Zaki, S., AlHassany, J., Al-Obaidy, A.H., Awad, E., Maktoof, A., 2019. Geochemical Evaluation of Heavy Metals (Cd, Cr, Fe, and Mn) in Sediment of Shatt Al-

Basrah, Iraq. Eng. Technol. J. 37, 237–241. https://doi.org/10.30684/etj.37.2c.6

Al-Mur, B.A., 2020. Geochemical fractionation of heavy metals in sediments of the Red Sea, Saudi Arabia. Oceanologia 62, 31–44. https://doi.org/10.1016/j.oceano.2019.07.001

Ali, H., Khan, E., 2018. Bioaccumulation of non-essential hazardous heavy metals and metalloids in freshwater fish. Risk to human health. Environ. Chem. Lett. https://doi.org/10.1007/s10311-018-0734-7

Alkarkhi, A.F.M., Ahmad, A., Easa, A.M., 2009. Assessment of surface water quality of selected estuaries of Malaysia: Multivariate statistical techniques. Environmentalist 29, 255–262. https://doi.org/10.1007/s10669-008-9190-4

Altman, N., Krzywinski, M., 2017. Points of significance: clustering. Nat. Methods 14, 545–546. https://doi.org/10.1038/NMETH.4299

Altman, N., Krzywinski, M., 2015a. Points of significance: association, correlation and causation. Nat. Methods 12, 899–900. https://doi.org/10.1038/nmeth.3587

Altman, N., Krzywinski, M., 2015b. Points of significance: simple linear regression. Nat. Methods 12, 999–1000. https://doi.org/10.1038/nmeth.3627

Alves, D.D., Riegel, R.P., de Quevedo, D.M., Osório, D.M.M., da Costa, G.M., do Nascimento, C.A., Telöken, F., 2018. Seasonal assessment and apportionment of surface water pollution using multivariate statistical methods: Sinos River, southern Brazil. Environ. Monit. Assess. 190, 1–12. https://doi.org/10.1007/s10661-018-6759-3

Amiri, V., Rezaei, M., Sohrabi, N., 2014. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. Environ. Earth Sci. 72, 3479–3490. https://doi.org/10.1007/s12665-014-3255-0

Amor, R. Ben, Yahyaoui, A., Abidi, M., Chouba, L., Gueddari, M., 2019. Bioavailability and assessment of metal contamination in surface sediments of Rades-Hamam Lif Coast, around Meliane River (Gulf of Tunis, Tunisia, Mediterranean Sea). J. Chem. 2019. https://doi.org/10.1155/2019/4284987

Anandkumar, A., Nagarajan, R., Sellappa Gounder, E., Prabakaran, K., 2022. Seasonal variation and mobility of trace metals in the beach sediments of NW Borneo. Chemosphere 287, 132069. https://doi.org/10.1016/j.chemosphere.2021.132069

Anon, 2019. BP RISC: human health risk assessment and biodegradation software [WWW Document]. GroundwaterSoftware.com. URL https://www.groundwatersoftware.com/risc.htm (accessed 9.30.21).

Anttila, P., Paatero, P., Tapper, U., Järvinen, O., 1995. Source identification of bulk wet deposition in Finland by positive matrix factorization. Atmos. Environ. 29, 1705–1718. https://doi.org/10.1016/1352-2310(94)00367-T

APHA, 2012. Standard Methods for the Examination of Water and Wastewater, 22nd editi. ed, American Public Health Association. American Public Health Association (APHA), American Water Works Association (AWWA) and Water Environment Federation (WEF), Washington, D.C., USA. https://doi.org/10.2105/AJPH.51.6.940-a

Arisekar, U., Shakila, R.J., Shalini, R., Jeyasekaran, G., 2020. Human health risk assessment of heavy metals in aquatic sediments and freshwater fish caught from Thamirabarani River, the Western Ghats of South Tamil Nadu. Mar. Pollut. Bull. 159, 111496. https://doi.org/10.1016/j.marpolbul.2020.111496

Arjonilla, M., Forja, J.M., Gómez-Parra, A., 1994. Sediment analysis does not provide a good measure of heavy metal bioavailability to Cerastoderma glaucum (Mollusca: Bivalvia) in confined coastal ecosystems. Bull. Environ. Contam. Toxicol. 52, 810–817. https://doi.org/10.1007/BF00200688

Armagan, B., Gok, N., Ucar, D., 2008. Assessment of seasonal variations in surface water

quality of the Balikligol lakes, Sanliurfa, Turkey. Fresenius Environ. Bull. 17, 79–85.

Auret, L., Aldrich, C., 2012. Interpretation of nonlinear relationships between process variables by use of random forests. Miner. Eng. 35, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008

Ayyanar, A., Thatikonda, S., 2020. Distribution and ecological risks of heavy metals in Lake Hussain Sagar, India. Acta Geochim. 39, 255–270. https://doi.org/10.1007/s11631-019-00360-y

Azhar, S.C., Aris, A.Z., Yusoff, M.K., Ramli, M.F., Juahir, H., 2015. Classification of River Water Quality Using Multivariate Analysis. Procedia Environ. Sci. 30, 79–84. https://doi.org/10.1016/j.proenv.2015.10.014

Barrio-Parra, F., Izquierdo-Díaz, M., Dominguez-Castillo, A., Medina, R., De Miguel, E., 2019. Human-health probabilistic risk assessment: the role of exposure factors in an urban garden scenario. Landsc. Urban Plan. 185, 191–199. https://doi.org/10.1016/j.landurbplan.2019.02.005

Baruah, N.K., Kotoky, P., Bhattacharyya, K.G., Borah, G.C., 1996. Metal speciation in Jhanji River sediments. Sci. Total Environ. 193, 1–12. https://doi.org/10.1016/S0048-9697(96)05318-1

Başak, B., Alagha, O., 2010. Trace metals solubility in rainwater: evaluation of rainwater quality at a watershed area, Istanbul. Environ. Monit. Assess. 167, 493–503. https://doi.org/10.1007/s10661-009-1066-7

Basant, N., Gupta, S., Malik, A., Singh, K.P., 2010. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water - A case study. Chemom. Intell. Lab. Syst. 104, 172–180. https://doi.org/10.1016/j.chemolab.2010.08.005

Batayneh, A., Laboun, A., Qaisy, S., Ghrefat, H., Zumlot, T., Zaman, H., Elawadi, E., Mogren, S., Al-Qudah, K., 2012. Assessing groundwater quality of the shallow alluvial aquifer system in the Midyan Basin, northwestern Saudi Arabia. Arab Gulf J. Sci. Res. 30, 7–13.

Bazeli, J., Ghalehaskar, S., Morovati, M., Soleimani, H., Masoumi, S., Rahmani Sani, A., Saghi, M.H., Rastegar, A., 2020. Health risk assessment techniques to evaluate non-carcinogenic human health risk due to fluoride, nitrite and nitrate using Monte Carlo simulation and sensitivity analysis in Groundwater of Khaf County, Iran. Int. J. Environ. Anal. Chem. https://doi.org/10.1080/03067319.2020.1743280

Beamonte, E., Bermúdez, J.D., Casino, A., Veres, E., 2005. A global stochastic index for water quality: the case of the river Turia in Spain. J. Agric. Biol. Environ. Stat. 10, 424–439. https://doi.org/10.1198/108571105X80987

Belkhiri, L., Mouni, L., Sheikhy Narany, T., Tiri, A., 2017. Evaluation of potential health risk of heavy metals in groundwater using the integration of indicator kriging and multivariate statistical methods. Groundw. Sustain. Dev. 4, 12–22. https://doi.org/10.1016/j.gsd.2016.10.003

Bellman, R.E., Zadeh, L.A., 1970. Decision making in a fuzzy environment. Manage. Sci. 17, B-141-B-164. https://doi.org/10.1287/mnsc.17.4.b141

Bermudez, G.M.A., Jasan, R., Plá, R., Pignata, M.L., 2011. Heavy metal and trace element concentrations in wheat grains: assessment of potential non-carcinogenic health hazard through their consumption. J. Hazard. Mater. 193, 264–271. https://doi.org/10.1016/j.jhazmat.2011.07.058

Berrow, M.L., 1986. Chemical methods for assessing bio-available metals in sludges and soils. Agric. Wastes 15, 314–316. https://doi.org/10.1016/0141-4607(86)90029-6

Bezdek, J.C., Pal, S.K., 1992. Fuzzy models for pattern recognition.

Bhagat, S.K., Tiyasha, T., Awadh, S.M., Tung, T.M., Jawad, A.H., Yaseen, Z.M., 2021. Prediction of sediment heavy metal at the Australian Bays using newly developed hybrid artificial intelligence models. Environ. Pollut. 268, 115663. https://doi.org/10.1016/j.envpol.2020.115663

Bhargava, D.S., Saxena, B.S., Dewakar, 1990. A study of the geopollutants in the Godavary river basin of India. Asian Environ. 12, 36–59.

Bhat, N.A., Ghosh, P., Ahmed, W., Naaz, F., Darshinee, A.P., 2022. Heavy metal contamination in soils and stream water in Tungabhadra basin, Karnataka: environmental and health risk assessment. Int. J. Environ. Sci. Technol. https://doi.org/10.1007/s13762-022-04040-y

Bhat, S.A., Meraj, G., Yaseen, S., Pandit, A.K., 2014. Statistical assessment of water quality parameters for pollution source identification in Sukhnag stream: an inflow stream of lake Wular (Ramsar Site), Kashmir Himalaya. J. Ecosyst. 2014, 1–18. https://doi.org/10.1155/2014/898054

Bhuiyan, M.A.H., Islam, M.A., Dampare, S.B., Parvez, L., Suzuki, S., 2010. Evaluation of hazardous metal pollution in irrigation and drinking water systems in the vicinity of a coal mine area of northwestern Bangladesh. J. Hazard. Mater. 179, 1065–1077. https://doi.org/10.1016/j.jhazmat.2010.03.114

Bilgin, A., 2015. An assessment of water quality in the Coruh Basin (Turkey) using multivariate statistical techniques. Environ. Monit. Assess. 187. https://doi.org/10.1007/s10661-015-4904-9

Bo, L., Wang, D., Li, T., Li, Y., Zhang, G., Wang, C., Zhang, S., 2015a. Accumulation and risk assessment of heavy metals in water, sediments, and aquatic organisms in rural rivers in the Taihu Lake region, China. Environ. Sci. Pollut. Res. 22, 6721–6731. https://doi.org/10.1007/s11356-014-3798-3

Bo, L., Wang, D., Zhang, G., Wang, C., 2015b. Heavy metal speciation in sediments and the associated ecological risks in rural rivers in Southern Jiangsu Province, China. Soil Sediment Contam. 24, 90–102. https://doi.org/10.1080/15320383.2014.914465

Bodek, I., Lyman, W.., Reehl, W.., 1988. Environmental inorganic chemistry : properties, processes, and estimation methods, New York : Pergamon.

Bora, M., Goswami, D.C., 2017. Channel morphology and hydraulic geometry of River Kolong, Nagaon district, Assam, India: a study from the standpoint of river restoration. Curr. Sci. 113, 743–751. https://doi.org/10.18520/cs/v113/i04/743-751

Bora, M., Goswami, D.C., 2015. A study on seasonal and temporal variation in physico-chemical and hydrological characteristics of River Kolong at Nagaon Town , Assam , India. Sch. Res. Libr. Arch. Appl. Sci. Res. 7, 110–117.

Bora, M., Goswami, D.C., 2014. Study for restoration using field survey and geoinformatics of the Kolong River, Assam, India. J. Environ. Res. Dev. 8, 997–1004.

Borah, K.K., Bhuyan, B., Sarma, H.P., 2009. Heavy metal contamination of groundwater in the tea garden belt of darrang district, Assam, India. E-Journal Chem. 6. https://doi.org/10.1155/2009/760953

Boszke, L., Kowalski, A., Astel, A., Barański, A., Gworek, B., Siepak, J., 2008. Mercury mobility and bioavailability in soil from contaminated area. Environ. Geol. 55, 1075–1087. https://doi.org/10.1007/s00254-007-1056-4

Bouguerne, A., Boudoukha, A., Benkhaled, A., Mebarkia, A.-H., 2017. Assessment of surface water quality of Ain Zada dam (Algeria) using multivariate statistical techniques. Int. J. River Basin Manag. 15, 133–143. https://doi.org/10.1080/15715124.2016.1215325

Bouragba, S., Komai, K., Nakayama, K., 2020. Empirical approach for modeling of partition coefficient on lead concentrations in riverine sediment. Int. J. Environ. Sci. Dev. 11, 352–357. https://doi.org/10.18178/IJESD.2020.11.7.1275

Boyacioglu, H., 2007. Development of a water quality index based on a European classification scheme. Water SA 33, 101–106. https://doi.org/10.4314/wsa.v33i1.47882

Braga, M.C.B., Birkett, J.W., Shaw, G., Lester, J.N., 2010. Modelling the long-term fate of mercury in a Lowland Tidal River. II. Calibration and comparison of two models with field data. Arch. Environ. Contam. Toxicol. 58, 383–393. https://doi.org/10.1007/s00244-009-9378-8

Brahma, N.K., Misra, A.K., 2014. Study of heavy metals content in water , water hyacinth and soil of Rupahi. Arch. Appl. Sci. Res. 6, 7–11.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brown, R.M., McClelland, N.I., Deininger, R.A., Tozer, R.G., 1970. A water quality index - do we dare? Water Sew. Work. 117, 339–343.

Bryce-Smith, D., 1989. Chromium in the natural and human environments. Endeavour 13, 45. https://doi.org/10.1016/0160-9327(89)90073-2

Burton, G.A., 2002. Sediment quality criteria in use around the world. Limnology 3, 65–75. https://doi.org/10.1007/s102010200008

Buszewski, B., Kowalkowski, T., 2006. A new model of heavy metal transport in the soil using nonlinear artificial neural networks. Environ. Eng. Sci. 23, 589–595. https://doi.org/10.1089/ees.2006.23.589

Bzdusek, P.A., Christensen, E.R., Lee, C.M., Pakdeesusuk, U., Freedman, D.L., 2006. PCB congeners and dechlorination in sediments of Lake Hartwell, South Carolina, determined from cores collected in 1987 and 1998. Environ. Sci. Technol. 40, 109–119. https://doi.org/10.1021/es050194o

Cai, S., Zhou, S., Cheng, J., Wang, Q., Dai, Y., 2021. Distribution, bioavailability and ecological risk of heavy metals in surface sediments from the wujiang river basin, southwest of china. Polish J. Environ. Stud. 30, 5479–5491. https://doi.org/10.15244/pjoes/136185

Cao, H., Chen, J., Zhang, J., Zhang, H., Qiao, L., Men, Y., 2010. Heavy metals in rice and garden vegetables and their potential health risks to inhabitants in the vicinity of an industrial zone in Jiangsu, China. J. Environ. Sci. 22, 1792–1799. https://doi.org/10.1016/S1001-0742(09)60321-1

Carr, C.J., 1994. American industrial health council: exposure factors sourcebook. Regul. Toxicol. Pharmacol. https://doi.org/10.1006/rtph.1994.1071

CCME, 2017. Canadian environmental quality guidelines.

CCME, 2001. Canadian sediment quality guidelines for the protection of aquatic life: summary tables. Updated., Canadian environmental quality guidelines. http://www.ccme.ca/assets/pdf/e1_06.pdf.

Celen, M., Oruc, H.N., Adiller, A., Yıldız Töre, G., Onkal Engin, G., 2022. Contribution for pollution sources and their assessment in urban and industrial sites of Ergene River Basin, Turkey. Int. J. Environ. Sci. Technol. 1–20. https://doi.org/10.1007/s13762-022-03919-0

Chanapathi, T., Thatikonda, S., 2019. Fuzzy based regional water quality index for surface water quality assessment. J. Hazardous, Toxic, Radioact. Waste 23, 04019010. https://doi.org/10.1061/(asce)hz.2153-5515.0000443

Chang, N. Bin, Chen, H.W., Ning, S.K., 2001. Identification of river water quality using the fuzzy synthetic evaluation approach. J. Environ. Manage. 63, 293–305. https://doi.org/10.1006/jema.2001.0483

Chapman, D., 2021. Water quality assessments - A guide to use of biota , sediments and water in environmental monitoring, Second Edi. ed.

Chau, K., 2006. A review on the integration of artificial intelligence into coastal modeling. J.

Environ. Manage. 80, 47–57. https://doi.org/10.1016/j.jenvman.2005.08.012

Chen, C.F., Dong, C. Di, Chen, C.W., 2013. Metal speciation and contamination in dredged harbor sediments from Kaohsiung harbor, Taiwan. Soil Sediment Contam. 22, 546–561. https://doi.org/10.1080/15320383.2013.750268

Chen, G., Wang, X., Wang, R., Liu, G., 2019. Health risk assessment of potentially harmful elements in subsidence water bodies using a Monte Carlo approach: An example from the Huainan coal mining area, China. Ecotoxicol. Environ. Saf. 171, 737–745. https://doi.org/10.1016/j.ecoenv.2018.12.101

Chen, J., Li, F., Fan, Z., Wang, Y., 2016. Integrated application of multivariate statistical methods to source apportionment ofwatercourses in the liao river basin, northeast China. Int. J. Environ. Res. Public Health 13, 1035. https://doi.org/10.3390/ijerph13101035

Chen, M., Ma, L.Q., Harris, W.G., 1999. Baseline concentrations of 15 trace elements in Florida surface soils. J. Environ. Qual. 28, 1173–1181. https://doi.org/10.2134/jeq1999.00472425002800040018x

Chen, R., Teng, Y., Chen, H., Hu, B., Yue, W., 2019. Groundwater pollution and risk assessment based on source apportionment in a typical cold agricultural region in Northeastern China. Sci. Total Environ. 696, 133972. https://doi.org/10.1016/j.scitotenv.2019.133972

Chen, Y., Song, L., Liu, Y., Yang, L., Li, D., 2020. A review of the artificial neural network models for water quality prediction. Appl. Sci. https://doi.org/10.3390/app10175776

Cheng, W.H., Yap, C.K., 2015. Potential human health risks from toxic metals via mangrove snail consumption and their ecological risk assessments in the habitat sediment from Peninsular Malaysia. Chemosphere 135, 156–165. https://doi.org/10.1016/j.chemosphere.2015.04.013

Cheng, W.L., Kuo, Y.C., Lin, P.L., Chang, K.H., Chen, Y.S., Lin, T.M., Huang, R., 2004. Revised air quality index derived from an entropy function. Atmos. Environ. 38, 383–391. https://doi.org/10.1016/j.atmosenv.2003.10.006

Chester, R., 2018. The concentration, mineralogy, and chemistry of total suspended matter in seawater, in: Pollutant Transfer and Transport in the Sea: Volume II. pp. 67–99. https://doi.org/10.1201/9781351075848

Cho, E., Arhonditsis, G.B., Khim, J., Chung, S., Heo, T.Y., 2016. Modeling metal-sediment interaction processes: parameter sensitivity assessment and uncertainty analysis. Environ. Model. Softw. 80, 159–174. https://doi.org/10.1016/j.envsoft.2016.02.026

Chou, C.L., Haya, K., Paon, L.A., Moffatt, J.D., 2004. A regression model using sediment chemistry for the evaluation of marine environmental impacts associated with salmon aquaculture cage wastes. Mar. Pollut. Bull. 49, 465–472. https://doi.org/10.1016/j.marpolbul.2004.02.039

Chowdhury, S., Mazumder, M.A.J., Al-Attas, O., Husain, T., 2016. Heavy metals in drinking water: occurrences, implications, and future needs in developing countries. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.06.166

Ciffroy, P., 2020. A comprehensive probabilistic approach for integrating and separating natural variability and parametric uncertainty in the prediction of distribution coefficient of radionuclides in rivers. J. Environ. Radioact. 225. https://doi.org/10.1016/j.jenvrad.2020.106371

Comero, S., Locoro, G., Free, G., Vaccaro, S., De Capitani, L., Gawlik, B.M., 2011. Characterisation of Alpine lake sediments using multivariate statistical techniques. Chemom. Intell. Lab. Syst. 107, 24–30. https://doi.org/10.1016/j.chemolab.2011.01.002

Concas, A., Ardau, C., Cristini, A., Zuddas, P., Cao, G., 2006. Mobility of heavy metals from tailings to stream waters in a mining activity contaminated site. Chemosphere 63, 244–253.

https://doi.org/10.1016/j.chemosphere.2005.08.024

CPCB, 2015. River stretches for restoration of water quality.

Cui, H., Sivakumar, B., Singh, V.P., 2018. Entropy applications in environmental and water engineering. Entropy. https://doi.org/10.3390/e20080598

Cui, J., Zang, S., Zhai, D., Wu, B., 2014. Potential ecological risk of heavy metals and metalloid in the sediments of Wuyuer River basin, Heilongjiang Province, China. Ecotoxicology 23, 589–600. https://doi.org/10.1007/s10646-014-1182-1

Dahiya, S., Singh, B., Gaur, S., Garg, V.K., Kushwaha, H.S., 2007. Analysis of groundwater quality using fuzzy synthetic evaluation. J. Hazard. Mater. 147, 938–946. https://doi.org/10.1016/j.jhazmat.2007.01.119

Dan, S.F., Udoh, E.C., Wang, Q., 2022. Contamination and ecological risk assessment of heavy metals, and relationship with organic matter sources in surface sediments of the Cross River Estuary and nearshore areas. J. Hazard. Mater. 438, 129531. https://doi.org/10.1016/J.JHAZMAT.2022.129531

Daşbaşi, T., Saçmaci, Ş., Çankaya, N., Soykan, C., 2016. A new synthesis, characterization and application chelating resin for determination of some trace metals in honey samples by FAAS. Food Chem. 203, 283–291. https://doi.org/10.1016/j.foodchem.2016.02.078

Dash, S., Borah, S.S., Kalamdhad, A.S., 2020. Application of environmetrics tools for geochemistry, water quality assessment and apportionment of pollution sources in Deepor beel, Assam, India. Water Pract. Technol. 15, 973–992. https://doi.org/10.2166/wpt.2020.078

de Almeida, R.C.M., Botero, W.G., de Oliveira, L.C., 2022. Natural and anthropogenic sources of potentially toxic elements to aquatic environment: a systematic literature review. Environ. Sci. Pollut. Res. 29, 51318–51338. https://doi.org/10.1007/s11356-022-20980-x

de Blois, C.J., Wind, H.G., de Kok, J.L., Koppeschaar, K., 2003. Robustness of river basin water quality models. J. Water Resour. Plan. Manag. 129, 189–199. https://doi.org/10.1061/(asce)0733-9496(2003)129:3(189)

de Campos Souza, P.V., 2020. Fuzzy neural networks and neuro-fuzzy networks: a review the main techniques and applications used in the literature. Appl. Soft Comput. J. 92. https://doi.org/10.1016/j.asoc.2020.106275

de Groot, A.J., 2018. Metals and sediments: a global perspective, in: Metal Contaminated Aquatic Sediments. Routledge, pp. 1–20. https://doi.org/10.1201/9780203747643-1

Dedić, A., Gerhardt, A., Kelly, M.G., Stanić-koštroman, S., Šiljeg, M., Stroil, B.K., Kamberović, J., Mateljak, Z., Pešić, V., Vučković, I., Snigirova, A., Bogatova, Y., Barinova, S., Radulović, S., Cvijanović, D., Lasić, A., Škobić, D., Sudar, A., 2020. Innovative methods and science and policy to fill knowledge gaps in approaches for WFD: ideas. Water Solut. 3, 30–42.

Dehghanian, N., Ghaedi, M., Ansari, A., Ghaedi, A., Vafaei, A., Asif, M., Agarwal, S., Tyagi, I., Gupta, V.K., 2016. A random forest approach for predicting the removal of Congo red from aqueous solutions by adsorption onto tin sulfide nanoparticles loaded on activated carbon. Desalin. Water Treat. 57, 9272–9285. https://doi.org/10.1080/19443994.2015.1027964

DeLaune, R.D., Jugsujinda, A., Gambrell, R.P., Miao, S., 2008. Metal concentrations and trace metal Al and Fe ratios in soil of the Chenier Plain, Southwest Louisiana coastal zone. J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng. 43, 300–312. https://doi.org/10.1080/10934520701792837

Delgado, A., Romero, I., 2016. Environmental conflict analysis using an integrated grey clustering and entropy-weight method: A case study of a mining project in Peru. Environ. Model. Softw. 77, 108–121. https://doi.org/10.1016/j.envsoft.2015.12.011

Deng, A., Ye, C., Liu, W., 2018. Spatial and seasonal patterns of nutrients and heavy metals in twenty-seven rivers draining into the South China Sea. Water (Switzerland) 10. https://doi.org/10.3390/w10010050

Dhanakumar, S., Murthy, K.R., Solaraj, G., Mohanraj, R., 2013. Heavy-metal fractionation in surface sediments of the Cauvery River estuarine region, southeastern coast of India. Arch. Environ. Contam. Toxicol. 65, 14–23. https://doi.org/10.1007/s00244-013-9886-4

Dinius, S.H., 1987. Design of an index of water quality. J. Am. Water Resour. Assoc. 23, 833–843.

Dinius, S.H., 1972. Social accounting system for evaluating water resources. Water Resour. Res. 8, 1159–1177. https://doi.org/10.1029/WR008i005p01159

Dixit, A., Siddaiah, N.S., 2021. Health and ecological risk assessment of metals in surface water from urban wetlands of Gurugram, India. Int. J. Environ. Anal. Chem. https://doi.org/10.1080/03067319.2021.1974012

Dixon, W., Chiswell, B., 1996. Review of aquatic monitoring program design. Water Res. https://doi.org/10.1016/0043-1354(96)00087-5

Djadé, P.J.O., Keuméan, K.N., Traoré, A., Soro, G., Soro, N., 2021. Assessment of health risks related to contamination of groundwater by trace metal elements (Hg, Pb, Cd, As and Fe) in the Department of Zouan-Hounien (West Côte D'Ivoire). J. Geosci. Environ. Prot. 09, 189–210. https://doi.org/10.4236/gep.2021.98013

Dooyema, C.A., Neri, A., Lo, Y.C., Durant, J., Dargan, P.I., Swarthout, T., Biya, O., Gidado, S.O., Haladu, S., Sani-Gwarzo, N., Nguku, P.M., Akpan, H., Idris, S., Bashir, A.M., Brown, M.J., 2012. Outbreak of fatal childhood lead poisoning related to artisanal gold mining in northwestern Nigeria, 2010. Environ. Health Perspect. 120, 601–607. https://doi.org/10.1289/ehp.1103965

Duodu, G.O., Goonetilleke, A., Ayoko, G.A., 2016. Comparison of pollution indices for the assessment of heavy metal in Brisbane River sediment. Environ. Pollut. 219, 1077–1091. https://doi.org/10.1016/j.envpol.2016.09.008

Dwivedi, S., Tiwari, I.C., Bhargava, D.S., 1997. Water quality of the river Ganga at Varanasi. J. Inst. Eng. Environ. Eng. Div. 78, 1–4.

Eberhardt, L.L., 1967. Restoring the quality of our environment. J. Wildl. Manage. 31, 846. https://doi.org/10.2307/3797998

Edokpayi, J.N., Odiyo, J.O., Durowoju, O.S., 2017. Impact of wastewater on surface water quality in developing countries: a Case study of South Africa, in: Water Quality. https://doi.org/10.5772/66561

Ehteram, M., Ahmed, A.N., Latif, S.D., Huang, Y.F., Alizamir, M., Kisi, O., Mert, C., El-Shafie, A., 2021. Design of a hybrid ANN multi-objective whale algorithm for suspended sediment load prediction. Environ. Sci. Pollut. Res. 28, 1596–1611. https://doi.org/10.1007/s11356-020-10421-y

Ekwutosi, O.K., Chudi, O.P.-A., Charity, E.N., 2020. Chemical speciation and potential mobility of heavy metals in the soils of Onitsha South Local Government Area Anambra Nigeria. Am. J. Appl. Chem. 8, 74. https://doi.org/10.11648/j.ajac.20200803.12

El-Ouaty, O., El-M'rini, A., Nachite, D., Marrocchino, E., Marin, E., Rodella, I., 2022. Assessment of the heavy metal sources and concentrations in the Nador Lagoon sediment, Northeast-Morocco. Ocean Coast. Manag. 216, 105900. https://doi.org/10.1016/j.ocecoaman.2021.105900

Elias, Hamzah, M.S., Rahman, S.A., Siong, W.B., Salim, N.A.A., 2012. Assessment of sediment quality collected from Tunku Abdul Rahman Park, Sabah. J. Sains Nukl. Malaysia 24, 59–70.

Emamgholizadeh, S., Kashi, H., Marofpoor, I., Zalaghi, E., 2014. Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. Int. J. Environ. Sci. Technol. 11, 645–656. https://doi.org/10.1007/s13762-013-0378-x

EPA, 1996. Method 3050B: Acid digestion of sediments, sludges, and soils. Enivromental Prot. Agency.

Ewaid, S.H., Abed, S.A., 2017. Water quality index for Al-Gharraf River, southern Iraq. Egypt. J. Aquat. Res. 43, 117–122. https://doi.org/10.1016/j.ejar.2017.03.001

Ewaid, S.H., Abed, S.A., Al-Ansari, N., Salih, R.M., 2020. Development and evaluation of a water quality index for the Iraqi rivers. Hydrology 7. https://doi.org/10.3390/HYDROLOGY7030067

Fagbote, E.O., Olanipekun, E.O., Uyi, H.S., 2014. Water quality index of the ground water of bitumen deposit impacted farm settlements using entropy weighted method. Int. J. Environ. Sci. Technol. 11, 38–127. https://doi.org/10.1007/s13762-012-0149-0

Fall, A., Fall, J., 2001. A model to calculate heavy metal load to lakes dominated by urban runoff and diffuse inflow. Ecol. Modell. 137, 1–21. https://doi.org/10.1016/S0304-3800(00)00370-7

Fan, X., Cui, B., Zhao, H., Zhang, Z., Zhang, H., 2010. Assessment of river water quality in Pearl River Delta using multivariate statistical techniques, in: Procedia Environmental Sciences. Elsevier, pp. 1220–1234. https://doi.org/10.1016/j.proenv.2010.10.133

Farahat, E.A., Mahmoud, W.F., Awad, H.E.A., Farrag, H.F., Arshad, M., Eid, E.M., Fahmy, G.M., 2021. Prediction models for evaluating the uptake of heavy metals by the invasive grass vossia cuspidata (Roxb.) griff. in the river nile, egypt: A biomonitoring approach. Sustain. 13, 10558. https://doi.org/10.3390/su131910558

Fernandes, L.L., Kessarkar, P.M., Rao, V.P., Suja, S., Parthiban, G., Kurian, S., 2019. Seasonal distribution of trace metals in suspended particulate and bottom sediments of four microtidal river estuaries, west coast of India. Hydrol. Sci. J. 64, 1519–1534. https://doi.org/10.1080/02626667.2019.1655147

Fernandes, M.C., Nayak, G.N., 2015. Speciation of metals and their distribution in tropical estuarine mudflat sediments, southwest coast of India. Ecotoxicol. Environ. Saf. 122, 68–75. https://doi.org/10.1016/j.ecoenv.2015.07.016

Fernández-Luqueño, F., López-Valdez, F., Gamero-Melo, P., Luna-Suárez, S., Aguilera-González, E., Martínez, A., García-Guillermo, M., Hernández-Martínez, G., Herrera-Mendoza, R., Álvarez-Garza, M., Pérez-Velázquez, I., 2013. Heavy metal pollution in drinking water - a global risk for human health: A review. African J. Environ. Sci. Technol. 7, 567–584. https://doi.org/10.5897/AJEST12.197

Fleming, S.W., 2007. An information theoretic perspective on mesoscale seasonal variations in ground-level ozone. Atmos. Environ. 41, 5746–5755. https://doi.org/10.1016/j.atmosenv.2007.02.027

Förstner, U., 1993. Metal speciation-general concepts and applications. Int. J. Environ. Anal. Chem. 51, 5–23. https://doi.org/10.1080/03067319308027608

Förstner, U., 1982. Accumulative phases for heavy metals in limnic sediments. Hydrobiologia 91–92, 269–284. https://doi.org/10.1007/bf00940118

Förstner, U., Salomons, W., 1980a. Trace metal analysis on polluted sediments. Environ. Technol. Lett. 1, 494. https://doi.org/10.1080/09593338009384006

Förstner, U., Salomons, W., 1980b. Trace metal analysis on polluted sediments: Part I: Assessment of sources and intensities. Environ. Technol. Lett. 1, 494–505. https://doi.org/10.1080/09593338009384006

Fulton, B.A., Meyer, J.S., 2014. Development of a regression model to predict copper toxicity to

Daphnia magna and site-specific copper criteria across multiple surface-water drainages in an arid landscape. Environ. Toxicol. Chem. 33, 1865–1873. https://doi.org/10.1002/etc.2631

Gad, M., El-Safa, M.M.A., Farouk, M., Hussein, H., Alnemari, A.M., Elsayed, S., Khalifa, M.M., Moghanm, F.S., Eid, E.M., Saleh, A.H., 2021. Integration of water quality indices and multivariate modeling for assessing surface water quality in qaroun lake, Egypt. Water (Switzerland) 13, 2258. https://doi.org/10.3390/w13162258

Gadde, R.R., Laitinen, H.A., 1974. Studies of heavy metal adsorption by Hydrous Iron and Manganese Oxides. Anal. Chem. 46, 2022–2026. https://doi.org/10.1021/ac60349a004

Gadh, R., Tandon, S.N., Mathur, R.P., Singh, O. V., 1993. Speciation of metals in Yamuna river sediments. Sci. Total Environ. 136, 229–242. https://doi.org/10.1016/0048-9697(93)90311-S

Gadkar, N.S., Nayak, G.N., Nasnodkar, M.R., 2019. Assessment of metal enrichment and bioavailability in mangrove and mudflat sediments of the tropical (Zuari) estuary, west coast of India. Environ. Sci. Pollut. Res. 26, 24998–25011. https://doi.org/10.1007/s11356-019-05733-7

Gandhimathi, A., Meenambal, T., 2012. Spatial prediction of heavy metal pollution for soils in Coimbatore, India based on ANN and kriging model. Eur. J. Sci. Res. 79, 198–207.

Gao, X., Chen, C.T.A., 2012. Heavy metal pollution status in surface sediments of the coastal Bohai Bay. Water Res. 46, 1901–1911. https://doi.org/10.1016/j.watres.2012.01.007

Gardiner, J., 1974. The chemistry of cadmium in natural water-II. The adsorption of cadmium on river muds and naturally occurring solids. Water Res. 8, 157–164. https://doi.org/10.1016/0043-1354(74)90038-4

Garneau, C., Sauvage, S., Sánchez-Pérez, J.M., Lofts, S., Brito, D., Neves, R., Probst, A., 2017. Modelling trace metal transfer in large rivers under dynamic hydrology: A coupled hydrodynamic and chemical equilibrium model. Environ. Model. Softw. 89, 77–96. https://doi.org/10.1016/j.envsoft.2016.11.018

Gharibi, H., Mahvi, A.H., Nabizadeh, R., Arabalibeik, H., Yunesian, M., Sowlat, M.H., 2012a. A novel approach in water quality assessment based on fuzzy logic. J. Environ. Manage. 112, 87–95.

Gharibi, H., Sowlat, M.H., Mahvi, A.H., Mahmoudzadeh, H., Arabalibeik, H., Keshavarz, M., Karimzadeh, N., Hassani, G., 2012b. Development of a dairy cattle drinking water quality index (DCWQI) based on fuzzy inference systems. Ecol. Indic. 20, 228–237. https://doi.org/10.1016/j.ecolind.2012.02.015

Gholizadeh, M.H., Melesse, A.M., Reddi, L., 2016. Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. Sci. Total Environ. 566–567, 1552–1567. https://doi.org/10.1016/J.SCITOTENV.2016.06.046

Ghosh, S., Mujumdar, P.P., 2006. Risk minimization in water quality control problems of a river system. Adv. Water Resour. 29, 458–470. https://doi.org/10.1016/j.advwatres.2005.06.001

Ghosh, S.P., Maiti, S.K., 2018. Evaluation of heavy metal contamination in roadside deposited sediments and road surface runoff: a case study. Environ. Earth Sci. 77. https://doi.org/10.1007/s12665-018-7370-1

Gibbs, R.J., 1973. Mechanisms of trace metal transport in rivers. Science (80-.). 180, 71–73. https://doi.org/10.1126/science.180.4081.71

Gitter, A., Mena, K.D., Wagner, K.L., Boellstorff, D.E., Borel, K.E., Gregory, L.F., Gentry, T.J., Karthikeyan, R., 2020. Human health risks associated with recreational waters: Preliminary approach of integrating quantitative microbial risk assessment with microbial source

tracking. Water (Switzerland) 12. https://doi.org/10.3390/w12020327

Goldblum, D.K., Rak, A., Ponnapalli, M.D., Clayton, C.J., 2006. The Fort Totten mercury pollution risk assessment: A case history. J. Hazard. Mater. 136, 406–417. https://doi.org/10.1016/j.jhazmat.2005.11.047

Golian, M., Katibeh, H., Singh, V.P., Ostad-Ali-Askari, K., Rostami, H.T., 2020. Prediction of tunnelling impact on flow rates of adjacent extraction water wells. Q. J. Eng. Geol. Hydrogeol. 53, 236–251. https://doi.org/10.1144/qjegh2019-055

Golshan, M., Dastoorpour, M., Birgani, Y.T., 2020. Fuzzy environmental monitoring for the quality assessment: Detailed feasibility study for the Karun River basin, Iran. Groundw. Sustain. Dev. 10, 100324. https://doi.org/10.1016/j.gsd.2019.100324

Golterman, H., Stumm, W., Morgan, J.J., 1982. An introduction emphasizing chemical equilibria in natural waters. J. Ecol. 70, 924. https://doi.org/10.2307/2260132

Goonetilleke, A., Thomas, E., Ginn, S., Gilbert, D., 2005. Understanding the role of land use in urban stormwater quality management. J. Environ. Manage. 74, 31–42. https://doi.org/10.1016/j.jenvman.2004.08.006

Gorgoglione, A., Gioia, A., Iacobellis, V., 2019. A framework for assessing modeling performance and effects of rainfall-catchment-drainage characteristics on nutrient urban runoff in poorly gauged watersheds. Sustain. 11. https://doi.org/10.3390/su11184933

Guan, Q., Wang, F., Xu, C., Pan, N., Lin, J., Zhao, R., Yang, Y., Luo, H., 2018. Source apportionment of heavy metals in agricultural soil based on PMF: A case study in Hexi Corridor, northwest China. Chemosphere 193, 189–197. https://doi.org/10.1016/j.chemosphere.2017.10.151

Gujre, N., Rangan, L., Mitra, S., 2021. Occurrence, geochemical fraction, ecological and health risk assessment of cadmium, copper and nickel in soils contaminated with municipal solid wastes. Chemosphere 271. https://doi.org/10.1016/j.chemosphere.2021.129573

Gupta, I., Dhage, S., Kumar, R., 2009. Study of variations in water quality of Mumbai coast through multivariate analysis techniques. Indian J. Mar. Sci. 38, 170–177.

Gupta, N., Khan, D.K., Santra, S.C., 2008. An assessment of heavy metal contamination in vegetables grown in wastewater-irrigated areas of Titagarh, West Bengal, India. Bull. Environ. Contam. Toxicol. 80, 115–118. https://doi.org/10.1007/s00128-007-9327-z

Gustafson, D.E., Kessel, W.C., 1978. Fuzzy clustering with a fuzzy covariance matrix, in: Proceedings of the IEEE Conference on Decision and Control. pp. 761–766. https://doi.org/10.1109/cdc.1978.268028

Ha, S.R., Bae, M.S., 2001. Effects of land use and municipal wastewater treatment changes on stream water quality. Environ. Monit. Assess. 70, 135–151. https://doi.org/10.1023/A:1010649705723

Hakanson, L., 1980. An ecological risk index for aquatic pollution control.a sedimentological approach. Water Res. 14, 975–1001. https://doi.org/10.1016/0043-1354(80)90143-8

Haloi, N., Sarma, H.P., 2012. Heavy metal contaminations in the groundwater of Brahmaputra flood plain: An assessment of water quality in Barpeta District, Assam (India). Environ. Monit. Assess. 184, 6229–6237. https://doi.org/10.1007/s10661-011-2415-x

Hanoon, M.S., Ahmed, A.N., Fai, C.M., Birima, A.H., Razzaq, A., Sherif, M., Sefelnasr, A., El-Shafie, A., 2021. Application of artificial intelligence models for modeling water quality in groundwater: comprehensive review, evaluation and future trends. Water. Air. Soil Pollut. https://doi.org/10.1007/s11270-021-05311-z

Hao, Y., Miao, X., Liu, H., Miao, D., 2021. The variation of heavy metals bioavailability in sediments of liujiang river basin, sw china associated to their speciations and environmental fluctuations, a field study in typical karstic river. Int. J. Environ. Res. Public

Health 18, 3986. https://doi.org/10.3390/ijerph18083986

Härdle, W.K., Simar, L., 2013. Applied multivariate statistical analysis, Springer Nature. https://doi.org/10.1007/978-3-642-17229-8

Harikumar, P.S., Nasir, U.P., Mujeebu Rahman, M.P., 2009. Distribution of heavy metals in the core sediments of a tropical wetland system. Int. J. Environ. Sci. Technol. https://doi.org/10.1007/BF03327626

Harris, M.J., Stinson, J., Landis, W.G., 2017. A bayesian approach to integrated ecological and human health risk assessment for the South River, Virginia Mercury contaminated site. Risk Anal. 37, 1341–1357. https://doi.org/10.1111/RISA.12691

Hazarika, S., Bhuyan, B., 2013. Fluoride, arsenic and iron content of groundwater around six selected tea gardens of Lakhimpur District, Assam, India. Arch. Appl. Sci. Res. 5, 57–61.

He, Z., Li, F., Dominech, S., Wen, X., Yang, S., 2019. Heavy metals of surface sediments in the Changjiang (Yangtze River) Estuary: Distribution, speciation and environmental risks. J. Geochemical Explor. 198, 18–28. https://doi.org/10.1016/j.gexplo.2018.12.015

Helali, M.A., Oueslati, W., Zaaboub, N., Added, A., Aleya, L., 2016. Bioavailability and assessment of heavy metal pollution in sediment cores off the Mejerda River Delta (Gulf of Tunis): How useful is a multiproxy approach? Mar. Pollut. Bull. 105, 215–226. https://doi.org/10.1016/j.marpolbul.2016.02.027

Hing, L.S., Abd Halim Shah, M.N.M., Mohd Yusoff, N., Ong, M.C., 2020. Fractionation of As, Co, Cu and Zn by sequential extraction in surface sediment of Kuala Terengganu river Estuary. Malaysian J. Appl. Sci. 5, 57–68. https://doi.org/10.37231/myjas.2020.5.2.267

Horton, R., 1965. An index number system for rating water quality. J. Water Pollut. Control Fed. 37(3), 6–300.

Hotz, C., Brown, K., 2004. Assessment of the risk of zinc deficiency in populations. Food Nutr. Bull. 25. https://doi.org/10.1177/15648265040251s205

Hou, W., Sun, S., Wang, M., Li, X., Zhang, N., Xin, X., Sun, L., Li, W., Jia, R., 2016. Assessing water quality of five typical reservoirs in lower reaches of Yellow River, China: Using a water quality index method. Ecol. Indic. 61, 309–316. https://doi.org/10.1016/j.ecolind.2015.09.030

Hseu, Z.Y., Chen, Z.S., Tsai, C.C., Tsui, C.C., 2002. Baseline concentrations of ten metals in the freshwater sediments of a watershed in Taiwan. J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng. 37, 1633–1647. https://doi.org/10.1081/ESE-120015426

Hu, W., Wang, H., Dong, L., Huang, B., Borggaard, O.K., Bruun Hansen, H.C., He, Y., Holm, P.E., 2018. Source identification of heavy metals in peri-urban agricultural soils of southeast China: An integrated approach. Environ. Pollut. 237, 650–661. https://doi.org/10.1016/j.envpol.2018.02.070

Huang, S., Rahn, K.A., Arimoto, R., 1999. Testing and optimizing two factor-analysis techniques on aerosol at Narragansett, Rhode Island. Atmos. Environ. 33, 2169–2185. https://doi.org/10.1016/S1352-2310(98)00324-0

Huang, Y., Zuo, R., Li, J., Wu, J., Zhai, Y., Teng, Y., 2018. The spatial and temporal variability of groundwater vulnerability and human health risk in the Limin District, Harbin, China. Water (Switzerland) 10. https://doi.org/10.3390/w10060686

Hue, N.H., Thanh, N.H., 2020. Surface water quality analysis using fuzzy logic approach: A case of inter-provincial irrigation network in Vietnam, in: IOP Conference Series: Earth and Environmental Science. IOP Publishing, p. 012017. https://doi.org/10.1088/1755-1315/527/1/012017

Hunsaker, V.E., Pratt, P.F., 1971. Calcium Magnesium exchange equilibria in soils. Soil Sci. Soc.

Am. J. 35, 151–152. https://doi.org/10.2136/sssaj1971.03615995003500010044x

Hussein, K., Al-Bayati, S., Al-Bakri, S.A.A.A.-B., 2019. Assessing Water Quality for Al-Diwaniyah River, Iraq Using GIS Technique. Eng. Technol. J. 37, 256–264. https://doi.org/10.30684/etj.37.7a.6

Icaga, Y., 2007. Fuzzy evaluation of water quality classification. Ecol. Indic. 7, 710–718. https://doi.org/10.1016/j.ecolind.2006.08.002

Idriss, A.A., Ahmad, A.K., 2013. Heavy metals Nickel and Chromiumin sediments in the Juru river, Penang, Malaysia. J. Environ. Prot. (Irvine,. Calif). 04, 1245–1250. https://doi.org/10.4236/jep.2013.411144

Ikem, A., Egiebor, N.O., Nyavor, K., 2003. Trace elements in water, fish and sediment from Tuskegee Lake, Southeastern USA. Water. Air. Soil Pollut. https://doi.org/10.1023/A:1025694315763

Ipeaiyeda, A.R., Onianwa, P.C., 2016. Sediment quality assessment and dispersion pattern of toxic metals from brewery effluent discharged into the Olosun river, Nigeria. Environ. Earth Sci. 75, 1–12. https://doi.org/10.1007/s12665-015-5143-7

IS 10500, 2012. Indian Standard Drinking Water Specification (Second Revision). Bur. Indian Stand. IS 10500, 1–11.

Iscen, C.F., 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. Environ. Monit. Assess. 144, 269–276. https://doi.org/10.1007/s10661-007-9989-3

Islam, M.S., Ahmed, M.K., Habibullah-Al-Mamun, M., 2015. Geochemical speciation and risk assessment of heavy metals in sediments of a river in Bangladesh. Soil Sediment Contam. 24, 639–655. https://doi.org/10.1080/15320383.2015.997869

Islam, M.S., Proshad, R., Ahmed, S., 2018. Ecological risk of heavy metals in sediment of an urban river in Bangladesh. Hum. Ecol. Risk Assess. 24, 699–720. https://doi.org/10.1080/10807039.2017.1397499

Israeli, M., Nelson, C.B., 1992. Distribution and expected time of residence for U.S. households. Risk Anal. 12, 65–72. https://doi.org/10.1111/j.1539-6924.1992.tb01308.x

Iwegbue, C.M.A., 2007. Metal fractionation in soil profiles at automobile mechanic waste dumps. Waste Manag. Res. 25, 585–593. https://doi.org/10.1177/0734242X07080761

Jafarabadi, A.R., Raudonytė-Svirbutavičienė, E., Shadmehri Toosi, A., Riyahi Bakhtiari, A., 2021. Positive matrix factorization receptor model and dynamics in fingerprinting of potentially toxic metals in coastal ecosystem sediments at a large scale (Persian Gulf, Iran). Water Res. 188, 116509. https://doi.org/10.1016/j.watres.2020.116509

Jain, C.K., Malik, D.S., Yadav, R., 2007. Metal fractionation study on bed sediments of Lake Nainital, Uttaranchal, India. Environ. Monit. Assess. 130, 129–139. https://doi.org/10.1007/s10661-006-9383-6

Jang, J., Sun, C., 1995. Neuro-fuzzy modeling and control. Proc. IEEE 83, 378–406.

Jena, P., Rahaman, S.M., DasMohapatra, P.K., Barik, D.P., Patra, D.S., 2022. Surface water quality assessment, prediction & modelling of river Daya in Odisha. https://doi.org/10.21203/rs.3.rs-1181043/v1

Jha, M.K., Shekhar, A., Jenifer, M.A., 2020. Assessing groundwater quality for drinking water supply using hybrid fuzzy-GIS-based water quality index. Water Res. 179. https://doi.org/10.1016/j.watres.2020.115867

Jha, P.K., Subramanian, V., Sitasawad, R., Van Grieken, R., 1990. Heavy metals in sediments of the Yamura River (a tributary of the Ganges), India. Sci. Total Environ. 95, 7–27. https://doi.org/10.1016/0048-9697(90)90049-Z

Ji, H., Li, H., Zhang, Y., Ding, H., Gao, Y., Xing, Y., 2018. Distribution and risk assessment of heavy metals in overlying water, porewater, and sediments of Yongding River in a coal mine brownfield. J. Soils Sediments 18, 624–639. https://doi.org/10.1007/s11368-017-1833-y

Jiang, F., Ren, B., Hursthouse, A.S., Zhou, Y., 2018. Trace metal pollution in topsoil surrounding the Xiangtan Manganese mine area (South-central China): Source identification, spatial distribution and assessment of potential ecological risks. Int. J. Environ. Res. Public Health 15. https://doi.org/10.3390/ijerph15112412

Jiang, J., Khan, A.U., Shi, B., Tang, S., Khan, J., 2019. Application of positive matrix factorization to identify potential sources of water quality deterioration of Huaihe River, China. Appl. Water Sci. 9. https://doi.org/10.1007/S13201-019-0938-4

Jiang, R., Pan, Z., Lin, C., Wang, W., Wang, L., Liu, Y., Chen, J., Zhou, K., Lin, H., 2022. Integrated insights into potentially hazardous metals in sediments of a typical bay under long-term human impacts: Implications for coastal management. J. Clean. Prod. 364, 132566. https://doi.org/10.1016/j.jclepro.2022.132566

Jiang, X., Wang, W., Wang, S., Zhang, B., Hu, J., 2012. Initial identification of heavy metals contamination in Taihu Lake, a eutrophic lake in China. J. Environ. Sci. (China) 24, 1539–1548. https://doi.org/10.1016/S1001-0742(11)60986-8

Jiang, Y., Chao, S., Liu, J., Yang, Y., Chen, Y., Zhang, A., Cao, H., 2017. Source apportionment and health risk assessment of heavy metals in soil for a township in Jiangsu Province, China. Chemosphere 168, 1658–1668. https://doi.org/10.1016/j.chemosphere.2016.11.088

Jiang, Y., Li, C., Song, H., Wang, W., 2022. Deep learning model based on urban multi-source data for predicting heavy metals (Cu, Zn, Ni, Cr) in industrial sewer networks. J. Hazard. Mater. 432, 128732. https://doi.org/10.1016/j.jhazmat.2022.128732

Jianhua, W.U., Peiyue, L.I., Hui, Q., 2011. Groundwater quality in Jingyuan County, a semi-humid area in northwest China. E-Journal Chem. 8, 787–793.

Jiménez-Oyola, S., Chavez, E., García-Martínez, M.J., Ortega, M.F., Bolonio, D., Guzmán-Martínez, F., García-Garizabal, I., Romero, P., 2021. Probabilistic multi-pathway human health risk assessment due to heavy metal(loid)s in a traditional gold mining area in Ecuador. Ecotoxicol. Environ. Saf. 224. https://doi.org/10.1016/j.ecoenv.2021.112629

Jolly, Y.N., Islam, A., Akbar, S., 2013. Transfer of metals from soil to vegetables and possible health risk assessment. Springerplus 2, 1–8. https://doi.org/10.1186/2193-1801-2-385

Jordão, C.P., Nickless, G., 1989. Chemical associations of Zn, Cd, Pb and Cu in soils and sediments determined by the sequential extraction technique. Environ. Technol. Lett. 10, 743–752. https://doi.org/10.1080/09593338909384793

Jorfi, S., Maleki, R., Jaafarzadeh, N., Ahmadi, M., 2017. Pollution load index for heavy metals in Mian-Ab plain soil, Khuzestan, Iran. Data Br. 15, 584–590. https://doi.org/10.1016/j.dib.2017.10.017

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. Environ. Monit. Assess. 173, 625–641. https://doi.org/10.1007/s10661-010-1411-x

Jun, R., Zhen, S., Ling, T., Jianxiu, H., 2017. Speciation and contamination assessment of metals in the sediments from the lanzhou section of the yellow river, China. Environ. Prot. Eng. 43, 113–124. https://doi.org/10.5277/epe170307

Juvela, M., Lehtinen, K., Paatero, P., 1996. The use of Positive Matrix Factorization in the analysis of molecular line spectra. Mon. Not. R. Astron. Soc. 280, 616–626.

Kabala, C., Singh, B.R., 2001. Fractionation and mobility of copper, lead, and zinc in soil

profiles in the vicinity of a copper smelter. J. Environ. Qual. 30, 485–492. https://doi.org/10.2134/jeq2001.302485x

Kale, S.S., Kadam, A.K., Kumar, S., Pawar, N.J., 2010. Evaluating pollution potential of leachate from landfill site, from the Pune metropolitan city and its impact on shallow basaltic aquifers. Environ. Monit. Assess. 162, 327–346. https://doi.org/10.1007/s10661-009-0799-7

Kang, X., Song, J., Yuan, H., Duan, L., Li, X., Li, N., Liang, X., Qu, B., 2017. Speciation of heavy metals in different grain sizes of Jiaozhou Bay sediments: Bioavailability, ecological risk assessment and source analysis on a centennial timescale. Ecotoxicol. Environ. Saf. 143, 296–306. https://doi.org/10.1016/j.ecoenv.2017.05.036

Karbassi, A.R., Monavari, S.M., Nabi Bidhendi, G.R., Nouri, J., Nematpour, K., 2008. Metal pollution assessment of sediment and water in the Shur River. Environ. Monit. Assess. 147, 107–116. https://doi.org/10.1007/s10661-007-0102-8

Karbassi, A.R., Shankar, R., 2005. Geochemistry of two sediment cores from the west coast of India. Int. J. Environ. Sci. Technol. 1, 307–316. https://doi.org/10.1007/bf03325847

Karbassi, A.R., Torabi, F., Ghazban, F., Ardestani, M., 2011. Association of trace metals with various sedimentary phases in dam reservoirs. Int. J. Environ. Sci. Technol. 8, 841–852. https://doi.org/10.1007/BF03326267

Kardam, A., Raj, K.R., Arora, J.K., Srivastava, S., 2012. Artificial neural network modeling for biosorption of Pb (II) ions on nanocellulose fibers. Bionanoscience 2, 153–160. https://doi.org/10.1007/s12668-012-0045-6

Kaushik, A., Kansal, A., Santosh, Meena, Kumari, S., Kaushik, C.P., 2009. Heavy metal contamination of river Yamuna, Haryana, India: Assessment by metal enrichment factor of the sediments. J. Hazard. Mater. 164, 265–270. https://doi.org/10.1016/j.jhazmat.2008.08.031

Kayhanian, M., Fruchtman, B.D., Gulliver, J.S., Montanaro, C., Ranieri, E., Wuertz, S., 2012. Review of highway runoff characteristics: Comparative analysis and universal implications. Water Res. 46, 6609–6624. https://doi.org/10.1016/j.watres.2012.07.026

Kazemi, S.M., Hosseini, S.M., 2011. Comparison of spatial interpolation methods for estimating heavy metals in sediments of Caspian Sea. Expert Syst. Appl. 38, 1632–1649. https://doi.org/10.1016/j.eswa.2010.07.085

Ke, X., Gui, S., Huang, H., Zhang, H., Wang, C., Guo, W., 2017. Ecological risk assessment and source identification for heavy metals in surface sediment from the Liaohe River protected area, China. Chemosphere 175, 473–481. https://doi.org/10.1016/j.chemosphere.2017.02.029

Keller, J.M., Hunt, D.J., 1986. Incorporating fuzzy membership functions into the perceptron algorithm. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-7, 693–699. https://doi.org/10.1109/TPAMI.1985.4767725

Kersten, M., Forstner, U., 1986. Chemical fractionation of heavy metals in anoxic estuarine and coastal sediments. Water Sci. Technol. 18, 121–130. https://doi.org/10.2166/wst.1986.0187

Khan, R., Saxena, A., Shukla, S., Sekar, S., Senapathi, V., Wu, J., 2021. Environmental contamination by heavy metals and associated human health risk assessment: a case study of surface water in Gomti River Basin, India. Environ. Sci. Pollut. Res. 28, 56105–56116. https://doi.org/10.1007/s11356-021-14592-0

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput. Electron. Agric. 153, 213–225. https://doi.org/10.1016/j.compag.2018.07.016

Khosravi, K., Barzegar, R., Miraki, S., Adamowski, J., Daggupati, P., Alizadeh, M.R., Pham, B.T., Alami, M.T., 2020. Stochastic modeling of groundwater fluoride contamination: introducing lazy learners. Groundwater 58, 723–734. https://doi.org/10.1111/gwat.12963

Khosravi, K., Mao, L., Kisi, O., Yaseen, Z.M., Shahid, S., 2018. Quantifying hourly suspended sediment load using data mining models: Case study of a glacierized Andean catchment in Chile. J. Hydrol. 567, 165–179. https://doi.org/10.1016/j.jhydrol.2018.10.015

Khozani, Z.S., Bonakdari, H., Zaji, A.H., 2016. Application of a genetic algorithm in predicting the percentage of shear force carried by walls in smooth rectangular channels. Meas. J. Int. Meas. Confed. 87, 87–98. https://doi.org/10.1016/j.measurement.2016.03.018

Kim, K.W., Chanpiwat, P., Hanh, H.T., Phan, K., Sthiannopkao, S., 2011. Arsenic geochemistry of groundwater in Southeast Asia. Front. Med. China. https://doi.org/10.1007/s11684-011-0158-2

Kisi, O., Dailr, A.H., Cimen, M., Shiri, J., 2012. Suspended sediment modeling using genetic programming and soft computing techniques. J. Hydrol. 450–451, 48–58. https://doi.org/10.1016/j.jhydrol.2012.05.031

Kittrick, J.A., 1971. Handbook of geochemistry. Soil Sci. Soc. Am. J. 35, viii–viii. https://doi.org/10.2136/sssaj1971.03615995003500030012x

Koelmel, J., Amarasiriwardena, D., 2012. Imaging of metal bioaccumulation in Hay-scented fern (Dennstaedtia punctilobula) rhizomes growing on contaminated soils by laser ablation ICP-MS. Environ. Pollut. 168, 62–70. https://doi.org/10.1016/j.envpol.2012.03.035

Krishan, G., Rao, M.S., Loyal, R.S., Lohani, A.K., Tuli, N.K., Takshi, K.S., Kumar, C.P., Semwal, P., Kumar, S., 2016. Assessment of vegetables and soils for some heavy metals from irrigated farmlands irrigated with industrial effluents of hpc, nagaon, assam (india), Octa Journal of Environment Research.

Krishna, A.K., Mohan, K.R., 2014. Risk assessment of heavy metals and their source distribution in waters of a contaminated industrial site. Environ. Sci. Pollut. Res. 21, 3653–3669. https://doi.org/10.1007/s11356-013-2359-5

Krzywinski, M., Altman, N., 2015. Points of significance: multiple linear regression. Nat. Methods. https://doi.org/10.1038/nmeth.3665

Kukreja, H., Bharath, N., Siddesh, C.S., Kuldeep, S., 2016. An introduction to artificial neural networks. Int. J. Adv. Res. Innov. ideas Educ. 1, 27–30. https://doi.org/10.3390/mol2net-07-11843

Kulkarni, A.R., 2020. Water qaulity indices for Panchaganga river basin. Poll Res. 39, 424–428.

Kumar, R., Mittal, S., Peechat, S., Sahoo, P.K., Sahoo, S.K., 2020. Quantification of groundwater–agricultural soil quality and associated health risks in the agri-intensive Sutlej River Basin of Punjab, India. Environ. Geochem. Health 42, 4245–4268. https://doi.org/10.1007/s10653-020-00636-w

Kumar, V., Thakur, R.K., 2017. Pollution load of SIDCUL effluent with reference to heavy metals accumulated in sediments using pollution load index (PLI) and geo-accumulation index (I-geo) at Haridwar (Uttarakhand), India. J. Environ. Biosci. 31, 163–168.

La Colla, N.S., Negrin, V.L., Marcovecchio, J.E., Botté, S.E., 2015. Dissolved and particulate metals dynamics in a human impacted estuary from the SW Atlantic. Estuar. Coast. Shelf Sci. 166, 45–55. https://doi.org/10.1016/j.ecss.2015.05.009

Lai, T.M., Lee, W., Hur, J., Kim, Y., Huh, I.A., Shin, H.S., Kim, C.K., Lee, J.H., 2013. Influence of sediment grain size and land use on the distributions of heavy metals in sediments of the han river basin in korea and the assessment of anthropogenic pollution. Water. Air. Soil Pollut. 224, 1–12. https://doi.org/10.1007/s11270-013-1609-y

Landis, W, Sofield, R., Yu, M., Landis, WG, 1995. Introduction to environmental toxicology:

impacts of chemicals upon ecological systems, Choice Reviews Online. https://doi.org/10.5860/choice.33-1539

Landwehr, J.M., 1979. A statistical view of a class of water quality indices. Water Resour. Res. 15, 460–468. https://doi.org/10.1029/WR015i002p00460

Landwehr, J.M., Deininger, R.A., McClelland, N.L., Brown, R.M., 1974. An objective water quality index. J. Water Pollut. Control Fed. 46, 1804–1807.

Lee, D.H., Kim, J.H., Mendoza, J.A., Lee, C.H., Kang, J.H., 2016. Characterization and source identification of pollutants in runoff from a mixed land use watershed using ordination analyses. Environ. Sci. Pollut. Res. 23, 9774–9790. https://doi.org/10.1007/s11356-016-6155-x

Lee, K., 2005. First course on fuzzy theory and applications. Choice Rev. Online 42, 42-5917-42–5917. https://doi.org/10.5860/choice.42-5917

Lei, P., Zhang, H., Shan, B., Lv, S., Tang, W., 2016. Heavy metals in estuarine surface sediments of the Hai River Basin, variation characteristics, chemical speciation and ecological risk. Environ. Sci. Pollut. Res. 23, 7869–7879. https://doi.org/10.1007/s11356-016-6059-9

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. Ecol. Indic. 9, 1188–1197. https://doi.org/10.1016/j.ecolind.2009.02.006

Li, A., Jang, J.K., Scheff, P.A., 2003. Application of EPA CMB8.2 model for source apportionment of sediment PAHS in Lake Calumet, Chicago. Environ. Sci. Technol. 37, 2958–2965. https://doi.org/10.1021/es026309v

Li, H., Hopke, P.K., Liu, X., Du, X., Li, F., 2015. Application of positive matrix factorization to source apportionment of surface water quality of the Daliao River basin, northeast China. Environ. Monit. Assess. 187, 1–12. https://doi.org/10.1007/s10661-014-4154-2

Liang, B., Han, G., Liu, M., Yang, K., Li, X., Liu, J., 2018. Distribution, sources, and water quality assessment of dissolved heavy metals in the Jiulongjiang river water, southeast China. Int. J. Environ. Res. Public Health 15. https://doi.org/10.3390/ijerph15122752

Liang, G., Zhang, B., Lin, M., Wu, S., Hou, H., Zhang, J., Qian, G., Huang, X., Zhou, J., 2017. Evaluation of heavy metal mobilization in creek sediment: Influence of RAC values and ambient environmental factors. Sci. Total Environ. 607–608, 1339–1347. https://doi.org/10.1016/j.scitotenv.2017.06.238

Liao, J., Qian, X., Liu, F., Deng, S., Lin, H., Liu, X., Wei, C., 2021. Multiphase distribution and migration characteristics of heavy metals in typical sandy intertidal zones: insights from solid-liquid partitioning. Ecotoxicol. Environ. Saf. 208, 111674. https://doi.org/10.1016/j.ecoenv.2020.111674

Lim, W.Y., Aris, A.Z., Tengku Ismail, T.H., 2013. Spatial geochemical distribution and sources of heavy metals in the sediment of Langat River, Western Peninsular Malaysia. Environ. Forensics 14, 133–145. https://doi.org/10.1080/15275922.2013.781078

Liou, S.M., Lo, S.L., Wang, S.H., 2004. A generalized water quality index for Taiwan. Environ. Monit. Assess. 96, 35–52. https://doi.org/10.1023/B:EMAS.0000031715.83752.a1

Liu, B., Hu, K., Jiang, Z., Qu, F., Su, X., 2011. A 50-year sedimentary record of heavy metals and their chemical speciations in the Shuangtaizi River estuary (China): Implications for pollution and biodegradation. Front. Environ. Sci. Eng. China 5, 435–444. https://doi.org/10.1007/s11783-011-0352-0

Liu, C., Lin, K., Kuo, Y., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Total Environ. 313, 77–89.

Liu, E., Shen, J., 2014. A comparative study of metal pollution and potential eco-risk in the sediment of Chaohu Lake (China) based on total concentration and chemical speciation.

Environ. Sci. Pollut. Res. 21, 7285–7295. https://doi.org/10.1007/s11356-014-2639-8

Liu, J., Xu, Y., Cheng, Y., Zhao, Y., Pan, Y., Fu, G., Dai, Y., 2017. Occurrence and risk assessment of heavy metals in sediments of the Xiangjiang River, China. Environ. Sci. Pollut. Res. 24, 2711–2723. https://doi.org/10.1007/s11356-016-8044-8

Liu, L., Ouyang, W., Wang, Y., Tysklind, M., Hao, F., Liu, H., Hao, X., Xu, Y., Lin, C., Su, L., 2020. Heavy metal accumulation, geochemical fractions, and loadings in two agricultural watersheds with distinct climate conditions. J. Hazard. Mater. 389, 122125. https://doi.org/10.1016/j.jhazmat.2020.122125

Liu, R., Men, C., Yu, W., Xu, F., Wang, Q., Shen, Z., 2018. Uncertainty in positive matrix factorization solutions for PAHs in surface sediments of the Yangtze River Estuary in different seasons. Chemosphere 191, 922–936. https://doi.org/10.1016/j.chemosphere.2017.10.070

Liu, Z., Kuang, Y., Lan, S., Cao, W., Yan, Z., Chen, L., Chen, Q., Feng, Q., Zhou, H., 2021. Pollution distribution of potentially toxic elements in a Karstic river affected by manganese mining in Changyang, Western Hubei, Central China. Int. J. Environ. Res. Public Health 18, 1–15. https://doi.org/10.3390/ijerph18041870

Long, E.R., Macdonald, D.D., Smith, S.L., Calder, F.D., 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. Environ. Manage. 19, 81–97. https://doi.org/10.1007/BF02472006

Low, K.H., Zain, S.M., Abas, M.R., Md. Salleh, K., Teo, Y.Y., 2015. Distribution and health risk assessment of trace metals in freshwater tilapia from three different aquaculture sites in Jelebu Region (Malaysia). Food Chem. 177, 390–396. https://doi.org/10.1016/j.foodchem.2015.01.059

Lu, C., Xu, Z., Dong, B., Zhang, Y., Wang, M., Zeng, Y., Zhang, C., 2022. Machine learning for the prediction of heavy metal removal by chitosan-based flocculants. Carbohydr. Polym. 285, 119240. https://doi.org/10.1016/j.carbpol.2022.119240

Lu, H., Li, H., Liu, T., Fan, Y., Yuan, Y., Xie, M., Qian, X., 2019. Simulating heavy metal concentrations in an aquatic environment using artificial intelligence models and physicochemical indexes. Sci. Total Environ. 694, 133591. https://doi.org/10.1016/j.scitotenv.2019.133591

Luo, P., Xu, C., Kang, S., Huo, A., Lyu, J., Zhou, M., Nover, D., 2021. Heavy metals in water and surface sediments of the Fenghe River Basin, China: assessment and source analysis. Water Sci. Technol. 84, 3072–3090. https://doi.org/10.2166/wst.2021.335

Luoma, S.N., 1989. Can we determine the biological availability of sediment-bound trace elements? Sediment/water Interact. Proc 4th Symp. Melbourne, 1987 379–396. https://doi.org/10.1007/978-94-009-2376-8_35

Lyon, F., 1994. Monographs on the evaluation of carcinogenic risks to humans, IARC. https://doi.org/10.1136/jcp.48.7.691-a

M. Mijwil, M., 2018. Artificial neural networks advantages and disadvantages. LinkedIn 1–2.

Macdonald, D.D., Carr, R.S., Calder, F.D., Long, E.R., Ingersoll, C.G., 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. Ecotoxicology 5, 253–278. https://doi.org/10.1007/BF00118995

MacDonald, D.D., Ingersoll, C.C.G., 2003. Development and applications of sediment quality criteria for managing contaminated sediment in British Columbia, British Columbia Ministry of Water, Land and Air Protection. Victoria.

MacDonald, D.D., Ingersoll, C.G., Berger, T.A., 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch. Environ. Contam. Toxicol. 39, 20–31. https://doi.org/10.1007/s002440010075

Magesh, N.S., Chandrasekar, N., Elango, L., 2017. Trace element concentrations in the groundwater of the Tamiraparani river basin, South India: Insights from human health risk and multivariate statistical techniques. Chemosphere 185, 468–479. https://doi.org/10.1016/j.chemosphere.2017.07.044

Magesh, N.S., Tiwari, A., Botsa, S.M., da Lima Leitao, T., 2021. Hazardous heavy metals in the pristine lacustrine systems of Antarctica: Insights from PMF model and ERA techniques. J. Hazard. Mater. 412, 125263. https://doi.org/10.1016/j.jhazmat.2021.125263

Mahapatra, S.S., Nanda, S.K., Panigrahy, B.K., 2011. A Cascaded Fuzzy Inference System for Indian river water quality prediction. Adv. Eng. Softw. 42, 787–796. https://doi.org/10.1016/j.advengsoft.2011.05.018

Maharana, C., Srivastava, D., Tripathi, J.K., 2018. Geochemistry of sediments of the Peninsular rivers of the Ganga basin and its implication to weathering, sedimentary processes and provenance. Chem. Geol. 483, 1–20. https://doi.org/10.1016/j.chemgeo.2018.02.019

Mahmoudi, E., Ng, L.Y., Mohammad, A.W., Ba-Abbad, M.M., Razzaz, Z., 2018. Enhancement of polysulfone membrane with integrated ZnO nanoparticles for the clarification of sweetwater. Int. J. Environ. Sci. Technol. 15, 561–570. https://doi.org/10.1007/s13762-017-1413-0

Mahurpawar, M., 2015. Effects of heavy metals on human health. Int. J. Res. Granthaalayah 3, 1–7. https://doi.org/10.29121/granthaalayah.v3.i9se.2015.3282

Maity, S., Biswas, R., Sarkar, A., 2020. Comparative valuation of groundwater quality parameters in Bhojpur, Bihar for arsenic risk assessment. Chemosphere 259, 127398. https://doi.org/10.1016/j.chemosphere.2020.127398

Mamun, M., An, K.G., 2021. Application of multivariate statistical techniques and water quality index for the assessment of water quality and apportionment of pollution sources in the yeongsan river, south korea. Int. J. Environ. Res. Public Health 18, 8268. https://doi.org/10.3390/ijerph18168268

Manivasakam, N., 2016. Industrial effluents: origin, characteristics, effects, analysis & treatment. Chem. Publ.

Marchini, A., Facchinetti, T., Mistri, M., 2009. F-IND: A framework to design fuzzy indices of environmental conditions. Ecol. Indic. 9, 485–496. https://doi.org/10.1016/j.ecolind.2008.07.004

Mariappan, P., Yegnaraman, V., Vasudevan, T., 2000. Correlation between fluoride and alkalinity in groundwaters of fluorosis endemic - Salem District. Indian J. Environ. Prot. 20, 182–187.

Martin, T.R., Holdich, D.M., 1986. The acute lethal toxicity of heavy metals to peracarid crustaceans (with particular reference to fresh-water asellids and gammarids). Water Res. 20, 1137–1147. https://doi.org/10.1016/0043-1354(86)90060-6

Massadeh, A.M., Alomary, A.A., Mir, S., Momani, F.A., Haddad, H.I., Hadad, Y.A., 2016. Analysis of Zn, Cd, As, Cu, Pb, and Fe in snails as bioindicators and soil samples near traffic road by ICP-OES. Environ. Sci. Pollut. Res. 23, 13424–13431. https://doi.org/10.1007/s11356-016-6499-2

McKone, T.E., Deshpande, A.W., 2005. Can fuzzy logic bring complex environmental problems into focus?. Environ. Sci. Technol. https://doi.org/10.1021/es0531632

Medeiros, A.C., Faial, K.R.F., do Carmo Freitas Faial, K., da Silva Lopes, I.D., de Oliveira Lima, M., Guimarães, R.M., Mendonça, N.M., 2017. Quality index of the surface water of Amazonian rivers in industrial areas in Pará Brazil. Mar. Pollut. Bull. 123, 156–164. https://doi.org/10.1016/j.marpolbul.2017.09.002

Mena, I., Horiuchi, K., Burke, K., Cotzias, G.C., 1969. Chronic manganese poisoning: Individual susceptibility and absorption of iron. Neurology 19, 1000–1006.

https://doi.org/10.1212/wnl.19.10.1000

Mendivil-Garcia, K., Amabilis-Sosa, L.E., Rodríguez-Mata, A.E., Rangel-Peraza, J.G., Gonzalez-Huitron, V., Cedillo-Herrera, C.I.G., 2020. Assessment of intensive agriculture on water quality in the Culiacan River basin, Sinaloa, Mexico. Environ. Sci. Pollut. Res. 27, 28636–28648. https://doi.org/10.1007/s11356-020-08653-z

Miao, X., Song, M., Xu, G., Hao, Y., Zhang, H., 2022. The accumulation and transformation of heavy metals in sediments of Liujiang River Basin in southern China and their threatening on water security. Int. J. Environ. Res. Public Health 19, 1619. https://doi.org/10.3390/ijerph19031619

Michalopoulos, C., Tzamtzis, N., Liodakis, S., 2014. Effects of an intensive hog farming operation on groundwater in East Mediterranean (II): A study on K+ , Na+ , Cl- , PO43 -- P, Ca2+ , Mg2+ , Fe3+/Fe2+, Mn2+, Cu2+, Zn2+ and Ni2+. Bull. Environ. Contam. Toxicol. 93, 688–93. https://doi.org/10.1007/s00128-014-1402-7

Mikhailenko, A. V., Ruban, D.A., Ermolaev, V.A., van Loon, A.J., 2020. Cadmium pollution in the tourism environment: A literature review. Geosci. 10, 1–18. https://doi.org/10.3390/geosciences10060242

Miller, C. V, Gutierrez-Magness, A., Feit Majedi, B., Foster, G., 2007. Water quality in the Upper Anacostia River, Maryland: continuous and discrete monitoring with simulations to estimate concentrations and yields, 2003-05., Usgs Survey.

Mirabbasi, R., Mazloumzadeh, S.M., Rahnama, M.B., 2008. Evaluation of irrigation water quality using fuzzy logic. Res. J. Environ. Sci. 2, 340–352. https://doi.org/10.3923/rjes.2008.340.352

Mishra, A.K., Özger, M., Singh, V.P., 2009. An entropy-based investigation into the variability of precipitation. J. Hydrol. 370, 139–154. https://doi.org/10.1016/j.jhydrol.2009.03.006

Mitra, S., Sarkar, S.K., Raja, P., Biswas, J.K., Murugan, K., 2018. Dissolved trace elements in Hooghly (Ganges) River Estuary, India: Risk assessment and implications for management. Mar. Pollut. Bull. 133, 402–414. https://doi.org/10.1016/j.marpolbul.2018.05.057

Mna, H. Ben, Helali, M.A., Oueslati, W., Amri, S., Aleya, L., 2021. Spatial distribution, contamination assessment and potential ecological risk of some trace metals in the surface sediments of the Gulf of Tunis, North Tunisia. Mar. Pollut. Bull. 170. https://doi.org/10.1016/j.marpolbul.2021.112608

Mna, H. Ben, Oueslati, W., Helali, M.A., Zaaboub, N., Added, A., Aleya, L., 2017. Distribution and assessment of heavy metal toxicity in sediment cores from Bizerte Lagoon, Tunisia. Environ. Monit. Assess. 189, 1–18. https://doi.org/10.1007/s10661-017-6073-5

Mohammed, H., Michel Tornyeviadzi, H., Seidu, R., 2022. Emulating process-based water quality modelling in water source reservoirs using machine learning. J. Hydrol. 609, 127675. https://doi.org/10.1016/j.jhydrol.2022.127675

Mokarram, M., Pourghasemi, H.R., Huang, K., Zhang, H., 2022. Investigation of water quality and its spatial distribution in the Kor River basin, Fars province, Iran. Environ. Res. 204, 112294. https://doi.org/10.1016/j.envres.2021.112294

Morillo, J., Usero, J., Gracia, I., 2002. Partitioning of metals in sediments from the Odiel River (Spain). Environ. Int. 28, 263–271. https://doi.org/10.1016/S0160-4120(02)00033-8

Mortvedt, J.J., 1995. Heavy metal contaminants in inorganic and organic fertilizers. Fertil. Res. 43, 55–61. https://doi.org/10.1007/BF00747683

Mpimpas, H., Anagnostopoulos, P., Ganoulis, J., 2001. Modelling of water pollution in the Thermaikos Gulf with fuzzy parameters. Ecol. Modell. 142, 91–104. https://doi.org/10.1016/S0304-3800(01)00281-2

Mukherjee, I., Singh, U.K., Singh, R.P., Anshumali, Kumari, D., Jha, P.K., Mehta, P., 2020.

Characterization of heavy metal pollution in an anthropogenically and geologically influenced semi-arid region of east India and assessment of ecological and human health risks. Sci. Total Environ. 705, 135801. https://doi.org/10.1016/j.scitotenv.2019.135801

Müller, B., Berg, M., Yao, Z.P., Zhang, X.F., Wang, D., Pfluger, A., 2008. How polluted is the Yangtze river? Water quality downstream from the Three Gorges Dam. Sci. Total Environ. 402, 232–247. https://doi.org/10.1016/j.scitotenv.2008.04.049

Müller, G., 1969. Index of geoaccumulation in sediments of the Rhine River. Geol. J. 2, 108–118.

Mustafa, M.R., Rezaur, R.B., Saiedi, S., Isa, M.H., 2012. River suspended sediment prediction using various multilayer perceptron neural network training algorithms-A case study in Malaysia. Water Resour. Manag. 26, 1879–1897. https://doi.org/10.1007/s11269-012-9992-5

Muyessar, T., Jilili, A., JIANG, F.-Q., 2013. Distribution characteristics of soil heavy metal content in northern slope of Tianshan Mountains and its source explanation. Chinese J. Eco-Agriculture 21, 883–890. https://doi.org/10.3724/sp.j.1011.2013.00883

Najah, A., El-Shafie, A., Karim, O.A., El-Shafie, A.H., 2013. Application of artificial neural networks for water quality prediction. Neural Comput. Appl. 22, 187–201. https://doi.org/10.1007/s00521-012-0940-3

Naji, A., Sohrabi, T., 2015. Distribution and contamination pattern of heavy metals from surface sediments in the southern part of caspian sea, Iran. Chem. Speciat. Bioavailab. 27, 29–43. https://doi.org/10.1080/09542299.2015.1023089

Nayak, J.G., Patil, L.G., Patki, V.K., 2020. Development of water quality index for Godavari River (India) based on fuzzy inference system. Groundw. Sustain. Dev. 10, 100350. https://doi.org/10.1016/j.gsd.2020.100350

Nikoo, M.R., Kerachian, R., Malakpour-Estalaki, S., Bashi-Azghadi, S.N., Azimi-Ghadikolaee, M.M., 2011. A probabilistic water quality index for river water quality assessment: A case study. Environ. Monit. Assess. 181, 465–478. https://doi.org/10.1007/s10661-010-1842-4

Nimyel, N., Egila, J., Lohdip, Y., 2015. Heavy metal speciation in some selected farms treated with urban solid waste, in Jos South, Plateau State, Nigeria. Int. Res. J. Pure Appl. Chem. 5, 342–351. https://doi.org/10.9734/irjpac/2015/14179

Niu, L., Li, J., Luo, X., Fu, T., Chen, O., Yang, Q., 2021. Identification of heavy metal pollution in estuarine sediments under long-term reclamation: Ecological toxicity, sources and implications for estuary management. Environ. Pollut. 290, 118126. https://doi.org/10.1016/j.envpol.2021.118126

Njuguna, S.M., Onyango, J.A., Githaiga, K.B., Gituru, R.W., Yan, X., 2020. Application of multivariate statistical analysis and water quality index in health risk assessment by domestic use of river water. Case study of Tana River in Kenya. Process Saf. Environ. Prot. 133, 149–158. https://doi.org/10.1016/j.psep.2019.11.006

Noori, R., Sabahi, M.S., Karbassi, A.R., Baghvand, A., Zadeh, H.T., 2010. Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. Desalination 260, 129–136. https://doi.org/10.1016/j.desal.2010.04.053

Norwich, A.M., Turksen, I.B., 1984. A model for the measurement of membership and the consequences of its empirical implementation. Fuzzy Sets Syst. 12, 1–25. https://doi.org/10.1016/0165-0114(84)90047-2

Nouh, M., Al-Noman, N., 2009. Regression models for the prediction of water quality in the stormwater of urban arid catchments. Can. J. Civ. Eng. 36, 331–344. https://doi.org/10.1139/S08-048

Nowicki, S., Koehler, J., Charles, K.J., 2020. Including water quality monitoring in rural water services: why safe water requires challenging the quantity versus quality dichotomy. NPJ

Clean Water 3. https://doi.org/10.1038/s41545-020-0062-x

Nriagu, J.O., Coker, R.D., 1980. Trace metals in humic and fulvic acids from lake Ontario Sediments. Environ. Sci. Technol. 14, 443–446. https://doi.org/10.1021/es60164a001

Ocampo-Duque, W., Ferré-Huguet, N., Domingo, J.L., Schuhmacher, M., 2006. Assessing water quality in rivers with fuzzy inference systems: A case study. Environ. Int. 32, 733–742. https://doi.org/10.1016/j.envint.2006.03.009

Ocampo-Duque, W., Osorio, C., Piamba, C., Schuhmacher, M., Domingo, J.L., 2013. Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems: Application to the Cauca River, Colombia. Environ. Int. 52, 17–28. https://doi.org/10.1016/j.envint.2012.11.007

Olivares-Rieumont, S., De La Rosa, D., Lima, L., Graham, D.W., D'Alessandro, K., Borroto, J., Martínez, F., Sánchez, J., 2005. Assessment of heavy metal levels in Almendares River sediments - Havana City, Cuba. Water Res. 39, 3945–3953. https://doi.org/10.1016/j.watres.2005.07.011

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 Res. 46, 3110–3122. https://doi.org/10.1016/j.watres.2012.03.028

Ömer Faruk, D., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Eng. Appl. Artif. Intell. 23, 586–594. https://doi.org/10.1016/j.engappai.2009.09.015

Omwene, P.I., Öncel, M.S., Çelen, M., Kobya, M., 2019. Influence of arsenic and boron on the water quality index in mining stressed catchments of Emet and Orhaneli streams (Turkey). Environ. Monit. Assess. 191. https://doi.org/10.1007/s10661-019-7337-z

Osei, J., Nyame, F.K., Armah, T.K., Osae, S.K., Dampare, S.B., Fianko, J.R., Adomako, D., Bentil, N., 2010. Application of multivariate analysis for identification of pollution sources in the densu delta wetland in the vicinity of a landfill site in Ghana. J. Water Resour. Prot. 02, 1020–1029. https://doi.org/10.4236/jwarp.2010.212122

Ostad-Ali-Askari, K., Shayannejad, M., 2021. Quantity and quality modelling of groundwater to manage water resources in Isfahan-Borkhar Aquifer. Environ. Dev. Sustain. 1–17. https://doi.org/10.1007/s10668-021-01323-1

Ozel, H.U., Gemici, B.T., Gemici, E., Ozel, H.B., Cetin, M., Sevik, H., 2020. Application of artificial neural networks to predict the heavy metal contamination in the Bartin River. Environ. Sci. Pollut. Res. 27, 42495–42512. https://doi.org/10.1007/s11356-020-10156-w

Paatero, P., 1997. Least squares formulation of robust non-negative factor analysis, in: Chemometrics and Intelligent Laboratory Systems. pp. 23–35. https://doi.org/10.1016/S0169-7439(96)00044-5

Paatero, P., Tapper, U., 1994. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111–126. https://doi.org/10.1002/env.3170050203

Paes, É. de C., Veloso, G.V., Fonseca, A.A. da, Fernandes-Filho, E.I., Fontes, M.P.F., Soares, E.M.B., 2022. Predictive modeling of contents of potentially toxic elements using morphometric data, proximal sensing, and chemical and physical properties of soils under mining influence. Sci. Total Environ. 817, 152972. https://doi.org/10.1016/j.scitotenv.2022.152972

Pal, D., Maiti, S.K., 2018. Heavy metal speciation, leaching and toxicity status of a tropical rain-fed river Damodar, India. Environ. Geochem. Health 40, 2303–2324. https://doi.org/10.1007/s10653-018-0097-9

Palani, S., Liong, S.Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Mar.

Pollut. Bull. 56, 1586–1597. https://doi.org/10.1016/j.marpolbul.2008.05.021

Palanques, A., Diaz, J.I., 1994. Anthropogenic heavy metal pollution in the sediments of the Barcelona continental shelf (Northwestern Mediterranean). Mar. Environ. Res. 38, 17–31. https://doi.org/10.1016/0141-1136(94)90043-4

Pan, B., Han, X., Chen, Y., Wang, L., Zheng, X., 2022. Determination of key parameters in water quality monitoring of the most sediment-laden Yellow River based on water quality index. Process Saf. Environ. Prot. 164, 249–259. https://doi.org/10.1016/j.psep.2022.05.067

Pandey, A., Manjare, S.D., Joshi, A., Singh, A.P., 2022. Quantification of environmental impact of water pollutants using fuzzy comprehensive model. Int. J. Environ. Sci. Technol. 1–12. https://doi.org/10.1007/s13762-022-03970-x

Pandey, M., Pandey, A.K., Mishra, A., Tripathi, B.D., 2015. Assessment of metal species in river Ganga sediment at Varanasi, India using sequential extraction procedure and SEM-EDS. Chemosphere 134, 466–474. https://doi.org/10.1016/j.chemosphere.2015.04.047

Pandey, M., Tripathi, S., Pandey, A.K., Tripathi, B.D., 2014. Risk assessment of metal species in sediments of the river ganga. Catena 122, 140–149. https://doi.org/10.1016/j.catena.2014.06.012

Pardo, R., Barrado, E., Castrillejo, Y., Velasco, M.A., Vega, M., 1993. Study of the Contents and Speciation of Heavy Metals in River Sediments by Factor Analysis. Anal. Lett. 26, 1719–1739. https://doi.org/10.1080/00032719308021492

Parizanganeh, A., Lakhan, V.C., Jalalian, H., 2007. A geochemical and statistical approach for assessing heavy metal pollution in sediments from the southern Caspian coast. Int. J. Environ. Sci. Technol. 4, 351–358. https://doi.org/10.1007/BF03326293

Park, S., Choi, M., Jang, D., Joe, D., Park, K., 2020. Distribution and Sources of Dissolved and Particulate Heavy Metals (Mn, Co, Ni, Cu, Zn, Cd, Pb) in Masan Bay, Korea. Ocean Sci. J. 55, 49–67. https://doi.org/10.1007/s12601-020-0001-2

Park, S., Kazama, F., Lee, S., 2014. Assessment of water quality using multivariate statistical techniques: A case study of the Nakdong river basin, Korea. Environ. Eng. Res. 19, 197–203. https://doi.org/10.4491/eer.2014.008

Parween, M., Ramanathan, A.L., Raju, N.J., 2021. Assessment of toxicity and potential health risk from persistent pesticides and heavy metals along the Delhi stretch of river Yamuna. Environ. Res. 202, 111780. https://doi.org/10.1016/j.envres.2021.111780

Pawar, N.J., Nikumbh, J.D., 1999. Trace element geochemistry of groundwater from Behedi basin, Nasik district, Maharashtra. J. Geol. Soc. India 54, 501–514.

Pcba, 2013. City Sanitation Plan, Pollution Control Board, Assam.

Peizhuang, W., 1983. Pattern recognition with fuzzy objective function algorithms (James C. Bezdek). SIAM Rev. 25, 442–442. https://doi.org/10.1137/1025116

Perin, G., Craboledda, L., Lucchese, M., Cirillo, R., Dotta, L., Zanetta, M.L., Oro, A., 1985. Heavy metal speciation in the sediments of northern Adriatic Sea. A new approach for environmental toxicity determination. In: Lakkas TD (ed) Heavy metals in the environment, vol 2. CEP Consult. Edinbg. Poult. DJ, Morris WA, Coakley JP 110, 195–205.

Perry, H.M., Tipton, I.H., Schroeder, H.A., Steiner, R.L., Cook, M.J., 1961. Variation in the concentration of cadmium in human kidney as a function of age and geographic origin. J. Chronic Dis. 14, 259–271. https://doi.org/10.1016/0021-9681(61)90157-6

Pham, D., Pham, P., 1999. Artificial intelligence in engineering. Int. J. Mach. Tools Manuf. 39, 937–949.

Phillips, L., Moya, J., 2013. The evolution of EPA's Exposure Factors Handbook and its future

as an exposure assessment resource. J. Expo. Sci. Environ. Epidemiol. https://doi.org/10.1038/jes.2012.77

Phillips, L.J., Moya, J., 2014. Exposure factors resources: Contrasting EPA's Exposure Factors Handbook with international sources. J. Expo. Sci. Environ. Epidemiol. 24, 233–243. https://doi.org/10.1038/jes.2013.17

Poblete, T., Ortega-Farías, S., Moreno, M.A., Bardeen, M., 2017. Artificial neural network to predict vine water status spatial variability using multispectral information obtained from an unmanned aerial vehicle (UAV). Sensors (Switzerland) 17. https://doi.org/10.3390/s17112488

Poikane, S., Salas Herrero, F., Kelly, M.G., Borja, A., Birk, S., van de Bund, W., 2020. European aquatic ecological assessment methods: A critical review of their sensitivity to key pressures. Sci. Total Environ. 740. https://doi.org/10.1016/j.scitotenv.2020.140075

Polissar, A. V., Hopke, P.K., Malm, W.C., Sisler, J.F., 1996. The ratio of aerosol optical absorption coefficients to sulfur concentrations, as an indicator of smoke from forest fires when sampling in polar regions. Atmos. Environ. 30, 1147–1157. https://doi.org/10.1016/1352-2310(95)00334-7

Poulton, D.J., Paatero, P., Coakley, J.P., Morris, W.A., 1995. Evaluation of several multivariate statistical techniques... - Google Scholar, in: Proceedings of the 38th Conference of the International Association of Great Lakes Research. pp. 39–40.

Prasad, S., Saluja, R., Joshi, V., Garg, J.K., 2020. Heavy metal pollution in surface water of the Upper Ganga River, India: human health risk assessment. Environ. Monit. Assess. 192, 1–15. https://doi.org/10.1007/s10661-020-08701-8

Prati, L., Pavanello, R., Pesarin, F., 1971. Assessment of surface water quality by a single index of pollution. Water Res. 5, 741–751. https://doi.org/10.1016/0043-1354(71)90097-2

Qin, W., Han, D., Song, X., Liu, S., 2021. Sources and migration of heavy metals in a karst water system under the threats of an abandoned Pb–Zn mine, Southwest China. Environ. Pollut. 277. https://doi.org/10.1016/j.envpol.2021.116774

Qingjie, G., Jun, D., Yunchuan, X., Qingfei, W., Liqiang, Y., 2008. Calculating Pollution Indices by Heavy Metals in Ecological Geochemistry Assessment and a Case Study in Parks of Beijing. J. China Univ. Geosci. 19, 230–241. https://doi.org/10.1016/S1002-0705(08)60042-4

RAIS, 1992. The risk assessment information system [WWW Document]. RAIS. URL https://rais.ornl.gov/ (accessed 9.18.21).

Rajasekhar, B., Nambi, I.M., Govindarajan, S.K., 2020. Human health risk assessment for exposure to BTEXN in an urban aquifer using deterministic and probabilistic methods: A case study of Chennai city, India. Environ. Pollut. 265. https://doi.org/10.1016/j.envpol.2020.114814

Rajeshkumar, S., Liu, Y., Zhang, X., Ravikumar, B., Bai, G., Li, X., 2018. Studies on seasonal pollution of heavy metals in water, sediment, fish and oyster from the Meiliang Bay of Taihu Lake in China. Chemosphere 191, 626–638. https://doi.org/10.1016/j.chemosphere.2017.10.078

Ralph, K., Diane, G., 1980. Occurrence, mineral chemistry, and metamorphism of precambrian carbonate rocks in a portion of the Grenville province. J. Petrol. 21, 573–620. https://doi.org/10.1093/petrology/21.3.573

Ramos-Gómez, M., Avelar, J., Medel-Reyes, A., Yamamoto, L., Godinez, L., Ramirez, M., Guerra, R., Rodríguez, F., 2011. Movilidad de metales en jales procedentes del distrito minero de Guanajuato, México. Rev. Int. Contam. Ambient. 28, 49–59.

Ran, H., Guo, Z., Yi, L., Xiao, X., Zhang, L., Hu, Z., Li, C., Zhang, Y., 2021. Pollution characteristics and source identification of soil metal(loid)s at an abandoned arsenic-

containing mine, China. J. Hazard. Mater. 413. https://doi.org/10.1016/j.jhazmat.2021.125382

Ranjith, S., Shivapur, A. V., Shiva Keshava Kumar, P., Hiremath, C.G., Dhungana, S., 2019. Water quality evaluation in term of wqi river Tungabhadra, Karnataka, India. Int. J. Recent Technol. Eng. 8, 2065–2073. https://doi.org/10.35940/ijrte.B1202.0982S1119

Ranković, V., Radulović, J., Radojević, I., Ostojić, A., Čomić, L., 2012. Prediction of dissolved oxygen in reservoirs using adaptive network-based fuzzy inference system. J. Hydroinformatics 14, 167–179. https://doi.org/10.2166/hydro.2011.084

Rasaei, Z., Bogaert, P., 2019. Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. Geoderma 344, 50–62. https://doi.org/10.1016/j.geoderma.2019.02.031

Rate, A.W., Robertson, A.E., Borg, A.T., 2000. Distribution of heavy metals in near-shore sediments of the Swan River estuary, Western Australia. Water. Air. Soil Pollut. 124, 155–168. https://doi.org/10.1023/A:1005289203825

Reghunath, R., Murthy, T., Raghavan, B., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. Water Res. 36, 2437–42.

Roach, A.C., 2005. Assessment of metals in sediments from Lake Macquarie, New South Wales, Australia, using normalisation models and sediment quality guidelines. Mar. Environ. Res. 59, 453–472. https://doi.org/10.1016/j.marenvres.2004.07.002

Rode, M., Suhr, U., 2007. Uncertainties in selected river water quality data. Hydrol. Earth Syst. Sci. 11, 863–874. https://doi.org/10.5194/hess-11-863-2007

Romano, D., Bernetti, A., De Lauretis, R., 2004. Different methodologies to quantify uncertainties of air emissions. Environ. Int. 30, 1099–1107. https://doi.org/10.1016/j.envint.2004.06.006

Rooki, R., Ardejani, F.D., Aryafar, A., Asadi, A.B., 2011. Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran. Environ. Earth Sci. 64, 1303–1316. https://doi.org/10.1007/s12665-011-0948-5

Ross, T.J., 2010. Fuzzy logic with engineering applications, Third Edit. ed, John Wiley & Sons. https://doi.org/10.1002/9781119994374

Roudposhti, G.M., Karbassi, A., Baghvand, A., 2016. A pollution index for agricultural soils. Arch. Agron. Soil Sci. 62, 1411–1424. https://doi.org/10.1080/03650340.2016.1154542

Ruppert, D., 2004. The elements of statistical learning: data mining, inference, and prediction. J. Am. Stat. Assoc. 99, 567–567. https://doi.org/10.1198/jasa.2004.s339

Sadeghi, S.H.R., Kiani Harchegani, M., Younesi, H.A., 2012. Suspended sediment concentration and particle size distribution, and their relationship with heavy metal content. J. Earth Syst. Sci. 121, 63–71. https://doi.org/10.1007/s12040-012-0143-4

Sadiq, R., Rodriguez, M.J., 2004. Fuzzy synthetic evaluation of disinfection by-products - A risk-based indexing system. J. Environ. Manage. 73, 1–13. https://doi.org/10.1016/j.jenvman.2004.04.014

Saha, N., Rahman, M.S., 2020. Groundwater hydrogeochemistry and probabilistic health risk assessment through exposure to arsenic-contaminated groundwater of Meghna floodplain, central-east Bangladesh. Ecotoxicol. Environ. Saf. 206. https://doi.org/10.1016/j.ecoenv.2020.111349

Saha, N., Rahman, M.S., Ahmed, M.B., Zhou, J.L., Ngo, H.H., Guo, W., 2017. Industrial metal pollution in water and probabilistic assessment of human health risk. J. Environ. Manage. 185, 70–78. https://doi.org/10.1016/J.JENVMAN.2016.10.023

Sahuquillo, A., Rigol, A., Rauret, G., 2003. Overview of the use of leaching/extraction tests for

risk assessment of trace metals in contaminated soils and sediments. TrAC - Trends Anal. Chem. https://doi.org/10.1016/S0165-9936(03)00303-0

Sakan, S.M., Dordević, D.S., Manojlović, D.D., Predrag, P.S., 2009. Assessment of heavy metal pollutants accumulation in the Tisza river sediments. J. Environ. Manage. 90, 3382–3390. https://doi.org/10.1016/j.jenvman.2009.05.013

Salim, I., Sajjad, R.U., Paule-Mercado, M.C., Memon, S.A., Lee, B.Y., Sukhbaatar, C., Lee, C.H., 2019. Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. Sci. Total Environ. 663, 764–775. https://doi.org/10.1016/j.scitotenv.2019.01.377

Salomons, W., Forstner, U., 2012. Metals in the hydrocycle. Springer Sci. Bus. Media. https://doi.org/10.2307/2403266

Salomons, W., Förstner, U., 1980. Trace metal analysis on polluted sediments: Part II: Evaluation of environmental impact. Environ. Technol. Lett. 1, 506–517. https://doi.org/10.1080/09593338009384007

Sawyer, C.N., McCarty, P.L., Parkin, G.F., 1978. Chemistry for environmental engineers. McGraw-Hill.

Schlosser, I.J., Karr, J.R., 1981. Riparian vegetation and channel morphology impact on spatial patterns of water quality in agricultural watersheds. Environ. Manage. 5, 233–243. https://doi.org/10.1007/BF01873282

Sebei, A., Helali, M.A., Oueslati, W., Abdelmalek-Babbou, C., Chaabani, F., 2018. Bioavailability of Pb, Zn, Cu, Cd, Ni and Cr in the sediments of the Tessa River: A mining area in the North-West Tunisia. J. African Earth Sci. 137, 1–8. https://doi.org/10.1016/j.jafrearsci.2017.09.005

Selvam, S., Jesuraja, K., Roy, P.D., Venkatramanan, S., Khan, R., Shukla, S., Manimaran, D., Muthukumar, P., 2022. Human health risk assessment of heavy metal and pathogenic contamination in surface water of the Punnakayal estuary, South India. Chemosphere 298, 134027. https://doi.org/10.1016/J.CHEMOSPHERE.2022.134027

Semirom, F.B., Hassan, A.H., Torabia, A., Karbass, A.R., Hosseinzadeh Lotf, F., 2011. Water quality index development using fuzzy logic: A case study of the Karoon River of Iran. African J. Biotechnol. 10, 10125–10133. https://doi.org/10.5897/ajb11.1608

Seppä, L., Kärkkäinen, S., Hausen, H., 1998. Caries frequency in permanent teeth before and after discontinuation of water fluoridation in Kuopio, Finland. Community Dent. Oral Epidemiol. 26, 256–262. https://doi.org/10.1111/j.1600-0528.1998.tb01959.x

Sergeant, C.J., Starkey, E.N., Bartz, K.K., Wilson, M.H., Mueter, F.J., 2016. A practitioner's guide for exploring water quality patterns using principal components analysis and procrustes. Environ. Monit. Assess. 188. https://doi.org/10.1007/s10661-016-5253-z

Setia, R., Dhaliwal, S.S., Kumar, V., Singh, R., Kukal, S.S., Pateriya, B., 2020. Impact assessment of metal contamination in surface water of Sutlej River (India) on human health risks. Environ. Pollut. 265, 114907. https://doi.org/10.1016/j.envpol.2020.114907

Sharafati, A., Khosravi, K., Khosravinia, P., Ahmed, K., Salman, S.A., Yaseen, Z.M., Shahid, S., 2019. The potential of novel data mining models for global solar radiation prediction. Int. J. Environ. Sci. Technol. 16, 7147–7164. https://doi.org/10.1007/s13762-019-02344-0

Sheng, D., Wen, X., Wu, J., Wu, M., Yu, H., Zhang, C., 2021. Comprehensive probabilistic health risk assessment for exposure to arsenic and cadmium in groundwater. Environ. Manage. 67, 779–792. https://doi.org/10.1007/s00267-021-01431-8

Shi, G.L., Zeng, F., Li, X., Feng, Y.C., Wang, Y.Q., Liu, G.X., Zhu, T., 2011. Estimated contributions and uncertainties of PCA/MLR-CMB results: Source apportionment for synthetic and ambient datasets. Atmos. Environ. 45, 2811–2819.

https://doi.org/10.1016/j.atmosenv.2011.03.007

Shi, W., Li, T., Feng, Y., Su, H., Yang, Q., 2022. Source apportionment and risk assessment for available occurrence forms of heavy metals in Dongdahe Wetland sediments, southwest of China. Sci. Total Environ. 815, 152837. https://doi.org/10.1016/j.scitotenv.2021.152837

Shihabudheen, K. V., Pillai, G.N., 2018. Recent advances in neuro-fuzzy system: A survey. Knowledge-Based Syst. 152, 136–162. https://doi.org/10.1016/j.knosys.2018.04.014

Shil, S., Singh, U.K., 2019. Health risk assessment and spatial variations of dissolved heavy metals and metalloids in a tropical river basin system. Ecol. Indic. 106, 105455. https://doi.org/10.1016/j.ecolind.2019.105455

Shin, J.H., Jo, D.H., Kim, Y., 2021. Mobility and source apportionment of As and heavy metals in the Taehwa River sediment, South Korea: anthropogenic and seasonal effects. Environ. Earth Sci. 80, 1–12. https://doi.org/10.1007/s12665-021-09371-6

Shrestha, S., Kazama, F., 2006. Multivariate statistical techniques for the assessment of surface water quality of Fuji River Basin, Japan, in: Water Science and Technology: Water Supply. pp. 59–67. https://doi.org/10.2166/ws.2006.802

Sibley, D.F., 1979. The chemistry of the atmosphere and oceans. Geochim. Cosmochim. Acta 43, 179. https://doi.org/10.1016/0016-7037(79)90059-0

Siddique, M.A.B., Islam, A.R.M.T., Hossain, M.S., Khan, R., Akbor, M.A., Hasanuzzaman, M., Sajid, M.W.M., Mia, M.Y., Mallick, J., Rahman, M.S., Rahman, M.M., Bodrud-Doza, M., 2021. Multivariate statistics and entropy theory for irrigation water quality and entropy-weighted index development in a subtropical urban river, Bangladesh. Environ. Sci. Pollut. Res. 1–20. https://doi.org/10.1007/s11356-021-16343-7

Silvert, W., 2000. Fuzzy indices of environmental conditions. Ecol. Modell. 130, 111–119. https://doi.org/10.1016/S0304-3800(00)00204-0

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M., Kouimtzis, T., 2003. Assessment of the surface water quality in Northern Greece. Water Res. 37, 4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

Singh, A., Sharma, R.K., Agrawal, M., Marshall, F.M., 2010. Health risk assessment of heavy metals via dietary intake of foodstuffs from the wastewater irrigated site of a dry tropical area of India. Food Chem. Toxicol. 48, 611–619. https://doi.org/10.1016/j.fct.2009.11.041

Singh, A.P., Dhadse, K., Ahalawat, J., 2019. Managing water quality of a river using an integrated geographically weighted regression technique with fuzzy decision-making model. Environ. Monit. Assess. 191, 1–17. https://doi.org/10.1007/s10661-019-7487-z

Singh, D.D., Thind, P.S., Sharma, M., Sahoo, S., John, S., 2019. Environmentally sensitive elements in groundwater of an industrial town in India: Spatial distribution and human health risk. Water (Switzerland) 11. https://doi.org/10.3390/w11112350

Singh, J., Kalamdhad, A.S., 2013. Chemical Speciation of Heavy Metals in Compost and Compost Amended Soil -A Review. Int. J. Environ. Eng. Res. 2, 27–37.

Singh, K., Malik, A., Sinha, S., 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. Anal. Chim. Acta 538, 74–355.

Singh, K.P., Basant, A., Malik, A., Jain, G., 2009. Artificial neural network modeling of the river water quality-A case study. Ecol. Modell. 220, 888–895. https://doi.org/10.1016/j.ecolmodel.2009.01.004

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 Res. 38, 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Singh, K.P., Mohan, D., Singh, V.K., Malik, A., 2005. Studies on distribution and fractionation of heavy metals in Gomti river sediments - A tributary of the Ganges, India. J. Hydrol. 312, 14–27. https://doi.org/10.1016/j.jhydrol.2005.01.021

Singh, K.R., Dutta, R., Kalamdhad, A.S., Kumar, B., 2019. Information entropy as a tool in surface water quality assessment. Environ. Earth Sci. 78, 15. https://doi.org/10.1007/s12665-018-7998-x

Singh, K.R., Goswami, A.P., Kalamdhad, A.S., Kumar, B., 2021a. Water quality evaluation and apportionment of pollution sources: A case study of the Baralia and Puthimari river (India). Water Pract. Technol. 16, 692–706. https://doi.org/10.2166/wpt.2021.020

Singh, K.R., Goswami, A.P., Kalamdhad, A.S., Kumar, B., 2020a. Assessment of surface water quality of Pagladia, Beki and Kolong rivers (Assam, India) using multivariate statistical techniques. Int. J. River Basin Manag. 18, 511–520. https://doi.org/10.1080/15715124.2019.1566236

Singh, K.R., Goswami, A.P., Kalamdhad, A.S., Kumar, B., 2020b. Surface water quality and health risk assessment of kameng river (Assam, india). Water Pract. Technol. 15, 1190–1201. https://doi.org/10.2166/wpt.2020.090

Singh, K.R., Kalamdhad, A.S., Kumar, B., 2021b. Source apportionment for spatial variation of surface water quality using chemometric techniques. Environ. Forensics. https://doi.org/10.1080/15275922.2021.1913675

Singh, M., Müller, G., Singh, I.B., 2002. Heavy metals in freshly deposited stream sediments of rivers associated with urbanisation of the Ganga Plain, India. Water. Air. Soil Pollut. https://doi.org/10.1023/A:1021339917643

Singh, V.P., 2015. Entropy theory in hydraulic engineering: An introduction, ASCE Press. American Society of Civil Engineers (ASCE). https://doi.org/10.1061/9780784412725

Singh, V.P., 2013. Entropy theory and its application in environmental and water engineering. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118428306

Siraj, K., Kitte, S.A., 2013. Analysis of copper, zinc and lead using atomic absorption spectrophotometer in ground water of Jimma town of Southwestern Ethiopia. Int. J. Chem. Anal. Sci. 4, 201–204. https://doi.org/10.1016/j.ijcas.2013.07.006

Smith, D.G., 1990. A better water quality indexing system for rivers and streams. Water Res. 24, 1237–1244. https://doi.org/10.1016/0043-1354(90)90047-A

Soumaila, K.I., Niandou, A.S., Naimi, M., Mohamed, C., Schimmel, K., Luster-Teasley, S., Sheick, N.N., 2019. A systematic review and meta-analysis of water quality indices. J. Agric. Sci. Technol. B 9, 1–14. https://doi.org/10.17265/2161-6264/2019.01.001

Sreenivasa Rao, K., Balaji, T., Prasada Rao, T., Babu, Y., Naidu, G.R.K., 2002. Determination of iron, cobalt, nickel, manganese, zinc, copper, cadmium and lead in human hair by inductively coupled plasma-atomic emission spectrometry. Spectrochim. Acta - Part B At. Spectrosc. 57, 1333–1338. https://doi.org/10.1016/S0584-8547(02)00045-9

Stanly, R., Yasala, S., Oliver, D.H., Nair, N.C., Emperumal, K., Subash, A., 2021. Hydrochemical appraisal of groundwater quality for drinking and irrigation: a case study in parts of southwest coast of Tamil Nadu, India. Appl. Water Sci. 11. https://doi.org/10.1007/s13201-021-01381-w

Stoichev, T., Coelho, J.P., De Diego, A., Valenzuela, M.G.L., Pereira, M.E., de Chanvalon, A.T., Amouroux, D., 2020. Multiple regression analysis to assess the contamination with metals and metalloids in surface sediments (Aveiro Lagoon, Portugal). Mar. Pollut. Bull. 159, 111470. https://doi.org/10.1016/j.marpolbul.2020.111470

Stoner, J.D., 1978. Water-quality indices for specific water uses, US Geol Surv Circ.

Su, S., Xiao, R., Mi, X., Xu, X., Zhang, Z., Wu, J., 2013. Spatial determinants of hazardous

chemicals in surface water of Qiantang River, China. Ecol. Indic. 24, 375–381. https://doi.org/10.1016/j.ecolind.2012.07.015

Sun, C., Zhang, Z., Cao, H., Xu, M., Xu, L., 2019. Concentrations, speciation, and ecological risk of heavy metals in the sediment of the Songhua River in an urban area with petrochemical industries. Chemosphere 219, 538–545. https://doi.org/10.1016/j.chemosphere.2018.12.040

Sun, R., Gao, Y., Xu, J., Yang, Y., Zhang, Y., 2021. Contamination Features and Source Apportionment of Heavy Metals in the River Sediments around a Lead-Zinc Mine: A Case Study in Danzhai, Guizhou, China. J. Chem. 2021. https://doi.org/10.1155/2021/9946026

Sundaray, S.K., Nayak, B.B., Lin, S., Bhatta, D., 2011. Geochemical speciation and risk assessment of heavy metals in the river estuarine sediments-A case study: Mahanadi basin, India. J. Hazard. Mater. 186, 1837–1846. https://doi.org/10.1016/j.jhazmat.2010.12.081

Sutadian, A.D., Muttil, N., Yilmaz, A.G., Perera, B.J.C., 2018. Development of a water quality index for rivers in West Java Province, Indonesia. Ecol. Indic. 85, 966–982. https://doi.org/10.1016/j.ecolind.2017.11.049

Sutadian, A.D., Muttil, N., Yilmaz, A.G., Perera, B.J.C., 2016. Development of river water quality indices—a review. Environ. Monit. Assess. 188, 1–29. https://doi.org/10.1007/s10661-015-5050-0

Swarnalatha, K., Nair, A.G., 2017. Assessment of sediment quality of a tropical lake using sediment quality standards. Lakes Reserv. Res. Manag. 22, 65–73. https://doi.org/10.1111/lre.12162

Sylaios, G., Kamidis, N., Stamatis, N., 2012. Assessment of trace metals contamination in the suspended matter and sediments of a semi-enclosed Mediterranean Gulf. Soil Sediment Contam. 21, 673–700. https://doi.org/10.1080/15320383.2012.691128

Takagi, H., Hayashi, I., 1991. NN-driven fuzzy reasoning. Int. J. Approx. Reason. 5, 191–212.

Tang, W., Zhang, C., Zhao, Y., Shan, B., Song, Z., 2017. Pollution, toxicity, and ecological risk of heavy metals in surface river sediments of a large basin undergoing rapid economic development. Environ. Toxicol. Chem. 36, 1149–1155. https://doi.org/10.1002/etc.3650

Tao, H., Keshtegar, B., Yaseen, Z.M., 2019. The feasibility of integrative radial basis M5Tree predictive model for river suspended sediment load simulation. Water Resour. Manag. 33, 4471–4490. https://doi.org/10.1007/s11269-019-02378-6

Tatone, L.M., Bilos, C., Skorupka, C.N., Colombo, J.C., 2016. Comparative approach for trace metal risk evaluation in settling particles from the Uruguay River, Argentina: enrichment factors, sediment quality guidelines and metal speciation. Environ. Earth Sci. 75, 1–7. https://doi.org/10.1007/s12665-016-5265-6

Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy metal toxicity and the environment. EXS. https://doi.org/10.1007/978-3-7643-8340-4_6

Teodorof, L., Burada, A., Despina, C., Seceleanu-Odor, D., Spiridon, C., Ţigănuş, M., Tudor, M.I., Tudor, M., Ene, A., Zubcov, E., Spanos, T., Bogdevich, O., 2020. Sediments quality assessment in terms of single and integrated indices from Romanian MONITOX network (2019 – 2020). Ann. "Dunarea Jos" Univ. Galati. Fascicle II, Math. Physics, Theor. Mech. 43, 175–183. https://doi.org/10.35219/ann-ugal-math-phys-mec.2020.2.16

Tessier, A., Campbell, P.G.C., Bisson, M., 1979. Sequential extraction procedure for the speciation of particulate trace metals. Anal. Chem. 51, 844–851. https://doi.org/10.1021/ac50043a017

Thi Minh Hanh, P., Sthiannopkao, S., The Ba, D., Kim, K.-W., 2011. Development of water quality indexes to identify pollutants in Vietnam's surface water. J. Environ. Eng. 137, 273–283. https://doi.org/10.1061/(asce)ee.1943-7870.0000314

Thorslund, J., Jarsjö, J., Wällstedt, T., Mörth, C.M., Lychagin, M.Y., Chalov, S.R., 2017. Speciation and hydrological transport of metals in non-acidic river systems of the Lake Baikal basin: Field data and model predictions. Reg. Environ. Chang. 17, 2007–2021. https://doi.org/10.1007/s10113-016-0982-7

Tiyasha, Tung, T.M., Yaseen, Z.M., 2020. A survey on river water quality modelling using artificial intelligence models: 2000–2020. J. Hydrol. 585, 124670. https://doi.org/10.1016/J.JHYDROL.2020.124670

Tomlinson, D.L., Wilson, J.G., Harris, C.R., Jeffrey, D.W., 1980. Problems in the assessment of heavy-metal levels in estuaries and the formation of a pollution index. Helgoländer Meeresuntersuchungen 33, 566–575. https://doi.org/10.1007/BF02414780

Tong, R., Cheng, M., Zhang, L., Liu, M., Yang, X., Li, X., Yin, W., 2018. The construction dust-induced occupational health risk using Monte-Carlo simulation. J. Clean. Prod. 184, 598–608. https://doi.org/10.1016/j.jclepro.2018.02.286

Turksen, I.B., 1991. Measurement of membership functions and their acquisition. Fuzzy Sets Syst. 40, 5–38. https://doi.org/10.1016/0165-0114(91)90045-R

Uddin, M.G., Nash, S., Olbert, A.I., 2021. A review of water quality index models and their use for assessing surface water quality. Ecol. Indic. https://doi.org/10.1016/j.ecolind.2020.107218

US EPA, 2005. Ecological soil screening level for iron interim final. US Environ. Prot. Agency - Off. Solid Waste Emerg. 211.

US EPA, 2004. Risk assessment guidance for superfund. volume i: Human health evaluation manual — part E, supplemental guidance for dermal risk assessment. https://doi.org/[EPA/540/R/99/005. OSWER 9285. 7-02EP PB99-963312]

US EPA, 2001. Risk assessment guidance for superfund: volume III-part A, process for conducting probabilistic risk assessment, Office of Emergency and Remedial Response. https://doi.org/EPA 540-R-02-002 OSWER 9285.7-45 PB2002 963302

US EPA, 1992. Guidelines for exposure assessment. Risk assessment forum.

US EPA, 1989. Risk assessment guidance for superfund (RAGS): volume I. Human health evaluation manual (HHEM) — part A, baseline risk assessment. https://doi.org/[EPA/540/1-89/002]

US EPA, EPA, U.S., 1991. Risk Assessment Guidance for Superfund, Human Health Evaluation Manual (Part B, Development of Risk-based Preliminary Remediation Goals).

US EPA, O., 2002. Risk Assessment Guidance.

USEPA, 2011. Exposure Factors Handbook.

USEPA, 2005. Guidelines for Carcinogen Risk Assessment. Washington, DC, USA.

Utgikar, V.P., Chaudhary, N., Koeniger, A., Tabak, H.H., Haines, J.R., Govind, R., 2004. Toxicity of metals and metal mixtures: Analysis of concentration and time dependence for zinc and copper. Water Res. 38, 3651–3658. https://doi.org/10.1016/j.watres.2004.05.022

Uzun, H., Yıldız, Z., Goldfarb, J.L., Ceylan, S., 2017. Improved prediction of higher heating value of biomass using an artificial neural network model based on proximate analysis. Bioresour. Technol. 234, 122–130. https://doi.org/10.1016/j.biortech.2017.03.015

Vadde, K.K., Wang, J., Cao, L., Yuan, T., McCarthy, A.J., Sekar, R., 2018. Assessment of water quality and identification of pollution risk locations in Tiaoxi River (Taihu Watershed), China. Water (Switzerland) 10. https://doi.org/10.3390/w10020183

Vadiati, M., Asghari-Moghaddam, A., Nakhaei, M., Adamowski, J., Akbarzadeh, A.H., 2016. A fuzzy-logic based decision-making approach for identification of groundwater quality based on groundwater quality indices. J. Environ. Manage. 184, 255–270. https://doi.org/10.1016/j.jenvman.2016.09.082

van der Weijden, C.H., 1976. Experiments on the uptake of zinc and cadmium by manganese oxides. Mar. Chem. 4, 377–387. https://doi.org/10.1016/0304-4203(76)90022-0

van der Weijden, C.H., Arnoldus, M.J.H.L., Meurs, C.J., 1977. Desorption of metals from suspended material in the rhine estuary. Netherlands J. Sea Res. https://doi.org/10.1016/0077-7579(77)90002-3

Varol, M., Gökot, B., Bekleyen, A., Şen, B., 2012. Water quality assessment and apportionment of pollution sources of Tigris River (Turkey) using multivariate statistical techniques-a case study. River Res. Appl. 28, 1428–1438. https://doi.org/10.1002/rra.1533

Vemula, V.R.S., Mujumdar, P.P., Ghosh, S., 2004. Risk evaluation in water quality management of a river system. J. Water Resour. Plan. Manag. 130, 411–423. https://doi.org/10.1061/(asce)0733-9496(2004)130:5(411)

Verma, K., Pandey, J., 2019. Heavy metal accumulation in surface sediments of the Ganga River (India): speciation, fractionation, toxicity, and risk assessment. Environ. Monit. Assess. 191, 1–21. https://doi.org/10.1007/s10661-019-7552-7

Vieira, R., Fernandes, J.N., Barbosa, A.E., 2013. Evaluation of the impacts of road runoff in a Mediterranean reservoir in Portugal. Environ. Monit. Assess. 185, 7659–7673. https://doi.org/10.1007/s10661-013-3126-2

Wagh, V.M., Panaskar, D.B., Mukate, S.V., Gaikwad, S.K., Muley, A.A., Varade, A.M., 2018. Health risk assessment of heavy metal contamination in groundwater of Kadava River Basin, Nashik, India. Model. Earth Syst. Environ. 4, 969–980. https://doi.org/10.1007/s40808-018-0496-z

Walling, D.E., 2009. The Impact of global change on erosion and sediment transport by rivers: current progress and future challenges, in: United Nations Educational, Scientific and Cultural Organization. pp. 1–23.

Wan, X., Dong, H., Feng, L., Lin, Z., Luo, Q., 2017. Comparison of three sequential extraction procedures for arsenic fractionation in highly polluted sites. Chemosphere 178, 402–410. https://doi.org/10.1016/j.chemosphere.2017.03.078

Wang, G., Cheng, S., Wei, W., Zhou, Y., Yao, S., Zhang, H., 2016. Characteristics and source apportionment of VOCs in the suburban area of Beijing, China. Atmos. Pollut. Res. 7, 711–724. https://doi.org/10.1016/j.apr.2016.03.006

Wang, J., Liu, G., Liu, H., Lam, P.K.S., 2017. Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China. Sci. Total Environ. 583, 421–431. https://doi.org/10.1016/j.scitotenv.2017.01.088

Wang, J., Liu, R., Wang, H., Yu, W., Xu, F., Shen, Z., 2015. Identification and apportionment of hazardous elements in the sediments in the Yangtze River estuary. Environ. Sci. Pollut. Res. 22, 20215–20225. https://doi.org/10.1007/s11356-015-5642-9

Wang, M., Hu, K., Zhang, D., Lai, J., 2019. Speciation and spatial distribution of heavy metals (Cu and Zn) in wetland soils of Poyang Lake (China) in wet seasons. Wetlands 39, 89–98. https://doi.org/10.1007/s13157-017-0917-1

Wang, P., Li, Z., Liu, J., Bi, X., Ning, Y., Yang, S., Yang, X., 2019. Apportionment of sources of heavy metals to agricultural soils using isotope fingerprints and multivariate statistical analyses. Environ. Pollut. 249, 208–216. https://doi.org/10.1016/j.envpol.2019.03.034

Wang, Shaofeng, Jia, Y., Wang, Shuying, Wang, X., Wang, H., Zhao, Z., Liu, B., 2010. Fractionation of heavy metals in shallow marine sediments from Jinzhou Bay, China. J. Environ. Sci. 22, 23–31. https://doi.org/10.1016/S1001-0742(09)60070-X

Wang, Y., Liu, R. hai, Zhang, Y. qing, Cui, X. qing, Tang, A. kun, Zhang, L. jun, 2016. Transport of heavy metals in the Huanghe River estuary, China. Environ. Earth Sci. 75, 1–11. https://doi.org/10.1007/s12665-015-4908-3

Wang, Y., Wang, R., Fan, L., Chen, T., Bai, Y., Yu, Q., Liu, Y., 2017. Assessment of multiple exposure to chemical elements and health risks among residents near Huodehong lead-zinc mining area in Yunnan, Southwest China. Chemosphere 174, 613–627. https://doi.org/10.1016/j.chemosphere.2017.01.055

Ward, J.H., 1963. Ward's method. J. Am. Stat. Assoc. 58, 236–246.

Wątor, K., Zdechlik, R., 2021. Application of water quality indices to the assessment of the effect of geothermal water discharge on river water quality – case study from the Podhale region (Southern Poland). Ecol. Indic. 121. https://doi.org/10.1016/j.ecolind.2020.107098

Weimin, Y., Batley, G.E., Ahsanullah, M., 1994. Metal bioavailability to the soldier crab Mictyris longicarpus. Sci. Total Environ. 141, 27–44. https://doi.org/10.1016/0048-9697(94)90015-9

Welling, R., Beaumont, J.J., Petersen, S.J., Alexeeff, G. V, Steinmaus, C., 2015. Chromium VI and stomach cancer: A meta-analysis of the current epidemiological evidence. Occup. Environ. Med. 72, 151–159. https://doi.org/10.1136/oemed-2014-102178

Wen, C., Zhu, S., Li, N., Luo, X., 2022. Source apportionment and risk assessment of metal pollution in natural biofilms and surface water along the Lancang River, China. Sci. Total Environ. 843, 156977. https://doi.org/10.1016/J.SCITOTENV.2022.156977

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028

WHO, 2013. WHO guidelines on good agricultural and collection practices (GACP).

Winton, E., Tardiff, R., Mccab, L., 1971. Nitrate in drinking water. Am. Water Work. Assoc. 63, 95–98. https://doi.org/10.1002/j.1551-8833.1971.tb04035.x

Wu, B., Zhang, Y., Zhang, X., Cheng, S., 2010. Health risk from exposure of organic pollutants through drinking water consumption in Nanjing, China. Bull. Environ. Contam. Toxicol. 84, 46–50. https://doi.org/10.1007/s00128-009-9900-8

Wu, H., Xu, C., Wang, J., Xiang, Y., Ren, M., Qie, H., Zhang, Y., Yao, R., Li, L., Lin, A., 2021. Health risk assessment based on source identification of heavy metals: A case study of Beiyun River, China. Ecotoxicol. Environ. Saf. 213, 112046. https://doi.org/10.1016/j.ecoenv.2021.112046

Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. Environ. Model. Softw. https://doi.org/10.1016/j.envsoft.2013.12.016

Wu, Z., He, M., Lin, C., 2012. Environmental impacts of heavy metals (Co, Cu, Pb, Zn) in surficial sediments of estuary in Daliao River and Yingkou Bay (northeast China): Concentration level and chemical fraction. Environ. Earth Sci. 66, 2417–2430. https://doi.org/10.1007/s12665-011-1466-1

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 Res. 35, 2881–2894. https://doi.org/10.1016/S0043-1354(00)00592-3

Xia, F., Fan, T., Chen, Y., Ding, D., Wei, J., Jiang, D., Deng, S., 2022. Prediction of Heavy Metal Concentrations in Contaminated Sites from Portable X-ray Fluorescence Spectrometer Data Using Machine Learning. Processes 10, 536. https://doi.org/10.3390/pr10030536

Xia, F., Zhang, C., Qu, L., Song, Q., Ji, X., Mei, K., Dahlgren, R.A., Zhang, M., 2020. A comprehensive analysis and source apportionment of metals in riverine sediments of a rural-urban watershed. J. Hazard. Mater. 381.

https://doi.org/10.1016/j.jhazmat.2019.121230

Xiao, H., Leng, X., Qian, X., Li, S., Liu, Y., Liu, X., Li, H., 2022. Prediction of heavy metals in airborne fine particulate matter using magnetic parameters by machine learning from a metropolitan city in China. Atmos. Pollut. Res. 13, 101347. https://doi.org/10.1016/j.apr.2022.101347

Xiao, R., Bai, J., Lu, Q., Zhao, Q., Gao, Z., Wen, X., Liu, X., 2015. Fractionation, transfer, and ecological risks of heavy metals in riparian and ditch wetlands across a 100-year chronosequence of reclamation in an estuary of China. Sci. Total Environ. 517, 66–75. https://doi.org/10.1016/j.scitotenv.2015.02.052

Xu, L., Wang, T., Wang, J., Lu, A., 2017. Occurrence, speciation and transportation of heavy metals in 9 coastal rivers from watershed of Laizhou Bay, China. Chemosphere 173, 61–68. https://doi.org/10.1016/j.chemosphere.2017.01.046

Xu, Y., Shi, H., Fei, Y., Wang, C., Mo, L., Shu, M., 2021. Article identification of soil heavy metal sources in a large-scale area affected by industry. Sustain. 13, 1–18. https://doi.org/10.3390/su13020511

Xu, Z., Lu, Q., Xu, X., Feng, X., Liang, L., Liu, L., Li, C., Chen, Z., Qiu, G., 2020. Multi-pathway mercury health risk assessment, categorization and prioritization in an abandoned mercury mining area: A pilot study for implementation of the Minamata Convention. Chemosphere 260. https://doi.org/10.1016/j.chemosphere.2020.127582

Yager, R.R., Filev, D.P., 1994. Generation of fuzzy rules by mountain clustering. J. Intell. Fuzzy Syst. 2, 209–219. https://doi.org/10.3233/IFS-1994-2301

Yan, D., Wu, S., Zhou, S., Tong, G., Li, F., Wang, Y., Li, B., 2019. Characteristics, sources and health risk assessment of airborne particulate PAHs in Chinese cities: A review. Environ. Pollut. https://doi.org/10.1016/j.envpol.2019.02.068

Yang, D., Liu, J., Wang, Q., Hong, H., Zhao, W., Chen, S., Yan, C., Lu, H., 2019. Geochemical and probabilistic human health risk of chromium in mangrove sediments: A case study in Fujian, China. Chemosphere 233, 503–511. https://doi.org/10.1016/j.chemosphere.2019.05.245

Yang, F., Geng, D., Wei, C., Ji, H., Xu, H., 2016. Distribution of arsenic between the particulate and aqueous phases in surface water from three freshwater lakes in China. Environ. Sci. Pollut. Res. 23, 7452–7461. https://doi.org/10.1007/s11356-015-5998-x

Yang, J., Ma, S., Zhou, J., Song, Y., Li, F., 2018. Heavy metal contamination in soils and vegetables and health risk assessment of inhabitants in Daye, China. J. Int. Med. Res. 46, 3374–3387. https://doi.org/10.1177/0300060518758585

Yang, W., Zhao, Y., Wang, D., Wu, H., Lin, A., He, L., 2020. Using principal components analysis and idw interpolation to determine spatial and temporal changes of Surfacewater quality of Xin'Anjiang river in huangshan, china. Int. J. Environ. Res. Public Health 17. https://doi.org/10.3390/ijerph17082942

Yang, Z., Zou, L., Xia, J., Qiao, Y., Bai, F., Wang, Q., Cai, D., 2022. Spatiotemporal variation characteristics and source identification of water pollution: Insights from urban water system. Ecol. Indic. 139, 108892. https://doi.org/10.1016/j.ecolind.2022.108892

Yao, X., Luo, K., Niu, Y., Li, Y., Ren, B., 2021. Ecological Risk from Toxic Metals in Sediments of the Yangtze, Yellow, Pearl, and Liaohe Rivers, China. Bull. Environ. Contam. Toxicol. 107, 140–146. https://doi.org/10.1007/s00128-021-03229-0

Yaseen, Z.M., Ramal, M.M., Diop, L., Jaafar, O., Demir, V., Kisi, O., 2018. Hybrid adaptive neuro-fuzzy models for water quality index estimation. Water Resour. Manag. 32, 2227–2245. https://doi.org/10.1007/s11269-018-1915-7

Yin, H., Gao, Y., Fan, C., 2011. Distribution, sources and ecological risk assessment of heavy metals in surface sediments from Lake Taihu, China. Environ. Res. Lett. 6.

https://doi.org/10.1088/1748-9326/6/4/044012

Yuan, F., Chaffin, J.D., Xue, B., Wattrus, N., Zhu, Y., Sun, Y., 2018. Contrasting sources and mobility of trace metals in recent sediments of western Lake Erie. J. Great Lakes Res. 44, 1026–1034. https://doi.org/10.1016/j.jglr.2018.07.016

Yuan, X., Zhang, L., Li, J., Wang, C., Ji, J., 2014. Sediment properties and heavy metal pollution assessment in the river, estuary and lake environments of a fluvial plain, China. Catena 119, 52–60. https://doi.org/10.1016/j.catena.2014.03.008

Zadeh, L.A., 1999. Fuzzy logic = computing with words. Physica, Heidelberg, pp. 3–23. https://doi.org/10.1007/978-3-7908-1873-4_1

Zadeh, L.A., 1965. Fuzzy sets. Inf. Control 8, 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhang, C., Shan, B., Tang, W., Dong, L., Zhang, W., Pei, Y., 2017. Heavy metal concentrations and speciation in riverine sediments and the risks posed in three urban belts in the Haihe Basin. Ecotoxicol. Environ. Saf. 139, 263–271. https://doi.org/10.1016/j.ecoenv.2017.01.047

Zhang, G., Bai, J., Xiao, R., Zhao, Q., Jia, J., Cui, B., Liu, X., 2017. Heavy metal fractions and ecological risk assessment in sediments from urban, rural and reclamation-affected rivers of the Pearl River Estuary, China. Chemosphere 184, 278–288. https://doi.org/10.1016/j.chemosphere.2017.05.155

Zhang, H., Cheng, S., Li, H., Fu, K., Xu, Y., 2020. Groundwater pollution source identification and apportionment using PMF and PCA-APCA-MLR receptor models in a typical mixed land-use area in Southwestern China. Sci. Total Environ. 741, 140383. https://doi.org/10.1016/J.SCITOTENV.2020.140383

Zhang, H., Yin, A., Yang, X., Fan, M., Shao, S., Wu, J., Wu, P., Zhang, M., Gao, C., 2021. Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. Ecol. Indic. 122, 107233. https://doi.org/10.1016/j.ecolind.2020.107233

Zhang, X., Wei, S., Sun, Q., Wadood, S.A., Guo, B., 2018. Source identification and spatial distribution of arsenic and heavy metals in agricultural soil around Hunan industrial estate by positive matrix factorization model, principle components analysis and geo statistical analysis. Ecotoxicol. Environ. Saf. 159, 354–362. https://doi.org/10.1016/j.ecoenv.2018.04.072

Zhang, X., Yang, H., Cui, Z., 2017. Migration and speciation of heavy metal in salinized mine tailings affected by iron mining. Water Sci. Technol. 76, 1867–1874. https://doi.org/10.2166/wst.2017.369

Zhang, X.Y., Lin, F.F., Wong, M.T.F., Feng, X.L., Wang, K., 2009. Identification of soil heavy metal sources from anthropogenic activities and pollution assessment of Fuyang County, China. Environ. Monit. Assess. 154, 439–449. https://doi.org/10.1007/s10661-008-0410-7

Zhang, Z., Juying, L., Mamat, Z., 2016. Sources identification and pollution evaluation of heavy metals in the surface sediments of Bortala River, Northwest China. Ecotoxicol. Environ. Saf. 126, 94–101. https://doi.org/10.1016/J.ECOENV.2015.12.025

Zhao, C.P., Zhou, X.J., Dong, K., Fu, J., Zhu, D. Di, An, S.Q., Zhu, H.L., 2017. Potential ecological risk and speciation analysis of heavy metals in sediments from the Jialu River, China. Int. J. Environ. Pollut. 61, 72–88. https://doi.org/10.1504/IJEP.2017.082700

Zhao, J., Fu, G., Lei, K., Li, Y., 2011. Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management. J. Environ. Sci. 23, 1460–1471. https://doi.org/10.1016/S1001-0742(10)60599-2

Zhao, S., Feng, C., Yang, Y., Niu, J., Shen, Z., 2012. Risk assessment of sedimentary metals in

the Yangtze Estuary: New evidence of the relationships between two typical index methods. J. Hazard. Mater. 241–242, 164–172. https://doi.org/10.1016/j.jhazmat.2012.09.023

Zhou, P., Zhao, Y., Zhao, Z., Chai, T., 2015. Source mapping and determining of soil contamination by heavy metals using statistical analysis, artificial neural network, and adaptive genetic algorithm. J. Environ. Chem. Eng. 3, 2569–2579. https://doi.org/10.1016/j.jece.2015.08.003

Zhou, T., Wu, J., Peng, S., 2012. Assessing the effects of landscape pattern on river water quality at multiple scales: A case study of the Dongjiang River watershed, China. Ecol. Indic. 23, 166–175. https://doi.org/10.1016/j.ecolind.2012.03.013

Zhuang, P., Zou, B., Li, N.Y., Li, Z.A., 2009. Heavy metal contamination in soils and food crops around Dabaoshan mine in Guangdong, China: Implication for human health. Environ. Geochem. Health 31, 707–715. https://doi.org/10.1007/s10653-009-9248-3

Zimmermann, H., 1996. Fuzzy Set Theory—and Its Applications, Fuzzy Set Theory—and Its Applications. https://doi.org/10.1007/978-94-015-8702-0

Zor, K., Timur, O., Teke, A., 2017. A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting, in: 2017 6th International Youth Conference on Energy, IYCE 2017. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IYCE.2017.8003734

RESEARCH OUTPUTS

Journal Publications (Published articles)

1. **Goswami, A.P.** and Kalamdhad, A.S., 2022. Water quality variation and source apportionment using multivariate statistical analysis. *Environmental Forensics*, pp.1-23.
2. **Goswami, A.P.** and Kalamdhad, A.S., 2022. Mobility and risk assessment of heavy metals in benthic sediments using contamination factors, positive matrix factorisation (PMF) receptor model, and human health risk assessment. *Environmental Science and Pollution Research*, pp.1-19.
3. **Goswami, A.P.,** Das, S. and Kalamdhad, A.S., 2021. Assessment of possible pollution risk using spatial distribution and temporal variation of heavy metals in river sediments. *Environmental Earth Sciences*, *80*(19), pp.1-15.
4. **Goswami, A.P.**, Das, S. and Kalamdhad, A.S., 2021. Monitoring and risk assessment of heavy metals surficial sediments using the 5-step sequential extraction process. *International Journal of Environmental Analytical Chemistry*, pp.1-22.
5. Singh, K.R., **Goswami, A.P.,** Kalamdhad, A.S. and Kumar, B., 2020. Assessment of surface water quality of Pagladia, Beki and Kolong rivers (Assam, India) using multivariate statistical techniques. *International Journal of River Basin Management*, *18*(4), pp.511-520

Book chapters (Articles published)

1. **Goswami, A.P**. and Kalamdhad, A.S., 2022. Trend of Water Quality of a River Flowing through Densely Populated Area of Assam, India using Water Quality Index (WQI). In *Environmental Degradation: Monitoring, Assessment and Treatment Technologies* (pp. 215-224). Springer, Cham.

International/National Conferences

1. **Goswami, A.P.**, Das, S. and Kalamdhad, A.S., (2019). Bioavailability and leachability study of sediments in Kolong River, Assam. Poster presentation in *National Environmental Conference 2019,* held on 31st Jan-2nd Feb 2019 at Indian Institute of Technology Bombay.
2. **Goswami, A.P.**, and Kalamdhad, A.S., (2019). Metal speciation study of Benthic Sediments in Kolong River, Assam. Poster presentation in *Research Conclave'19* held on March 14-17 at Indian Institute of Technology Guwahati.
3. **Goswami, A.P.**, and Kalamdhad, A.S., (2021). Assessment of the Ecological Risks Associated with Heavy Metals in river bed Sediments. Paper presentation in 2nd International Conference on *Futuristic and Sustainable Aspects in Engineering and Technology* (FSAET 2021) held on 24-25th December 2021 at GLA University, Mathura. (**Best Paper Award)**