HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

DOCTORAL (PhD) DISSERTATION

INTEGRATED GEOCHEMICAL, STATISTICAL, AND SIMULATION-BASED APPROACHES FOR ASSESSING WATER QUALITY AND HUMAN HEALTH RISKS IN SURFACE AND GROUNDWATER SYSTEMS OF THE DANUBE AND AL-JAWF BASINS

A Dissertation submitted for the degree of Doctor of Philosophy at the Doctoral School of Environmental Sciences, Hungarian University of Agriculture and Life Sciences

By

OMAR ABDULHAKIM HIZAM SAEED

GÖDÖLLŐ, HUNGARY

2025

Title: Integrated geochemical, statistical, and simulation-based approaches for assessing water quality and health risks in surface and groundwater systems of the Danube and Al-Jawf basins

Discipline: Hydrology and Water Resources Engineering

Name of Doctoral School: Environmental Sciences

Head: Dr. Erika Csákiné Michéli,
Professor, DSc MATE, Institute of Environmental Sciences,
Department of Soil Science.

Supervisor: Dr. András Székács
Professor, DSc. MATE, Institute of Environmental Sciences,
Agro-Environmental Research Centre.

Co-supervisor: Dr. Mária Mörtl
Senior Researcher, PhD, MATE. Institute of Environmental Sciences,
Agro-Environmental Research Centre.

Approval

……………………………… ………………………………

Approval of the School Leader Approval of the Supervisor(s)

DECLARATION

This dissertation is my original work and has not been presented for a degree in any other university. No part of this dissertation may be reproduced without prior permission of the author and/or Hungarian University of Agriculture and Life Sciences.

………OMARSAEED………… 2025/08/09

Omar Abdulhakim Hizam Saeed Date

DECLARATION BY SUPERVISOR
This dissertation has been submitted with my approval as supervisor
…………………………………………………
……………….
Dr. András Székács Date
Professor, DSc. MATE, Institute of Environmental
Sciences, Agro-Environmental Research Centre.
…………………………………….. …………………..
Dr. Mária Mörtl Date
Senior Researcher, PhD, MATE, Institute of Environmental
Sciences, Agro-Environmental Research Centre.

DEDICATION

This work is lovingly dedicated to those who have been the pillars of my existence –
to my beloved wife, whose unwavering love, patience, and strength have been my constant source of courage and hope.
To my family, whose boundless support, sacrifices, and faith have nurtured my dreams and lifted me beyond every obstacle.
To the mentors, friends, and companions who journeyed alongside me – your wisdom, kindness, and encouragement have illuminated my path and fueled my determination.

With profound gratitude and deepest affection, I dedicate this achievement to all of you – my heart’s true foundation.

ACKNOWLEDGEMENTS

Completing this Ph.D. has been one of the most challenging yet rewarding journeys of my life, and it would not have been possible without the unwavering support, guidance, and encouragement of many remarkable individuals and institutions. I wish to express my deepest gratitude to all who have been part of this path.

I wish to thank the Head of the Doctoral School of Environmental Sciences, Prof. Dr. Erika Csákiné Michéli, Professor, DSc, for granting me the opportunity to pursue my studies at MATE and for fostering an environment where research can flourish.

I am profoundly grateful to my supervisor, Prof. Dr. András Székács, Professor, DSc, for his outstanding mentorship, expert guidance, and steadfast dedication to my academic growth. His ability to challenge my thinking while providing consistent support has been invaluable in shaping both my research and my scholarly development.

My sincere appreciation goes to my co-supervisor, Dr. Mária Mörtl, Senior Researcher, PhD, for her patient guidance, insightful feedback, and constant encouragement, which have enriched my work immeasurably.

I also extend my heartfelt thanks to Prof. Győző Jordán, whose initial acceptance of my application and early supervision were crucial in enabling me to obtain the Stipendium Hungaricum Scholarship and embark on this academic journey. His support at the outset laid the foundation for my doctoral studies, and I remain truly grateful for his efforts.

To my colleagues at the Agro-Environmental Research Centre, I am thankful for the warm welcome, collaborative spirit, and stimulating discussions that have broadened my academic horizons.

Special thanks go to Ms. Zsuzsanna Tassy and Ms. Edit Dolányi Simáné for their tireless administrative support, efficiency, and kindness throughout the program.

I gratefully acknowledge the Stipendium Hungaricum Scholarship Program for their financial support, without which this journey would not have been possible. Hungary has been not only my place of study but also a source of cultural and personal enrichment.

I extend my deepest appreciation to the Republic of Yemen, represented by the Ministry of Higher Education and Scientific Research, for their trust and support in granting me the opportunity to pursue this academic journey.

To my beloved wife, whose unwavering love, patience, and sacrifices have been my constant source of strength – thank you for walking beside me through every challenge, for celebrating every small victory, and for believing in me even when the road was difficult.

To my family, whose encouragement, prayers, and unconditional support have been the foundation of my resilience, I am forever indebted. This achievement is as much yours as it is mine.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Full form
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
BW	Body Weight
CA	Cluster Analysis
CAI	Chlor-Alkali Index
CDF	Cumulative Distribution Function
CDI	Chronic Daily Intake
CR	Carcinogenic Risk
CSF	Cancer Slope Factor
DWQI	Drinking Water Quality Index
EC	Electrical Conductivity
ED	Exposure Duration
EF	Exposure Frequency
ELM	Extreme Learning Machine
ET	Exposure Time
FIS	Fuzzy Inference System
GWL	Groundwater Level
GIS	Geographic Information System
HCA	Hierarchical Cluster Analysis
HHRA	Human Health Risk Assessment
HI	Health Risk Index
HM	Heavy metal
HPI	Heavy Metal Pollution Index
HQ	Hazard Quotient
IR	Intake Rate
IDW	Inverse Distance Weighting
IWQI	Irrigation Water Quality Index
KR	Kelley Ratio Index

Kp	Permeability Coefficient
MCS	Monte Carlo Simulation
MF	Membership Function
MI	Metal Index
PS	Potential Salinity
RF	Random Forest
RfD	Reference Dose
RMSE	Root Mean Square Error
RSBC	Residual Sodium Bicarbonate
RSC	Residual Sodium Carbonate
SA	Skin Surface Area
SAR	Sodium Adsorption Ratio
SI	Saturation Index
SSP	Soluble Sodium Percentage
SW	Surface Water
SVM	Support Vector Machine
TDS	Total Dissolved Solid
WQI	Water Quality Index

1. INTRODUCTION

1.1. Background

Water is a vital and valuable natural resource that is essential to all environmental systems. The worldwide dilemma of insufficient access to clean and safe drinking water affects over a billion people (Proshad et al. 2021; Hoque et al. 2023). Tragically, water-related diseases and natural calamities claim the lives of 6-8 million people each year, highlighting the critical need of tackling global water supply challenges. Surface water (SW) is the most important freshwater supply for a variety of uses, including residential, recreational, agricultural, and commercial (Islam et al. 2022). Nonetheless, groundwater (GW) is also an essential freshwater supply in numerous regions of the world, particularly where SW is scarcely available. The endpoint of spent water, generally called wastewater, manifests inside the aquatic environment, including rivers, ponds, or other aqueous reservoirs (Muhammad et al. 2022; Shammi et al. 2023). Rivers are important among these water resources since they provide the principal water source for human use, agriculture cultivation, and industrial processing. The rising contamination of SW caused by human activities, as well as the deposition of pollutants from the atmosphere, such as heavy metals (HMs) (metallic or metalloid elements of high relative atomic weight and of relatively high density, mostly transitional metals, certain metalloids, as well as lanthanoids and actinoids, which are generally poorly uptaken from soil by plants), are major problems, particularly in metropolitan areas, emphasizing the critical need for comprehensive environmental management measures. The existence of the physicochemical and HM characteristics of water in aquatic habitats has prompted substantial worldwide concerns owing to their potential negative impacts on human health and ecosystems. HMs, such as copper (Cu), manganese (Mn), iron (Fe), and zinc (Zn), are essential for human metabolism at certain levels but become toxic when concentrations exceed permissible limits. Metals like arsenic (As) and lead (Pb) have no physiological role and can disrupt the endocrine system, leading to various health issues including teratogenic and carcinogenic effects (Chowdhury et al. 2016). These contaminants enter human bodies through dermal contact, ingestion, and inhalation, posing significant health risks (Rezaei et al. 2019). The pathways through which humans are exposed to these metals are varied. Ingestion of contaminated water and food, inhalation of polluted air, and dermal contact are common routes. In aquatic environments, metals can accumulate in sediments and enter the food chain, affecting fish and other aquatic organisms that humans consume. This bioaccumulation can lead to higher concentrations of metals in higher trophic levels, posing significant health risks to humans who consume these organisms (Izah et al. 2017). Monitoring and assessing the amount of HMs in water bodies like the Danube river is crucial. The Heavy Metal Pollution Index (HPI) is one such method, providing a comprehensive assessment of the overall HM pollution status of water bodies. Techniques such as cluster analysis (CA) help in identifying pollution sources and patterns, while Geographic Information System (GIS) methods assist in visualizing spatial distribution and potential hotspots of contamination. The Monte Carlo simulation (MCS) method can predict hazard quotients for heavy metals, helping in risk assessment and management (Sheng et al. 2021). The Danube river, Europe's second-longest river, supports a diverse aquatic ecosystem but is vulnerable to HMs pollution, affecting fish populations, benthic fauna, and aquatic plants (Saeed et al. 2023a). Significant sources of HM pollution in the Lower Danube region include sewage discharge, municipal waste, fertilizers, pesticides, fossil fuel combustion, and activities related to navigation and mining (Calmuc et al. 2021; Saeed et al. 2023a). To protect the environment and human health, it is imperative to monitor water quality regularly. Techniques such as the HPI, CA, GIS methods, and the MCS are used to assess and manage pollution levels effectively (Sheng et al. 2021). The Danube river is not only a crucial waterway for transportation and trade but also a lifeline for millions of people living along its banks. It provides drinking water, supports agriculture, and sustains fisheries and biodiversity. However, the river's health is under constant threat from both point and non-point sources of pollution. Point sources include industrial discharges and wastewater treatment plants, while non-point sources encompass agricultural runoff and atmospheric deposition. To comprehensively assess water quality and associated health risks under differing environmental contexts, this study selected two contrasting hydrological basins: the Danube river in Hungary and the Al-Jawf basin in Yemen. These sites were chosen due to their representativeness of two distinct geo-environmental and socio-economic settings, one located in a temperate, industrialized European region with relatively strong regulatory frameworks, and the other in an arid, agriculturally reliant Middle Eastern region where the primary water resource is GW. By applying comparable analytical methods, including HM assessments, geochemical modeling, and predictive machine learning (ML) models, to both systems, the research aims to test the transferability of methodologies, highlight context-specific pollution drivers, and inform regionally appropriate water management strategies. This cross-basin design strengthens the study's generalizability while also exposing site-specific challenges in water quality monitoring and policy planning. The cumulative effect of these pollutants can lead to significant degradation of water quality, affecting not only human health but also the ecological balance of the river basin (Saeed et al. 2023a). GW, especially in arid regions like Yemen's Al-Jawf basin, is a vital resource for socioeconomic development, agriculture being the primary consumer. However, GW quality has deteriorated due to population growth, industrial expansion, agricultural activities, and climate change (Siebert et al. 2010; Al-Mashreki et al. 2023). Yemen, one of the most water-stressed countries, faces severe challenges due to the excessive extraction of GW insufficient infrastructure, and bad management techniques, leading to the rapid depletion of aquifers (Al-Mashreki et al. 2023). The chemical composition of GW results from long-term interactions with geological, climatic, and human factors, affecting its suitability for use. Evaluative procedures such as statistical analysis, Piper diagrams, and geochemical models are essential for understanding the chemical features of GW (Al-Mashreki et al. 2023). The evaluation of groundwater quality involves a comprehensive analysis of its physicochemical properties. Parameters such as pH, electrical conductivity (EC), total dissolved solid (TDS), and concentrations of major ions (e.g., calcium, magnesium, sodium, potassium, bicarbonate, sulfate, and chloride) are essential indicators of water quality. GW suitability for irrigation is assessed using various indices, including the Sodium Adsorption Ratio (SAR), Residual Sodium Carbonate (RSC), and Permeability Index (PI). These indices help determine the potential impact of irrigation water on soil properties and crop health. High concentrations of sodium, for example, can lead to soil dispersion, reducing its permeability and affecting plant growth. Therefore, regular monitoring and evaluation of GW quality are necessary to ensure its sustainable use for agriculture (Al-Mashreki et al. 2023). In recent years, artificial intelligence (AI) and ML models have been increasingly used to predict irrigation water quality indices (IWQIs) and assess water quality due to their ability to handle complex data interactions and improve prediction accuracy (Al-Mashreki et al. 2023). The integration of AI and ML models in water quality assessment has revolutionized the field by providing more accurate and reliable predictions. These models can handle large datasets and complex interactions between various water quality parameters, improving the prediction of IWQIs. ML algorithms such as artificial neural networks (ANNs), support vector machines (SVMs), and decision trees have been successfully applied to predict water quality indices and assess the suitability of water for different uses (Khadr et al. 2020; Al-Mashreki et al. 2023). These advanced models not only enhance the accuracy of predictions but also provide insights into the factors influencing water quality. For instance, ML models can identify the most significant parameters affecting GW quality and predict future trends based on historical data. This information is critical for policymakers and water resource managers, enabling them to make informed decisions and implement effective management strategies (Saeed et al. 2023a). Water quality is a fundamental component of environmental sustainability and public health, particularly in regions where SW and GW serve as primary sources for drinking and agriculture. Previous studies have explored HM contamination and physicochemical characteristics in both the Danube river and arid zones like Yemen's Al-Jawf basin. However, most of these investigations have been geographically or thematically narrow, often applying distinct methodologies to different hydrological systems and rarely integrating SW and GW assessments in a comparative context. In addition, few have combined conventional water quality indices with advanced AI, ML, and MCS models to enhance predictive accuracy and risk evaluation. This study fills that gap by employing a unified analytical approach across two hydrologically contrasting systems, Hungary's Danube river lower watershed and Yemen's Al-Jawf GW basin. Through the integration of traditional geochemical analyses, health risk assessments, and predictive models (ANFIS, RF, and SVM), this research provides a novel, cross-regional framework for understanding spatial contamination dynamics and improving water management strategies, particularly in data-scarce or resource-limited settings.

1.2. Problem statement

Despite growing environmental pressures in both the lower Danube river basin (Hungary) and the Al-Jawf GW basin (Yemen), comprehensive and comparative assessments of water quality and related health risks remain limited. Although the Danube river and Al-Jawf basin are geographically and climatically distinct, analyzing both provides a valuable contrast between humid and arid water systems. This comparative approach enhances our understanding of how different environmental and anthropogenic factors influence water quality, and how advanced methods like simulation and ML can be applied universally to manage water-related risks. Most prior studies focus on localized contexts, rarely integrating HM analysis, physicochemical characterization, and health risk evaluation across contrasting hydrological regions. Additionally, industrial waste events, such as the 2010 Ajka red mud disaster in Hungary, have shown how hazardous byproducts like red mud can significantly impact soil and water quality. Although not directly assessed in this study, such incidents highlight the need for integrated approaches that consider both point and non-point pollution sources in future water quality assessments. Additionally, advanced predictive tools like Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Random Forest (RF), and SVM, along with MCS for quantifying uncertainty in hazard and cancer risk estimates, are underutilized. This study fills that gap by applying unified water quality indices and machine learning models across two diverse environments, aiming to enhance scientific understanding and support more accurate, data-driven water management strategies.

1.3. Objectives

This study aims to assess and compare the water quality of SW in Hungary's lower Danube river basin and GW in Yemen's Al-Jawf basin, focusing on environmental sustainability and public health. The specific objectives are:

- To characterize water chemistry and geochemical processes using physicochemical parameters and HM concentrations.
- To evaluate agricultural water suitability through the application of indices for irrigation water quality, including IWQI, SAR, SSP, KR, PS, RSC, and RSBC.
- To conduct non-carcinogenic and carcinogenic health risk assessments of HM, using MCS to quantify uncertainty.
- To enhance the predictive assessment of irrigation water quality and health risks by applying and evaluating multiple machine learning algorithms–RF, SVM, and ANFIS–selected for their complementary strengths in handling nonlinear patterns, small datasets, and complex geochemical interactions.

1.4. Research Questions

- What are the key geochemical characteristics of water in the lower Danube and Al-Jawf basins?
- How suitable are these water sources for agricultural use based on IWQI and related indices?
- What are the potential non-carcinogenic and carcinogenic health risks associated with HM in these water sources?
- Can ML models effectively predict irrigation water quality and health risk indicators in both basins?

1.5. Hypotheses

- H1: Water quality in both basins is significantly impacted by anthropogenic activities, with elevated concentrations of HMs beyond acceptable limits.
- H2: The Al-Jawf GW is less suitable for irrigation compared to the Danube SW due to higher salinity and sodium hazard indicators.
- H3: ML models (RF, SVM) can accurately predict water quality and health risk indices, reducing uncertainty in risk assessments.

2. LITERATURE OVERVIEW

2.1. Groundwater management

GW reserves are extensive natural reservoirs located beneath the Earth's surface. In arid regions, these reserves play a crucial role in supplying water for various purposes such as drinking, irrigation, and industrial activities (Ghosh et al. 2022; Ahmed et al. 2024). Approximately half of the drinking water supply and nearly 43% of irrigation needs are met through GW (Ahmed et al. 2024). However, both natural phenomena and human activities have exerted significant pressure on GW resources. Factors such as climate change, population growth, and agricultural demands pose substantial threats to both the quantity and quality of GW. Many studies employing ML techniques have focused primarily on forecasting groundwater levels (GWLs). Understanding the dynamics of GW flow is crucial for assessing water availability and scarcity. Therefore, achieving sustainable management of GW resources in arid and semi-arid regions hinges upon accurately predicting GWL, which serves as a direct indicator of GW availability and provides insights into its hydrodynamics (Ahmed et al. 2024). However, characterizing GW dynamics is challenging due to spatial and temporal variations influenced by factors like slope, geology, rainfall patterns, and soil composition.

2.2. Predictive modeling and management of water resources

Access to safe drinking water can be hindered by various factors, such as inadequate water infrastructure, climate-related issues like droughts, and the overexploitation of freshwater resources. Recently, researchers have applied ML techniques to forecast droughts. (Wang et al. 2023) for instance, investigated the use of extreme learning machines (ELMs) for predicting hydrological droughts. They employed SVM models and ELMs in approximately 144 different configurations, demonstrating that ELMs can forecast the standardized hydrological drought index with high accuracy. Their study highlighted the effectiveness of self-adaptive differential evolution ELMs over other models, especially when combined with wavelet hybrid techniques, which minimized prediction errors. Global water scarcity is a significant challenge recognized by the United Nations within its Sustainable Development Goals for 2030. The gap between water demand and sustainable limits has narrowed dangerously in many regions, with water demand rates surpassing population growth rates over the past century (Ahmed et al. 2024). Arid regions face the most severe impacts of this water crisis, underscoring the critical importance of accurately monitoring water demand for informed decision-making. Effective forecasting of water demand can contribute to the sustainable management of limited water resources, thereby mitigating overexploitation.

2.3. Water quality monitoring

Ensuring water quality is crucial for public health and safety. However, since the 1990s, water pollution has escalated in most rivers across Africa, Latin America, and Asia (Programme United Nations Environment 2016). The increasing global population has intensified the demand for clean drinking water, placing significant stress on water treatment facilities. Major urban centers like London, Hong Kong, and New York contend with substantial stormwater runoff carrying diverse pollutants, often compromising drinking water quality and endangering marine life. (Yu and Qu 2020) employed radial basis function neural networks, and utilizing difference-gated neural networks. Ma and co-workers (Ma et al. 2020) developed a robust deep matrix factorization combined with DNNs for water quality assessments in New York City harbors, demonstrating superior performance over traditional methods like Ridge, LR, and LASSO. Efficient wastewater treatment is essential for public health and environmental safety. In 2015, UNICEF reported that 1.8 billion people rely on contaminated water sources, exposing them to diseases like polio, cholera, and typhoid. UNESCO (2017) highlighted that globally, 80% of wastewater is inadequately treated before release, leading to severe impacts on aquatic life and ecosystem health due to pollutants like organic matter and reduced oxygen levels (Ahmed et al. 2024).

2.4. Water quality models, challenges, and limitations

Water quality modeling (WQM) is a crucial tool that supports environmentalists, policymakers, and water resource managers in strategic water resource management. However, WQM presents challenges within the scientific community due to various constraints and limitations. Water quality models are generally classified based on the type of receiving water, model complexity, and the water quality parameters they can predict, such as nutrients, dissolved oxygen, and biological oxygen demand. Effective WQM requires proper standardization, identification of pollution hotspots, recognition of common features, and the development of policy-relevant models. These models help reduce labor costs, materials, and time (Kumar et al. 2022a), facilitating effective pollution mitigation in watersheds. Recently, numerous models have been used to simulate the water quality of freshwater bodies (streams, rivers, reservoirs, and lakes), estuaries, coastal waters, and marine ecosystems (Loucks and van Beek 2017). However, differing theories and algorithms in these models result in varying outputs, leading to significant differences in results. Consequently, models are most useful when applied to specific environmental problems (Kumar et al. 2022a).

2.5. Water quality mitigation measures

Water quality mitigation measures or strategies are designed to help communities identify potential alternatives to minimize the adverse impacts of pollutants on water quality and ensure water safety for community use (Liu et al. 2016). These measures aim to protect, restore, preserve, and improve the water quality of receiving bodies. Water quality protection involves adequately treating runoff to prevent downstream water quality degradation (Dujmović and Allen 2021). When protection strategies fail to maintain water quality standards, restoration efforts are implemented, requiring collaboration among stakeholders from various fields to achieve restoration goals (Ball 2008). Preservation of water quality requires a decision-support framework to evaluate, monitor, and optimize the effects of different factors on water quality (Kumar et al. 2022a). Water quality standards can be addressed through regulation, remediation, and watershed management (Kumar et al. 2022a). Regulation in a specific area controls the discharge of waste from industries or sewage treatment plants by setting standards for each pollutant released into SW (Guo et al. 2019). Remediation efforts, which include biological, chemical, and physical methods, help clean water contamination. Biological remediation, is cost-efficient and involves using naturally occurring organisms such as plants, bacteria, and fungi to remove or neutralize pollutants and break down hazardous substances into less toxic or non-toxic forms. This method is commonly used to treat human sewage and agricultural chemicals that leach into GW (Song et al. 2022). Watershed management strategies focus on reducing the chemicals applied to land, making them more effective for nonpoint source pollution control than merely setting pollution standards (Kumar et al. 2022a).

2.6. Environmental impacts and management of red mud: a global perspective

Red mud holds importance because of its potential for resource recovery, as it contains valuable materials like iron, titanium, and rare earth elements. The storage of red mud, a hazardous byproduct of aluminum refining, presents substantial ecological risks due to its alkalinity, toxic trace elements, and radionuclides. Typically stored in open-air reservoirs, the global stockpile of red mud exceeds 2.7 billion tons and continues to grow annually (Klauber et al. 2011; Power et al. 2011). Wet storage methods heighten the risk of dam failures, leading to severe environmental disasters, as seen in the 2010 Kolontár dam spill (Hungary) and the 2015 Fundão dam collapse (Brazil). Post-disaster cleanup involves either removing or integrating red mud into the soil. Studies conducted after these events highlight varied impacts on soil organisms, underscoring the need for further extensive ecological investigations over the long term. The global aluminum refining industry produces red mud, a hazardous waste containing toxic elements such as Cd, Cr, Hg, and radionuclides like Ra-226 and Th-230 (Klauber et al. 2011; Winkler et al. 2018). Typically stored in aqueous solutions with sodium hydroxide, red mud has accumulated to more than 2.7 billion tons globally, with an annual increase of 120 million tons (Power et al. 2011; Klauber et al. 2011). The 2010 disaster in Ajka, Hungary, where more than a million cubic meters of toxic sludge inundated the environment, remains a striking example (Appendix Figure 1) (Mayes et al. 2016; Winkler et al. 2018). Cleanup efforts included both the removal and integration of red mud into affected areas (Uzinger et al. 2015). Studies conducted after the disaster underscored responses in soil biota, with investigations on organisms like Folsomia candida revealing diverse impacts (Winkler 2014; Winkler et al. 2018). Red mud alters soil texture due to its fine particle size, influencing the accessibility and function of soil biota (Winkler et al. 2018). Its elevated alkalinity and sodium levels further disturb both underground and surface-level communities (Stenchly et al. 2017). Trace metal contamination from red mud affects invertebrate soil organisms differently (Hodson 2013). Despite its capability to immobilize certain trace elements in soil, the risks associated with its toxic components remain significant (Garau et al. 2007; Winkler et al. 2018).

2.7. Sources and impacts of heavy metal pollution in the environment

Appendix Figure 2A depicts the origins of HM pollution in the environment. Industries such as mining, milling, plating, and surface finishing are primary contributors to HM pollution, releasing a variety of harmful metals (e.g. Ni, Cr, Cd, Cu, Co, Zn, and Pb) into the atmospheric (Cheng et al. 2018; An et al. 2020; Saravanan et al. 2024). The levels of these HMs in water bodies and sediments have increased dramatically over the past few decades. Consequently, concentrations of toxic metals in crops grown in polluted soils have risen significantly, posing a significant threat to human health and the environment due to the toxic nature, non-biodegradability, and bioaccumulation of HMs. Appendix Figure 2B illustrates the detrimental effects of various HMs on human health.

2.7.1. Impacts and health effects of lead (Pb) toxicity

Pb is considered a highly hazardous HM. Its prolonged exposure is commonly known as lead poisoning, affecting various human systems including the nervous, hematopoietic, and cardiovascular systems (Tao et al. 2021). At high doses, lead can cause neural, hematological, and renal disorders. In children, lead exposure can lead to developmental issues such as behavioral problems, cognitive impairments, and learning disabilities. Lead poisoning also results in reduced fertility, miscarriages, spontaneous abortions, neurological disorders, cardiovascular diseases, gastrointestinal disturbances, and has mutagenic and carcinogenic effects. Lead is toxic to many aquatic organisms, particularly freshwater species. Marine organisms generally exhibit greater tolerance to lead exposure, with some species showing resilience or adaptability to its effects (Qu et al. 2021).

2.7.2. Impacts and health effects of arsenic (As) toxicity

As contamination in GW is a significant global concern affecting numerous countries worldwide. Over a hundred million people are at risk of arsenic poisoning due to arsenic contamination in GW (Ravindra and Mor 2019). Arsenic is a toxic and carcinogenic substance. Its presence in water is primarily due to natural geological processes, namely mineral dissolution/precipitation and adsorption/desorption processes. Human activities, for example pesticide use contribute also significantly to arsenic contamination in the environment. Prolonged exposure to arsenic in drinking water has been associated with various cancers including liver, lung, kidney, bladder, and skin cancers (Saravanan et al. 2024).

2.7.3. Impacts and health effects of chromium (Cr) toxicity

Cr is recognized as a potent toxicant known for its carcinogenic and teratogenic properties. Cr III compounds are insoluble in water, thereby posing negligible risks compared to more soluble chromium components. Industrial activities like chromite ore refining, ferrochromium production, manufacturing of Cr chemicals, steel production, leather tanning, coal and oil combustion, cement manufacturing, domestic wastewater, cooling towers, and coke ovens are responsible for Cr contamination in the environment. Cr (VI) is highly toxic due to its strong oxidizing properties (Sarojini et al. 2021). Upon entering the bloodstream, it adversely affects the kidneys, liver, and blood cells, leading to conditions like renal failure, hemolysis, liver failure, and necessitating aggressive dialysis for management. Exposure to Cr-containing materials can lead to allergic contact dermatitis and irritant dermatitis, often resulting in skin ulceration, commonly referred to as "chrome ulcers," particularly among workers exposed to concentrated chromate solutions in electroplating, tanning, and chrome production facilities (Kumaraguru et al. 2023).

2.7.4. Impacts and health effects of copper (Cu) toxicity

Cu is naturally present in a variety of minerals found in the environment, including native Cu, CuS_2, $Cu_2(OH)_2CO_3$, and CuO (Hama Aziz et al. 2023). It serves as an essential micronutrient vital for all organisms. In addition to proteins and carbohydrates, Cu plays crucial roles in cellular metabolism, chlorophyll formation, photosynthesis, respiratory electron transport chains, and protection against oxidative stress in plants. Consequently, agricultural practices involving copper-based chemicals have led to environmental contamination from human activities. The phytotoxic effects of excessive and widespread Cu exposure include increased production of reactive oxygen species (ROS) and damage to sugars, lipids, proteins, and DNA (Moyé et al. 2017). High levels of Cu in plants, animals, and humans further contribute to environmental contamination. Sources of Cu pollution from natural and anthropogenic origins include copper mining, petroleum refining, waste incineration, domestic waste disposal, industrial effluents, phosphate fertilizer production, wood combustion, volcanic eruptions, wind-blown debris, marine aerosols, forest fires, and decomposition of vegetation. Over recent decades, cumulative copper deposits have reached 939,000 metric tons (Moyé et al. 2017).

2.7.5. Impacts and health effects of zinc (Zn) toxicity

Zn is widely distributed throughout the environment and is considered an essential metal for all living organisms. It finds extensive use in various applications such as galvanizing, alloying, and manufacturing antibacterial and UV-protective textiles (Huang et al. 2014). Naturally, zinc is abundant in the earth's crust, typically occurring at a concentration of 70 mg/kg. In its +2 oxidation state, zinc exists in minerals like smithsonite (zinc carbonate), sphalerite (zinc sulfide), and zincite (zinc oxide). Both natural processes and human activities contribute to zinc emissions into the environment, with anthropogenic sources outweighing natural sources. Exceeding concentrations of 20 mg/kg, zinc adversely impacts aquatic species through various biochemical, physiological, and ecological effects. Toxicity levels can vary depending on water temperature, hardness, and alkalinity. Mining activities contribute significantly to zinc contamination of soils and water in affected regions. Excessive intake, whether intentional through supplements or unintentional through exposure to contaminated environments, poses various health risks including organ damage, anosmia (loss of smell), fatigue, neurological disorders, genetic abnormalities, fungal infections, and urinary complications (Chen et al. 2018).

2.7.6. Impacts and health effects of nickel (Ni) toxicity

Ni is a lightweight metal widely utilized in industrial applications due to its unique physicochemical properties. It is employed in manufacturing various items such as rechargeable batteries, coins, electroplating, and stainless steel. Nickel-plated pipes, fixtures, cookware made from nickel-containing stainless steel, and items coated with nickel-based alloys contribute to environmental contamination through soil contact and leaching of nickel compounds into water sources.

Wastewater and residues generated during nickel mining and refining processes also contribute significantly to environmental pollution. Consequently, individuals are often exposed to nickel through inhalation, ingestion of contaminated water, and direct contact with skin (Jacob et al. 2015). Despite being one of the lighter metals in the HM category, Ni exposure through prolonged skin contact can lead to hypersensitivity reactions, most commonly manifested as contact dermatitis. Additionally, ingestion of Ni-contaminated solutions can induce symptoms such as nausea, vomiting, and diarrhea. Prolonged inhalation of Ni in its metallic or monoxide forms may also result in conditions such as asthma and bronchial irritation.

2.8. Advancements in detection technologies for heavy metal and hazardous element pollution: current methods and emerging innovations

Traditional analytical methods for detecting potential toxic elements, including HM detection, include inductively coupled plasma mass spectrometry, inductively coupled plasma atomic emission spectrometry, X-ray fluorescence spectrometry, atomic fluorescence spectrometry, atomic absorption spectrometry, infrared spectrometry, ultraviolet-visible spectrometry, and ion chromatography. While these methods provide accurate results, they are constrained to laboratory settings, involve lengthy detection periods, require expensive equipment, and necessitate complex sample preparation. Previous studies have utilized flame atomic absorption spectrometry for quantitative analysis of HMs in water and soil, despite limitations such as narrow linear response range and single-element detection capability (Saravanan et al. 2024). Similarly, graphite flame atomic absorption spectrometry has been employed for soil analysis of HMs like Hg (Saravanan et al. 2024), while cold vapor atomic absorption spectrometry and hydride-generating atomic absorption spectrometry are utilized for detecting elements such as As and Se in natural and potable water sources. Inductively coupled plasma mass spectrometry is widely recognized for its capability in trace metal analysis (Al, As, Pb, Cd, Ag, U, Zn, Mn, Ni, Co, and V), yet it requires specialized expertise, multiple high-purity gases, and expensive operational costs (Ar). Inductively coupled plasma-optical emission spectrometry has also been validated as a multielement method for identifying HMs (Cd, Co, Cr, Ni, Pb, and Zn) in water samples (Kobylinska et al. 2020). Therefore, ongoing advancements in detection technologies are crucial for addressing the challenges posed by HMs pollution, ensuring effective environmental monitoring, and safeguarding public health. Emerging new methodologies for rapid identification include electrochemical analysis approaches such as dissolution voltammetry, potentiometric dissolution, polarimetric analysis, and electrochemical sensor methods. Additionally, biological detection techniques including immunoassays, enzyme analysis, biosensor techniques, and microbial detection methods are under active investigation. Laser-induced breakdown spectroscopy has identified HMs like Cr, Cu, Fe, Mn, Ni, Pb, and Zn in soil samples (Nawar et al. 2019). These cutting-edge approaches aim to overcome the limitations of conventional methods by offering faster and more convenient HM detection tailored to the type of HMs, matrix complexity, and concentration range being targeted (Helaluddin et al. 2016).

2.9. GIS for water quality mapping

GIS have become an essential tool for water quality assessment and management by enabling spatial analysis and visualization of water quality indicators. GIS integrates spatial data with statistical and database operations, allowing researchers and managers to monitor and analyze environmental variables with respect to their geographic distribution (Mtetwa et al. 2003). This technology integrates standard database operations and statistical analyses with the unique benefits of visualization and geographic analysis provided by maps and spatial databases (Khouni et al. 2021). It enables the exploration of cause-and-effect relationships through visual interpretation, and supports complex analyses and simulations, making it invaluable for researchers and natural resource managers (Khouni et al. 2021). Within GIS, spatial interpolation techniques are employed to predict attribute values at unsampled locations, thereby creating spatially continuous data (Li and Heap 2011). These techniques are categorized into deterministic and geostatistical methods (Qu et al. 2017), with IDW, kriging, Spline, Trend, and cokriging among the commonly used methods (Khouni et al. 2021). The selection of interpolation models should align with specific research objectives and characteristics of the study area (Qiao et al. 2018), although no single method has consistently shown superiority over others (Khouni et al. 2021). IDW and kriging are extensively applied in water assessment and contamination mapping, with IDW often preferred for its precision in certain applications over kriging (Khouni et al. 2021). Studies have demonstrated that IDW can be more accurate than kriging in predicting pollutant levels in GW. The IDW method assigns higher weights to nearby sample points compared to those farther away, thereby influencing calculations more significantly, reflecting its suitability for enhancing monitoring and assessment in river basin management systems aimed at improving pollution control (Khouni et al. 2021). Therefore, due to its ability to address sampling regularity, temporal correlations, and associated uncertainties, IDW emerges as a preferred interpolation method for assessing water quality in the Danube River and Al-

Jawf basin. The WQI maps were generated using QGIS software version 3.36.3. QGIS, a widely used open-source GIS software, is particularly suitable for mapping and analyzing river water quality. Its advanced cartographic capabilities allow users to create detailed datasets for river water quality mapping and effectively visualize water quality data on maps. QGIS offers a range of tools including various color ramps, symbology styles, and graduated symbols to depict water quality parameters, facilitating the identification of spatial patterns and trends. Integration of diverse data sources within QGIS enables the incorporation of Water Quality Index (WQI) attributes. For generating WQI maps, an equal interval mode was used with a graded visualization technique employing five distinct colors to represent different grades of river water quality based on the WQI.

2.10. Human health risk assessment (HHRA) of heavy metal in water resources

The severity of the adverse health consequences is influenced by several variables, including the sort of HM, duration and intensity of exposure, individual vulnerability, and concurrent exposure to other pollutants. Regular monitoring of water sources and food for HMs and the implementation of mitigation strategies are crucial to prevent adverse health effects in the human population (Guzzi et al. 2021; Swain 2024). Given the importance impact of quality of water on public health, assessing the human health risks associated with HM in water sources has garnered widespread attention. Human health risk (HHRA) assessment serves as a method to evaluate the potential adverse health effects on the public over a defined period. The health assessment of each contaminant primarily relies on assessing the magnitude of risk and whether the substance is carcinogenic (Liu et al. 2020; Jia et al. 2024). The primary routes of human exposure to HMs are through ingestion, inhalation, and dermal absorption via contaminated water and food sources. Both carcinogenic and non-carcinogenic risks associated with these exposure routes are assessed individually (Mahato et al. 2016). HM, when ingested through contaminated water, fish, or seafood, pose various acute and chronic health risks including carcinogenicity, mutagenicity, neurotoxicity, and endocrine disruption. Common health effects include acute HM poisoning, neurological symptoms, chronic conditions such as cancer and neurological disorders, renal and cardiovascular damage, reproductive and developmental effects, and disruption of endocrine function (González-Feijoo et al. 2024).

2.11. The presence of heavy metal in water resources - policies and regulations

In many regions, people directly collect water from local water bodies. Access to water purification facilities prior to distribution is limited or impractical in developing countries like India. Due to rapid urban expansion and industrial development, ensuring the quality and quantity of water is challenging. The World Health Organization (Saravanan et al. 2024) identifies several hazardous HMs including Al, Be, Fe, Cr, Mn, Co, Ni, Mo, Cu, Ag, Zn, Cd, and Se. This list encompasses elements ranging from beryllium, the second lightest metallic element after lithium (atomic number 4), to aluminum, the widely used lightweight metal with a density of 2.7 g/cm^3, as well as arsenic classified as a metalloid and selenium as a non-metal (Kumar et al. 2022b). These metals are non-biodegradable and are naturally distributed throughout the Earth's crust. They enter the human body through inhalation and ingestion of water and food. Both natural and human activities contribute metals to the environment, with anthropogenic sources often outweighing natural ones. Some metals such as Fe, Zn, Co, Cu, Ni, Mn, Mo, and Se are essential nutrients for living organisms. However, excessive levels of these metals, and that of non-essential metals, including antimony, As, Be, Cd, Pb, and Hg, pose health risks (Saravanan et al. 2024). Environmental regulations establish maximum allowable limits and quality standards for several hazardous substances and elements in different environmental compartments (Sun et al. 2021).

2.12. Regional studies on heavy metal pollution in the Hungarian Danube basin

HM pollution in the Hungarian Danube basin has been extensively studied, revealing significant spatial and geomorphological influences on contamination patterns. Near Dunaújváros, an industrialized area along the Danube, sediment samples showed elevated copper (Cu) and zinc (Zn) concentrations exceeding Hungarian environmental standards, with lower but detectable levels of cadmium (Cd) and nickel (Ni). These findings highlight the strong anthropogenic influence from nearby metal-processing industries, emphasizing the need for continuous monitoring to mitigate long-term environmental degradation (Kovács-Bokor et al. 2021). In the Tisza River floodplain, oxbow lakes, formed from abandoned meander loops, act as natural sediment traps, accumulating HMs (Cd, Co, Cr, Cu, Ni, Pb, Zn) in layers up to 10 cm deep. A study of seven oxbow lakes (four on the active floodplain and three outside the levee) found that contamination sources could be attributed to either fluvial deposition during floods or localized anthropogenic activities such as sewage discharge and fishing (Babcsányi et al. 2020). Further research on the Tisza floodplain demonstrated that fluvial landforms, particularly swales and point bars, significantly influence HM distribution in soils. Swales exhibited higher concentrations of Fe, Mn, Cr, Cu, Ni, Pb, and Zn at all depths compared to other landforms ($p < 0.05$), underscoring the role of geomorphology in sediment deposition and HM retention (Szabó et al. 2020). Beyond HM contamination, long-term environmental changes in the region have been observed. A 150-year reconstruction of SW temperatures in the Pannonian Ecoregion (including the Danube and Tisza Rivers) revealed a gradual warming trend of 0.05°C per decade (1870–2021), intensifying to 0.32°C per decade over the last 40 years (Li et al. 2024). This warming may exacerbate HM mobility and ecological risks, necessitating integrated environmental assessments to address both pollution and climate-driven changes in river systems.

2.13. Machine learning applications in science and engineering

ML represents a critical application of AI across various scientific and engineering fields. Its primary role involves capturing and understanding intricate relationships between diverse parameters within complex systems, all without direct human intervention. A significant advantage of ML lies in its capability to forecast future events and conditions across different domains, thereby facilitating practical applications (Korkmaz and Correia 2019). In water resource management, statistical and ML techniques are widely utilized for prediction purposes, encompassing pure forecasting to the estimation of optimization model parameters. ML algorithms are designed to process data, enabling them to discern various patterns that can be applied in other processes (Sarker 2021). ML serves as a potent tool for comprehending system behaviors from large datasets where identifying relevant patterns may be challenging or impractical (Rezaei and Javadi 2024). It operates in two primary modes: unsupervised learning, where the machine autonomously evaluates data to predict trends, and supervised learning, where patterns are derived from input and output data to forecast future patterns across different datasets. Given the exponential growth in data generation, robust and rapid tools are indispensable for the assessment and analysis of big data (Rezaei and Javadi 2024). Recent environmental monitoring studies increasingly leverage RF algorithms to enhance traditional water quality evaluation and source apportionment methods. For instance, (Zhang et al. 2024) fused RF with WQI and Positive Matrix Factorization models, using data from six sites along China’s Minjiang River (2015–2020). Their RF-optimized framework identified total phosphorus, total nitrogen, chemical oxygen demand, dissolved oxygen, and biological oxygen demand as the five most influential indicators, enabling exceptionally accurate WQI predictions ($R^2 \approx 0.97$) and improving the realism of pollution source apportionment (e.g. agriculture 30 %, domestic sewage 29 %, industrial wastewater 26 %). Similarly, (Ahmadi et al. 2023), studying marine waters at the Abrolhos reefs, applied RF and Boosted Regression Trees to a dataset of 665 seawater samples, which revealed that hydrodynamic forcing, dissolved organic carbon, and total nitrogen best explained microbial abundance variations. These applications demonstrate RF’s ability to handle complex, multidimensional environmental datasets, quantify variable importance, and improve predictive performance, justifying its selection for modeling water quality indices and health risk indicators in our study.

2.14. Refining the adaptive neuro-fuzzy inference system (ANFIS)

ANFIS was originally introduced and developed by Jang in 1993. It represents an advanced modeling technique that integrates fuzzy logic with neuro-computing and genetic algorithms to handle complex nonlinear systems. ANFIS constructs input–output models using specified data and human knowledge in the form of fuzzy if-then rules (Jang, 1993), essentially implementing a fuzzy inference system (FIS) within an artificial neural network (ANN) framework. ANNs simulate human brain information processing by learning tasks from input–output datasets without explicit programming of task-specific rules. They adapt their structure through learning algorithms during training, typically comprising interconnected neurons organized in input, output, and one or more hidden layers (Irshaid and Abu-Eisheh 2023). In contrast, FIS is a computing framework that mimics human reasoning using fuzzy logic, which operates on degrees of truth rather than binary true or false values. Fuzzy logic allows for multiple degrees of truth between 0 (completely false) and 1 (completely true), introduced by Zadeh in 1965 (Zadeh 1989). Knowledge in FIS is encoded through if-then rules where input conditions (antecedents) and output actions (consequents) are defined. Membership Functions (MFs) assign membership values to input variables, with different shapes such as triangle, trapezoidal, or Gaussian representing linguistic terms (Jang et al. 2005). In practice, FIS converts crisp inputs into fuzzy inputs using MFs, processes them using rule bases, and de-fuzzifies the output to continuous real values. The Takagi-Sugeno FIS is widely used in engineering applications of ANFIS (Irshaid and Abu-Eisheh 2023). ANFIS functions as a multilayer feed-forward ANN that parallels the Takagi-Sugeno FIS model. It integrates the learning capabilities of ANN with the reasoning capabilities of FIS within a unified framework (Jang et al. 2005). ANFIS initializes a FIS using an input–output training dataset and employs ANN's learning capability to optimize the fuzzy rules and membership function (MF) parameters. This optimization process involves iterative training using a designated learning algorithm over multiple epochs. The learning algorithm aims to minimize the difference between actual and predicted values, typically measured using RMSE (root mean square error). This study employs a hybrid learning algorithm, combining back-propagation gradient descent for input MF parameters and least squares estimation for optimizing output MF parameters. More details on this approach can be found in the literature (Irshaid and Abu-Eisheh 2023).

2.15. Applications of machine learning in water quality and health risk assessment

ML techniques have gained significant traction in water quality and health risk assessment due to their ability to model complex, nonlinear relationships and process large, multifaceted datasets effectively. Recent research demonstrates the successful application of various ML algorithms in predicting IWQIs and assessing GW suitability. For instance, Shaw and Sharma (2025) (Shaw and Sharma 2025) applied multiple regression alongside machine learning models, including ANN, Support Vector Machines (SVM), Classification and Regression Trees (CART), Conditional RF, and k-Nearest Neighbors (KNN), to predict irrigation water quality parameters such as SAR, KR, and the IWQI. Their study found that ML models, especially ANN, significantly outperformed traditional regression techniques in predictive accuracy, enhancing irrigation suitability assessments. Similarly, Singh et al. (2024) (Singh et al. 2024) utilized ML algorithms to evaluate GW quality in agricultural regions, highlighting the potential of these approaches to improve water resource management strategies by providing more reliable water quality evaluations. In arid environments, Mohammed et al. (2023) (Mohammed et al. 2023) integrated GIS data with ML models to spatially predict irrigation GW quality indices in the Nubian aquifer system. Their work illustrates how coupling GIS with ML facilitates detailed spatial analysis, which is critical for sustainable water management in resource-limited regions. An overview of these models and their applications is summarized in Appendix Table 1. Collectively, these studies exemplify the increasing adoption of ML techniques to improve the precision and efficiency of water quality and health risk assessments, thereby supporting more informed decision-making in environmental and water resource management.

2.16. Monte Carlo simulations (MCS) in hydrology: methodology and application

The MCS is employed in hydrology and other disciplines for conducting uncertainty analyses. This method utilizes statistical techniques involving a large set of random samples to approximate actual outcomes (Etu and Oyedepo 2018). By testing variables across a sufficient number of samples, MCS enhances precision and probability in results. Notably, the complexity of the problem determines the requisite number of samples/simulations. The MCS approach involves randomly sampling hypothetical observations and assigning specific probability distribution functions to each observation (Kabakuş and Tortum 2019). This process entails describing the studied process, its governing equations, and choosing a suitable probability distribution matching the data format.

Mathematically, MCS examines a flow comprising deterministic and stochastic processes, utilizing the law of large numbers and asymptotic theorems. Stochastic processes involve sequences of random variables indexed by time and event, whereas deterministic processes yield predictable outputs with known inputs. Conversely, stochastic processes are probabilistic and uncertain (Irshaid and Abu-Eisheh 2023). To describe statistical functions, three key terms from probability theory are employed: the sample space (Ω); the sample (A); and the probability (P[ ]).

The sample space (Ω) represents the entire population, while the sample (A) refers to events or subsets of the sample space (when the entire population cannot be evaluated). The function P[ ] is a probability function with a domain in the sample (Alfaica et al. 2023). In random experiments inherent to MCS, the sample space constitutes all possible outcomes. The function X(.) maps each outcome in the sample space (Ω) to a real number. The objective of MCS is to simulate this random variable multiple times, ensuring $X(.) \in R$ (Kabakuş and Tortum 2019). To apply MCS effectively, a cumulative distribution function (CDF) of the random variable X, denoted as FX(x), is determined. FX(x) operates within the domain of X and codomain [0,1], satisfying the condition $FX(x) = P[X \leq x]$ for all real numbers x. Based on observations of X, a probability distribution is inferred, and random numbers are simulated from the codomain [0,1]. The aim is to generate random observations "x" from the random variable X using the CDF F(x). Given that the distribution function is defined probabilistically ($F_X(x) = p\ [X \leq x]$), random observations are obtained via an inverse transform,

$$x = F^{-1}(p) \qquad \textit{Eq. 1}$$

The process of extracting random observations from the cumulative distribution function CDF F(x) starts by generating a uniform random number “p” between 0 and 1, which serves as the probability input for the inverse transform function F-1(p). Evaluating F-1(p) provides a random observation “x”. This process is repeated multiple times to approximate the true mean and standard deviation of the population. In hydrology, Monte Carlo Simulation (MCS) is used to compute flow by considering a set of flows characterized by time, positional components, and velocity vectors. Random processes are initiated for each set, generating new synthetic values that match the frequency distributions of observed flows. By incorporating multiple flow values and parameters, a collection of flows and new parameter values is obtained, with the goal of sampling distributions at time $t+\Delta t$. This process is repeated until there are minimal changes in the distribution across iterations, thus avoiding the need for complex integrations or matrix multiplications, focusing instead on a sequence of straightforward events (Alfaica et al. 2023)

3. MATERIALS AND METHODS

3.1. Study area description

3.1.1. Danube river lower basin in Hungary (SW)

This study assessed surface water quality in Hungary’s Danube River, focusing on three key regions, Hercegszántó, Dunaföldvár, and Baja, over the period from January 2013 to December 2019. Water samples were collected from seven monitoring sites along a 120-km stretch of the southern Danube (18°57′15.50″ E, 46°10′54.45″ N), covering the right, left, and middle stream sections at Dunaföldvár (S1–S3), Baja (S4), and Hercegszántó (S5–S7) (Appendix Figure 3). These locations were selected to encompass a variety of land use types: industrial (Dunaföldvár), urban (Baja), and agricultural (Hercegszántó), which are known potential sources of heavy metals and other contaminants. Sampling methodology, including timing, frequency, composite sampling techniques, and analytical procedures for heavy metals and physicochemical parameters, is fully described in Sections 3.2.1 and 3.2.2. In brief, samples were collected seasonally and processed following standardized protocols from USEPA (2007) and APHA (2017), ensuring data quality and comparability. The dominant soil types in the region include alluvial loams and clay-loam soils, with sandy riverbank substrates that influence the mobility and retention of pollutants such as Pb, Zn, and Mn (Fuchs et al. 2015). The Danube's lower basin has a continental climate, with average summer temperatures reaching 29 °C and winter lows around −2.6 °C. Annual precipitation averages 619 mm, primarily between May and July, potentially intensifying pollutant loads through surface runoff. These geographical, hydrological, and climatic characteristics form a comprehensive framework for interpreting spatial and seasonal variations in water quality across the study area (Saeed et al. 2023a).

To support the interpretation of potential non-point sources of contamination in the study area, a land use and land cover (LULC) classification map was prepared using recent satellite imagery and GIS analysis. The classification identified major land cover categories, including dense vegetation, built-up areas, barren land, and agricultural zones within the river catchment. Appendix Figure 4 illustrates the spatial distribution of these land cover types along the Danube river in the study area. The LULC information provides essential context for linking observed variations in water quality with surrounding land use practices, particularly in identifying areas where anthropogenic influence is likely to contribute to pollutant loading through surface runoff, leaching, or erosion.

3.1.2. Al-Jawf basin in Yemen (GW)

The Al-Jawf Governorate, situated in northwest Yemen near the Saudi Arabian border and about 170 km from the capital Sana'a, covers an area of roughly 30,620 km^2 and has a population of approximately 663,147. The region's economy is mainly based on agriculture, with key crops including wheat, corn, fodder, and barley, alongside small-scale industries such as textiles and food production. The landscape consists primarily of plains and stretches toward the Empty Quarter desert. The region lies within an arid climatic zone, characterized by low precipitation and high evapotranspiration. The Al-Jawf Governorate is segmented into 12 districts administratively, and the designated study region within Al-Jawf Governorate includes nine districts: Alhazm, Alkhaleq, Alghail, Almasloob, Almutun, Alzahir, Alhamidat, and Almatamah. This area lies geographically between longitudes 44°20′ to 45°0′ E and latitudes 16°0′ to 16°20′ N, as well as between longitudes 44°20′ to 44°40′ E and latitudes 16°10′ to 16°20′ N (Appendix Figure 5 A-B). Water sampling in this region involves wells with depths ranging from 10 to 165 meters, tapping into the shallow Quaternary aquifer, which is unconfined and composed of sands interspersed with clays and carbonates. Various studies have observed well depths ranging from 18 to 150 meters below the surface, penetrating similar geological layers within the aquifer (Alaug and Al-Wosabi 2015).

The Al-Jawf basin's subsurface geology is characterized by a complex sequence of sedimentary formations that reflect alternating transgressive and regressive depositional cycles, influenced by various rift phases. This sedimentary succession includes a mix of clastic rocks and carbonates, ranging from the Paleozoic to the Quaternary period. The stratigraphic sequence, listed from youngest to oldest, encompasses the Nayfa Formation (Berriasian-Hauterivian), the untied Tawilah Group (Cretaceous), Shuqra Formation (Bathonian-Oxfordian), Madbi Formation (Kimmeridgian), Wajid and/or Akbara Formations (Paleozoic), and Kuhlan Formation (Early Middle Jurassic) (Appendix Figure 6 A, B) (Alaug and Al-Wosabi 2015). The elevation of the GWLs in the Al-Jawf basin varies between 1086 and 1259 meters above mean sea level (AMSL), with GW flowing from the northwest to the southeast. Therefore, excessive extraction of GW, particularly in the middle parts of the basin, has led to a rapid decline in water levels, creating a cone of depression that aligns with the natural flow direction of the GW. The area, particularly Wadi Al-Jawf, receives minimal rainfall, restricting the replenishment of GW through direct precipitation. Instead, the main contributor to GW recharge is SW flowing from nearby mountainous regions into the plains and valleys. This SW gradually infiltrates the ground, restoring aquifers via a process called diffuse recharge. This method is vital in dry and semi-dry regions with limited rainfall, though it heavily relies on the intensity of rainfall in the mountainous zones and the soil's capability to absorb water. The main aquifer in the Al-Jawf basin comprises sandstone interbedded with shale, clay, and carbonate minerals such as limestone (Alaug and Al-Wosabi 2015). The subsurface geological formations provide a framework that significantly influences GW storage and flow dynamics in the region. Despite the vital role of diffuse recharge, the unsustainable rate of groundwater extraction threatens the long-term viability of the aquifer, emphasizing the need for careful management of water resources in this arid region

3.2. Water sampling

3.2.1. Heavy metal analysis of the lower Danube river basin (HMs-SW)

To evaluate seasonal variability in HM concentrations, the dataset was divided into two hydrologically meaningful periods: April to September and October to March. This classification was adopted to reflect typical hydrological and ecological dynamics in the region. The April–September interval corresponds to spring and summer, characterized by higher river discharge, enhanced surface runoff, and greater biological activity, all of which can affect contaminant transport and dilution. Conversely, the October–March period represents autumn and winter, when lower flow rates and reduced biological activity often result in different contaminant retention and mobilization behavior. This binning approach was preferred over the conventional four-season division (spring, summer, autumn, winter) to maintain statistical robustness and clarity in interpretation, given the limited number of samples per season. The two-period grouping thus provides a scientifically justified and practical framework for analyzing seasonal patterns in water quality without over-fragmenting the dataset. To determine the concentrations of HMs in SW, samples were collected in accordance with the United States Environmental Protection Agency (USEPA 2007) and APHA Standard Methods (2017) (APHA, 2017) for water sampling and preservation. All samples were collected in acid-washed polyethylene bottles, pre-rinsed with deionized water and 1% nitric acid to avoid contamination. Upon collection, water samples were immediately filtered through 0.45 µm cellulose nitrate membranes, acidified to $pH < 2$ using ultrapure nitric acid (HNO_3), and stored at 4 °C until analysis. The concentrations of heavy metals, including As, Cu, Cr, Pb, and Ni, were quantified using Electrothermal Atomic Absorption Spectroscopy, while Fe, Mn, and Zn were measured using Flame Atomic Absorption Spectroscopy. These analyses followed APHA Standard Methods 3113B (ETA-AAS) and 3111B (FAAS), as outlined in APHA (2017). Each sample (500 mL) was split into two 250 mL aliquots. For arsenic (As), a pre-treatment step was applied involving the addition of 0.5 mL of ascorbic acid, 0.5 mL of potassium iodide, and 3 mL of hydrochloric acid (HCl), allowing the solution to react for 2 hours before analysis. To ensure data accuracy and reproducibility, all measurements were conducted in triplicate, with a relative standard deviation (RSD) $\leq 5\%$. Calibration was performed using certified standard solutions and method blanks were run regularly to monitor instrument drift and contamination.

3.2.2. Physicochemical parameters of the lower Danube river basin (SW)

In 2019, a total of 85 surface water samples were collected seasonally (spring, summer, autumn, and winter) from seven fixed locations along a 120-km stretch of the lower Danube river in Hungary. At each site, 3-4 individual grab samples were taken per season, approximately 30 cm below the water surface in flowing midstream zones to ensure representative conditions. These samples were analyzed separately, and their values were averaged post-analysis to identify seasonal water quality trends. All samples were collected in 500 mL polyethylene containers, pre-cleaned with 10% nitric acid and rinsed with deionized water. Field parameters including pH, temperature, electrical conductivity, and total dissolved solids were measured in situ using calibrated portable instruments. Further chemical analyses were conducted on filtered samples. Cations and anions were measured using EDTA titration, flame photometry, titrimetric, and spectrophotometric methods. Target analytes included Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, HCO_3^{-}, CO_3^{2-}, NO_3^{-}, and SO_4^{2-}. All instruments were calibrated, and standard QA/QC procedures were followed to ensure accuracy and reliability.

3.2.3. Heavy metals and physicochemical characteristics of the Al-Jawf basin (GW)

In 2021 and 2022, GW samples were collected from the Al-Jawf region's quaternary aquifer in the for chemical analysis. In 2021, 27 samples were taken from wells and boreholes, while in 2022, 33 samples were obtained from boreholes. Both sets of samples were gathered in approximately 1.5 dm^3 polyethylene bottles, kept at 4 °C and 3 °C respectively, without acidification. The chemical composition of the water was analyzed for the presence of K^{+}, Ca^{2+}, Mg^{2+}, Na^{+}, SO_4^{2-}, Cl^{-}, HCO_3^{-}, and NO_3^{-}. The water analyses were conducted using various methods. In 2021, bicarbonates were determined by titration with a methyl orange endpoint, chlorides by Mohr titration with AgCl precipitation until silver chromate appeared, sulfates by nephelometry, nitrates by colorimetric analysis, calcium and magnesium by complexometric titration, and sodium and potassium by emission spectrometry (Al-Mashreki et al. 2023). In 2022, similar procedures were employed. Bicarbonate levels were measured by titration with a methyl orange endpoint, chloride content by argentometric titration with silver chromate formation indicating the endpoint, sulfates by nephelometry, nitrates by colorimetric analysis, calcium and magnesium by complexometric titration, and sodium and potassium by emission spectrometry. Mn and Fe were evaluated through the flame atomic absorption spectrometry. A summary of the sampling years, site numbers, and purpose of collection for Al-Jawf and other regions is presented in Appendix Table 2. These comprehensive analyses aimed to evaluate and ensure the groundwater quality in the Quaternary Aquifer across both years.

3.4. Quality assurance and control

The water quality examination followed the standard procedures established by the American Public Health Association (APHA) in 2012 (Rice et al. 2012). To ensure precise results from the field instruments, all devices were calibrated using buffer solutions and deionized water before beginning sample analysis. Several quality control techniques were used throughout the water testing process. The testing methods were verified by calibrating the equipment and evaluating their accuracy and dependability. Charge balance errors (CBE) were assessed using field data and confirmed in the lab. Each sample underwent triplicate analysis, with the mean values reported. Anion-cation balance errors were calculated using Eq.2 (Al-Mashreki et al. 2023), adhering to the neutrality principle, which holds that the total cations should equal the total anions in milliequivalents per liter. The CBE for all samples fell within the acceptable range of ±5%. The ionic balance (IB) was determined by calculating the proportion difference between the total cations (positively charged ions) and anions (negatively charged ions) in the water using the following formula:

$$IB = (TC - TA)/(TC + TA) \times 100 \qquad \text{Eq. 2}$$

where TC stands for the total cations and TA represents the total anions, both expressed in milliequivalents per liter. Furthermore, the accuracy of the analytical process was confirmed through thorough checks using the blank analysis method and Certified Reference Material.

3.5. Statistical analysis and data processing

3.5.1. Multivariate statistical methods

A correlation matrix, which is a statistical tool displaying correlation coefficients between different variables, typically results in a large matrix. However, this approach may not be the most effective for summarizing data in a large dataset. An alternative method, multivariate statistical techniques like Principal Component Analysis (PCA), can significantly reduce the complexity of datasets with numerous variables. PCA accomplishes this by transforming the original variables into a smaller set of new variables while preserving the key information from the original data. These new variables, known as principal components, are linear combinations of the initial variables (Helena et al. 2000; Shrestha and Kazama 2007; Allafta and Opp 2020). To conduct a valid PCA, several conditions must be satisfied, including the Kolmogorov-Smirnov test, Kaiser-Meyer-Olkin (KMO) test, and Bartlett's test. The Kolmogorov-Smirnov test was used to assess whether the data followed a log-normal distribution. This analysis showed that all variables under investigation fit a log-normal distribution with 95% confidence or greater. The KMO test was also employed to evaluate the dataset's suitability for PCA. This test measures sampling adequacy by determining the proportion of variance among the variables that might be shared. High KMO values suggest that PCA is appropriate, with a recommended threshold above 0.5 (Allafta and Opp 2020). In this research, the KMO value was found to be 0.65, indicating sufficient adequacy. Additionally, Bartlett's test of sphericity was applied to determine if the correlation matrix is an identity matrix. The null hypothesis of Bartlett's test states that the correlation matrix equals the unit matrix, while the alternative hypothesis suggests they are unequal. In this work, the null hypothesis was rejected as the significance level was 0 (less than 0.05), indicating significant relationships among variables (Allafta and Opp 2020). While PCA reduces the contribution of less significant variables, factor analysis (FA), when applied after PCA, can further streamline the data structure by diminishing the impact of minor variables even more. This is achieved by rotating the axes defined by PCA according to well-established rules, resulting in new variable sets called varifactors (VFs). The rotation of principal components makes the depiction of underlying factors more straightforward and expressive by reducing the contribution of less significant variables and enhancing the contribution of more significant ones. The rotation produces a new set of factors, each primarily comprising a subgroup of the original variables with minimal overlap, thereby partitioning the original variables into new, independent groups. While principal components are linear combinations of observable variables, varifactors can include unobservable, latent, or hypothetical variables. Before performing PCA and FA, we standardized our data using z-scale transformation to avoid misclassifications due to differences in the magnitudes and variances of numerical data values. PCA and FA were conducted using SPSS version 29 to determine the clustering of elements and ultimately identify the sources of elements in the river surface sediments (Vega et al. 1998; Liu et al. 2003; Simeonov et al. 2003; Yan et al. 2010; Zhang et al. 2015). CA is used for unsupervised pattern recognition, focused at dividing large volumes of data from one entity into several clusters. This method allows for the identification of the unique characteristics of each cluster group, as indicated by sources (Chen et al. 2022). In hydrochemical investigations, CA is especially useful for identifying hydrogeochemical processes in GW. Grouping gathered water samples allows researchers to discover major geological and hydrogeological groups (Barkat et al. 2021). The results of this clustering process are typically represented in a dendrogram, a tree-like diagram that significantly simplifies the complexity of the original dataset. Despite the reduction in complexity, the dendrogram effectively illustrates the relationships and proximities between the groups, offering a clear visual representation of the clustering activities (Mutea et al. 2021; Al-Mashreki et al. 2023).

3.5.2. Data analysis, processing, and spatial distribution

The physicochemical properties and IWQIs were analyzed using SPSS software version 29. The hydrogeochemical processes and water classifications were determined by examining the cation and anion compositions, with the use of Piper, Chadha, and Gibbs diagrams. The DIAGRAM tool was employed to demonstrate the connection between aquifer features and water chemistry (Vasanthavigar et al. 2012). To pinpoint significant elements in the water and improve water quality assessments through simplified data analysis, CA and PCA were applied using SPSS 29. These methods are essential for extracting valuable insights from large, complex datasets and for identifying underlying trends and relationships among variables (Rakotondrabe et al. 2018; Al-Mashreki et al. 2023).

3.5.2.1. GIS mapping procedure and IDW interpolation analysis

The GIS mapping procedure involved establishing spatial relationships among chemical properties using various water quality indices (IWQIs). These indices were developed to connect chemical properties spatially. QGIS 3.36.3. was utilized to visualize these indices, generating maps through raster interpolation and spatial analysis techniques, particularly employing the IDW method. IDW interpolation was implemented using the Spatial Analyst Extension in QGIS 3.36.3. Experimental results from laboratory analysis of water samples collected from seven sites along the Danube River in Hungary and sixty sites within the Al-Jawf basin in Yemen were compiled into an Excel file and converted into a shapefile. Average data values for each parameter at specific points were used to calculate values for each interpolated cell, with the river network serving as a mask. Cells corresponding to the Danube River and Al-Jawf watershed were assigned values based on the first input raster in the output raster. Water quality classifications were determined by mapping pollutant distributions spatially, using a legend corresponding to the data range for each parameter. IDW was chosen as the preferred interpolation method for its superior accuracy over kriging. It effectively handles widely spaced data points and local variations, preserving abrupt changes, which is crucial for this study. In contrast, kriging is better suited for smoothly varying data with clear spatial correlation. The decision was based on evaluating mean squared error, the primary criterion for comparing measured and predicted values across most physicochemical properties (Khouni et al. 2021). The general equation for IDW (Eq. (3)) is as follows:

$$Z(x)_0 = \frac{\sum_{i=1}^{n} z(x_i).d_{ij}^{-p}}{\sum_{i=1}^{n} d_{ij}^{-p}} \qquad \text{Eq. 3}$$

In the context of IDW interpolation:

Z_i represents the interpolated value of a grid node,

Z_j denotes the neighboring data points,

d_{ij} signifies the distances between the grid node and the data points.

This formula outlines how IDW interpolation computes the value at a grid node based on nearby data points, weighting them inversely according to their distances.

3.6. Indexing techniques and modeling approaches

3.6.1. Heavy metal pollution analysis index (HPI) and metal index (MI)

The HPI is widely considered the most efficient method for assessing the contamination levels in water samples caused by HMs (Al-Hejuje[1] et al. 2017; Saeed et al. 2023a). This index evaluates water quality for human consumption by analyzing metal contamination and takes into account the overall presence of HMs in the water. The HPI is computed using a weighted arithmetic mean method, which includes assigning a rating scale for each selected parameter and applying weights to each contamination factor. The rating scale spans from 0 to 1, determined based on the importance of each quality indicators or by comparing values to recommended standards (Shankar 2019). The HPI offers a quantitative evaluation of water quality, focusing on the concentration of HMs in the sample. The formulas used to calculate the HPI are outlined below (Eq. 4 and 5).

$$HPI = \frac{\sum_{i=1}^{n} W_i Q_i}{\sum_{i=1}^{n} Wi} \qquad \text{Eq. 4}$$

Here, Q_i) represents th sub-index element, n is the number of parameters analyzed, and wi denotes the weight of each parameter, calculated as $1/S_i$. S_i symbolizes the standard value of each parameter. Q_i also signifies the sub-index of the boundary, determined by Eq. 5.

$$Q_i = \sum_{i=1}^{n} 100 \times \frac{C_i}{S_i} \qquad \text{Eq. 5}$$

The HPI was determined based on the levels of 8 HMs: Cr, Cu, Mn, Fe, Ni, Zn, Pb, and one metalloid, As, due to its toxic nature. A three-tier modified scale is commonly employed to accurately represent various levels of HM contamination. This scale classifies pollution as high (HPI > 30), moderate (15 ≤ HPI ≤ 30), or low (HPI < 15) (Edet and Offiong 2002; Qu et al. 2018).

In contrast, the MI for drinking water evaluates the combined potential effects of HMs on human health, offering a general assessment of water quality (Ojekunle et al. 2016). The MI is based on the principle that the toxicity of HMs is linearly related to their concentration. HMs can cause a variety of chronic and acute toxic effects on different organs. The MI is computed through a thorough evaluation of the current metal concentrations. If any metal's concentration surpasses its upper allowable limit (UAL), the water quality is deemed compromised. The MI concept was initially proposed by Tamasai and Cini (Shankar 2019; Saeed et al. 2023a) and can be represented as follows (Eq. 6):

$$M_i = \sum_{i=1}^{i} \frac{C_{ave}}{UAL_i} \quad Eq.\ 6$$

Where, C_{ave} signifies the mean concentration of each analyzed HM, while UAL_i refers to the maximum permissible limit for the *i*th metal in the sample. The MI is categorized into six levels: severely impacted (MI > 6), highly impacted (4 > MI > 6), moderately impacted (2 > MI > 4), slightly impacted (1 > MI > 2), clean (0.3 > MI > 1), and very clean (MI < 0.3) (Withanachchi et al. 2018).

3.6.2. The potential ecological risk index of heavy metals

The RI for heavy metals, developed by Hakanson in 1980, is used for evaluating the potential threats that HMs may pose to a given ecosystem. This index takes into account elements like concentration levels, types of metals, their sensitivity, toxicity, and natural background concentrations(Xie et al. 2013). While it has broad application across different scientific disciplines, this study focused on using it to assess the ecological risks of HMs in river water. The equation is provided as follows (Eq. 7):

$$RI = \sum E_r^i = T_r^i \times \left\{\frac{C_{ave}}{C_{bg}}\right\} \qquad \text{Eq. 7}$$

In this context: E_r refers to the potential ecological risk factor of a substance; T_r represents the toxic response factor for the specific HM; C_{ave} indicates the mean amount of each HM within the given sample; and C_{bg} signifies the background concentrations of each HM. RI reflects the total potential ecological risk posed by contamination. The RI is divided into four risk categories: extremely high (< 120), significant (60-120), moderate (30-60), and low (> 30) (Yuan et al. 2015).

3.6.3. Human health risk assessment of heavy metals (HHRA)

3.6.3.1. Non-carcinogenic human health risk method

The application of drinking water polluted with hazardous substances heightens the likelihood of both carcinogenic and non-carcinogenic conditions in humans (Qu et al. 2018; Bineshpour et al. 2021). In this research, the non-carcinogenic risks of Cr, As, Fe, Cu, Pb, Mn, Ni, and Zn were evaluated following methods proposed by the U.S. Environmental Protection Agency (Selvam et al. 2022). The USEPA (2004) (USEPA 2004) designed a HHRA framework to estimate non-carcinogenic HHRA from HM exposure in SW and GW through inhalation, ingestion, and skin contact. The main risks come from direct digestion and skin absorption of water (Qu et al. 2018; Mukherjee et al. 2020; Selvam et al. 2022; Saeed et al. 2023a). This approach computes the amount of contaminants ingested by humans using the chronic daily intake (CDI) method, which estimates the daily pollutant intake through ingestion (CDI ingestion) and through skin absorption (CDI dermal), using Eqs. 8 and 9, respectively(USEPA 2004; Tokatli and Ustaoğlu 2020; Jehan et al. 2020).

$$CDI_{oral} = \frac{C_{ave} \times IR \times EF}{BW \times AT} \times ED \qquad \text{Eq. 8}$$

$$CDI_{dermal} = \frac{C_{ave} \times ET \times EF \times Kp \times SA \times CF}{BW \times AT} \times ED \qquad \text{Eq. 9}$$

Where, CDI is measured in mg/kg/day, C_{ave} refers to the average level of each HM, expressed in mg/L, IR indicates the intake rate, which is 2.2 L/day for adults and 1.8 L/day for children. EF signifies the exposure frequency, set at 350 days per year for both adults and children. ED represents exposure duration, which is 70 years for adults and 6 years for children. ET is the exposure time, with adults exposed for 0.58 hours per day and children for 1 hour per day. Kp stand for the permeability coefficient, measured in cm/Hr. SA refers to the skin surface area, which is 18,000 square centimeters for adults and 6,600 square centimeters for children. BW is body weight, with adults weighing 70 kilograms and children 15 kg. CF represents the conversion factor, 1×10^{-3} liters per cubic centimeter. AT represents the average time for carcinogenic risks, with adults having 25,550 days and children 2,190 days. (Saleem et al. 2019; Xu et al. 2020; Selvam et al. 2022). In the subsequent step, the HQ (Hazard Quotient) was computed by dividing the CDI by the Reference Dose (RfD) for both oral and dermal exposure using Eqs. 10 and 11, respectively (Saha and Paul 2019; Mthembu et al. 2022).

$$HQ_{dermal/oral} = \frac{CDI_{dermal}/CDI_{oral}}{RfD_{dermal}/RfD_{oral}} \qquad \text{Eq. 10}$$

$$RfD_{dermal} = RfD_{oral} \times ABS \qquad \text{Eq. 11}$$

In the last step, the overall non-carcinogenic risks were assessed by calculating the HI (Hazard Index) using Eq.12 (Jehan et al. 2020).

$$HI = HQ_{oral} + HQ_{dermal} \qquad \text{Eq. 12}$$

Toxic metals with (HI or HQ > 1) have the potential to negatively affect HHRA, whereas those with (HI or HQ < 1) are deemed to have no harmful effects (Selvam et al. 2022). All constant values are demonstrated in (Appendix Table 6).

3.6.3.2. Carcinogenic human health risk method

According to Li and Zhang's approach (Li and Zhang 2010), the level of carcinogenic risk (CR) was calculated from the CDI value using the carcinogenic slope factor (CSF) by Eq. 13. The outcome indicates the possibility of developing cancer over a lifetime due to exposure to carcinogenic substances. Generally, the acceptable risk range is between 1×10^{-6} and 1×10^{-4} (Li and Zhang 2010; Saeed et al. 2023a).

$$CR = CDI \times CSF \qquad \text{Eq. 13}$$

Where CSF represents the carcinogenic slope factor.

3.6.4. Geochemical modeling and saturation index (SI)

3.6.4.1. The index of processes influencing surface water and groundwater chemistry

A different method for determining the basis and connections between main components includes using various ion concentration comparisons such as [$SO4^{2-}$] versus [Ca^{2+}], [$HCO3^{-} + SO4^{2-}$] versus [$Mg^{2+} + Ca^{2+}$], [Cl^{-}] versus [Na^{+}], and [$HCO3^{-}$] versus [$Mg^{2+} + Ca^{2+}$]. Additionally, the two chloralkali indix (CAI) values (CAI-I and CAI-II) (Eqs. 14 and 15) were employed in this research to identify the minerals present in the aquifers and to assess GW exchange ions (Al-Mashreki et al. 2023).

$$CAI\text{-}I = Cl^{-} - (Na^{+} + K^{+})(_{meqL}) / Cl^{-}(_{meqL}) \qquad \text{Eq. 14}$$

$$CAI\text{-}II = Cl^{-} - (Na^{+} + K^{+})(_{meqL}) / (SO_4^{2-} + HCO_3^{-} + CO_3^{2-} + NO_3^{-})\ (_{meqL}) \qquad \text{Eq. 15}$$

3.6.4.2. Saturation Index (SI)

A speciation method was applied to calculate the SI of minerals in SW samples from the lower Danube river basin in Hungary and GW samples from the Al-Jawf basin. The SI of a mineral indicates its saturation level based on the existing conditions of the surrounding environment. In this research, the SI was applied as a predictive tool to evaluate the potential presence of reactive minerals in the water, relying on water samples rather than solid-phase samples or detailed mineral analysis (Khadr et al. 2020; Al-Mashreki et al. 2023). Eq. (16) was applied for SI computation:

$$SI = \log\left\{\frac{IAP}{K_{sp}}\right\} \qquad \textit{Eq. 16}$$

Where, IAP stands for "ion activity product," and K_{sp} represents the "solubility product" at a specified temperature. A saturation index (*SI*) of zero indicates equilibrium, $SI > 0$ indicates oversaturation, and $SI < 0$ signifies undersaturation with respect to mineral equilibrium.

3.6.5. Drinking and irrigation water quality indices

3.6.5.1. Drinking water quality index calculations (DWQI) for the Danube river

The WQI is an important tool for evaluating the total condition of SW, especially in terms of its potentiality for human consumption (Khadr and Schlenkhoff 2021). Using the WQI for initial water quality evaluations plays a key role in assisting policymakers with decisions related to sustainable water resource management and policy-making. As a rating system, the WQI provides a broad summary of how various water quality factors combine to affect the overall condition of the water (Alquraish et al. 2021). In contrast to traditional approaches, which concentrate on specific contamination levels or individual water quality factors, the WQI offers an innovative method for assessing the overall health of rivers (Al-Mashreki et al. 2023). Its distinctiveness lies in its ability to merge multiple environmental variables into a single numeric value, representing the current state of water quality. This shift from traditional methodologies underscores the WQI's potential to deliver a more holistic and integrated approach to water quality evaluation. Consequently, instead of comparing inconsistent assessment outcomes from multiple individual indicators, the WQI method proves to be a reliable method for effectively assessing and managing water quality, offering a clear understanding of overall conditions (Karunanidhi et al. 2021b). In Hungary's lower Danube river watershed, the calculation of WQI considered a range of key parameters, including pH, temperature, TDS, EC, Na^+, Ca^{2+}, K^+, Mg^{2+}, Cl^-, SO_4^{2-}, NO_3^-, and HCO_3^-. This comprehensive approach to various indicators underscores the WQI's ability to encompass a wide array of water quality factors, enabling a more holistic and integrated assessment approach (Eq. 17).

$$DWQI = \left\{\frac{w_i}{\sum_{i=1}^{n} wi}\right\} \times \left\{100 \times \frac{C_i}{S_i}\right\} \quad \text{Eq. 17}$$

w indicates the weight allocated to each parameter, and "n" signifies the total number of variables considered in this study. C_i refers to the concentration of each chemical parameter, expressed in mg/L for each water sample, while S_i denotes the relevant WHO water quality standards for each chemical parameter, also provided in mg/L as per WHO guidelines. The calculation of w_i for each parameter is determined in accordance with the prescribed standards eq. (18):

$$w_i = K/S_i \quad \text{Eq. 18}$$

Here, the constant of proportionality, denoted as "K," is computed using eq. (19):

$$K = 1/\sum 1/S_i \quad \text{Eq. 19}$$

The DWQI involves assigning a weight (w_i) to each surface water variable, followed by the determination of the relative weight (W_i).

3.6.5.2. Irrigation indices (IWQIs)

The provided equations in Appendix Table 3 represent the employed indices used to assess the appropriateness of GW from the quaternary aquifer for irrigation applications.

The evaluation of water quality intended for irrigation is confined within a scale ranging from 0 to 100. This assessment is based on various physicochemical factors, including SAR, EC, Na^+, Cl^-, and HCO^{3-} (Meireles et al. 2010; Abbasnia et al. 2019). The IWQI, a unitless index ranging from 0 to 100, is determined by considering multiple parameters such as EC, Na^+, SAR, Cl^-, and HCO^{3-}. It can be calculated as illustrated in eqs. 20 and 21:

$$IWQI = \sum_{i=1}^{n} Q_i W_i \qquad \text{Eq. 20}$$

Here, the values for (Q_i) and (W_i) signify the assessed quality within predefined tolerance limits and the respective weighting assigned to each criterion, as outlined in Supplementary Appendix Table 4.

$$Q_i = Q_{max} - \left(\left[\left(X_{ij} - X_{inf}\right) \times Q_{imap}\right]/X_{amp}\right) \qquad \text{Eq. 21}$$

In the given context:

Xij represents the measured values of the parameter.

Xinf denotes the lower limit of the classes.

Qimap signifies the range of class amplitudes.

Xamp pertains to the class amplitudes specific to the parameter being analyzed.

Finally, the W_i is obtained as follows (Eq. 22):

$$W_i = \frac{\sum_{j=1}^{k} F_j A_{ij}}{\sum_{j=1}^{k} \sum_{i=1}^{n} F_j A_{ij}} \qquad \text{Eq. 22}$$

where, F represents the automated logging of element j, while A denotes the chosen parameter *i* primarily influenced by factor *j*. Here, *i* spans from 1 to n, indicating the selected physicochemical parameters, and j spans from 1 to k_{ij}, representing the chosen factors.

3.6.6. Adaptive neuro-fuzzy inference system (ANFIS)

The ANFIS is a form of ANN that integrates fuzzy logic for decision-making (Veza et al. 2021). It merges the learning abilities of neural networks with the clarity of fuzzy logic. ANFIS is adaptable, capable of handling both classification and regression tasks, and is particularly effective in modeling complex nonlinear relationships between inputs and outputs. The process of using ANFIS involves several key stages (Adnan Ikram et al. 2022; Al-Mashreki et al. 2023). Model Design: This initial stage involves creating the ANFIS model by specifying the amount of input and output factors, the amount of fuzzy rules, and the setup of the adaptive network layer. The model's structure is selected depending on the intricacy of the relationships between input and output factors, and the number of fuzzy rules is adjusted to align with the input variables and the desired complexity of the model. Fuzzy Partitioning: This phase involves partitioning the input space into fuzzy regions through methods such as clustering or grid division. The objective is to separate the input space into parts where the input variables show similar patterns in relation to the output variable. Rule-Based Generation: In this stage, a set of fuzzy rules is created based on the fuzzy partitions and the connections between input and output variables. Each rule includes an antecedent (the "if" condition) and a consequent (the "then" outcome), with the antecedent usually featuring fuzzy membership functions that describe how input variables relate to fuzzy regions. Model Training: During this phase, the ANFIS model is trained using a dataset and an optimization method, such as gradient descent or particle swarm optimization. This training process adjusts the model's parameters to improve its performance by reducing the error between predicted and actual output values. Different optimization techniques and loss functions may be used to achieve this goal. The ANFIS was employed to determine the connections between IWQIs and water quality indicators, which were subsequently represented as fuzzy if-then rules (Appendix Figure 7). The ANFIS models used a Sugeno-type FIS with bell-shaped functions for the input membership functions, utilizing five such functions. In contrast, the outputs were defined by linear membership functions. Appendix Figure 8 depicts the ANFIS model, which includes a multilayer feedforward structure with separate x- and y-input networks.

The ANFIS was employed in this study to investigate the connections between IWQIs and water quality metrics. These connections were articulated as fuzzy if-then rules, enabling the prediction of IWQIs from input variables. The rule foundation of the Sugeno model is delineated by Eqs. 23-24.

$$\text{When } a = A_1 \text{ and } b = B1\text{ , } f_1 = p1 \times a + q_1 \times b + r_1 \qquad \text{Eq. 23}$$

$$\text{Conversely, when } a = A_2 \text{ and } b = B_2,\ f_2 = p_2 \times a + q_2 \times b + r_2. \qquad \text{Eq. 24}$$

The ANFIS system comprises five layers, illustrated in Appendix Figure 9. Membership functions for fuzzy sets *A and B* are represented by μA_i and μB_i respectively. *A and B* signify orders of magnitude, while a and b serve as indicators for an indirect identification function. During the forward pass of the training algorithm, relevant constraints p_i, q_i, and r are adjusted, and the output fi falls within an inconsistent region defined by the FIS concept. Additional details regarding the ANFIS can be found in Khadr (Athamena et al. 2023).

3.6.7. Performance evaluation of the simulation models

To evaluate the precision and performance of the ANFIS model in predicting IWQs, several statistical metrics were used for performance assessment. These metrics are commonly applied to compare a model's predictions with actual data or those from other models, offering quantitative insights into the model's predictive power, dependability, and accuracy. One of these metrics is the Nash-Sutcliffe efficiency coefficient (NSE), represented by the following equations (Eqs. (25-28)):

Nash–Sutcliffe efficiency coefficient (NSE)

$$E = 1 - \left[\frac{\sum_{i=1}^{n}\left(IW_{oi}-IW_{fi}\right)^2}{\sum_{i=1}^{n}(IW_{oi}-IW_{O})^2}\right] \qquad \text{Eq. 25}$$

The mean absolute error (MAD)

$$MAD = \frac{\sum_{i=1}^{n}\left|IW_{oi}-IW_{fi}\right|}{n} \qquad \text{Eq. 26}$$

The absolute variance fraction, R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(IW_{oi}-IW_{fi}\right)^2}{\sum_{i=1}^{n}(IW_{oi})^2} \qquad \text{Eq. 27}$$

The root-mean-square error (RMSE)

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}\left(IW_{oi}-IW_{fi}\right)^2}{n}} \qquad \text{Eq. 28}$$

IW_o is the observed IWQ index, n is the number of data points, IW_f is the predicted IWQ index, and IW is the average of the observed Irrigation water quality index. Appendix Figure10 illustrates the model architecture utilised by the ML algorithms. Used ML approached are illustrated in Appendix Table 5.

3.6.8. Random forest algorithm (RF)

The RF algorithm is a robust and versatile ensemble learning technique widely utilized in various fields, such as water quality evaluation, air pollution forecasting, and ecological monitoring (Wongnakae et al. 2023; Ahmadi et al. 2023; Zhang et al. 2024). Its core principle involves using bootstrap resampling to create multiple subsets from the original dataset, with each subset serving as the basis for an individual decision tree. These trees collectively form the random forest, enabling sophisticated classification and regression tasks. In this model, each decision tree is constructed using randomly selected samples and features, while the out-of-bag error provides an unbiased estimate of the model's generalization error. This approach minimizes the impact of noise variables and mitigates overfitting risks (Montes et al. 2021).

3.6.9. Model Evaluation and Performance Metrics

After determining the importance rankings of the features, those with the highest relevance were selected for model construction. To assess the stability and accuracy of the different models, we utilized key performance metrics, including the coefficient of determination (R^2), mean squared error (MSE), mean absolute percentage error (MAPE), and accuracy (ACC). The equations for these metrics are defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}-y_i)^2}{\sum_{i=1}^{n}(\bar{y}-y_i)^2} \qquad \text{Eq. 29}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}-y_i|^2 \qquad \text{Eq. 30}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}-y_i}{y_i}\right| \times 100\% \qquad \text{Eq. 31}$$

where y_i refers to the real value of each sample and $\hat{y}$ represents the predicted value obtained using the RF model, $\bar{y}$ is the average of the measured values of the indicator.

3.7. Utilization of Monte Carlo simulation model (MCS)

In our study, the main aim of the MCS was to assess the probabilistic distributions of various factors, including concentrations of HMs and NO_3^-, IR, ET, ED, EF, BW, SA, and K_p (Appendix Figure 11). This analysis aimed to outline the probabilistic distributions associated with uncertainties in the assessment metrics (Vasanthavigar et al. 2013). Integrating the MCS model with the USEPA's HHRA framework allows for a comprehensive analysis of uncertainties connected to HHRA from HMs exposure. This involves generating a probabilistic distribution for CR readings. In this invistigation, the MCS method was employed to predict oral HQ readings for both children and adults, thereby reducing uncertainty and improving the credibility of non-carcinogenic HHRA evaluations for HMs and NO_3^- in the lower Danube river basin of Hungary. The input parameters included concentrations of HMs and NO_3^-, as well as relevant variables from Equations 11 and 12. To ensure the accuracy of the MCS, Python was used to perform 10,000 iterations, resulting in consistent observed and predicted HQ values, thus validating the model's effectiveness. While distributions for heavy metal and NO_3^- concentrations were based on existing data, other factors such as ET, IR, SA, and BW were modeled as normal distributions to more accurately represent their true distributions (Qu et al. 2018; Saeed et al. 2023b). The MCS was implemented using Python 3.9 programming language. Key Python libraries used in the simulation and data analysis included:

- NumPy (Harris et al. 2020) for numerical operations and random sampling,
- Pandas (McKinney 2010) for data handling and management,
- Matplotlib (Hunter, 2007) and Seaborn (Waskom 2021) for visualization,
- SciPy (Virtanen et al. 2020) for statistical distributions and calculations.

These libraries facilitated efficient simulation runs, data processing, and graphical presentation of results. All code was documented to ensure reproducibility and transparency.

4. RESULTS AND DISCUSSION

4.1 Heavy metal assessment of the Danube river

4.1.1 Heavy metal concentrations in surface water

This research investigated the distribution of eight HMs at seven representative locations (Appendix Table 7). The concentrations of Cu, Cr, As, Mn, Fe, Ni, Zn, and Pb were recorded and averaged over two distinct periods: April to September and October to March. During the April–September period, the average levels of Cr, Cu, As, Mn, Fe, Ni, Zn, and Pb were 1.43, 4.03, 1.29, 53.02, 520.63, 2.70, 15.00, and 1.31 µg/L, respectively. The concentration hierarchy from highest to lowest was Fe > Mn > Zn > Cu > Ni > Cr > Pb > As. In the October–March period, the mean concentrations were 1.36, 3.69, 1.44, 39.71, 403.36, 2.51, 1.27, and 15.59 µg/L, respectively. The order of concentrations, from highest to lowest, was Fe > Mn > Zn > Cu > Ni > As > Cr > Pb. It is important to note that the average amounts of Mn and Fe surpassed the regulatory limits set by the EU Directive and WHO in 2017.

4.1.2 Geospatial modeling and analysis

Elevated levels of most HMs were detected in both the upstream and downstream sections of the study region, areas marked by high population density and significant industrial operations. These activities lead to the release of both anthropogenic and industrial waste into the river. Furthermore, the widespread application of fertilizers in farming practices alongside the riverbank adds to the increased concentrations of HMs in the water. The application of GIS for water quality information systems and spatial analysis, employing IDW interpolation, facilitated the mapping of water quality indicators across the Danube and Al-Jawf basins. This spatial modeling helped pinpoint critical areas requiring targeted management intervention and guided the refinement of river basin management strategies. GIS integrates spatial and temporal data, allowing effective visualization and interpretation of water quality variations, making it a powerful tool for environmental monitoring and decision-making.

4.1.3 Specific heavy metals analysis

Although the Danube river is not directly used as a primary drinking water source, health-based thresholds such as those from the World Health Organization (WHO, 2017) are applied in this study for comparative evaluation. These benchmarks allow an assessment of potential risks under hypothetical exposure conditions, such as incidental ingestion, indirect usage, or potential abstraction and treatment. This approach also supports consistency in interpreting HM concentrations across spatial and seasonal patterns. To provide a clear overview of HM distribution in the Danube, Appendix Table 8 presents the raw concentration values for all sampling sites (S1–S7) during the April–September and October–March periods. To visualize spatial variability and identify pollution hotspots, Appendix Figures 12(A–h) display the interpolated HM concentration maps for As, Cr, Cu, Fe, Mn, Ni, Pb, and Zn for the April–September season. Corresponding maps for the October–March period are presented in Appendix Figures 13(A–H). These maps reveal clear spatial trends, with higher concentrations of Fe, Mn, and Zn particularly evident in the southern and southwestern zones. This suggests possible inputs from localized industrial or land-use sources. The seasonal variation captured through this mapping supports a more nuanced interpretation of HM dynamics in the Danube River basin.

4.1.3.1 Arsenic (As)

High levels of As in water pose serious risks to both human health and the environment. As, a highly hazardous element, has been associated with immune system disorders, reproductive issues , and skin cancer (Tokatli and Ustaoğlu 2020; Selvam et al. 2022). The levels of As in water were monitored during two time periods: April to September and October to March. In the April–September period (Appendix Figure 12A), concentrations varied between 0.9 and 3.90 µg/L, with an average of 1.29 µg/L, while from October to March (Appendix Figure 13A), levels ranged from 0.1 to 3.79 µg/L, with an average of 1.44 µg/L. These concentrations remained below the WHO (2017) guideline of 10 µg/L for drinking water, indicating low potential health risks under hypothetical human exposure scenarios (Appendix Table 7).

4.1.3.2 Chromium (Cr)

Although Cr concentrations peaked at 3.49 µg/L, with a mean of 1.43 µg/L from April to September (Appendix Figure 12B) and increased to a peak of 4.89 µg/L, with a mean of 1.36 µg/L from October to March (Appendix Figure 13B), these levels remained well within the WHO (2017) drinking water guideline of 50 µg/L. The slight decrease in Cr concentrations during the April–September period may reflect seasonal changes in anthropogenic inputs or dilution processes, contributing positively to water quality(Saeed et al. 2023a).

4.1.3.3 Copper (Cu)

Cu is a vital element for human health (Samantara et al. 2017; Selvam et al. 2022). In this study, Cu concentrations were higher during the April–September period, ranging from 2.6 to 8.59 µg/L, with an average of 4.03 µg/L (Appendix Figure 12C). The levels slightly declined during the October–March period, reaching a maximum of 8.09 µg/L, with an average of 3.69 µg/L (Appendix Figure 13C). All values were far below the WHO (2017) drinking water guideline of 2000 µg/L, indicating that Cu concentrations do not pose a health concern in the context of potential human exposure.

4.1.3.4 Iron (Fe)

The high Fe concentrations found in SW are likely due to multiple sources, such as industrial waste discharge and runoff from rainwater. Iron levels above 300 µg/L are known to cause harm to plumbing and laundry systems (WHO 2017; Mollo et al. 2022). In this study, Fe concentrations ranged from 10 to 2250 µg/L, with a mean of 520.6 µg/L during April–September (Appendix Figure 12D), and from 1 to 2900 µg/L, with an average of 403.3 µg/L during October–March (Appendix Figure 13D) (Appendix Table 7). These values exceeded the WHO (2017) aesthetic-based guideline of 300 µg/L. The elevated Fe levels observed may be attributed to multiple sources such as industrial discharges, urban runoff, and sediment interactions. In particular, one plausible contributing factor is the 2010 Ajka alumina plant disaster in western Hungary. The accident resulted in the release of red mud, a by-product of bauxite refining, into the Torna Creek and subsequently into the Marcal and Rába rivers, which are hydrologically connected to the Danube river. Previous studies, including (Uzinger et al. 2015), reported that this red mud contained iron oxide (Fe_2O_3) concentrations ranging from 40% to 45%, suggesting that a significant iron load was introduced into the river system. While this study does not directly detect red mud residues at the sampling sites, the spatial proximity and hydrological linkage support the possibility that legacy contamination from the Ajka spill may still influence iron concentrations in downstream segments of the Danube (Saeed et al. 2023a). Further investigations would be needed to confirm this connection with greater certainty.

4.1.3.5 Manganese (Mn)

Although Mn is a necessary nutrient, excessive exposure can negatively impact human health (Obasi and Akudinobi 2020). In our study, Mn levels ranged from 5 to 350 µg/L, with a mean of 53.02 µg/L during the April–September period (Appendix Figure 12E), and from 5 to 350 µg/L, averaging 39.7 µg/L from October to March (Appendix Figure 13E). In several sampling locations, concentrations exceeded the WHO (2017) guideline of 50 µg/L. These exceedances suggest localized sources or mobilization under changing redox conditions and highlight the need for continued monitoring and evaluation (Appendix Table 7).

4.1.3.6 Nickel (Ni)

As shown in Appendix Table 7, Ni concentrations ranged from 1.6 to 9.59 µg/L, with an average of 2.70 µg/L during April–September (Appendix Figure 12F), and from 0.7 to 6.39 µg/L, with an average of 2.51 µg/L during October–March (Appendix Figure 13F). These values remained below the WHO (2017) drinking water guideline of 70 µg/L. Although not alarming in magnitude, nickel is classified as a Group 1 human carcinogen by the International Agency for Research on Cancer (IARC, 2012), and its presence in aquatic systems warrants precautionary observation (WHO 2017; Mollo et al. 2022).

4.1.3.7 Lead (Pb)

During April–September, Pb levels ranged from 0.2 to 4.19 µg/L, averaging 1.31 µg/L (Appendix Figure 12G), while in the October–March period, levels ranged from 0.3 to 8.99 µg/L, with an average of 1.27 µg/L (Appendix Figure 13G). All observed concentrations were below the WHO (2017) drinking water guideline of 10 µg/L. The minor seasonal decline in Pb levels may reflect changes in industrial discharge and reduced runoff of petroleum-based pollutants into the Danube river (Saeed et al. 2023a).

4.1.3.8 Zinc (Zn)

Zn is essential for immune function and human metabolism (Samantara et al. 2017; Karunanidhi et al. 2021a; Selvam et al. 2022). Zn concentrations in this study ranged from 2.1 to 46.99 µg/L, with an average of 15.0 µg/L during April–September (Appendix Figure 12H), and from 7.2 to 40 µg/L, averaging 15.59 µg/L during October–March (Appendix Figure 13H). All values were well below the WHO (2017) guideline of 3000 µg/L for drinking water. This suggests low potential risk associated with zinc under hypothetical exposure and highlights stable seasonal behavior. Variations between periods may reflect both anthropogenic and natural geochemical processes (Sharma et al. 2019). Natural factors such as shifts in geological formations, weather patterns, and soil composition can impact HM levels in SW. On the other hand, human actions, such as industrial and agricultural practices, contribute to HM contamination. Therefore, ongoing monitoring and evaluation of HM levels in SW are essential for identifying potential threats to human health and the environment.

The first three sites (S1, S2, and S3) are located downstream from the site of the Ajka red mud spill, which occurred upstream and may represent a potential point source. However, other upstream or nearby sources could also contribute, and Ajka is not assumed to be the only possible origin. At these sites, concentrations of Fe and Mn are notably higher than the limits established by the WHO and the EU. This increase is likely due to the Ajka accident, which discharged large amounts of these metals into the river. As the contaminated water flows downstream, it reaches these initial sites, resulting in the observed elevated metal levels. At the midpoint (S4), the concentrations of Fe and Mn decrease, likely due to dilution from water mixing with additional sources, which lowers the metal levels. Nevertheless, as the river continues to the ultimate sites (S5, S6, and S7), Mn and Fe amounts rise again, suggesting that other sources beyond the Ajka incident are contributing to the higher metal levels. The land use and land cover (LULC) map (Appendix Figure 4) was referenced to explore potential sources of diffuse pollution along the river corridor. Specific land cover types, such as barren land, built-up areas, and agricultural zones, are known to contribute to pollutant loading through surface runoff, soil erosion, and leaching, especially during rainfall events. Although soil concentrations were not directly measured in this study, the spatial distribution of these land use categories provides an important context for understanding potential pollutant pathways into the aquatic system. The presence of urban and agricultural areas along the riverbanks in the study region suggests that these land uses may be responsible for the elevated concentrations of Mn and Fe observed at the downstream sites. Moreover, while trace element patterns reveal higher Fe concentrations in certain areas, there is no direct evidence linking these increases to the Ajka incident. This ambiguity highlights the need for further analysis, such as REE (rare earth element) fingerprinting, to determine whether residual red mud contamination persists or if other anthropogenic sources (e.g., industrial discharge, agricultural runoff) are responsible. Without such confirmatory testing, the exact origin of these elevated metal levels remains uncertain.

4.1.4 Heavy metal pollution index (HPI) and metal index (MI)

The HPI model is an important method for assessing the overall contamination level in SW. This tool helps assess the impact of HMs on water quality and helps in monitoring and managing potential health risks linked to HM exposure (Rahman et al. 2022). The highest HPI readings for the HMs examined in the lower Danube river basin were recorded between October and March. The average HPI for April through September was 21.91, with values ranging from 17.63 to 25.17, while the average HPI from October to March was 19.42, with values ranging from 15.82 to 25.59. According to the results and referencing Edet and Offiong (2002) (Edet and Offiong 2002), almost all samples from both periods fell within the moderate pollution range of the HPI (15–30). It is important to note that the SW was generally not contaminated with the HMs studied, except for Fe and Mn. Nevertheless, the substantial presence of Mn and Fe at all sites led to higher HPI values at those sampling points (Appendix Figure 14A). To assess the extent of HM contamination in water and its effect on water quality, the MI approach was utilized along with the HPI index, applying the maximum permissible limits specified by the EU- Water Framework Directive (WFD) (The European Parliament and the Council of the European Union 2020). The MI averages for the periods April–September and October–March were found to be 4.13 and 3.28, respectively. These results point to serious pollution during April–September and moderate pollution from October–March. When compared to the maximum permissible values set by the EU-Directive Framework, the level of contamination indicates a substantial presence of HMs in the water samples. This highlights the pressing need for ongoing monitoring and efforts to enhance water quality in the lower Danube River basin in Hungary (Appendix Figure 14B).

4.1.5 Spearman correlation and heatmap cluster analysis

Before performing the statistical analysis, the normality of the input was tested using the Kolmogorov–Smirnov test. The interactions among different HM parameters in the Danube SW can offer insights into their origins and distribution. For elements that did not follow a normal distribution, the nonparametric Spearman correlation matrix was used, as illustrated in Appendix Figure 15. It is important to note that statistical significance (e.g., $p < 0.01$ or $p < 0.05$) does not necessarily indicate a strong correlation. Many of the detected correlations, while significant, have low to moderate strength (e.g., $r < 0.4$), and should be interpreted with caution in the context of environmental processes and potential sources. During the April–September period (Appendix Figure 15A), with a significance threshold of $p < 0.01$, several positive correlations were detected. Specifically, Cr showed positive correlations with Fe ($r = 0.33$), Mn ($r = 0.23$), and Zn ($r = 0.37$). Likewise, Fe was positively correlated with Mn ($r = 0.40$) and Zn ($r = 0.39$), while Mn was positively correlated with Pb ($r = 0.32$) and Zn ($r = 0.31$). At a significance level of $p < 0.05$, positive correlations were also found between As and Fe ($r = 0.17$), Cu and Ni ($r = 0.19$), and Ni and Pb ($r = 0.17$). Although statistically significant at $p < 0.01$, the observed positive correlations among Fe, Cr, Zn, Mn, and Pb (with correlation coefficients ranging from 0.31 to 0.40) are generally weak to moderate in strength. These results suggest a potential association or co-occurrence, but caution is needed when inferring shared sources solely based on such correlation values.

During the period from October to March (Appendix Figure 15B), positive correlations were found at a significance level of $p < 0.01$. Specifically, lead (Pb) had positive correlations with Cr ($r = 0.24$), Fe ($r = 0.40$), and Mn ($r = 0.272$), while Fe and Mn also displayed a positive correlation ($r = 0.515$). Additionally, at a significance level of $p < 0.05$, Zn showed a positive correlation with As ($r = 0.19$), Mn was positively correlated with Cr ($r = 0.199$), and Fe had a positive correlation with Cu ($r = 0.17$). The substantial positive correlations among Cr, Fe, Mn, and Pb during October–March, significant at $p < 0.01$, suggest that these elements might originate from a common source. In contrast, the positive correlations among As, Cu, Zn, and Ni at a significance level of $p < 0.05$ imply that these elements may have different sources. The heatmap cluster tree provides an effective visualization for determining groups of variables with similar correlation patterns within a correlation matrix. The results obtained from this heatmap cluster tree were in agreement with those from the spearman correlation matrix, which then was applied to non-normally distributed data. The positive correlations observed in the Spearman matrix among HMs like Zn, Cr, Mn, Fe, and Pb were also represented in the heatmap cluster tree. Likewise, the diverse origins of metals such as Zn, As, Cu, and Ni identified in the Spearman matrix were evident in the heatmap cluster tree as well. These results underscore the usefulness of the heatmap cluster tree as a complementary tool to the correlation analysis matrix, especially for non-normally distributed data, in uncovering correlation patterns among the variables. Storm runoff plays a significant role in transporting HMs from diverse sources, such as industrial operations, transportation systems, and human activities, into rivers (Xiong et al. 2021). Consequently, HMs may enter river waters from reservoir banks and areas affected by water level fluctuations. Under certain environmental conditions, like specific redox potentials and pH levels, HMs that are bound in sediments can be released into the water, resulting in pollution (Gao et al. 2018). Therefore, it is crucial to adopt strategies to reduce HM contamination in riverbank soils. These strategies include ecosystem preservation and rehabilitation through methods like reforestation, limiting adjacent commercial activities, and converting agricultural land into forest areas (Xiong et al. 2021). Mn, which is widely found in the Earth's crust and frequently occurs alongside iron, indicates a potential link between Mn and Fe. CA identified three main groups of HMs: Cluster 1, which includes Fe, Cr, Mn, and Zn, likely represents HMs that share identical sources or behaviors within the river, possibly connected to industrial activities, urban discharge, or agricultural practices. In contrast, Cluster 2, comprising Pb, Ni, and Cu, exhibits distinct patterns from Cluster 1, suggesting different sources or behaviors. These elements are often associated with industrial discharges (e.g., metal plating, machinery manufacturing), traffic-related emissions (e.g., brake and tire wear, fuel combustion), and urban runoff, indicating anthropogenic influences typical of built-up areas. Arsenic, forming a separate Cluster 3, shows a unique pattern that may reflect localized sources. These could include geological contributions from naturally enriched sediments, agricultural inputs such as pesticides or irrigation runoff, or isolated industrial activities. The clustering suggests that As follows a distinct pathway or mobilization process compared to other metals in the dataset. A positive correlations, such as those between Mn and Fe (0.4), Mn and Zn (0.32), and Mn and Pb (0.31), suggest that these metals may have shared sources or interconnected behaviors within the river. Moderate positive correlations, like those between Cr and Fe (0.33) and Cr and Zn (0.37), also imply co-occurrence due to similar sources or environmental conditions. Low correlations (below 0.1) between Ni and As with other metals indicate that their presence may be influenced by distinct factors or sources, showing limited interdependence with other metals in the dataset. The clustering suggests that metals in Cluster 1 may originate from industrial sources, while Cu, Pb, and Zn could be associated with urban and agricultural runoff. As might be linked to geological sources. These findings offer valuable insights into potential pollution sources and can aid in developing environmental management strategies and mitigate HM contamination in the Danube River.

4.1.6 Potential ecological risk index (RI)

Developed in 1980, Hakanson's potential ecological HI is a widely identified approach for assessing the impact of HM pollution on sediment and aquatic environments. This index integrates several factors, including HM concentrations, their ecological and toxicological effects, and broader environmental consequences. By combining these factors, the index provides a comprehensive evaluation of the ecological risks associated with HM contamination in sediment and water systems. This research focuses on evaluating the potential RI for HMs, as shown in Appendix Figure 16. The mean RI values calculated for samples collected between April and September and October and March were 3.35 and 3.34, respectively. These findings suggest a relatively low RI ($RI < 30$) from the HMs studied in the lower Danube river basin in Hungary. Appendix Figure 16 demonstrates the relationship between RI values and sampling locations within the study area, highlighting the link between pollution levels and ecological risks.

The region under study, characterized by extensive agricultural and residential areas adjacent to the river and its tributaries, may contribute to higher concentrations of HMs in sediments, posing risks to the Danube river ecosystem. Research by Szabó et al. (Szabó et al. 2020) highlights key factors influencing HM distribution in floodplain soils, demonstrating that organic matter primarily governs macro-elements (e.g., Fe, Mg, K), while clay content and landform type significantly control micro-elements (e.g., Cr, Cu, Ni, Pb, Zn). Their findings emphasize that geomorphic features, particularly clay-rich swales, enhance HM retention in soils, suggesting similar depositional environments along the Danube may act as HM sinks. Given these dynamics, assessing the ecological risks associated with HM contamination in aquatic environments is essential. Such evaluations are critical for developing targeted management strategies and policies to mitigate the threats posed by trace contaminants, ensuring the long-term health of riverine ecosystems.

4.1.7 Health risk assessment

The HI for both non-carcinogenic and carcinogenic risks were assessed by calculating oral and dermal HQ. The results demonstrate the overall health hazards to people from various HM exposures between April and September, as well as October and March, impacting both children and adults.

4.1.7.1 Non-carcinogenic health risk

To assess non-carcinogenic risks in both children and adults, the concentrations of toxic elements including As, Cr, Cu, Fe, Mn, Ni, Pb, and Zn were analyzed (Appendix Figure 17). During the April–September period, the oral hazard quotient (HQ) for adults ranged as follows: As (1.2×10^{-1} to 1.4×10^{-1}), Cr (1.1×10^{-2} to 1.8×10^{-2}), Cu (2.6×10^{-3} to 3.3×10^{-3}), Fe (7.3×10^{-3} to 2.9×10^{-2}), Mn (5.0×10^{-2} to 9.1×10^{-2}), Ni (3.4×10^{-3} to 4.7×10^{-3}), Pb (2.5×10^{-2} to 3.5×10^{-2}), and Zn (1.2×10^{-3} to 1.8×10^{-3}). For children in the same period, the oral HQ varied within the following ranges: As (4.6×10^{-1} to 5.3×10^{-1}), Cr (4.3×10^{-2} to 6.9×10^{-2}), Cu (1.0×10^{-2} to 1.3×10^{-2}), Fe (2.8×10^{-2} to 1.1×10^{-1}), Mn (1.9×10^{-1} to 3.5×10^{-1}), Ni (1.3×10^{-2} to 1.8×10^{-2}), Pb (9.4×10^{-2} to 1.3×10^{-1}), and Zn (4.6×10^{-3} to 6.8×10^{-3}) (Fig. 17A). Conversely, during the October–March period, oral HQ values for adults were as follows: As (1.3×10^{-1} to 1.7×10^{-1}), Cr (9.3×10^{-3} to 1.8×10^{-2}), Cu (2.5×10^{-3} to 3.1×10^{-3}), Fe (9.0×10^{-3} to 2.3×10^{-2}), Mn (3.3×10^{-2} to 8.3×10^{-2}), Ni (3.5×10^{-3} to 4.2×10^{-3}), Pb (2.1×10^{-2} to 3.3×10^{-2}), and Zn (1.3×10^{-3} to 1.9×10^{-3}). In children, HQ values during this period ranged from: As (4.9×10^{-1} to 6.5×10^{-1}), Cr (3.6×10^{-2} to 7.0×10^{-2}), Cu (9.6×10^{-3} to 1.2×10^{-2}), Fe (3.4×10^{-2} to 8.7×10^{-2}), Mn (1.2×10^{-1} to 3.2×10^{-1}), Ni (1.3×10^{-2} to 1.6×10^{-2}), Pb (8.1×10^{-2} to 1.3×10^{-1}), and Zn (4.9×10^{-3} to 7.4×10^{-3}) (Appendix Figure 17B).

From the HQ oral data, it appears that the risks to human health from As, Cr, Cu, Fe, Mn, Ni, Pb, and Zn oral are typically greater for children than for adults during the April–September timeframe. It's important to emphasize that while the HQ oral values remain below the acceptable threshold of 1, these values are contextual to the studied location and timeframe. Human health risks can vary based on factors such as the length and frequency of exposure, individual susceptibility, and environmental conditions. During the April–September period, the HQ for adults ranged between 6×10^{-4} and 7×10^{-4} for arsenic (As), 4.2×10^{-3} to 6.9×10^{-3} for chromium (Cr), 1×10^{-5} to 4×10^{-4} for copper (Cu), 2×10^{-4} to 7×10^{-4} for iron (Fe), 6×10^{-3} to 1.07×10^{-2} for manganese (Mn), 1×10^{-4} to 3×10^{-4} for nickel (Ni), 0 to 1×10^{-5} for lead (Pb), and 1×10^{-5} to 3×10^{-5} for zinc (Zn). For children during the same period, dermal HQ values ranged between 1.7×10^{-3} and 1.9×10^{-3} for As, 1.25×10^{-2} to 2.03×10^{-2} for Cr, 1×10^{-4} to 2×10^{-4} for Cu, 5×10^{-4} to 2×10^{-3} for Fe, 1.76×10^{-2} to 3.17×10^{-2} for Mn, 2×10^{-4} to 3×10^{-4} for Ni, 1×10^{-4} to 2×10^{-4} for Pb, and 1×10^{-4} for Zn (Appendix Figure 17A). In contrast, during the October–March period, dermal HQ values for adults ranged from 6×10^{-4} to 8×10^{-4} for As, 3.5×10^{-3} to 7×10^{-3} for Cr, 2×10^{-5} to 4×10^{-5} for Cu, 2×10^{-4} to 5×10^{-4} for Fe, 3.9×10^{-3} to 9.8×10^{-3} for Mn, 1×10^{-4} to 2×10^{-4} for Ni, 1×10^{-5} to 2×10^{-4} for Pb, and 1×10^{-5} to 2×10^{-5} for Zn. For children in the same period, the dermal HQ values ranged between 1.8×10^{-3} and 2.4×10^{-3} for As, 1.05×10^{-2} to 2.06×10^{-2} for Cr, 1×10^{-4} for Cu, 6×10^{-4} to 1.6×10^{-3} for Fe, 1.14×10^{-2} to 2.89×10^{-2} for Mn, 2×10^{-4} to 3×10^{-4} for Ni, 1×10^{-4} to 2×10^{-4} for Pb, and 1×10^{-4} for Zn (Appendix Figure 17B). According to both oral and dermal HQ values, children are at greater risk from exposure to Cr, As, Cu, Fe, Ni, Mn, Zn, and Pb through skin contact compared to adults during October–March. Nonetheless, it is important to emphasize that while dermal HQ values are below the acceptable threshold of 1, these results are specific to the location and period studied. Actual health risks can differ based on different factors such as exposure frequency, duration, individual sensitivity, and environmental conditions. Continuous monitoring of these metals in the Danube river is crucial, especially for at-risk groups like children, given that the river serves as a major source of drinking and irrigation water in Hungary. Recent research has shown that consuming vegetables cultivated on Danube river sediment contaminated by industrial activity poses health risks, particularly to children. Although daily intake metrics for adults remained within safe limits, the health risk index (HRI) exceeded 1 for children exposed to cadmium in radish and to copper and zinc in green peas. These findings emphasize the vulnerability of children to metal toxicity and suggest that even moderate exposure from locally grown crops can present long-term risks (Kovács-Bokor et al. 2021).

4.1.7.2 Monte Carlo simulation approach (MCS)

The MCS was utilized to forecast the HQ (oral & dermal) values for Cu, As, Fe, Cr, Mn, Ni, Pb, and Zn, as well as the CR (oral and dermal) values for Cr, As, and Pb, considering both adults and children.

4.1.7.3 Non-carcinogenic health risk

The outcomes of the MCS revealed that the projected HQ values for all HMs remained below the standard thresholds (HQ < 1). Hence, the anticipated levels of exposure through both oral and dermal pathways are unlikely to present substantial health risks for either adults or children. Moreover, it's significant to acknowledge that hazard evaluations often depend on conservative assumptions and entail uncertainties in available data. Therefore, continual monitoring of exposure levels and regular updates to risk assessments are imperative as fresh information emerges (Appendix Figures 18 and 19).

4.1.7.4 Carcinogenic health risk

This work employed the MCS approach to replicate carcinogenic risk assessments derived from the deterministic method across oral and dermal exposure routes in both adults and children. The 95th and 5th percentile risk exposures delineated the most favorable and adverse scenarios, respectively.

4.1.7.4.1. Arsenic

The probability analysis (Appendix Figures 20 and 21 A–D) indicates that oral CR values are notably higher in children compared to adults within the studied groups. During both the October–March and April–September periods, the 95th percentile CR values for children were recorded as 3.6×10^{-4} and 3.3×10^{-4}, respectively, whereas for adults, these values were 9.65×10^{-5} and 8.636×10^{-5}, respectively. This data underscores a heightened health risk for children, with the most pronounced risk occurring in October–March, where the 95th percentile value of 3.6×10^{-4} was the highest observed among all groups. Additionally, the 95th percentile dermal CR values for children were 4.537×10^{-5} and 4.055×10^{-5}, compared to 1.55×10^{-5} and 1.380×10^{-5} for adults. Overall, the carcinogenic risk for children exceeded the acceptable carcinogenic risk threshold of 1.0×10^{-4}, indicating a potential increased risk of developing cancer in the future due to prolonged arsenic exposure throughout the watershed. The elevated CR in children may be partly attributed to their lower body weight (Fallahzadeh et al. 2019). As, a major carcinogen typically present in low concentrations in natural environments, is largely introduced into SW through human activities such as arsenic and arsenic-containing metal mining, the use of arsenides and arsenic in various processes, and coal combustion (Zhang et al. 2022).

4.1.7.4.2. Chromium (Cr)

Appendix Figures 20 and 21 A–D display histograms that depict the estimated CR from Cr exposure through dermal and oral routes for both adults and children. The data for oral CR shows that children face a higher cancer risk compared to adults, with their oral CR values exceeding the acceptable risk threshold of 1.0E−04 (oral CR 95% = 1.2×10^{-4}), whereas the CR levels for adults were minimal during both April–September and October–March. In contrary, the dermal CR results for the 95th percentile reveal cancer susceptibility in both groups. For children, the dermal CR values during April–September and October–March were 9.03×10^{-4} and 8.53×10^{-4}, respectively, while for adults, they were 3.050×10^{-4} and 2.92×10^{-4}, respectively. Therefore, lowering Cr levels in the environment is expected to decrease the related health hazards from exposure (Fallahzadeh et al. 2019).

4.1.7.4.3. Lead (Pb)

The histograms shown in Appendix Figures 20 and 21 A–D were generated through 10,000 iterations to predict the CR from lead exposure via dermal and oral pathways for both children and adults. The analysis indicates that children are at a greater risk of developing cancer than adults, with oral CR values surpassing the critical limit of 1.0×10^{-4} (oral CR 95% = 1.1×10^{-4}) in the April–September period. In contrast, during October–March, cancer risk was minimal for both children and adults, with CR values being very low (2.922×10^{-5}) for children and for adults in both periods (2.922×10^{-5} and 2.818×10^{-5}, respectively). Additionally, the dermal CR values at the 95th percentile show no significant CR for either group. For children, the CR values for dermal exposure were 4.134×10^{-5} during April–September and 4.004×10^{-5} during October–March, while for adults, these values were 1.400×10^{-5} and 1.363×10^{-5}, respectively. Therefore, reducing environmental lead levels is essential for alleviating health risks, particularly for children.

4.2 Physical chemical properties of the Danube river

4.2.1 Hydrochemical properties of Hungary's Danube river

The spatial distribution of maximum concentrations of key hydrochemical parameters is shown in Appendix Figure 22. This visualization highlights regional variations in water quality across the Danube river Basin in southern Hungary. The values represent the highest concentration recorded at each sampling location during the monitoring period. These maps support the interpretation of spatial trends discussed throughout this section, including north-to-south gradients in Ca^{2+}, Mg^{2+}, and Cl^-, as well as localized enrichments in NO_3^- and Fe, potentially linked to agricultural and urban influences. When evaluating the SW quality of the Danube river in Hungary, it is useful to classify the collected samples based on their physicochemical characteristics and ion compositions, and then compare these against established irrigation water standards (Appendix Table 8) (Ayers and Westcot 1985). The pH levels of the samples (Appendix Figure 22A) vary between 7.6 and 8.6, with an average value of 8.1, reflecting a slightly alkaline nature that aligns with the recommended pH range (6.5-8.4) for irrigation (Ayers and Westcot 1985; Salem et al. 2023). The EC values (Appendix Figure 22L) show considerable fluctuation, ranging from 360 to 680 µS/cm, but all measurements are well below the maximum irrigation standard of 3000 µS/cm. TDS concentrations (Appendix Figure 22B) range from 199 to 370 mg/L, with all samples meeting irrigation standards. Ca^{2+} is the most prevalent cation (Appendix Figure 22C), followed by Mg^{2+} (Appendix Figure 22D), Na^+ (Appendix Figure 22E), and K^+ (Appendix Figure 22F), with average concentrations of 53.8, 15.37, 16.06, and 2.6 mg/L, respectively. Among the anions, HCO_3^- is the most abundant (Appendix Figure 22G), followed by SO_4^{2-} (Appendix Figure 22J), Cl^- (Appendix Figure 22I), CO_3^{2-} (Appendix Figure 22H), and NO_3^- (Appendix Figure 22K), with average levels of 189.5, 36.2, 23.36, 9.59, and 7.7 mg/L, respectively. These elevated concentrations are likely due to the dissolution of evaporite minerals and cation exchange with mineral clays (Bel Hadj Salem et al. 2012; Ahmed et al. 2020; Mhamdi et al. 2022). Na^+ levels, ranging from 11 to 25.5 mg/L, are below the FAO's irrigation threshold (Ayers and Westcot 1985), though higher concentrations in northern samples suggest possible agricultural runoff accumulation. K^+ levels range from 0.1 to 4 mg/L, averaging 2.6 mg/L, which is suitable for irrigation, with elevated levels in southern regions potentially resulting from potash feldspar weathering or the application of fertilizers (Salem et al. 2023). Ca^{2+} concentrations exhibit a gradual rise from southern to northern regions, ranging between 39 and 68 mg/L. Importantly, all samples show Ca^{2+} amounts below the irrigation guidelines threshold, with these notable levels possibly originating from the dissolving of gypsum and the incongruent dissolution of dolomite, a process commonly referred to as dedolomitization (Blubaugh et al. 1991; Al-Mashreki et al. 2023; Salem et al. 2023). The dedolomitization process is described by the following equation (Eq.32):

$$CaMg^{2+}(CO_3^{2-})_2(s) + Ca\,SO_4^{2-}.2H_2O(s) + 2H^+ \rightleftharpoons CaCO_3(s) + Ca^{2+} + Mg^{2+} + SO_4^{2-} + 2HCO_3^- + 2H_2O$$
Eq. 32

Mg^{2+} levels ranged widely from 12 to 24 mg/L, showing an increasing trend from south to north along the research area. Despite this variation, all samples remained below the FAO standards (Ayers and Westcot 1985; Ben Moussa et al. 2021). The low Mg^{2+} concentrations might be related to the dissolution of ferromagnetic minerals and cation-exchange processes (El Mejri et al. 2014; Ben Moussa et al. 2021). HCO_3^- levels ranged from 125 to 255 mg/L, and all measurements were within the permissible irrigation limit of approximately 610 mg/L. Cl^- concentrations ranged from 15 to 40 mg/L across all samples, all of which were below the irrigation limit. However, higher chloride levels were observed in the northern region, likely due to increased human activities in agrarian and industrial/urban areas. SO_4^{2-} concentrations in the SW ranged from 27 to 51 mg/L, with most samples being below the FAO's standard limit. The presence of SO_4^{2-} is likely due to anthropogenic activities and/or the dissolution of gypsum (Eq. 33) and/or anhydrite minerals (Eq. 34), as described by the following equations:

$$CaSO_4^{2-}, (2H_2O) \rightleftharpoons Ca^{2+} + SO_4^{2-} \qquad \text{Eq. 33}$$

$$CaSO_4^{2-} \rightleftharpoons Ca^{2+} + SO_4^{2-} \qquad \text{Eq. 34}$$

Fe and Mn (Appendix Figure 22N and 22O, respectively) exhibit localized concentrations possibly linked to redox conditions. Recent long-term hydrological studies in the Pannonian Ecoregion, which includes the Hungarian Danube river, have revealed that surface water warming has been most pronounced during the winter and spring seasons, with relatively lower but still notable increases in summer and autumn. Over the past 30 to 40 years, both lakes and rivers in the region have experienced a substantial intensification in the frequency and duration of SW heatwaves. Specifically, rivers showed an increase of approximately 1.2 heatwave events per decade, contributing to an additional 22 days per decade of cumulative heatwave duration. These thermal shifts are expected to significantly influence the hydrochemical behavior of river systems, affecting parameters such as dissolved oxygen, solubility of HMs, and nutrient cycling (Li et al. 2024).

4.2.2 Surface water facies and source determination

4.2.2.1 Surface water categories

SW can be categorized into specific water types based on ion dominance. Tools such as Piper diagram are useful for determining these chemical water types (Piper 1944). According to the Piper diagram, the Danube River's surface water falls into a single chemical category, namely the Ca-Mg-

HCO_3 type (Appendix Figure 23A). This classification is marked by a high concentration of bicarbonate, signifying ongoing replenishment, while the majority of cations tend to align with the calcium end of the spectrum in most samples. Conversely, the Gibbs diagram, which examines the link between water chemistry and geological factors, shows that all samples fall within the rock weathering domain. This is attributed to water-rock interactions occurring along the flow route from south to north (Appendix Figure 23B). It implies that the composition of SW is predominantly influenced by solute enrichment from carbonate mineral dissolution, ion exchange with silicate minerals, and the return of irrigation water (Derwich et al. 2012; Bel Hadj Salem et al. 2012; Gad et al. 2020a). Additionally, the water samples show a tendency towards the rock weathering pole, reflecting the impact of cation-exchange processes.

4.2.2.2 Ion exchange

The study utilized statistical methods to examine the connection and ratios of key ions, aiming to determine the main factors influencing SW chemistry in the area of investigation (Appendix Figure 24). In aquifer systems with a significant presence of clay minerals, ion exchange is a critical process affecting SW mineral content. These clay minerals typically balance their electrical charge by capturing monovalent cations like Na^+ and K^+, and simultaneously releasing divalent cations such as Ca^{2+} and $Mg^{2+,}$ or the reverse. Appendix Figure 24A depicting $[(Ca^{2+}+Mg^{2+}) - (Na^{+}+ K^{+})]$ vs. $[HCO^{3-} - (SO_4^{2} + Cl)]$ illustrates that all SW samples fall within the Ca^{2+}-Mg^{2+}-HCO_3^- class, indicating continuous recharge of the river and dominance of Ca^{2+}, Mg^{2+}-and HCO_3^- ions in the water chemistry along the flow path. This underscores the significance of ion exchange in SW mineralization in the Danube river (Gad et al. 2020a), a process closely associated with cationic interactions with clay minerals observed in the studied river network. Throughout this process, Ca^{2+} and Mg^{2+} ions originally adsorbed onto the outer layer of clay minerals interchange with Na^+ and K^+ ions in SW. This ionic exchange mechanism can be represented by the following eq. 35 (Gad et al. 2020a; Gaagai et al. 2023).

$$(Ca^{2+}, Mg^{2+})\text{-Clay} + 2(Na^{+}, K^{+}) \rightleftharpoons (Na^{+}, K^{+})\text{-Clay} + (Ca^{2+}, Mg^{2+}) \qquad \text{Eq. 35}$$

Ion exchange processes or silicate weathering are responsible for decreasing the $(Ca^{2+} + Mg^{2+})/HCO_3^-$ ratio and the $Ca^{2+}+Mg^{2+} / (HCO_3^- + SO_4^{2-})$ ratio to values below 1 (Appendix Figure 24 B-C). In ion exchange, $Ca^{2+}+ Mg^{2+}$ ions are extracted from water while Na^+ and K^+ ions are introduced (Salem et al. 2023). Plotting the summation of Ca^{2+} and Mg^{2+} ions against $HCO_3^- + SO_4^{2-}$ ions on a linear graph helps identify the source of Mg^{2+} and Ca^{2+} in water samples (Al-Mashreki et al. 2023; Salem et al. 2023). Ratios below 0.5 may indicate ion exchange or bicarbonate enrichment as the primary causes of Mg^{2+} + Ca^{2+} depletion. All samples fall below the 1:1 line, indicating higher proportions of HCO_3^- than Ca^{2+}+ Mg^{2+}, suggesting sources of HCO_3^- ions other than the dissolution of calcite and dolomite. The linear connection observed between Na^+ and Cl^- (Appendix Figure 24D) suggests a discrepancy in the concentrations of these ions within the water samples. This connection indicates that the dissolution of halite is not a major source of Na^+ and Cl^- ions. Instead, the higher concentration of chloride ions compared to Na^+ points to anthropogenic sources, such as agricultural runoff or industrial discharges. Most samples fall to the right of the 1:1 line, suggesting the presence of additional sources of Cl^- or possible Na^+ removal due to recharge water (Appendix Figure 24D). The analysis of the ionic balance between Ca^{2+} and Mg^{2+} (Appendix Figure 24E) demonstrates that all samples are positioned above the 1:1 line, which indicates an abundance of Ca^{2+} relative to Mg^{2+}. This observation suggests that the river in the study area acquires Ca^{2+} from sources other than dolomite dissolution, potentially from the dissolution of calcite or from agricultural runoff containing Ca^{2+} from fertilizers.

4.2.2.3 Chlor-alkali indices

The CAI approach was employed to determine the primary mechanism–whether it is ion exchange or reverse ion exchange–affecting the mineral composition in the lower Danube river Basin. Overall, the CAI values, including CAI-I and CAI-II, were negative for all water samples (Appendix Figure 25 A-B). These negative values suggest a strong tendency towards ion exchange processes, especially between K^+ and Na^+ in the SW of the study area, and between Ca^{2+} and Mg^{2+} in the surrounding geological formations. The analysis of various ionic ratios led to the conclusion that ion exchange primarily influences the SW chemistry in this region.

4.2.3 Geochemical modeling and mineral saturation state

We utilized the PHREEQC model (Parkhurst and Appelo1 1999) to analyze mineral concentrations, SIs, and the ability of SW to either precipitate or dissolve minerals. This approach calculated saturation indices for important minerals such as gypsum, dolomite, calcite, halite, aragonite, and anhydrite, as well as the potential partial pressure of CO_2. The model inputs consisted of pH, temperature (in °C), EC, TDS, and major anions and cations (Appendix Figure 26). The results revealed a negative partial pressure of CO_2 below saturation levels in the lower Danube river watershed, suggesting that water recharge is insufficient compared to extraction from the surrounding aquifer. Additionally, it was observed that CO_2 concentrations decrease along the water flow direction, which is consistent with a concurrent reduction in SW recharge in that direction.

4.2.4 Statistical analysis

4.2.4.1 Cluster analysis

To assess the similarity among SW samples in the study region, we applied hierarchical CA using the Mahalanobis distance metric combined with Ward's linkage method. This approach was selected to better account for correlations among variables, providing a more robust clustering outcome than the Euclidean distance. As shown in Appendix Figure 27A, the dendrogram reflects relationships among physicochemical parameters and HMs, capturing variance structures inherent to the dataset. All parameters were logarithmically transformed and standardized (Z-scores) prior to analysis to approximate normality and ensure comparability. The dendrogram identifies three principal clusters but with clearer separation and internal consistency. Notably, Cluster 1 groups HCO_3^- and TDS, indicating a strong association between alkalinity and overall salinity. This reflects the buffering role of bicarbonate and its contribution to dissolved solids, possibly linked to carbonate weathering and GW–SW interactions. Cluster 2 comprises Na^+, Cl^-, Mg^{2+}, and SO_4^{2-}, which may represent mineral dissolution from evaporitic or saline sources (Eid et al. 2024, 2025a). The grouping suggests a dominant geogenic influence, possibly from natural saline intrusion or lithological contributions. Cluster 3 includes Ca^{2+}, NO_3^-, Fe, and Mn, suggesting a mixed influence. The presence of nitrate along with Fe and Mn may reflect agricultural runoff or redox-driven mobilization of metals under variable oxidation states (Saeed et al. 2024).

4.2.4.2 Principal component analysis (PCA)

In our research, we applied the PCA to pinpoint the potential influences affecting the water quality parameters under investigation. PCA revealed four main factors with eigenvalues > 1, which had a considerable impact on the collected data during the sampling period. Appendix Table 9 presents a detailed account of the PCA results. By using PCA with Kaiser normalization (≥ 0.75), we identified 4 PC at all sampling sites. These components were crucial in determining the condition of the Danube SW in the study area. The PCA analysis yielded four primary components that together explained 82.098% of the total variance, each with eigenvalues greater than 1 (Appendix Figure 27 B-C). PC1 alone accounted for 49.76% of the overall variance and had an eigenvalue of 7.47. Variables associated with PC1 were significant indicators, representing both human activities and natural processes influencing SW chemistry. Key variables such as Ca^{2+}, Mg^{2+}, Na^{+}, Cl^{-}, SO_4^{2-}, HCO_3^{-}, EC, NO_3^{-}, and TH showed strong positive loadings.

On the other hand, PC2, which accounted for 13.53% of the variance with an eigenvalue of 2.029, highlighted the effects of anthropogenic activities on SW resources. It showed a positive correlation with pH levels and CO_3^{2-} concentrations, suggesting that localized human activities within the study area could impact water quality by altering pH and CO_3^{2-} amounts (Uddin et al. 2023). Although silicate weathering is likely the primary natural driver of pH and CO_3^{2-} variations in the study area, the influence of anthropogenic activities, particularly agriculture and industrial discharge, cannot be excluded. These activities are known to introduce ions such as nitrates, phosphates, and sulfates into aquatic systems, which can alter carbonate equilibrium and affect pH levels (Yu et al. 2016; Li et al. 2023). While direct source measurements were beyond the scope of this study, the land use context and observed ion patterns suggest that both natural and human-induced factors likely interact to shape SW chemistry. Further targeted research, including isotope tracing or source-specific sampling, would be valuable to more definitively determine the sources and mechanisms involved.

PC3, which explained 10.51% of the total variance and had an eigenvalue of 1.58, was notably associated with iron, showing a negative correlation. On the other hand, PC4, with an eigenvalue of 1.24 and accounting for 8.30% of the variance, was positively correlated with K^{+} and Mn, while negatively correlated with TDS.

4.2.5 Water quality indices

4.2.5.1 Drinking water quality index (DWQI)

The WQI ratings, determined according to aquatic environment criteria established by the (WHO 2017), were employed to evaluate the drinking water quality at seven sampling locations in the lower Danube river basin. While the Danube river is not commonly used as a direct source of untreated drinking water, it is abstracted for potable use in some regions after proper treatment. Therefore, the use of WHO drinking water guidelines in this context provides a precautionary and comparative benchmark to assess overall water quality and potential health risks in scenarios involving indirect exposure, abstraction, or emergency use. The results showed considerable differences in the water's suitability for such uses. Specifically, the WQI values at sites S2, S3, S4, S6, and S7 ranged from 76 to 100, suggesting that the water was appropriate for irrigation (Appendix Figure 28A). Conversely, sites S1 and S5 had WQI values above 100, indicating unsuitability for irrigation due to lower water quality. The unsuitability of the water for drinking purposes was primarily attributed to elevated concentrations of Fe and HCO_3^-, which exceeded the allowable limits set by the (WHO 2017). These findings highlight the importance of targeted monitoring and pollution control strategies to safeguard both ecological and human health across the basin.

4.2.5.2 Irrigation water quality indices (IWQIs)

Evaluating soil water quality and its impact on crop requirements involves analysis of several key factors. These factors may include distinct chemical indicators (Li et al. 2013; Kawo and Karuppannan 2018) to a set of indicators (Meireles et al. 2010). Understanding the results of these indicators allows decision-makers to develop appropriate strategies for managing irrigation water. This study specifically examines these indicators. Water quality for irrigation was classified according to defined parameter values, along with the six IWQIs detailed in Appendix Table 4.

WQI and Categorization

The IWQI, derived from the equation presented Appendix Table 3 (Meireles et al. 2010; Al-Mashreki et al. 2023), plays a pivotal role in assessing water quality within the study area, particularly regarding agricultural irrigation. This index carefully evaluates various characteristics of water composition that can profoundly affect both irrigation suitability and soil conditions. The assessment classifies water quality into five distinct categories (Ayers and Westcot 1985), each with unique implications for irrigation practices and soil health. Interestingly, the dataset reveals that 100% of the samples indicate no limitations for irrigation (Appendix Figure 28B) highlighting the consistent water quality across the study area. The IWQI numerical values show significant variability, ranging from 99.6 to 107.6, with an average of 104.36. This range reflects the multiple factors that affect water quality. The spatial analysis (Appendix Figure 28B) pinpoints specific locations with diminished water quality, raising concerns about their suitability for irrigation and the potential effects on soil health. Notably, all samples fall into the category of excellent water quality, indicating that they are suitable for unrestricted irrigation. This finding is crucial, as such water conditions could negatively impact soil permeability, nutrient absorption, and overall agricultural output. Therefore, based on the IWQI results, there seems to be no requirement for costly water treatment or fertilization before irrigation in the examined areas along the Danube river.

4.2.6 Impact on soil composition

The quality of water used in agriculture greatly impacts soil composition, particularly affecting aspects such as permeability, filtration rate, and aeration, all of which are influenced by the water's chemical properties (Todd and Mays 2005). One key factor is the sodium ion concentration. High sodium levels can interfere with soil filtration processes by replacing calcium and magnesium ions through adsorption (Salem et al. 2023). When evaluating water suitability for irrigation, metrics such as Na%, SAR, and SSP are essential tools. These metrics, derived from Na^+, Mg^{2+}, and Ca^{2+} levels, not only help pinpoint areas at risk of soil deterioration but also inform effective soil management strategies. The recorded values for SAR, Na%, and SSP span a broad range. The minimum values are 0.37, 13.75, and 12.52, respectively, while the maximum values reach 0.68, 18.72, and 17.59. The average values for these metrics are 0.49, 16.11, and 14.88, respectively. According to these metrics, all samples are deemed suitable for irrigation (Appendix Figure 28C-E). While no direct soil sampling was conducted, the evaluation of these indices allows for an informed interpretation of how sodium levels in irrigation water may potentially influence soil properties. Therefore, it is suggested, based on established guidelines and index values, that water treatment may not be necessary in areas where sodium levels remain within acceptable ranges.

4.2.7 Potential salinity index (PS)

The PS, calculated from Cl^- and SO_4^{2-} ions levels, is employed to evaluate the suitability of irrigation water for agricultural use (Salem et al. 2023). The PS index divides result into 3 categories: excellent to good ($PS < 3.0$), good to potentially harmful ($PS = 3.0\text{-}5.0$), and potentially harmful to unsatisfactory ($PS > 5.0$). Analysis of the PS index shows a consistent pattern: all water samples collected, representing 100% of the total, were rated as suitable to excellent for irrigation (Appendix Figure 28F). This finding goes beyond water quality evaluation; it also suggests potential improvements in soil structure, nutrient absorption, and agricultural productivity. The significance of this classification implies that measures to maintain soil sustainability may not be required. Using water samples with PS values classified as potentially harmful to unsatisfactory could negatively impact soil conditions, decrease crop yields, and undermine agricultural sustainability. However, such conditions were not observed in this study.

4.2.8 Precipitation of alkali elements and residual sodium carbonate (RSC)

Excessive presence of carbonates and bicarbonates in irrigation water can adversely affect its quality when combined with divalent cations such as Mg^{2+} and Ca^{2+}. This imbalance can lead to the precipitation of alkali metals, particularly Mg^{2+} and Ca^{2+}, thereby deteriorating the overall quality of irrigation water. The precipitation of CO_3^{2-} minerals with Ca^{2+} and Mg^{2+} ions results in elevated concentrations of Na^+ and SAR (Eaton 1950). This phenomenon holds significant implications, especially in regions characterized by arid to semi-arid climates. In such environments, high levels of RSC can disrupt soil physical properties, often leading to the decomposition of organic matter, leaving noticeable black marks on the soil surface after drying (Gad et al. 2021; Zhang et al. 2023). In this work, RSC calculations were employed to forecast the precipitation of Mg^{2+} and Ca^{2+} on soil particle surfaces, a method widely acknowledged and utilized in arid and semi-arid regions where higher levels are associated with soil salinization (Prasad et al. 2001). This work, conducted in a humid climate, revealed a noteworthy classification based on RSC values (Appendix Figure 28G). According to this classification scheme, irrigation water with an RSC exceeding 2.5 is deemed unsuitable for irrigation, whereas water with an RSC below 1.25 is considered acceptable. Water falling within the RSC range of 1.25 to 2.5 is considered questionable for irrigation use (Gad et al. 2021; Al-Mashreki et al. 2023). A significant finding of our study is that all samples had an RSC value of below 1.25 strongly indicating that the SW is safe and effective for irrigation

4.2.9 Human health risk assessment

4.2.9.1 non-carcinogenic health risk

Although untreated SW is not typically used for direct consumption in Hungary, health risk assessments were conducted to evaluate potential non-carcinogenic effects under hypothetical exposure scenarios. These include accidental ingestion during recreational activities, agricultural contact, or emergency use in rural or remote areas. This approach follows standard USEPA methodology, which models oral exposure to assess public health risks, especially for vulnerable populations such as children. The daily consumption of Fe, Mn, and NO_3^- is calculated using key parameters established by the USEPA (2004) (USEPA 2004; Xu et al. 2020). Appendix Figure 29 illustrates the assessments of non-carcinogenic risks associated with these factors in SW for children and adults across different sampling locations, while Appendix Table 10 details the HQ values and the resulting non-carcinogenic risks. For children, HQ values for Fe, Mn, and NO_3^- range between 0.002 and 0.477, 0.096 and 1.678, and 0.190 and 0.927, respectively, with average values of 0.092, 0.281, and 0.523. In adults, the HQ values range from 0.001 to 0.125 for Fe, 0.025 to 0.439 for Mn, and 0.050 to 0.243 for NO_3^-, with corresponding averages of 0.024, 0.074, and 0.173 (Appendix Table 10). Contaminants in SW present a significantly higher average HI for children compared to adults, highlighting a considerable non-carcinogenic risk in the study region (Appendix Figure 29). Notably, the concentrations of NO_3^- and Mn are the primary contributors to non-carcinogenic risk, followed by Fe. Additionally, the HI readings for children via the oral exposure route are higher than those for adults, and the total HI for non-carcinogenic risk is 2.1 times greater for children than for adults (Appendix Table 10).

A MCS was conducted to estimate the oral HQ readings for Mn, Fe, and NO_3^- for both adults and children, using 10,000 iterations in Python. The results of the simulation showed that the predicted HQ readings for Mn, Fe, and NO_3^- remained below the acceptable threshold ($HQ < 1$). Therefore, based on the MCS, the anticipated exposure levels are not expected to present a significant health risk for either group through oral intake. However, it is important to note that risk assessments are often based on conservative assumptions and data uncertainties. As a result, continuous monitoring of exposure levels and regular updates to risk assessments are essential as new data becomes available (Appendix Figure 29).

4.3 Heavy metal and physical chemical properties of the Al-Jawf basin

4.3.1 Characterization of groundwater chemistry

The characteristics of GW within an agricultural zone in the Al-Jawf Governorate (Yemen), were evaluated by collecting and analyzing thirty-three samples. Appendix Table 11 provides a statistical overview of the tested sample characteristics (pH, Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, CO_3^{2-}, SO_4^{2-}, Cl^-, NO_3^-, Fe, Mn, TDS, and EC) and the FAO irrigation water standards (Ayers and Westcot 1985). Additionally, the regional distribution maps are presented in Appendix Figure 30.

The GW pH ranged from 7.00 to 8.20, indicating slightly alkaline conditions suitable for irrigation within acceptable limits. TDS levels varied significantly, with all samples classified as moderately saline, and 46% exceeding recommended limits. EC ranged from 1030 µS/cm to 5930 µS/cm, with 46% of samples exceeding the irrigation threshold. Ca^{2+} concentrations ranged from 40 to 359.64 mg/L, while Mg^{2+} levels varied from 41.31 to 359.64 mg/L, both exceeding recommended limits, likely due to mineral dissolution processes. Mg^{2+} concentrations in subsurface water predominantly result from ion exchange involving iron-magnesium bearing minerals, the chemical breakdown of soil matrices, and the decomposition of saline and carbonaceous formations. The dissolution of dolomite ($MgCO_3$) particularly contributes to the release of Mg^{2+} into the aqueous solution (Salem et al. 2023)).

Na^+ in groundwater ranged from 51.73 to 689.70 mg/L, staying below the FAO's irrigation limit. However, elevated Na^+ levels in the eastern regions may pose long-term risks of sodification and soil salinization. Groundwater K^+ concentrations ranged from 2 to 23.46 mg/L, making it unsuitable for irrigation, likely due to leaching from potash feldspars and agrochemicals. HCO_3^- content in all samples was within the acceptable range for irrigation, between 170.86 and 463.75 mg/L. Cl^- concentrations ranged from 127.62 to 1013.87 mg/L, with all samples remaining below recommended irrigation limits. High chloride levels in the east suggest halite (NaCl) dissolution. While most groundwater samples had sulphate levels below FAO standards, 13% exceeded these limits. The dissolution of gypsum and anhydrite minerals (Eqs. 33) could be the source of the sulphate content:

Fe levels in the GW ranged from 0.01 to 3 mg/L, making it suitable for irrigation. Mn levels varied from 0.001 to 4 mg/L, with 13% of samples exceeding drinking water standards. Elevated Mg^{2+} concentrations were particularly noted in the western region, potentially due to industrial waste and manufacturing outputs, as reported by Ahmad et al. (2021) (Ahmad et al. 2021).

4.3.2 Geochemical processes influencing groundwater facies

By ranking ions according to their dominance, GW may be divided into several water facies. Piper's diagram was applied in several research studies to detect chemical water facies (Piper 1944). The Piper diagram allowed the GW of the Al-Jawf plain to be divided into two distinct chemical water facies or types, (1) Na-Cl and (2) Ca-Mg-SO_4 (Appendix Figure 31). An examination of local elemental molar concentrations revealed the following order for cations: $Na^+ > Ca^{2+} > Mg^{2+} > K^+$. For anions, the order was $SO_4^{2-} > Cl^- > HCO_3^- > NO_3^-$. Unlike cations, the dominant water facies in most samples are characterized by higher proportions of anions, particularly SO_4^{2-} and Cl^-. Some GW samples show a mixed water type, likely due to the mixing of water from various sources via vertical interactions between superimposed aquifer levels. Lenticular clay or semi-permeable layers separating upper and lower aquifer levels enable vertical communication, altering the physicochemical characteristics of GW. The diagram developed by Gibbs is a valuable analytical framework for examining the relationship between the mineral constituents of aquifers and the chemical properties of water. This framework indicates that all analyzed samples fall within the region dominated by evaporative processes. This placement suggests that the primary influences on the composition of the subsurface water are the influx of water enriched with dissolved solutes from agricultural drainage and the dissolution processes of minerals after evaporation(Gad et al. 2020b).

4.3.2.1 Ion exchange processes

Ion exchange can be a significant factor in regulating GW mineralization, especially when clay minerals are abundant in the system. Clay minerals often fix Na^+ and K^+ while releasing Ca^{2+} and Mg^{2+}, or vice versa. (Salem et al. 2023) have explained how clay minerals stabilize their electrostatic charge. The schematic developed by Chadha (1999) (Chadha D. 1999) provides a conceptual framework for interpreting hydrochemical data, allowing for a comparative analysis of different water types (Appendix Figure 32A). In this study, the collected specimens are depicted in the quartile indicative of Ca-Mg-SO4 dominance in sub-quadrant D and Cl-Na prevalence in sub-quadrant C. This area is characterized by a predominance of strong acid anions over weak ones, suggesting an inverse connection between bivalent and monovalent ions. This highlights the significant role of ionic exchange in the mineralization of GW in the Al-Jawf plain (Gaagai et al. 2023). To further understand the source and connections between key elements, additional methods are necessary. Examining the interplay between Na^+ and Cl^- ions is crucial for understanding the water quality dynamics in the study area (Appendix Figure 32B) (Gad et al. 2023). The positive correlation ($R = 0.81$) between GW samples and the halite dissolution line suggests silicate mineral weathering or an ion exchange mechanism (Appendix Figure 32B). Most samples (68%) on the Na+ vs. Cl− plot were above the equiline (1:1), indicating that chloride-bearing salt dissolution played a minor role in the elevated Cl^- levels in GW (Singh et al. 2018). However, 32% of GW samples were below the equiline (1:1), signifying that ion exchange and silicate weathering mechanisms contributed to Na^+ enrichment in the GW of the study area (Jacobson and Wasserburg 2005). Appendix Figure 32C illustrates the Mg^{2+}/SO_4^{2-} ratio, revealing a strong correlation between Mg^{2+} and SO_4^{2-} with a correlation coefficient of 0.76. On the scatter plot of SO_4^{2-} and Mg^{2+} (Appendix Figure 32C), most samples are positioned above the one-to-one line, indicating the dissolution of both gypsum (calcium sulfate) and anhydrite in the subsurface water, as reported by Ayers and Westcott (1994) (Ayers and Westcot 1985). Furthermore, samples above the 1:1 line are enriched in Ca^{2+} and Mg^{2+}, suggesting that Mg^{2+} dissolution from evaporites is more significant than from carbonates. The levels of Mg^{2+} and SO_4^{2-} are primarily associated with the dissolution of evaporites, including anhydrite, gypsum, and magnesium sulfate (Eqs. 32-36).

$$CaMg\,(CO_3)_2 \longrightarrow Ca^{2+} + Mg^{2+} + 2CO_3^{2-} \qquad \text{Eq. 36}$$

However, the presence of 8 samples showing substantial deviation from the (1:1) line indicates the influence of additional sources of SO_4^{2-}, such as mineral weathering, ion exchange processes, and agricultural practices (Athamena and Menani 2018). The relationship between ($Na^+ + K^+$) versus ($Ca^{2+} + Mg^{2+}$) in the water samples reflects ion exchange processes (Appendix Figure 32D). Within the scope of this research, the ion exchange interactions between aquifer minerals and GW have been identified as chloro-alkaline. Notably, around 60% of the samples analyzed exhibited negative CAI values, encompassing both CAI-I and CAI-II (Appendix Figure 32E), indicating diverse cation exchanges at specific locations within the study area, as reported by Salem et al. in 2023 (Salem et al. 2023). Concurrently, approximately 40% of the samples exhibited a significant trend of cation exchange involving shifts between K^+ and Na^+ in GW, as well as Mg^{2+} and Ca^{2+} in nearby Quaternary rocks. Interestingly, these cation exchanges were associated with CAI values showing contrasting trends (Al-Mashreki et al. 2023).

4.3.3 Statistical analysis

4.3.3.1 Cluster analysis

CA was utilized to categorize the hydrochemical parameters, employing Ward's linkage method and squared Euclidean distances. The findings are presented in a dendrogram (Appendix Figure 33), where clusters are structured according to similarities among the 12 physicochemical parameters (Snousy et al. 2023). In terms of horizontal orientation, two main clusters were predominantly influenced by TDS. Cluster1 includes Ca^{2+}, Mg^{2+}, CO_3^{2-}, K^+, Mn, Fe, and NO_3^-.

The cluster can be divided into two distinct sub-clusters: one containing Ca^{2+} and Mg^{2+}, indicating a shared origin responsible for GW hardness, and another comprising CO_3^{2-}, K^+, Mn^+, Fe^+, and NO_3^-, suggesting anthropogenic sources such as potassium-rich fertilizers and SW infiltration contributing to these parameters in GW. Cluster 2 includes Cl^-, SO_4^{2-}, HCO_3^-, and Na^+, associated with mineral dissolution and rock weathering via water-rock interactions, underscoring their influence on GW salinity. Cluster 3 (G3) represented by TDS originates from two sources: evaporitic and carbonate processes.

4.3.3.2 Principal component analysis (PCA)

The PCA was utilized to analyze data from a total of 33 samples across 12 variables. After selecting three principal components, it was found that these components collectively explained more than 89.6% of the total variance in the dataset. The results of this analysis are detailed with a tabular representation of parameter loadings corresponding to the three principal components, available in Appendix Table 12. Notably, loadings close to 1 indicate strong correlations between components and observed variables. PC1 accounted for 55.20% of the variance and showed significant loadings for Mg^{2+} (0.81), Ca^{2+} (0.85), Na^+ (0.82), Cl^- (0.86), SO_4^{2-} (0.88), HCO_3^- (0.73), Fe (0.68), Mn (0.79), TDS (0.99), and K^+ (0.66). This component suggests that ion exchange processes, limestone weathering, and mineral dissolution (e.g., pyrite, halite, gypsum) contribute to water salinity. Sources such as chloride formation and soil weathering also play roles, while SO_4^{2-} may originate from fertilizer use or sulfur oxidation. Elevated Ca^{2+}, Na^+, and Mg^{2+} levels may reflect poor irrigation water quality, household waste, and excessive fertilization (Athamena and Menani 2018; Gaagai et al. 2023), indicating a salinization factor. PC2 contributed 14.07% of the variance, with notable loadings for NO_3^- (0.84) and CO_3^{2-} (0.73). This suggests nitrate presence is influenced by septic systems, surface water inputs, agricultural runoff, soil mineralization, and intensive fertilizer use (Athamena et al. 2023). The widespread use of urea and composite fertilizers has notably increased NO_3^- concentrations over time (Eq. 38).

$$NH_4^+ + 2O_2 \longrightarrow NO_3 + H_2O + 2H^+ \qquad \text{Eq. 38}$$

PC3, contributing 7.99% of the cumulative variance, shows a strong positive loading for K^+ (0.66), indicating a distinct anthropogenic influence related to potassium-rich fertilizers, consistent with observations by (Snousy et al. 2023). This component reflects localized human impacts on groundwater chemistry. This component reflects localized human impacts on groundwater chemistry. Thus, the three principal components can be interpreted as distinct environmental influence factors: PC1 represents salinization and mineral weathering from both geogenic and anthropogenic sources; PC2 corresponds to agricultural nitrate contamination; and PC3 reflects localized anthropogenic contributions from potassium-rich fertilizers.

4.3.4 Water quality indices for irrigation and agricultural purposes

This study utilized multiple water quality indices to assess the suitability of water for agricultural purposes. These indices, including the IWQI, SSP, SAR, PS, KI, and RSC, were crucial in evaluating the potential risks of soil salinization and the adverse impacts on soil and plant health associated with irrigation practices. A comprehensive statistical analysis was performed on the data from these indices to determine the appropriateness of water for agricultural use. Detailed results can be found in Appendix Figure 34 and the distribution maps of the irrigation indices presented in Appendix Figure 35 (A-F).

4.3.4.1 Irrigation water quality index (IWQI)

GW quality assessment for irrigation utilized the IWQI, considering individual chemical indices or their combinations to provide comprehensive insights for decision-makers. The GW was classified into five hazard groups to evaluate its suitability for irrigation, resulting in IWQI values ranging from 16.09 to 92.12, with an average of 48.03 (Appendix Figure 35). Significantly, approximately 63.5% of samples were categorized in the high to severe restriction range, indicating suitability for salt-tolerant crops, while 23.4% showed low to moderate restrictions, and about 3.03% had no irrigation limitations (Appendix Figure 35). An overall index map (Appendix Figure 35A) depicted declining water quality in the eastern and western study areas, potentially influenced by anthropogenic activities and geogenic factors.

4.3.4.2 Sodium adsorption ratio (SAR) and salinity hazards

In the realm of irrigation water management, the SAR parameter plays a crucial role as it indicates the soil's ability to release Mg^{2+} and Ca^{2+} while absorbing Na^+ at specific ion exchange sites. This complex process significantly impacts soil structure, potentially causing soil particle dispersion and reducing water infiltration capacity (McGeorge 1954; Hanson and Grattan 1991; Eid et al. 2023). The use of highly saline irrigation water introduces a nuanced balance of benefits and drawbacks. On one hand, it can improve soil structure by increasing water infiltration rates; however, on the other hand, it intensifies water stress on plants and crops. When exposed to saline irrigation water, plants must exert more effort to extract water from the soil, leading to conditions conducive to water stress. Achieving a delicate equilibrium between soil structure and plant water stress is crucial in agricultural practices.

The USSL scatter diagram in Appendix Figure 36 presents the classification of irrigation water based on the relationship between SAR and EC, following the system developed by the U.S. Salinity Laboratory (1954) (McGeorge 1954). The majority of samples (42.4%) fall into the C4-S2 class, which shows very high salinity and medium SAR. About 15.15% of the samples are classified as C4-S3, indicating very high salinity and high SAR.

Additionally, 33.3% of the samples are within the C3-S1 and C3-S2 categories, representing high salinity combined with low to medium SAR, respectively. These results clearly demonstrate the dominance of high salinity levels in the sampled waters, with SAR levels varying across low to high classes depending on the category. Interestingly, only three samples were categorized as C4-

S1, indicating exceptionally high salinity levels combined with a low SAR. Importantly, the highest observed SAR value among all samples remained below the critical threshold of 10, which is considered suitable for irrigation purposes (Appendix Figure 35B). This empirical investigation underscores the potential adverse effects of elevated salinity levels in irrigation water on plant growth. Furthermore, when SAR values are in the low to medium range (Appendix Figure 35B), and no specific calcium-related issues are documented, their impact on soil infiltration capacity is generally considered minimal according to established literature. Based on these findings, the most effective strategy for managing GW in this degraded region for irrigation involves selecting plant varieties and crops that are tolerant to high water salinity. Choosing suitable plant species presents a practical approach to mitigate the negative impacts of salinity on crop yields and overall agricultural productivity.

4.3.4.3 Soluble sodium percentage index (SSP)

In this study, we have independently conducted an SSP analysis to assess salinity risks related to sodium content, specifically by comparing Na^+ concentrations with those of Mg^{2+} and Ca^{2+}. When sodium dominates over these divalent cations, it can negatively affect plant health, causing symptoms such as leaf burn and tissue necrosis. The SSP values calculated by the team ranged from 21.44 to 66.04, with an average of 48.03. Based on these results, the team determined that 90.9% of the GW samples are suitable for irrigation, while 9.09% are unsuitable (Appendix Figure 35C). This original interpretation underscores a significant risk in the western part of the study area, where elevated SSP levels suggest a need for water treatment interventions to protect crop productivity and soil health.

4.3.4.4 Potential salinity (PS)

The PS index is widely used to assess GW suitability for irrigation, based on chloride ion concentrations and half of the SO_4^{2-} concentration. In this study, PS values ranged from 4.75 to 37.95, averaging 19.38. Remarkably, 93.9% of water samples were categorized as Injurious to Unsatisfactory, while 6.06% fell into the Good to Injurious range. Addressing the challenges posed by high water salinity is therefore critical, necessitating the selection of salt-tolerant plant species and implementation of additional treatment measures before irrigation. This is particularly crucial in the eastern and western regions of the study area (Appendix Figure 35D).

4.3.4.5 Kelley ratio index (KR)

In this study, GW suitability for irrigation was evaluated using the KR to identify elevated sodium levels. KR values observed ranged from 0.27 to 1.94, with a mean KR value of 0.93. Notably, approximately 60.6% of GW samples were classified as suitable, falling into the "good class" for irrigation purposes. Conversely, 39.39% of samples were deemed unsuitable for irrigation (Appendix Figure 35E) and detailed in Appendix Figure 34. Interpreting the KR value is crucial for assessing GW suitability: a KR value exceeding one ($KR > 1$) indicates excessive Na^+ content, rendering the water unsuitable for irrigation. Conversely, a KR value below one ($KR < 1$) signifies suitability for irrigation purposes. Interestingly, unsuitable GW samples were predominantly located in the western regions of the study area. This spatial distribution highlights the critical need for ongoing monitoring and effective management of sodium levels in water resources to ensure their suitability for irrigation in specific geographical contexts.

4.3.4.6 Residual sodium bicarbonate (RSBC)

The RSBC quantifies the surplus concentration of HCO_3^- over Ca^{2+} and is crucial for evaluating irrigation water quality. Its adverse effects include altering soil physical characteristics, leading to organic matter decomposition and the formation of darkened surface discoloration upon soil drying. Generally, the RSBC values are classified as suitable (<5) or unsuitable (>5) for irrigation purposes. In this study, the RSBC values ranged from -9.60 to 2.92, with a mean of -2.01. The findings revealed that all tested GW samples were suitable for irrigation, as indicated by their negative RSBC values, signifying an excess of HCO_3^- in the GW. These indices play a critical role in assessing GW suitability for irrigation and underscore the importance of effectively managing salinity levels to ensure the safety and effectiveness of irrigation practices (Appendix Figure 35F).

4.3.5 Non-carcinogenic health risk

The assessment of health risks associated with HM exposure included evaluating CDI, HQ, and HI as key indicators. CDI measured daily human exposure to metal contaminants, while HQ and HI assessed potential health risks, where values above one indicated potential harm. Specifically, dermal CDI values for Fe and Mn were found to be below acceptable limits ($CDI < 1$), suggesting no significant health threats. However, Mn dermal CDI values were higher than those for Fe, with adults and children averaging 6.4×10^{-5} mg L−1 day^{-1} and 0.0002 mg L−1 day^{-1}, respectively (Appendix Figure 37), emphasizing varying impacts of HM exposure, particularly through dermal contact. In terms of HQ dermal values, adults generally had values below one for both Fe and Mn, but Mn exhibited notably higher values, with maximum HQ dermal values of 0.6 for adults and 1.76 for children, indicating potentially increased non-carcinogenic risks from Mn dermal exposure, especially in specific study locations. The assessment of HI Dermal provided further insights, showing adults had a mean HI Dermal of 0.07 (range: 0.0002 - 0.6), while children had a slightly higher mean of 0.2 (range: 0.0005 - 1.77), highlighting elevated risks for children, particularly in certain samples where HI Dermal exceeded 1, indicating significant health risks (Appendix Figure 37). Overall, 90.9% of GW samples showed HI Dermal values for children below 1, suggesting a generally low health risk from HM exposure through skin contact. However, three samples exhibited HI Dermal values above 1, indicating heightened risks, particularly for children. Conversely, all HI Dermal values for adults remained within safe limits, indicating low risk based on these findings (Appendix Figure 37). These findings underscore the critical necessity for targeted interventions to mitigate health risks, particularly in areas of elevated HM concentrations, focusing on protecting vulnerable populations like children. Comprehensive water treatment and ongoing monitoring are recommended to reduce these risks effectively.

4.3.6 RF simulation model

In this study, a RF simulation model was developed to predict irrigation water quality indices (IWQI, SAR, KR, SSP, PS, RSBC) and non-carcinogenic health hazard indices (HI for adults and children) using 60 groundwater (GW) samples, each comprising 12 hydro-geochemical input variables, including major cations and anions. The dataset was partitioned into training (80%) and testing (20%) subsets to evaluate the model's performance and generalizability, with accuracy and feature importance assessed using 5-fold cross-validation. A key limitation of traditional approaches, such as IWQI and HI, lies in their reliance on expert-assigned weights for variable scoring, which introduces subjectivity and uncertainty into the final assessments. These indices represent aggregated values of physicochemical components into a single numerical form indicating water quality suitability for specific uses. To address this challenge, several studies have proposed integrating objective, data-driven modeling approaches. In this context, the RF model was implemented using High-Grade Vegetation (HGV) data to eliminate subjectivity, enhance predictive performance, and identify influential features through entropy-informed variable weighting. The outcomes revealed that the RF-3HG model exhibited outstanding capability in forecasting integrated IWQI, primarily based on three critical HGV-inferred parameters, achieving R^2 values of 0.988 (RMSE = 2.239) for training and 0.940 (RMSE = 3.976) for validation. For SAR predictions, the

RF-2HG model demonstrated the highest accuracy, with R^2 values of 0.971 (RMSE = 0.404) for training and 0.869 (RMSE = 0.667) for validation. Similarly, KR predictions were most accurately captured by the RF-2HG model, which yielded R^2 values of 0.952 (RMSE = 0.099) for training and 0.665 (RMSE = 0.203) for validation. The RF-4HG model was most effective for SSP estimation, producing R^2 values of 0.959 (RMSE = 2.453) for training and 0.671 (RMSE = 5.394) for validation. For PS, the RF-2HG model demonstrated exceptional reliability, with R^2 values of 0.994 (RMSE = 0.744) and 0.982 (RMSE = 0.960) for training and validation, respectively. RSBC predictions were best performed by the RF-12HG model, with R^2 values of 0.924 (RMSE = 0.768) for training and 0.751 (RMSE = 1.042) for validation. Regarding the non-carcinogenic health risk assessment, the RF-1HG model achieved the most robust results for both adults and children. For adult HI prediction, R^2 values were 0.988 (RMSE = 0.018) and 0.901 (RMSE = 0.041) for training and validation, respectively, while for children, the model yielded identical R^2 values of 0.988 (RMSE = 0.041) for training and 0.901 (RMSE = 0.048) for validation. These results are presented in Appendix Table 13, while the measured vs. predicted comparisons are illustrated in Appendix Figures 38, 39, and 40. These findings highlight the effectiveness of RF modeling in capturing complex relationships among hydrochemical variables, driven by careful hyperparameter tuning and optimized feature selection based on HGV data. The process ensures enhanced model fairness, accuracy, and interpretability, crucial for sustainable groundwater quality monitoring and policy development. Moreover, this approach aligns with previous research indicating that machine learning methods with optimized feature selection–such as the RF models used here–offer superior performance in hydro-environmental assessments (Bai et al. 2019; Saleh et al. 2022; Abou El-Safa et al. 2022; Galal et al. 2022; Qiu et al. 2023; El-Hendawy et al. 2024; Eid et al. 2025b). The model's reliability is further supported by its application in other environmental domains, including accurate pollution and ecological risk prediction in Gamasa coastal sediment samples (Abou El-Safa et al. 2022). Additionally, the efficacy of similar ML methods, such as Recurrent Neural Networks, in groundwater quality assessments has been documented (Gad et al. 2024), further reinforcing the importance of informed feature selection and model design in understanding contamination risks. While the focus of this study was on RF due to its interpretability, computational efficiency, and widespread application in hydrology (Zavareh et al. 2023), future research will incorporate comparative analysis with advanced boosting algorithms such as XGBoost to further validate and refine performance.

4.4 Physical chemical properties of the Al-Jawf basin

4.4.1 Hydrochemical characteristics of groundwater

The physicochemical properties of the Al-Jawf GW were analyzed to assess the quality of quaternary aquifer waters. GW samples had a mean temperature of 20.7 °C. The pH of the water demonstrated variability influenced by the geological terrain, resulting in both acidic and alkaline conditions. TDS amounts ranged from 378.42 to 5012 mg/L, categorizing the GW samples as fresh to moderately saline, with 18.5% exceeding FAO (Ayers and Westcot 1985) water criteria. The pH ranged from 6.5 to 7.5, averaging 7.13, meeting FAO (Ayers and Westcot 1985) standards (Appendix Table 14). The pH values indicated low alkalinity, primarily influenced by dolomite and calcite fixation mechanisms within the pH range of 6.5 to 8.5. Notably, well S19 recorded the highest pH value, while S7 exhibited the lowest.

Therefore, EC serves to quantify water mineralization levels. In our research, 18.5% of the GW samples exhibited EC values, which surpassed the FAO recommended limit of 3000 µS/cm (Ayers and Westcot 1985). All GW in the Al-Jawf region displayed elevated EC values when underground flows directed towards discharge areas southwest of Al-Jawf. Ca^{2+} adhered to FAO's 400 mg/L maximum threshold for irrigation water. Mg^{2+} levels varied generally below FAO guidelines. Water hardness primarily stems from alkaline earth elements, influenced by soil type. The highest concentrations of Ca^{2+} and Mg^{2+} were observed in the southeast and southwest of Al-Jawf, with Mg^{2+} peaking at 114.5 mg/L due to dissolution of dolomites from the Cretaceous and Tawilah groups. Overall, the breakdown of dolomitic rocks ($CaMg(CO_3)^2$), gypsum ($CaSO_4$), and calcium carbonate ($CaCO_3$) in the aquifer generates Ca^{2+} solved in the GW (Eqs 33-41).

$$CaCO_3 + H_2CO_3 \rightarrow Ca^{2+} + 2HCO_3^- \qquad \text{Eq. 39}$$

The alkaline elements K^+ and Na^+ are naturally present in both irrigation water and the Earth's crust. Na^+ concentrations in Al-Jawf GW varied from 24.15 to 724.5 mg/L, all falling below the FAO limit of 919 mg/L (Ayers and Westcot 1985). K^+ concentrations ranged from 2.34 to 23.40 mg/L, with the FAO standard setting a high allowable amount of 2 mg/L for irrigation water. The enrichment of these elements is attributed to interactions between rock and water, saline seeps, and minor atmospheric inputs (Panno et al. 2002). Higher potassium concentrations were observed in wells located in the central and southern parts of the study area. Cl^- concentrations ranged from 35.5 to 1136 mg/L, with the latter being the highest value. Previous studies (Lima et al. 2021) have suggested that high chloride levels may result from geological formations through which water flows or inadequate treatment of sewage waste. Some samples exceeded the FAO recommended limit of 1036 mg/L (Ayers and Westcot 1985), particularly in two boreholes in the southwest of the study area. Elevated Cl^- levels can cause metal corrosion, impart a salty taste to water, and reduce the durability of concrete. The FAO guideline (Ayers and Westcot 1985) specifies that HCO_3^- concentrations should ideally be around 610 mg/L. However, GW samples from this research showed levels ranging from 150 to 915 mg/L, mostly within the specified range except for four samples that exceeded the limit. Samples with the highest HCO_3^- content were found in wells S7, S17, S22, and S23, with values of 915, 730, 780, and 823.5 mg/L, respectively. SO_4^{2-} levels varied from 19.2 to 1939 mg/L and generally complied with FAO standards, except for two samples in the southwest of Al-Jawf that exceeded the limits (Ayers and Westcot 1985). High concentrations of SO_4^{2-} in irrigation water can seriously harm public health (Ibrahim Mohamed et al. 2019). The intensive use of chemical fertilizers by farmers has contributed to increased SO_4^{2-} concentrations (Mahapatra et al. 2022), posing potential health risks. Higher concentrations were notably found in the southwestern part of the study area, aligning with the direction of subsurface flow. Gypsum and anhydrite dissolution were identified as significant contributors to elevated SO_4^{2-} concentrations, with potential input from sulfide mineral weathering such as pyrite, warranting further investigation (Baragaño et al. 2020). NO_3^- concentrations in Al-Jawf water samples ranged from 0.1 to 6 mg/L, all below the FAO recommended limit of 10 mg/L (Ayers and Westcot 1985). The low NO_3^- levels suggest minimal contamination from anthropogenic sources in the study area, as depicted in Appendix Table 12. Distribution maps of these physicochemical parameters across the study area are presented in Appendix Figure 41.

4.4.2 Groundwater facies and processes influencing groundwater chemistry

Aquifer hydrochemical processes influence GW chemistry (Jain et al. 2018), with the expansion of GW facies primarily determined by local lithology and geology, especially in arid to semi-arid regions (Maghrebi et al. 2021). Major ions play a crucial role in governing these hydrochemical procedures, which is why Chadha's and Piper's diagrams (Piper 1944; Chadha D. 1999) were utilized in this study (Appendix Figures 42 A-B). These diagrams assess GW composition through cation-anion pairings and plot positioning (Thilagavathi et al. 2012). In Chadha's diagram (Appendix Figure 42B), most samples analyzed fell within Field 3, with only two falling into Field 2. In the Piper diagram (Appendix Figure 42A), 25 samples were categorized in zone 1 and two in zone 2 of the diamond shape. Both diagrams indicate similar water types: Ca-Mg-Cl/SO_4 and Ca-Mg-HCO_3 types, respectively, characterized by higher concentrations of alkaline metals over alkaline earth metals and a prevalence of weakly acidic anions compared to strongly acidic ones. Various diagrams, including the TIS salinity diagram (Apollaro et al. 2019), ionic ratio relationships, and Gibbs and Chadha's diagrams, have been employed by different researchers to determine GW chemical compositions or water types. These diagrams, alongside the Piper diagram, illustrate the geochemical evolution of GW from recharge to discharge areas. The study's findings indicate that GW in the recharge area is characterized as Ca-Mg-HCO_3 type, whereas water-rock interactions during flow transform it into Ca-Mg-Cl/SO4 type in the discharge zone. Additionally, SIs provide insights into aquifer mineral compositions, while Gibbs diagrams depict relationships between ionic ratios and salinity levels. Local element concentrations analysis showed a cation sequence of $Ca^{2+} > Mg^{2+} > Na^{+} > K^{+}$, while anion evolution followed $SO_4^{2-} > Cl^{-} > HCO_3^{-} > NO_3^{-}$. The presence of Mg^{2+} and Ca^{2+} sulfate facies in the chemical profile is attributed to the breakdown of evaporitic deposits. The region's lithology and anthropogenic factors such as intensive fertilization and irrigation water quality significantly influence the distribution of major ions (Ca^{2+}, Mg^{2+}, SO_4^{2-}). Ca^{2+}, Mg^{2+} and SO_4^{2-} contribute both Mg^{2+} and Ca^{2+} upon dissolution, although shifts between dominant ions may occur due to dilution following mixing or precipitation. The primary sources of SO_4^{2-} identified include organic matter decomposition in soil and leachable organic matter from highly cultivated areas of the Al-Jawf plain. Gypsum dissolution serves as a secondary source of sulfate (Pant et al. 2020), facilitated by the presence of evaporite sequences.

The study utilized Gibbs plots Gibbs RJ (1970) to investigate the fundamental mechanisms governing water evolution and the various hydrogeochemical processes influencing groundwater chemistry in the region. These plots (Appendix Figure 42 C-D) illustrate that most of the GW samples analyzed are situated in the upper portion of the diagrams, indicating dominance of evaporation processes, typical in arid climates with shallow GW depths. Additionally, five samples were categorized within the rock weathering zone. The prevalence of evaporation processes in the Al-Jawf plain is evident from Appendix Figure 42 C-D, where GW is influenced by both natural factors and human activities that alter hydrochemical evolution, not fully captured by Gibbs diagrams (Li et al. 2016). Shallow water depths in the area have improved due to elevated irrigation and effective rainfall, impacting ion concentrations primarily through evaporation. The approaching saturation zone, particularly moist and rich in evaporites, leads to their precipitation via evaporation, resulting in increased salinity (maximum TDS = 5012 mg/L), with sulfate-rich GW more affected by evaporation than bicarbonate-rich GW (Belhadj et al. 2017). These findings are consistent with previous studies highlighting the high evaporation influence on sulfate-type GW in alluvial plains (Chebotarev 1955). The lack of equilibrium between Na^+ and Cl^- ions suggests diverse sources for these ions, including halite dissolution. Only a minority of samples aligned on the 1:1 line graph (Appendix Figure 43A), indicating excess chloride possibly due to additional ion sources or Na^+ removal from GW through ion exchange. Elevated Cl^- levels could stem from anthropogenic activities such as waste disposal, seepage from irrigated agricultural land, or atmospheric deposition. Anthropogenic impacts, rather than climate variability, are identified as the primary cause of freshwater scarcity in the region (Maghrebi et al. 2023). Ion exchange processes significantly affect aquifer chemistry in Al-Jawf, illustrated in Appendix Figure 43B through the scatter plot of Ca^{2+} + Mg^{2+} versus HCO_3^- + SO_4^{2-}. Samples near the 1:1 line indicate prevalent reactions involving dolomite, calcite, and gypsum dissolution. Below the 1:1 line, ion exchange is evident with Ca^{2+} retained in soil and Na^+ returning to GW, whereas above the line indicates reverse ion exchange where Na+ remains in soil and Ca^{2+} enters GW. Samples above the 1:1 line are enriched in calcium and magnesium, suggesting evaporites as a more significant source of Ca^{2+} and Mg^{2+} compared to carbonates. The interaction of Ca^{2+} + Mg^{2+} with HCO_3^- + SO_4^{2-} (Appendix Figure 43B) does not support carbonate weathering as the primary process in the study area. Appendix Figure 43C, displaying Ca^{2+} + Mg^{2+} versus HCO_3^-, reveals excess Ca^{2+} and Mg^{2+} attributed to silicate weathering, a prominent process releasing HCO_3^- into GW (Kumar et al. 2022c). Appendix Figure 43D further corroborates this with ratios of SO_4^{2-}/Ca^{2+}, showing a strong association (correlation coefficient of 0.6) between SO_4^{2-} and Ca^{2+}, indicative of gypsum and anhydrite dissolution. Most samples above the 1:1 equiline in the Ca^{2+} versus SO_4^{2-} scatter plot (Appendix Figure 43D) suggest additional sources of calcium from calcite and dolomite dissolution in GW (Athamena et al. 2023), while deviations from the 1:1 line point to alternative sources of SO_4^{2-}, potentially from mineral weathering, ion exchange reactions, or other geological processes.

The CAI was employed to assess whether ion exchange or reverse ion exchange predominantly controlled interactions between minerals in the aquifer and groundwater (Wu et al. 2015; Qian et al. 2016). Overall, CAI values, including CAI-I and CAI-II, were positive across all water samples (Appendix Figure 43 E and F), indicating a pronounced tendency for reverse ion exchange involving Na^+ and K^+ in the GW of the study area, as well as Mg^{2+} and Ca^{2+} in the surrounding rock formations.

Analysis of various ionic ratios further confirmed that reverse ion exchange was the primary mechanism governing GW chemistry.

4.4.3 Statistical analysis

4.4.3.1 Cluster analysis

Integrating the Ward linkage approach with Euclidean distance was employed to assess the similarity among GW samples. Appendix Figure 44 presents a dendrogram categorizing the different physicochemical factors observed in the collected GW samples. For statistical analysis, standard scores (Z-scores) were computed for each variable (Güler et al. 2002). All variables underwent logarithmic transformation to achieve approximate normality in the data distribution. The dendrogram identified three main clusters based on nine physicochemical parameters (HCO_3^-, Cl^-, SO_4^{2-}, Mg^{2+}, K^+, Ca^{2+}, TDS, and Na^+). A specific phenon line was selected at a connection distance of 5, delineating distinct hydrochemical characteristics within each cluster (Gad et al. 2023). The clusters were further divided based on their TDS management. Here are the delineated groups: G1: This group comprises Ca^{2+}, Mg^{2+}, Na^+, K^+, and NO_3^-, predominantly influenced by carbonate components. The substantial presence of Mg^{2+} and Ca^{2+} indicates the influence of minerals such as sulfates and anhydrites, reflecting a carbonate-dominated hydrochemical profile (Gad et al. 2023); G2: Characterized by elevated levels of SO_4^{2-} and Cl^-, this group is associated with evaporitic characteristics. The prevalence of chlorides and salts underscores their primary role in the salinity of GW in the study region; and G3: TDS in this group originates from both evaporitic and carbonate sources, exhibiting varied associations with salinity across the studied area. Overall, these findings illustrate that the quaternary aquifer waters in Al-Jawf are mineralized primarily due to their lithological components, with distinct hydrochemical profiles evident in different clusters.

4.4.3.2 Principal component analysis (PCA)

The PCA was performed to reduce dimensionality and identify dominant hydrochemical processes. The scree plot (Appendix Figure 45A) indicated that three principal components with eigenvalues >1 were retained (Appendix Figure 45A), explaining a cumulative 86.02% of the variance (PC1: 55.56%, PC2: 17.38%, PC3: 13.09%) (Appendix Table 15). According to (Hinge et al. 2022), these loadings are categorized as high (>0.75), moderate (0.75 to 0.50), and weak (0.50 to 0.30). The rotated component matrix (Appendix Table 16) revealed strong loadings for EC, TDS, SO_4^{2-}, Cl^-, Na^+, Mg^{2+}, and Ca^{2+} on PC1, highlighting salinity and mineral dissolution processes. PC2 was dominated by HCO_3^- and K^+, reflecting bicarbonate buffering and agricultural influence. PC3 showed strong positive loadings for CO_3^{2-} and NO_3^-, indicating rock weathering and anthropogenic nitrate input. The 3D loading plot (Appendix Figure 45B) visualizes these relationships in the factor space.

4.4.4 Geochemical modeling and mineral saturation

The PHREEQC model was applied to assess mineral concentrations, SIs, and the likelihood of GW to either dissolve or precipitate minerals (Parkhurst and Appelo1 1999). The approach's outputs involved SIs for minerals such as halite, calcite, dolomite, gypsum, anhydrite, and aragonite, as well as the partial pressure of CO_2. The input data consisted of physical and chemical parameters, including temperature, EC, pH, TDS, and major cations and anions (Appendix Figure 46). The results for selected GW samples from the quaternary aquifer are displayed in Appendix Table 16 and Fig. 46. The CO_2 partial pressure was found to be negative and below saturation levels, indicating that GW recharge was lower than the extraction from production wells in the quaternary aquifer. Additionally, the study observed a decline in CO_2 along the water flow direction, which aligns with reduced GW recharge in that direction.

Water quality is impacted by various solutes originating from processes such as soil erosion, rock weathering, and atmospheric deposition (Saleh et al. 1999). Mineral dissolution in water typically occurs when water interacts with rocks, and when the water becomes oversaturated with a particular mineral, precipitation may follow. Appendix Table 16 provides the statistical values of SI values for minerals like calcite, dolomite, gypsum, halite, aragonite, anhydrite, and CO_2. According to SI outputs for GW in the quaternary aquifer, all water samples were found to be undersaturated with anhydrite, halite, and gypsum, indicating the GW capacity to dissolve more of these minerals. Conversely, most samples were oversaturated with aragonite, calcite, and dolomite, suggesting these minerals could precipitate from the water. This observation is compatible with the Gibbs plot analysis. Specifically, only S9, S8, and S12 showed negative *SI* values for calcite, aragonite, and dolomite, respectively, and these samples were mainly situated in the central region of the Al-Jawf area. The semi-arid climate of the study region likely promotes the precipitation of dolomite, aragonite, and calcite due to low rainfall and high evaporation. The presence of Ca^{2+}, Na^+, SO_4^{2-}, and Cl^- does not prevent equilibrium with anhydrite, gypsum, and halite minerals (Güler and Thyne 2004).

4.4.5 Irrigation water quality indices

Various critical indicators must be investigated to assess the influence of agricultural soil water quality on crop quality. These indicators may encompass individual chemical metrics (Li et al. 2013; Kawo and Karuppannan 2018) or groups thereof (Meireles et al. 2010). Decision-makers can derive an effective irrigation water management strategy based on the outcomes of these indicators. In this study, the focus is on utilizing six IWQIs to classify irrigation water quality, applying typical parameter value ranges as a basis (Appendix Table 17).

The IWQI, computed using the formula detailed in Appendix Table 3 (Meireles et al. 2010), categorized water quality in the research area into five distinct criteria (Ayers and Westcot 1985). These categories were arranged based on the percentage distribution across the total samples: 14.8% of samples were classified as having low restrictions for irrigation, another 14.8% as moderately restricted, 25.9% as strongly restricted, and 18.5% as severely restricted (Appendix Figure 47A). The indicator values ranged from a minimum to a maximum, as detailed in Appendix Table 17, spanning from 17.03 to 96.77, with an average of 61.03. Areas showing water degradation based on the IWQ indicator were identified primarily in the northern and southwestern regions of the study area. The quality of irrigation water significantly influences soil composition and its agricultural productivity by affecting permeability, filtration rate, and aeration through its chemical properties (Todd and Mays 2005). Among these properties, sodium concentration is particularly critical as it can reduce soil permeability by displacing Mg^{2+} and Ca^{2+} ions through adsorption, thereby affecting soil structure and fertility. Therefore, indicators such as SAR, Na%, and SSP were calculated to assess water quality based on sodium, calcium, and magnesium concentrations, aiming to determine suitability for agricultural irrigation (Eaton 1950; McGeorge 1954; Ravikumar et al. 2013a). These indicators also pinpoint areas most susceptible to soil degradation, guiding appropriate management practices. Appendix Table 3 outlines the methodology for computing these indicators and mapping their values to identify areas prone to soil deterioration. According to the results summarized in Appendix Table 17, all water samples were rated as permissible to excellent for Na%, excellent for SAR, and suitable for the SSP index. The average values for SAR, SSP, and Na%, were 2.56, 27.57%, and 28.29%, respectively. Appendix Figures 47B-D illustrates the distribution maps of these indices, confirming that all water samples are suitable for agricultural irrigation, with no adverse effects on soil permeability or infiltration rates, and indicating minimal need for calcium fertilization due to low Na^{+} concentrations in irrigation water. The quality of irrigation water can degrade when alkali elements precipitate, especially calcium and magnesium carbonates, leading to increased Na^+ ion concentrations and SAR in irrigation water (Kumar et al. 2007; Srinivasamoorthy et al. 2014). Therefore, the RSC index was calculated to evaluate irrigation suitability, as an elevated RSC can degrade soil physical properties and cause surface crusting when the soil dries. Appendix Table 3 details the RSC index calculation to assess the likelihood of Mg^{2+} and Ca^{2+} ion precipitation on soil particles and their removal from the soil solution. Based on the results in Appendix Table 17 and its spatial distribution in Appendix Figure 47F, all samples in the study area were classified as good, with RSC values ranging from −43.21 to −1.96, confirming suitability for agricultural irrigation. Another critical index, the PS, evaluates water suitability based on Cl^- and SO_4^{2-} ion concentrations. According to the classification criteria (Appendix Table 17), PS values are categorized as Excellent to Good (PS below 3.0), Good to Injurious ($3.0 < PS < –5.0$), and Injurious to Unsatisfactory ($PS > 5.0$). The PS index results (Appendix Figure 47E), indicate that two samples were rated as Excellent to Good, three as Good, and 22 as Injurious to Unsatisfactory, highlighting areas where irrigation water may pose challenges due to salinity concerns.

4.4.6 Simulation model (ANFIS)

The ANFIS model relies on data patterns to make predictions, starting with the selection of input variables crucial for accurate predictions. These variables undergo a correlation analysis with the desired output to identify the most influential ones, a process repeated for each IWQ index (output) (Appendix Figure 8). Following preprocessing, the ANFIS model undergoes training to determine the optimal configuration for predicting each IWQ index. Subsequently, the model is validated using new data to assess its accuracy in predicting future values. The predicted IWQ indices during both training and testing phases are detailed in Appendix Table 18, while Appendix Figures 48-50 illustrate the prediction results for IWQI, SAR, and CO_2 across these phases. Appendix Figures 48-50 demonstrate substantial alignment between predicted and actual values, although some deviations from measured values are observed. Particularly noteworthy is the high R2 value (> 0.95) indicating excellent agreement between observed and predicted IWQI values. The ANFIS model exhibits robust performance across all indices, achieving high accuracy metrics (E values > 0.90) in both training and testing stages, as shown in Appendix Table 18. However, there is a slight decrease in model performance metrics such as R2, RMSE, and MAD from the training to testing phase. Appendix Figures 48-50 illustrate the discrepancy between predicted and measured IWQIs during training and testing, accompanied by comparative scatter plots. The ANFIS model effectively captures the dynamic patterns observed in IWQI data over time. These findings corroborate earlier studies (Khadr et al., 2020; Khadr and Elshemy 2017; Elsayed et al. 2021; Khadr and Schlenkhoff 2021), affirming that the ANFIS model enhances predictive accuracy compared to standalone models. Overall, ANFIS proves to be a potent predictive tool owing to its capability to handle nonlinear relationships between inputs and outputs, while also adapting swiftly to changes in training data. Additionally, One of the main limitations of this study lies in the sample size available for ML modeling. The relatively modest sample size for ML modeling in the Al-Jawf aquifer may constrain predictive resolution. However, the use of cross-validation, rigorous performance evaluation, and careful model selection helps mitigate this limitation. Future studies should aim to increase spatial and temporal sampling to further enhance model robustness and generalizability.

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

This study undertook separate assessments of water quality in two distinct hydrological settings: the Danube river in Hungary (SW) and the Al-Jawf Basin in Yemen (GW). Through HM analysis, physicochemical assessments, and geochemical modeling, the research aimed to explore spatial variation and potential risks associated with water use in each system. In the Danube river, elevated levels of manganese (Mn) and iron (Fe) were detected at multiple sites, with concentrations exceeding permissible drinking water limits. These levels are likely influenced by industrial discharges and agricultural runoff. Other metals such as Cr, Cu, Pb, Ni, As, and Zn remained below ecological risk thresholds. The HM pollution index indicated moderate pollution, particularly in the central sections (e.g., Baja). SW quality, evaluated using irrigation water quality indices (IWQIs), supported by statistical analysis and GIS tools, showed that certain river sections could be suitable for irrigation. The major ion chemistry and water types reflected both natural and anthropogenic influences, such as water–rock interactions and ion exchange processes. However, because this analysis was limited to SW, the results should not be extrapolated to GW suitability. Further studies are needed to address this gap.

In contrast, the GW assessment in Yemen's Al-Jawf basin revealed spatial heterogeneity in hydrochemical composition, driven by both geological formations and human activity. Multivariate statistical tools such as principal component analysis (PCA), combined with geochemical modeling, helped identify dominant processes, including rock-water interaction and ion exchange. The IWQIs varied across the basin, indicating that some sources may require treatment or conditional use for irrigation. Predictive modeling tools such as ANFIS and RF were employed successfully to estimate IWQI values, validating their applicability in GW assessments. Health risk assessments, while included, should be interpreted cautiously due to assumptions in exposure models and the absence of carcinogenic risk modeling.

The cross-basin comparison was exploratory and methodological in nature, aimed at evaluating the robustness and transferability of modern analytical and predictive tools–such as PCA, geochemical modeling, machine learning (RF and ANFIS), and MCS–across distinct hydrological settings. Rather than asserting equivalency between SW and GW systems, the study sought to test these methods in diverse contexts. These findings reinforce the importance of integrated, context-specific water quality assessments and suggest future research directions for improving water resource monitoring frameworks in both SW and GW environments.

5.2. Recommendations

To enhance water quality management in diverse hydrogeological contexts, this study recommends the establishment of region-specific, multi-seasonal monitoring programs. These should capture both spatial and temporal variability, with sampling designs informed by land use patterns and potential pollution sources. Effective mitigation of anthropogenic contamination requires stricter regulation of industrial discharges in the Danube basin and promotion of sustainable agricultural practices in the Al-Jawf basin. Constructed wetlands or vegetative buffers should be implemented where feasible to intercept pollutants before reaching water bodies. Advanced analytical tools, including PCA, geochemical modeling, and water quality indices, should be routinely integrated into assessment frameworks to identify dominant geochemical processes and pollution sources. MS, as demonstrated in this study, are vital for probabilistic health risk assessments and should focus particularly on vulnerable groups such as children. Predictive models (e.g., ANFIS, RF) should be refined with broader datasets and validated through field observations to support GW resource planning and irrigation suitability decisions. In regions where contaminant levels exceed safe thresholds, small-scale water treatment solutions such as filtration or ion exchange should be prioritized, particularly in rural or decentralized areas. Future research should explore the impacts of extreme climatic events and investigate emerging contaminants like PFAS, microplastics, and pharmaceuticals. Interdisciplinary collaboration, combining hydrogeology, environmental science, public health, and community engagement, is essential for addressing complex water quality challenges and developing sustainable, science-based policy responses.

6. Key Scientific Results (Thesis Points)

1. **Multivariate Fingerprinting of Heavy Metal Sources in the Danube**
   I established a multivariate fingerprinting framework for identifying heavy metal (HM) sources in Danube surface waters using Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). I identified three source-specific clusters: industrial (Fe, Cr, Mn, Zn), urban (Pb, Ni, Cu), and geogenic (As). I demonstrated that co-occurrence patterns such as Fe–Mn ($r = 0.515$) revealed shared geochemical behaviors, enabling site-specific pollution source attribution at an unprecedented regional resolution.

2. **Seasonal Dynamics of Heavy Metal Pollution Linked to Hydrology**
   I quantified a 12.8% increase in the Heavy Metal Pollution Index (HPI) during the April–September compared to the October-March season, accompanied by elevated conductivity and reduced river flow. I statistically linked these seasonal HPI variations to river discharge and land use patterns, demonstrating that reduced flow intensifies pollutant concentrations in specific reaches of the Danube.

3. **Geochemical Evolution and Lithological Control Along the Danube Flow Path**
   I revealed that geochemical composition along the northward Danube flow path in northwestern Hungary shifts due to dedolomitization processes, as shown by Gibbs diagrams and major ion ratios (Ca^{2+}, Mg^{2+}, HCO_3^-). I demonstrated that Ca-Mg-HCO_3 water types dominate, reflecting lithological control over solute chemistry and highlighting natural buffering mechanisms in the basin.

4. **Machine Learning–Based Prediction of Groundwater Quality in Arid Regions**
   I applied Random Forest (RF) and Adaptive Neuro-Fuzzy Inference System (ANFIS) models to predict Irrigation Water Quality Index (IWQI) values and salinity hazard levels in the Al-Jawf aquifer (arid zone). I achieved high predictive accuracy (RF $R^2 = 0.988$) despite limited sample size, and showed that ANFIS outperformed in nonlinear classification tasks. I demonstrated that AI-based approaches are viable for groundwater quality forecasting in data-scarce arid environments.

5. **Geochemical Characterization and Processes Governing Groundwater Quality in the Al-Jawf Basin**

I identified the dominant hydrochemical facies of the Al-Jawf aquifer using Piper and Gibbs diagrams, revealing Ca–Mg–HCO_3 and Na–Cl water types as the most prevalent. By integrating PCA with geochemical modeling, I demonstrated that water quality is primarily shaped by carbonate dissolution, silicate weathering, and ion-exchange reactions. This was the first systematic identification of geochemical processes influencing groundwater quality in this arid basin, providing a mechanistic basis for interpreting salinity hazards and irrigation risks.

7. SUMMARY

This study provides a comprehensive evaluation of water quality and related health risks in two distinct geographic regions: the Danube river in Hungary and the Al-Jawf basin in Yemen. The assessment encompasses both surface water and groundwater, employing a variety of advanced analytical methods, statistical tools, and predictive modeling techniques to address environmental sustainability and public health concerns. In the Danube river study, HM concentrations (Fe, Mn, Cr, Zn, Cu, Ni, Pb, and As) were analyzed from 2013 to 2019. Utilizing Spearman correlation, heatmap cluster analysis (CA), and various pollution indices such as the Heavy Metal Pollution Index (HPI), Metal Index (MI), hazard quotient (HQ), hazard index (HI), and carcinogenic risk (CR), the research identified moderate pollution levels, with significant local contamination in Dunaföldvár and Hercegszántó. The findings highlighted potential carcinogenic risks from chromium exposure, especially for children, emphasizing the need for targeted mitigation strategies. Furthermore, the evaluation of physicochemical properties revealed that natural processes and human activities influence water quality. The dominant water type was identified as $Ca\text{-}Mg\text{-}HCO_3$, and statistical analyses indicated the impact of water-rock interaction, dolomite dissolution, and ion exchange. Although the water was generally suitable for irrigation, it was deemed unfit for drinking based on WQI values, with children facing higher non-carcinogenic risks compared to adults. In the Al-Jawf basin study, groundwater quality was assessed for irrigation purposes. Techniques such as water quality indices (IWQIs), geochemical modeling, multivariate statistical analysis, GIS, and adaptive neuro-fuzzy inference systems (ANFIS) were employed. The groundwater composition followed the order: $Ca^{2+} > Mg^{2+} > Na^{+} > K^{+}$ and $SO_4^{2-} > Cl^{-} > HCO_3^{-} > NO_3^{-}$. The evaluation highlighted moderate-to-severe restrictions for irrigation in some samples. The ANFIS model demonstrated high accuracy in predicting IWQIs, providing critical guidance for sustainable water resource management. Additionally, the study assessed HM contamination and health risks, utilizing indices such as IWQIs and Health Risk Indices (HI), and predictive models like Random Forest (RF). The research traced hydrochemical signatures to natural processes and human activities, identifying some groundwater sources that posed dermal exposure risks, particularly to children. Overall, this study underscores the importance of continuous monitoring and the implementation of effective mitigation strategies to ensure the sustainability of water resources and protect public health. The integrated approach and findings offer valuable insights for environmental and public health policy-making in both regions.

APPENDIX A – FIGURES

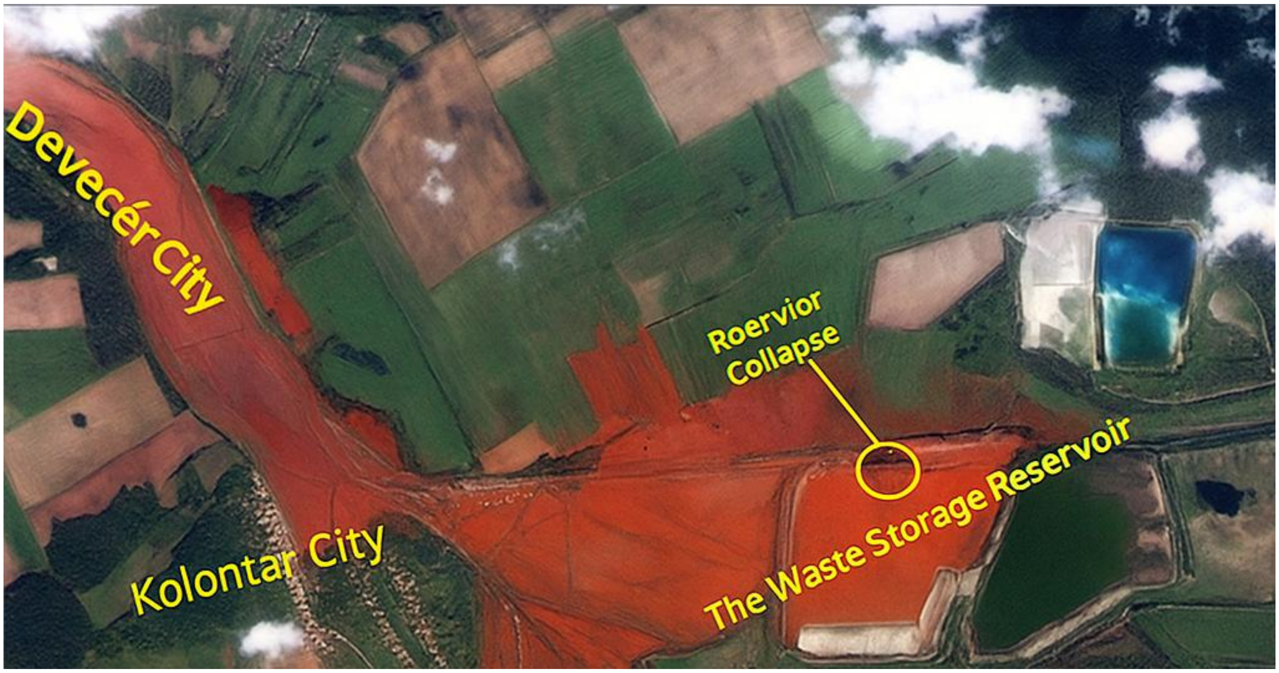


Figure 1 Devastating toxic sludge spill in Hungary, 2010 (Chemistry World, 2011)

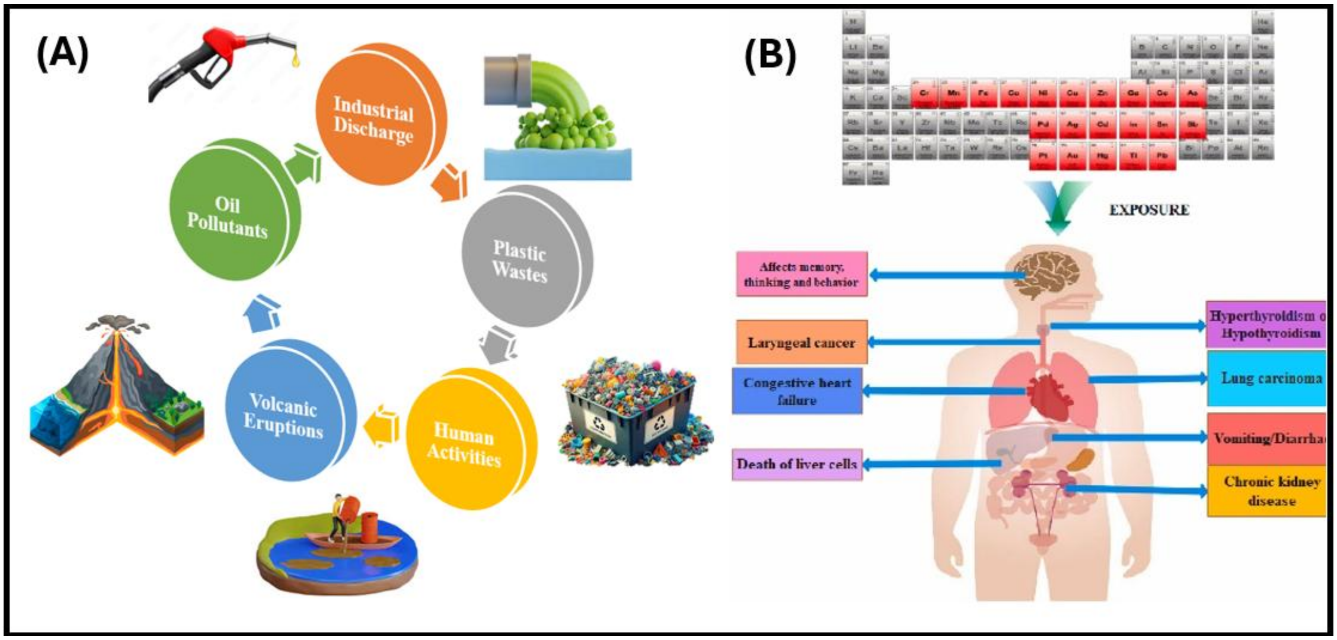


Figure 2 (A) Sources contributing to heavy metal pollution in the environment; (B) Toxic effects of heavy metals on human health (Saravanan et al. 2024).

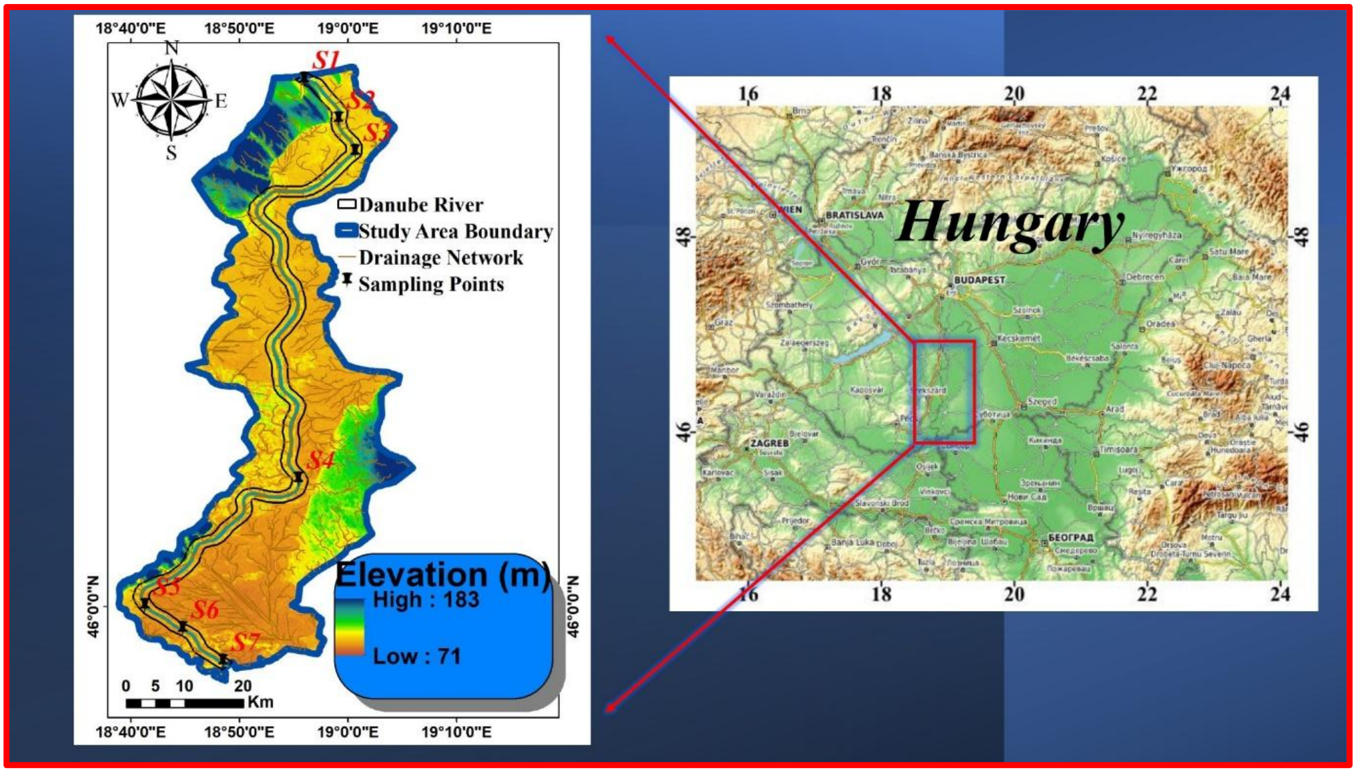


Figure 3 Study area and the distribution of sampling points in the Danube River

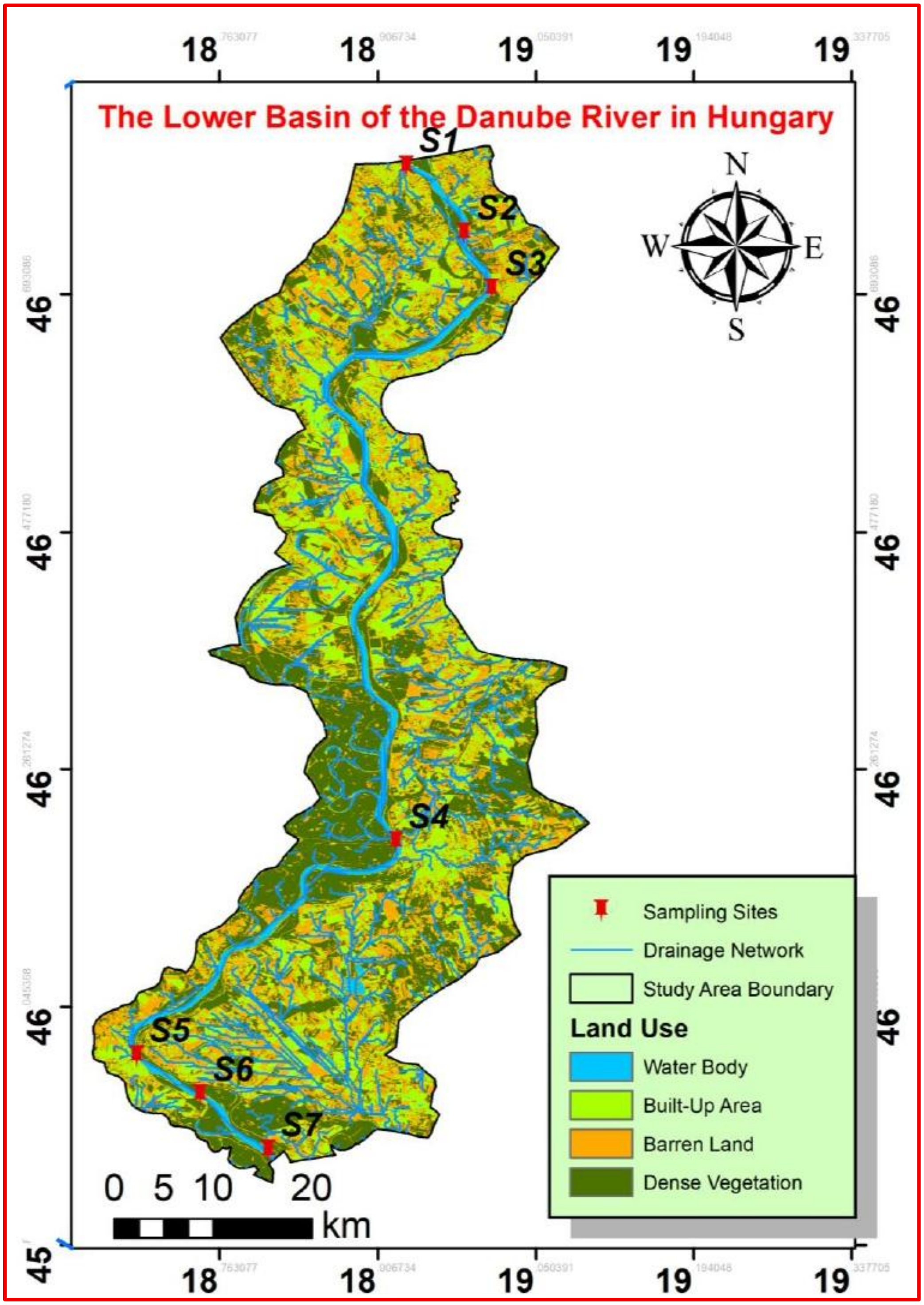


Figure 4 Land use map of the study area

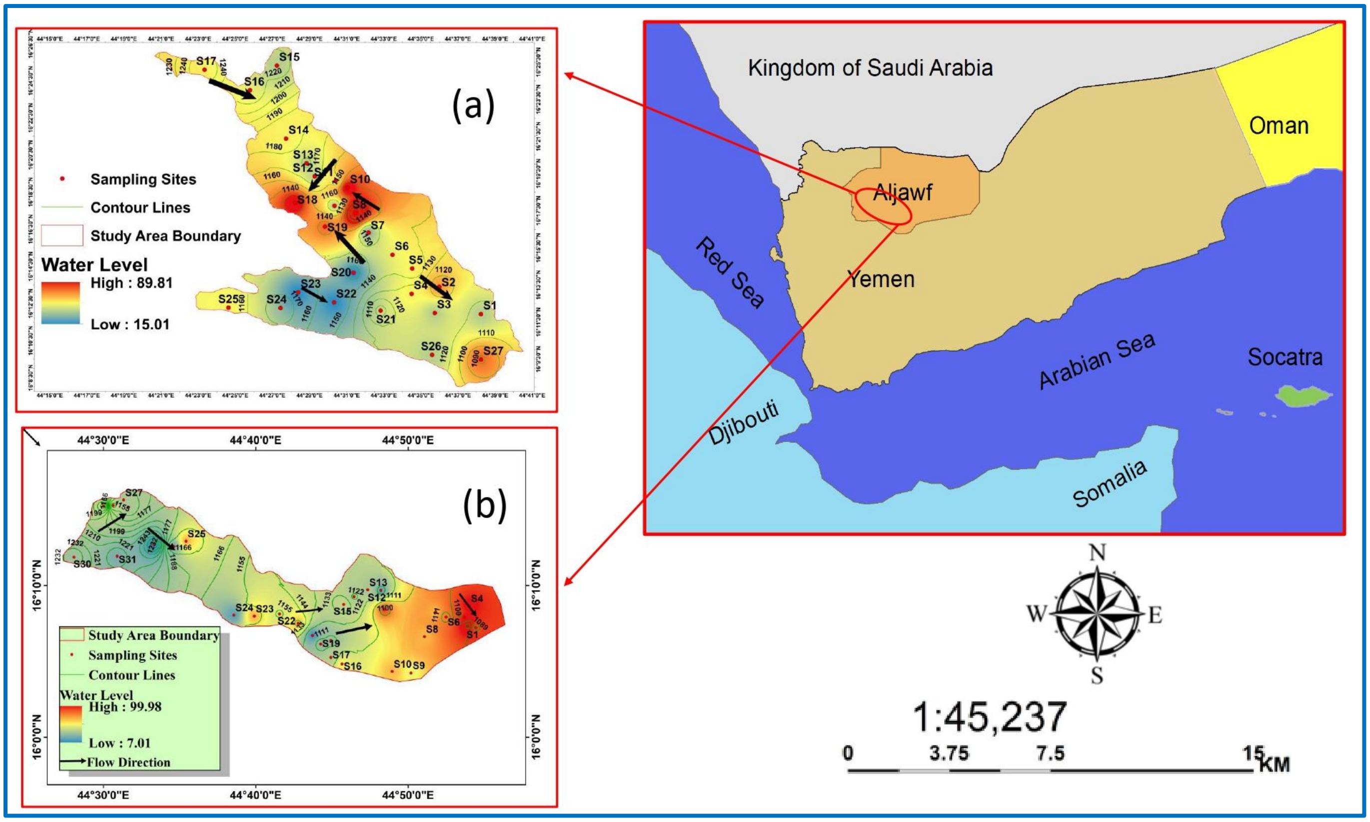


Figure 5 Location and sampling sites within the Al-Jawf basin, Yemen

(The right panel shows the location of Al-Jawf Governorate in Northern Yemen. Sub-panels (a)(2021) and (b)(2022) illustrate two different parts of the basin where groundwater sampling and contour mapping were conducted, including water level gradients and site codes. Arrows indicate the groundwater flow within the study zones.)

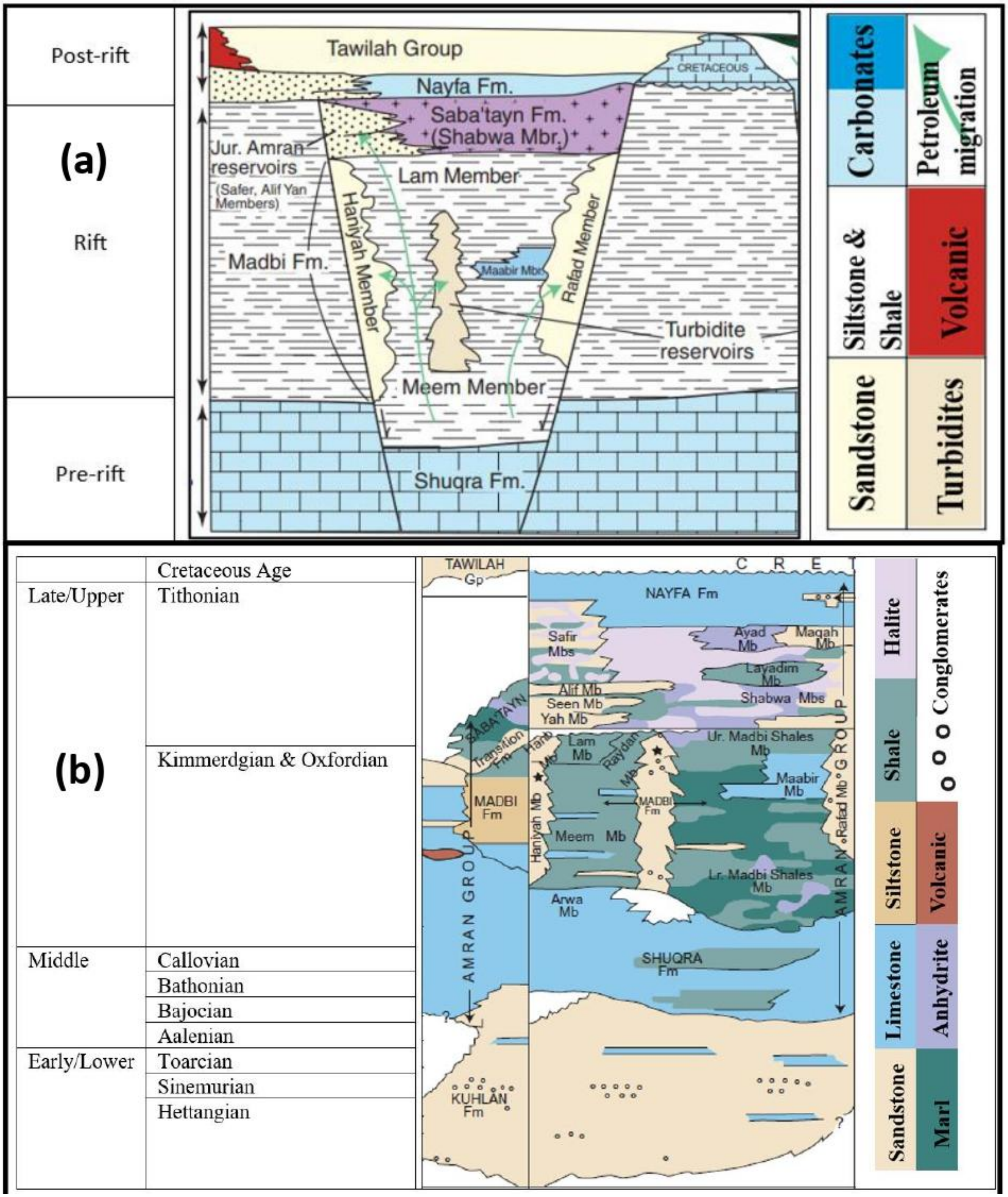


Figure 6 Geological, Hydrogeological units in the Al-Jawf basin (source: modified after (Ahlbrandt 2002)

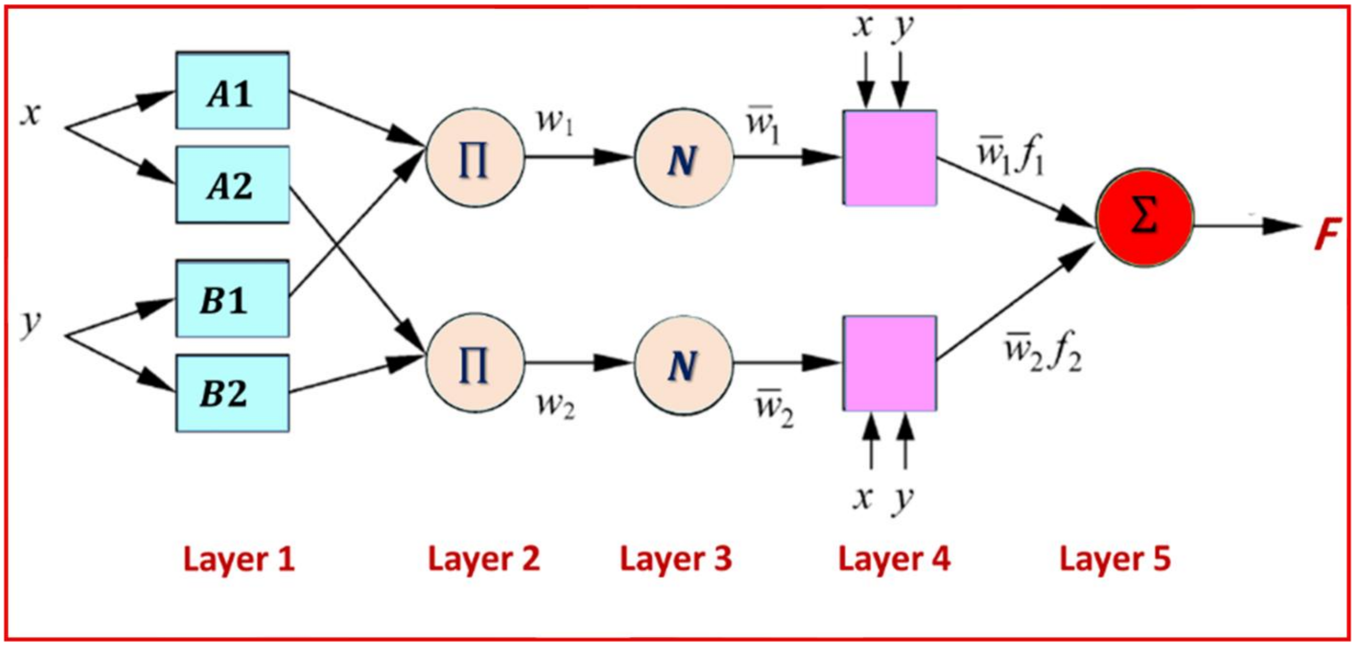


Figure 7 ANFIS system with a two-rule Sugeno system.

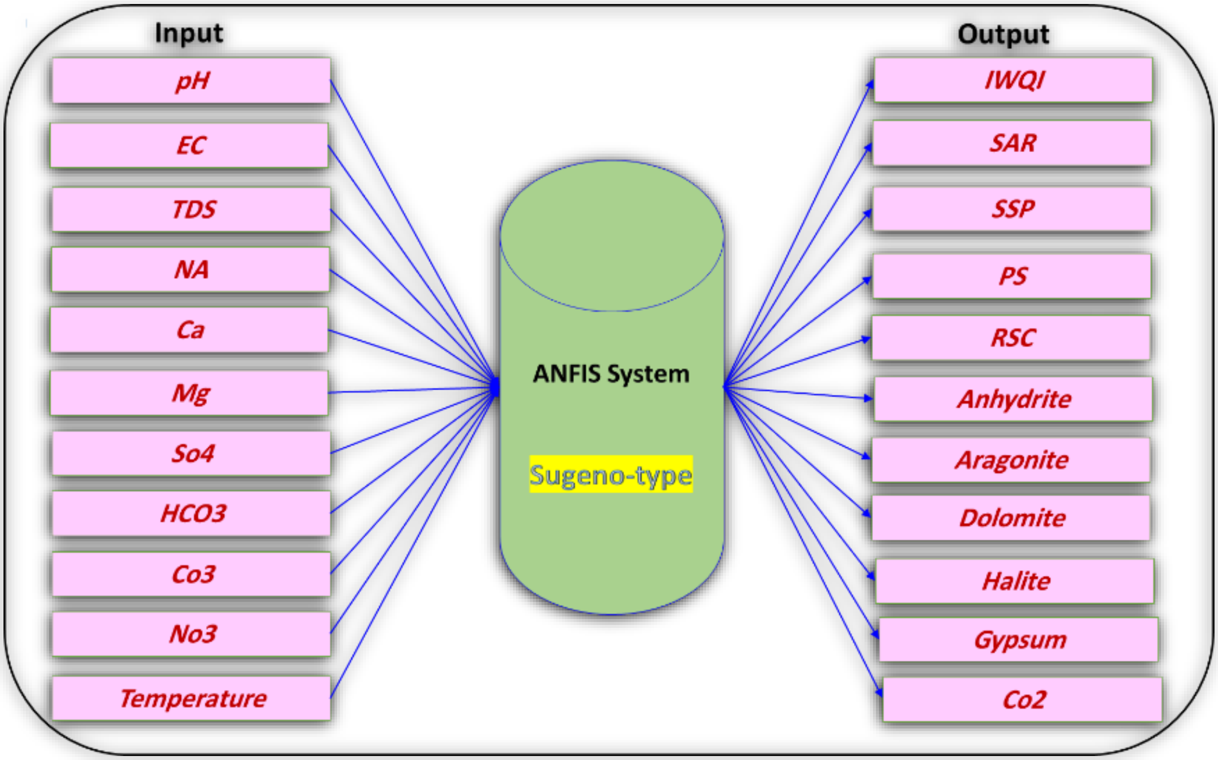
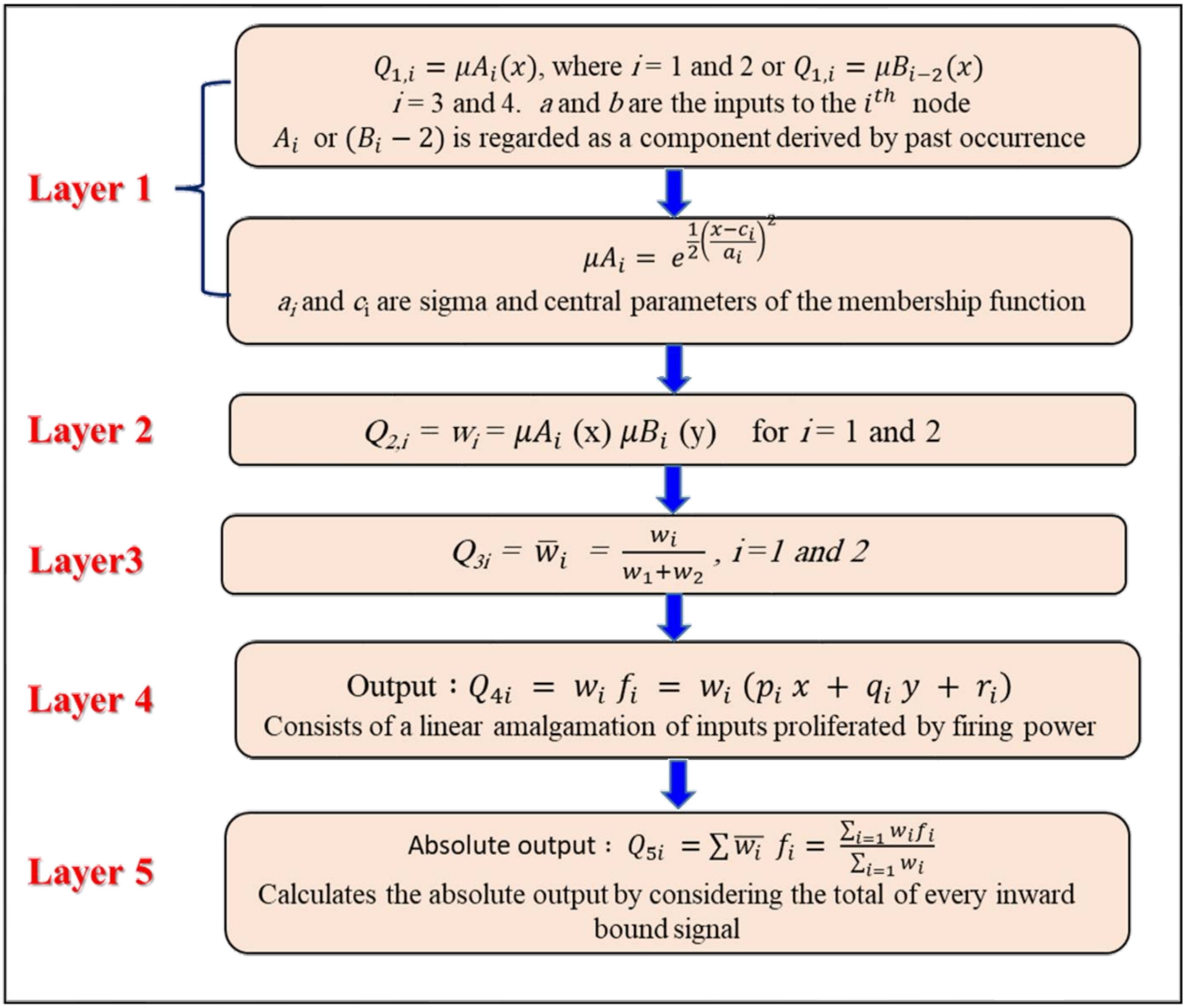


Figure 8 An ANFIS architecture for IWQs prediction.



$Q_{1,i} = \mu A_i(x)$, where i = 1 and 2 or $Q_{1,i} = \mu B_{i-2}(x)$
i = 3 and 4. a and b are the inputs to the i^{th} node
A_i or $(B_i - 2)$ is regarded as a component derived by past occurrence

$$\mu A_i = e^{\frac{1}{2}\left(\frac{x-c_i}{a_i}\right)^2}$$

a_i and c_i are sigma and central parameters of the membership function

$Q_{2,i} = w_i = \mu A_i$ (x) μB_i (y) for i = 1 and 2

$Q_{3i} = \overline{w}_i = \frac{w_i}{w_1+w_2}$, i=1 and 2

Output : $Q_{4i} = w_i f_i = w_i (p_i x + q_i y + r_i)$
Consists of a linear amalgamation of inputs proliferated by firing power

Absolute output : $Q_{5i} = \sum \overline{w_i} f_i = \frac{\sum_{i=1} w_i f_i}{\sum_{i=1} w_i}$
Calculates the absolute output by considering the total of every inward bound signal

Figure 9 Conceptual illustration of the five layers that comprise ANFIS

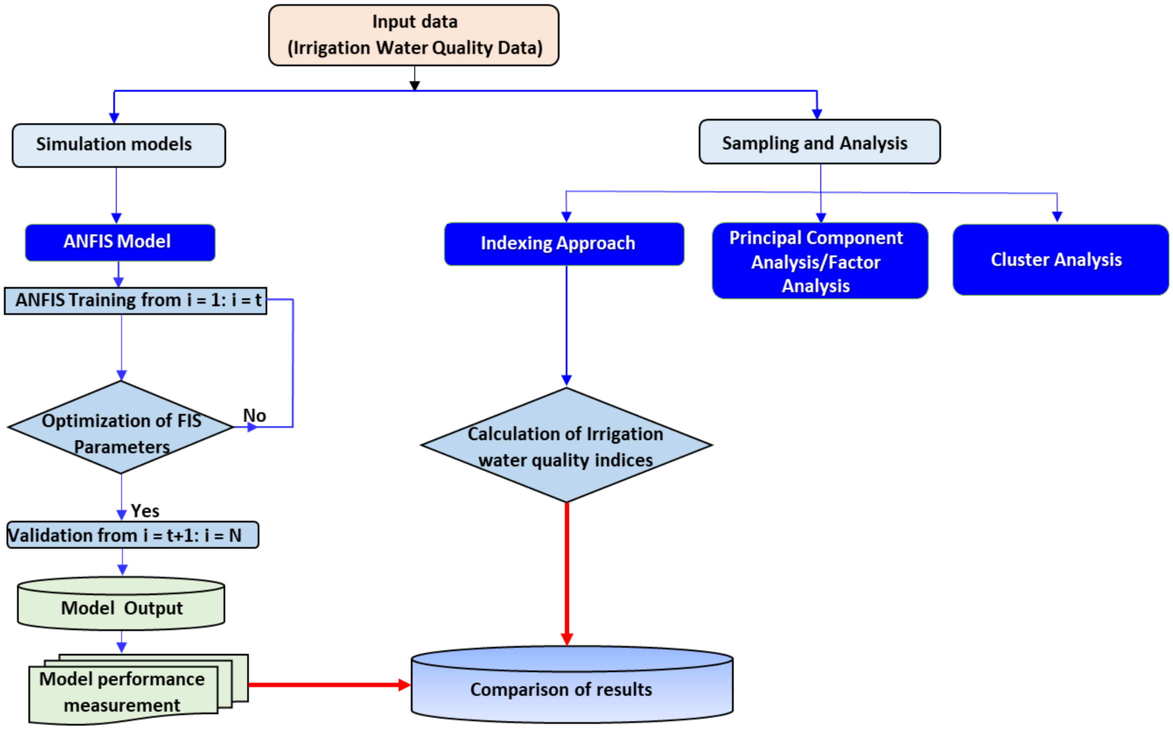


Figure 10 Schematic diagram of the methodology presented in this study.

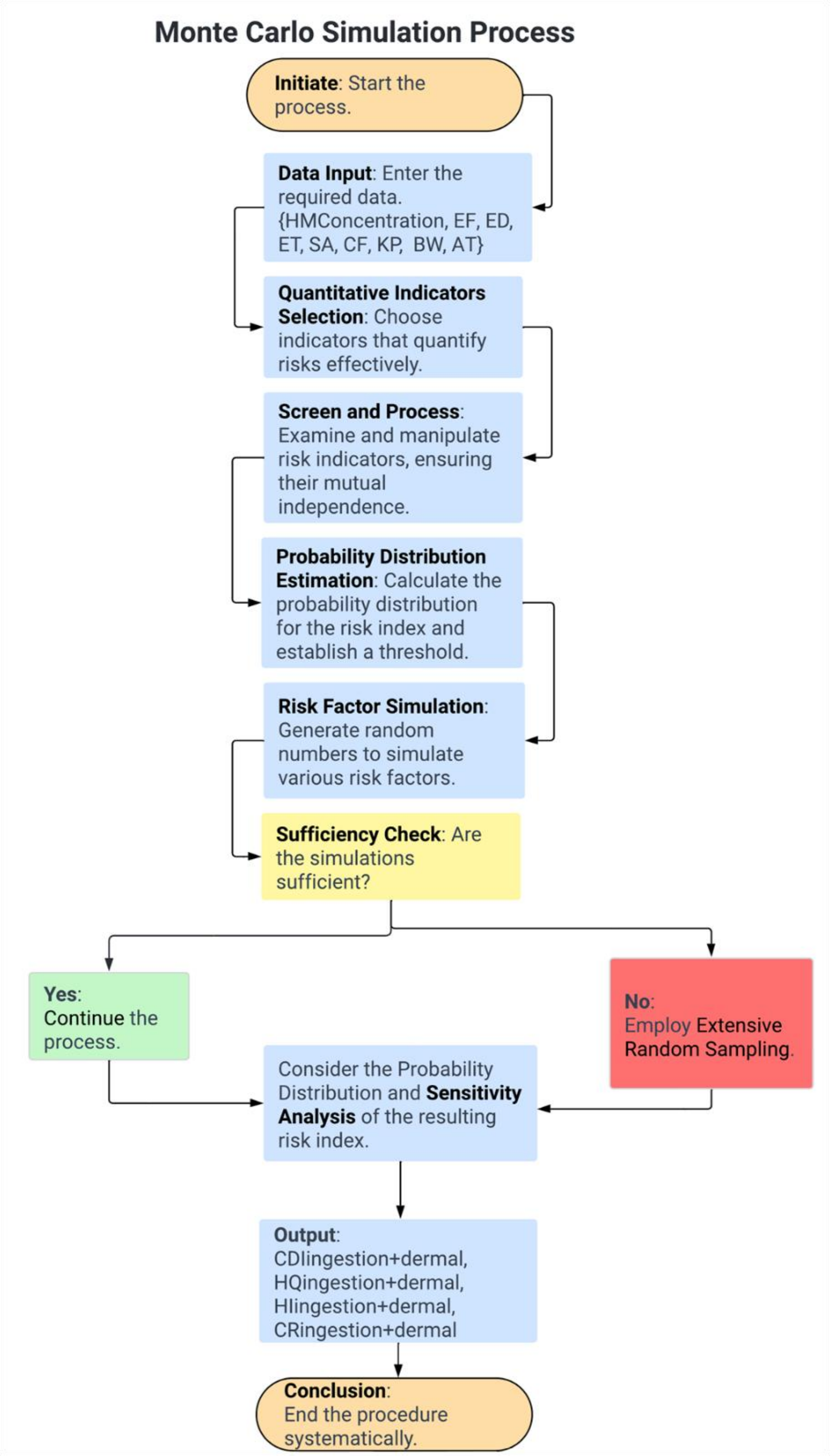


Figure 11 Operational sequence of a Monte Carlo simulation model

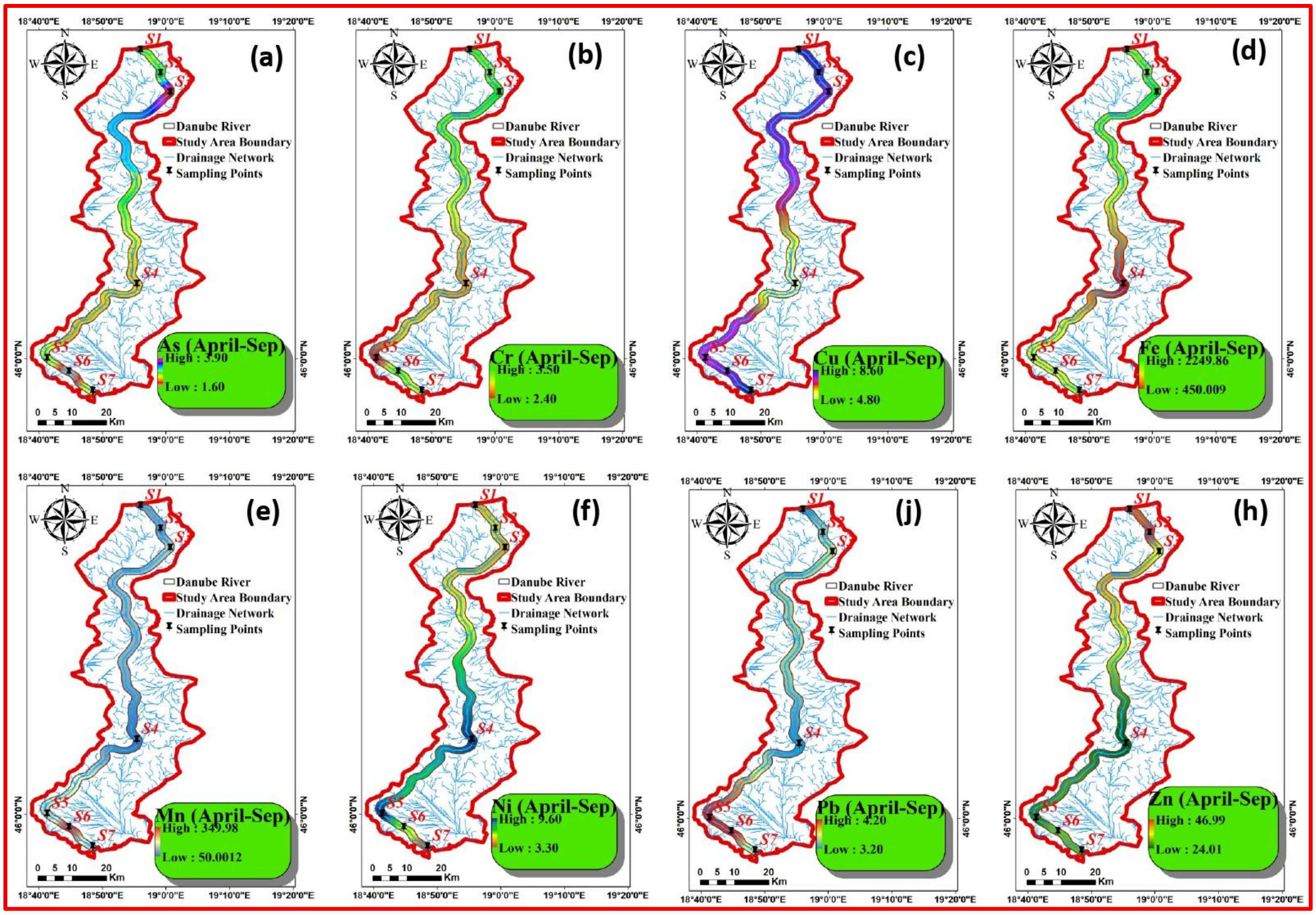


Figure 12 Spatial distribution maps of heavy metal concentrations in surface water samples from the Danube river basin during the April–September period. The maps illustrate interpolated values for As, Cr, Cu, Fe, Mn, Ni, Pb, and Zn using Inverse Distance Weights

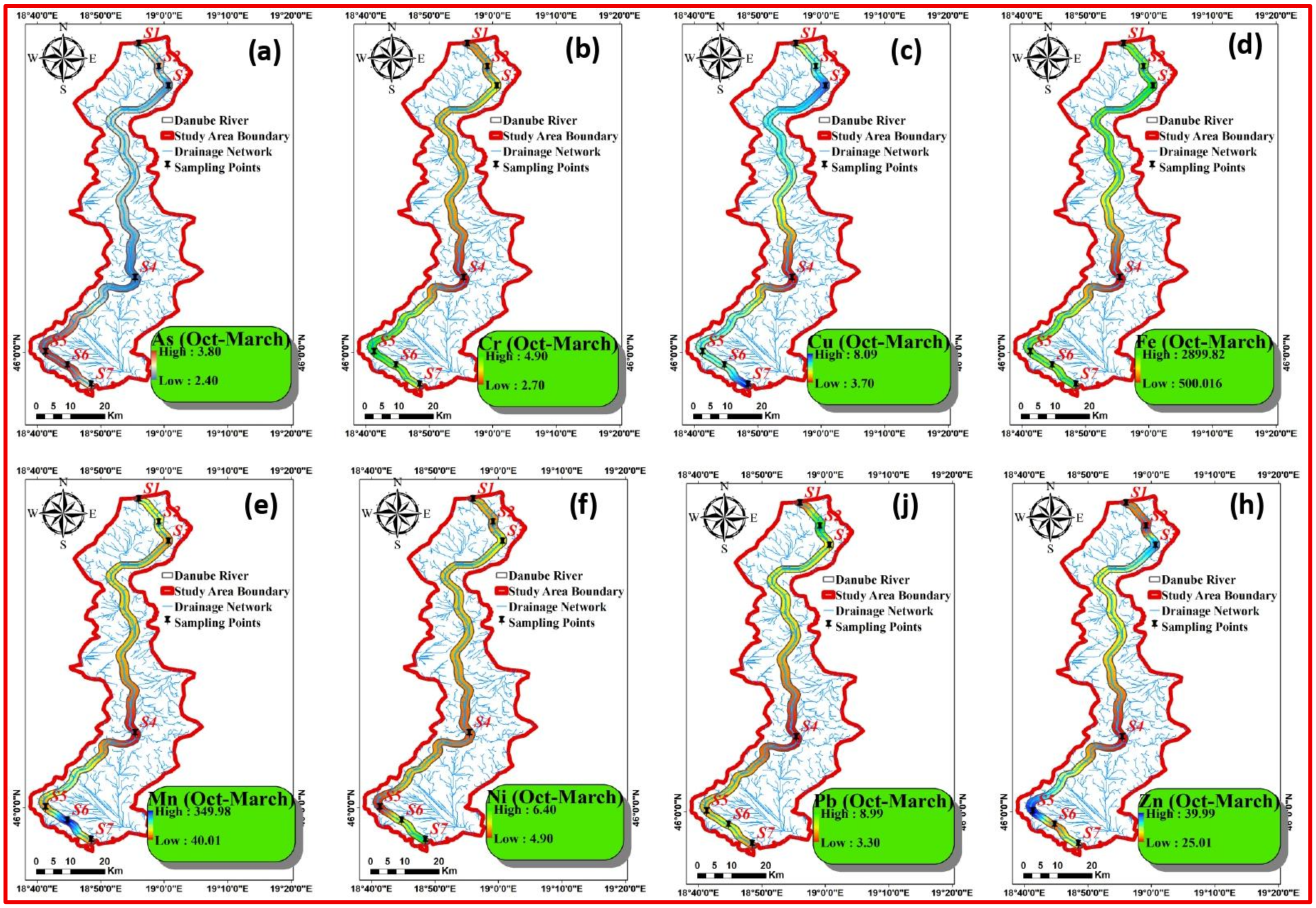


Figure 13 Spatial distribution maps of heavy metal concentrations in surface water samples from the Danube river basin during the October–March period. The maps illustrate interpolated values for As, Cr, Cu, Fe, Mn, Ni, Pb, and Zn using Inverse Distance Weights

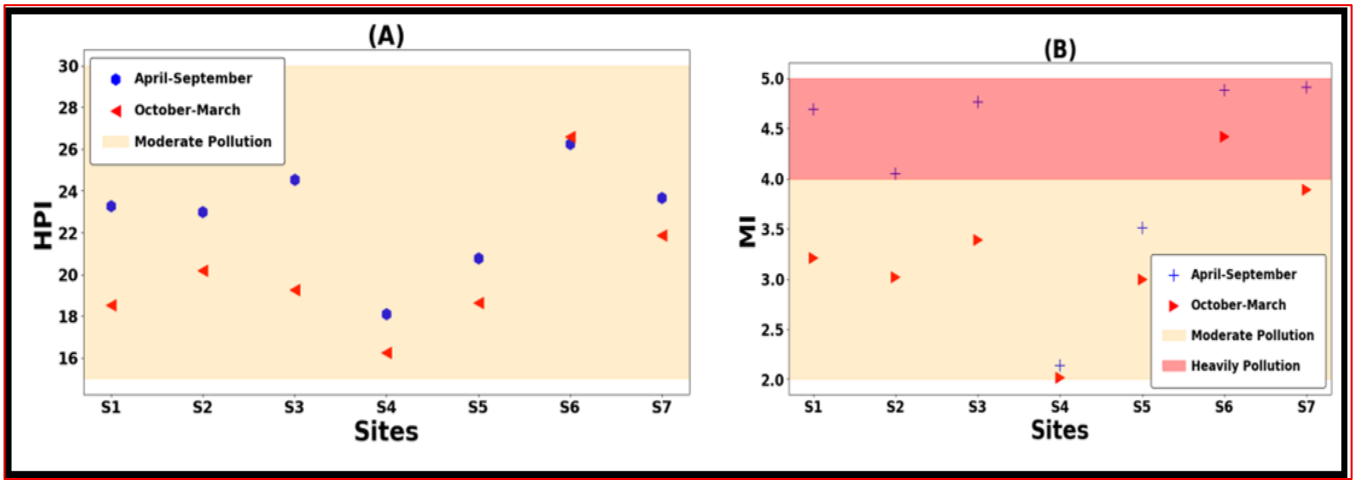
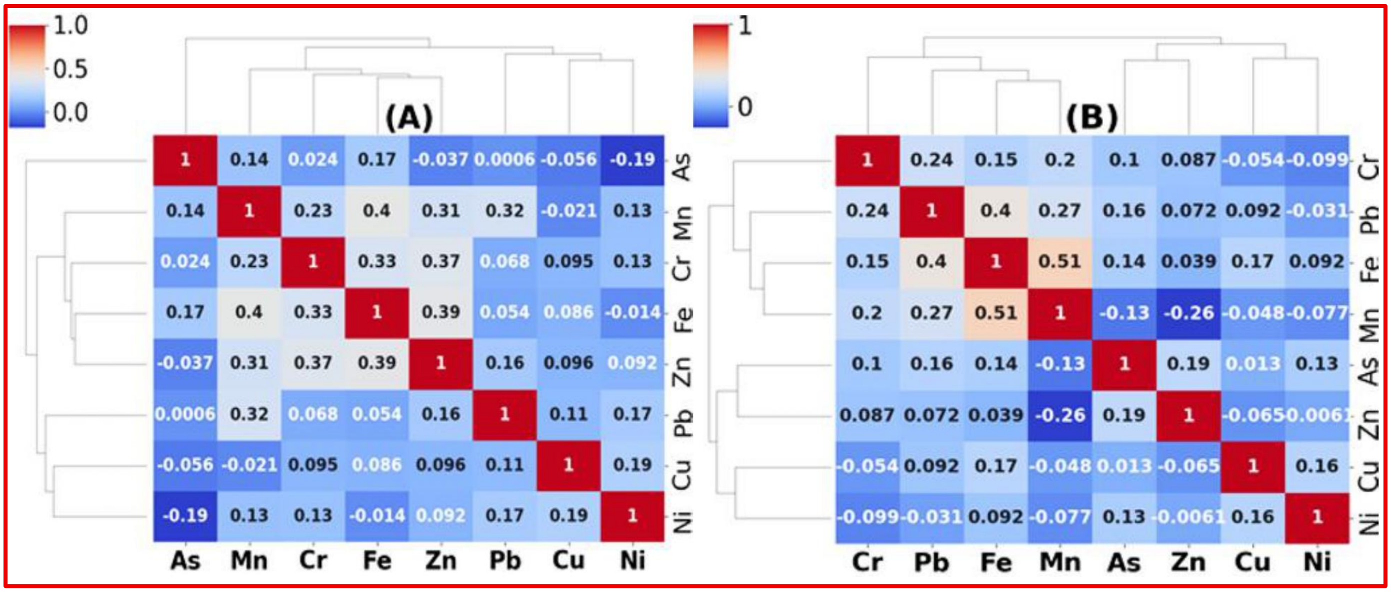


Figure 14 Heavy metal pollution index (HPI) (A) and metal index (MI) (B)



(A)

	As	Mn	Cr	Fe	Zn	Pb	Cu	Ni
As	1	0.14	0.024	0.17	-0.037	0.0006	-0.056	-0.19
Mn	0.14	1	0.23	0.4	0.31	0.32	-0.021	0.13
Cr	0.024	0.23	1	0.33	0.37	0.068	0.095	0.13
Fe	0.17	0.4	0.33	1	0.39	0.054	0.086	-0.014
Zn	-0.037	0.31	0.37	0.39	1	0.16	0.096	0.092
Pb	0.0006	0.32	0.068	0.054	0.16	1	0.11	0.17
Cu	-0.056	-0.021	0.095	0.086	0.096	0.11	1	0.19
Ni	-0.19	0.13	0.13	-0.014	0.092	0.17	0.19	1

(B)

	Cr	Pb	Fe	Mn	As	Zn	Cu	Ni
Cr	1	0.24	0.15	0.2	0.1	0.087	-0.054	-0.099
Pb	0.24	1	0.4	0.27	0.16	0.072	0.092	-0.031
Fe	0.15	0.4	1	0.51	0.14	0.039	0.17	0.092
Mn	0.2	0.27	0.51	1	-0.13	-0.26	-0.048	-0.077
As	0.1	0.16	0.14	-0.13	1	0.19	0.013	0.13
Zn	0.087	0.072	0.039	-0.26	0.19	1	-0.065	-0.0061
Cu	-0.054	0.092	0.17	-0.048	0.013	-0.065	1	0.16
Ni	-0.099	-0.031	0.092	-0.077	0.13	-0.0061	0.16	1

Figure 15 Heatmap and Spearman correlation for the studied heavy metal: (A) during April-September; (B) during October-March

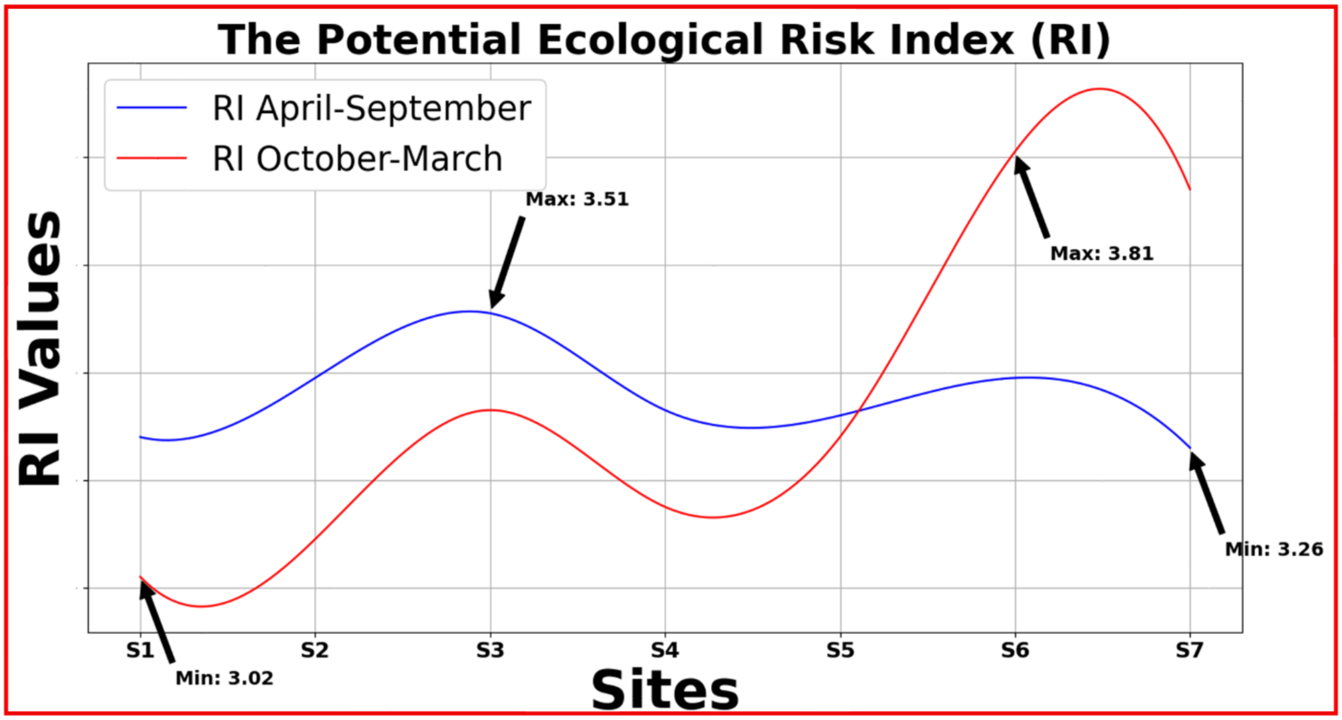


Figure 16 Ecological risk index (RI)

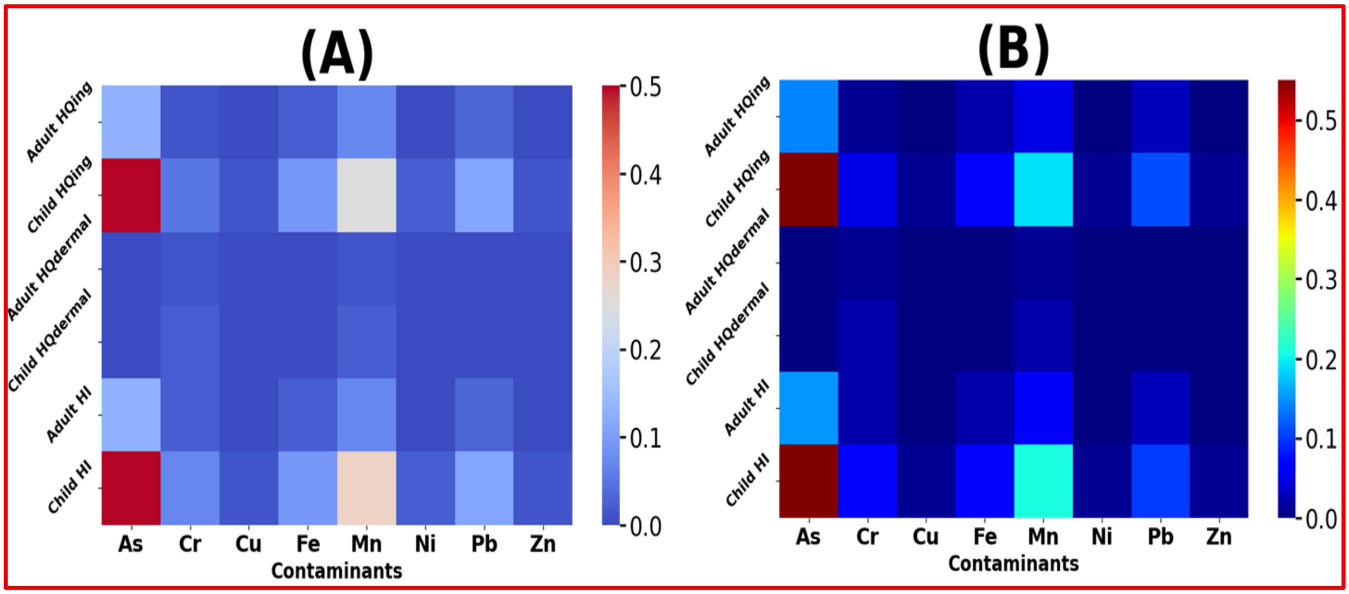


Figure 17 Hazard quotient and Hazard index for oral and dermal pathways: (A) during April-September; (B) during October-March

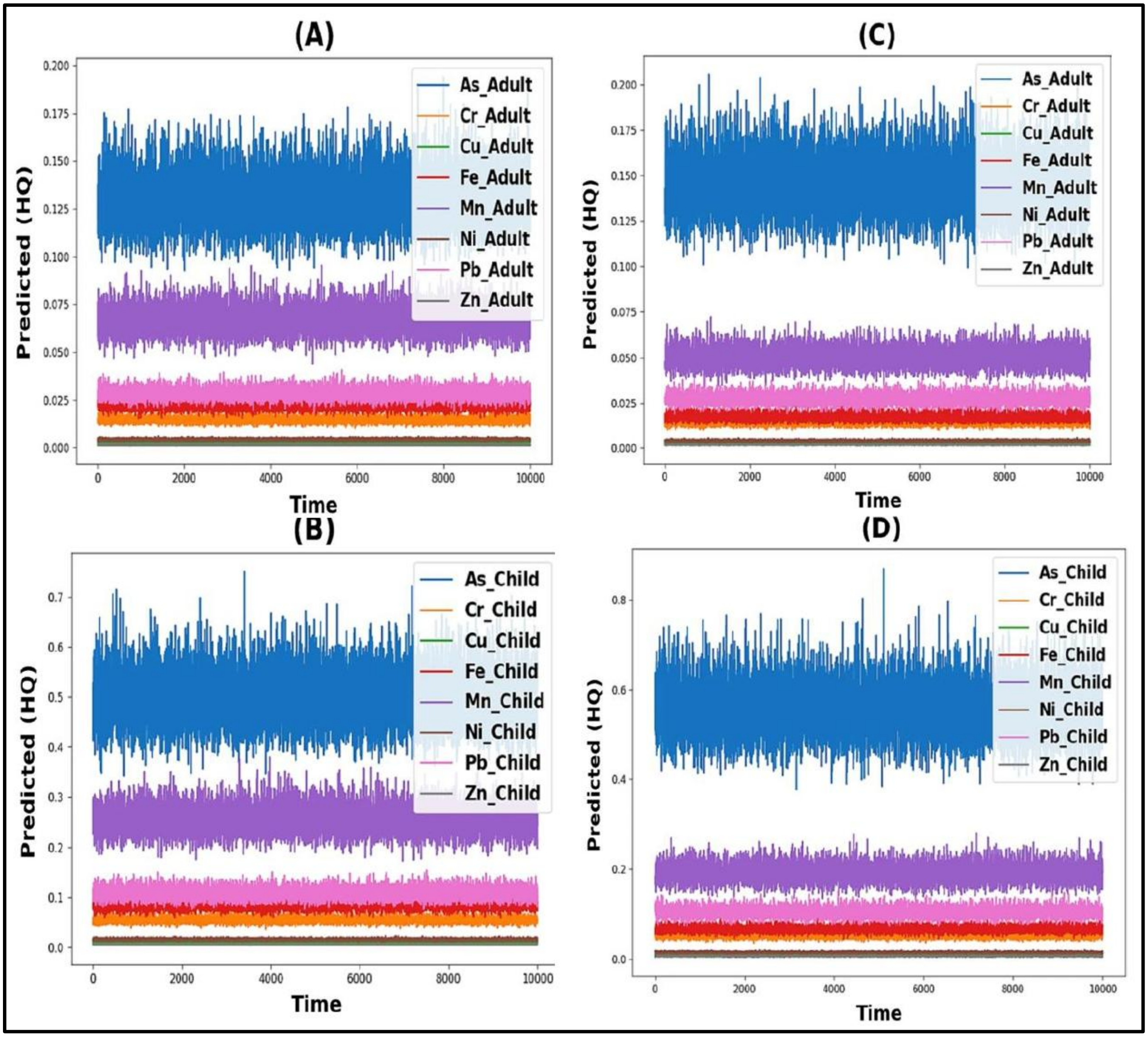


Figure 18 Predicted oral hazard quotient (A) Adults oral during April-September; (B) Children oral during April-September; (C) Adults oral during October-March & (D) Children oral during October-March

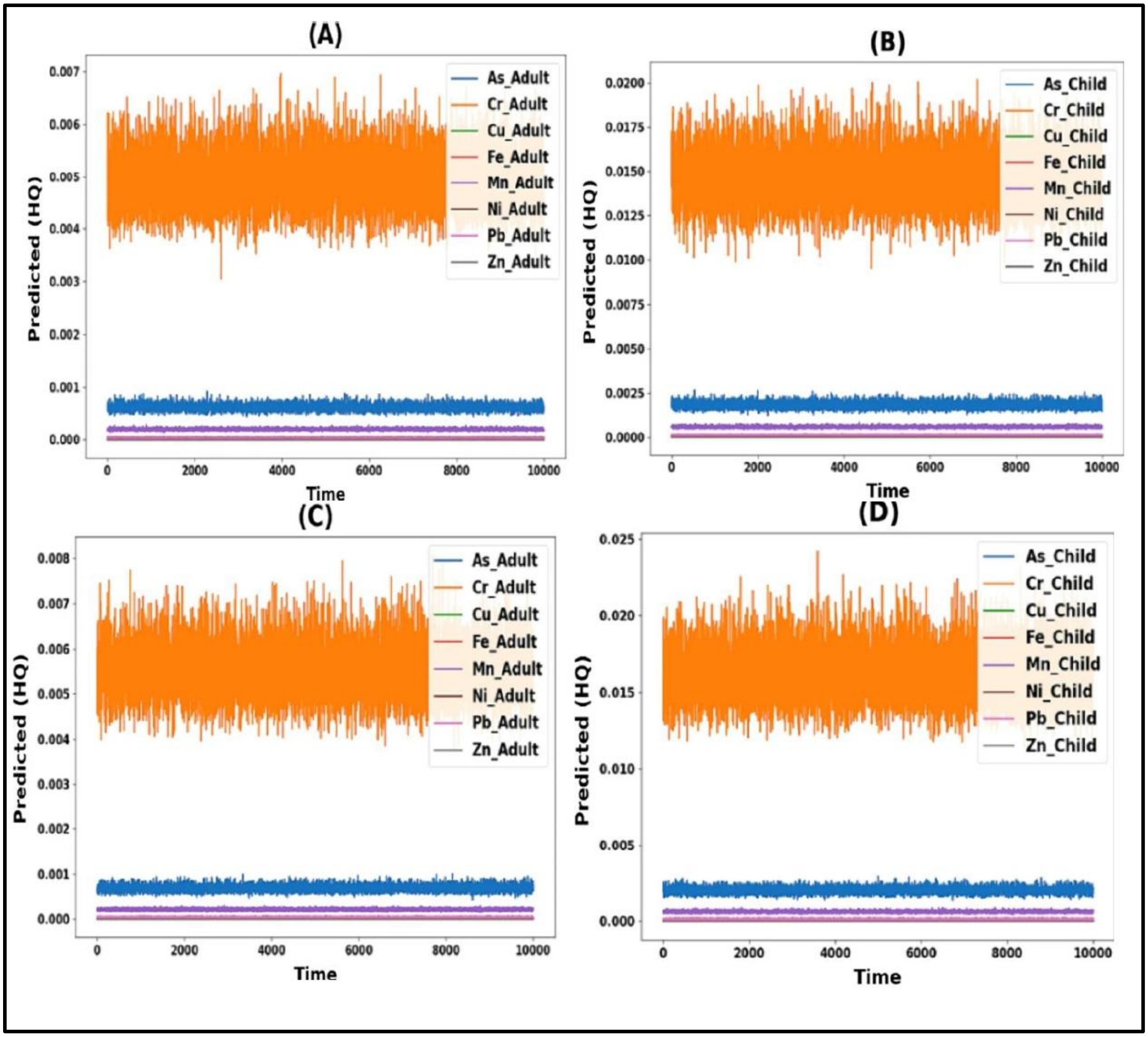


Figure 19 Predicted dermal hazard quotient (A) Adults dermal during April-September; (B) Children dermal during April-September; (C) Adults dermal during October-March & (D) Children dermal during October-March

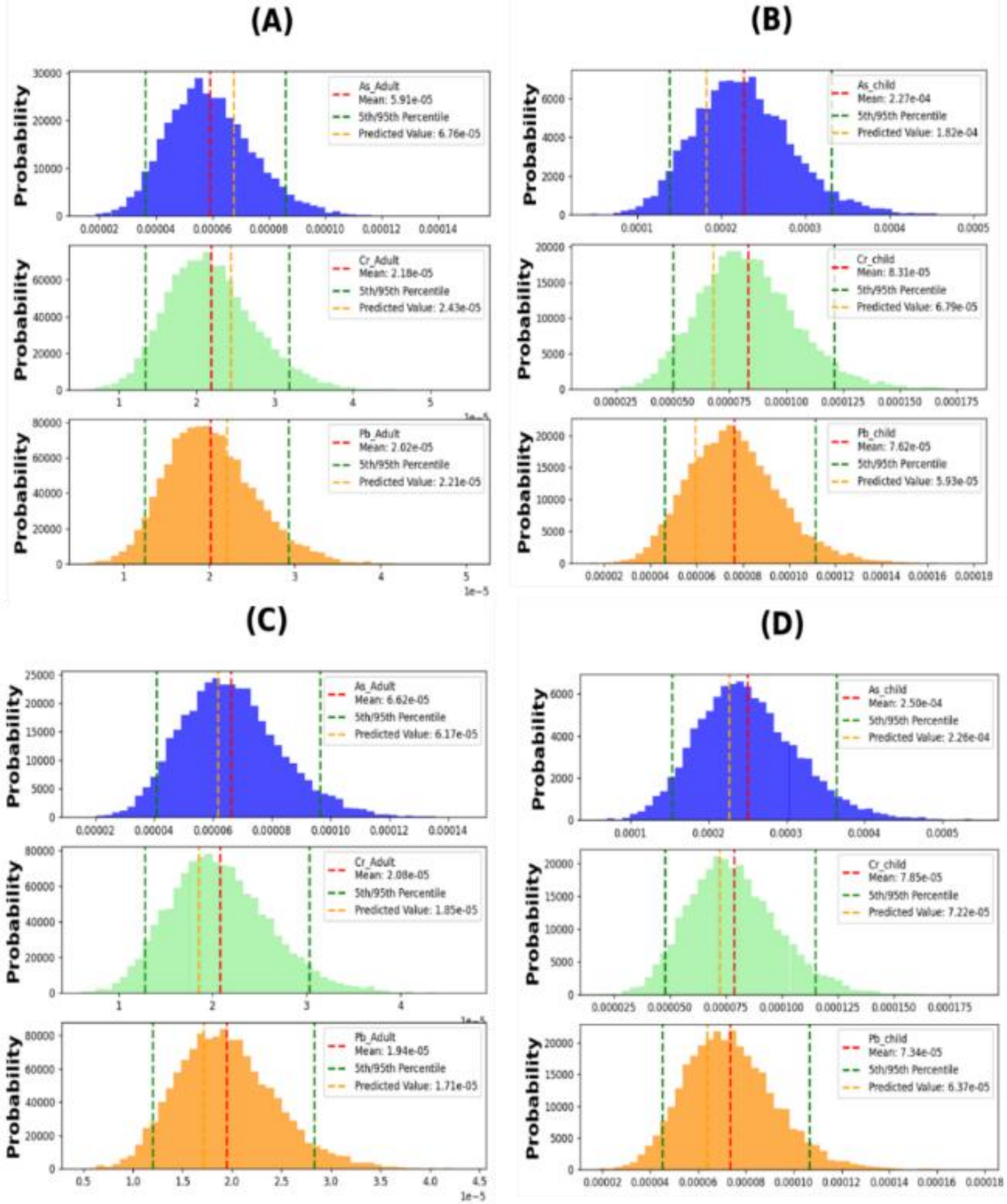


Figure 20 Carcinogenic risk oral: (A) adults in April-September; (B) children in April-September; (C) adults in October-March; (D) children in October-March

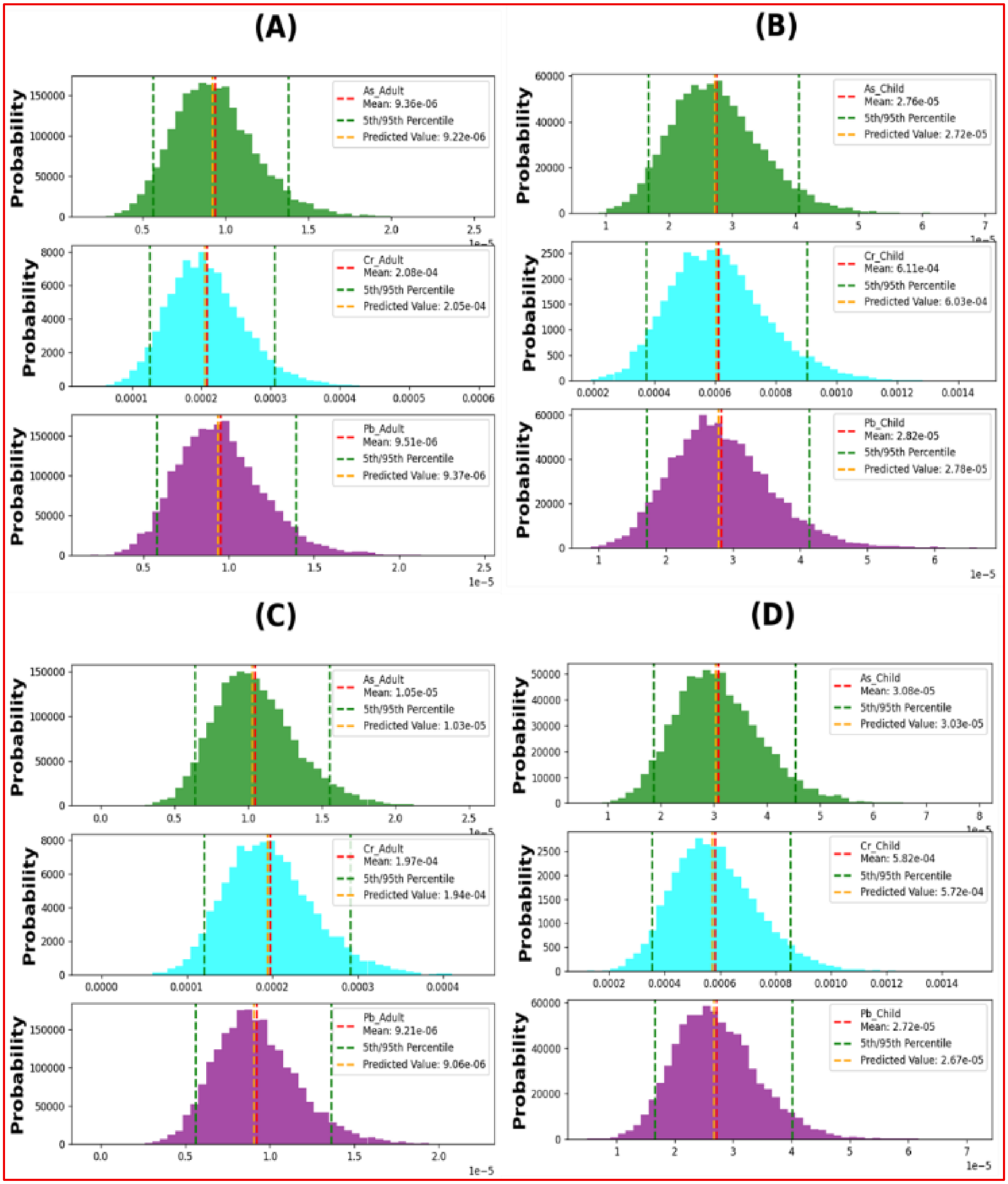


Figure 21 Carcinogenic risk dermal: (A) adults in April-September; (B) children in April-September; (C) adults in October-March; (D)

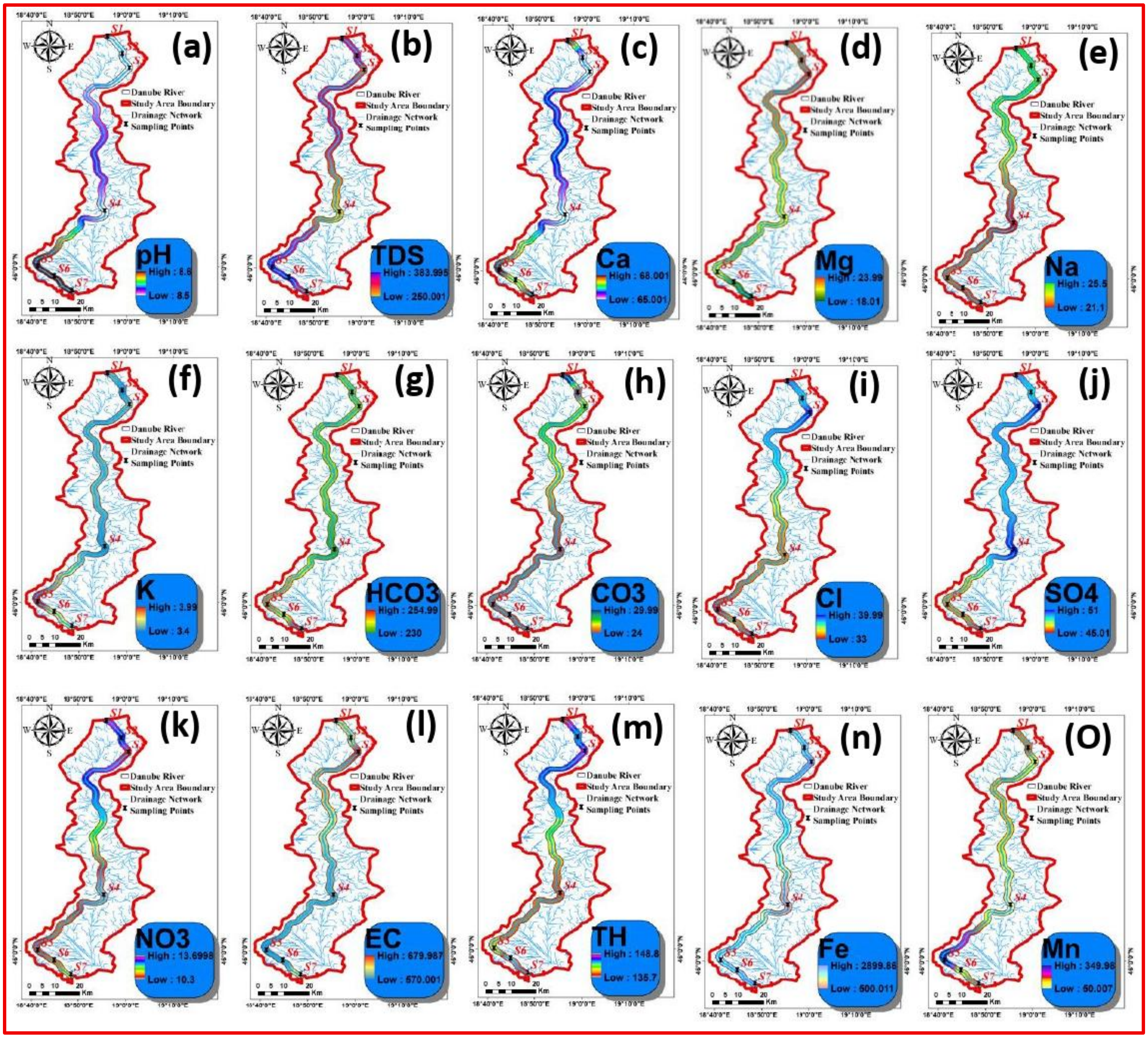


Figure 22 Distribution map of the HM and physicochemical parameters (maximum): (a) pH, (b)TDS, (c) Ca^{2+}, (d) Mg^{2+}, (e) Na^{+}, (f) K^{+}, (g) HCO_3^{-}, (h) CO_3^{-}, (i) Cl^{-}, (j) SO_4^{2-}, (k) NO_3^{-}, (l) EC, (m) TH, (n) Fe and (o) Mn.

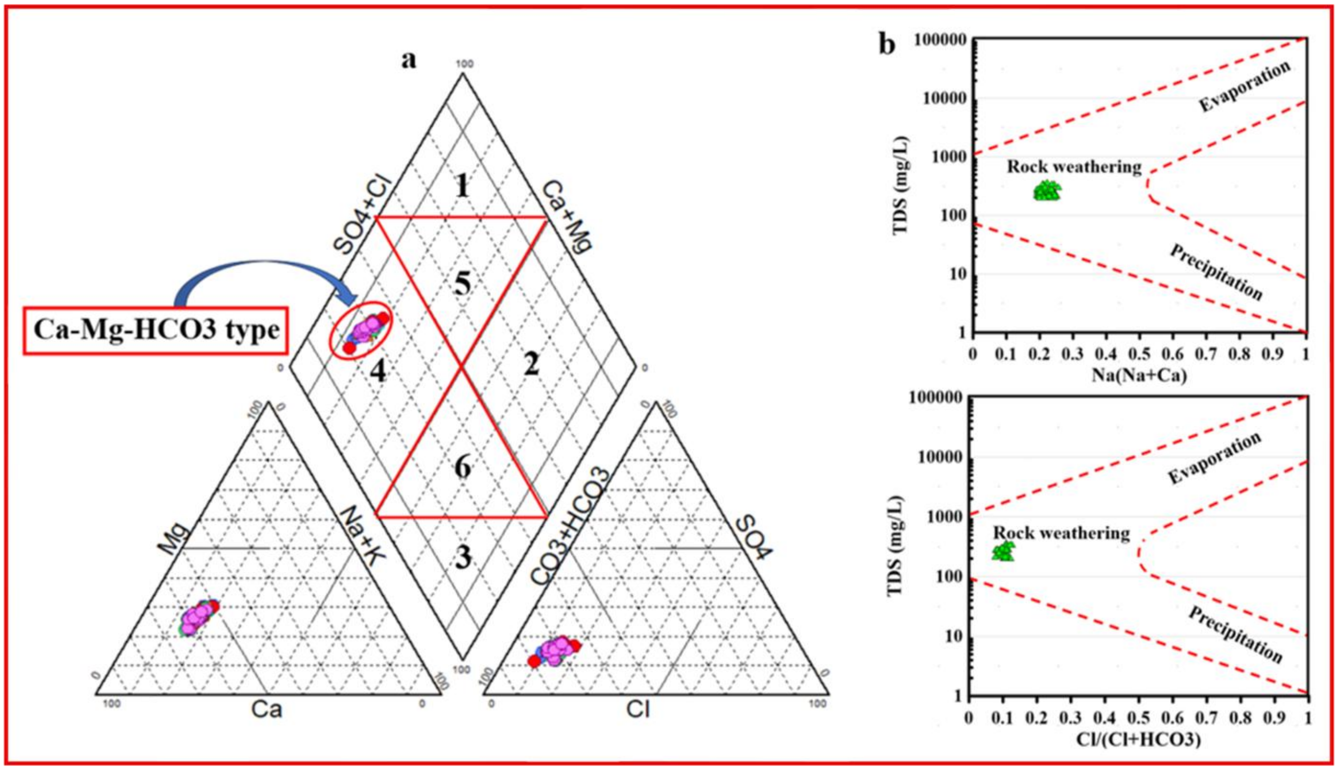


Figure 23 Surface water facies according to Piper diagram (a) and geochemical controlling mechanisms according to Gibbs diagram (b)

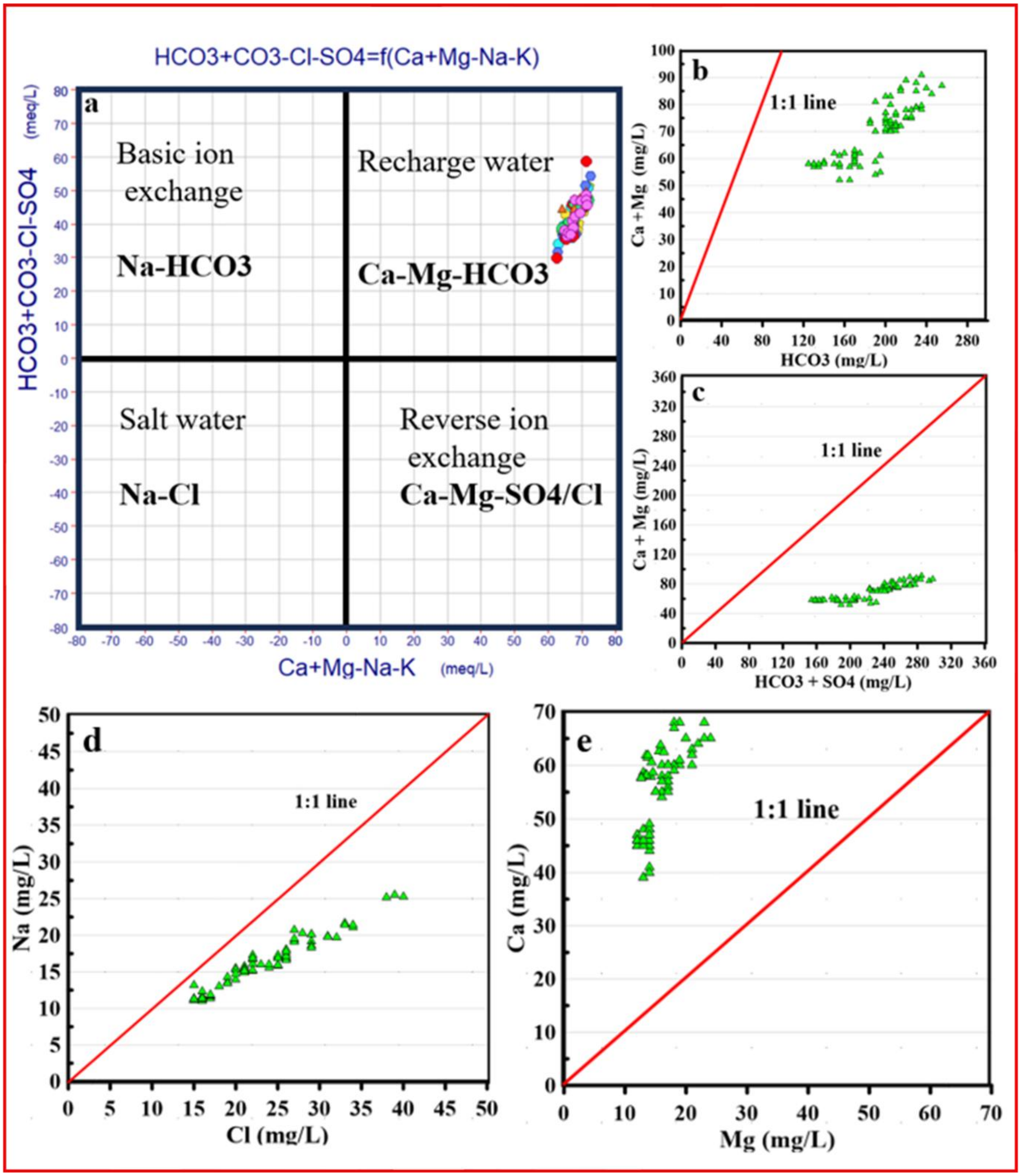


Figure 24 Relationships between the main cations and anions in the sample water using stoichiometry: (a) Ca^{2+}+Mg^{2+}-(Na^{+}+K^{+}) vs. (HCO_3^{-}(SO_4^{2-}+Cl^{-}), (b) Ca^{2+}+Mg^{2+} vs. HCO_3^{-}, (c) Ca^{2+}+Mg^{2+}vs. HCO_3^{-}+SO_4^{2-}, (d) Na^{+} vs Cl^{-}, (e) Ca^{2+} vs Mg^{2+}

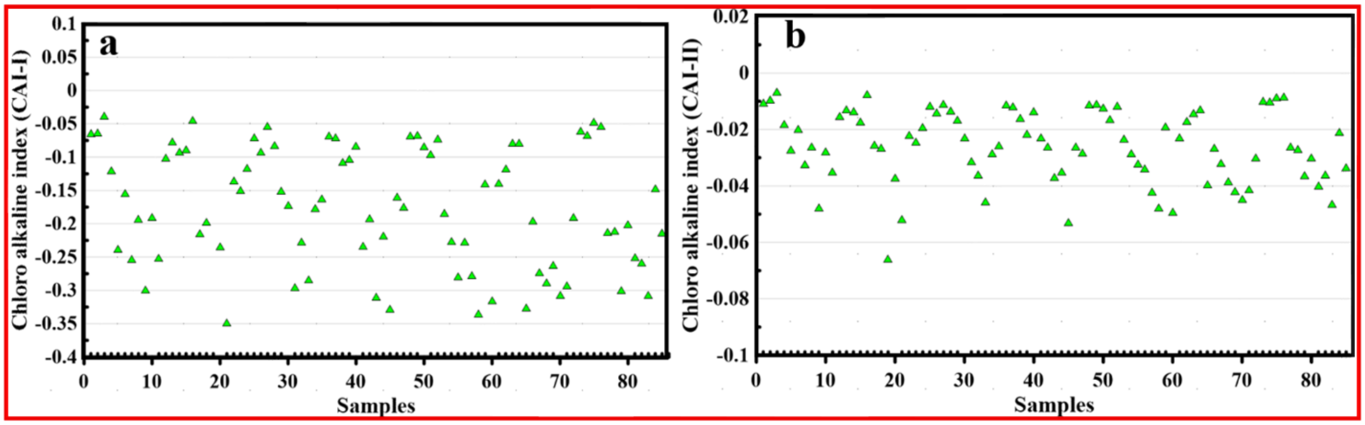


Figure 25 Relationships between Samples vs. CAI-I, and (f) Samples vs. CAI-II.

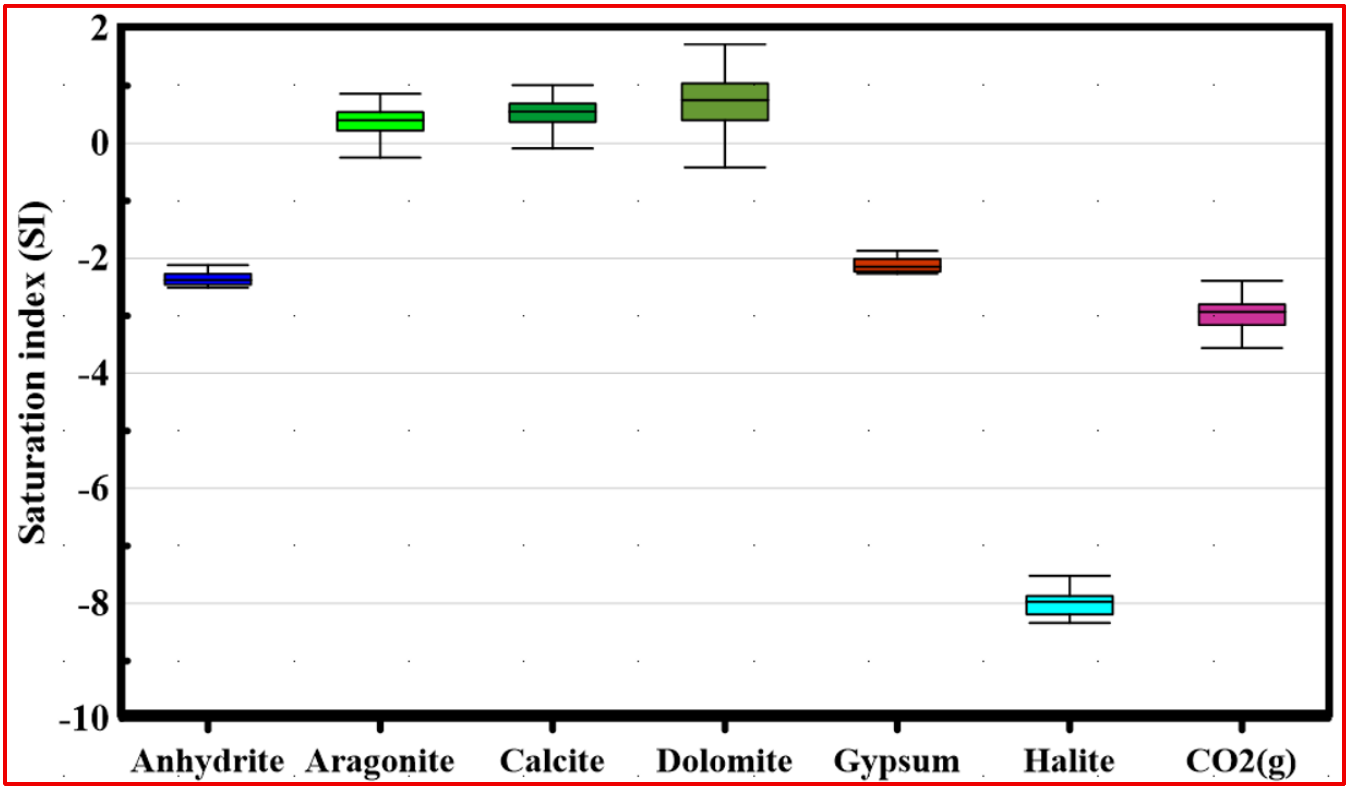


Figure 26 Mineral saturation state showing the ability of precipitation and dissolution of minerals in Danube River

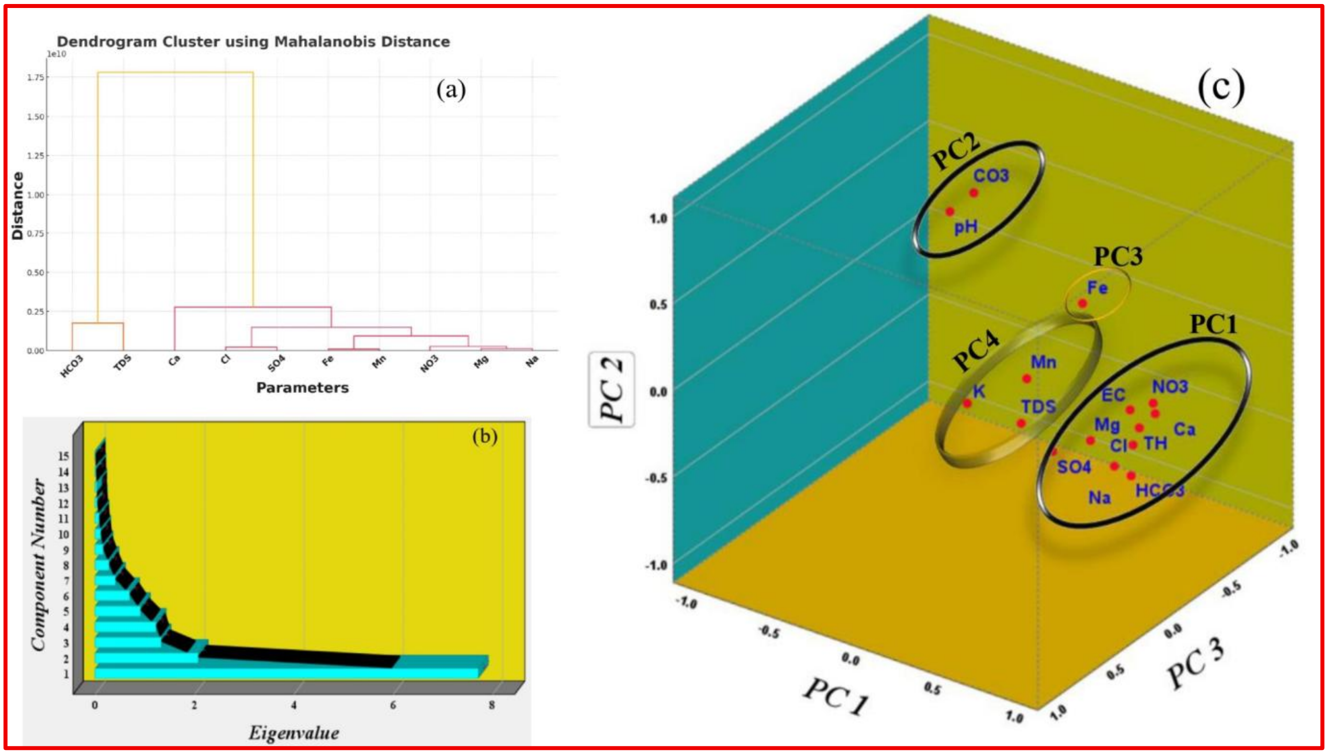


Figure 27 Multivariate statistical analysis: (a) Cluster dendrogram for variables; (b) Scree plot and (c) Principal coordinate analysis scores for PC1 vs. PC2 vs. PC3.

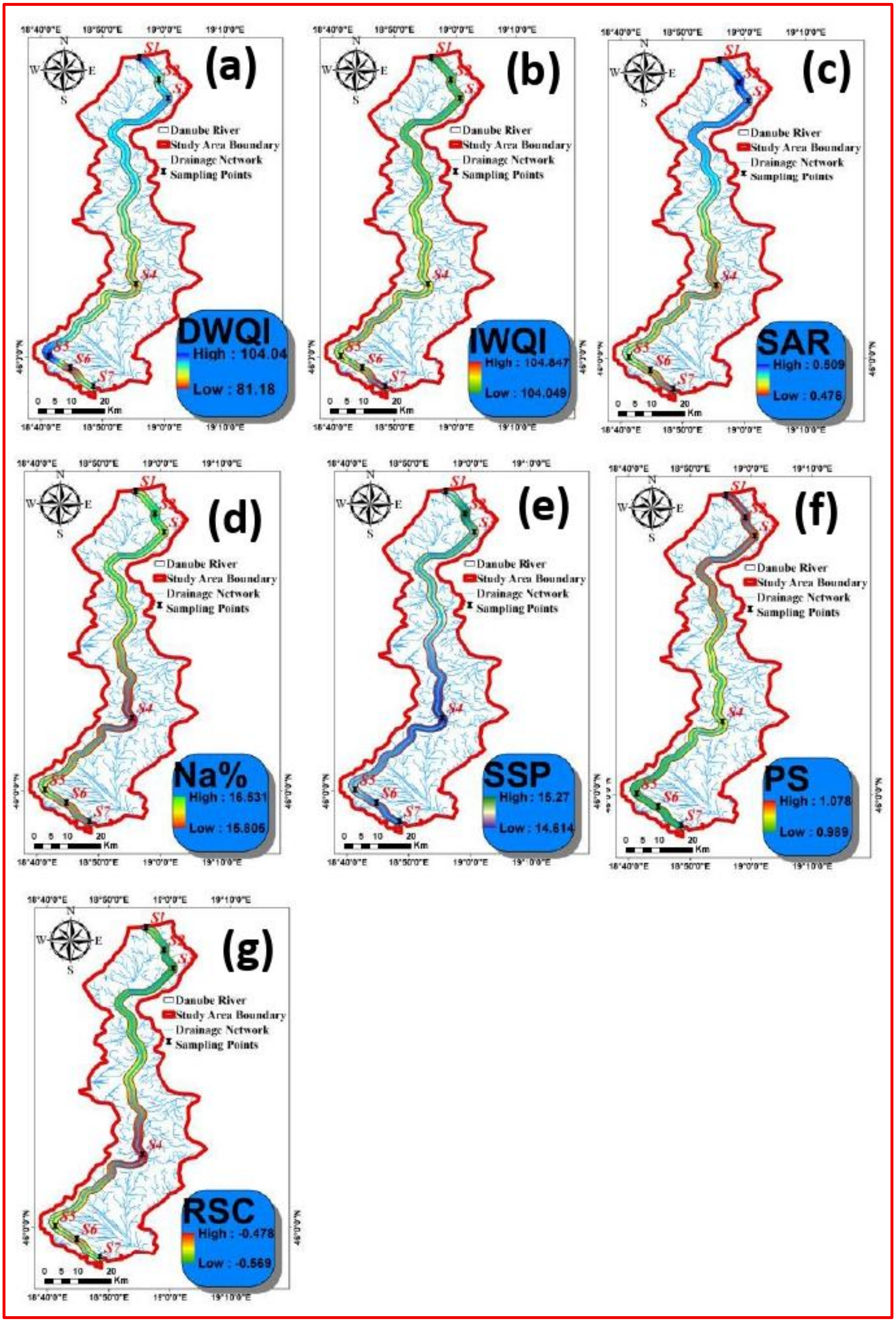

(a)
DWQI
High : 104.04
Low : 81.18
(b)
IWQI
High : 104.847
Low : 104.049
(c)
SAR
High : 0.509
Low : 0.478
(d)
Na%
High : 16.531
Low : 15.605
(e)
SSP
High : 15.27
Low : 14.614
(f)
PS
High : 1.078
Low : 0.989
(g)
RSC
High : -0.478
Low : -0.569
Danube River
Study Area Boundary
Drainage Network
Sampling Points
S1
S2
S3
S4
S5
S6
S7
N
S
E
W
0 5 10 20 Km
18°40'0"E
18°50'0"E
19°0'0"E
19°10'0"E
46°0'0"N

Figure 28 The spatial variation maps of the WQI and IWQIs for the lower Danube river basin: (a) DWQI (b) IWQI, (c) SAR, (d) Na %, (e) SSP, (f) PS, and (g) RSC. Values were computed using equations summarized in Appendix Table 3

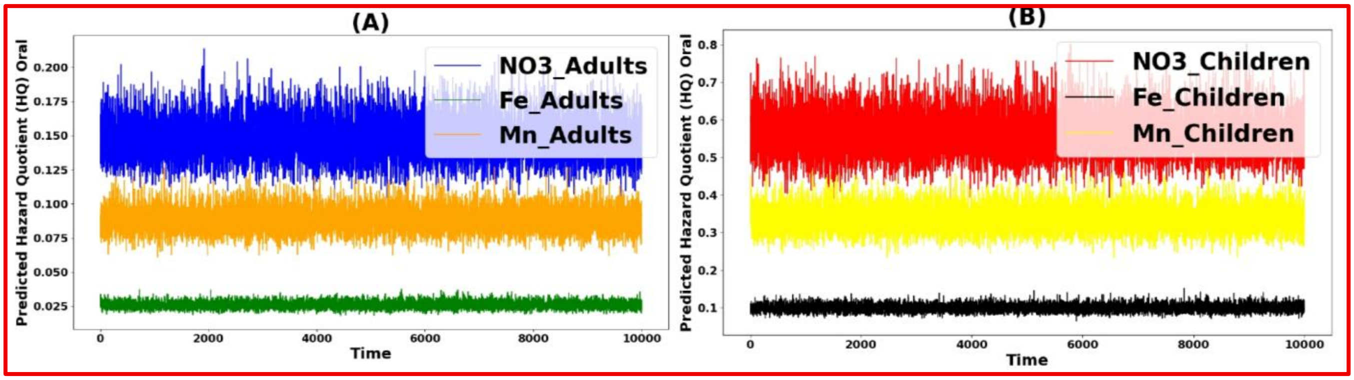
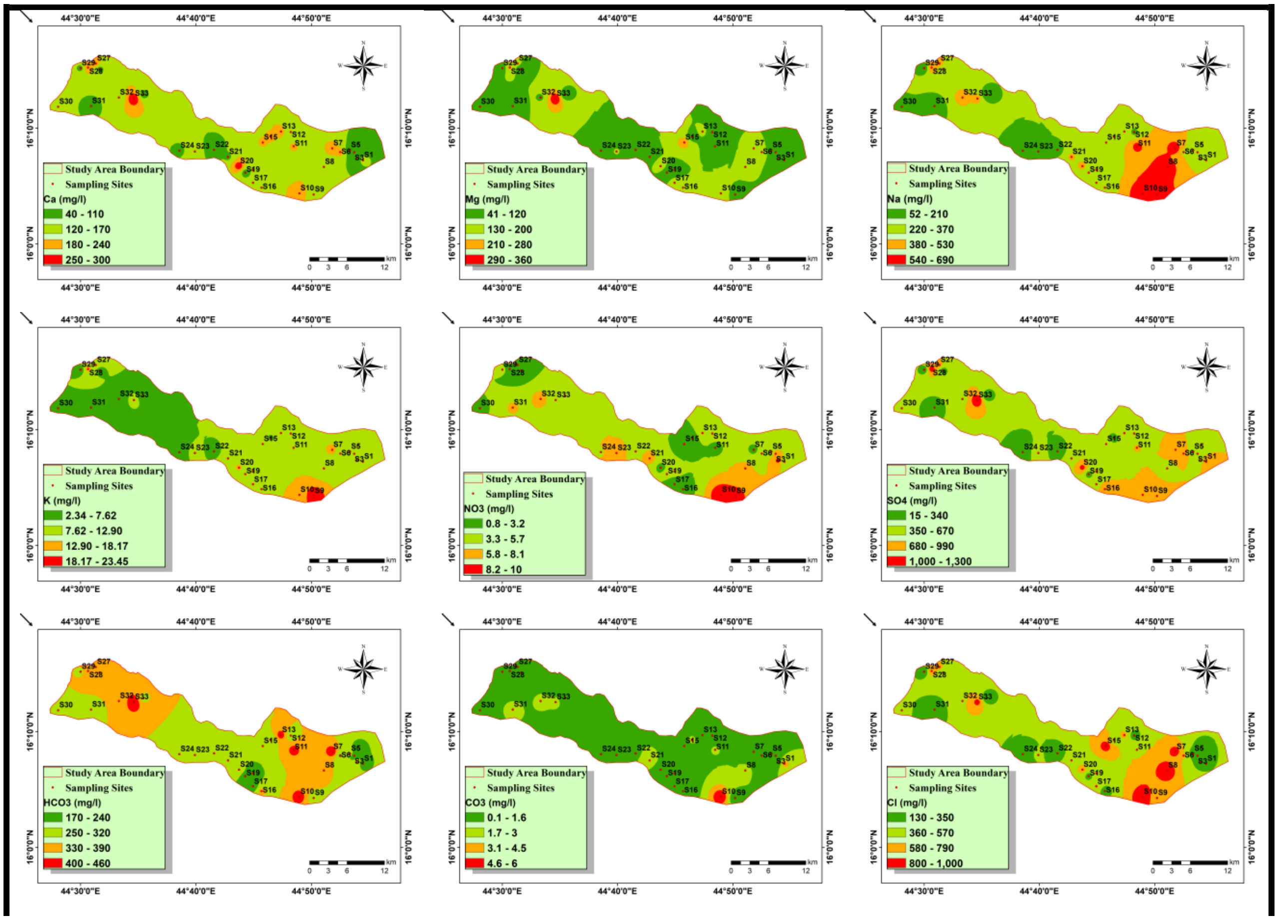


Figure 29 Predicted oral hazard quotient: (A) adults, (B) children



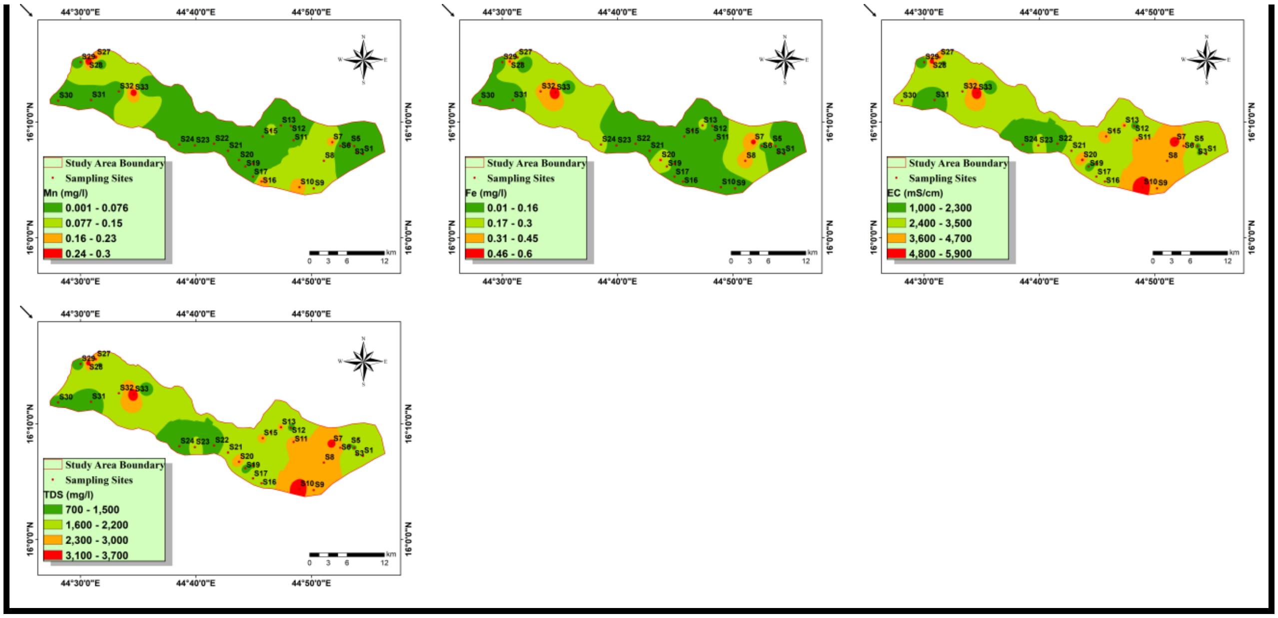


Figure 30 Distribution map of the physicochemical parameters in Al-jawf basin, Yemen

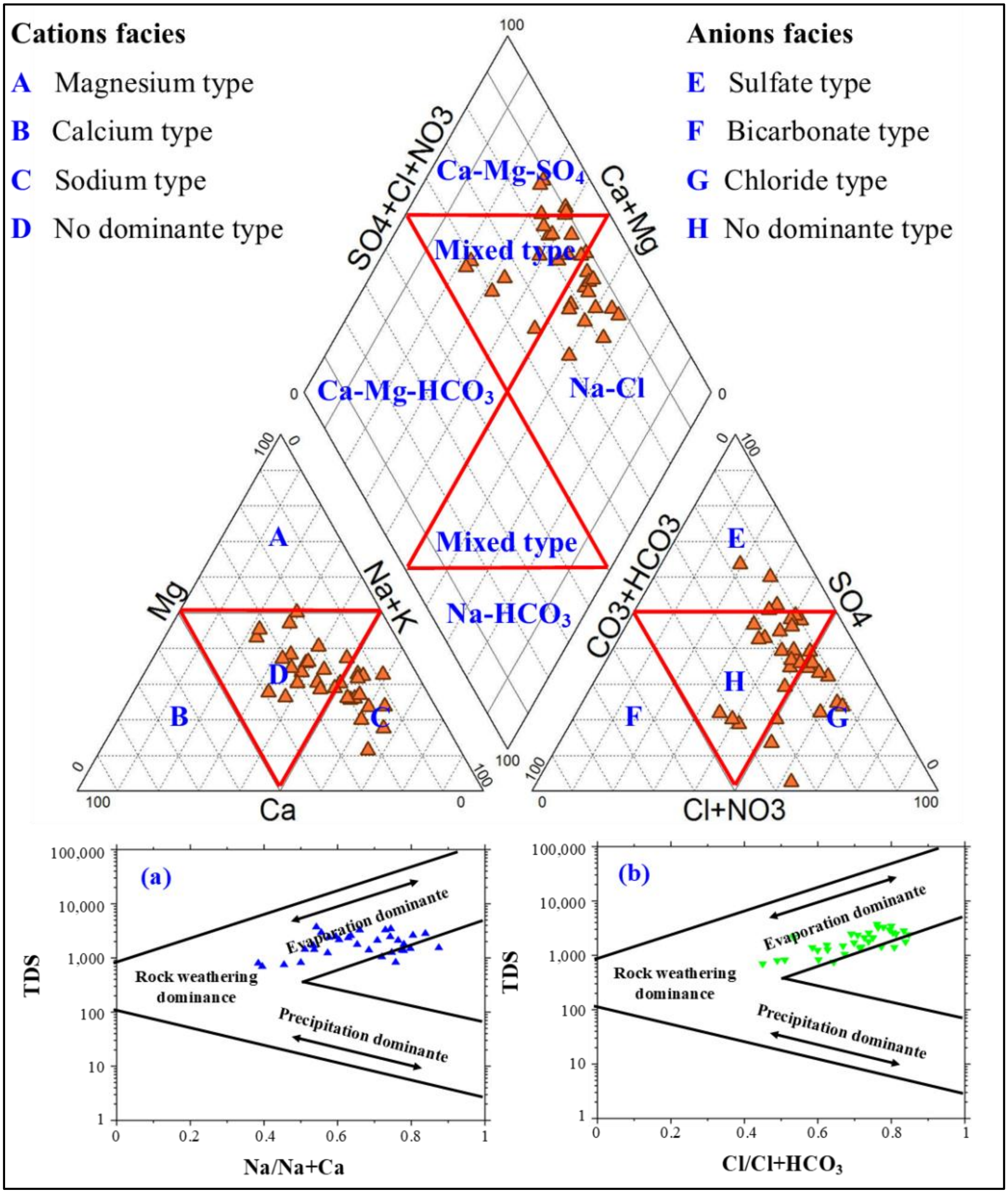


Figure 31 Piper diagram and Gibbs diagram showing the groundwater evolution

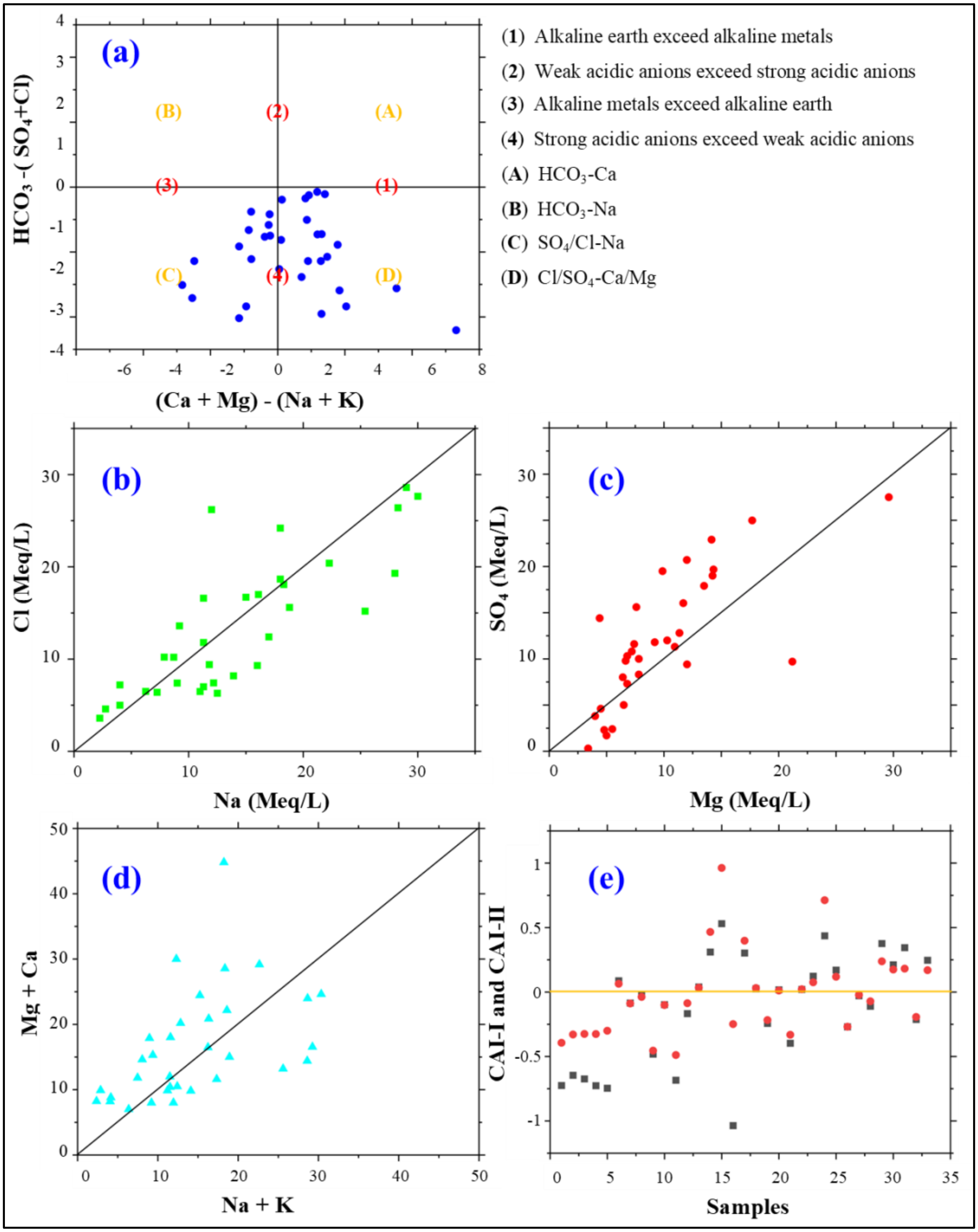

(a)
HCO$_3$ -(SO$_4$+Cl)
(Ca + Mg) - (Na + K)
(B)
(2)
(A)
(3)
(1)
(C)
(4)
(D)
(1) Alkaline earth exceed alkaline metals
(2) Weak acidic anions exceed strong acidic anions
(3) Alkaline metals exceed alkaline earth
(4) Strong acidic anions exceed weak acidic anions
(A) HCO$_3$-Ca
(B) HCO$_3$-Na
(C) SO$_4$/Cl-Na
(D) Cl/SO$_4$-Ca/Mg
(b)
Cl (Meq/L)
Na (Meq/L)
(c)
SO$_4$ (Meq/L)
Mg (Meq/L)
(d)
Mg + Ca
Na + K
(e)
CAI-I and CAI-II
Samples

Figure 32 Relationships between the main investigated parameters (a) $Ca^{2+}+Mg^{2+}-(Na^{+}+K^{+})$ vs. ($HCO_3^{-}(SO_4^{2-}+Cl)$, (b) Cl^{-} vs Na^{+}, (c) SO_4^{2-} vs. Mg^{2+}, (d) $Mg^{2+}+Ca^{2+}$ vs. $Na^{+}K^{+}$, (e) Samples vs. CAI-I and CAI-II.

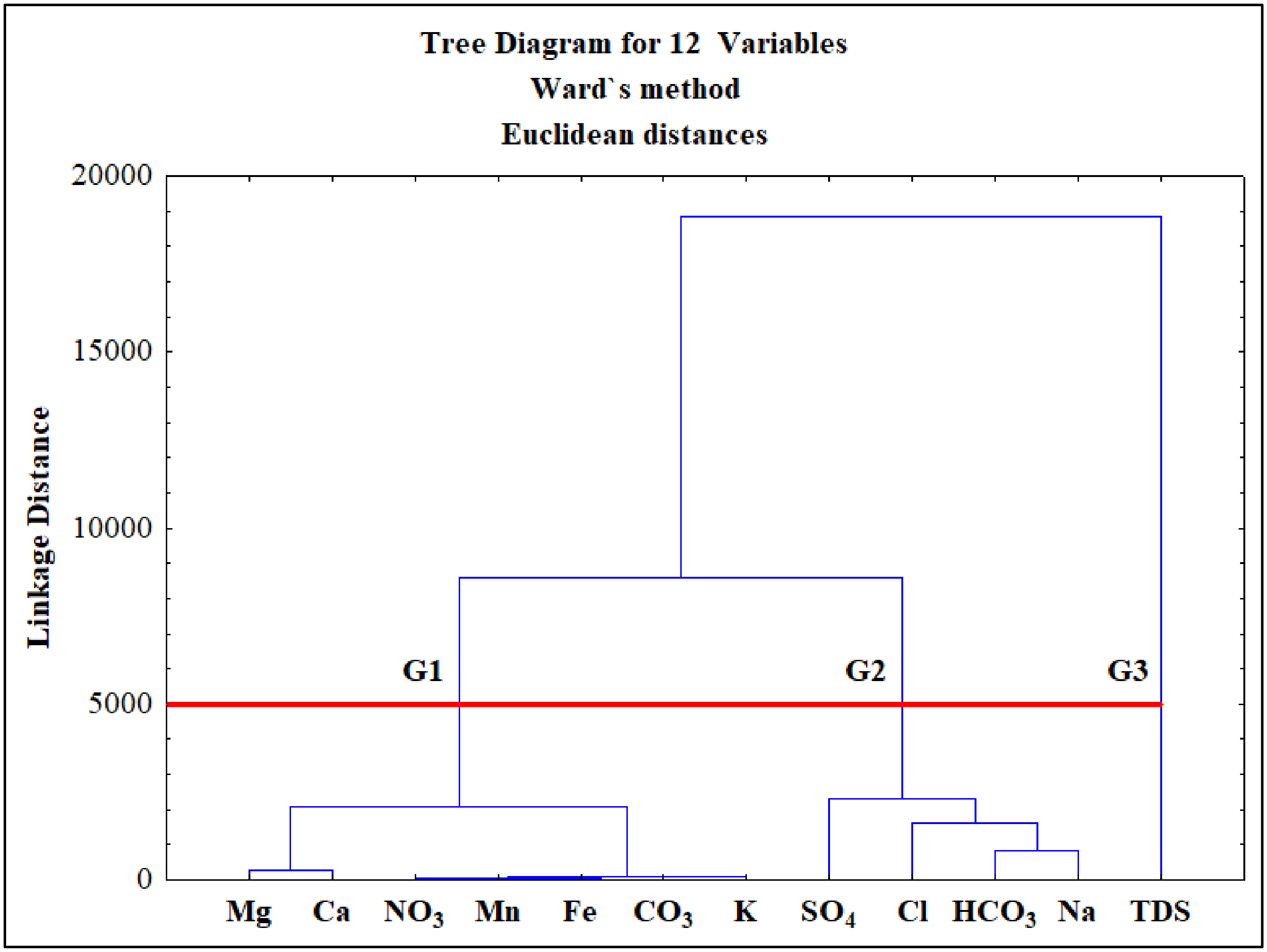


Figure 33 The three groups extracted from cluster dendrogram for variables

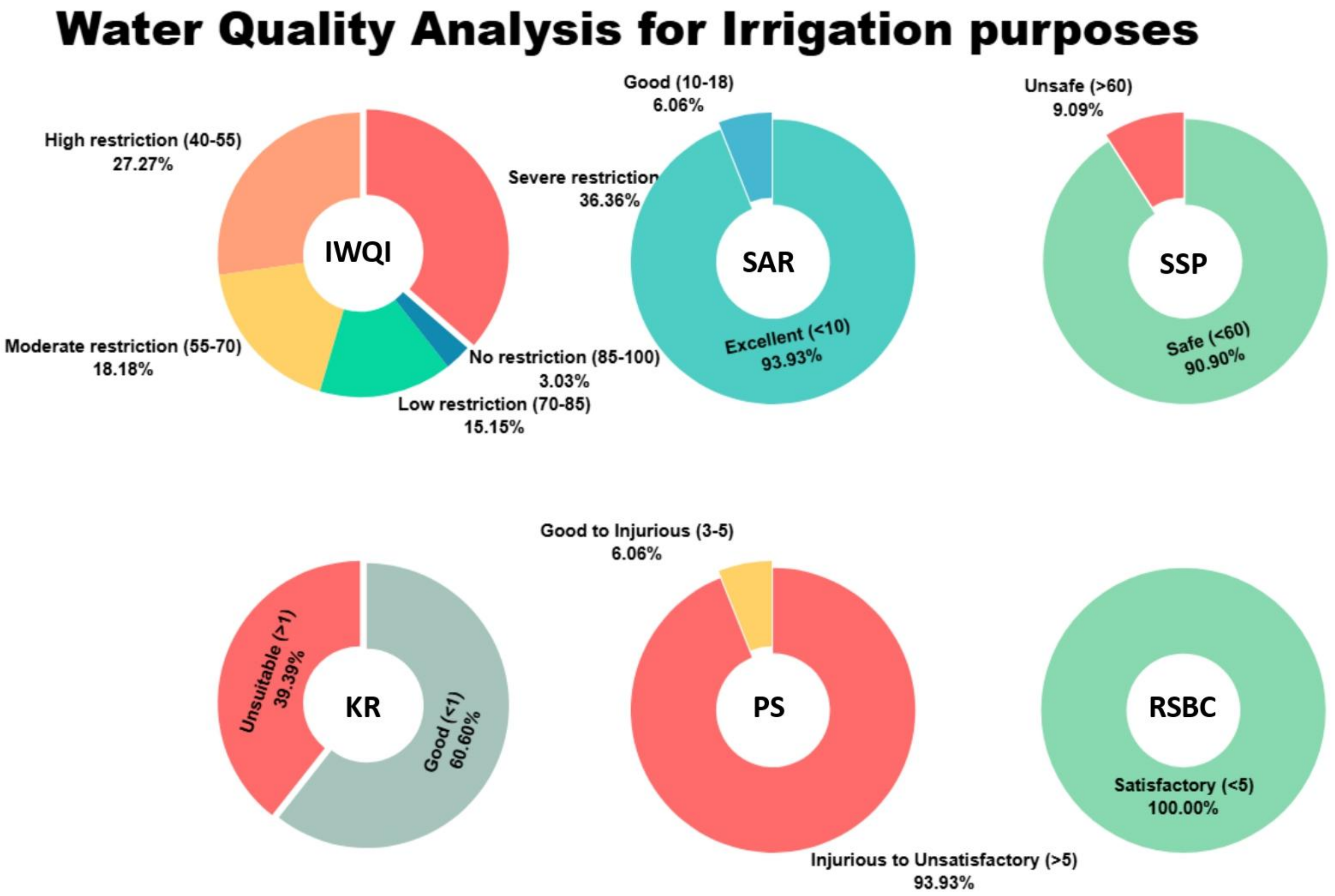


Figure 34 Classification of irrigation water quality based on six indices.

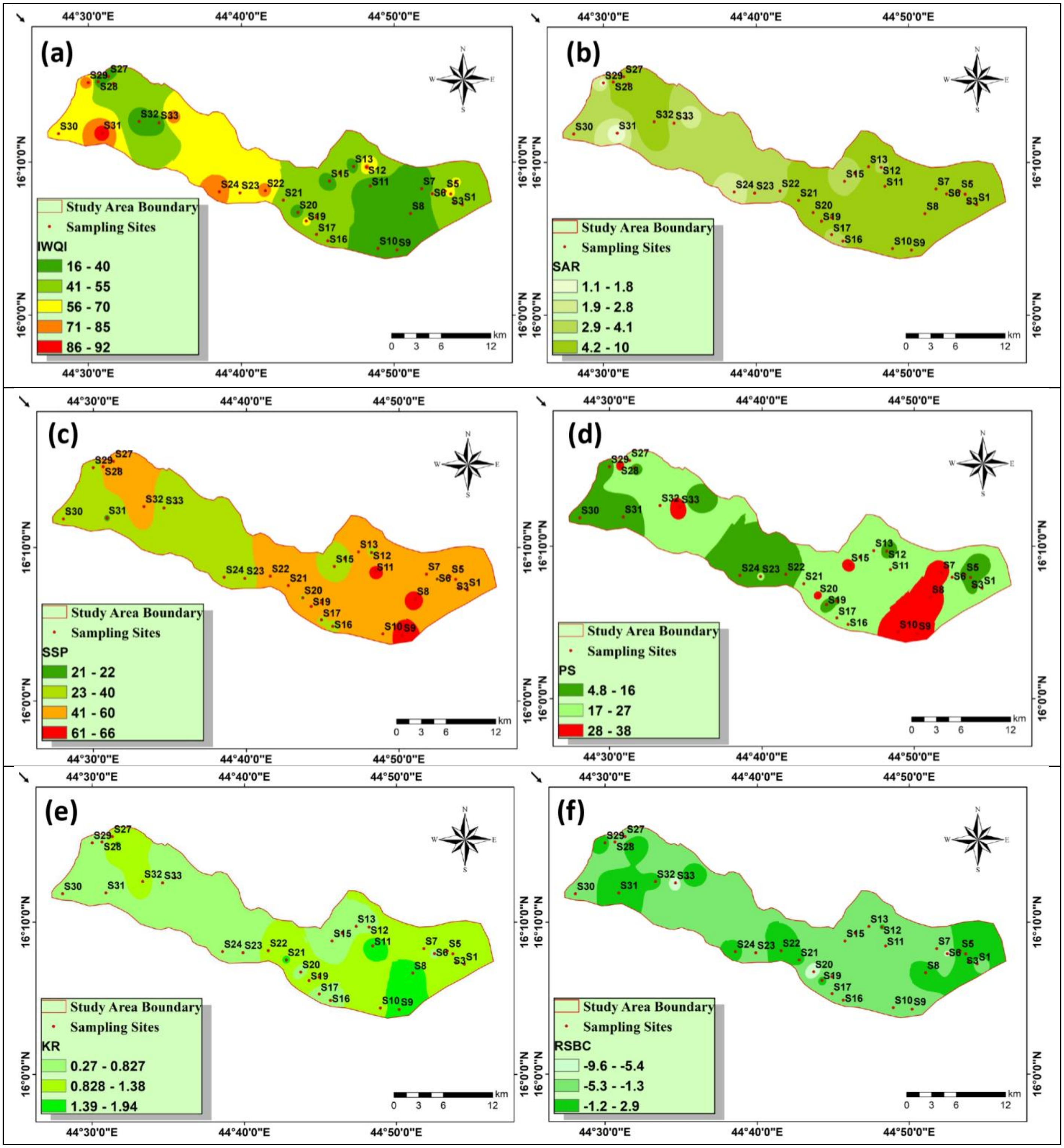


Figure 35 GIS maps of various IWQIs in the Al-Jawf basin: (a) IWQI, (b) SAR, (c) SSP, (d) KR, (e) PS, and (f) RSC.

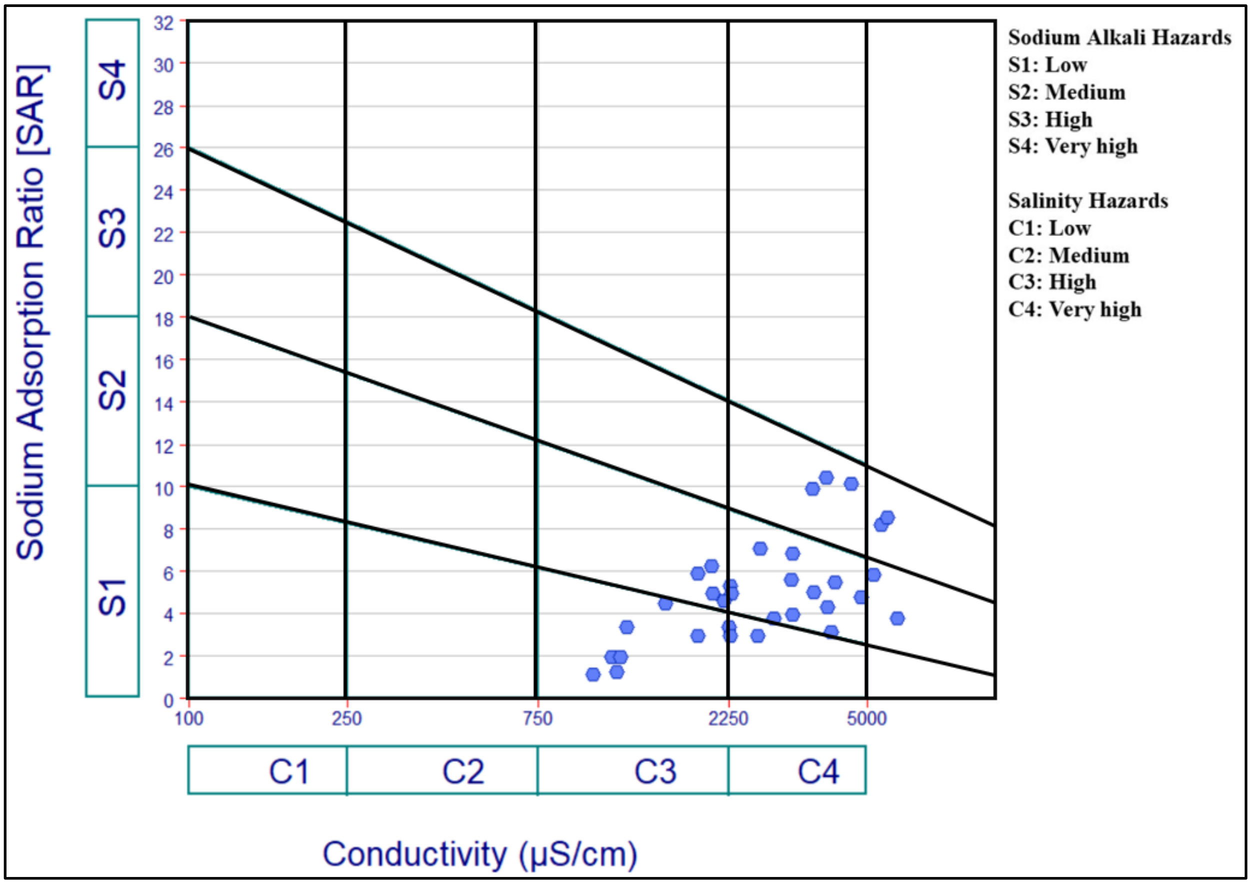
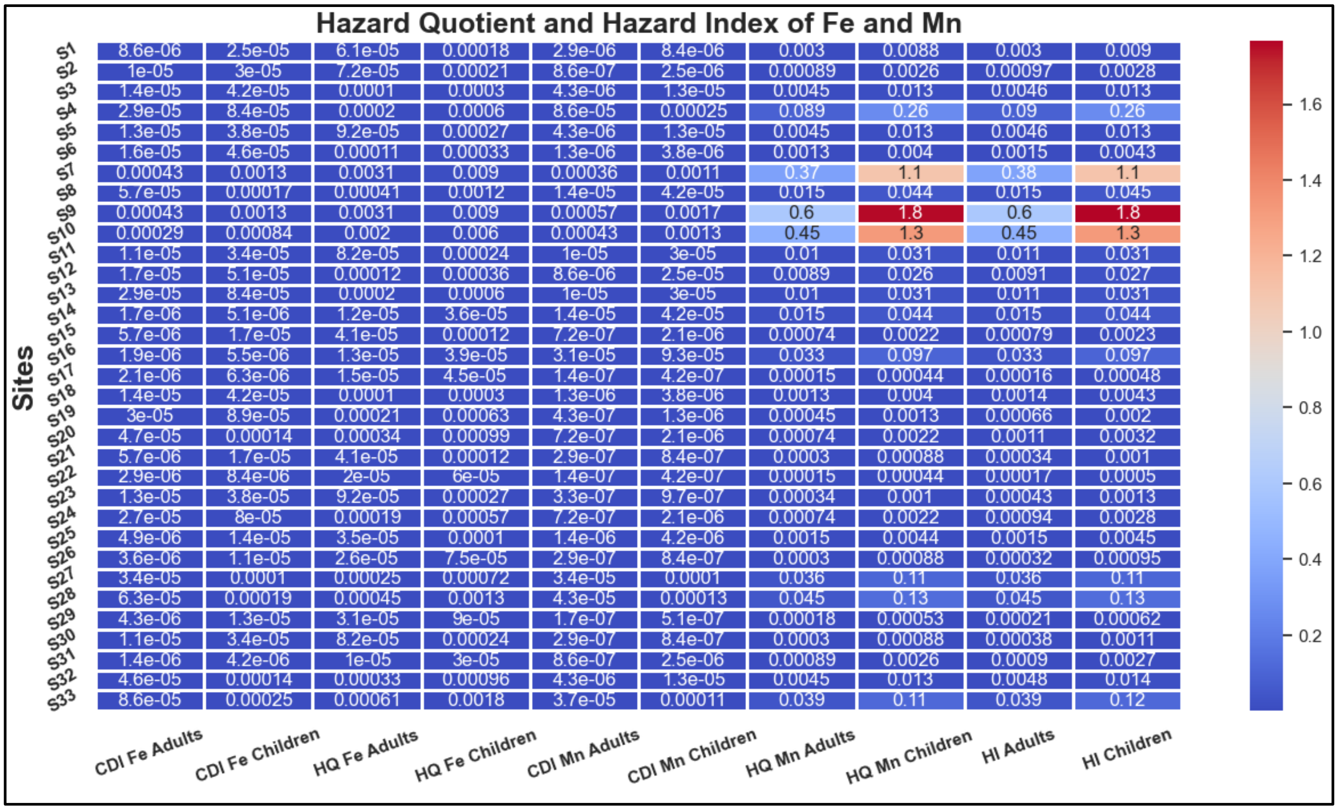


Figure 36 USSL diagram for irrigation purposes



Sites	CDI Fe Adults	CDI Fe Children	HQ Fe Adults	HQ Fe Children	CDI Mn Adults	CDI Mn Children	HQ Mn Adults	HQ Mn Children	HI Adults	HI Children
S1	8.6e-06	2.5e-05	6.1e-05	0.00018	2.9e-06	8.4e-06	0.003	0.0088	0.003	0.009
S2	1e-05	3e-05	7.2e-05	0.00021	8.6e-07	2.5e-06	0.00089	0.0026	0.00097	0.0028
S3	1.4e-05	4.2e-05	0.0001	0.0003	4.3e-06	1.3e-05	0.0045	0.013	0.0046	0.013
S4	2.9e-05	8.4e-05	0.0002	0.0006	8.6e-05	0.00025	0.089	0.26	0.09	0.26
S5	1.3e-05	3.8e-05	9.2e-05	0.00027	4.3e-06	1.3e-05	0.0045	0.013	0.0046	0.013
S6	1.6e-05	4.6e-05	0.00011	0.00033	1.3e-06	3.8e-06	0.0013	0.004	0.0015	0.0043
S7	0.00043	0.0013	0.0031	0.009	0.00036	0.0011	0.37	1.1	0.38	1.1
S8	5.7e-05	0.00017	0.00041	0.0012	1.4e-05	4.2e-05	0.015	0.044	0.015	0.045
S9	0.00043	0.0013	0.0031	0.009	0.00057	0.0017	0.6	1.8	0.6	1.8
S10	0.00029	0.00084	0.002	0.006	0.00043	0.0013	0.45	1.3	0.45	1.3
S11	1.1e-05	3.4e-05	8.2e-05	0.00024	1e-05	3e-05	0.01	0.031	0.011	0.031
S12	1.7e-05	5.1e-05	0.00012	0.00036	8.6e-06	2.5e-05	0.0089	0.026	0.0091	0.027
S13	2.9e-05	8.4e-05	0.0002	0.0006	1e-05	3e-05	0.01	0.031	0.011	0.031
S14	1.7e-06	5.1e-06	1.2e-05	3.6e-05	1.4e-05	4.2e-05	0.015	0.044	0.015	0.044
S15	5.7e-06	1.7e-05	4.1e-05	0.00012	7.2e-07	2.1e-06	0.00074	0.0022	0.00079	0.0023
S16	1.9e-06	5.5e-06	1.3e-05	3.9e-05	3.1e-05	9.3e-05	0.033	0.097	0.033	0.097
S17	2.1e-06	6.3e-06	1.5e-05	4.5e-05	1.4e-07	4.2e-07	0.00015	0.00044	0.00016	0.00048
S18	1.4e-05	4.2e-05	0.0001	0.0003	1.3e-06	3.8e-06	0.0013	0.004	0.0014	0.0043
S19	3e-05	8.9e-05	0.00021	0.00063	4.3e-07	1.3e-06	0.00045	0.0013	0.00066	0.002
S20	4.7e-05	0.00014	0.00034	0.00099	7.2e-07	2.1e-06	0.00074	0.0022	0.0011	0.0032
S21	5.7e-06	1.7e-05	4.1e-05	0.00012	2.9e-07	8.4e-07	0.0003	0.00088	0.00034	0.001
S22	2.9e-06	8.4e-06	2e-05	6e-05	1.4e-07	4.2e-07	0.00015	0.00044	0.00017	0.0005
S23	1.3e-05	3.8e-05	9.2e-05	0.00027	3.3e-07	9.7e-07	0.00034	0.001	0.00043	0.0013
S24	2.7e-05	8e-05	0.00019	0.00057	7.2e-07	2.1e-06	0.00074	0.0022	0.00094	0.0028
S25	4.9e-06	1.4e-05	3.5e-05	0.0001	1.4e-06	4.2e-06	0.0015	0.0044	0.0015	0.0045
S26	3.6e-06	1.1e-05	2.6e-05	7.5e-05	2.9e-07	8.4e-07	0.0003	0.00088	0.00032	0.00095
S27	3.4e-05	0.0001	0.00025	0.00072	3.4e-05	0.0001	0.036	0.11	0.036	0.11
S28	6.3e-05	0.00019	0.00045	0.0013	4.3e-05	0.00013	0.045	0.13	0.045	0.13
S29	4.3e-06	1.3e-05	3.1e-05	9e-05	1.7e-07	5.1e-07	0.00018	0.00053	0.00021	0.00062
S30	1.1e-05	3.4e-05	8.2e-05	0.00024	2.9e-07	8.4e-07	0.0003	0.00088	0.00038	0.0011
S31	1.4e-06	4.2e-06	1e-05	3e-05	8.6e-07	2.5e-06	0.00089	0.0026	0.0009	0.0027
S32	4.6e-05	0.00014	0.00033	0.00096	4.3e-06	1.3e-05	0.0045	0.013	0.0048	0.014
S33	8.6e-05	0.00025	0.00061	0.0018	3.7e-05	0.00011	0.039	0.11	0.039	0.12

Figure 37 Heat map showing the health risk indices including CDI, HQ, and HI for adult and children through dermal contact.

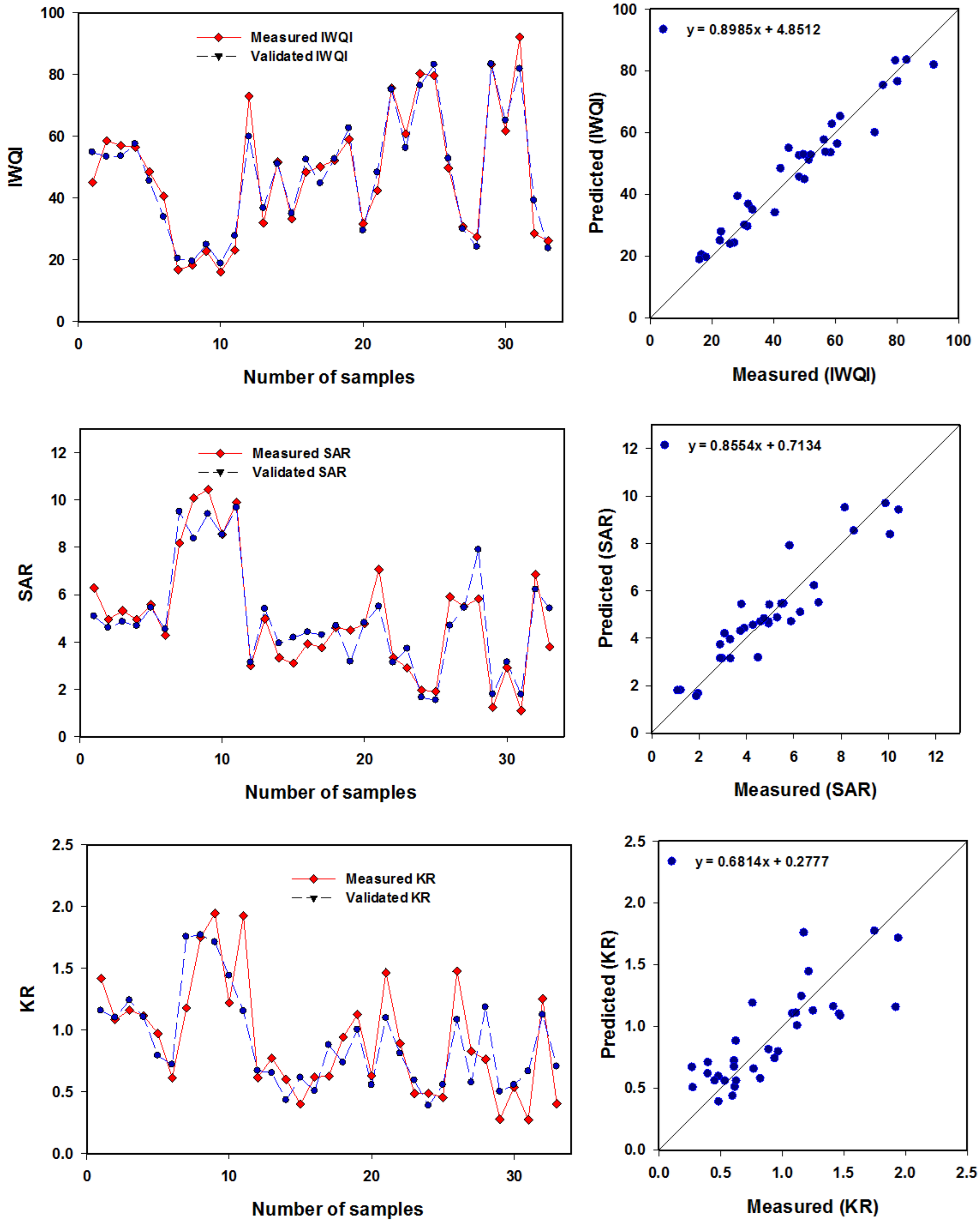
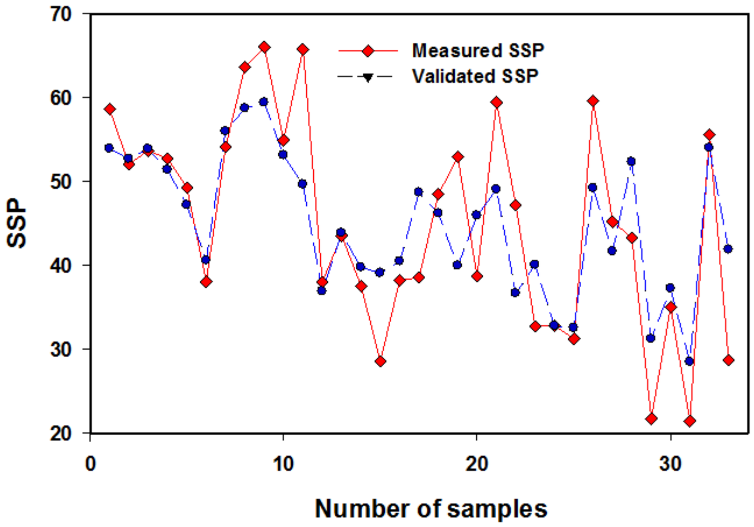
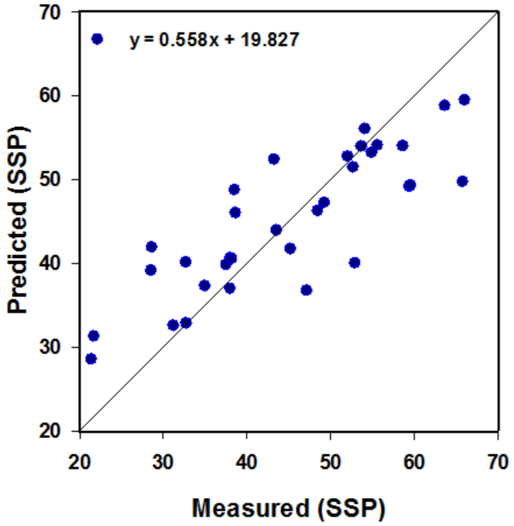
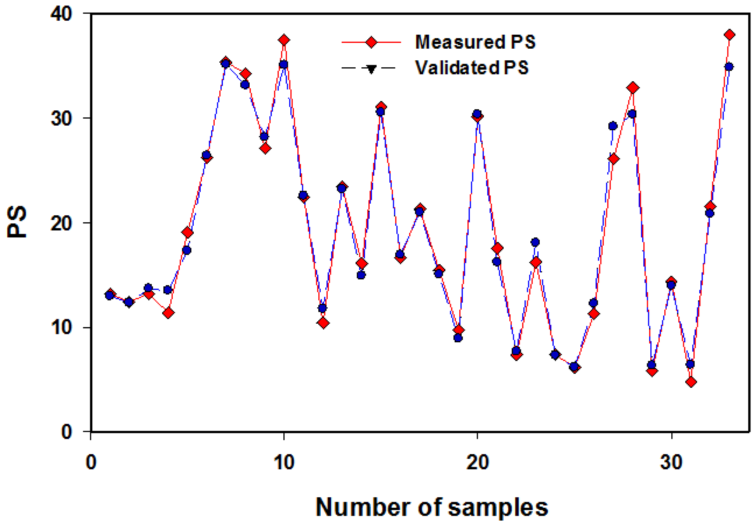
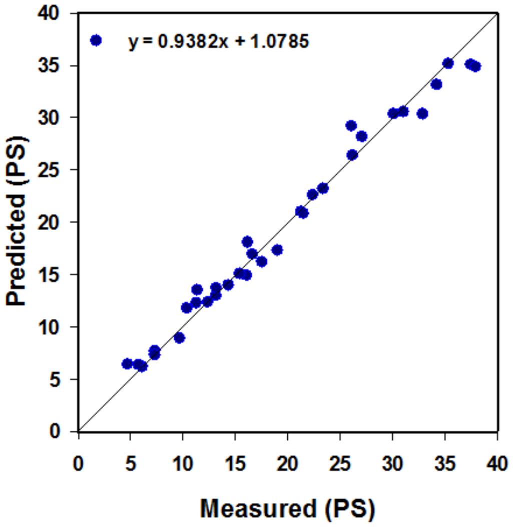


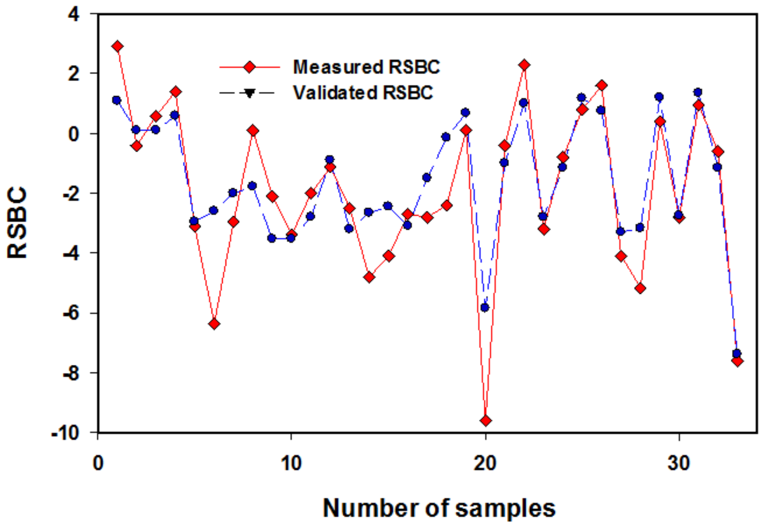
Figure 38 Comparison between measured series and predicted series for IWQI, SAR, KR using the developed RF models


SSP
70
60
50
40
30
20
0
10
20
30
Number of samples
Measured SSP
Validated SSP


y = 0.558x + 19.827
Predicted (SSP)
70
60
50
40
30
20
30
40
50
60
70
Measured (SSP)


PS
40
30
20
10
0
0
10
20
30
Number of samples
Measured PS
Validated PS


y = 0.9382x + 1.0785
Predicted (PS)
40
35
30
25
20
15
10
5
0
0
5
10
15
20
25
30
35
40
Measured (PS)


RSBC
4
2
0
-2
-4
-6
-8
-10
0
10
20
30
Number of samples
Measured RSBC
Validated RSBC

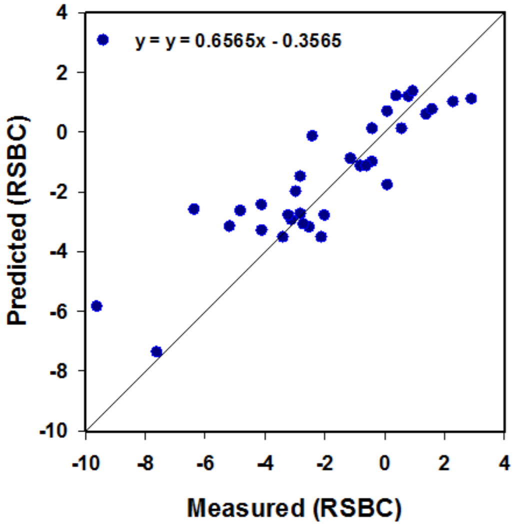

y = y = 0.6565x - 0.3565
Predicted (RSBC)
4
2
0
-2
-4
-6
-8
-10
-10
-8
-6
-4
-2
0
2
4
Measured (RSBC)

Figure 39 Comparison between measured series and predicted series for SSP, PS, RSBC using the developed RF models.

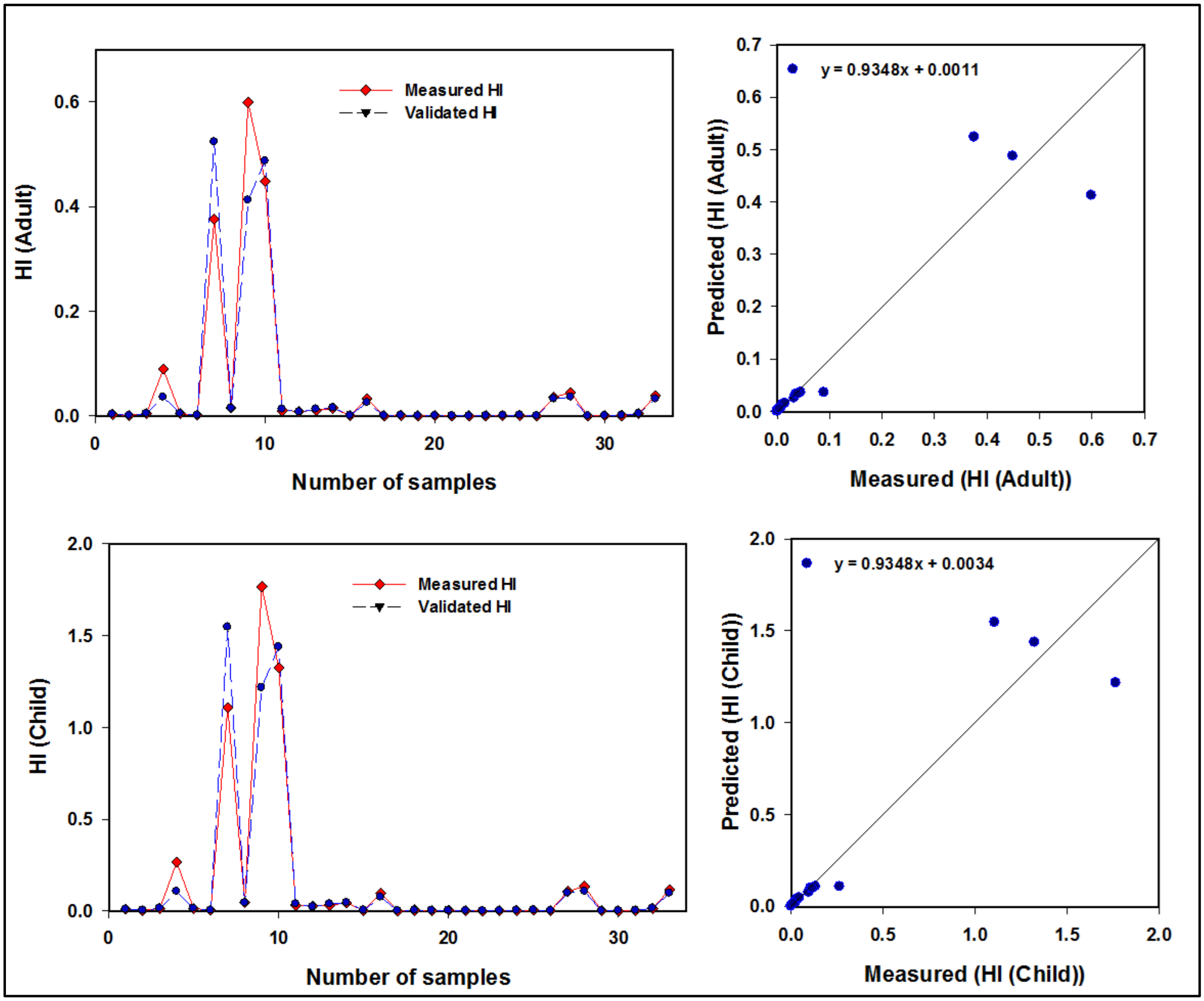


Figure 40 Comparison between measured series and predicted series for HI (Adult), and HI (Child) using the developed RF models.

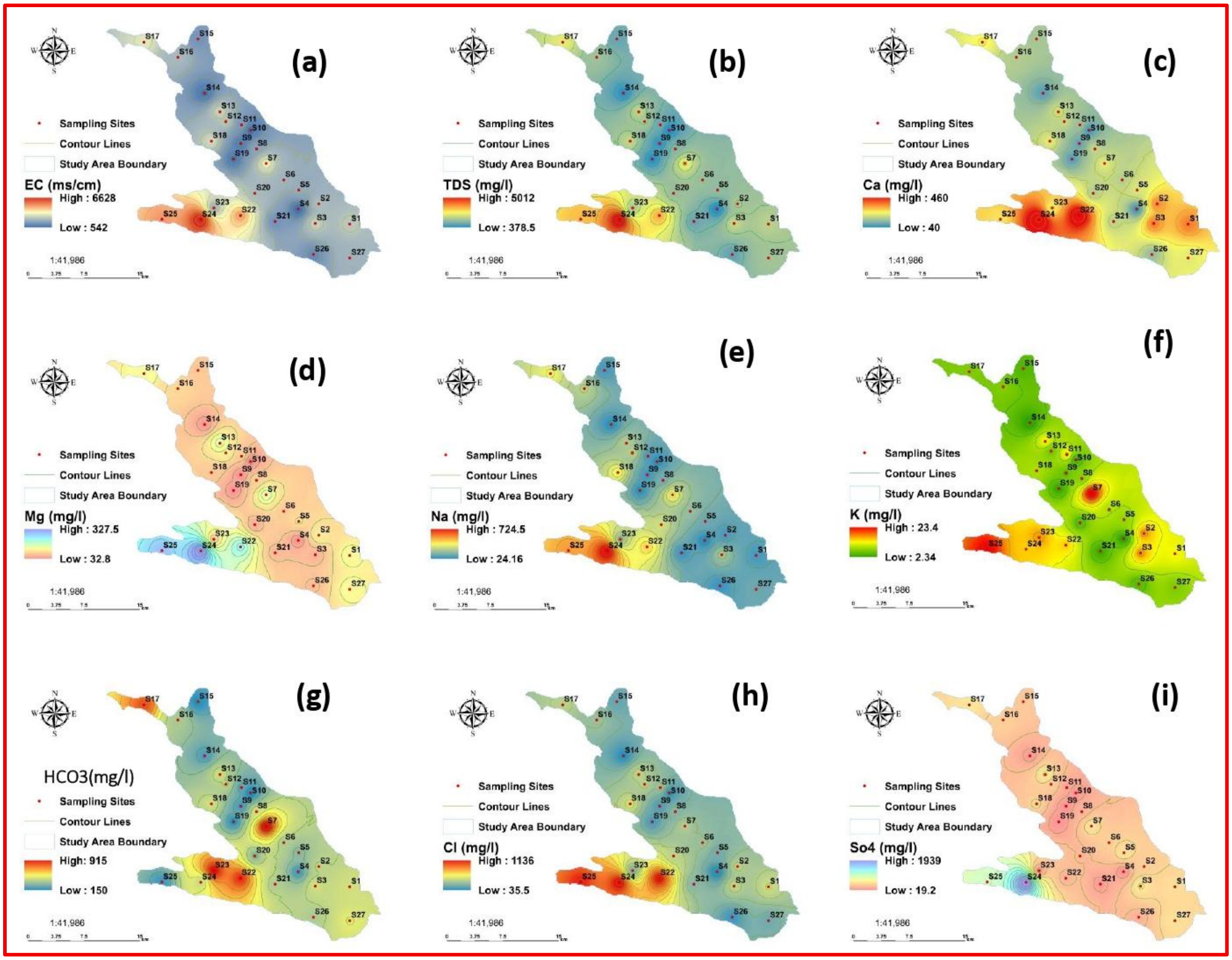


Figure 41 Distribution map of the physicochemical parameters: (a) EC, (b)TDS, (c) Ca^{2+}, (d) Mg^{2+}, (e) Na^{+}, (f) K^{+}, (g) HCO_3^{-}, (h) Cl^{-}, and (i) SO_4^{2-}.

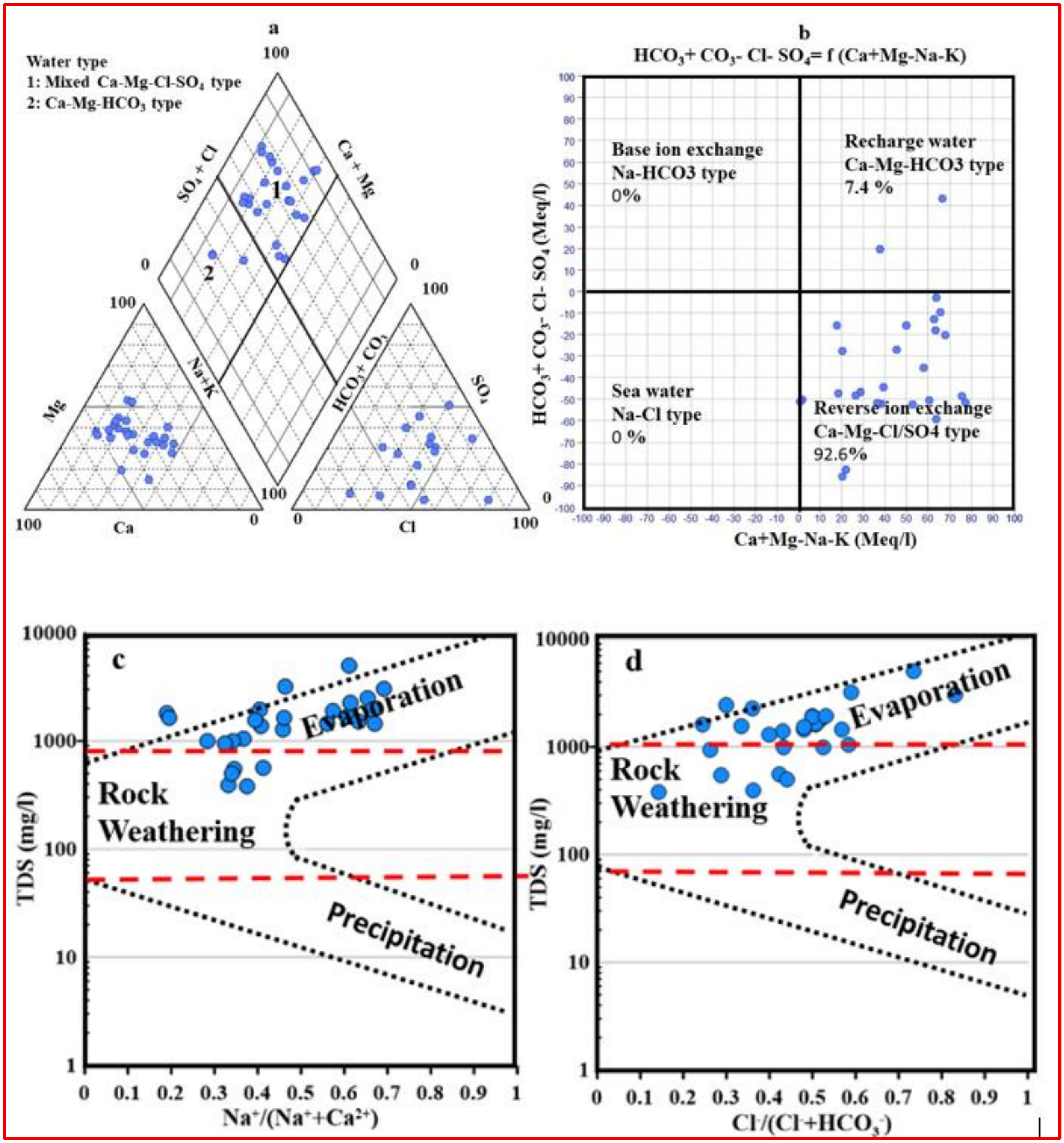


Figure 42 Plotting water samples on piper diagram (a), Chadha diagram (b), (c,d) Gibbs diagram.

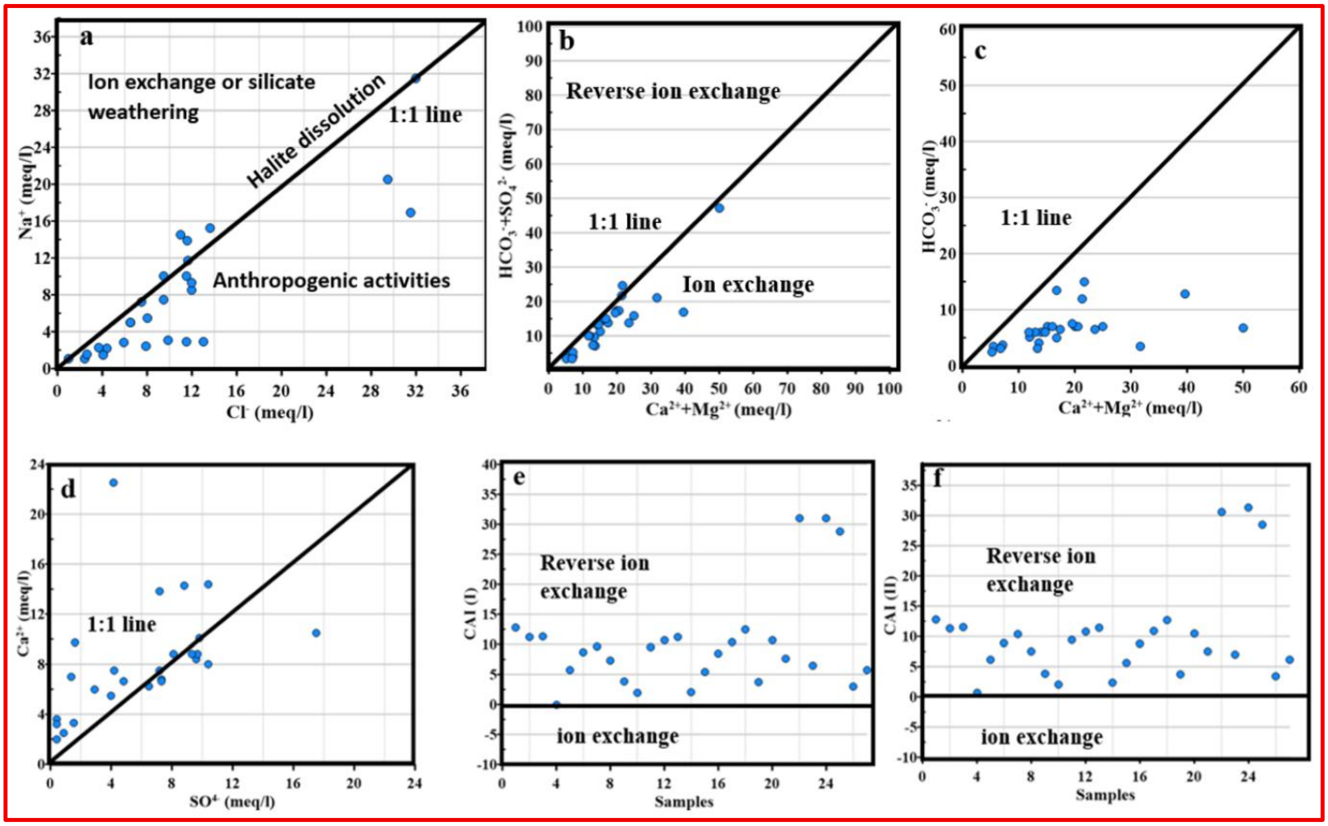


Figure 43 Relationships between the main cations and anions in the sample water using stoichiometry: (a) Na^+ vs. Cl^-, (b) $Ca^{2+} + Mg^{2+}$ vs. $HCO_3^- + SO_4^{2-}$, (c) $Ca^{2+} + Mg^{2+}$ vs. HCO_3^-, (d) Ca^{2+} vs. SO_4^{2-}, (e) Samples vs. CAI-I, and (f) Samples vs. CAI-II.

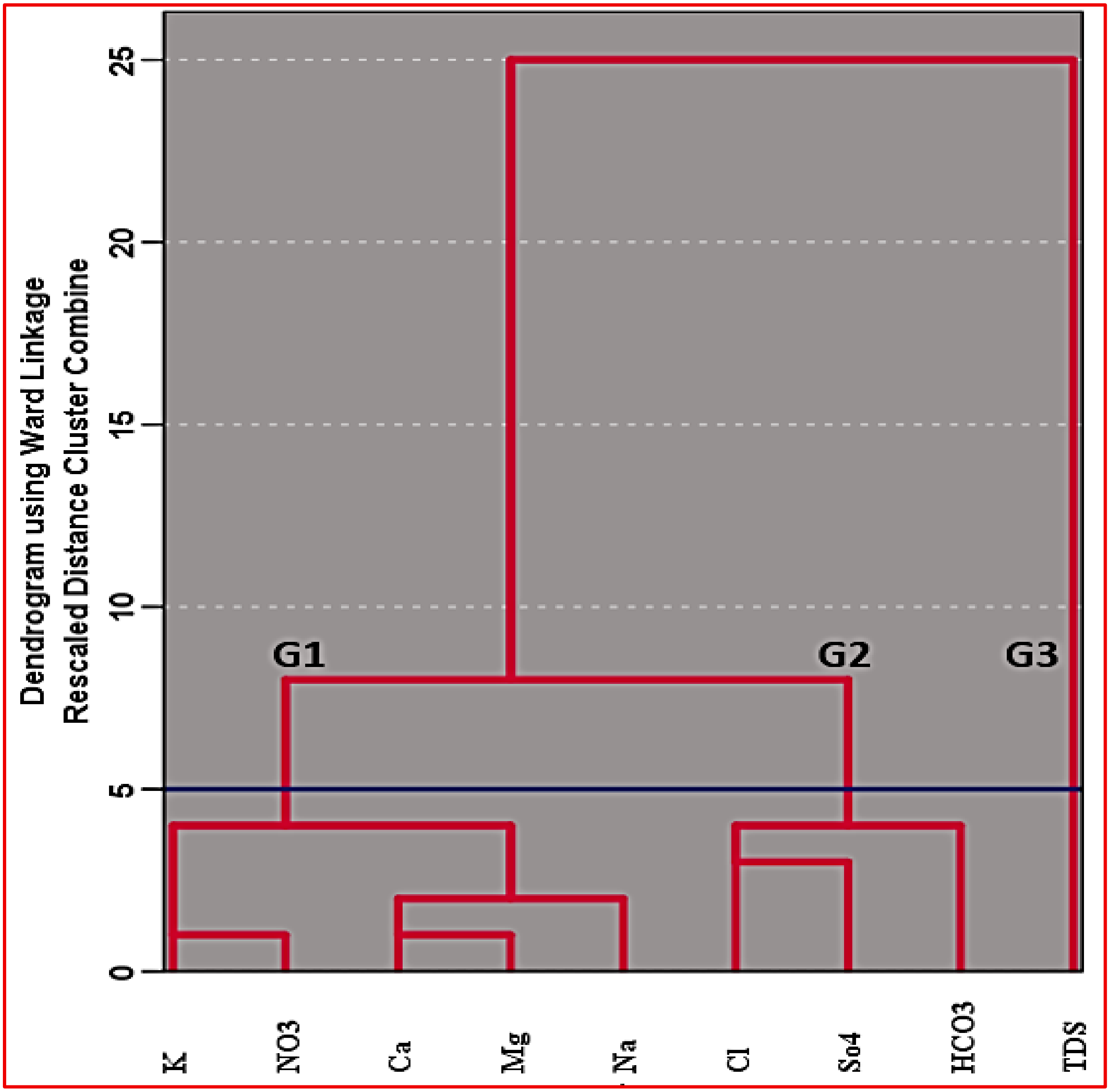


Figure 44 Cluster dendrogram for variables.

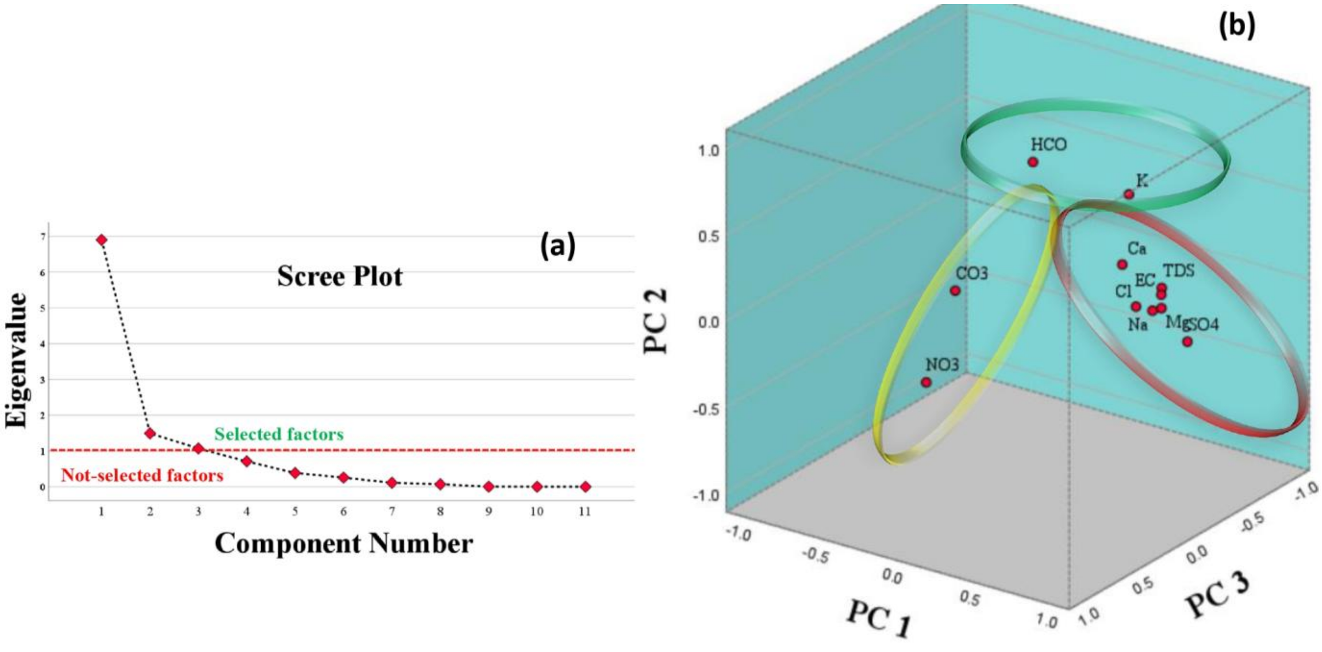


Figure 45 Multivariate statistical analysis: (a), Scree plot and (b) PCA scores for F2 vs F1vs F3.

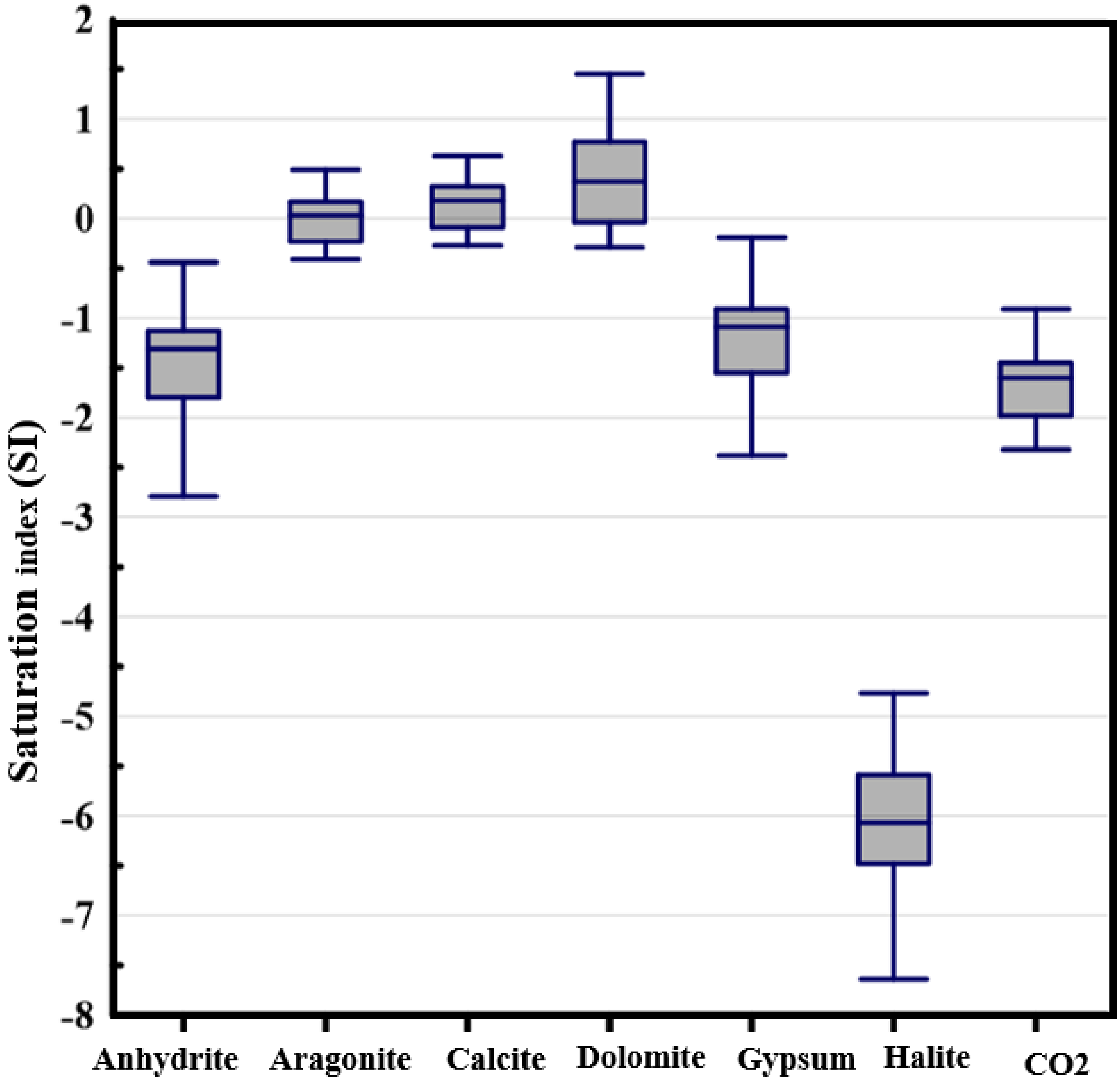


Figure 46 Box plot of the SI results for the quaternary aquifer in the study area.

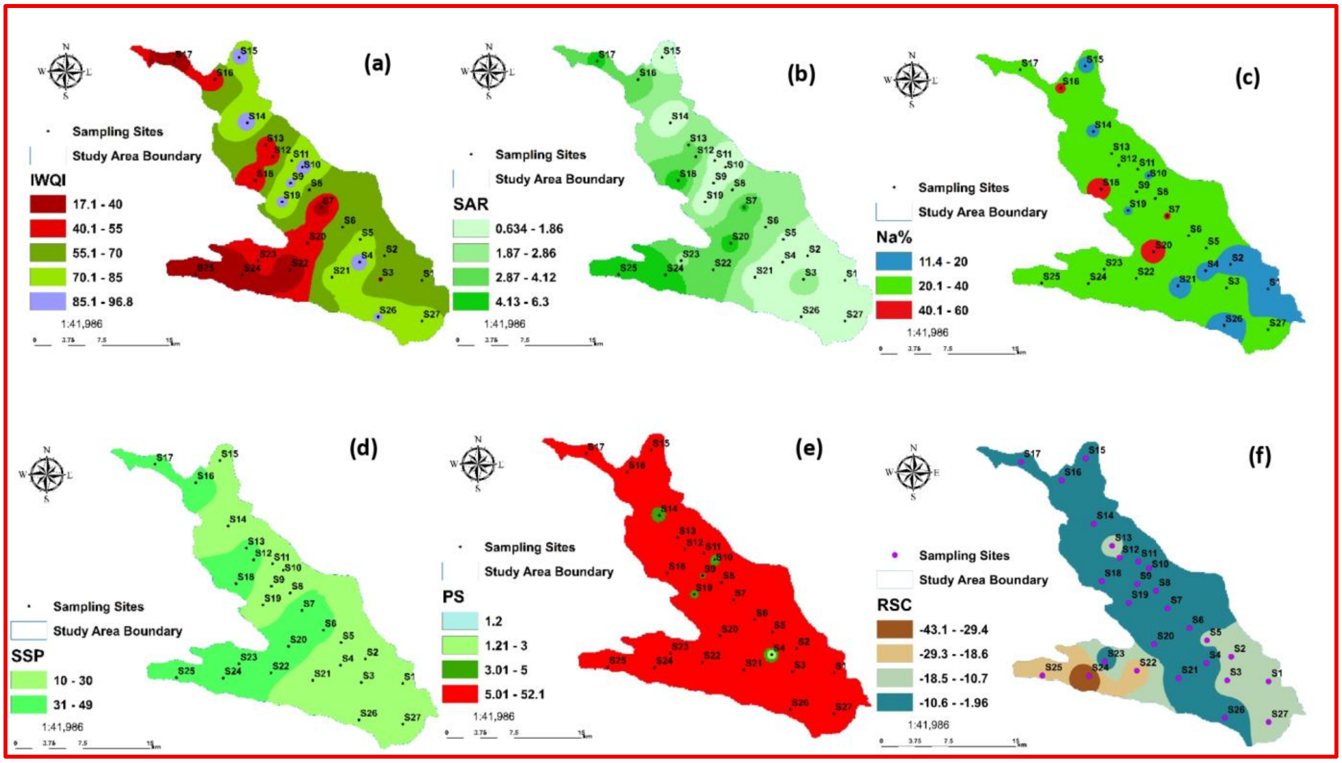


Figure 47 The spatial variation maps of the IWQIs for Al-Jawf plain: (a) IWQI, (b) SAR, (c) Na %, (d) SSP, (e) PS, and (f) RSC.

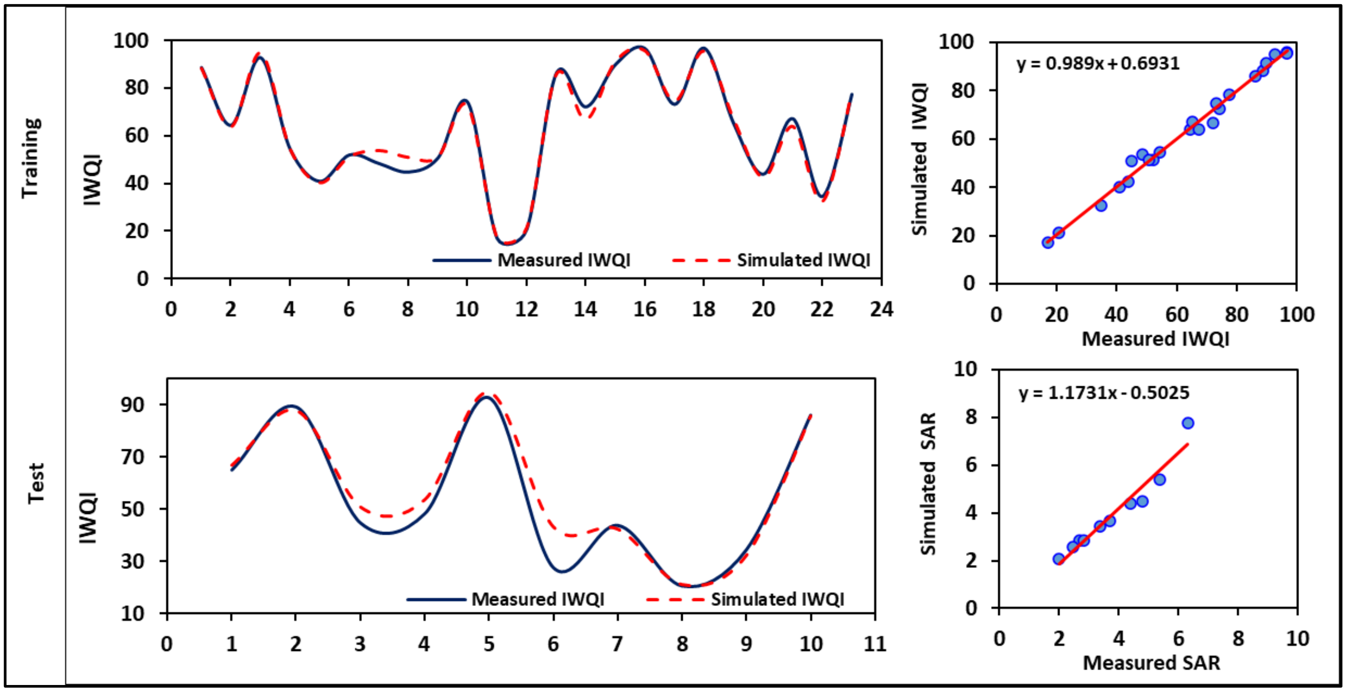


Figure 48 Results of the ANFIS model-based simulated IWQI.

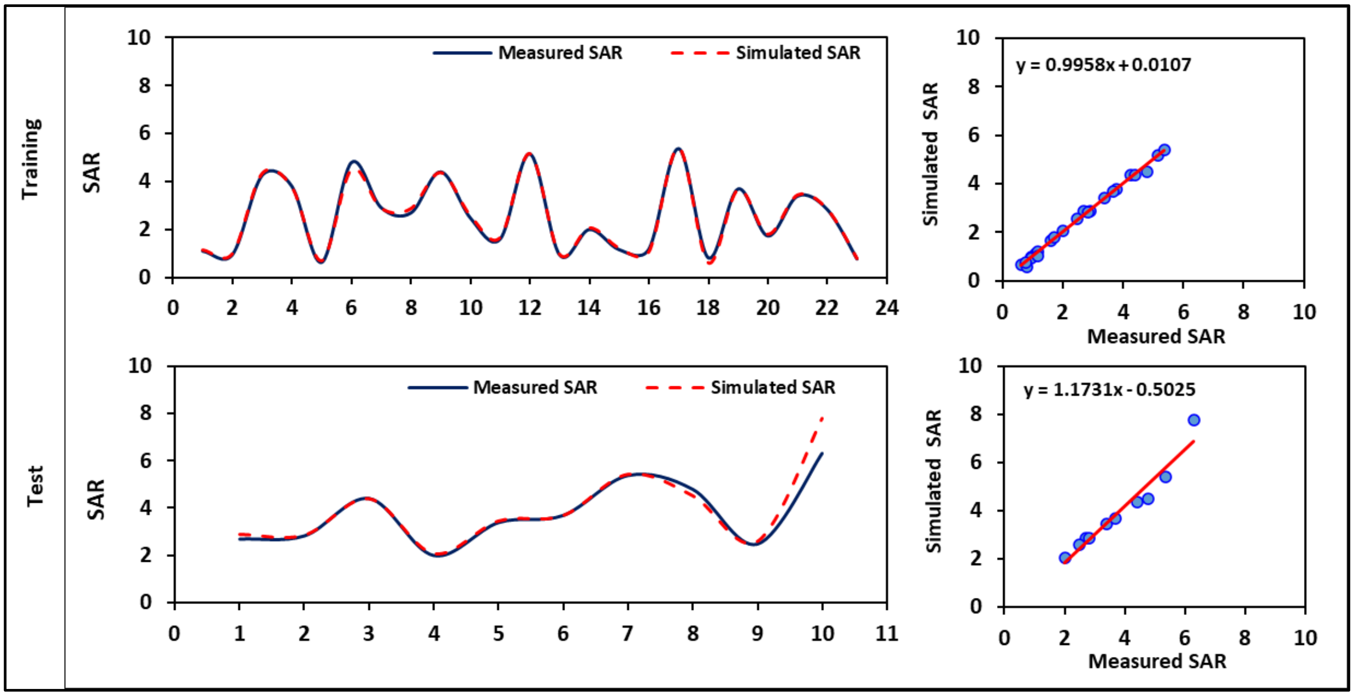


Figure 49 Results of the ANFIS model-based simulated SAR

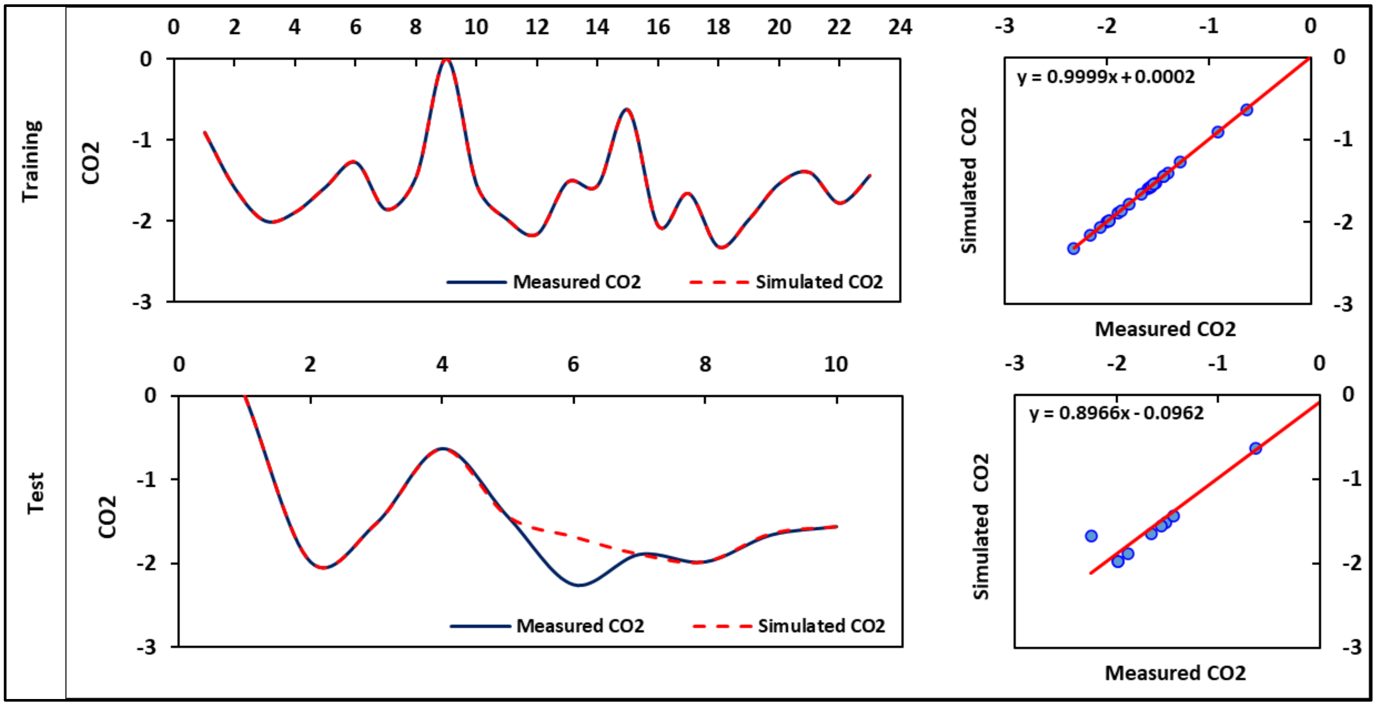


Figure 50 Results of the ANFIS model-based simulated CO₂

APPENDIX B – TABLES

Table 1 Summary of key studies on machine learning (ML) applications in water quality and health risk assessment

ML Techniques Used	Study Focus	Key Findings	Reference
ANN, SVM, CART, CRRF, KNN	Groundwater irrigation suitability	ML models, especially ANN, improved prediction accuracy over regression models	Shaw & Sharma (2025)
ML techniques	Groundwater quality for irrigation	ML models effectively assessed water quality parameters	Singh et al. (2024)
GIS-based ML algorithms	Irrigation groundwater quality indices	Integration of GIS with ML models enabled spatial prediction of water quality	Mohammed et al. (2023)

Table 2 Clarifying summary table of sampling numbers and periods

Region	Water Type	Sampling Years	Number of Sites	Samples Collected	Purpose
Danube, Hungary	Surface water	2013–2019 (monthly)	7	280	Heavy metal analysis
Danube, Hungary	Surface water	2019	7	85	Physicochemical properties
Al-Jawf, Yemen	Groundwater	2021–2022	9 districts	60	Heavy metals + physicochemical

Table 3 The calculation methods of irrigation indices.

IWQIs	Equation	References
IWQI	$\sum_{i=1}^{n} Q_i W_i$	(Meireles et al. 2010)
SAR	$(\frac{Na^+}{\sqrt{(Ca^{2+}+Mg^{2+})/2}})$ x 100	(McGeorge 1954)
NA%	$Na\% = (Na^+ + K^+)(Ca^{2+} + Mg^{2+}) + (Na^+ + K^+) \times 100$	(Ravikumar et al. 2013)
SSP	$[Na^{2+}/(Ca^{2+}+Mg^{2+}+ Na^{2+})]\times 100$	(Eaton 1950)
KR	$KI= Na^+/ (Ca^{2+} + Mg^{2+})$	(Al-Mashreki et al. 2023)
PS	$Cl^- + (SO_4^{2-}/2)$	(Al-Mashreki et al. 2023)
RSC	$(HCO_3^- + CO_3^-) - (Ca^{2+} + Mg^{2+})$	(Eaton 1950)

Table 4 The range of limit values of the parameters used in the computation of quality measurement (Qi)

Q_i	SAR	EC (µs/cm)	HCO_3^- (meq/L)	Na^+ (meq/L)	Cl^- (meq/L)
0 - 35	≥ 12	$EC < 200$ or $EC \geq 3000$	$HCO_3^- < 1$ or $HCO_3^- \geq 8.5$	$Na^+ < 2$ or $SAR \geq 9$	$Cl^- < 1$ or $Cl^- \geq 10$
35 – 60	$6 \leq SAR < 12$	$1500 \leq EC < 3000$	$4.5 \leq HCO_3^- < 8.5$	$6 \leq Na^+ < 9$	$7 \leq Cl^- < 10$
60 – 85	$3 \leq SAR < 6$	$750 \leq EC < 1500$	$1.5 \leq HCO_3^- < 4.5$	$3 \leq Na^+ < 6$	$4 \leq Cl^- < 7$
85 – 100	$2 \leq SAR < 3$	$200 \leq EC < 750$	$1 \leq HCO_3^- < 1.5$	$2 \leq Na^+ < 3$	$1 \leq Cl^- < 4$

Table 5 Comparative summary of considered machine learning models

Model	Considered	Used	Justification
Random Forest (RF)	✓	✓	Robust to overfitting, handles non-linear relationships, performs well with moderate datasets, and allows interpretation of variable importance.
ANFIS	✓	✓	Effective in modeling complex non-linear systems with fuzzy logic, suitable for small to medium datasets, and widely used in water quality modeling.
Support Vector Machine (SVM)	✓	✕	Considered, but excluded due to sensitivity to hyperparameter tuning and performance variability on small datasets.
Artificial Neural Network (ANN)	✓	✕	Avoided to reduce model complexity and overfitting risk due to limited sample size.
k-Nearest Neighbors (kNN)	✓	✕	Rejected due to sensitivity to noise and reduced performance in high-dimensional spaces with small datasets.

Table 6 The parameters for the computation of HQ, HI, RI and CR

HM	As	Cr	Cu	Fe	Mn	Ni	Pb	Zn	Ref.
RfD oral (mg/kg/day)	0.0003	0.003	0.04	0.7	0.024	0.02	0.0014	0.3	(Xu et al. 2020)
ABS	1	0.025	0.3	0.2	0.04	0.04	0.3	0.2	(Xu et al. 2020)
Rfd dermal(mg/kg/day)	0.0003	0.000075	0.012	0.14	0.00096	0.0008	0.00042	0.06	(Xu et al. 2020)
CSFing mg/kg/day	1.5	0.5	–	–	–	–	0.5	–	(Xu et al. 2020)
CSFderm	50	500	–	–	–	–	500	–	(Xu et al. 2020)
Kp	0.001	0.002	0.001	0.001	0.001	0.0002	0.0001	0.0006	(USEPA 2004)
Background (µg/g)	10	30	30	15000	500	20	20	100	(Woitke et al. 2003)
Tr	10	2	5	1	1	5	5	1	(Hakanson 1980)

Table 7 The overview of average HM levels in µg L^{-1} at the lower watershed of the Danube River, Hungary, compared with the threshold values for drinking water from the WHO Guidelines (WHO, 2017).

Sampling Sites	Period	As	Cr	Cu	Fe	Mn	Ni	Pb	Zn
S1	April-Sep	1.21	1.45	4.29	**626.8**	**55.26**	2.55	1.22	16.45
	Oct-March	1.30	1.07	3.33	**405.5**	38.88	2.41	0.93	14.35
S2	April-Sep	1.27	1.3	4.17	**518.3**	48.26	2.64	1.61	14.78
	Oct-March	1.27	0.93	3.37	**336.1**	44.78	2.29	1.37	14.44
S3	April-Sep	1.34	1.42	4.4	**634.73**	**55.26**	2.64	1.42	17.66
	Oct-March	1.45	1.01	3.71	**452.9**	34.37	2.71	1.03	16.25
S4	April-Sep	1.35	1.8	3.48	170.67	40	3.1	1.14	12.93
	Oct-March	1.28	1.58	3.81	**208.3**	26	2.45	1.3	12.79
S5	April-Sep	1.37	1.11	3.74	**436.76**	43.42	2.87	1.2	11.98
	Oct-March	1.46	1.42	3.62	**381.2**	31.91	2.3	1.12	16.13
S6	April-Sep	1.20	1.49	4.11	**590.5**	**72.1**	2.86	1.45	14.38
	Oct-March	1.70	1.83	3.79	**511**	**65.71**	2.6	1.55	15.8
S7	April-Sep	1.31	1.44	4.03	**666.67**	**56.84**	2.27	1.14	16.8
	Oct-March	1.60	1.65	4.17	**528.5**	36.3	2.8	1.34	19.39
WHO-2017		10	50.0	3000	300	50	70	10	1000

Table 8 Mean concentrations (mg/L) of major ions and related physicochemical parameters in surface water samples from the Danube river basin, south of Hungary, with reference to FAO and WHO guidelines for drinking and irrigation water.

Sampling sites	S1	S2	S3	S4	S5	S6	S7	FAO (mg/L)	WHO (2017) (mg/L)
Ca^{2+} (mg/L)	54.45	54.15	54.41	53.93	53.59	53.72	53.11	400	75
Mg^{2+} (mg/L)	16.13	15.59	15.70	15.31	15.25	14.76	14.50	60	50
Na^{+} (mg/L)	16.56	16.68	16.56	15.63	15.87	15.57	15.27	919	200
Cl^{-} (mg/L)	24.50	24.25	24.17	22.92	22.50	22.33	22.23	1036	250
SO_4^{2-} (mg/L)	37.25	37.08	37.08	36.67	35.08	35.00	34.85	960	250
HCO_3^{-} (mg/L)	195.42	188.33	193.33	190.83	190.00	188.33	181.15	610	120
CO_3^{2-} (mg/L)	9.17	10.08	8.08	10.42	8.75	9.17	10.23	–	350
T (°C)	14.53	14.06	14.93	14.48	15.51	15.44	16.21	–	–
EC (µS/cm)	467.92	466.67	462.08	464.58	463.75	464.17	451.15	3000	1500
TDS (mg/L)	272.50	263.38	251.88	255.71	296.36	254.70	257.44	2000	500
K^{+} (mg/L)	2.58	2.69	2.60	2.50	2.74	2.64	2.57	2	12
NO_3^{-} (mg/L)	8.37	7.86	8.11	8.04	7.13	7.35	7.28	–	50
pH	8.14	8.16	8.16	8.14	8.16	8.18	8.13	8.5	8.5
TH (mg/L as $CaCO_3$)	113.82	112.14	112.73	115.02	110.56	109.61	108.15	–	500

Fe (mg/L)	0.60	0.58	0.57	0.60	0.63	0.53	0.65	5	0.3
Mn (mg/L)	0.07	0.06	0.07	0.05	0.07	0.05	0.05	0.2	0.1

Table 9 Varimax rotated factor of principal component analysis

Rotated Component Matrix				
	Principal components			
	PC1	PC2	PC3	PC4
Ca^{2+} (mg/L)	0.945	-0.053	-0.149	-0.003
Mg^{2+} (mg/L)	0.897	0.011	0.347	0.040
Na^{+} (mg/L)	0.933	-0.199	0.190	0.003
Cl^{-} (mg/L)	0.957	-0.130	0.062	-0.013
SO_4^{2-} (mg/L)	0.791	-0.002	0.526	-0.031
HCO_3^{-} (mg/L)	0.843	-0.414	-0.084	-0.049
CO_3^{2-} (mg/L)	-0.076	0.940	-0.026	-0.044
EC (µS/cm)	0.926	0.054	0.043	-0.086
TDS (mg/L)	0.308	-0.197	0.110	-0.598
K^{+} (mg/L)	0.127	-0.042	0.321	0.774
NO_3^{-} (mg/L)	0.825	-0.104	-0.302	0.069
pH	-0.102	0.904	0.145	-0.009
TH (mg/L)	0.985	-0.029	0.047	0.014
Fe (mg/L)	-0.028	-0.096	-0.904	0.033
Mn (mg/L)	0.093	-0.179	-0.246	0.518
Eigenvalues	7.464	2.029	1.577	1.245
% of Variance	49.759	13.527	10.511	8.300
Cumulative %	49.759	63.287	73.798	82.098

Table 10 Mean values of HQ oral of Mn, Fe and NO_3^- for adults and children

Parameters		Minimum	Maximum	Mean
HQ oral for adult	Fe	0.001	0.125	0.024
	Mn	0.025	0.439	0.074
	NO_3^-	0.050	0.243	0.137
HQ oral for children	Fe	0.002	0.477	0.092
	Mn	0.096	1.678	0.281
	NO_3^-	0.190	0.927	0.523
Non-carcinogenic risk	HI oral for adult	0.096	0.577	0.235
	HI oral for children	0.366	2.201	0.896

Table 11 Comparative analysis of physicochemical properties and trace metal concentrations against FAO standards

Parameter	FAO	Min	Max	Average	SD	CV%
pH	8.50	7.00	8.20	7.20	0.29	0.04
EC (μS/cm)	3000	1030	5930	2710	1409.15	0.52
TDS (mg/L)	2000	697.93	3734.73	1773.39	853.03	0.48
K^+ (mg/L)	2	2.00	23.46	8.99	4.07	0.45
Na^+ (mg/L)	919	51.73	689.70	280.00	177.10	0.63
Mg^{2+} (mg/L)	60	41.31	359.64	95.00	67.00	0.71
Ca^{2+} (mg/L)	400	40.00	304.61	120.24	67.96	0.57
NO_3^- (mg/L)	10	0.80	10.50	4.00	2.50	0.62
Cl^- (mg/L)	1036	127.62	1013.87	418.31	260.89	0.62
SO_4^{2-} (mg/L)	960	14.41	1320.83	518.72	340.21	0.66
HCO_3^- (mg/L)	610	170.86	463.75	286.79	79.34	0.28
CO_3^{2-}(mg/L)	3	0.10	6.00	0.80	1.25	1.56
Fe (mg/L)	5	0.01	3.0	0.09	0.16	1.73
Mn (mg/L)	0.2	0.001	4.0	0.02	0.09	4.42

CV: Coefficient of Variation, SD: Standard Deviation. FAO: Food and Agriculture Organization (Ayers and Westcot 1985)

Table 12 The association among various parameters and factors.

	Principal components		
Parameters	PC1	PC2	PC3
TDS (mg/L)	0.99	0.09	0.06
Na^+ (mg/L)	0.82	0.40	0.11
K^+ (mg/L)	0.66	0.21	0.66
Ca^{2+} (mg/L)	0.85	-0.24	-0.01
Mg^{2+} (mg/L)	0.81	-0.21	-0.01
Cl^- (mg/L)	0.86	0.15	0.03
SO_4^{2-} (mg/L)	0.88	-0.01	0.16
HCO_3^- (mg/L)	0.73	-0.01	-0.44
CO_3^{2-} (mg/L)	0.16	0.73	-0.41
NO_3^- (mg/L)	-0.09	0.84	-0.02
Fe^+ (mg/L)	0.68	-0.30	-0.31
Mn^+ (mg/L)	0.79	-0.16	-0.17

Table 13 Evaluation metrics for RF models on measured parameters, including IWQI, SAR, KR, SSP, PS, RSBC, HI (adult), and HI (child)

Index	Input variables	Model Parameters (ntree, mtry)	Training R^2	Training RMSE	Testing R^2	Testing RMSE
IWQI	EC, Na^+, Cl^-	(1, 17)	0.988	2.239	0.940	3.976
SAR	Mg^{2+}, Na^+	(2, 9)	0.971	0.404	0.869	0.667
KR	Ca^{2+}, Na^+	(2, 8)	0.952	0.099	0.665	0.203
SSP	TDS, Cl^-, Ca^{2+}, Na^+	(3, 13)	0.959	2.453	0.671	5.394
PS	Cl^-, EC	(2, 6)	0.994	0.744	0.982	0.960
RSBC	Cl^-, Na^+, Mg^{2+}, pH, SO_4^{2-}, HCO_3^-, EC, Mn, K^+, Fe, NO_3^-, Ca^{2+}	(12, 2)	0.924	0.768	0.751	1.042
HI (adult)	Mn	(1, 2)	0.988	0.014	0.903	0.018
HI (child)	Mn	(1, 2)	0.988	0.041	0.903	0.048

Table 14 Descriptive results of the groundwater samples with the standard limit for irrigation purposes

Parameter	FAO	Min	Max	Average
pH	8.5	6.5	7.5	7.13
Temp. (°C)	-	11.5	27.1	20.7
TDS (mg/L)	2000	378.42	5012	1685
EC (µS/cm)	3000	542	6628	2361.72
Ca^{2+} (mg/L)	400	40	460	179.69
Mg^{+2} (mg/L)	60	32.81	328	114.5
K^+ (mg/L)	2	2.34	23.40	9.54
Na^+ (mg/L)	919	24.15	724.5	197.2
HCO_3^- (mg/L)	610	150	915	408.18
Cl^- (mg/L)	1036	35.5	1136	395
SO_4^{2-} (mg/L)	960	19.2	1939	392
NO_3^- (mg/L)	10	0.1	6	2.21

Table 15 Correlation between the parameters and factors

	Principal components		
Parameters	PC1	PC2	PC3
Ca^{2+} (mg/L)	0.770	0.450	0.133
Mg^{2+} (mg/L)	0.939	0.196	0.012
Na^{+} (mg/L)	0.912	0.191	0.056
K^{+} (mg/L)	0.514	0.637	-0.294
HCO_3^- (mg/L)	0.176	0.887	0.111
Cl^- (mg/L)	0.900	0.261	0.192
SO_4^{2-} (mg/L)	0.934	-0.090	-0.235
CO_3^{2-} (mg/L)	0.142	0.375	0.781
TDS (mg/L)	0.944	0.316	0.016
EC (µS/cm)	0.955	0.287	0.036
NO3- (mg/L)	-0.044	-0.205	0.785
Eigenvalue	6.111	1.911	1.439
% of Variance	55.556	17.376	13.086
Cumulative %	55.556	72.932	86.018

Table 16 Statistical description of the mineral SI for the obtained GW samples

SI	Anhydrite	Aragonite	Calcite	Dolomite	Gypsum	Halite	CO_2 (g)
Min.	-2.79	-0.41	-0.27	-0.29	-2.57	-7.64	-2.32
Max.	-0.44	0.49	0.63	1.45	-0.19	-4.77	-0.63
Mean	-1.501	0.003	0.151	0.404	-1.269	-6.075	-1.649

Table 17 Statistical analysis and classes of IWQIs.

Criteria	Min	Max	Mean	Range	Class	Number of Samples (%)
IWQI	17.03	96.77	61.03	85 - 100	No restriction	7 (25.920%)
				70 - 85	Low restriction	4 (14.81%)
				55 - 70	Moderate restriction	4 (14.81%)
				40 - 55	High restriction	7 (25.92%)
				0 - 40	Severe restriction	5 (18.51%)
SAR	0.63	6.30	2.56	<10	Excellent	27(100%)
				10–18	Good	0 (0%)
				19–26	Fair Poor	0 (0%)
				>26	Unsuitable	0 (0%)
Na%	11.35	49.73	28.29	<20%	Excellent	10 (37%)
				21%–40%	Good	13(48.14%)
				41%–60%	Permissible	4(14.81%)
				61%–80%	Doubtful	0 (0%)
				> 80%	Unsuitable	50 (100%)
SSP	10.39	49.45	27.57	<60	Suitable	27 (100%)
				>60	Unsuitable	0 (0%)
PS	1.2	52.20	15.23	PS < 3.0	Excellent to good	2(7.40%)
				PS = 3.0–5.0	Good to injurious	3(11.11%)
				PS > 5.0	Injurious to unsatisfactory	22(81.48%)
RCS	-43.21	-1.96	-11.89	<1.25	Good	27 (100%)

1.25-2.5	Doubtful	0 (0%)
>2.5	Unsuitable	0 (0%)

Table 18 Performance criteria of the simulation models for IWQIs prediction.

	Index	Performance criteria			
		R^2	*RMSE*	*MAD*	*E*
Training Series	IWQI	0.999	2.393	1.691	0.999
	SAR	0.973	0.098	0.067	0.973
	SSP	0.996	11.202	8.802	0.996
	Anhydrite	0.970	0.003	0.002	0.971
	Aragonite	0.955	0.001	0.001	0.955
	Dolomite	0.988	0.001	0.000	0.980
	Halite	0.985	0.007	0.003	0.980
	Gypsum	0.976	0.002	0.001	0.979
	CO_2	0.9 80	0.002	0.001	0.985
Testing Series	IWQI	0.960	5.670	3.665	0.949
	SAR	0.940	0.479	0.222	0.868
	SSP	0.892	11.667	9.880	0.864
	Anhydrite	0.895	0.354	0.159	0.775
	Aragonite	0.908	0.071	0.032	0.886
	Dolomite	0.878	0.318	0.114	0.443
	Halite	0.932	0.860	0.275	0.070
	Gypsum	0.964	0.174	0.056	0.958
	CO_2	0.879	0.019	0.007	0.999

APPENDIX C –PUBLICATIONS RELATED TO THE DISSERTATION

1. Eid, M., Saeed, O., Szűcs, P. et al. Impacts and sources of potential toxic elements on water quality and optimizing machine learning models for sustainable management. Model. Earth Syst. Environ. 11, 375 (2025). https://doi.org/10.1007/s40808-025-02548-z. (Q1)
2. Saeed, O.*, et al. Assessing surface water quality in Hungary's Danube basin using geochemical modeling, multivariate analysis, irrigation indices, and Monte Carlo simulation. Sci. Rep. 14, 18639 (2024). https://doi.org/10.1038/s41598-024-69312-8. (D1)
3. Saeed, O.*, et al. Correction: Investigating the impacts of heavy metal(loid)s on ecology and human health in the lower basin of Hungary's Danube River: A Python and Monte Carlo simulation-based study. Environ. Geochem. Health 45, 9757–9784 (2023). https://doi.org/10.1007/s10653-023-01769-4. (Q1)
4. Saeed, O.*, et al. Investigating the impacts of heavy metal(loid)s on ecology and human health in the lower basin of Hungary's Danube River: A Python and Monte Carlo simulation-based study. Environ. Geochem. Health 45, 9757–9784 (2023). https://doi.org/10.1007/s10653-023-01769-4. (Q1)
5. Al-Mashreki, M.H., Eid, M.H., Saeed, O. Integration of geochemical modeling, multivariate analysis, and irrigation indices for assessing groundwater quality in the Al-Jawf basin, Yemen. Water 15, 1496 (2023). https://doi.org/10.3390/W15081496. (Q1)

APPENDIX D – OTHER PUBLICATIONS

1. Eid, M.H., Saeed, O., Szűcs, P. et al. Seasonal hydrochemical characteristics of spring water in Southern Poland: integrating geochemical modeling, health risk analysis and mitigation strategies. Sci Rep **15**, 25459 (2025). https://doi.org/10.1038/s41598-025-10322-5
2. Eid, M.H., Saeed, O., Székács, A. et al. Integrating unsupervised machine learning, statistical analysis, and Monte Carlo simulation to assess toxic metal contamination and salinization in non-rechargeable aquifers. Results Eng 12, 104989 (2025). https://doi.org/10.1016/j.rineng.2025.104989
3. Alsakran, A.A., Younis, O.S., Székács, A. et al. CFD, energy, and exergy analysis and sustainability indicators of tilapia fish strips drying using an evacuated tubes indirect solar dryer. Sci Rep **15**, 25893 (2025). https://doi.org/10.1038/s41598-025-11230-4
4. Tantawy, A.A., Ali, M., Kurbonova, M. et al. Functional properties of sunflower protein concentrates extracted using different anti-greening agents: low-fat whipping cream preparation. LWT 211, 117456 (2025). https://doi.org/10.1016/j.lwt.2025.117456
5. Eid, M.H., Tamma, A.A., Saeed, O. et al. Advanced approach combines integrated weight water quality index and potential toxic elements for environmental and health risk assessment supported by simulation technique in Oued Souf, Algeria. Sci Rep **14**, 17805 (2024). https://doi.org/10.1038/s41598-024-68854-1
6. Lameck, A.S., Saeed, O., Justine, P.N. et al. Hydrochemical properties and heavy metal concentrations (ecological and human risk) of Lake Rukwa. Environ Challenges 14, 100940 (2024). https://doi.org/10.1016/j.envc.2024.100940
7. Lameck, A.S., Saeed, O., Boros, E. The chemical composition, classification, and geographical distributions of soda-saline lakes in Eastern Tanzania's Rift Valley. Eur J Remote Sens Hydrol 16, 101668 (2024). https://doi.org/10.1016/j.ejrh.2024.101668
8. Younis, O.S., Oraiath, A.A.T., Metwally, K.A. et al. Drying characteristics, environmental and economic analysis of a solar dryer with evacuated tube solar collector for drying Nile Tilapia slices. Sci Rep **15**, 9822 (2025). https://doi.org/10.1038/s41598-025-94194-9
9. Alnemari, A.M., Elmessery, W.M., Moghanm, F.S. et al. Energy optimization in large-scale recirculating aquaculture systems: Implementation and performance analysis of a hybrid deep learning approach. Aquac Eng 110, 102561 (2025). https://doi.org/10.1016/j.aquaeng.2025.102561

APPENDIX E – REFERENCES

Abbasnia A, Yousefi N, Mahvi AH, et al. (2019). Evaluation of groundwater quality using water quality index and its suitability for assessing water for drinking and irrigation purposes: Case study of Sistan and Baluchistan province (Iran). Hum Ecol Risk Assess 25:988–1005. https://doi.org/10.1080/10807039.2018.1458596

Abou El-Safa MM, Elsayed S, Elsherbiny O, et al. (2022). Environmental assessment of potentially toxic elements using pollution indices and data-driven modeling in surface sediment of the littoral shelf of the Mediterranean Sea coast and Gamasa estuary, Egypt. J Mar Sci Eng 10:816. https://doi.org/10.3390/jmse10060816

Adnan Ikram RM, Jaafari A, Milan SG, et al. (2022). Hybridized adaptive neuro-fuzzy inference system with metaheuristic algorithms for modeling monthly pan evaporation. Water 14:3519. https://doi.org/10.3390/w14213549

Ahlbrandt TS. (2002). Madbi Amran/Qishn total petroleum system of the Ma'rib–Al Jawf/Shabwah, and Masilah–Jeza basins, Yemen. US Geol Surv Bull 2202-F. US Department of the Interior.

Ahmad W, Alharthy RD, Zubair M, et al. (2021). Toxic and heavy metals contamination assessment in soil and water to evaluate human health risk. Sci Rep 11:94616. https://doi.org/10.1038/s41598-021-94616-4

Ahmadi RA, Varasteh T, Silveira CB, et al. (2023). Machine learning sheds light on physical-chemical and biological parameters leading to Abrolhos coral reef microbialization. Sci Total Environ 891:164465. https://doi.org/10.1016/j.scitotenv.2023.164465

Ahmed AA, Sayed S, Abdoulhalik A, et al. (2024). Applications of machine learning to water resources management: A review of present status and future opportunities. J Clean Prod 441:140715. https://doi.org/10.1016/j.jclepro.2024.140715

Ahmed MT, Hasan MY, Monir MU, et al. (2020). Evaluation of hydrochemical properties and groundwater suitability for irrigation uses in southwestern zones of Jashore, Bangladesh. Groundw Sustain Dev 11:100441. https://doi.org/10.1016/j.gsd.2020.100441

Al-Hejuje MM, Hussain NA, Al-Saad HT. (2017). Applied heavy metals pollution index (HPI) as a water pollution indicator of Shatt Al-Arab River, Basrah-Iraq. Int J Mar Sci 7:353–360. https://doi.org/10.5376/ijms.2017.07.0035

Al-Mashreki MH, Eid MH, Saeed O, et al. (2023). Integration of geochemical modeling, multivariate analysis, and irrigation indices for assessing groundwater quality in the Al-Jawf Basin, Yemen. Water 15:1496. https://doi.org/10.3390/w15081496

Alaug AS, Al-Wosabi KA. (2015). Organic geochemical evaluation of Madbi source rock, Al-Jawf Basin, NE central Yemen. Iran J Earth Sci 7:25–36.

Alfaica AO, Rampinelli CG, Tiago Filho GL, et al. (2023). Hydrological uncertainty analysis using Monte Carlo simulations to determine power purchasing agreements for small hydroelectric powerplants. Renew Energy 219:119349. https://doi.org/10.1016/j.renene.2023.119349

Ali MM, Ali ML, Islam MS, Rahman MZ. (2016). Preliminary assessment of heavy metals in water and sediment of Karnaphuli River, Bangladesh. Environ Nanotechnol Monit Manag 5:27–35. https://doi.org/10.1016/j.enmm.2016.01.002

Allafta H, Opp C. (2020). Spatio-temporal variability and pollution sources identification of the surface sediments of Shatt Al-Arab River, southern Iraq. Sci Rep 10:63893. https://doi.org/10.1038/s41598-020-63893-w

Alquraish M, Abuhasel K, Alqahtani AS, Khadr M. (2021). SPI-based hybrid hidden Markov–GA, ARIMA–GA, and ARIMA–GA–ANN models for meteorological drought forecasting. Sustain 13:12576. https://doi.org/10.3390/su132212576

An JM, Gu SY, Kim DJ, Shin HC, Hong KS, Kim YK. (2020). Arsenic, cadmium, lead, and mercury contents of mushroom species in Korea and associated health risk. J Food Prop 23:992–998. https://doi.org/10.1080/10942912.2020.1770786

Apollaro C, Tripodi V, Vespasiano G, et al. (2019). Chemical, isotopic and geotectonic relations of the warm and cold waters of the Galatro and Antonimina thermal areas, southern Calabria, Italy. Mar Pet Geol 109:469–483. https://doi.org/10.1016/j.marpetgeo.2019.06.020

Athamena A, Gaagai A, Aouissi HA, et al. (2023). Chemometrics of the environment: Hydrochemical characterization of groundwater in Lioua Plain (North Africa) using time series and multivariate statistical analysis. Sustain 15:20. https://doi.org/10.3390/su15010020

Athamena A, Menani MR. (2018). Nitrogen flux and hydrochemical characteristics of the calcareous aquifer of the Zana Plain, northeast Algeria. Arab J Geosci 11:3681. https://doi.org/10.1007/s12517-018-3681-5

Aullón Alcaine A, Schulz C, Bundschuh J, et al. (2020). Hydrogeochemical controls on the mobility of arsenic, fluoride and other geogenic co-contaminants in the shallow aquifers of northeastern La Pampa Province in Argentina. Sci Total Environ 715:136671. https://doi.org/10.1016/j.scitotenv.2020.136671

Ayers R, Westcot D. (1985). Water quality for agriculture. FAO Irrigation and Drainage Paper 29. Rome: FAO.

Babcsányi I, Tamás M, Szatmári J, Ringer M, Dobos E, Bidló A. (2020). Assessing the impacts of the main river and anthropogenic use on the degree of metal contamination of oxbow lake sediments (Tisza River Valley, Hungary). J Soils Sediments 20:1662–1675. https://doi.org/10.1007/s11368-019-02516-y

Bai B, Rao D, Chang T, Guo Z. (2019). A nonlinear attachment-detachment model with adsorption hysteresis for suspension-colloidal transport in porous media. J Hydrol 578:124080. https://doi.org/10.1016/j.jhydrol.2019.124080

Ball T. (2008). Management approaches to floodplain restoration and stakeholder engagement in the UK: A survey. Ecohydrol Hydrobiol 8:273–280. https://doi.org/10.2478/v10104-009-0021-0

Baragaño D, Boente C, Rodríguez-Valdés E, Pérez C, de la Fuente V. (2020). Arsenic release from pyrite ash waste over an active hydrogeological system and its effects on water quality. Environ Sci Pollut Res 27:10672–10684. https://doi.org/10.1007/s11356-019-07120-8

Barkat A, Bouaicha F, Bouteraa O, Kacimi A. (2021). Assessment of complex terminal groundwater aquifer for different use of Oued Souf Valley (Algeria) using multivariate statistical methods, geostatistical modeling, and water quality index. Water 13:1609. https://doi.org/10.3390/w13111609

Bel Hadj Salem S, Chkir N, Zouari K, et al. (2012). Natural and artificial recharge investigation in the Zéroud Basin, central Tunisia: Impact of Sidi Saad Dam storage. Environ Earth Sci 66:1099–1110. https://doi.org/10.1007/s12665-011-1316-1

Belhadj M, Boudoukha A, Amroune A, Gaagai A, Ziani D. (2017). Statistical characterization of groundwater quality of the northern area of the basin of Hodna, M'sila, southeastern Algeria. Larhyss J 31:177–194. http://www.larhyss.net

Belvederesi C, Dominic JA, Hassan QK, et al. (2020). Predicting river flow using an AI-based sequential adaptive neuro-fuzzy inference system. Water 12:1622. https://doi.org/10.3390/w12061622

Ben Moussa A, Chandoul S, Mzali H, et al. (2021). Hydrogeochemistry and evaluation of groundwater suitability for irrigation purpose in the Mornag region, northeastern Tunisia. Environ Dev Sustain 23:2698–2718. https://doi.org/10.1007/s10668-020-00696-z

Bineshpour M, Payandeh K, Nazarpour A, Sabzalipour S. (2021). Status, source, human health risk assessment of potential toxic elements (PTEs), and Pb isotope characteristics in urban surface soil: Case study, Arak city, Iran. Environ Geochem Health 43:4939–4958. https://doi.org/10.1007/s10653-020-00778-x

Blubaugh DJ, Atamian M, Babcock GT, et al. (1991). Photoinhibition of hydroxylamine-extracted Photosystem II membranes: Identification of the sites of photodamage. Biochemistry 30:7586–7597. https://doi.org/10.1021/bi00244a030

Calmuc VA, Calmuc M, Arseni M, et al. (2021). Assessment of heavy metal pollution levels in sediments and of ecological risk by quality indices, applying a case study: The lower Danube River, Romania. Water 13:1801. https://doi.org/10.3390/w13131801

Chadha D. (1999). Proposed new diagram for geochemical classification of natural waters and interpretation of chemical data. Hydrogeol J 7:431–439. https://doi.org/10.1007/s100400050216

Chebotarev II. (1955). Metamorphism of natural waters in the crust of weathering-3. Geochim Cosmochim Acta 8:198–212. https://doi.org/10.1016/0016-7037(55)90053-3

Chen K, Yu S, Ma T, et al. (2022). Modeling the water and nitrogen management practices in paddy fields with HYDRUS-1D. Agriculture 12:924. https://doi.org/10.3390/agriculture12070924

Chen Q, Yao Y, Li X, et al. (2018). Comparison of heavy metal removals from aqueous solutions by chemical precipitation and characteristics of precipitates. J Water Process Eng 26:289–300.

https://doi.org/10.1016/j.jwpe.2018.11.003

Cheng Z, Chen LJ, Li HH, et al. (2018). Characteristics and health risk assessment of heavy metals exposure via household dust from urban area in Chengdu, China. Sci Total Environ 619–620:621–629. https://doi.org/10.1016/j.scitotenv.2017.11.144

Cho YC, Choi H, Lee MG, et al. (2022). Identification and apportionment of potential pollution sources using multivariate statistical techniques and APCS-MLR model to assess surface water quality in Imjin River Watershed, South Korea. Water 14:793. https://doi.org/10.3390/w14050793

Chounlamany V, Tanchuling MA, Inoue T. (2017). Spatial and temporal variation of water quality of a segment of Marikina River using multivariate statistical methods. Water Sci Technol 76:1510–1522. https://doi.org/10.2166/wst.2017.279

Chowdhury S, Mazumder MAJ, Al-Attas O, Husain T. (2016). Heavy metals in drinking water: Occurrences, implications, and future needs in developing countries. Sci Total Environ 569–570:476–488. https://doi.org/10.1016/j.scitotenv.2016.06.166

Deepa A, Singh A, Singh A, Mishra BK. (2021). An experimental approach for the utilization of tannery sludge-derived Bacillus strain for biosorptive removal of Cr(VI)-contaminated wastewater. Environ Sci Pollut Res 28:9864–9876. https://doi.org/10.1007/s11356-020-11284-z

Derwich LJ, Zouar K, Michelot JL. (2012). Recharge and paleorecharge of the deep groundwater aquifer system in the Zeroud Basin (Kairouan plain, central Tunisia). Quat Int 257:56–63. https://doi.org/10.1016/j.quaint.2011.12.003

Dujmović J, Allen WL. (2021). Soft computing logic decision making in strategic conservation planning for water quality protection. Ecol Inform 61:101167. https://doi.org/10.1016/j.ecoinf.2020.101167

Eaton FM. (1950). Significance of carbonates in irrigation waters. Soil Sci 69:123–133. https://doi.org/10.1097/00010694-195002000-00004

Edet AE, Offiong OE. (2002). Evaluation of water quality pollution indices for heavy metal contamination monitoring: A study case from Akpabuyo-Odukpani area, Lower Cross River

Basin (southeastern Nigeria). GeoJournal 57:295–304. https://doi.org/10.1023/b:gejo.0000007250.92458.de

Eid MH, Elbagory M, Tamma AA, et al. (2023). Evaluation of groundwater quality for irrigation in deep aquifers using multiple graphical and indexing approaches supported with machine learning models and GIS techniques, Souf Valley, Algeria. Water 15:182. https://doi.org/10.3390/w15010182

Eid MH, Saeed O, Székács A, et al. (2025a). Integrating unsupervised machine learning, statistical analysis, and Monte Carlo simulation to assess toxic metal contamination and salinization in non-rechargeable aquifers. Results Eng 26:104989. https://doi.org/10.1016/j.rineng.2025.104989

Eid MH, Saeed O, Szűcs P, et al. (2025b). Impacts and sources of potential toxic elements on water quality and optimizing machine learning models for sustainable management. Model Earth Syst Environ 11:375. https://doi.org/10.1007/s40808-025-02548-z

Eid MH, Tamma AA, Saeed O, et al. (2024). Advanced approach combines integrated weight water quality index and potential toxic elements for environmental and health risk assessment supported by simulation technique in Oued Souf, Algeria. Sci Rep 14:68854. https://doi.org/10.1038/s41598-024-68854-1

El-Hendawy S, Tahir MU, Al-Suhaibani N, et al. (2024). Potential of thermal and RGB imaging combined with artificial neural networks for assessing salt tolerance of wheat genotypes grown in real-field conditions. Agronomy 14:1390. https://doi.org/10.3390/agronomy14071390

El Mejri H, Ben Moussa A, Zouari K. (2014). The use of hydrochemical and environmental isotopic tracers to understand the functioning of the aquifer system in the Bou Hafna and Haffouz regions, Central Tunisia. Quat Int 338:88–98. https://doi.org/10.1016/j.quaint.2014.04.046

Elsayed S, El-Hendawy S, Khadr M, et al. (2021). Integration of spectral reflectance indices and adaptive neuro-fuzzy inference system for assessing the growth performance and yield of potato under different drip irrigation regimes. Chemosensors 9:55. https://doi.org/10.3390/chemosensors9030055

Etu JE, Oyedepo OJ. (2018). Forecasting trip generation for high density residential zones of Akure, Nigeria: Comparability of artificial neural network and regression models. J Civ Eng Sci Technol 9:76–86. https://doi.org/10.33736/jcest.988.2018

Everest T, Özcan H. (2019). Applying multivariate statistics for identification of groundwater resources and qualities in NW Turkey. Environ Monit Assess 191:106. https://doi.org/10.1007/s10661-018-7165-6

Fallahzadeh RA, Almodaresi SA, Ghadirian D, et al. (2019). Spatial analysis and probabilistic risk assessment of exposure to nitrate in drinking water of Abarkouh, Iran. J Environ Health Sustain Dev 4:744–752. https://doi.org/10.18502/jehsd.v4i2.1054

Fuchs M, Láng V, Szegi T, Michéli E. (2015). Traditional and pedometric approaches to justify the introduction of swelling clay soils as a new soil type in the modernized Hungarian Soil Classification System. Catena 128:80–94. https://doi.org/10.1016/j.catena.2015.01.024

Gaagai A, Aouissi HA, Bencedira S, et al. (2023). Application of water quality indices, machine learning approaches, and GIS to identify groundwater quality for irrigation purposes: A case study of Sahara Aquifer. Water 15:289. https://doi.org/10.3390/w15020289

Gad M, El-Hendawy S, Al-Suhaibani N, et al. (2020a). Combining hydrogeochemical characterization and a hyperspectral reflectance tool for assessing quality and suitability of two groundwater resources for irrigation. Water 12:2169. https://doi.org/10.3390/w12082169

Gad M, El-Safa MMA, Farouk M, et al. (2021). Integration of water quality indices and multivariate modeling for assessing surface water quality in Qaroun Lake, Egypt. Water 13:2258. https://doi.org/10.3390/w13162258

Gad M, Gaagai A, Agrama AA, Belhadj M. (2024). Comprehensive evaluation and prediction of groundwater quality and risk indices using quantitative approaches, multivariate analysis, and machine learning models: An exploratory study. Heliyon 10:e36606. https://doi.org/10.1016/j.heliyon.2024.e36606

Gad M, Gaagai A, Eid MH, et al. (2023). Groundwater quality and health risk assessment using indexing approaches, multivariate statistical analysis, artificial neural networks, and GIS techniques in El Kharga Oasis, Egypt. Water 15:1216. https://doi.org/10.3390/w15061216

Gad M. (2020b). Geochemical controlling mechanisms and quality of the groundwater resources in

El Fayoum Depression, Egypt. Arab J Geosci 13:5882. https://doi.org/10.1007/s12517-020-05882-x

Galal H, Elsayed S, Elsherbiny O, et al. (2022). Using RGB imaging, optimized three-band spectral indices, and a decision tree model to assess orange fruit quality. Agriculture 12:1558. https://doi.org/10.3390/agriculture12101558

Gibbs RJ (1970). Mechanisms controlling world water chemistry. Science 170(3962):1088-1090. https://doi.org/10.1126/science.170.3962.108

Gao Q, He G, Fang H, et al. (2018). Numerical simulation of water age and its potential effects on the water quality in Xiangxi Bay of Three Gorges Reservoir. J Hydrol 566:484–499. https://doi.org/10.1016/j.jhydrol.2018.09.033

Garau G, Castaldi P, Santona L, et al. (2007). Influence of red mud, zeolite and lime on heavy metal immobilization, culturable heterotrophic microbial populations and enzyme activities in a contaminated soil. Geoderma 142:47–57. https://doi.org/10.1016/j.geoderma.2007.07.011

Ghosh A, Adhikary PP, Bera B, et al. (2022). Assessment of groundwater potential zone using MCDA and AHP techniques: Case study from a tropical river basin of India. Appl Water Sci 12:147. https://doi.org/10.1007/s13201-021-01548-5

González-Feijoo R, Santás-Miguel V, Arenas-Lago D, et al. (2024). Effectiveness of cork and pine bark powders as biosorbents for potentially toxic elements present in aqueous solution. Environ Res 250:118455. https://doi.org/10.1016/j.envres.2024.118455

Güler C, Thyne G, McCray J, et al. (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol J 10:455–474. https://doi.org/10.1007/s10040-002-0196-6

Güler C, Thyne GD. (2004). Hydrologic and geologic factors controlling surface and groundwater chemistry in Indian Wells–Owens Valley area, southeastern California, USA. J Hydrol 285:177–198. https://doi.org/10.1016/j.jhydrol.2003.08.019

Guo T, Gill D, Johengen TH, Cardinale BL. (2019). What determines the public's support for water quality regulations to mitigate agricultural runoff? Environ Sci Policy 101:323–330. https://doi.org/10.1016/j.envsci.2019.09.008

Guzzi G, Ronchi A, Pigatto P. (2021). Toxic effects of mercury in humans and mammals. Chemosphere 263:127990. https://doi.org/10.1016/j.chemosphere.2020.127990

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

Hama Aziz KH, Mustafa FS, Omer KM, et al. (2023). Heavy metal pollution in the aquatic environment: Efficient and low-cost removal approaches to eliminate their toxicity: A review. RSC Adv 13:17595–17610. https://doi.org/10.1039/d3ra00723e

Hanson BR, Grattan SR. (1991). Irrigation and drainage. In: Remote Sensing in Hydrology and Water Management. Chapter 17. Springer. pp. 377-400. https://doi.org/10.1007/978-3-642-59583-7_17

Harris CR, Millman KJ, van der Walt SJ, et al. (2020). Array programming with NumPy. Nature 585:357–362. https://doi.org/10.1038/s41586-020-2649-2

Hassan HS, Abdel Moamen OA, Zaher WF. (2020). Adaptive neuro-fuzzy inference system analysis on sorption studies of strontium and cesium cations onto a novel impregnated nano-zeolite. Adv Powder Technol 31:1125–1139. https://doi.org/10.1016/j.apt.2019.12.031

Helaluddin A, Khalid R, Alaama M. (2016). Main analytical techniques used for elemental analysis in various matrices. Trop J Pharm Res 15:427–434. https://doi.org/10.4314/tjpr.v15i2.29

Helena B, Pardo R, Vega M, et al. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Res 34:807–816. https://doi.org/10.1016/s0043-1354(99)00225-0

Hinge G, Bharali B, Baruah A, Sharma A. (2022). Integrated groundwater quality analysis using Water Quality Index, GIS and multivariate technique: A case study of Guwahati City. Environ Earth Sci 81:10544. https://doi.org/10.1007/s12665-022-10544-0

Hodson ME. (2013). Effects of heavy metals and metalloids on soil organisms. In: heavy metals in Soils. Chapter 5.

Springer, pp. 141–160. https://doi.org/10.1007/978-94-007-4470-7_5

Hoque MMM, Sarker A, Sarker ME, et al. (2023). Heavy metals in sediments of an urban river at the vicinity of tannery industries in Bangladesh: A preliminary study for ecological and human health risk. Int J Environ Anal Chem 103:7909–7927. https://doi.org/10.1080/03067319.2021.1977288

Huang M, Wang W, Chan CY, et al. (2014). Contamination and risk assessment (based on bioaccessibility via ingestion and inhalation) of metal(loid)s in outdoor and indoor particles from urban centers of Guangzhou, China. Sci Total Environ 479–480:117–124. https://doi.org/10.1016/j.scitotenv.2014.01.115

Ibrahim Mohamed A, Osman Mohamed A, Mohamed Ali O. (2019). Effect of sulphur application and water salinity on soil and plant properties. J Soil Sci Environ Manag 10:29–38. https://doi.org/10.5897/jssem2018.0625

Imran U, Ullah A, Shaikh K, et al. (2019). Health risk assessment of the exposure of heavy metal contamination in surface water of lower Sindh, Pakistan. SN Appl Sci 1:594. https://doi.org/10.1007/s42452-019-0594-1

Irshaid M, Abu-Eisheh S. (2023). Application of adaptive neuro-fuzzy inference system in modelling home-based trip generation. Ain Shams Eng J 14:102523. https://doi.org/10.1016/j.asej.2023.102523

Islam ARMT, Pal SC, Chakrabortty R, et al. (2022). A coupled novel framework for assessing vulnerability of water resources using hydrochemical analysis and data-driven models. J Clean Prod 336:130407. https://doi.org/10.1016/j.jclepro.2022.130407

Izah SC, Inyang IR, Angaye TCN, Okowa IP. (2017). A review of heavy metal concentration and potential health implications of beverages consumed in Nigeria. Toxics 5:1–15. https://doi.org/10.3390/toxics5010001

Jacob SE, Goldenberg A, Pelletier JL, et al. (2015). Nickel allergy and our children’s health: A review of indexed cases and a view of future prevention. Pediatr Dermatol 32:779–785. https://doi.org/10.1111/pde.12639

Jacobson AD, Wasserburg GJ. (2005). Anhydrite and the Sr isotope evolution of groundwater in a carbonate aquifer. Chem Geol 214:331–350. https://doi.org/10.1016/j.chemgeo.2004.10.006

Jain CK, Sharma SK, Singh S. (2018). Physico-chemical characteristics and hydrogeological mechanisms in groundwater with special reference to arsenic contamination in Barpeta District, Assam (India). Environ Monit Assess 190:417. https://doi.org/10.1007/s10661-018-6781-5

Jang JSR, Sun CT, Mizutani E. (2005). Neuro-fuzzy and soft computing—A computational approach to learning and machine intelligence [Book review]. IEEE Trans Autom Control 42:1482–1484. https://doi.org/10.1109/tac.1997.633847

Jang J-SR. (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23:665-685. https://ieeexplore.ieee.org/document/256541

Jehan S, Ullah I, Khan S, et al. (2020). Evaluation of the Swat River, Northern Pakistan, water quality using multivariate statistical techniques and water quality index (WQI) model. Environ Sci Pollut Res 27:38545–38558. https://doi.org/10.1007/s11356-020-09688-y

Jia R, Zhang Y, Wang Y, et al. (2024). Toxic effects on ciliates under nano-/micro-plastics coexist with silver nanoparticles. J Hazard Mater 465:133058. https://doi.org/10.1016/j.jhazmat.2023.133058

Kabakuş N, Tortum A. (2019). Comparative analysis of trip generation models according to household characteristics for developed, developing and non-developed provinces in Turkey. Sadhanā Acad Proc Eng Sci 44:122. https://doi.org/10.1007/s12046-019-1104-2

Kacmaz H. (2020). Assessment of heavy metal contamination in natural waters of Dereli, Giresun: an area containing mineral deposits in northeastern Turkey. Environ Monit Assess 192:8057. https://doi.org/10.1007/s10661-019-8057-0

Karunanidhi D, Aravinthasamy P, Deepali M, et al. (2021a). Groundwater pollution and human health risks in an industrialized region of Southern India: Impacts of the COVID-19 lockdown and the monsoon seasonal cycles. Arch Environ Contam Toxicol 80:259–276. https://doi.org/10.1007/s00244-020-00797-w

Karunanidhi D, Subramani T, Roy PD, Li H. (2021b). Impact of groundwater contamination on human health. Environ Geochem Health 43:643–647. https://doi.org/10.1007/s10653-021-00824-2

Kawo NS, Karuppannan S. (2018). Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia. J Afr Earth Sci 147:300–311.

https://doi.org/10.1016/j.jafrearsci.2018.06.034

Khadr M, Elshemy M. (2017). Data-driven modeling for water quality prediction: Case study of the drains system associated with Manzala Lake, Egypt. Ain Shams Eng J 8:549–557. https://doi.org/10.1016/j.asej.2016.08.004

Khadr M, Gad M, El-Hendawy S, et al. (2020). The integration of multivariate statistical approaches, hyperspectral reflectance, and data-driven modeling for assessing the quality and suitability of groundwater for irrigation. Water 13:35. https://doi.org/10.3390/w13010035

Khadr M, Schlenkhoff A. (2021). GA-based implicit stochastic optimization and RNN-based simulation for deriving multi-objective reservoir hedging rules. Environ Sci Pollut Res 28:19107–19120. https://doi.org/10.1007/s11356-020-12291-w

Khouni I, Louhichi G, Ghrabi A. (2021). Use of GIS-based Inverse Distance Weighted interpolation to assess surface water quality: Case of Wadi El Bey, Tunisia. Environ Technol Innov 24:101892. https://doi.org/10.1016/j.eti.2021.101892

Klauber C, Gräfe M, Power G. (2011). Bauxite residue issues: II. Options for residue utilization. Hydrometallurgy 108:11–32. https://doi.org/10.1016/j.hydromet.2011.02.007

Kobylinska N, Kostenko L, Khainakov S, Garcia-Granda S. (2020). Advanced core-shell EDTA-functionalized magnetite nanoparticles for rapid and efficient magnetic solid phase extraction of heavy metals from water samples prior to the multi-element determination by ICP-OES. Microchim Acta 187:423. https://doi.org/10.1007/s00604-020-04231-9

Korkmaz C, Correia AP. (2019). A review of research on machine learning in educational technology. Educ Media Int 56:250–267. https://doi.org/10.1080/09523987.2019.1669875

Kovács-Bokor É, Domokos E, Biró B. (2021). Toxic metal phytoextraction potential and health-risk parameters of some cultivated plants when grown in metal-contaminated river sediment of Danube, near an industrial town. Environ Geochem Health 43:2317–2330. https://doi.org/10.1007/s10653-021-00880-8

Kumar A, Palmate SS, Shukla R. (2022a). Water quality modelling, monitoring, and mitigation. Appl Sci 12:11403. https://doi.org/10.3390/app122211403

Kumar A, Subrahmanyam G, Mondal R, et al. (2021). Bio-remediation approaches for alleviation of cadmium contamination in natural resources. Chemosphere 268:128855. https://doi.org/10.1016/j.chemosphere.2020.128855

Kumar M, Kumari K, Ramanathan A, Saxena R. (2007). A comparative evaluation of groundwater suitability for irrigation and drinking purposes in two intensively cultivated districts of Punjab, India. Environ Geol 53:553–574. https://doi.org/10.1007/s00254-007-0672-3

Kumar P, Mishra V, Yadav S, et al. (2022b). Heavy metal pollution and risks in a highly polluted and populated Indian river–city pair using the systems approach. Environ Sci Pollut Res 29:60212–60231. https://doi.org/10.1007/s11356-022-20034-2

Kumar S, Kumar A, Prashant, et al. (2022c). Groundwater quality and its suitability for drinking and irrigational purpose in Bhojpur district: middle Gangetic plain of Bihar, India. Water Supply 22:7072–7084. https://doi.org/10.2166/ws.2022.317

Kumaraguru K, Saravanan P, Rajesh Kannan R, Saravanan V. (2023). A systematic analysis of hexavalent chromium adsorption and elimination from aqueous environment using brown marine algae (Turbinaria ornata). Biomass Convers Biorefin 13:8223–8238. https://doi.org/10.1007/s13399-021-01795-1

Li H, Sun J, Zhou Q, et al. (2024). 150-year daily data (1870–2021) in lakes and rivers reveals intensifying surface water warming and heatwaves in the Pannonian Ecoregion (Hungary). J Hydrol Reg Stud 56:101985. https://doi.org/10.1016/j.ejrh.2024.101985

Li J, Heap AD. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecol Inform 6:228–241. https://doi.org/10.1016/j.ecoinf.2010.12.003

Li J, Xie H, Li J, et al. (2023). Influences of anthropogenic acids on carbonate weathering and CO2 sink in an agricultural karst wetland (South China). Ecol Indic 150:110192. https://doi.org/10.1016/j.ecolind.2023.110192

Li P, Wu J, Qian H. (2013). Assessment of groundwater quality for irrigation purposes and identification of hydrogeochemical evolution mechanisms in Pengyang County, China. Environ Earth Sci 69:2211–2225. https://doi.org/10.1007/s12665-012-2049-5

Li P, Wu J, Qian H. (2016). Preliminary assessment of hydraulic connectivity between river water and shallow groundwater and estimation of their transfer rate during dry season in the Shidi. Environ Earth Sci 75:99. https://doi.org/10.1007/s12665-015-4949-7

Li S, Zhang Q. (2010). Risk assessment and seasonal variations of dissolved trace elements and heavy metals in the Upper Han River, China. J Hazard Mater 181:1051–1058. https://doi.org/10.1016/j.jhazmat.2010.05.120

Lima VR do N, Silva ÁGF da, Cruz RRP, et al. (2021). Nopalea cochenillifera biomass as bioadsorbent in water purification. Water 13:2012. https://doi.org/10.3390/w13152012

Liu CW, Lin KH, Kuo YM. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci Total Environ 313:77–89. https://doi.org/10.1016/s0048-9697(02)00683-6

Liu F, Wang S, Yeh TCJ, Zhen P, Wang L, Shi L. (2020). Using multivariate statistical techniques and geochemical modelling to identify factors controlling the evolution of groundwater chemistry in a typical transitional area. Hydrol Process 34:1888–1905. https://doi.org/10.1002/hyp.13701

Liu Y, Theller LO, Pijanowski BC, Engel BA. (2016). Optimal selection and placement of green infrastructure to reduce impacts of land use change and climate change on hydrology and water quality: An application to the Trail Creek Watershed, Indiana. Sci Total Environ 553:149–163. https://doi.org/10.1016/j.scitotenv.2016.02.116

Loucks DP, van Beek E. (2017). Water quality modeling and prediction. In: Water Resource Systems Planning and Management. Chapter 10. pp. 417-467. https://doi.org/10.1007/978-3-319-44234-1_10

Ma J, Ding Y, Cheng JCP, et al. (2020). Soft detection of 5-day BOD with sparse matrix in city harbor water using deep learning techniques. Water Res 170:115350. https://doi.org/10.1016/j.watres.2019.115350

Maghrebi M, Noori R, Mehr AD, et al. (2023). Spatiotemporal changes in Iranian rivers' discharge. Elementa 11:00002. https://doi.org/10.1525/elementa.2022.00002

Maghrebi M, Noori R, Partani S, Araghi A, Barati R, Farnoush H, Torabi Haghighi A. (2021). Iran's groundwater hydrochemistry. Earth Space Sci 8:001793. https://doi.org/10.1029/2021EA001793

Mahapatra DM, Satapathy KC, Panda B. (2022). Biofertilizers and nanofertilizers for sustainable agriculture: Phycoprospects and challenges. Sci Total Environ 803:149990. https://doi.org/10.1016/j.scitotenv.2021.149990

Mahato MK, Singh PK, Tiwari AK, Singh AK. (2016). Risk assessment due to intake of metals in groundwater of east Bokaro coalfield, Jharkhand, India. Expo Health 8:265–275. https://doi.org/10.1007/s12403-016-0201-2/tables/6

Mayes WM, Burke IT, Gomes HI, et al. (2016). Advances in understanding environmental risks of red mud after the Ajka spill, Hungary. J Sustain Metall 2:332–343. https://doi.org/10.1007/s40831-016-0050-z

McGeorge WT. (1954). Diagnosis and improvement of saline and alkaline soils. Soil Sci Soc Am J 18:348–348. https://doi.org/10.2136/sssaj1954.03615995001800030032x

McKinney W. (2010). Data structures for statistical computing in Python. Proc. 9th Python in Science Conf. (SCIPY 2010:56-61. https://doi.org/10.25080/majora-92bf1922-00a

Meireles ACM, Andrade EM de, Chaves LCG, et al. (2010). A new proposal of the classification of irrigation water. Rev Ciênc Agron 41:349–357. https://doi.org/10.1590/s1806-66902010000300005

Mhamdi A, Moussaoui I, Khalfi C, et al. (2022). Assessment of groundwater mineralization processes in the Garâat Douza-Tebeddit shallow aquifer, Southwestern Tunisia: An integrated geoelectrical and hydrochemical approach. Carbonates Evaporites 37:781. https://doi.org/10.1007/s13146-022-00781-6

Mohammadi AA, Zarei A, Majidi S, et al. (2019). Carcinogenic and non-carcinogenic health risk assessment of heavy metals in drinking water of Khorramabad, Iran. MethodsX 6:1642–1651. https://doi.org/10.1016/j.mex.2019.07.017

Mohammed MAA, Kaya F, Mohamed A, et al. (2023). Application of GIS-based machine learning algorithms for prediction of irrigational groundwater quality indices. Front Earth Sci 11:1274142. https://doi.org/10.3389/feart.2023.1274142

Mollo VM, Nomngongo PN, Ramontja J. (2022). Evaluation of surface water quality using various indices for heavy metals in Sasolburg, South Africa. Water 14:2375. https://doi.org/10.3390/w14152375

Montes C, Kapelan Z, Saldarriaga J. (2021). Predicting non-deposition sediment transport in sewer pipes using Random Forest. Water Res 189:116639. https://doi.org/10.1016/j.watres.2020.116639

Moyé J, Picard-Lesteven T, Zouhri L, et al. (2017). Groundwater assessment and environmental impact in the abandoned mine of Kettara (Morocco). Environ Pollut 231:899–907. https://doi.org/10.1016/j.envpol.2017.07.044

Mtetwa S, Kusangaya S, Schutte CF. (2003). The application of geographic information systems (GIS) in the analysis of nutrient loadings from an agro-rural catchment. Water SA 29:189–194. https://doi.org/10.4314/wsa.v29i2.4855

Mthembu PP, Elumalai V, Li P, et al. (2022). Integration of heavy metal pollution indices and health risk assessment of groundwater in semi-arid coastal aquifers, South Africa. Expo Health 14:487–502. https://doi.org/10.1007/s12403-022-00478-0

Muhammad S, Ullah S, Ali W, et al. (2022). Spatial distribution of heavy metal and risk indices of water and sediments in the Kunhar River and its tributaries. Geocarto Int 37:5985–6003. https://doi.org/10.1080/10106049.2021.1926557

Muhammad S, Usman QA. (2022). Heavy metal contamination in water of Indus River and its tributaries, Northern Pakistan: Evaluation for potential risk and source apportionment. Tox Rev 41:380–388. https://doi.org/10.1080/15569543.2021.1882499

Mukherjee I, Singh UK, Singh RP, et al. (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

Mutea FG, Nelson HK, Van Au H, et al. (2021). Assessment of water quality for aquaculture in Hau River, Mekong Delta, Vietnam using multivariate statistical analysis. Water 13:3307. https://doi.org/10.3390/w13223307

Nawar S, Cipullo S, Douglas R, et al. (2019). The applicability of spectroscopy methods for estimating potentially toxic elements in soils: State-of-the-art and future trends. Appl Spectrosc Rev 55:525–557. https://doi.org/10.1080/05704928.2019.1608110

Obasi PN, Akudinobi BB. (2020). Potential health risk and levels of heavy metals in water resources of lead–zinc mining communities of Abakaliki, southeast Nigeria. Appl Water Sci 10:1–23. https://doi.org/10.1007/s13201-020-01233-z

Obropta C, Niazi M, Kardos JS. (2008). Application of an environmental decision support system to a water quality trading program affected by surface water diversions. Environ Manage 42:946–956. https://doi.org/10.1007/s00267-008-9153-z

Ojekunle OZ, Ojekunle OV, Adeyemi AA, et al. (2016). Evaluation of surface water quality indices and ecological risk assessment for heavy metals in scrap yard neighbourhood. SpringerPlus 5:2158. https://doi.org/10.1186/s40064-016-2158-9

Panno SV, Hackley KC, Hwang HH, et al. (2002). Source identification of sodium and chloride contamination in natural waters: Preliminary results. Proc 12th Annu Illinois Groundwater Consort Symp. 1-26. https://hdl.handle.net/2142/55186

Pant D, Keesari T, Rishi MS, et al. (2020). Hydrochemical evolution of groundwater in the waterlogged area of southwest Punjab. Arab J Geosci 13:579. https://doi.org/10.1007/s12517-020-05795-9

Parkhurst DL, Appelo CAJ. (1999). User's guide to PHREEQC (Version 2): A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. Water-Resources Investigations Report 99-4259. https://doi.org/10.3133/wri994259

Pathak D, Maharjan R, Maharjan N, et al. (2021). Evaluation of parameter sensitivity for groundwater potential mapping in the mountainous region of Nepal Himalaya. Groundw Sustain Dev 13:100562. https://doi.org/10.1016/j.gsd.2021.100562

Patton H, Krometis LA, Sarver E. (2020). Springing for Safe Water: Drinking Water Quality and Source Selection in Central Appalachian Communities. Water 12:888. https://doi.org/10.3390/w12030888

Piper AM. (1944). A graphic procedure in the geochemical interpretation of water-analyses. Trans Am Geophys Union 25:914–928.

Power G, Gräfe M, Klauber C. (2011). Bauxite residue issues: I. Current management, disposal and storage practices. Hydrometallurgy 108:33–45. https://doi.org/10.1016/j.hydromet.2011.02.006

Prasad A, Kumar D, Singh DV. (2001). Effect of residual sodium carbonate in irrigation water on the soil sodication and yield of palmarosa (Cymbopogon martinii) and lemongrass (Cymbopogon flexuosus). Agric Water Manag 50:161–172. https://doi.org/10.1016/s0378-3774(01)00103-2

Programme United Nations Environment. (2016). A snapshot of the world's water quality: Towards a global assessment. Nairobi, Kenya: UNEP.

Proshad R, Islam S, Tusher TR, et al. (2021). Appraisal of heavy metal toxicity in surface water with human health risk by a novel approach: A study on an urban river in vicinity to industrial areas of Bangladesh. Tox Rev 40:803–819. https://doi.org/10.1080/15569543.2020.1780615

Qian C, Wu X, Mu WP, Zhang Y, Li J. (2016). Hydrogeochemical characterization and suitability assessment of groundwater in an agro-pastoral area, Ordos Basin, NW China. Environ Earth Sci 75:6123. https://doi.org/10.1007/s12665-016-6123-2

Qiao P, Lei M, Yang S, et al. (2018). Comparing ordinary kriging and inverse distance weighting for soil As pollution in Beijing. Environ Sci Pollut Res 25:15597–15608. https://doi.org/10.1007/s11356-018-1552-y

Qiu D, Zhu G, Bhat MA, et al. (2023). Water use strategy of Nitraria tangutorum shrubs in ecological water delivery area of the lower inland river: Based on stable isotope data. J Hydrol 624:129918. https://doi.org/10.1016/j.jhydrol.2023.129918

Qu J, Niu H, Wang J, et al. (2021). Potential mechanism of lead poisoning to the growth and development of ovarian follicle. Toxicol 457:152810. https://doi.org/10.1016/j.tox.2021.152810

Qu L, Huang H, Xia F, et al. (2018). Risk analysis of heavy metal concentration in surface waters across the rural-urban interface of the Wen-Rui Tang River, China. Environ Pollut 237:639–649. https://doi.org/10.1016/j.envpol.2018.02.020

Qu L, Xiao H, Zheng N, Zhang Z, Xu Y. (2017). Comparison of four methods for spatial interpolation of estimated atmospheric nitrogen deposition in South China. Environ Sci Pollut Res 24:2578–2588. https://doi.org/10.1007/s11356-016-7995-0

Rahman M, Rima SA, Saha SK, et al. (2022). Pollution evaluation and health risk assessment of heavy metals in the surface water of a remote island Nijhum Dweep, northern Bay of Bengal. Environ Nanotechnol Monit Manage 18:100706. https://doi.org/10.1016/j.enmm.2022.100706

Rakotondrabe F, Ndam Ngoupayou JR, Mfonka Z, et al. (2018). Water quality assessment in the Bétaré-Oya gold mining area (East-Cameroon): Multivariate statistical analysis approach. Sci Total Environ 610–611:831–844. https://doi.org/10.1016/j.scitotenv.2017.08.080

RamyaPriya R, Elango L. (2018). Evaluation of geogenic and anthropogenic impacts on spatio-temporal variation in quality of surface water and groundwater along Cauvery River, India. Environ Earth Sci 77:7176. https://doi.org/10.1007/s12665-017-7176-6

Ravikumar P, Aneesul Mehmood M, Somashekar RK. (2013a). Water quality index to determine the surface water quality of Sankey Tank and Mallathahalli Lake, Bangalore urban district, Karnataka, India. Appl Water Sci 3:247–261. https://doi.org/10.1007/s13201-013-0077-2

Ravikumar P, Aneesul Mehmood M, Somashekar RK. (2013b). Water quality index to determine the surface water quality of Sankey Tank and Mallathahalli Lake, Bangalore urban district, Karnataka, India. Appl Water Sci 3:247–261. https://doi.org/10.1007/s13201-013-0077-2

Ravindra K, Mor S. (2019). Distribution and health risk assessment of arsenic and selected heavy metals in groundwater of Chandigarh, India. Environ Pollut 250:820–830. https://doi.org/10.1016/j.envpol.2019.03.080

Rezaei H, Zarei A, Kamarehie B, et al. (2019). Levels, distributions and health risk assessment of lead, cadmium and arsenic found in drinking groundwater of Dehgolan's villages, Iran. Toxicol Environ Health Sci 11:54–62. https://doi.org/10.1007/s13530-019-0388-2

Rezaei T, Javadi A. (2024). Environmental impact assessment of ocean energy converters using quantum machine learning. J Environ Manage 362:121275. https://doi.org/10.1016/j.jenvman.2024.121275

Rice EW, Baird RB, Eaton AD, Clesceri LS (Eds.) (2012). Standard methods for the examination of water and wastewater. 22nd edition. American Public Health Association, Washington, D.C., USA.

Rupakheti D, Tripathee L, Kang S, et al. (2017). Assessment of water quality and health risks for toxic trace elements in urban Phewa and remote Gosainkunda lakes, Nepal. Hum Ecol Risk Assess 23:959–973. https://doi.org/10.1080/10807039.2017.1292117

Saeed O, Székács A, Jordán G, et al. (2023a). Investigating the impacts of heavy metal(loid)s on ecology and human health in the lower basin of Hungary’s Danube River: A Python and Monte Carlo simulation-based study. Environ Geochem Health 45:9757–9784. https://doi.org/10.1007/s10653-023-01769-4

Saeed O, Székács A, Jordán G, et al. (2023b). Correction: Investigating the impacts of heavy metal(loid)s on ecology and human health in the lower basin of Hungary’s Danube River: A Python and Monte Carlo simulation-based study. Environ Geochem Health 45:9785. https://doi.org/10.1007/s10653-023-01777-4

Saeed O, Székács A, Jordán G, et al. (2024). Assessing surface water quality in Hungary’s Danube basin using geochemical modeling, multivariate analysis, irrigation indices, and Monte Carlo simulation. Sci Rep 14:69312. https://doi.org/10.1038/s41598-024-69312-8

Saha P, Paul B. (2019). Assessment of heavy metal toxicity related with human health risk in the surface water of an industrialized area by a novel technique. Hum Ecol Risk Assess 25:966–987. https://doi.org/10.1080/10807039.2018.1458595

Saleem M, Iqbal J, Shah MH. (2019). Seasonal variations, risk assessment and multivariate analysis of trace metals in the freshwater reservoirs of Pakistan. Chemosphere 216:715–724. https://doi.org/10.1016/j.chemosphere.2018.10.173

Saleh A, Al-Ruwaih F, Shehata M. (1999). Hydrogeochemical processes operating within the main aquifers of Kuwait. J Arid Environ 42:195–209. https://doi.org/10.1006/jare.1999.0511

Saleh AH, Elsayed S, Gad M, et al. (2022). Utilization of pollution indices, hyperspectral reflectance indices, and data-driven multivariate modelling to assess the bottom sediment quality of Lake Qaroun, Egypt. Water 14:890. https://doi.org/10.3390/w14060890/s1

Salem SBH, Gaagai A, Ben Slimene I, et al. (2023). Applying multivariate analysis and machine learning approaches to evaluating groundwater quality on the Kairouan Plain, Tunisia. Water 15:3495. https://doi.org/10.3390/w15193495

Samantara MK, Padhi RK, Sowmya M, et al. (2017). Heavy metal contamination, major ion chemistry and appraisal of the groundwater status in coastal aquifer, Kalpakkam, Tamil Nadu, India. Groundw Sustain Dev 5:49–58. https://doi.org/10.1016/j.gsd.2017.04.001

Saravanan P, Saravanan V, Rajeshkannan R, et al. (2024). Comprehensive review on toxic heavy metals in the aquatic system: Sources, identification, treatment strategies, and health risk assessment. Environ Res 258:119440. https://doi.org/10.1016/j.envres.2024.119440

Sarker IH. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Comput Sci 2:160. https://doi.org/10.1007/s42979-021-00592-x

Sarojini G, Venkatesh Babu S, Rajamohan N, et al. (2021). Surface modified polymer-magnetic-algae nanocomposite for the removal of chromium: Equilibrium and mechanism studies. Environ Res 201:111626. https://doi.org/10.1016/j.envres.2021.111626

Selvam S, Jesuraja K, Roy PD, et al. (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

Shammi RS, Hossain MS, Kabir MH, et al. (2023). Hydrochemical appraisal of surface water from a subtropical urban river in southwestern Bangladesh using indices, GIS, and multivariate statistical analysis. Environ Sci Pollut Res 30:3467–3489. https://doi.org/10.1007/s11356-022-22384-3

Shankar BS. (2019). A critical assay of heavy metal pollution index for the groundwaters of Peenya Industrial Area, Bangalore, India. Environ Monit Assess 191.289. https://doi.org/10.1007/s10661-019-7453-9

Sharma S, Nagpal AK, Kaur I. (2019). Appraisal of heavy metal contents in groundwater and associated health hazards posed to human population of Ropar wetland, Punjab, India and its environs. Chemosphere 227:179–190. https://doi.org/10.1016/j.chemosphere.2019.04.009

Shaw SK, Sharma A. (2025). Evaluation and prediction of groundwater quality for irrigation using regression and machine learning models. Water Qual Res J 60:260–297.

https://doi.org/10.2166/wqrj.2025.075

Sheng D, Wen X, Wu J, et al. (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

Shrestha S, Kazama F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. Environ Model Softw 22:464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Siddique MAB, Khan R, Islam ARM T, et al. (2021). Quality assessment of freshwaters from a coastal city of southern Bangladesh: Irrigation feasibility and preliminary health risks appraisal. Environ Nanotechnol Monit Manage 16:100524. https://doi.org/10.1016/j.enmm.2021.100524

Siebert S, Burke J, Faures JM, et al. (2010). Groundwater use for irrigation: A global inventory. Hydrol Earth Syst Sci 14:1863–1880. https://doi.org/10.5194/hess-14-1863-2010

Simeonov V, Stratis JA, Samara C, et al. (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 G, Singh J, Wani OA, et al. (2024). Assessment of groundwater suitability for sustainable irrigation: A comprehensive study using indexical, statistical, and machine learning approaches. Groundw Sustain Dev 24:101059. https://doi.org/10.1016/j.gsd.2023.101059

Singh S, Ghosh NC, Gurjar S, et al. (2018). Index-based assessment of suitability of water quality for irrigation purposes under Indian conditions. Environ Monit Assess 190:1–14. https://doi.org/10.1007/s10661-017-6407-3

Siyal AA, Shamsuddin MR, Khan MI, et al. (2018). A review on geopolymers as emerging materials for the adsorption of heavy metals and dyes. J Environ Manage 224:327–339. https://doi.org/10.1016/j.jenvman.2018.07.046

Snousy MG, Alexakis DE, Gamvroula DE, et al. (2023). Multivariate statistical analysis and geospatial mapping for assessing groundwater quality in West El Minia District, Egypt. Water 15:2909. https://doi.org/10.3390/w15162909

Song H, Kumar A, Ding Y, et al. (2022). Removal of Cd^{2+} from wastewater by microorganism induced carbonate precipitation (MICP): An economic bioremediation approach. Sep Purif Technol 297:121540. https://doi.org/10.1016/j.seppur.2022.121540

Srinivasamoorthy K, Gopinath M, Chidambaram S, et al. (2014). Hydrochemical characterization and quality appraisal of groundwater from Pungar sub-basin, Tamil Nadu, India. J King Saud Univ Sci 26:37–52. https://doi.org/10.1016/j.jksus.2013.08.001

Stenchly K, Dao J, Lompo DJP, Buerkert A. (2017). Effects of wastewater irrigation on soil properties and soil fauna of spinach fields in a West African urban vegetable production system. Environ Pollut 222:58–63. https://doi.org/10.1016/j.envpol.2017.01.006

Sun XL, Kido T, Nakagawa H, et al. (2021). The relationship between cadmium exposure and renal volume in inhabitants of a cadmium-polluted area of Japan. Environ Sci Pollut Res 28:22372–22379. https://doi.org/10.1007/s11356-020-12278-7

Swain CK. (2024). Environmental pollution indices: A review on concentration of heavy metals in air, water, and soil near industrialization and urbanisation. Discover Environ 2:5. https://doi.org/10.1007/s44274-024-00030-8

Szabó Z, Buró B, Szabó J, et al. (2020). Geomorphology as a driver of heavy metal accumulation patterns in a floodplain. Water 12:563. https://doi.org/10.3390/w12020563

Tao Z, Guo Q, Wei R, et al. (2021). Atmospheric lead pollution in a typical megacity: Evidence from lead isotopes. Sci Total Environ 778:145810. https://doi.org/10.1016/j.scitotenv.2021.145810

Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ. (2012). Heavy metal toxicity and the environment. EXS 101:133–164. https://doi.org/10.1007/978-3-7643-8340-4_6

The European Parliament and the Council of the European Union. (2020). Directive (EU) 2020/2184 of the European Parliament and of the Council. Off J Eur Union 2019:1–62.

Thilagavathi R, Chidambaram S, Prasanna MV, et al. (2012). A study on groundwater geochemistry and water quality in layered aquifer systems of Pondicherry region, southeast India. Appl Water Sci 2:253–269. https://doi.org/10.1007/s13201-012-0045-2

Tiwari AK, De Maio M, Singh PK, Mahato MK. (2015). Evaluation of surface water quality by using GIS and a heavy metal pollution index (HPI) model in a coal mining area, India. Bull Environ Contam Toxicol 95:304–310. https://doi.org/10.1007/s00128-015-1558-9

Todd, D. K., & Mays, L. W. (2005). *Groundwater hydrology* (2nd ed.). John Wiley & Sons. 636 pp.

Tokatli C, Ustaoğlu F. (2020). Health risk assessment of toxicants in Meriç River Delta Wetland, Thrace Region, Turkey. Environ Earth Sci 79:426. https://doi.org/10.1007/s12665-020-09171-4

Uddin G, Talas M, Diganta M, et al. (2023). Assessment of hydrogeochemistry in groundwater using water quality index model and indices approaches. Heliyon 9:e19668. https://doi.org/10.1016/j.heliyon.2023.e19668

USEPA. (2004). Risk assessment guidance for Superfund (RAGS). Volume I. Human health evaluation manual (HHEM). Part E: Supplemental guidance for dermal risk assessment. U.S. Environmental Protection Agency. https://doi.org/EPA/540/1-89/002

Uzinger N, Anton ÁD, Ötvös K, et al. (2015). Results of the clean-up operation to reduce pollution on flooded agricultural fields after the red mud spill in Hungary. Environ Sci Pollut Res Int 22:9849–9857. https://doi.org/10.1007/s11356-015-4158-7

Varol S, Şekerci M. (2018). Hydrogeochemistry, water quality and health risk assessment of water resources contaminated by agricultural activities in Korkuteli (Antalya, Turkey) district center. J Water Health 16:574–599. https://doi.org/10.2166/wh.2018.003

Vasanthavigar M, Srinivasamoorthy K, Prasanna MV, et al. (2012). Evaluation of groundwater suitability for domestic, irrigational, and industrial purposes: A case study from Thirumanimuttar River Basin, Tamil Nadu, India. Environ Monit Assess 184:405–420. https://doi.org/10.1007/s10661-011-1977-y

Vasanthavigar M, Srinivasamoorthy K, Prasanna MV. (2013). Identification of groundwater contamination zones and its sources by using multivariate statistical approach in Thirumanimuthar sub-basin, Tamil Nadu, India. Environ Earth Sci 68:1783–1795. https://doi.org/10.1007/s12665-012-1868-8

Vega M, Pardo R, Barrado E, Debán L. (1998). Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Res 32:3581–3592.

https://doi.org/10.1016/S0043-1354(98)00138-9

Veza I, Muhamad Said MF, Abdul Latiff Z, Abas MA. (2021). Application of Elman and Cascade neural network (ENN and CNN) in comparison with adaptive neuro fuzzy inference system (ANFIS) to predict key fuel properties of ABE-diesel blends. Int J Green Energy 18:1510–1522. https://doi.org/10.1080/15435075.2021.1911807

Virtanen P, Gommers R, Oliphant TE, et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nat Methods 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang J, Wang X, Khu ST. (2023). A decomposition-based multi-model and multi-parameter ensemble forecast framework for monthly streamflow forecasting. J Hydrol 618:129083. https://doi.org/10.1016/J.JHYDROL.2023.129083

Wang Q, Li S, Jia P, Qi C, Ding F. (2013). A review of surface water quality models. Sci World J 2013:231768. https://doi.org/10.1155/2013/231768

Waskom ML. (2021). seaborn: Statistical data visualization. J Open Source Softw 6:3021. https://doi.org/10.21105/JOSS.03021

World Health Organization. (2011). Adverse health effects of heavy metals in children. Geneva: WHO.

WHO. (2017). Guidelines for drinking-water quality, 4th edition, incorporating the 1st addendum.

Winkler D. (2014). Collembolan response to red mud pollution in Western Hungary. Appl Soil Ecol 83:219–229. https://doi.org/10.1016/J.APSOIL.2013.07.006

Winkler D, Bidló A, Bolodár-Varga B, et al. (2018). Long-term ecological effects of the red mud disaster in Hungary: Regeneration of red mud flooded areas in a contaminated industrial region. Sci Total Environ 644:1292–1303. https://doi.org/10.1016/j.scitotenv.2018.07.059

Withanachchi SS, Ghambashidze G, Kunchulia I, et al. (2018). Water quality in surface water: A preliminary assessment of heavy metal contamination of the Mashavera river, Georgia. Int J Environ Res Public Health 15:1–25. https://doi.org/10.3390/ijerph15040621

Woitke P, Wellmitz J, Helm D, et al. (2003). Analysis and assessment of heavy metal pollution in suspended solids and sediments of the river Danube. Chemosphere 51:633–642.

https://doi.org/10.1016/S0045-6535(03)00217-0

Wongnakae P, Chitchum P, Sripramong R, Phosri A. (2023). Application of satellite remote sensing data and random forest approach to estimate ground-level PM2.5 concentration in Northern region of Thailand. Environ Sci Pollut Res 30:88905–88917. https://doi.org/10.1007/s11356-023-28698-0

Wu J, Li P, Qian H. (2015). Hydrochemical characterization of drinking groundwater with special reference to fluoride in an arid area of China and the control of aquifer leakage on its concentrations. Environ Earth Sci 73:8575–8588. https://doi.org/10.1007/S12665-015-4018-2

Xie XD, Min XB, Chai LY, Zhang Y. (2013). Quantitative evaluation of environmental risks of flotation tailings from hydrothermal sulfidation-flotation process. Environ Sci Pollut Res 20:6050–6058. https://doi.org/10.1007/s11356-013-1643-8

Xiong B, Li R, Johnson D, et al. (2021). Spatial distribution, risk assessment, and source identification of heavy metals in water from the Xiangxi River, Three Gorges Reservoir Region, China. Environ Geochem Health 43:915–930. https://doi.org/10.1007/S10653-020-00614-2

Xu J, Zheng L, Xu L, et al. (2020). Identification of dissolved metal contamination of major rivers in the southeastern hilly area, China: Distribution, source apportionment, and health risk assessment. Environ Sci Pollut Res 27:3908–3922. https://doi.org/10.1007/s11356-019-06774-8

Yan C, Li Q, Zhang X, Li G. (2010). Mobility and ecological risk assessment of heavy metals in surface sediments of Xiamen Bay and its adjacent areas, China. Environ Earth Sci 60:1469–1479. https://doi.org/10.1007/S12665-009-0282-3

Yu S, He S, Sun P, et al. (2016). Impacts of anthropogenic activities on weathering and carbon fluxes: A case study in the Xijiang River Basin, Southwest China. Environ Earth Sci 75:1–11. https://doi.org/10.1007/S12665-015-5226-5

Yu Y, Qu Y. (2020). Multi-component spectral detection based on neural network in water quality inspection. Optik (Stuttg) 217:164915. https://doi.org/10.1016/J.IJLEO.2020.164915

Yuan X, Leng L, Huang H, et al. (2015). Speciation and environmental risk assessment of heavy metal in bio-oil from liquefaction/pyrolysis of sewage sludge. Chemosphere 120:645–652.

https://doi.org/10.1016/J.CHEMOSPHERE.2014.10.010

Zadeh L. (1989). Knowledge representation in fuzzy logic. IEEE Trans Knowl Data Eng 1:89–100. https://doi.org/10.1109/69.43406

Zhang H, Ren X, Chen S, et al. (2024). Deep optimization of water quality index and positive matrix factorization models for water quality evaluation and pollution source apportionment using a random forest model. Environ Pollut 347:123771. https://doi.org/10.1016/J.ENVPOL.2024.123771

Zhang T, Song B, Han G, et al. (2023). Effects of coastal wetland reclamation on soil organic carbon, total nitrogen, and total phosphorus in China: A meta-analysis. Land Degrad Dev 34:3340–3349. https://doi.org/10.1002/ldr.4687

Zhang, Y., Sillanpää, M., Li, C., et al. (2015). River water quality across the Himalayan regions: Elemental concentrations in headwaters of Yarlung Tsangbo, Indus and Ganges River. Environ. Earth Sci., 73, 4151–4163. https://doi.org/10.1007/S12665-014-3702-Y

Zhang, Z., Guo, Y., Wu, J., & Su, F. (2022). Surface water quality and health risk assessment in Taizhou City, Zhejiang Province (China). Expo. Health, 14:1-16. https://doi.org/10.1007/S12403-021-00408-6

Zubaidi, S. L., Al-Bugharbee, H., Ortega-Martorell, S., et al. (2020). A novel methodology for prediction of urban water demand by wavelet denoising and adaptive neuro-fuzzy inference system approach. Water (Switzerland), 12:1628. https://doi.org/10.3390/W12061628