The Thesis of the Doctoral (Ph.D.) dissertation

OMAR ABDULHAKIM HIZAM SAEED

GÖDÖLLŐ, HUNGARY

2025

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)

1. BACKGROUND OF WORK AND ITS AIMS

Water is essential for human survival, ecosystems, and socioeconomic development, yet over one billion people lack access to safe water, with waterborne diseases causing 6–8 million deaths annually (Proshad et al., 2021; Hoque et al., 2023). Surface water (SW) is a primary freshwater source, while groundwater (GW) is crucial in regions with limited SW (Islam et al., 2022). Anthropogenic activities and natural processes have widely contaminated SW and GW, particularly with heavy metals (HMs) (Muhammad et al., 2022; Shammi et al., 2023). HMs persist in the environment, bioaccumulate, and are toxic: essential metals (Cu, Mn, Fe, Zn) are harmful at high concentrations, while toxic metals (Pb, As) pose serious health risks (Chowdhury et al., 2016; Rezaei et al., 2019; Izah et al., 2017).

Monitoring water quality, identifying pollution sources, and assessing health risks are crucial. The Heavy Metal Pollution Index (HPI) quantifies HM pollution, while cluster analysis (CA) and GIS map spatial patterns. Monte Carlo simulation (MCS) supports risk predictions (Sheng et al., 2021). The Danube River, Europe's second-longest, faces HM pollution from sewage, agriculture, navigation, and mining (Calmuc et al., 2021; Saeed et al., 2023a). Yemen's Al-Jawf GW basin exemplifies an arid, agriculturally dependent system, suffering from overextraction, poor management, and climate-driven aquifer depletion (Siebert et al., 2010; Al-Mashreki et al., 2023). GW quality depends on geological, climatic, and anthropogenic factors, assessed via parameters like pH, EC, TDS, major ions, and indices including SAR, RSC, and PI (Al-Mashreki et al., 2023). Recent advances in AI and machine learning (ML) – e.g., artificial neural networks (ANNs), support vector machines (SVMs), and decision trees – improve predictive assessment of water quality indices (IWQIs) and key pollution drivers (Khadr et al., 2020; Al-Mashreki et al., 2023; Saeed et al., 2023a). Previous studies on the Danube and Al-Jawf focused on either HM contamination or physicochemical properties. Few integrated conventional analyses with AI/ML and risk models. This study addresses this gap, combining geochemical characterization, HPI, MCS-based risk assessment, and predictive models (ANFIS, RF, SVM) to evaluate two contrasting basins.

1.2. Problem Statement

Comprehensive water quality and health risk assessments remain limited in both the Lower Danube (Hungary) and Al-Jawf (Yemen). Most studies are localized, focusing separately on physicochemical or HM contamination, with minimal integration of health risks across

hydrological systems. Industrial and agricultural pressures in the Danube and severe water scarcity in Yemen underscore the need for unified, cross-regional approaches. Advanced tools – HPI, MCS, ANFIS and RF remain underutilized in comparative studies. This research applies an integrated framework combining geochemistry, health risk, and ML modeling to enhance data-driven water management strategies.

1.3. Objectives

The study aims to assess and compare SW in Hungary's lower Danube basin and GW in Yemen's Al-Jawf basin. Objectives include:

- Characterizing water chemistry and geochemical processes via physicochemical parameters and HM concentrations.
- Evaluating irrigation suitability using indices such as IWQI, SAR, SSP, KR, PS, RSC, and RSBC.
- Conducting non-carcinogenic and carcinogenic HM risk assessments using MCS to quantify uncertainty.
- Applying ML algorithms (RF, ANFIS) to predict water quality and health risks.

2. MATERIALS AND METHODS

2.1. Study area description

2.1.1. Danube river lower basin in Hungary (SW)

Surface water was monitored at seven sites along a 120-km stretch (18°57′15.50″ E, 46°10′54.45″ N) from Dunaföldvár (S1–S3) to Hercegszántó (S5–S7), representing industrial, urban, and agricultural areas (Fig. 1A). Seasonal sampling (2013–2019) measured physicochemical parameters and HMs following USEPA (2007) and APHA (2017) protocols. Alluvial soils influence HM mobility (Fuchs et al., 2015). The region has a continental climate (summer 29 °C, winter −2.6 °C, annual precipitation ~619 mm). LULC mapping via GIS identified agriculture, urban, vegetation, and barren land to contextualize pollution sources (Fig. 1B) (Saeed et al., 2023a).

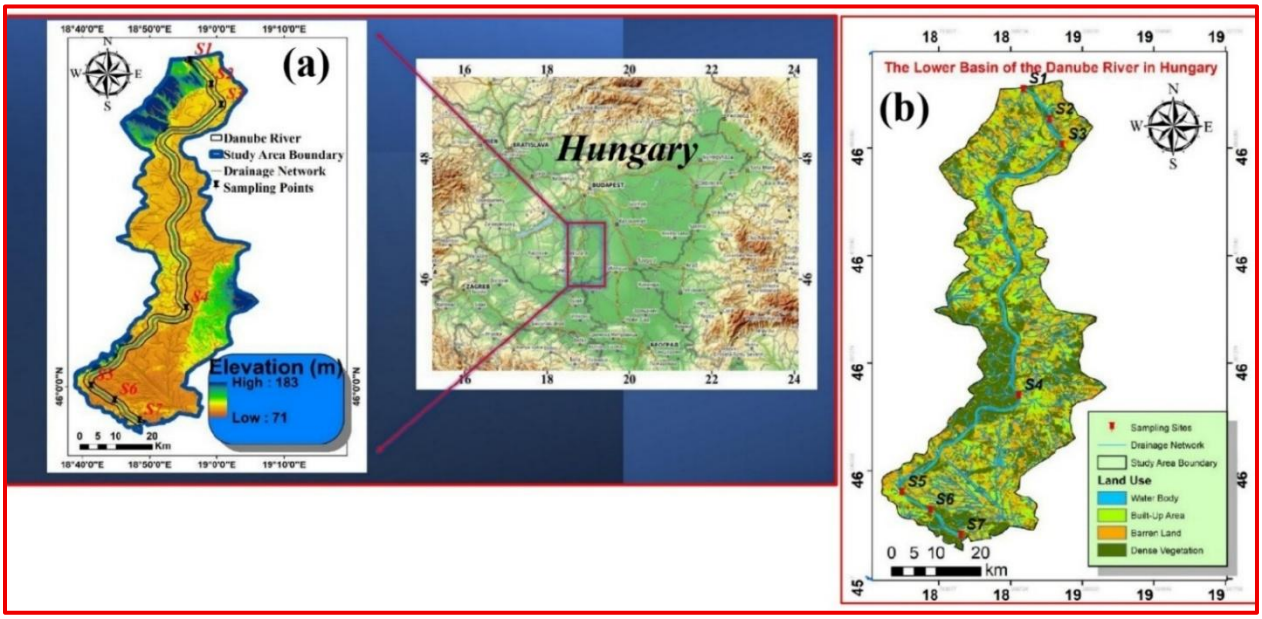

Figure 1 Map of the study area in the Danube River showing the distribution of sampling points and the corresponding land use classification

2.1.2. Al-Jawf basin in Yemen (GW)

Al-Jawf (~30,620 km², ~663,000 population) is agriculturally dominated with wheat, corn, fodder, and barley, plus small-scale industries. The study covered nine districts (longitudes 44°20′–45°0′ E; latitudes 16°0′–16°20′ N) (Alaug & Al-Wosabi, 2015) (Fig. 2). GW is extracted mainly from the shallow Quaternary aquifer (10–165 m depth) with sands, clays, and carbonates, underlain by

Paleozoic–Quaternary formations. GW levels (1086–1259 m AMSL) flow NW to SE, but overextraction forms cones of depression. Arid climate limits recharge; surface runoff into wadis is primary replenishment. Overextraction and limited recharge threaten sustainability, highlighting urgent management needs (Alaug & Al-Wosabi, 2015).

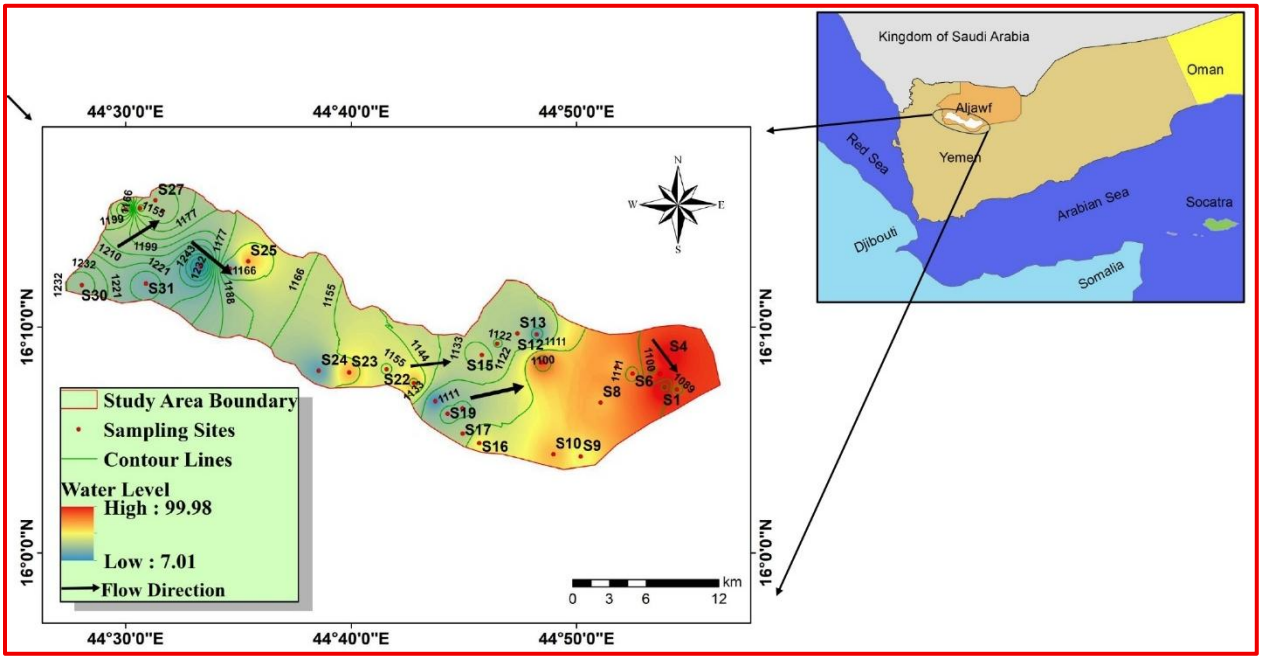

Figure 2 Location and sampling sites within the Al-Jawf basin, Yemen

2.2. Water Sampling

2.2.1. Heavy Metal Analysis of the Lower Danube River Basin (HMs-SW)

To assess seasonal variability, data were grouped into two hydrological periods: April–September (high discharge) and October–March (low flow). SW samples were collected using acid-washed polyethylene bottles, filtered (0.45 µm), acidified to pH < 2 with ultrapure HNO_3, and stored at 4 °C. HMs (As, Cu, Cr, Pb, Ni) were measured via Electrothermal AAS, and Fe, Mn, Zn via Flame AAS, following APHA 3111B/3113B protocols (APHA, 2017). Samples were analyzed in triplicate (RSD ≤ 5%) with calibration using certified standards and blanks to ensure accuracy.

2.2.2 Physicochemical Parameters of the Lower Danube River Basin (SW)

In 2019, 85 seasonal SW samples were collected at seven sites (~30 cm depth) and analyzed for pH, temperature, EC, TDS (in situ), major cations (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}), and anions (Cl^{-}, HCO_3^{-}, CO_3^{2-}, NO_3^{-}, SO_4^{2-}) using titrimetric, spectrophotometric, and flame photometry methods. QA/QC protocols included calibration, duplicate analyses, and averaging replicate samples.

2.2.3 Heavy Metals and Physicochemical Characteristics of the Al-Jawf Basin (GW)

GW samples were collected in 2021 (27 wells) and 2022 (33 boreholes) from the Quaternary aquifer, stored at 3–4 °C, and analyzed for major ions (K^+, Ca^{2+}, Mg^{2+}, Na^+, SO_4^{2-}, Cl^-, HCO_3^-, NO_3^-) and HMs (Mn, Fe) (Al-Mashreki et al., 2023). Standard titration, spectrometry, and FAAS methods were applied.

2.3. Quality Assurance and Control

Analyses followed APHA (Rice et al., 2012). Instruments were calibrated before sampling, samples analyzed in triplicate, and charge balance errors (CBE) calculated via:

$$IB = (TC - TA)/ (TC + TA) \times 100 \quad \text{Eq. 1}$$

All CBE values were within ±5%. Certified Reference Materials and blanks ensured accuracy and reproducibility.

2.4. Multivariate Statistical Methods

PCA reduced dataset complexity and identified dominant water quality factors, following z-score standardization. Data suitability was verified via Kolmogorov–Smirnov (log-normal), KMO = 0.65, and Bartlett's test ($p < 0.05$) (Allafta & Opp, 2020). Cluster Analysis (CA) grouped samples to reveal hydrogeochemical patterns, visualized through dendrograms (Chen et al., 2022; Al-Mashreki et al., 2023).

2.5. Data Analysis, Processing, and Spatial Distribution

IWQIs were analyzed in SPSS v29. Hydrogeochemical processes and water types were assessed via Piper, Chadha, and Gibbs diagrams, using DIAGRAM software (Vasanthavigar et al., 2012). PCA and CA extracted key patterns (Rakotondrabe et al., 2018). GIS (QGIS 3.36.3) visualized spatial distributions; IDW interpolation generated continuous maps from 7 Danube and 60 Al-Jawf sampling points, validated via mean squared error analysis (Khouni et al., 2021).

2.6. Indexing Techniques and Modeling Approaches

2.6.1. Heavy Metal Pollution Index (HPI) and Metal Index (MI)

HPI quantifies HM contamination for human use:

$$HPI = \frac{\sum_{i=1}^{n} W_i Q_i}{\sum_{i=1}^{n} Wi} \quad \text{Eq. 2}$$

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. 3.

$$Q_i = \sum_{i=1}^{n} 100 \times \frac{C_i}{S_i} \quad \text{Eq. 3}$$

MI evaluates cumulative HM effects:

$$M_i = \sum_{i=1}^{i} \frac{C_{ave}}{UAL_i} \quad \text{Eq. 4}$$

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 (Withanachchi et al. 2018).

2.6.2. Potential Ecological Risk Index (RI)

RI quantifies ecological risks:

$$RI = \sum E_r^i = T_r^i \times \left\{\frac{C_{ave}}{C_{bg}}\right\} \quad \text{Eq. 5}$$

Risk categories: low (>30), moderate (30–60), significant (60–120), extremely high (<120) (Yuan et al., 2015).

2.6.3. Human Health Risk Assessment (HHRA)

Non-carcinogenic risks:

$$CDI_{oral} = \frac{C_{ave} \times IR \times EF}{BW \times AT} \times ED \quad \text{Eq. 6}$$

$$CDI_{dermal} = \frac{C_{ave} \times ET \times EF \times Kp \times SA \times CF}{BW \times AT} \times ED \quad \text{Eq. 7}$$

$$HQ_{dermal/oral} = \frac{CDI_{dermal}/CDI_{oral}}{RfD_{dermal}/RfD_{oral}} \quad \text{Eq. 8}$$

$$RfD_{dermal} = RfD_{oral} \times ABS \quad \text{Eq. 9}$$

Hazard Quotient (HQ) and Index (HI) calculated (Eq. 8–9). Carcinogenic risks assessed via CDI and slope factor (CSF), acceptable range 10^{-6}–10^{-4}.

2.6.4. Geochemical Modeling and Saturation Index (SI)

SI assesses mineral equilibrium:

$$SI = log\left\{\frac{IAP}{K_{sp}}\right\} \quad \text{Eq. 10}$$

Where, *IAP* stands for "ion activity product," and K_{sp} represents the "solubility product" at a specified temperature.

2.6.5 Drinking and Irrigation Water Quality Indices

Water quality for human consumption and irrigation was evaluated using DWQI and IWQI.

DWQI: Integrates multiple parameters (HMs, NO_3^-) via weighted arithmetic method:

$$DWQI = \left\{\frac{w_i}{\sum_{i=1}^{n} wi}\right\} \times \left\{100 \times \frac{C_i}{S_i}\right\} \quad \text{Eq. 11}$$

Where: w_i = weight of parameter I; n = total number of parameters; C_i = observed concentration (mg/L); S_i = WHO standard (mg/L)

Weight Calculation:

$$w_i = K/S_i \text{ with } K = 1/\sum 1/S_i \quad \text{Eq. 12}$$

IWQI: Assesses irrigation suitability based on SAR, RSC, EC, and major ions:

$$IWQI = \sum_{i=1}^{n} Q_i W_i \quad \text{Eq. 13}$$

$$Q_i = Q_{max} - \left(\frac{[(X_{ij} - X_{inf}) \times Q_{imap}]}{X_{amp}}\right) \quad \text{Eq. 14}$$

$$W_i = \frac{\sum_{j=1}^{k} F_j A_{ij}}{\sum_{j=1}^{k}\sum_{i=1}^{n} F_j A_{ij}} \quad \text{Eq. 15}$$

Where: X_{ij} = measured value; X_{inf} = lower class limit; Q_{imap} = class amplitude range; X_{amp} = amplitude of parameter; F_j = automated logging factor; A_{ij} = influence of factor *j* on parameter *i*

2.7. Machine Learning-Based Water Quality Assessment

Random Forest (RF) models predicted IWQIs and identified feature importance. Performance metrics included:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y} - y_i)^2}{\sum_{i=1}^{n}(\bar{y} - y_i)^2} \quad \text{Eq. 16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y} - y_i|^2 \quad \text{Eq. 17}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\hat{y}-y_i}{y_i}\right| \times 100\% \quad \text{Eq. 18}$$

Complementary ANFIS and RF models ensured robust probabilistic and nonlinear water quality predictions (Athamena et al., 2023; Zhang et al., 2024).

2.8. Monte Carlo Simulation (MCS)

MCS quantified uncertainties in HHRA for HMs and NO_3^-. Input distributions (concentration, IR, ED, EF, BW, SA, Kp) were simulated with 10,000 iterations using Python 3.9, confirming that predicted HQ values matched observed data and remained below thresholds (HQ < 1) (Vasanthavigar et al., 2013).

3. RESULTS AND DISCUSSION

3.1. Heavy metal assessment of the Danube River (SW)

Mean HM concentrations: Fe > Mn > Zn > Cu > Ni > Cr > Pb > As; Fe and Mn occasionally exceeded WHO (2017) downstream of urban/industrial sites. Seasonal dilution slightly reduced Mn, Fe, and Cu in October–March. Other metals remained below guideline values, suggesting low immediate risk (Figs. 3-4; Table 1).

Table 1 The overview of average HM levels in µg L−1 at the lower watershed of the Danube River, Hungary

	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

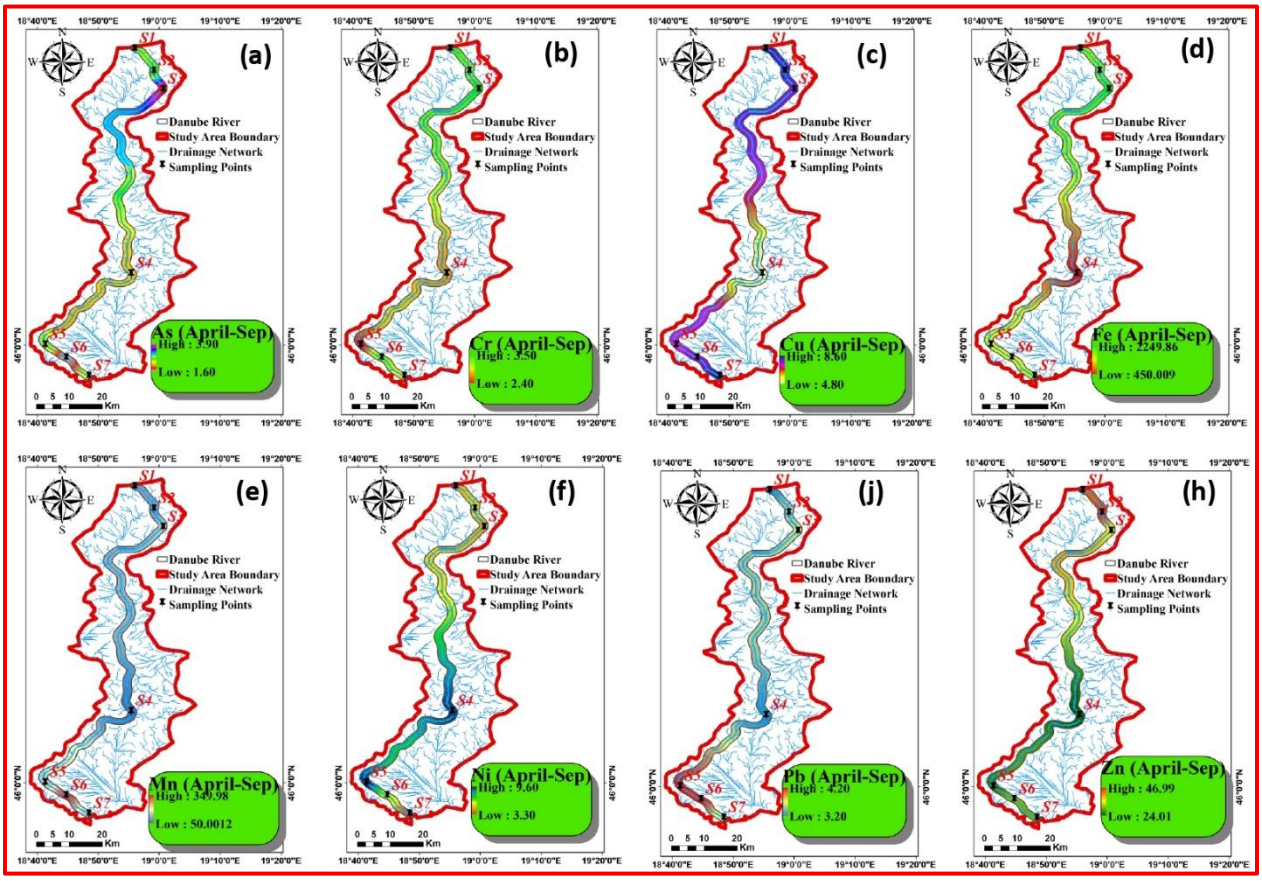

Figure *3* Spatial distribution maps of heavy metal concentrations in surface water samples from the Danube River basin during the April–September period.

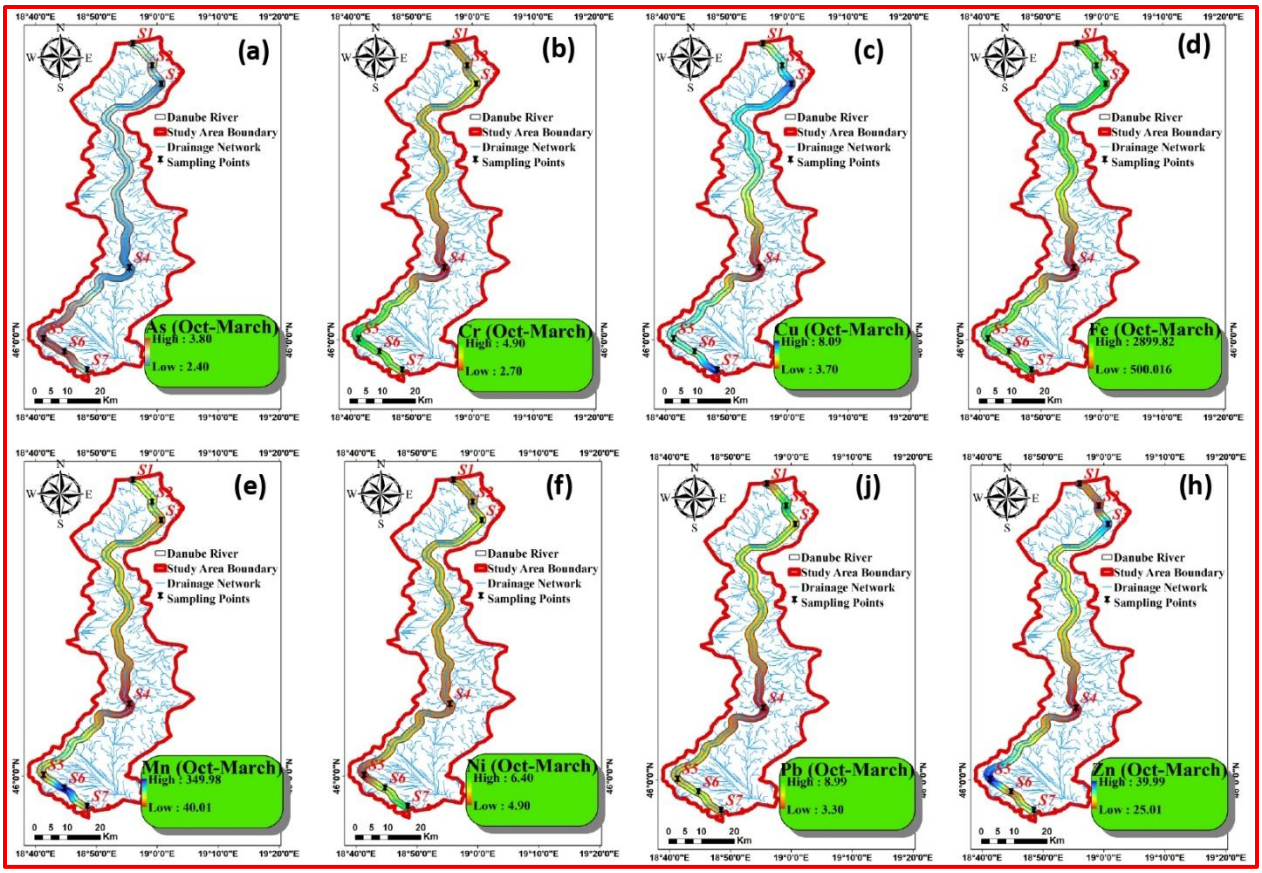

Figure *4* Spatial distribution maps of heavy metal concentrations in surface water samples from the Danube River basin during the October–March period.

3.2. Pollution & ecological risk indices (HPI, MI, RI)

HPI indicated moderate pollution (15–30), with Fe and Mn as major contributors. MI reflected serious pollution in April–September (4.13) and moderate in October–March (3.28). RI values remained low (3.34–3.35, RI < 30), indicating minimal ecological risk, but continuous monitoring is advised (Fig. 5).

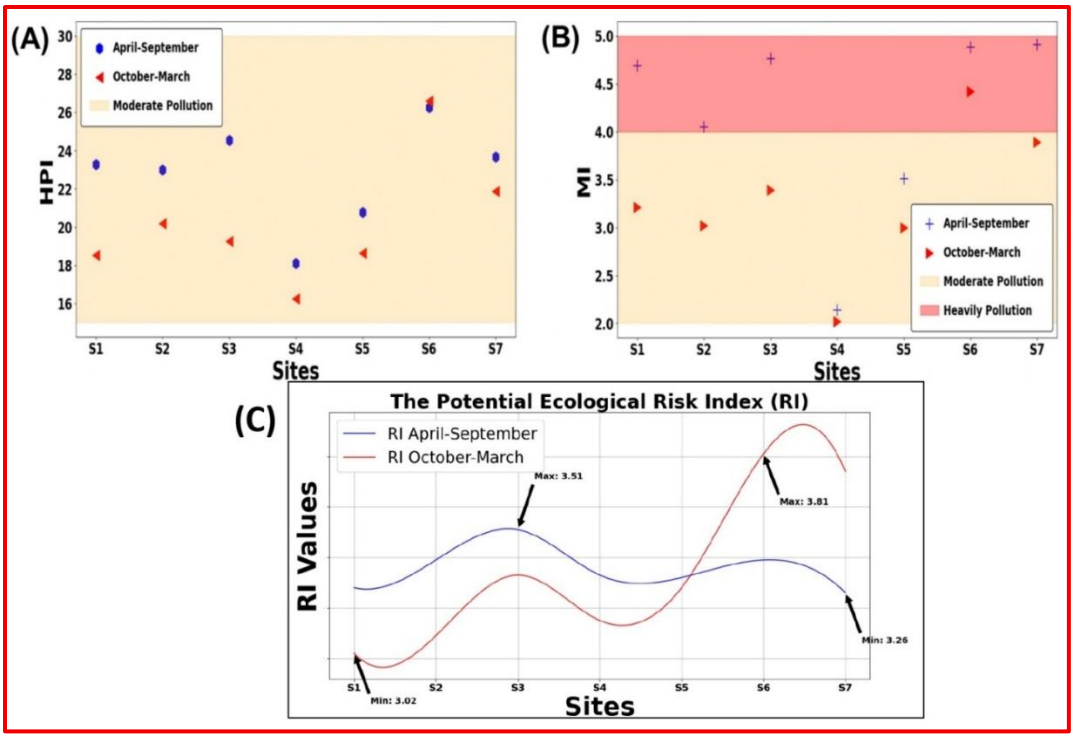

Figure 5 Seasonal analysis of heavy metal contamination: (A) Correlation heatmap for April-September, (B) Correlation heatmap for October-March, and (C) Assessment of ecological risk (RI)

3.3. Health Risk Assessment (HHRA & MCS)

Probabilistic HQ distributions confirmed all values <1 for adults and children, validating risk predictions (Figs. 8–9).

3.3.1. Non-Carcinogenic Health Risk

Children exhibited higher HQs than adults. Oral As and Mn HQs reached 0.5 and 0.35, respectively; dermal HQs followed a similar trend. All HQs <1, indicating no immediate risk but higher vulnerability in children (Figs. 6A–B).

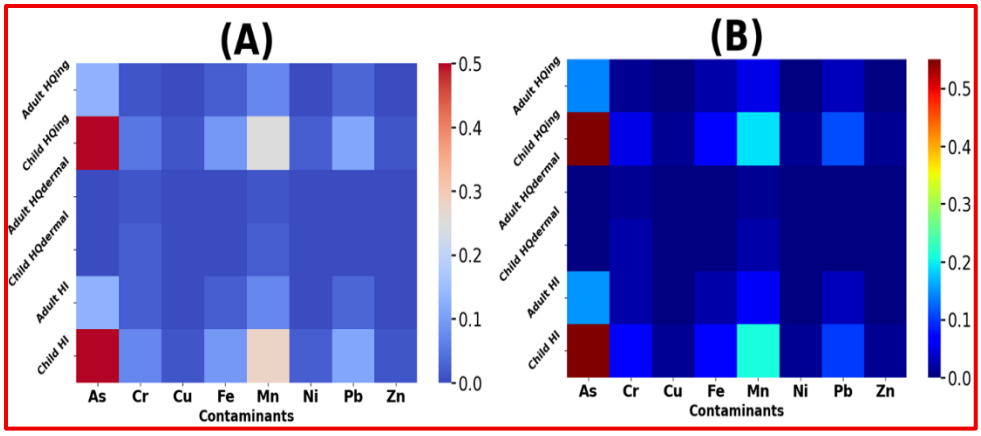

Figure 6 Hazard quotient and Hazard index for oral and dermal pathways: (A) during April-September; (B) during October-March

3.3.2. Monte Carlo Simulation (MCS)

Probabilistic HQ distributions confirmed all values <1 for adults and children, validating risk predictions.

3.3.3. Carcinogenic Health Risk

Children faced higher CR from As, Cr, Pb. Oral As CR exceeded 1×10^{-4} (October–March, 95th percentile 3.6×10^{-4}). Adults remained below thresholds (Figs. 7-8).

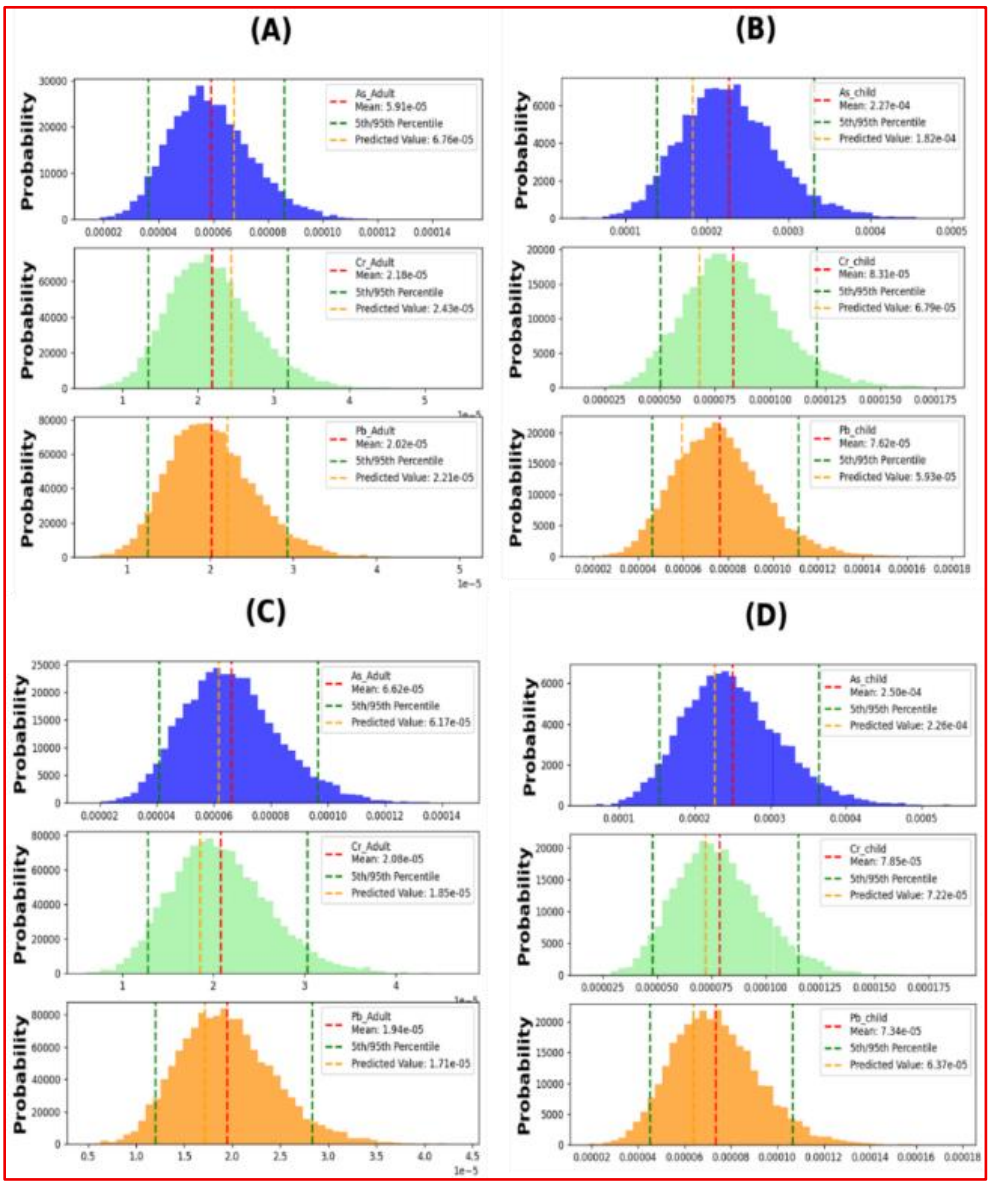

Figure 7 Carcinogenic risk oral: (A) adults in April-September; (B) children in April-September; (C) adults in October-March; (D) children in October-March

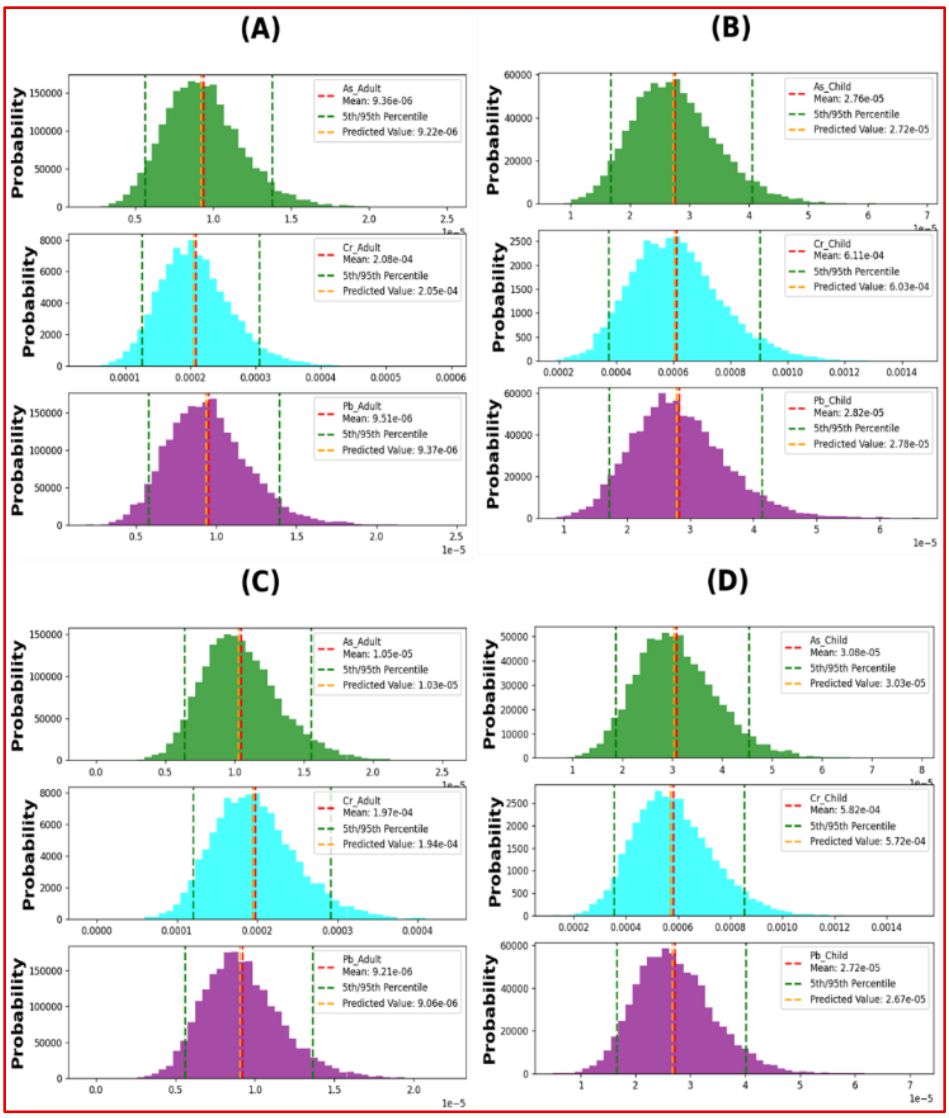

Figure 8 Carcinogenic risk dermal: (A) adults in April-September; (B) children in April-September; (C) adults in October-March; (D)

3.4. Physicochemical Properties & Hydrochemical Characteristics of the Danube River

pH: 7.6–8.6; EC: 360–680 µS/cm; TDS: 199–370 mg/L. Ca^{2+} (53.8 mg/L) > Mg^{2+} (15.37) > Na^{+} (16.06) > K^{+} (2.6). Dominant anions: HCO_3^- > SO_4^{2-} > Cl^- > CO_3^{2-} > NO_3^-. North-south gradients reflect mineral dissolution and anthropogenic inputs. Thermal warming trends affect hydrochemical processes (Table 2).

Table 2 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+}	54.45	54.15	54.41	53.93	53.59	53.72	53.11	400	75
Mg^{2+}	16.13	15.59	15.70	15.31	15.25	14.76	14.50	60	50
Na^{+}	16.56	16.68	16.56	15.63	15.87	15.57	15.27	919	200
Cl^{-}	24.50	24.25	24.17	22.92	22.50	22.33	22.23	1036	250
SO_4^{2-}	37.25	37.08	37.08	36.67	35.08	35.00	34.85	960	250
HCO_3^{-}	195.42	188.33	193.33	190.83	190.00	188.33	181.15	610	120
CO_3^{2-}	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	272.50	263.38	251.88	255.71	296.36	254.70	257.44	2000	500
K^{+}	2.58	2.69	2.60	2.50	2.74	2.64	2.57	2	12
NO_3^{-}	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	0.60	0.58	0.57	0.60	0.63	0.53	0.65	5	0.3
Mn	0.07	0.06	0.07	0.05	0.07	0.05	0.05	0.2	0.1

3.5. Surface Water Facies and Source Determination

3.5.1. Surface Water Categories

Piper diagram: Ca–Mg–HCO_3 type; Gibbs diagram confirms rock weathering as dominant process (Figs. 9A–B).

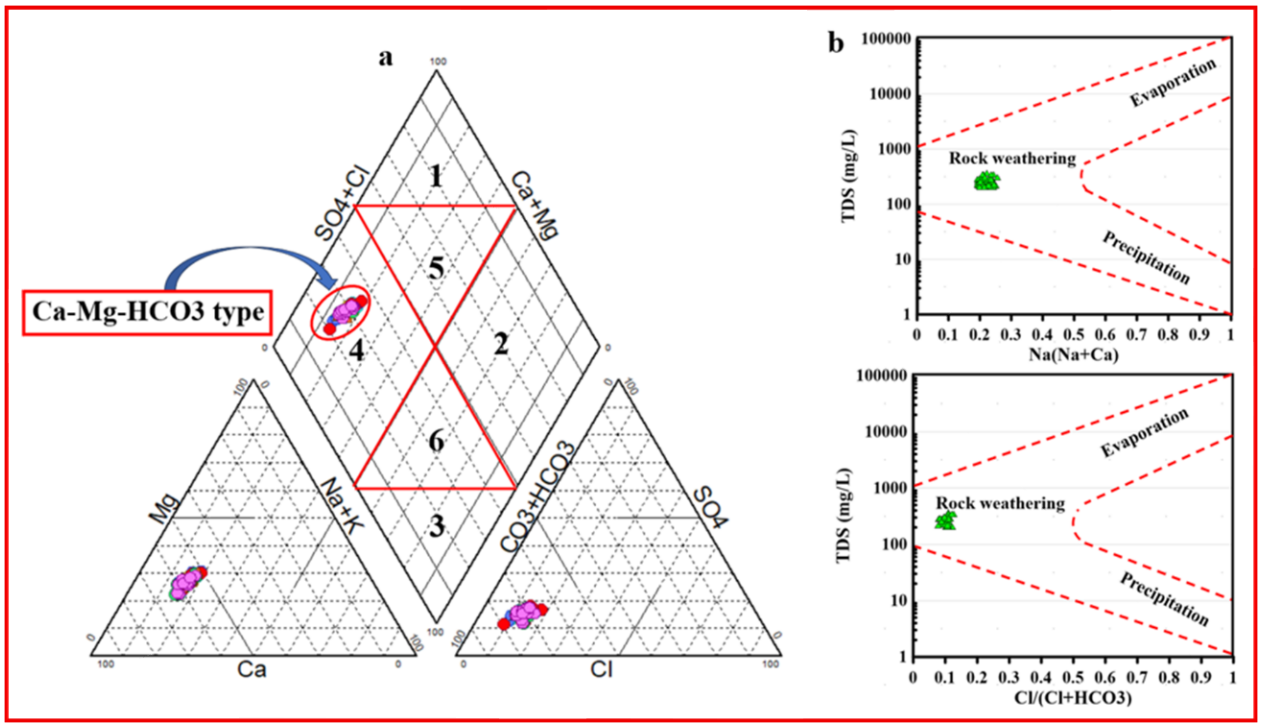

Figure 9 Surface water facies according to Piper diagram (a) and geochemical controlling mechanisms according to Gibbs diagram (b)

3.5.2. Ion Exchange

Clay minerals mediate cation exchange (Ca^{2+}/Mg^{2+} ↔ Na^{+}/K^{+}). Ratios $[(Ca^{2+}+Mg^{2+})/(HCO_3^{-}+SO_4^{2-})]$ <1 confirm ion exchange dominance (Fig. 10).

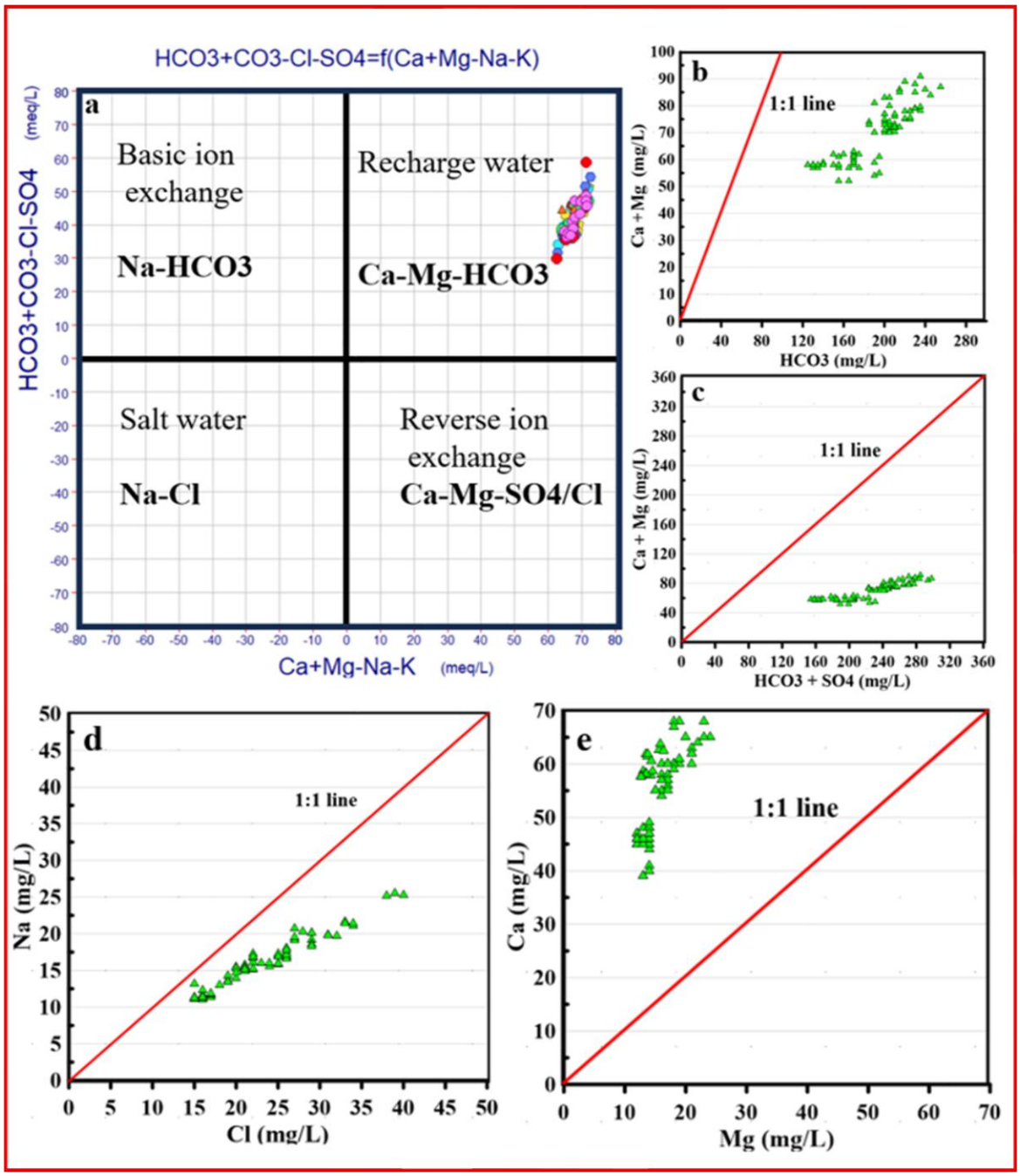

Figure 10 Relationships between the main cations and anions in the sample water using stoichiometry: (a) $Ca^{2+}+Mg^{2+}$-$(Na^{+}+K^{+})$ vs. $(HCO_3^{-}(SO_4^{2-}+Cl^{-})$, (b) $Ca^{2+}+Mg^{2+}$ vs. HCO_3^{-}, (c) $Ca^{2+}+Mg^{2+}$vs. $HCO_3^{-}+SO_4^{2-}$, (d) Na^{+} vs Cl^{-}, (e) Ca^{2+} vs Mg^{2+}

3.5.3. Chlor-Alkali Indices

Negative CAI-I/II values indicate strong ion exchange (Fig. 11).

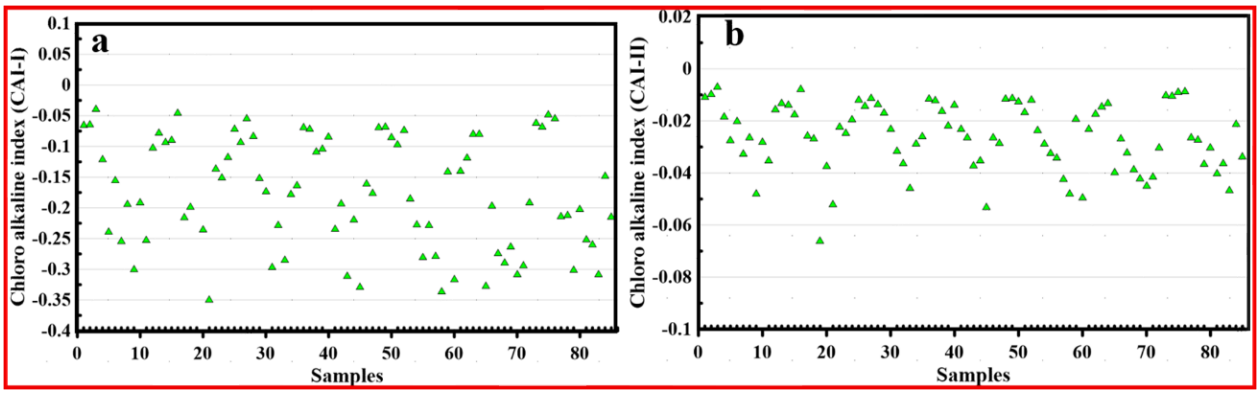

Figure 11 Relationships between Samples vs. CAI-I, and (f) Samples vs. CAI-II.

3.5.4. Geochemical Modeling and Mineral Saturation StatePHREEQC modeling shows SI <0 for gypsum, dolomite, calcite, suggesting undersaturation and active dissolution along flow path (Fig. 12).

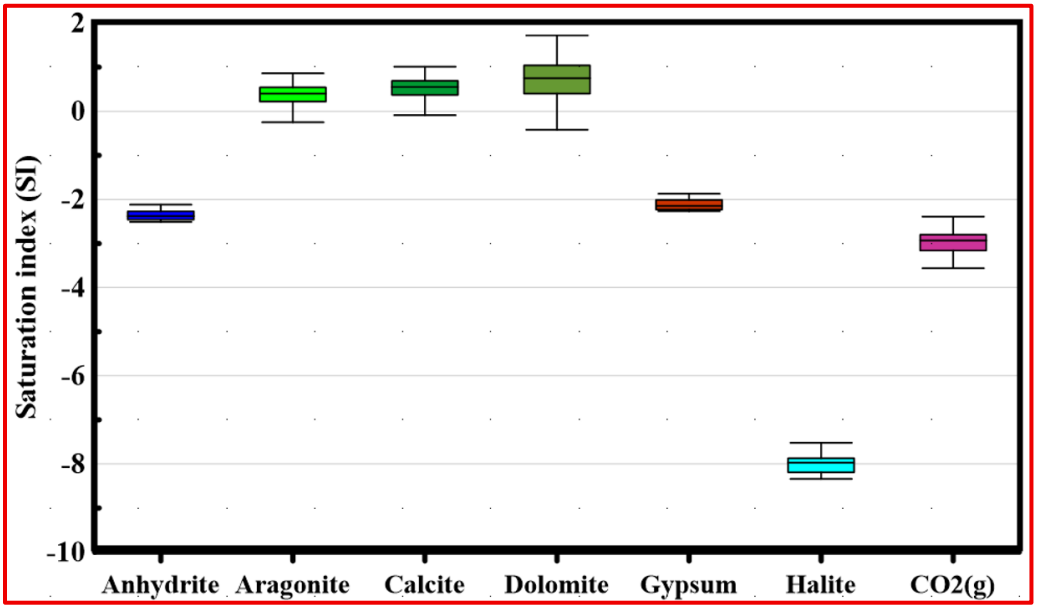

Figure 12 Mineral saturation state showing the ability of precipitation and dissolution of minerals in Danube River

3.6. Statistical Analysis

3.6.1. Cluster Analysis

Three clusters identified: Cluster 1 (HCO_3^-, TDS), Cluster 2 (Na^+, Cl^-, Mg^{2+}, SO_4^{2-}), Cluster 3 (Ca^{2+}, NO_3^-, Fe, Mn) (Fig. 13A).

3.6.2. Principal Component Analysis

Four PCs explained 82.1% variance. PC1: major ions & TDS; PC2: pH & CO_3^{2-}; PC3: Fe; PC4: K^+ & Mn vs. TDS (Fig. 13B–C).

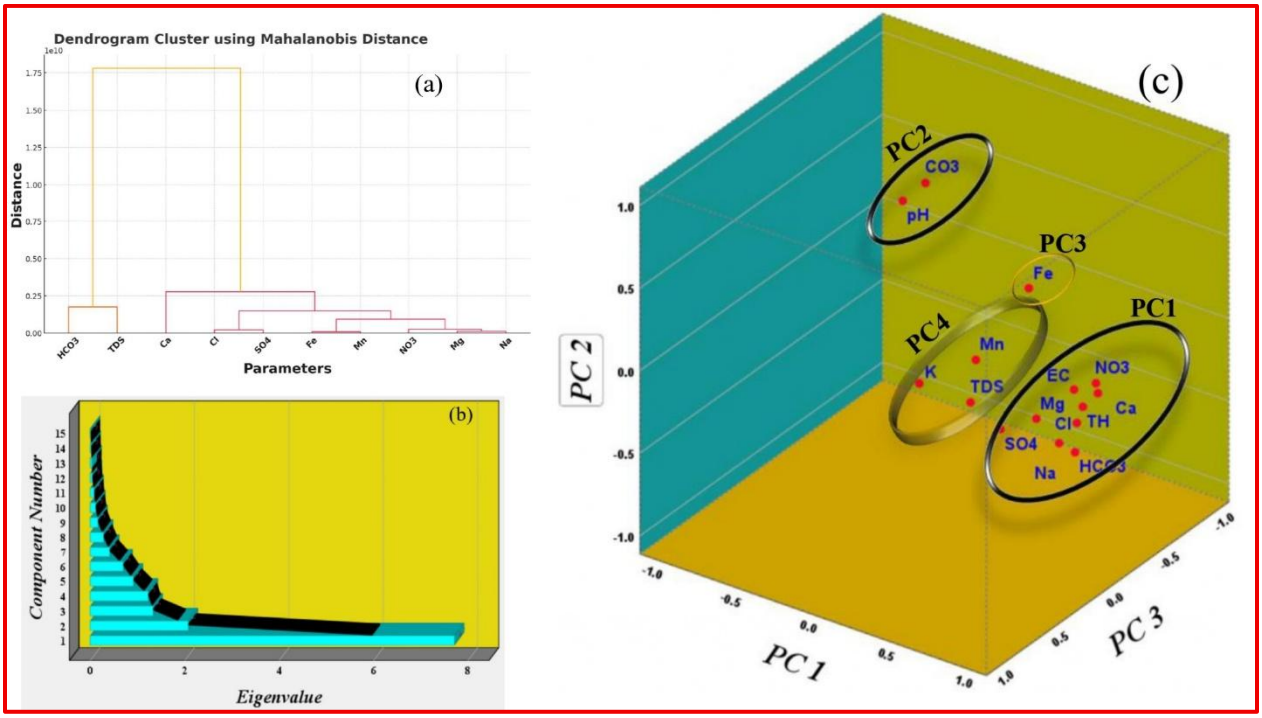

Figure 13 Multivariate statistical analysis: (a) Cluster dendrogram for variables; (b) Scree plot and (c) Principal coordinate analysis scores for PC1 vs. PC2 vs. PC3.

3.7. Water Quality Indices: Drinking and Irrigation

DWQI: S2–S7 acceptable (76–100); S1, S5 slightly above due to Fe/HCO_3^- (Fig. 13A). IWQI: excellent (99.6–107.6), SAR 0.37–0.68, Na% 13.75–18.72, RSC <1.25, indicating safe irrigation (Figs. 13B–G).

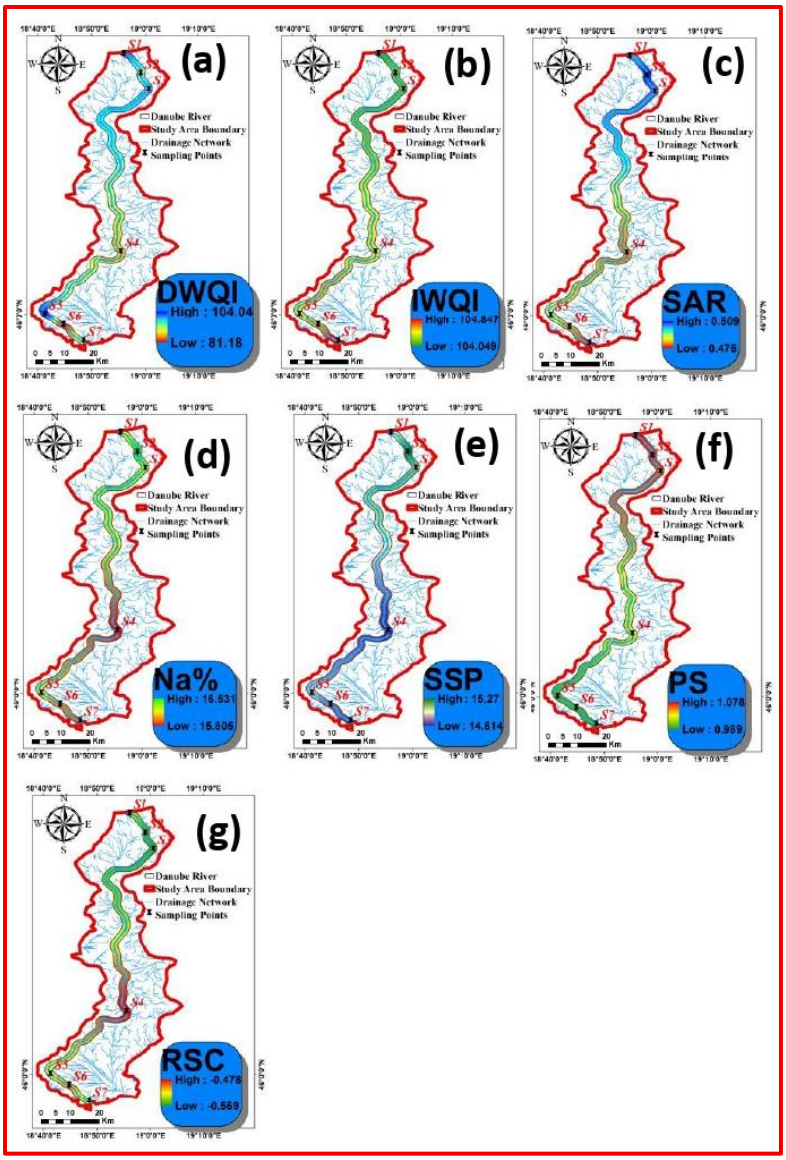

Figure 13 The spatial variation maps of the WQI and IWQIs for the lower Danube river basin: (a) DWQI (b) IWQI, (c) SAR, (d) Na %, (e) SSP, (f) PS, and (g) RSC

3.8. Heavy Metal and Physicochemical Properties of the Al-Jawf Basin

3.8.1. Groundwater Chemistry

pH 7–8.2, TDS 46% moderate salinity, EC 1030–5930 µS/cm. Ca^{2+} 40–360 mg/L, Mg^{2+} 41–360 mg/L, Na^{+} 52–690 mg/L, K^{+} 2–23 mg/L. HCO_3^{-} 171–464 mg/L, Cl^{-} 128–1014 mg/L, SO_4^{2-} exceeded FAO in 13% of samples. Fe 0.01–3 mg/L; Mn 0.001–4 mg/L, exceeding drinking limits in 13% (Table 3).

Table 3 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

3.8.2. Geochemical Processes Influencing Groundwater Facies Piper diagram: Na–Cl and Ca–Mg–SO_4 types; Gibbs diagram confirms evaporation, drainage, and mineral dissolution as controls (Fig. 14).

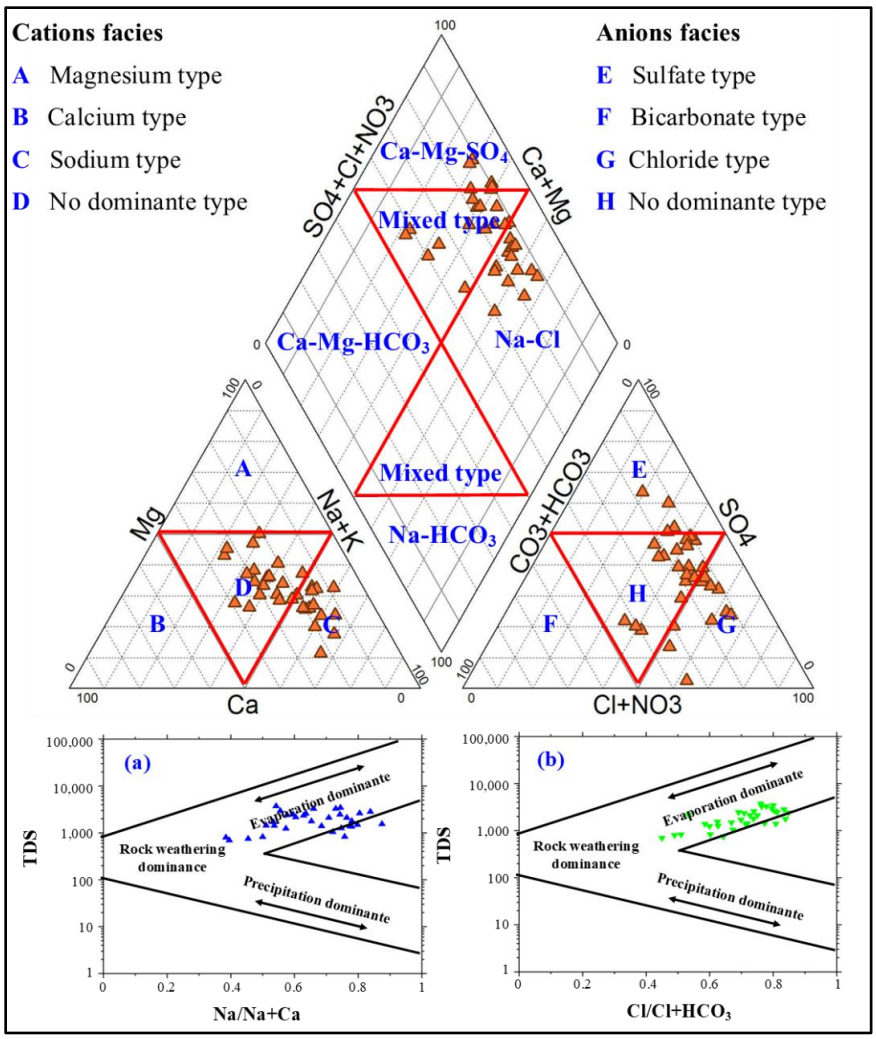

Figure 14 Piper diagram and Gibbs diagram showing the groundwater evolution

3.8.3. Ion Exchange Processes

Cl–alkali indices mostly negative (~60%), supporting active cation exchange. Na^+ vs. Cl^- (R=0.81) and SO_4^{2-}–Mg^{2+} (R=0.76) indicate silicate weathering and gypsum/anhydrite dissolution as dominant sources (Figs. 15A–E).

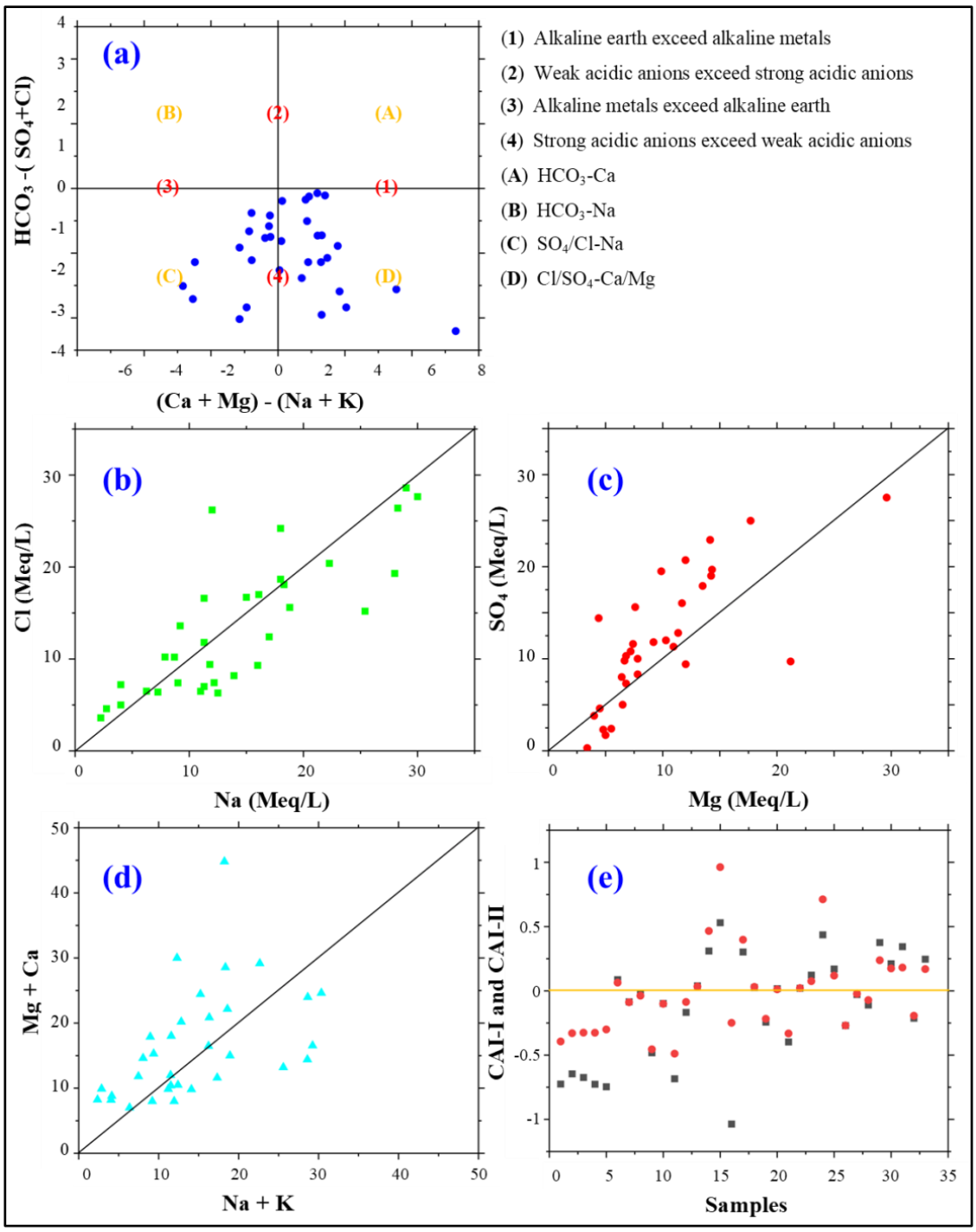

Figure 15 Relationships between the main investigated parameters (a) Ca^{2+}+Mg^{2+}-(Na^{+}+K^{+}) vs. (HCO_3-(SO_4^{2}+Cl), (b) Cl^{-} vs Na^{+}, (c) SO_4^{2-} vs. Mg^{2+}, (d) Mg^{2+}+Ca^{2+} vs. $Na^{+}K^{+}$, (e) Samples vs. CAI-I and CAI-II

3.8.4. Statistical Analysis

Cluster analysis (CA) grouped 33 groundwater samples into three distinct clusters. Cluster 1, characterized by Ca^{2+}, Mg^{2+}, CO_3^{2-}, K^+, Mn, Fe, and NO_3^-, reflects water hardness and the influence of fertilizers. Cluster 2, comprising Cl^-, SO_4^{2-}, HCO_3^-, and Na^+, is associated with mineral dissolution and water–rock interactions. Cluster 3, represented by TDS, indicates contributions from evaporitic and carbonate sources. Principal component analysis (PCA) extracted three components that together explained 89.6% of the total variance. PC1, accounting for 55.2% of the variance, is linked to salinization and mineral dissolution. PC2, representing 14.1%, reflects nitrate contamination from agricultural activities and soil mineralization, while PC3, explaining 8.0%, highlights localized anthropogenic influence, particularly from potassium-rich fertilizers.

3.8.5. water quality indices for irrigation and agricultural purposes

Groundwater suitability was assessed using IWQI, SAR, SSP, PS, KR, and RSBC indices. The IWQI ranged from 16.09 to 92.12, with a mean of 48.03, indicating that 63.5% of the samples impose high to severe restrictions suitable only for salt-tolerant crops, 23.4% show moderate restrictions, and 3% have no limitations. Analysis of SAR and USSL classifications revealed that 42.4% of samples have very high salinity with medium SAR (C4-S2), 15.15% exhibit very high salinity with high SAR (C4-S3), and 33.3% fall under high salinity with low to medium SAR (C3-S1/S2). The SSP ranged from 21.44 to 66.04, with 90.9% of samples suitable for use and 9.09% requiring treatment. PS assessment showed that 93.9% of samples are injurious or unsatisfactory due to chloride and sulfate content. KR values indicated 60.6% suitability, while 39.39% exceeded the threshold of 1. RSBC values ranged from -9.60 to 2.92, with a mean of -2.01, reflecting safe residual sodium levels. Overall, these indices highlight spatial variability, the importance of salinity management, and the need for strategic crop selection.

3.8.6. Non-Carcinogenic Health Risk Assessment

Dermal exposure assessment (CDI) showed that manganese (Mn) levels were higher than iron (Fe), with adults receiving approximately 6.4×10^{-5} mg/L/day and children around 2×10^{-4} mg/L/day. The dermal hazard quotient (HQ) was below 1 for adults, while children reached a maximum of 1.76 for Mn, indicating a potential non-carcinogenic risk. The dermal hazard index (HI) averaged 0.07 for adults (range 0.0002–0.6) and 0.2 for children (range 0.0005–1.77), with

three samples exceeding 1 for children. These results suggest that children are more vulnerable, emphasizing the need for targeted interventions and ongoing monitoring.

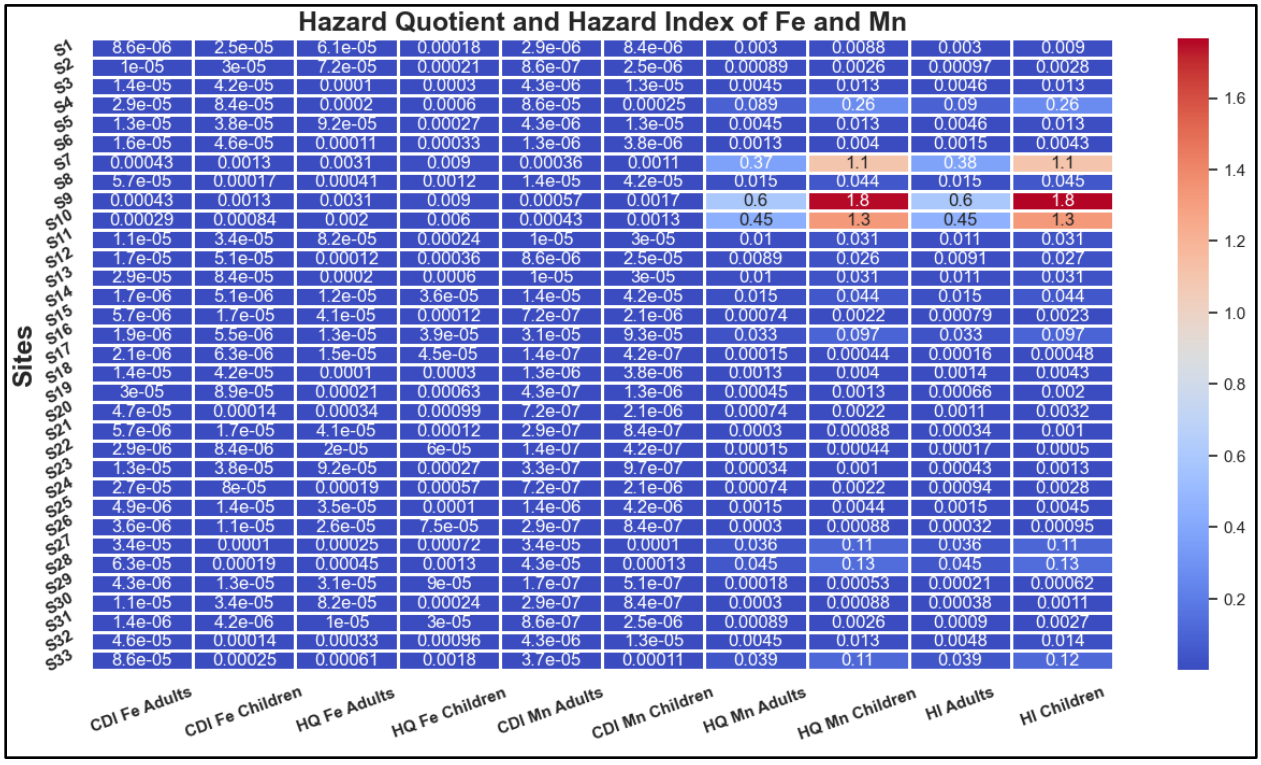

Figure 16 Heat map showing the health risk indices including CDI, HQ, and HI for adult and children through dermal contact

3.8.7. Random Forest Simulation Model

RF models predicted IWQI, SAR, KR, SSP, PS, RSBC, and HI using 12 hydrogeochemical variables (60 samples, 80/20 train-test split, 5-fold CV).

Table 4 Evaluation metrics for RF models on measured parameters, including IWQI, SAR, KR, SSP, PS, RSBC, HI (adult), and HI (child)

Index	Best RF Model	R^2 (Train)	RMSE (Train)	R^2 (Validation)	RMSE (Validation)
IWQI	RF-3HG	0.988	2.239	0.940	3.976
SAR	RF-2HG	0.971	0.404	0.869	0.667
KR	RF-2HG	0.952	0.099	0.665	0.203
SSP	RF-4HG	0.959	2.453	0.671	5.394
PS	RF-2HG	0.994	0.744	0.982	0.960
RSBC	RF-12HG	0.924	0.768	0.751	1.042
HI	RF-1HG	0.988	0.018	0.901	0.041
Index	Best RF Model	R^2 (Train)	RMSE (Train)	R^2 (Validation)	RMSE (Validation)
IWQI	RF-3HG	0.988	2.239	0.940	3.976
SAR	RF-2HG	0.971	0.404	0.869	0.667
KR	RF-2HG	0.952	0.099	0.665	0.203
SSP	RF-4HG	0.959	2.453	0.671	5.394
PS	RF-2HG	0.994	0.744	0.982	0.960
RSBC	RF-12HG	0.924	0.768	0.751	1.042
HI	RF-1HG	0.988	0.018	0.901	0.041

RF models improved predictive accuracy, reduced subjectivity, and identified key influential features, supporting sustainable groundwater management.

4. CONCLUSION AND RECOMMENDATIONS

4.1. Conclusion

- Danube River (SW, Hungary): Mn and Fe exceeded drinking water limits at several sites; other HMs below risk thresholds. Irrigation indices indicate moderate suitability. Hydrochemistry reflects both natural and anthropogenic influences.
- Al-Jawf Basin (GW, Yemen): Groundwater quality shows strong spatial variability. PCA and geochemical modeling reveal rock–water interactions and ion exchange as key processes. IWQIs indicate some sources need treatment.
- Machine Learning: RF models effectively predicted IWQIs and HI, demonstrating applicability across SW and GW systems.

4.2. Recommendations

- Monitoring: Multi-seasonal, region-specific monitoring to capture spatial/temporal variability.
- Management: Danube: stricter industrial discharge regulation.
- Al-Jawf: sustainable agriculture, crop selection (salt-tolerant), salinity mitigation.
- Mitigation: Constructed wetlands, vegetative buffers, small-scale filtration or ion exchange.
- Tools: Routine integration of PCA, geochemical modeling, IWQIs, MCS, and machine learning.
- Future Research: Address climate impacts, emerging contaminants (PFAS, microplastics, pharmaceuticals).
- Collaboration: Hydrogeology, environmental science, and public health integration for evidence-based water management.

5. 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.

6. 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).