

Petroleum-driven groundwater contamination in Azzan city, Yemen: Integrated assessment using hydrochemistry, trace metals, PAHs, and PCBs

Salem Salem Hussein Sofui

Department of Oil and Gas, Faculty of Oil and Minerals, Shabwah University, Shabwah, Yemen

Shaif Mohammed Kasem Saleh

Department of Chemistry, Faculty of Science, University of Aden, Aden, Yemen

Sabrina Jemai

Higher Institute of Applied Sciences and Technology, University of Gabès, Gabès, Tunisia

Safa Gamal Mohammed

Department of Chemistry, Faculty of Science, University of Aden, Aden, Yemen

Abdelmonem Marsas Habtoor

Department of Geological Engineering, Faculty of Oil and Minerals, Shabwah University, Shabwah, Yemen

Mahmoud Bali

`Mahmoud.Bali@isstegb.rnu.tn`

Higher Institute of Water Sciences and Techniques, University of Gabès, Gabès, Tunisia

<https://orcid.org/0000-0002-5288-9732>

Research Article

Keywords: Groundwater contamination, Petroleum pollution, Trace metals, PAHs, PCBs, Health risk assessment, Azzan City, Yemen

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10432540/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

**Petroleum-driven groundwater contamination in Azzan city, Yemen:
Integrated assessment using hydrochemistry, trace metals, PAHs, and
PCBs**

**Salem Salem Hussein Sofui ^a, Shaif Mohammed Kasem Saleh ^b, Sabrine Jemai ^c,
Safa Gamal Mohammed ^d, Abdelmonem Marsas Habtoor ^e, Mahmoud Bali ^{f*}**

^a Department of Oil and Gas, Faculty of Oil and Minerals, Shabwah University,
Shabwah, Yemen.

^b Department of Chemistry, Faculty of Science, University of Aden, Aden, Yemen

^c Higher Institute of Applied Sciences and Technology, University of Gabès, Gabès,
Tunisia.

^d Department of Chemistry, Faculty of Science, University of Aden, Aden, Yemen.

^e Department of Geological Engineering, Faculty of Oil and Minerals, Shabwah
University, Shabwah, Yemen.

^f Higher Institute of Water Sciences and Techniques, University of Gabès, Gabès,
Tunisia.

***Corresponding Author:** Mahmoud Bali (Email: Mahmoud.Bali@isstegb.rnu.tn)

Petroleum-driven groundwater contamination in Azzan city, Yemen:

Integrated assessment using hydrochemistry, trace metals, PAHs, and PCBs

Abstract

This study presents an integrated assessment of petroleum-induced groundwater contamination in Azzan City, Shabwah Governorate, Yemen, an arid area affected by petroleum-related activities. Seven groundwater wells located along a crude oil pipeline corridor were investigated to evaluate hydrochemistry, trace metals, petroleum-related organic pollutants, and associated health risks. Physicochemical parameters and major ions were determined using standard analytical methods, trace metals by inductively coupled plasma–mass spectrometry (ICP–MS), and polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) by gas chromatography–mass spectrometry (GC–MS).

Groundwater showed pronounced spatial variability. Electrical conductivity ranged from 1015 to 4740 $\mu\text{S}/\text{cm}$, and total dissolved solids ranged from 406 to 3078 mg/L, indicating elevated mineralization. The hydrochemical composition was dominated by a Ca–Mg– SO_4 facies, reflecting water–rock interaction and carbonate–sulfate mineral dissolution. Toxic trace metals, particularly Pb, As, and Cd, exceeded health-based drinking-water guideline values in several wells, while integrated pollution indices confirmed severe groundwater deterioration, especially in Mahdi and Atash.

Total PAHs and PCBs were detected in all samples. PAHs were dominated by low-molecular-weight compounds, especially naphthalene, which accounted for 89.08% of total PAHs, indicating a petrogenic signature. Total PCB concentrations reached

0.374 mg/L and were dominated by tetrachlorinated congeners. A strong positive correlation between electrical conductivity and total PAHs ($r = 0.961$) suggests that salinity-related parameters may support screening of petroleum-related contamination. Health-risk assessment revealed substantial non-carcinogenic and carcinogenic risks, with $HI = 16.19$ and $TCR = 8.79 \times 10^{-3}$. Overall, the results indicate a multi-component contamination system controlled by geogenic processes and localized petroleum-related inputs, requiring continuous monitoring and targeted groundwater protection.

Keywords: Groundwater contamination; Petroleum pollution; Trace metals; PAHs; PCBs; Health risk assessment; Azzan City, Yemen

1. Introduction

Environmental degradation in oil-producing regions has emerged as a critical environmental challenge due to contamination associated with petroleum industry activities, including improper waste disposal and accidental leakages from crude oil transportation pipelines. Among these sources, pipeline failures and petroleum leakage represent major drivers of localized groundwater contamination, particularly in arid and semi-arid environments where groundwater is often the main water resource and natural dilution is limited. Such impacts extend to multiple environmental compartments, including water, soil, air, and biota, posing significant risks to surrounding communities, agricultural productivity, and water resources (Smith, 2001; Moeini and Alizadeh Paenafrakaty, 2008; Jughmam and Habtour, 2020).

Petroleum-derived contaminants include both organic and inorganic pollutants. Polycyclic aromatic hydrocarbons (PAHs) are among the most important petroleum-

related organic contaminants because of their toxicity, persistence, mobility, and carcinogenic potential. Their occurrence in groundwater may reflect crude oil leakage, petroleum-product migration, or combustion-related inputs, depending on their molecular distribution and diagnostic ratios (Neff, 1979; Wang et al., 2009; Yunker et al., 2002; Ugochukwu, 2018). Polychlorinated biphenyls (PCBs), although not naturally present in petroleum, are persistent organic pollutants that may occur in industrially affected environments and are characterized by long environmental residence times, low degradation rates, and significant ecological and human health implications (ATSDR, 2000). In addition to organic pollutants, petroleum-related activities are frequently associated with trace-metal enrichment through produced-water discharge, corrosion products, drilling-related residues, industrial wastes, and interactions with metal-bearing geological formations, further complicating groundwater quality and increasing environmental and health risks (Alloway, 2013; Tchounwou et al., 2012).

Groundwater systems are particularly vulnerable to persistent contamination because pollutants may remain in the subsurface for long periods with limited natural attenuation. In arid and semi-arid regions, this vulnerability is intensified by low recharge rates, high evaporation, and strong dependence on groundwater for domestic and agricultural use (Huling and Weaver, 1991; Al Manmi et al., 2019). Groundwater chemistry is controlled not only by anthropogenic inputs but also by natural hydrogeochemical processes, including water–rock interaction, carbonate dissolution, sulfate mineral dissolution, evaporation, and ion exchange. These geogenic processes define the natural hydrochemical baseline, whereas petroleum-related contamination can superimpose additional organic and inorganic pollution signatures (Freeze and Cherry, 1979; Hem, 1985; Appelo and Postma, 2005; Todd and Mays, 2005).

Despite the growing number of studies addressing groundwater contamination in petroleum-producing regions worldwide, there remains a lack of integrated investigations that simultaneously evaluate hydrochemical characteristics, trace metals, and petroleum-derived organic contaminants, particularly in arid environments where groundwater serves as the primary water resource. In Yemen, previous investigations have evaluated heavy metal distribution in Al-Dhalia (Saleh et al., 2020), microbiological and physicochemical risks in the Tuban Delta (Saleh and Al-Sallami, 2022), and water-quality indices in coastal agricultural basins such as Wadi Zabid (Saleh et al., 2017b; Al-Alimi et al., 2023). However, most available studies have focused on isolated contamination indicators rather than integrated forensic assessments combining hydrochemistry, toxic trace metals, PAHs, PCBs, multivariate statistics, and human-health risk analysis. This gap is especially critical in petroleum-stressed aquifers such as Azzan City, Shabwah Governorate, where crude oil infrastructure and pipeline corridors may directly affect groundwater quality.

Furthermore, the interaction between geogenic mineral dissolution and anthropogenic petroleum-related contamination remains insufficiently understood in Yemen, particularly regarding how these processes jointly control contaminant distribution, mobility, hotspot formation, and associated health risks in groundwater systems (Freeze and Cherry, 1979; Appelo and Postma, 2005). Distinguishing natural hydrochemical evolution from petroleum-induced contamination is therefore essential for accurate source interpretation, risk assessment, and groundwater management.

Therefore, this study aims to provide an integrated assessment of petroleum-induced groundwater contamination in the vicinity of a crude oil pipeline corridor in Azzan City, Shabwah Governorate, Yemen. Specifically, the study aims to: (i) characterize the physicochemical properties and major-ion composition of groundwater; (ii)

evaluate trace-metal contamination and integrated pollution indices; (iii) determine the occurrence, composition, and source characteristics of PAHs and PCBs; (iv) assess non-carcinogenic and carcinogenic health risks; and (v) identify the dominant geogenic and anthropogenic controls using Pearson correlation, principal component analysis, and hierarchical cluster analysis. This integrated approach provides a comprehensive understanding of groundwater quality in petroleum-impacted arid environments and contributes to the development of effective monitoring and management strategies for protecting groundwater resources.

2. Materials and Methods

2.1 Study Area

The study was conducted in Azzan City, Shabwah Governorate, southeastern Yemen, near a crude oil pipeline corridor in one of the major oil-producing regions of the country. The study area lies within an arid to semi-arid climatic zone characterized by high temperatures, low precipitation, and limited natural groundwater recharge, conditions that increase the vulnerability of groundwater resources to persistent contamination. The location of the study area and the distribution of groundwater sampling wells are shown in Figure 1.

Geologically, the area is dominated by formations ranging from the Cretaceous Tawilah Group to recent Quaternary deposits. The Tawilah Group consists mainly of porous and permeable sandstones interbedded with clay layers, which may influence groundwater flow, solute transport, and contaminant migration. These hydrogeological conditions, combined with the presence of petroleum infrastructure, increase the potential for groundwater contamination in zones affected by leakage or localized anthropogenic inputs (Beydoun et al., 1998; Al-Kharbash et al., 2024).

2.2 Groundwater Sampling and Field Procedures

Groundwater sampling was conducted in December 2024. Seven groundwater wells were selected based on their proximity to the crude oil pipeline corridor and their use as local groundwater sources. Prior to sampling, wells were purged to remove stagnant water and obtain representative aquifer water.

For physicochemical and major-ion analyses, groundwater samples were collected in pre-cleaned polyethylene bottles. Samples for trace-metal analysis were filtered through 0.45 μm membrane filters, acidified to $\text{pH} < 2$ using ultrapure nitric acid (HNO_3), and stored at 4 $^{\circ}\text{C}$ until analysis. Samples for PAHs and PCBs were collected separately in pre-cleaned amber glass bottles to minimize photodegradation, volatilization, and adsorption losses. Organic-contaminant samples were stored at 4 $^{\circ}\text{C}$ and transported to the laboratory for extraction and analysis. Sampling, preservation, and handling procedures followed standard protocols for water-quality assessment (Estefan et al., 2013; APHA, 2017).

Geographical coordinates of each well were recorded using a Global Positioning System (GPS). Spatial distribution maps were prepared using ArcGIS 10.8, and interpolation was performed using the inverse distance weighting (IDW) method. The IDW maps were used to visualize spatial trends of selected groundwater-quality parameters and should be interpreted as estimated spatial patterns rather than exact concentration distributions (ESRI, 2011).

2.3 Physicochemical and Major-Ion Analysis

Temperature, pH, and electrical conductivity (EC) were measured in situ using calibrated portable instruments. Total dissolved solids (TDS) were determined using

the gravimetric method. Major cations, including Ca^{2+} , Mg^{2+} , Na^{+} , and K^{+} , were determined together with trace elements using inductively coupled plasma–mass spectrometry (ICP–MS). Major anions and nitrogen-related species, including SO_4^{2-} , NO_3^{-} , NO_2^{-} , PO_4^{3-} , and NH_4^{+} , were analyzed using a HACH DR/2800 spectrophotometer following standard methods (APHA, 2017).

2.4 Determination of Trace Metals by ICP–MS

Trace metals, including Ti, Fe, Al, Cu, Zn, As, Pb, V, Cr, Ni, Cd, and Ba, were analyzed using inductively coupled plasma–mass spectrometry (ICP–MS). Filtered and acidified samples were diluted with deionized water where necessary before analysis. Calibration was performed using multi-element standard solutions, with calibration correlation coefficients maintained at $R^2 > 0.995$. Limits of detection (LOD) and limits of quantification (LOQ) were calculated as three and ten times the standard deviation of blank measurements, respectively. Each sample was analyzed in triplicate.

Quality assurance and quality control procedures included procedural blanks, duplicate samples, spiked samples, and calibration verification standards. Recovery rates ranged between 90% and 110%, indicating acceptable analytical performance (Hou and Jones, 2000; Skoog et al., 2014).

2.5 Determination of Organic Pollutants by GC–MS

Polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) were extracted from groundwater samples using liquid–liquid extraction with dichloromethane and n-hexane at a 1:1 volume ratio (Net et al., 2015). The extracts were dried, concentrated, and analyzed using gas chromatography–mass spectrometry

(GC–MS) with an Agilent 6890 gas chromatograph coupled to a 5973 mass-selective detector. The system operated in electron impact mode at 70 eV, and chromatographic separation was achieved using a DB-5MS capillary column with helium as the carrier gas.

Compound identification was performed by comparing mass spectra with the NIST library and confirmed using retention times of calibration standards where available. Quantification was performed using external calibration curves (USEPA, 2007; Yang et al., 2018). Non-detected compounds were treated as half the limit of detection (LOD/2) for statistical calculations.

PAHs were classified into low-molecular-weight (LMW) and high-molecular-weight (HMW) groups. Diagnostic ratios, including LMW/HMW, Ant/(Ant+Phe), and Phe/(Phe+Ant), were calculated to infer PAH source characteristics (Yunker et al., 2002). The carcinogenic potential of PAHs was assessed using benzo[a]pyrene equivalent concentration (BaP_{eq}) based on toxic equivalency factors (Nisbet and LaGoy, 1992).

2.6 Trace-Metal Pollution Indices

Groundwater contamination by trace metals was evaluated using the contamination factor (CF), degree of contamination (Cdeg), Nemerow pollution index (NPI), heavy metal pollution index (HPI), and heavy metal evaluation index (HEI). These indices were calculated using international drinking-water guideline values and Yemeni standards where applicable.

$$CF_i = \frac{C_i}{C_{ref}} \quad (1)$$

$$Cd = \Sigma CF_i \quad (2)$$

$$NPI = \sqrt{\left[\frac{(CF_{\max}^2 + CF_{\text{mean}}^2)}{2} \right]} \quad (3)$$

$$HPI = \Sigma \frac{(W_i \times Q_i)}{\Sigma W_i} \quad (4)$$

$$Q_i = \left[\frac{(C_i - I_i)}{(S_i - I_i)} \right] \times 100 \quad (5)$$

$$HEI = \Sigma \left(\frac{C_i}{MAC_i} \right) \quad (6)$$

where (C_i) is the measured concentration of metal (i), (C_{ref}) is the reference guideline value, (CF_i) is the contamination factor, (W_i) is the unit weight, (Q_i) is the sub-index of metal (i), (I_i) is the ideal value, (S_i) is the standard permissible value, and (MAC_i) is the maximum admissible concentration. The symbol Cdeg was used for degree of contamination to avoid confusion with Cd as cadmium. Index interpretation followed established classification criteria (Prasad and Bose, 2001; Edet and Offiong, 2002; WHO, 2017).

2.7 Human Health Risk Assessment

Human health risks were assessed through the ingestion exposure pathway following U.S. Environmental Protection Agency guidelines. Non-carcinogenic risk was

evaluated using chronic daily intake (CDI), hazard quotient (HQ), and hazard index (HI), whereas carcinogenic risk was evaluated using cancer risk (CR), total cancer risk (TCR), and incremental lifetime cancer risk (ILCR) for PAHs.

$$CDI = \frac{(C \times IR \times EF \times ED)}{(BW \times AT)} \quad (7)$$

$$HQ = \frac{CDI}{RfD} \quad (8)$$

$$HI = \Sigma HQ \quad (9)$$

$$CR = CDI \times SF \quad (10)$$

$$TCR = \Sigma CR \quad (11)$$

where (C) is the contaminant concentration in groundwater, (IR) is the ingestion rate, (EF) is the exposure frequency, (ED) is the exposure duration, (BW) is body weight, (AT) is the averaging time, (RfD) is the oral reference dose, and (SF) is the oral cancer slope factor. The exposure and toxicity parameters used for the health-risk calculations are provided in Supplementary Table S1. HQ and HI values greater than 1 indicate potential non-carcinogenic health risk, whereas the commonly acceptable carcinogenic risk range is 10^{-6} – 10^{-4} (USEPA, 1989; USEPA, 2011; WHO, 2017; Li et al., 2021).

For PAHs, carcinogenic risk was assessed using benzo[a]pyrene equivalent concentration (BaP_{eq}), chronic daily intake, and incremental lifetime cancer risk (ILCR). ILCR values exceeding 10^{-4} were considered indicative of elevated carcinogenic risk.

2.8 Statistical Analysis

Descriptive statistics, including minimum, maximum, mean, standard deviation, and median values, were calculated to summarize groundwater-quality parameters. Pearson correlation analysis was applied to evaluate relationships among physicochemical parameters, trace metals, PAHs, and PCBs. Principal component analysis (PCA) was used to identify the dominant geogenic and anthropogenic controls on groundwater chemistry and contamination patterns. Hierarchical cluster analysis (HCA), based on Euclidean distance and the average linkage method, was used to classify groundwater samples according to their hydrochemical and contamination similarities.

Before multivariate analysis, variables were standardized to reduce the influence of different measurement scales. Statistical analyses were performed using IBM SPSS Statistics software. Non-detected values were treated as LOD/2 to minimize bias in statistical calculations (Shrestha and Kazama, 2007; Zhang et al., 2020).

2.9 Quality Assurance and Quality Control

Strict quality assurance and quality control procedures were applied throughout sampling and laboratory analysis. Instruments were calibrated using certified standards before analysis, and calibration verification was performed during analytical runs. Procedural blanks, duplicate samples, and spiked samples were included in each

analytical batch. Recovery values between 90% and 110% were considered acceptable. Charge balance error for major ions was maintained within $\pm 5\%$, confirming the reliability of the hydrochemical dataset. Detection and quantification limits were sufficiently low to support trace-level measurement of metals and organic contaminants.

3. Results and Discussion

3.1 Groundwater Quality and Pollution Assessment

3.1.1 Physicochemical Properties and Dominant Major-Ion Composition

Groundwater physicochemical properties and dominant major-ion composition in the study area showed marked spatial variability, particularly in pH, electrical conductivity (EC), total dissolved solids (TDS), Ca^{2+} , Mg^{2+} , and SO_4^{2-} . This variability reflects the combined influence of hydrogeological conditions, aquifer lithology, water–rock interactions, groundwater flow dynamics, and possible localized contamination sources (Freeze and Cherry, 1979; Todd and Mays, 2005). These key parameters are summarized in Table 1. The spatial variation of pH, EC, and TDS is illustrated in Figure 2, while the spatial distribution of the dominant major ions is shown in Figure 3. The complete physicochemical and major cation–anion dataset is provided in Supplementary Table S2. Major-cation concentration plots, selected anion and nitrogen-related species plots, and spatial maps of selected secondary ions are presented in Supplementary Figures S1–S3. Spatial interpolation was performed using the inverse distance weighting (IDW) method; therefore, the generated maps should be interpreted as estimated spatial trends rather than exact concentration distributions.

Groundwater temperature ranged from 31.5 to 33.5 °C, with a mean value of 32.5 °C and a standard deviation of 0.8 °C, indicating relatively stable subsurface thermal conditions primarily controlled by aquifer properties rather than short-term atmospheric variability. Minor inter-well differences may be attributed to variations in groundwater flow regimes, residence time, and lithological conditions. The highest temperatures were recorded in the Mahdi and Bathnin wells, whereas the lowest value was observed in Al-Mashroa B. Although temperature was not retained in the main figure to improve readability, the complete temperature dataset is reported in Supplementary Table S2.

The pH values ranged from 6.9 to 9.7, with a mean of 8.31, indicating neutral to strongly alkaline groundwater conditions. Several samples exceeded the recommended drinking-water pH range of 6.5–8.5 (WHO, 2022). The alkaline character of the groundwater is mainly associated with carbonate mineral dissolution, particularly calcite and dolomite, which enhances bicarbonate buffering and regulates groundwater pH within the aquifer system (Hem, 1985; Appelo and Postma, 2005).

EC values ranged from 1015 to 4740 $\mu\text{S}/\text{cm}$, with a mean of 2746.4 $\mu\text{S}/\text{cm}$ and a standard deviation of 1369.5 $\mu\text{S}/\text{cm}$, indicating substantial spatial variability in groundwater salinity. The highest EC value was recorded in the Mahdi well, reflecting intensive mineralization associated with the dissolution of geological salts, with possible contributions from localized anthropogenic sources. TDS ranged from 406 to 3078 mg/L, with a mean of 1975.1 mg/L, confirming high groundwater mineralization in several samples. The close association between EC and TDS indicates that groundwater salinity is mainly controlled by dissolved ionic constituents derived from geochemical processes within the aquifer (Hem, 1985; Todd and Mays, 2005).

The dominant cationic composition was characterized by elevated Ca^{2+} and Mg^{2+} concentrations. Calcium ranged from 209.66 to 595.59 mg/L, with a mean of 347.19 mg/L, while magnesium ranged from 52.20 to 253.64 mg/L, with a mean of 153.96 mg/L. The highest Ca^{2+} and Mg^{2+} concentrations were recorded in the Mahdi well, indicating intensive water–rock interaction and the dissolution of carbonate and sulfate-bearing minerals. This pattern is consistent with the dissolution of calcite, dolomite, and gypsum, which commonly controls Ca^{2+} and Mg^{2+} enrichment in groundwater systems (Hem, 1985; Subramani et al., 2005). The complete cation dataset, including K^{+} and Na^{+} , is provided in Supplementary Table S2.

Among the measured anions, SO_4^{2-} was the dominant species. Sulfate concentrations ranged from 260 to 2100 mg/L, with a mean of 904.29 mg/L. The highest sulfate concentration was observed in the Atash well, followed by Mahdi, suggesting the dissolution of sulfate minerals such as gypsum and anhydrite (Appelo and Postma, 2005; Zhang et al., 2020; Li et al., 2021). Possible additional contributions from hydrocarbon-related or anthropogenic sources cannot be excluded, particularly in wells showing elevated salinity and major-ion enrichment. The complete anion dataset, including NO_3^{-} , PO_4^{3-} , and NO_2^{-} , as well as NH_4^{+} as a nitrogen-related cationic species, is reported in Supplementary Table S2.

Elevated Ca^{2+} , Mg^{2+} , and SO_4^{2-} concentrations corresponded closely with high EC and TDS values, particularly in Mahdi, Atash, Bathnin, Qurayna A, and Qurayna B. This relationship indicates that groundwater salinity is primarily controlled by dissolved ions produced through carbonate dissolution, sulfate mineral dissolution, and water–rock interaction within the aquifer system (Appelo and Postma, 2005; Todd and Mays, 2005). The hydrochemical composition is therefore characterized by

a dominant Ca–Mg–SO₄ facies, reflecting the influence of carbonate and sulfate mineral dissolution processes (Subramani et al., 2005).

Spatially, Mahdi, Atash, Bathnin, Qurayna A, and Qurayna B exhibited higher EC, TDS, pH, Ca²⁺, Mg²⁺, and SO₄²⁻ values than Al-Mashroa A and Al-Mashroa B. This spatial heterogeneity highlights the combined role of natural geochemical processes, local lithological variability, groundwater flow dynamics, and possible hydrocarbon-related contamination sources in shaping groundwater quality within the study area (Zhang et al., 2020; Li et al., 2021).

3.1.2 Trace Metals, Pollution Indices, and Health Risks

3.1.2.1 Distribution of Trace Metals in Groundwater

Trace-metal concentrations in groundwater showed pronounced spatial variability, reflecting the combined influence of geogenic processes and localized anthropogenic inputs (Appelo and Postma, 2005). The concentrations of the analyzed elements, including Ti, Fe, Al, Cu, Zn, As, Pb, V, Cr, Ni, Cd, and Ba, are summarized in Table 2 and illustrated in Figure 4. The spatial distribution of selected toxic metals is shown in Figure 5, whereas the spatial distribution of additional trace metals is provided in Supplementary Figure S6.

The results indicate considerable heterogeneity among the studied elements, particularly Ti, Cu, Zn, Pb, and Cd, as reflected by their high inter-well variability. Titanium exhibited the highest mean concentration, reaching 11.31 mg/L, with a maximum value of 37.44 mg/L recorded in the Atash well, suggesting a strong lithogenic contribution from mineral-rich geological formations. Iron also showed elevated concentrations, with a mean value of 3.72 mg/L, exceeding permissible

drinking-water limits in all samples. This pattern reflects the dissolution of iron-bearing minerals within the aquifer system (Hem, 1985).

Toxic trace elements, particularly Pb, As, and Cd, showed substantial exceedances relative to international drinking-water guidelines (WHO, 2022). Lead concentrations reached up to 0.347 mg/L, with a mean value of 0.151 mg/L, far exceeding the guideline value of 0.01 mg/L, especially in the Atash, Bathnin, and Qurayna B wells. Arsenic recorded a mean concentration of 0.135 mg/L, which is markedly higher than the guideline value of 0.01 mg/L and represents a serious health concern because of its carcinogenic toxicity. Cadmium concentrations also exceeded the guideline value of 0.003 mg/L in several wells, particularly in Atash, suggesting localized anthropogenic contamination that may be linked to petroleum-related activities.

To place these findings within a regional hydrochemical context, the Azzan results were compared with groundwater studies from other Yemeni basins. The concentrations of Pb and As observed in Azzan were substantially higher than those reported for Al-Dhalia (Saleh et al., 2020) and Bir Ahmed (Saleh and Al-Sallami, 2022). Previous investigations in the Wadi Zabid basin generally reported trace-metal concentrations within WHO guideline limits and an overall good water-quality status (Saleh et al., 2017a; Al-Alimi et al., 2023). This contrast suggests that the elevated metal concentrations in Azzan are unlikely to be explained solely by regional lithogenic background conditions. Instead, they indicate localized anthropogenic inputs, most plausibly associated with petroleum-related activities along the pipeline corridor.

The spatial distribution maps reveal distinct contamination hotspots. Mahdi and Atash consistently exhibited elevated concentrations of several metals, particularly Pb and

Cd, whereas Al-Mashroa A and Al-Mashroa B showed comparatively lower concentrations. This spatial pattern supports the interpretation that groundwater quality deterioration is controlled by both natural mineral dissolution and localized contamination inputs.

3.1.2.2 Pollution Indices

Integrated pollution indices were calculated to provide a quantitative assessment of trace-metal contamination in groundwater. The Heavy Metal Pollution Index (HPI), Heavy Metal Evaluation Index (HEI), degree of contamination (Cdeg), and Nemerow Pollution Index (NPI) are summarized in Table 3, while element-specific contamination factor (CF) values are provided in Supplementary Table S3.

The HPI reached 738.42 based on WHO guideline values and 529.12 based on Yemeni standards, both far exceeding the critical threshold of 100. This indicates that groundwater in the study area is unsuitable for drinking due to severe trace-metal pollution (Prasad and Bose, 2001). Similarly, HEI values exceeded the critical threshold of 20 in all wells, with the highest values recorded in Atash and Mahdi, confirming very high contamination levels. These results demonstrate that groundwater deterioration is cumulative and driven by multiple trace metals, particularly Pb, As, Cd, and Fe (Edet and Offiong, 2002).

The CF results further indicate that Pb, As, Fe, and Cd are the dominant contributors to groundwater contamination, whereas Zn, Ba, and Cu show comparatively lower contributions. The Cdeg values reached 49.91 based on WHO guideline values and 38.48 based on Yemeni standards, indicating a high degree of contamination. NPI values were 11.23 and 9.94, respectively, far exceeding the severe-pollution threshold

of 3. These results confirm severe groundwater pollution and support the classification of the studied wells as highly contaminated.

3.1.2.3 Health Risk Assessment

Non-carcinogenic and carcinogenic health risks associated with trace-metal exposure are summarized in Table 4, while the exposure and toxicity parameters used for risk calculation are provided in Supplementary Table S1. Non-carcinogenic risk was evaluated using chronic daily intake (CDI), hazard quotient (HQ), and hazard index (HI), whereas carcinogenic risk was assessed using cancer risk (CR) and total cancer risk (TCR).

Arsenic exhibited the highest HQ value of 12.87, exceeding the safe threshold by more than twelvefold. Lead also exceeded the acceptable limit, with an HQ value of 1.23. The overall HI reached 16.19, indicating a significant cumulative non-carcinogenic health risk for exposed populations (USEPA, 2011). These results identify As and Pb as the principal contributors to non-carcinogenic risk in the study area.

The carcinogenic risk assessment showed that arsenic contributed the highest CR value, reaching 5.79×10^{-3} , followed by cadmium and chromium. The TCR reached 8.79×10^{-3} , greatly exceeding the commonly acceptable risk range of 10^{-6} – 10^{-4} . This indicates a very high long-term carcinogenic risk associated with the consumption of contaminated groundwater (USEPA, 2011).

3.1.2.4 Integrated Interpretation

The integrated assessment of trace-metal concentrations, pollution indices, and health-risk parameters demonstrates that groundwater in the study area is severely

contaminated and poses significant environmental and public-health concerns. The elevated concentrations of toxic metals, together with high HPI, HEI, CF, Cdeg, NPI, HQ, HI, and TCR values, confirm that contamination is cumulative and results from multiple interacting sources.

Mahdi and Atash represent the principal contamination hotspots, as indicated by both concentration patterns and pollution-index results. The co-occurrence of elevated toxic metals with high salinity and major-ion enrichment suggests the combined influence of natural geochemical processes, including mineral dissolution, and localized anthropogenic inputs, most plausibly related to petroleum activities along the pipeline corridor. These findings highlight the need for continuous monitoring, source-control measures, and sustainable groundwater management to reduce environmental and human-health risks in Azzan City.

3.1.3 Organic Contaminants in Groundwater: PAHs and PCBs

Organic contaminants were detected in all groundwater samples, indicating that groundwater quality in the study area is affected not only by inorganic pollutants but also by petroleum-related and persistent organic compounds. The integrated summary of PAH and PCB concentrations, compositional indicators, source-diagnostic ratios, and carcinogenic risk parameters is presented in Table 5. Variations in total PAHs and total PCBs across wells are illustrated in Figure 6 (a and b), while their spatial distribution patterns are shown in Figure 7 (a and b). Detailed compound-specific PAH and PCB data are provided in Supplementary Tables S4–S6.

Total PAH concentrations ranged from 0.014 to 0.104 mg/L, with a mean value of 0.068 mg/L (Table 5; Supplementary Table S4). The highest concentrations were recorded in the Mahdi, Qurayna B, and Atash wells, whereas the lowest concentration

was observed in Al-Mashroa A. This distribution, also illustrated in Figure 6a and Figure 7a, indicates localized organic-contamination hotspots rather than uniform regional contamination.

The PAH compositional profile was strongly dominated by low-molecular-weight compounds, particularly naphthalene, which accounted for approximately 89.08% of total PAHs (Table 5; Supplementary Table S4; Supplementary Figure S4). Phenanthrene represented about 7.14%, followed by acenaphthene at approximately 3.57%, whereas benzo[b]fluoranthene was detected only in trace amounts. The dominance of low-molecular-weight PAHs indicates relatively high solubility and mobility in groundwater and is consistent with petroleum-derived contamination rather than combustion-derived inputs (Neff, 1979; Yunker et al., 2002).

The molecular distribution and diagnostic ratios further support a petrogenic PAH source (Table 5; Supplementary Table S5). Low-molecular-weight PAHs dominated all samples, with LMW/HMW ratios exceeding unity in every well and reaching 93 in the Atash well. Anthracene was not detected in any sample, resulting in Ant/(Ant+Phe) ratios of 0 and Phe/(Phe+Ant) ratios of 1 across all wells. These diagnostic patterns strongly indicate a petroleum-related origin of PAHs. The high proportion of naphthalene, particularly in Atash and Mahdi, suggests recent or continuous input of petroleum hydrocarbons into the groundwater system.

The carcinogenic risk associated with PAH exposure was evaluated using benzo[a]pyrene equivalent concentration (BaP_{eq}), chronic daily intake (CDI), and incremental lifetime cancer risk (ILCR) (Table 5; Supplementary Table S5). ILCR values ranged from 1.20×10^{-5} to 1.64×10^{-4} , with the highest value recorded in Atash, followed by Mahdi and Qurayna B. Several values exceeded the upper

acceptable risk limit of 10^{-4} , indicating significant long-term carcinogenic risk associated with groundwater consumption (USEPA, 2011; WHO, 2022). The spatial correspondence between elevated ILCR values and high PAH concentrations confirms that petroleum-related organic contamination represents a relevant public-health concern in the study area.

PCBs were detected in all groundwater samples, with total concentrations ranging from 0.085 to 0.374 mg/L and a mean value of 0.216 mg/L (Table 5; Supplementary Table S6A). The highest PCB concentration was recorded in Qurayna B, followed by Bathnin, Atash, and Mahdi. This pattern, shown in Figure 6b and Figure 7b, indicates spatially variable but widespread contamination by persistent organic pollutants.

The PCB compositional profile was dominated by tetrachlorinated congeners, particularly 2,2',4,5'-tetrachlorobiphenyl, which accounted for approximately 70.93% of total PCBs, followed by 2,2',3,5'-tetrachlorobiphenyl at approximately 14.93% (Supplementary Table S6A; Supplementary Figure S4). When grouped by degree of chlorination, tetrachlorinated PCBs dominated all wells, while LowCl/TetraCl ratios remained below unity (Table 5; Supplementary Table S6B). This pattern indicates the prevalence of relatively persistent PCB congeners. The presence of pentachlorinated congeners, particularly in Mahdi, further suggests long-term environmental persistence and possible historical contamination (ATSDR, 2000; WHO, 2017).

The integrated spatial analysis of PAHs and PCBs reveals a consistent organic-contamination pattern (Figure 7). Mahdi, Atash, Qurayna B, and Bathnin represent important organic-contamination hotspots, although the relative dominance differs between PAHs and PCBs. PAHs were highest in Mahdi, Qurayna B, and Atash, whereas PCBs were highest in Qurayna B and Bathnin. The spatial overlap between

organic contaminants and trace-metal enrichment indicates a multi-component contamination system involving petroleum-derived hydrocarbons, persistent organic pollutants, and toxic metals.

Overall, the dominance of low-molecular-weight PAHs, high LMW/HMW ratios, petrogenic diagnostic signatures, elevated PCB concentrations, and partial spatial correspondence with trace-metal hotspots strongly suggest that groundwater contamination in Azzan City is influenced by localized anthropogenic sources, most plausibly related to petroleum infrastructure and associated industrial activities. These findings highlight the need for continuous monitoring of organic contaminants, source-control measures, and groundwater protection strategies to reduce long-term environmental and public-health risks.

3.2 Statistical Analysis of Groundwater Samples

3.2.1 Pearson Correlation Analysis

Pearson correlation analysis was performed to evaluate the relationships among physicochemical parameters, trace metals, and organic contaminants in groundwater. The significant correlation coefficients are provided in Supplementary Table S7A.

The correlation structure showed strong associations between salinity-related parameters and contaminant variables. A particularly strong positive relationship was observed between Σ PAHs and EC ($r = 0.961$, $p < 0.01$), as well as between Σ PAHs and TDS ($r = 0.914$, $p < 0.01$), suggesting that elevated organic contamination is closely associated with increased groundwater salinity and reflects the interaction between natural hydrogeochemical processes and localized anthropogenic inputs (Freeze and Cherry, 1979; Appelo and Postma, 2005; Zhang et al., 2020).

The strong positive correlation between ΣPAHs and ΣPCBs ($r = 0.882$, $p < 0.01$) indicates either a common source or similar transport behavior within the aquifer system. This association is consistent with petroleum-related contamination and the co-occurrence of persistent organic pollutants in affected groundwater zones (ATSDR, 2000; Yunker et al., 2002). In contrast, nitrate (NO_3^-) showed negative correlations with TDS and ΣPAHs , suggesting that nitrate may originate from different source pathways, such as surface-derived inputs, agricultural activities, or wastewater discharge, rather than petroleum-related contamination (WHO, 2022).

3.2.2 Principal Component Analysis

Principal component analysis (PCA) was applied to distinguish the dominant geogenic and anthropogenic controls on groundwater chemistry. Detailed PCA loadings are provided in Supplementary Tables S7B and S7C, while the PCA biplots are presented in Figure 8.

The first PCA model, based on physicochemical parameters and major ions, showed that PC1 explained 82.17% of the total variance, while PC2 accounted for 14.00%, with a cumulative variance of 96.18%. EC, TDS, Ca^{2+} , Mg^{2+} , Na^+ , and SO_4^{2-} exhibited strong positive loadings on PC1, indicating that this component represents natural geochemical processes, particularly carbonate and sulfate mineral dissolution. These processes control groundwater salinity and ionic composition within the aquifer system (Hem, 1985; Appelo and Postma, 2005). The close alignment of EC, TDS, Ca^{2+} , Mg^{2+} , Na^+ , and SO_4^{2-} in Figure 8a further confirms their common geogenic origin and strong interdependence.

The second PCA model, incorporating trace metals and organic contaminants, showed that PC1 explained 63.68% of the total variance. In this model, ΣPAHs , ΣPCBs , Pb,

and NH_4^+ showed strong positive loadings along the same axis, indicating a close association among petroleum-related organic contaminants, toxic metals, and nitrogen-related species. This pattern suggests a common anthropogenic influence, most plausibly related to petroleum activities and localized contamination inputs (Alloway, 2013; Zhang et al., 2020). In contrast, NO_3^- and As showed negative loadings, suggesting different source controls or distinct geochemical behavior. The opposite orientation of NO_3^- relative to ΣPAHs , ΣPCBs , and Pb in Figure 8b further supports the separation between surface-derived inputs and petroleum-related contamination.

Overall, PCA results indicate that groundwater quality in the study area is controlled by two dominant processes. The first process is geogenic and reflects mineral dissolution and water–rock interaction, which define the hydrochemical baseline of the aquifer. The second process is anthropogenic and is associated with petroleum-derived contamination, persistent organic pollutants, and toxic trace metals.

3.2.3 Hierarchical Cluster Analysis

Hierarchical cluster analysis (HCA) was performed using Euclidean distance and the average linkage method to classify groundwater samples according to their hydrochemical and contamination similarities (Hair et al., 2010). The dendrogram is provided in Supplementary Figure S5. The sample codes SW-1 to SW-7 correspond respectively to Al-Mashroa A, Al-Mashroa B, Mahdi, Atash, Bathnin, Qurayna A, and Qurayna B.

The HCA results separated the groundwater samples into three main clusters. The first cluster includes Al-Mashroa A and Al-Mashroa B, representing wells with relatively low contamination levels. The second cluster includes Mahdi and Atash, which are

characterized by the highest contamination levels and elevated concentrations of trace metals and organic pollutants. The third cluster includes Bathnin, Qurayna A, and Qurayna B, representing intermediate contamination conditions. This classification confirms the spatial heterogeneity of groundwater quality and supports the presence of localized contamination sources within the study area.

3.2.4 Quantitative Assessment of Anthropogenic Contamination Signatures

Beyond conventional concentration-based interpretation, the interaction among different contaminant groups provides a more comprehensive characterization of the groundwater contamination fingerprint. The quantitative assessment of anthropogenic contamination signatures, derived from PCA loadings, contaminant ratios, and correlation patterns, is summarized in Table 6.

Total PCBs exhibited higher mean concentrations than total PAHs, with mean values of 0.216 mg/L and 0.068 mg/L, respectively. This indicates a greater mass contribution of persistent organic pollutants to the overall contaminant load. However, PCA results showed that PAHs had a higher loading value (0.980) than PCBs (0.872), suggesting that PAHs are more strongly associated with recent contamination dynamics and spatial variability across the study area. This interpretation is supported by the strong positive correlation between $\sum$ PAHs and EC ($r = 0.961$; Supplementary Table S7A), indicating a close linkage between organic contamination and groundwater salinity.

In the identified contamination hotspots, particularly Mahdi and Atash, the interaction between trace metals and nitrogen-related species further highlights the complexity of the contamination system. Ammonium showed a mean concentration of 0.350 mg/L, whereas Pb exhibited a higher PCA loading of 0.838, indicating a stronger statistical

influence on the overall contamination structure and associated health risks. This finding is consistent with the elevated non-carcinogenic risk levels reported in the health-risk assessment, where Pb contributed to the cumulative hazard index (Table 4).

These results demonstrate that groundwater contamination in the study area is governed not only by the absolute concentration of pollutants but also by their relative statistical influence, persistence, and interaction within the hydrochemical system. The combined interpretation of concentration levels, correlation patterns, and PCA loadings provides a quantitative basis for distinguishing dominant contamination drivers from secondary contributors.

3.2.5 Integrated Statistical Interpretation and Monitoring Implications

The combined results of Pearson correlation analysis, PCA, and HCA provide consistent evidence that groundwater quality in the study area is controlled by a dual system involving natural geogenic processes and localized anthropogenic contamination. Correlation analysis highlights the relationships among salinity, organic contaminants, and trace metals (Supplementary Table S7A), while PCA differentiates between geogenic controls related to mineral dissolution and anthropogenic controls linked to petroleum-derived contaminants (Figure 8; Supplementary Tables S7B and S7C). HCA further confirms spatial variability by grouping wells according to contamination intensity (Supplementary Figure S5).

The strong positive correlation between EC and Σ PAHs ($r = 0.961$; Supplementary Table S7A) suggests that easily measurable physicochemical parameters may provide useful screening indicators for identifying petroleum-related contamination trends in groundwater. Although predictive models were not implemented in this study, the

observed statistical relationships provide a conceptual basis for developing future data-driven monitoring approaches. Previous studies have demonstrated the applicability of machine learning and soft-computing techniques, including artificial neural networks and random forest models, for predicting groundwater quality under complex nonlinear conditions (Haghiabi et al., 2018; Rajaei et al., 2019).

In the present study, the dominance of low-molecular-weight PAHs, particularly naphthalene, together with elevated salinity levels and strong EC–PAH correlation, supports the potential use of EC as an indirect screening indicator of organic-contamination dynamics. However, EC should not be considered a substitute for direct PAH and PCB analysis; rather, it may serve as a rapid field-based indicator to guide targeted laboratory testing.

Overall, the integration of hydrochemical parameters, correlation analysis, PCA, HCA, and quantitative contamination signatures provides a robust framework for understanding groundwater contamination in Azzan City. The results confirm that groundwater quality is controlled by the interaction between natural mineral dissolution and anthropogenic inputs, particularly petroleum-related activities. This dual-control mechanism is summarized conceptually in Figure 9, which illustrates the interaction between the Ca–Mg–SO₄ hydrochemical baseline, petroleum-derived organic contaminants, toxic trace metals, salinity indicators, contamination hotspots, and associated health risks. This integrated statistical interpretation supports the development of focused monitoring programs, source-control strategies, and groundwater management measures in petroleum-affected regions.

4. Conclusion

Groundwater quality in Azzan City is controlled by the interaction between natural hydrogeochemical processes and anthropogenic contamination, forming a complex multi-component pollution system. Elevated EC and TDS values, together with the dominance of the Ca–Mg–SO₄ hydrochemical facies, indicate that mineral dissolution and water–rock interaction govern the natural hydrochemical baseline of the aquifer.

Superimposed on this geogenic background, trace-metal contamination represents a major deterioration factor. Concentrations of Pb, As, Cd, and Fe exceeded international drinking-water guideline values, while integrated pollution indices, including HPI, HEI, CF, Cdeg, and NPI, consistently confirmed severe groundwater contamination. Mahdi and Atash were identified as the principal contamination hotspots, reflecting the cumulative impact of multiple toxic elements.

Human-health risk assessment further confirmed the seriousness of groundwater contamination. The non-carcinogenic hazard index exceeded the acceptable threshold, mainly due to As and Pb exposure, while the total carcinogenic risk substantially exceeded the acceptable range, indicating significant long-term health risks associated with groundwater consumption.

Organic contaminants intensified the overall pollution status. PAHs were dominated by low-molecular-weight compounds, especially naphthalene, and diagnostic ratios indicated a predominantly petrogenic origin. Elevated ILCR values, particularly in Atash, Mahdi, and Qurayna B, indicated carcinogenic risk associated with long-term exposure to PAH-contaminated groundwater. PCBs were also detected in all samples, with tetrachlorinated congeners dominating the PCB profile, reflecting persistent organic contamination with long environmental residence times.

Multivariate statistical analysis, including Pearson correlation, PCA, and HCA, provided consistent evidence for a dual-control mechanism. Geogenic processes define the baseline groundwater chemistry, whereas anthropogenic inputs, most plausibly associated with petroleum-related activities and localized wastewater influences, drive the observed enrichment of toxic metals, PAHs, and PCBs. The co-occurrence of salinity, trace metals, and organic pollutants in specific wells confirms the presence of localized contamination hotspots.

This study provides an integrated assessment framework combining hydrochemistry, trace metals, petroleum-related organic pollutants, health-risk assessment, and multivariate statistical analysis. The findings highlight the urgent need for continuous groundwater monitoring, stricter control of petroleum-related activities, targeted remediation of hotspot wells, and protection measures to reduce environmental and public-health risks. Without effective intervention, groundwater resources in the study area may undergo progressive degradation, limiting their suitability for human consumption and increasing long-term ecological and health concerns.

References

- Al Manmi DAMA, Abdullah TO, Al-Jaf PM, Al-Ansari N (2019) Soil and groundwater pollution assessment and delineation of intensity risk map in Sulaymaniyah City, NE of Iraq. *Water* 11(10):2158. <https://doi.org/10.3390/w11102158>
- Al-Alimi AK, Rimawi O, Al-Alaiy SHG, Saleh SMK (2023) Assessment of household drinking water quality in Zabied City, Yemen. *International Journal of Water Resources and Arid Environments* 12(1):31–45.

Al-Kharbash SA, Al-Wosabi MA, Khanbari KM, Al-Abbawi MI, Habtour AM (2024) *Geology of Yemen*. 2nd edn. Arabian Gulf Library for Printing, Publishing and Distribution, Sana'a, Yemen.

Alloway BJ (2013) *Heavy Metals in Soils: Trace Metals and Metalloids in Soils and Their Bioavailability*. 3rd edn. Springer, Dordrecht, Netherlands.
<https://doi.org/10.1007/978-94-007-4470-7>

APHA (2017) *Standard Methods for the Examination of Water and Wastewater*. 23rd edn. American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, USA. <https://www.standardmethods.org/>

Appelo CAJ, Postma D (2005) *Geochemistry, Groundwater and Pollution*. 2nd edn. A.A. Balkema Publishers, Rotterdam, Netherlands.

ATSDR (2000) *Toxicological Profile for Polychlorinated Biphenyls (PCBs)*. Agency for Toxic Substances and Disease Registry, U.S. Department of Health and Human Services, Atlanta, GA, USA.

Beydoun ZR, As-Saruri ML, Baraba RS, Nani AO, Al-Aawah MH, Al-Hiti MA (1998) *International Lexicon of Stratigraphy. Volume III, Asia, Fascicle 10b1: Republic of Yemen*. UNESCO, Paris, France.

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(4):295–304.
<https://doi.org/10.1023/B:GEJO.0000007250.92458.de>

ESRI (2011) *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, CA, USA.

Estefan G, Sommer R, Ryan J (2013) *Methods of Soil, Plant, and Water Analysis: A Manual for the West Asia and North Africa Region*. International Center for

Agricultural Research in the Dry Areas (ICARDA), Beirut, Lebanon.

<https://repo.mel.cgiar.org/items/14bb2a3f-dc9e-455b-99e3-97f27ceb9e42>

Freeze RA, Cherry JA (1979) *Groundwater*. Prentice-Hall, Englewood Cliffs, NJ, USA.

Haghiabi AH, Nasrolahi AH, Parsaie A (2018) Water quality prediction using machine learning methods. *Water Quality Research Journal* 53(1):3–13.

<https://doi.org/10.2166/wqrj.2018.025>

Hair JF, Black WC, Babin BJ, Anderson RE (2010) *Multivariate Data Analysis*. 7th edn. Pearson Prentice Hall, Upper Saddle River, NJ, USA.

Hem JD (1985) *Study and Interpretation of the Chemical Characteristics of Natural Water*. 3rd edn. U.S. Geological Survey Water-Supply Paper 2254. U.S. Government Printing Office, Washington, DC, USA. <https://pubs.usgs.gov/publication/wsp2254>

Hou X, Jones BT (2000) Inductively coupled plasma/optical emission spectrometry. In: Meyers RA (ed) *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Chichester, pp 9468–9485.

Huling SG, Weaver JW (1991) *Dense Nonaqueous Phase Liquids (DNAPLs)*. EPA Ground Water Issue, EPA/540/4-91-002. U.S. Environmental Protection Agency, Washington, DC, USA.

Jughmam A, Habtour A (2020) *Environmental Oil Pollution and Its Effects in Yemen: A Technical Environmental Study*. Utara Company, Sana'a, Yemen.

Moeini A, Alizadeh Paenafrakaty E (2008) Groundwater quality and the sources of pollution in Baghan watershed, Iran. *International Journal of Environmental, Ecological, Geological and Geophysical Engineering* 2(7):12–17.
<https://publications.waset.org/15773/groundwater-quality-and-the-sources-of-pollution-in-baghan-watershed-iran>

Neff JM (1979) *Polycyclic Aromatic Hydrocarbons in the Aquatic Environment: Sources, Fates and Biological Effects*. Applied Science Publishers, London, UK.

Net S, Sempéré R, Delmont A, Paluselli A, Ouddane B (2015) Occurrence, fate, behavior and ecotoxicological state of phthalates in different environmental matrices. *Environmental Science & Technology* 49(7):4019–4035. <https://doi.org/10.1021/es505233b>

Nisbet ICT, LaGoy PK (1992) Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regulatory Toxicology and Pharmacology* 16(3):290–300. [https://doi.org/10.1016/0273-2300\(92\)90009-X](https://doi.org/10.1016/0273-2300(92)90009-X)

Prasad B, Bose J (2001) Evaluation of the heavy metal pollution index for surface and spring water near a limestone mining area of the lower Himalayas. *Environmental Geology* 41(1–2):183–188. <https://doi.org/10.1007/s002540100380>

Rajae T, Ebrahimi H, Nourani V (2019) A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology* 572:336–351. <https://doi.org/10.1016/j.jhydrol.2018.12.037>

Saleh SMK, Al-Alaiy SHG, Abdul-Razzak BI (2017b) Application of water quality index to assessment of groundwater quality. *Journal of University of Aden for Basic and Applied Sciences* 21(1):125–134. <https://doi.org/10.47372/uajnas.2017.n1.a13>

Saleh SMK, Al-Alaiy SHG, Abdul-Razzak BI, Nasher GSA (2017a) Evaluation of groundwater quality and its suitability for drinking and agricultural use of rural areas for Zabid Directorate-Wadi Zabid, Hodeidah, Yemen. *Journal of Scientific and Engineering Research* 4(7):10–24.

Saleh SMK, Taher MM, Saif BOA (2020) Estimation of heavy metals in the groundwater of selected areas from Al-Dhalia District, Al-Dhalia Governorate,

Yemen. *Electronic Journal of University of Aden for Basic and Applied Sciences* 1(3):151–158. <https://doi.org/10.47372/ejua-ba.2020.3.37>

Saleh SMK, Al-Sallami AMA (2022) Assessment of the level of physicochemical and microbiological contamination of groundwater in parts of Bir Nasser and Bir Ahmed water fields in Tuban Delta in Aden and Lahej Governorates, Yemen. *Electronic Journal of University of Aden for Basic and Applied Sciences* 3(2):101–116. <https://doi.org/10.47372/ejua-ba.2022.2.158>

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques. *Environmental Modelling & Software* 22(4):464–475. <https://doi.org/10.1016/j.envsoft.2006.02.001>

Skoog DA, Holler FJ, Crouch SR (2014) *Principles of Instrumental Analysis*. 6th edn. Cengage Learning, Boston, MA, USA.

Smith K (2001) *Environmental Hazards: Assessing Risk and Reducing Disaster*. 3rd edn. Routledge, London, UK.

Subramani T, Elango L, Damodarasamy SR (2005) Groundwater quality and its suitability for drinking and irrigation purposes in Chithar River Basin, Tamil Nadu, India. *Environmental Geology* 47(8):1099–1110. <https://doi.org/10.1007/s00254-005-1243-0>

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

Todd DK, Mays LW (2005) *Groundwater Hydrology*. 3rd edn. John Wiley & Sons, New York, NY, USA.

Ugochukwu UC, Ochonogor A (2018) Groundwater contamination by polycyclic aromatic hydrocarbons due to diesel spill from a telecom base station in a Nigerian

city: assessment of human health risk exposure. *Environmental Monitoring and Assessment* 190:249. <https://doi.org/10.1007/s10661-018-6626-2>

USEPA (1989) *Risk Assessment Guidance for Superfund (RAGS), Volume I: Human Health Evaluation Manual (Part A)*. EPA/540/1-89/002. U.S. Environmental Protection Agency, Washington, DC, USA.

USEPA (2007) *Method 8270D: Semivolatile Organic Compounds by Gas Chromatography/Mass Spectrometry (GC/MS)*. Revision 4. U.S. Environmental Protection Agency, Washington, DC, USA. <https://www.epa.gov/hw-sw846/sw-846-test-method-8270d-semivolatile-organic-compounds-gas-chromatographymass>

USEPA (2011) *Exposure Factors Handbook: 2011 Edition*. EPA/600/R-09/052F. U.S. Environmental Protection Agency, Washington, DC, USA. <https://www.epa.gov/expobox/exposure-factors-handbook-2011-edition>

Wang L, Yang Z, Niu J, Wang J (2009) Characterization, ecological risk assessment and source diagnostics of polycyclic aromatic hydrocarbons in water column of the Yellow River Delta, one of the most biodiversity-rich zones in the world. *Journal of Hazardous Materials* 169(1–3):460–465. <https://doi.org/10.1016/j.jhazmat.2009.03.125>

WHO (2017) *Guidelines for Drinking-water Quality*. 4th edn. World Health Organization, Geneva, Switzerland. <https://www.who.int/publications/i/item/9789241549950>

WHO (2022) *Guidelines for Drinking-water Quality: Fourth Edition Incorporating the First and Second Addenda*. World Health Organization, Geneva, Switzerland. <https://www.who.int/publications/i/item/9789240045064>

Yang C, Wang Z, Brown CE, Landriault M, Yang Z, Hollebone B, Lambert P, Zhang G (2018) Fingerprinting analysis and source differentiation of petroleum-

contaminated environmental samples. In: Wang Z, Stout SA (eds) *Oil Spill Environmental Forensics Case Studies*. Elsevier, Oxford, pp 49–65.
<https://doi.org/10.1016/B978-0-12-804434-6.00003-3>

Yunker MB, Macdonald RW, Vingarzan R, Mitchell RH, Goyette D, Sylvestre S (2002) PAHs in the Fraser River basin: A critical appraisal of PAH ratios. *Organic Geochemistry* 33(4):489–515. [https://doi.org/10.1016/S0146-6380\(02\)00002-5](https://doi.org/10.1016/S0146-6380(02)00002-5)

Zhang Q, Xu P, Qian H (2020) Groundwater quality assessment using improved water quality index (WQI) and human health risk (HHR) evaluation in a semi-arid region of northwest China. *Exposure and Health* 12(3):487–500.
<https://doi.org/10.1007/s12403-020-00345-w>

Declarations

Acknowledgments

The authors sincerely thank the Higher Institute of Water Sciences and Techniques (ISSTEG), University of Gabès, Tunisia, and the Department of Oil and Gas, Faculty of Oil and Minerals, Shabwah University, Shabwah, Yemen, for their support and facilities provided during this study.

Funding:

The authors declare that no funds were obtained for this study.

Competing Interests:

The authors declare that there are no conflicts of interest to disclose.

Author contributions:

Salem Salem Hussein Sofui: Conceptualization; Methodology; Investigation; Data curation; Formal statistical analysis (Pearson, PCA, HCA); Formal analysis; Visualization (graphs); Writing – original draft.

Shaif Mohammed Kasem Saleh: Conceptualization; Methodology validation (chemical fingerprinting and predictive modeling logic); Formal analysis; Data validation (PCA and HCA); Writing – review and editing.

Sabrine Jemai: Visualization (maps); Cartography; Data presentation.

M. Safa Gamal: Physicochemical analysis; Laboratory investigation.

Abdelmonem M. Habtoor: Field investigation; Sample collection.

Mahmoud Bali: Supervision; Writing – review and editing.

Ethical Approval:

This study did not involve human participants, human tissues, vertebrate animals, or identifiable personal data. Therefore, ethical approval was not required.

Consent to Participate:

Not applicable.

Consent to Publish:

Not applicable.

Data Availability Statement:

The data supporting the findings of this study are available within the article. Additional data are available from the corresponding author upon reasonable request.

Clinical trial number: not applicable.

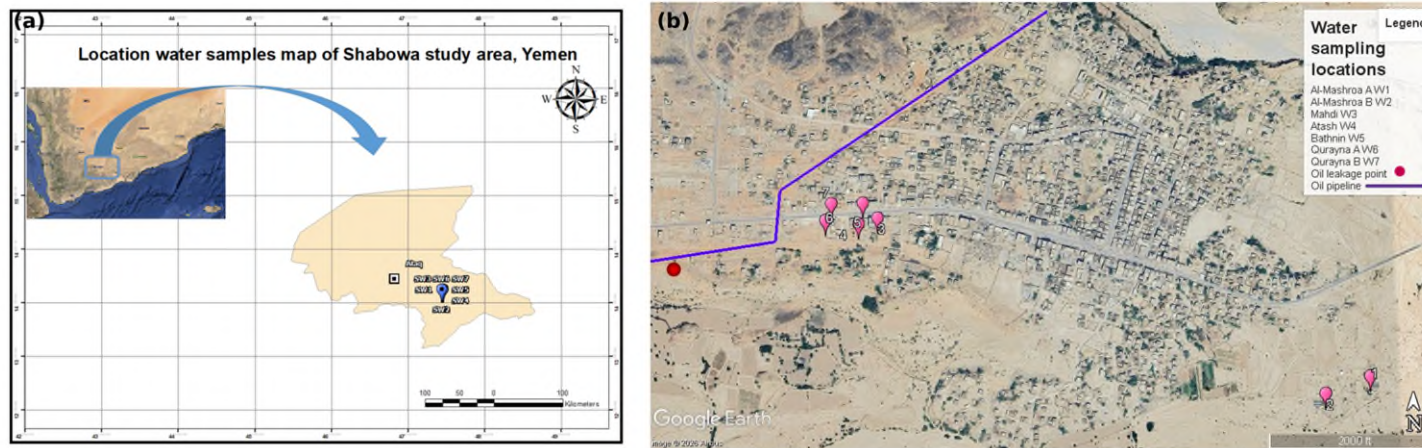


Figure 1. Study area and sampling locations in Azzan City, Shabwah Governorate, Yemen.

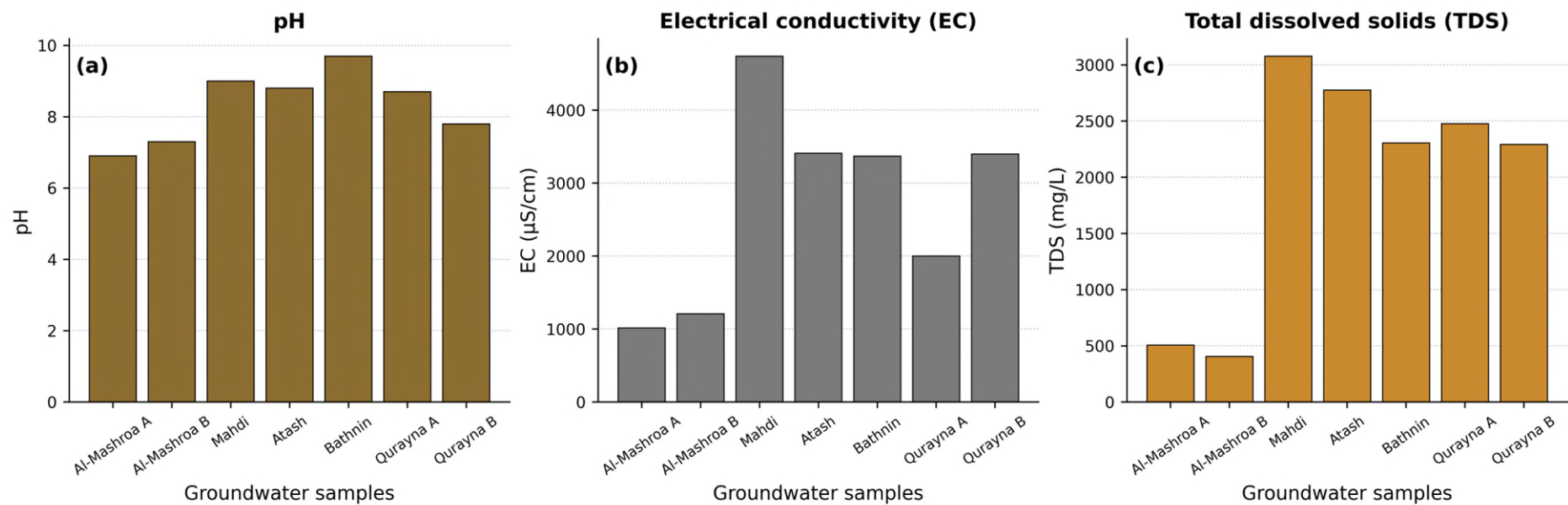


Figure 2. Spatial distribution of physicochemical parameters in groundwater: (a) pH, (b) EC, and (c) TDS.

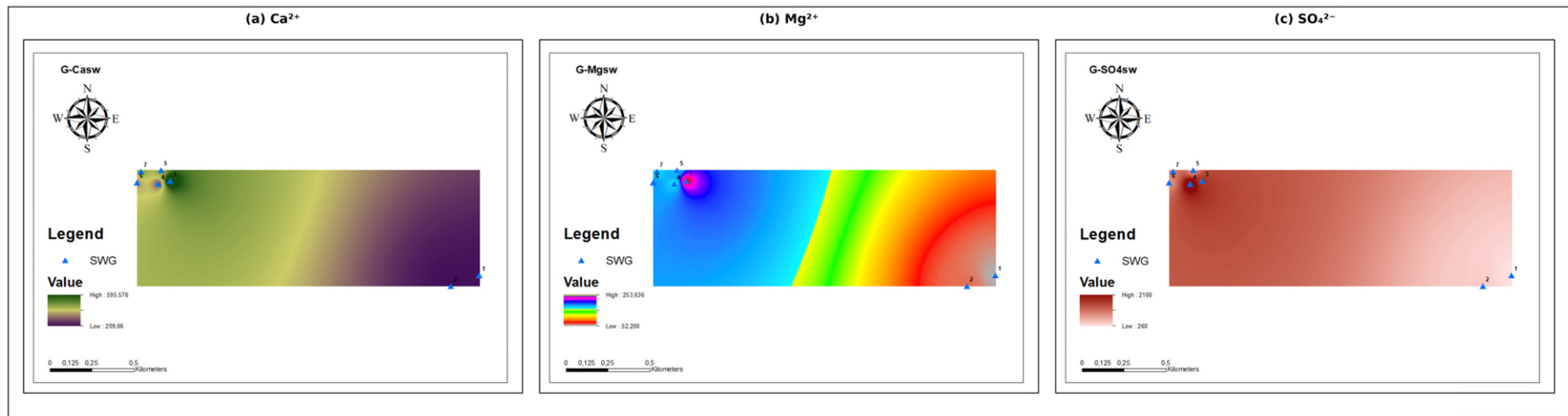


Figure 3. Spatial distribution of major ions in groundwater using IDW interpolation: (a) Ca^{2+} , (b) Mg^{2+} , and (c) SO_4^{2-} .

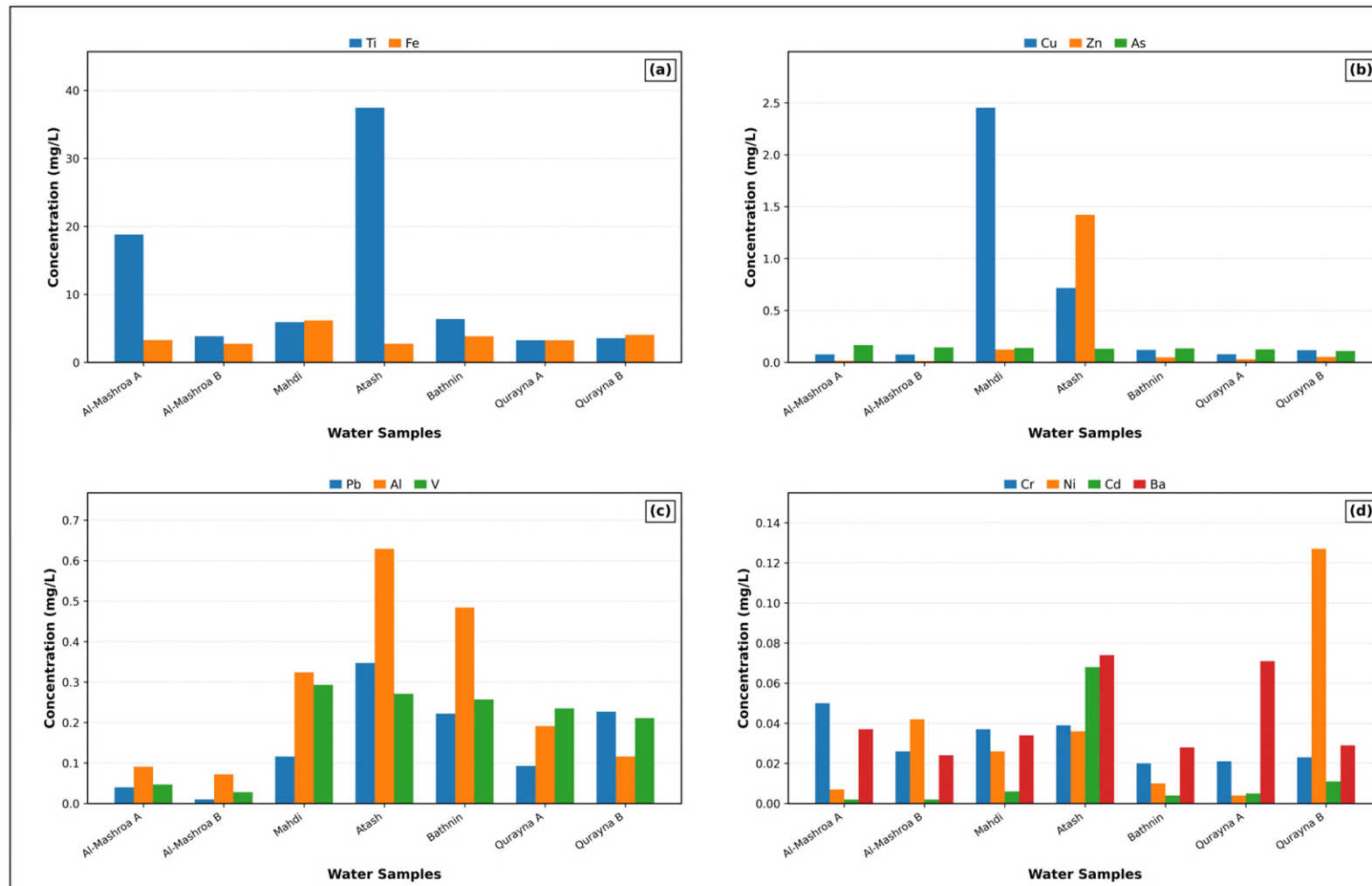


Figure 4. Trace-metal concentrations in groundwater samples: (a) Ti and Fe; (b) Cu, Zn, and As; (c) Pb, Al, and V; and (d) Cr, Ni, Cd, and Ba.

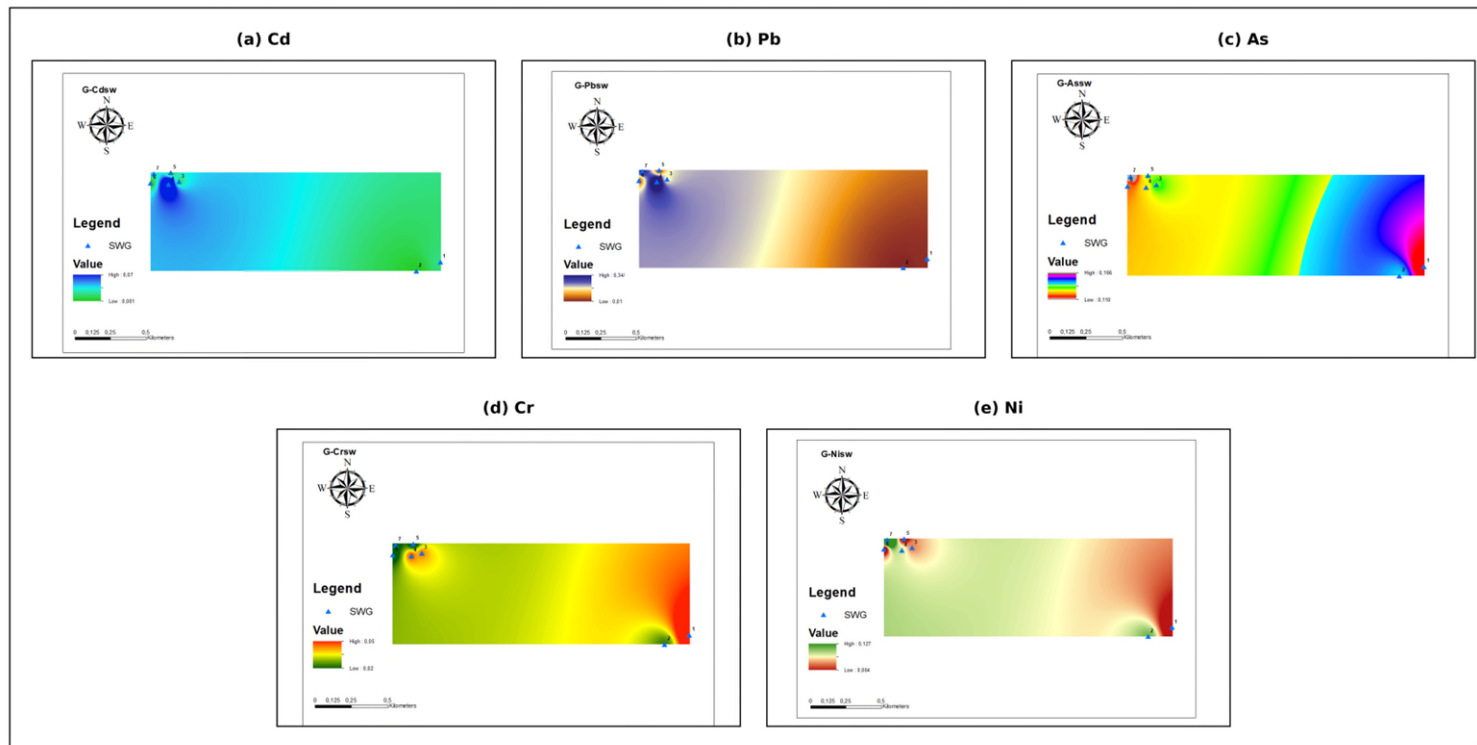


Figure 5. Spatial distribution of selected toxic trace metals in groundwater using IDW interpolation: (a) Cd, (b) Pb, (c) As, (d) Cr, and (e) Ni.

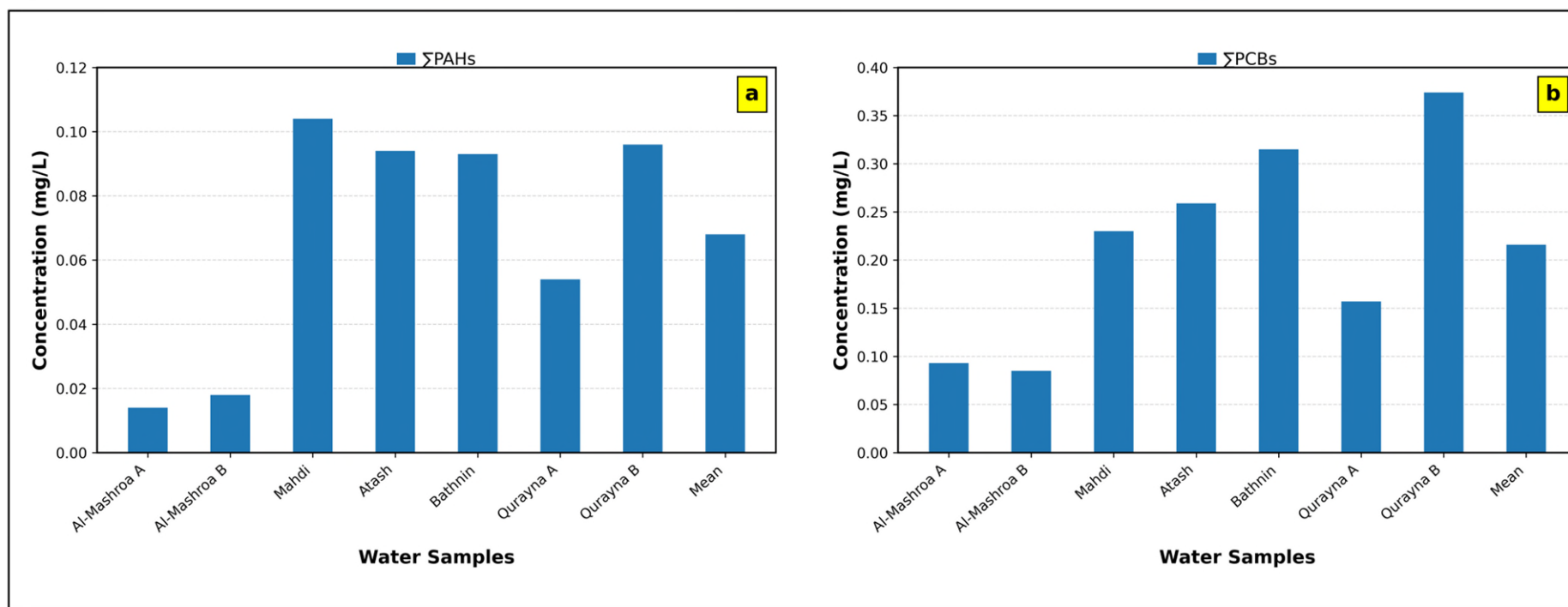


Figure 6. Total organic contaminants in groundwater samples: (a) Σ PAHs and (b) Σ PCBs.

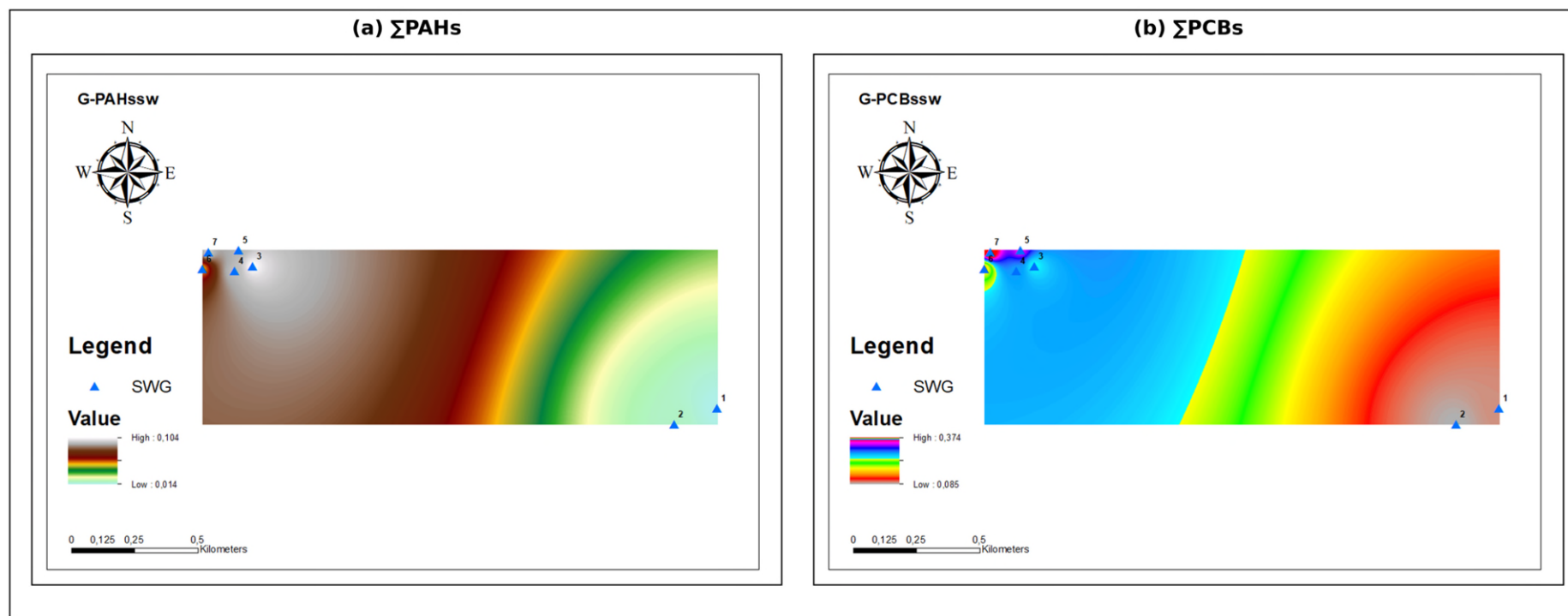


Figure 7. Spatial distribution of petroleum-related and persistent organic contaminants in groundwater using IDW interpolation: (a) Σ PAHs and (b) Σ PCBs.

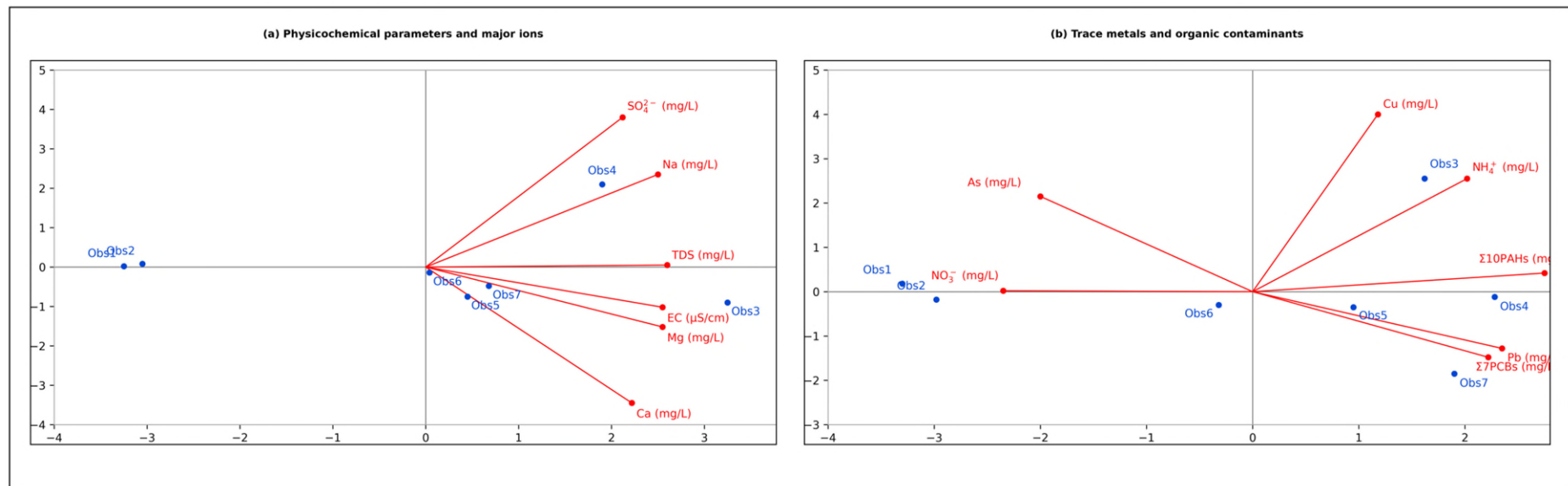


Figure 8. PCA biplots showing geogenic and anthropogenic controls on groundwater quality: (a) physicochemical parameters and major ions; (b) trace metals and organic contaminants.

Dual-Control Mechanism of Groundwater Contamination in Azzan City, Yemen

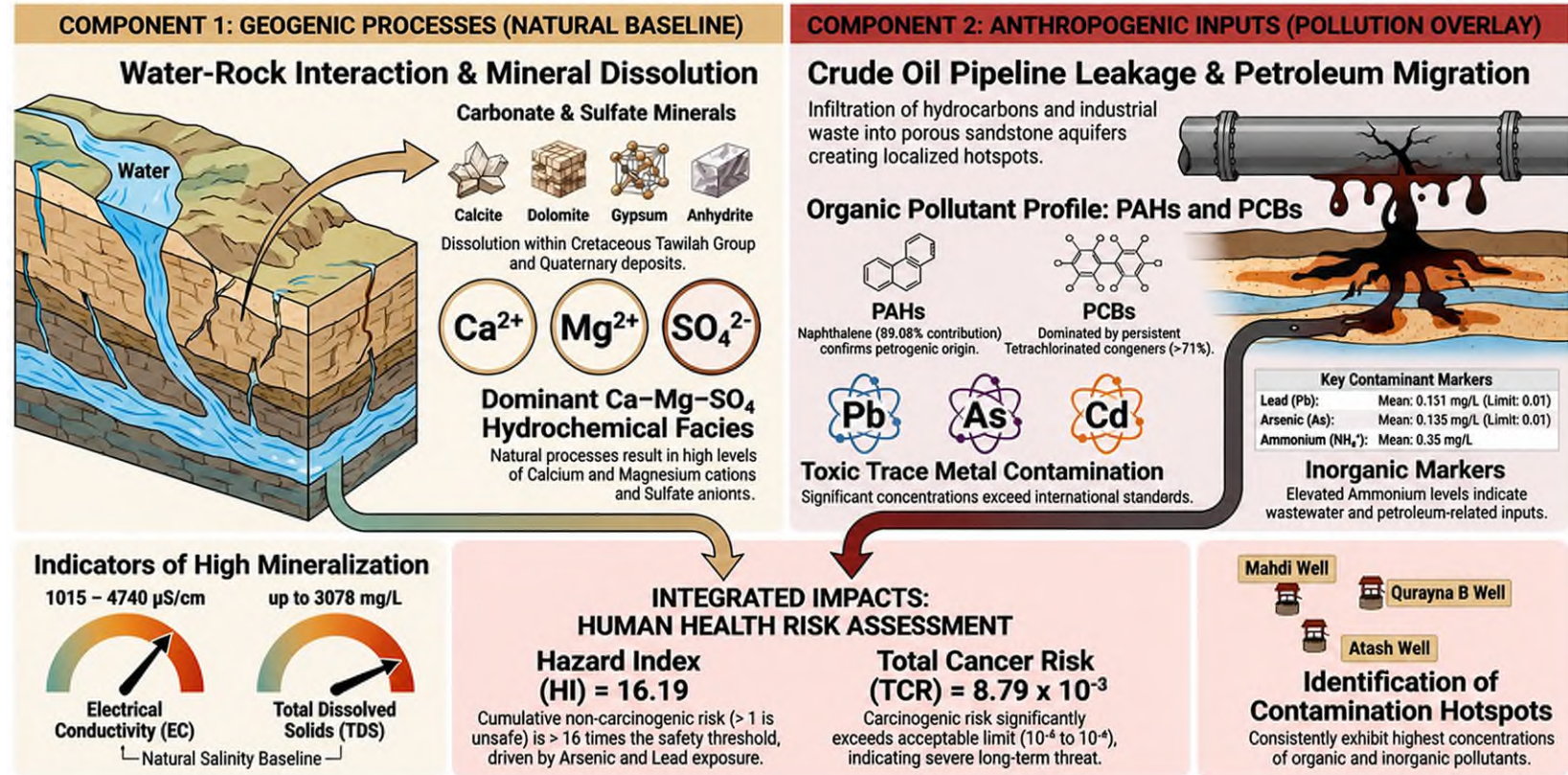


Figure 9. Integrated conceptual framework of geogenic and petroleum-related controls on groundwater contamination in Azzan City, Yemen.

Table 1. Summary of physicochemical parameters and dominant major ions in groundwater samples from Azzan City.

Sample	T (°C)	pH	EC (μS/cm)	TDS (mg/L)	Ca²⁺ (mg/L)	Mg²⁺ (mg/L)	SO₄²⁻ (mg/L)
Al-Mashroa A	32.0	6.9	1015.0	506.6	212.77	52.20	260.00
Al-Mashroa B	31.5	7.3	1210.0	406.0	209.66	63.69	380.00
Mahdi	33.5	9.0	4740.0	3078.0	595.59	253.64	1380.00
Atash	32.8	8.8	3410.0	2777.0	284.10	168.80	2100.00
Bathnin	33.5	9.7	3400.0	2291.0	391.32	177.14	620.00
Qurayna A	32.0	8.7	2000.0	2476.0	327.31	189.57	700.00
Qurayna B	32.0	7.8	3450.0	2291.0	409.55	172.69	890.00
Min	31.5	6.9	1015.0	406.0	209.66	52.20	260.00
Max	33.5	9.7	4740.0	3078.0	595.59	253.64	2100.00
Mean	32.5	8.3	2746.4	1975.1	347.19	153.96	904.29
SD	0.8	1.0	1369.5	1074.6	134.72	71.67	641.45
Median	32.0	8.7	3400.0	2291.0	327.31	172.69	700.00

Note: EC = electrical conductivity; TDS = total dissolved solids; SD = standard deviation.

Table 2. Trace-metal concentrations in groundwater samples from Azzan City.

Sample	Ti	Fe	Al	Cu	Zn	As	Pb	V	Cr	Ni	Cd	Ba
Al-Mashroa A	18.794	3.268	0.091	0.076	0.015	0.166	0.040	0.047	0.050	0.007	0.002	0.037
Al-Mashroa B	3.847	2.744	0.072	0.075	0.013	0.143	0.010	0.028	0.026	0.042	0.002	0.024
Mahdi	5.919	6.154	0.324	2.453	0.123	0.138	0.116	0.293	0.037	0.026	0.006	0.034
Atash	37.444	2.747	0.629	0.717	1.421	0.130	0.347	0.271	0.039	0.036	0.068	0.074
Bathnin	6.363	3.846	0.484	0.120	0.048	0.134	0.222	0.257	0.020	0.010	0.004	0.028
Qurayna A	3.255	3.236	0.191	0.078	0.030	0.125	0.093	0.235	0.021	0.004	0.005	0.071
Qurayna B	3.551	4.044	0.116	0.118	0.053	0.110	0.227	0.211	0.023	0.127	0.011	0.029
Min	3.255	2.744	0.072	0.075	0.013	0.110	0.010	0.028	0.020	0.004	0.002	0.024
Max	37.444	6.154	0.629	2.453	1.421	0.166	0.347	0.293	0.050	0.127	0.068	0.074
Mean	11.310	3.720	0.272	0.520	0.243	0.135	0.151	0.192	0.031	0.036	0.014	0.042
SD	12.737	1.182	0.216	0.884	0.521	0.017	0.120	0.109	0.011	0.043	0.024	0.021
Median	5.919	3.268	0.191	0.118	0.048	0.134	0.116	0.235	0.026	0.026	0.005	0.034

Note: All concentrations are expressed in mg/L. SD = standard deviation.

Table 3. Integrated trace-metal pollution indices in groundwater samples from Azzan City.

Index	Standard / Basis	Value / Range	Classification / Interpretation
HPI	WHO guideline values	738.42	Unsuitable for drinking
HPI	Yemeni standards	529.12	Unsuitable for drinking
HEI	WHO guideline values	24.56–82.90	High contamination in all wells
HEI	Yemeni standards	28.23–52.09	High contamination in all wells
Cdeg	WHO guideline values	49.91	Very high contamination
Cdeg	Yemeni standards	38.48	Very high contamination
NPI	WHO guideline values	11.23	Severe pollution
NPI	Yemeni standards	9.94	Severe pollution

Note: HPI = Heavy Metal Pollution Index; HEI = Heavy Metal Evaluation Index; Cdeg = degree of contamination; NPI = Nemerow Pollution Index. HPI > 100 indicates water unsuitable for drinking; HEI > 20 indicates high contamination; NPI > 3 indicates severe pollution.

Table 4. Non-carcinogenic and carcinogenic health risks associated with trace metals in groundwater from Azzan City.

Element	Concentration (mg/L)	CDI (mg/kg/day)	HQ	Slope factor	CR
Pb	0.1507	0.00431	1.23	0.0085	3.7×10^{-5}
Al	0.2724	0.00778	0.008	—	—
V	0.1917	0.00548	0.78	—	—
Cr	0.0309	0.00088	0.29	0.5	4.4×10^{-4}
Ni	0.0360	0.00103	0.05	—	—
Cd	0.0140	0.00040	0.40	6.3	2.52×10^{-3}
Ba	0.0424	0.00121	0.006	—	—
Fe	3.7199	0.10628	0.15	—	—
Cu	0.5196	0.01485	0.37	—	—
Zn	0.2433	0.00695	0.02	—	—
As	0.1351	0.00386	12.87	1.5	5.79×10^{-3}
Total	—	—	HI = 16.19	—	TCR = 8.79×10^{-3}

Note: CDI = chronic daily intake; HQ = hazard quotient; HI = hazard index; CR = cancer risk; TCR = total cancer risk. HQ and HI > 1 indicate potential non-carcinogenic health risk. The acceptable cancer-risk range is 10^{-6} – 10^{-4} ; values above this range indicate elevated carcinogenic risk.

Table 5. Integrated summary of PAHs and PCBs in groundwater samples from Azzan City.

Sample	Σ PAHs (mg/L)	Dominant PAH	Naphthalene (%)	LMW/HMW	PAH source	PAH-ILCR	Σ PCBs (mg/L)	Dominant PCB pattern
Al-Mashroa A	0.014	Naphthalene	64.29	>>1	Petrogenic	1.20×10^{-5}	0.093	Tetra-chlorinated dominance
Al-Mashroa B	0.019	Phenanthrene / Naphthalene	31.58	>>1	Petrogenic	1.63×10^{-5}	0.086	Tetra-chlorinated dominance
Mahdi	0.104	Naphthalene	92.31	>>1	Petrogenic	8.91×10^{-5}	0.231	Tetra-chlorinated dominance
Atash	0.094	Naphthalene	94.68	93	Petrogenic	1.64×10^{-4}	0.259	Tetra-chlorinated dominance
Bathnin	0.093	Naphthalene	94.62	>>1	Petrogenic	7.96×10^{-5}	0.315	Tetra-chlorinated dominance
Qurayna A	0.055	Naphthalene	90.91	>>1	Petrogenic	4.72×10^{-5}	0.157	Tetra-chlorinated dominance
Qurayna B	0.097	Naphthalene	88.66	>>1	Petrogenic	8.30×10^{-5}	0.374	Tetra-chlorinated dominance
Min	0.014	—	31.58	—	—	1.20×10^{-5}	0.085	—
Max	0.104	—	94.68	—	—	1.64×10^{-4}	0.374	—
Mean	0.068	—	89.08*	—	Petrogenic dominance	—	0.216	Tetra-chlorinated dominance
SD	0.039	—	—	—	—	—	0.110	—
Median	0.093	—	—	—	—	—	0.230	—

Note: PAHs = polycyclic aromatic hydrocarbons; PCBs = polychlorinated biphenyls; LMW = low-molecular-weight PAHs; HMW = high-molecular-weight PAHs; ILCR = incremental lifetime cancer risk; SD = standard deviation. “>>1” indicates that HMW PAHs were not detected or were negligible. *The value 89.08% represents the overall contribution of naphthalene to total PAHs across all samples.

Table 6. Quantitative assessment of anthropogenic contamination signatures derived from PCA loadings, contaminant ratios, and correlation patterns.

Contaminant category	Mean concentration (mg/L)	PCA loading (weight)	Relative contribution (%)	Key diagnostic ratio / indicator
Total PAHs	0.068	0.980	28.7	PAHs/EC correlation: $r = 0.961$
Total PCBs	0.216	0.872	25.5	PCBs/PAHs mass ratio: 3.2:1
Lead (Pb)	0.151	0.838	24.5	Pb/ NH_4^+ PCA-loading ratio: 1.15:1
Ammonium (NH_4^+)	0.350	0.727	21.3	$r(\text{NH}_4^+ - \text{TDS}) = \text{weak}$

Note: Relative contribution values were estimated from normalized PCA loadings to represent the relative influence of each parameter within the overall contamination structure.

PAHs = polycyclic aromatic hydrocarbons; PCBs = polychlorinated biphenyls; EC = electrical conductivity; NH_4^+ = ammonium.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphicalAbstract.tiff](#)