



Integrated assessment of human health risks from groundwater pollutants in Nubian Aquifer, Sudan: Combining source apportionment and probabilistic analysis

Musaab A.A. Mohammed^{a,b,*}, Norbert P. Szabó^a, Elamin D. Suliman^c,
Magboul M.S. Siddig^{d,e}, Mohammed N.M. Hassan^f, Péter Szűcs^a

^a Faculty of Earth and Environmental Sciences and Engineering, University of Miskolc, 3515 Miskolc, Egyetemváros, Hungary

^b College of Petroleum Geology and Minerals, University of Bahri, Khartoum, Sudan

^c Faculty of Minerals and Earth Sciences, Nile Valley University, Nile River State, Sudan

^d Department of Soil and Environment Sciences, Faculty of Agriculture, University of Khartoum, Khartoum, Sudan

^e Department of Physical Geography, University of Göttingen, 37077, Göttingen, Germany

^f Faculty of Petroleum and Minerals, Al-Neelain University, Khartoum, Sudan

ARTICLE INFO

Keywords:

Human health
Self-organizing map
Principal components
Monte Carlo
Sobol analysis
Groundwater quality

ABSTRACT

Groundwater contamination is a significant global challenge, particularly in arid and semi-arid regions where groundwater is a primary source for drinking and irrigation. This contamination is closely associated with human health risks, potentially leading to severe diseases and long-term health consequences. In this study, the groundwater quality of the Nubian Aquifer System (NAS) in the Shendi area, Sudan, is assessed to evaluate health risks linked to nitrogen compounds (NO_2 , NO_3 , NH_3) and fluoride (F). The analysis integrates self-organizing maps (SOM), principal component analysis (PCA), and Monte Carlo-based health risk simulations. SOM analysis revealed distinct clustering patterns in groundwater samples, identifying three major hydrochemical trends. PCA indicated that elevated NO_3 concentrations were localized, primarily associated with agricultural runoff, while NO_2 and NH_3 reflected pollution from both agriculture and wastewater. High fluoride concentrations were linked to geogenic sources, particularly water-rock interactions with fluorine-bearing minerals. The Monte Carlo simulation assessed probabilistic health risks, revealing higher mean hazard quotients (HQs) and hazard index (HI) values for children compared to adults. Children's mean HI of 1.06 significantly exceeds the safe threshold, indicating potential non-carcinogenic health hazards. Sobol sensitivity analysis identified the most influential parameters in shaping health risks, including average exposure time, body weight, and exposure duration, with strong parameter interactions amplifying these effects. Among contaminants, NO_3 and F contributed the most to cumulative HI values. These findings underscore the urgent need for targeted interventions, such as advanced water treatment, stricter pollution controls, and public health awareness programs to mitigate groundwater contamination and protect vulnerable populations.

1. Introduction

Groundwater accounts for about 99 % of the Earth's total liquid freshwater (Famiglietti, 2014; Oki and Kanae, 2006), supplies approximately 50 % of the global water used for agricultural irrigation purposes, and represents primary sources of the drinking water for billions of world population (Jasechko and Perrone, 2021). Groundwater supports multiple systems, such as industry, agriculture production, sanitation, and other ecosystems (Griebler and Avramov, 2015; Saito et al.,

2021). Groundwater pollution is considered the most serious problem that hinders healthy life and sustainable development of the global society. Thus, preventing groundwater pollution is strongly related to the UN Sustainable Development Goals (SDGs), specifically goal 3 "Good Health and Well-being", goal 6 "clean water and sanitation", goal 14 "life below water", and goal 15 "life on land", for sustainable management of surface and groundwater, which can be achieved via the development of modern techniques. Additionally, goal 13 "climate change" has also direct connection with sustainable groundwater

* Corresponding author.

E-mail address: musab20501@gmail.com (M.A.A. Mohammed).

<https://doi.org/10.1016/j.envc.2025.101176>

Received 29 January 2025; Received in revised form 12 April 2025; Accepted 4 May 2025

Available online 5 May 2025

2667-0100/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

management globally. Two common sources of groundwater contamination have been identified globally: (1) natural (geogenic) sources (e.g., rock weathering, soil dissolution, seawater intrusion) and (2) anthropogenic sources (e.g., agricultural, industrial, mining, inaccurate waste disposal), and the latter is the most harmful sources, leading to elevate human health risks (Dharmani et al., 2024).

Groundwater pollution has become a critical global concern due to the presence of a wide range of contaminants that threaten water quality and public health (Foster et al., 2013; Karadeniz et al., 2024; Karunanidhi et al., 2021; Mohammed et al., 2025; Seben et al., 2021). Common pollutants include heavy metals, pesticides, industrial chemicals, and emerging contaminants such as pharmaceuticals and microplastics (Kumar et al., 2023; Nti et al., 2024; Rajan et al., 2024; Yüksel et al., 2021). Among these, nitrogen and fluoride compounds are often found in arid and semi-arid regions in elevated concentrations (Duggal and Sharma, 2022; Mohammed et al., 2023b; Yang et al., 2024). The primary drivers of this contamination include intensive agricultural practices (Xiao et al., 2022), where excessive use of nitrogen-based fertilizers leads to nitrate leaching (Moeini and Azhdarpoor, 2021; Patil et al., 2024), and the application of phosphate fertilizers, which can contribute to fluoride accumulation (Gupta et al., 2024; Kumari et al., 2021). Another possible source of these contaminants is geogenic processes, such as mineral dissolution (Su et al., 2021). On the other hand, rapid industrialization and urbanization have exacerbated fluoride contamination through industrial discharges and improper waste management. The over-extraction of groundwater, common in arid and semi-arid regions, further concentrates these pollutants, compounding the health risks and environmental challenges associated with groundwater pollution (Mohammed et al., 2022b).

Identifying groundwater pollutants, the occurrence and exposure characteristics of mixed contaminants, and the combination interactions between pollutants are important to reduce the challenges of distinguishing contamination response from different pollutants. However, understanding these complications requires advanced analytical tools that are capable of handling large datasets and investigating relationships between various hydrochemical parameters. Among these techniques, self-organizing maps (SOMs) and principal component analysis (PCA) are two powerful methods that have been used by multiple studies to accomplish such analyses. For example, Cauch-Kau et al. (2024) applied PCA to examine sources of several ground water pollutants in the North-Central Mexico including fluoride (F), and NO₃, and concluded that F originated from the weathering of felsic magmatic rocks, while agricultural inputs were the primary NO₃ sources. Urme et al. (2025) utilized SOMs to investigate the spatial distribution and related human health risk of heavy metals in the groundwater of NW Bangladesh. Other several studies used PCA and SOMs techniques to determine sources of pollutants in the groundwater and investigate their spatial distributions (Dey et al., 2024; Islam et al., 2025; Mia et al., 2023; Ming et al., 2024; Mohammed et al., 2022a; Shang et al., 2024; Yazman et al., 2024).

The utilization of probabilistic risk evaluation tools, such as Monte Carlo simulations in conjunction with groundwater pollutants, could provide deep understanding of health risks by accounting for uncertainties and variations in exposure conditions, such as water consumption rates, body weight, and contaminant concentrations. In this context, Yu et al. (2024) investigated human health risks-based groundwater pollutants in Chengdu city, China by applying several techniques including Monto Carlo simulation and sensitivity analysis. Jannat et al. (2022) utilized Monto Carlo simulation and sensitivity analysis for tracing human health risks posed by NO₂ and F in the groundwater at the coastal regions of Bangladesh and India. Numerous studies used Monto Carlo simulation and sensitivity analysis techniques to investigate human health risks from groundwater pollutants (Li et al., 2024; Ramesh et al., 2021; Yang et al., 2025). Together, the above-mentioned methodologies enable a comprehensive evaluation of groundwater quality and associated human health risks, facilitating the

development of targeted risk management strategies to protect vulnerable populations and ensure sustainable water resources.

The Shendi area in Sudan faces significant challenges in ensuring the quality of its groundwater resources, which serve as the primary source of drinking and irrigation for local communities. This arid region is characterized by limited rainfall, which exacerbates the reliance on groundwater for both domestic and agricultural purposes. However, rapid population growth, unregulated agricultural practices, and inadequate wastewater management have contributed to the degradation of groundwater quality (Omer and Alzahrane, 2013). These contamination issues are compounded by the lack of comprehensive groundwater monitoring and limited access to water treatment facilities, leaving vulnerable populations at higher risk of adverse health effects. The primary objectives of this study are to conduct a comprehensive hydrogeochemical analysis and health risk assessment focusing on nitrogen compounds and fluoride content in the extensive Nubian aquifer system in the Shendi area, Sudan. As the first study of its kind in this region, it seeks to identify contamination sources and patterns and address natural influences and anthropogenic impacts on groundwater quality in the study area. The findings of this study help to understand regional groundwater quality dynamics in the Nubian aquifer system, identifying contamination hotspots, and guiding sustainable water management practices to protect public health in the arid-zones aquifer system.

2. Study area

The Shendi Sub-Basin, a part of the larger Atbara Basin (Abdallrhman et al., 2022), is located in northern Sudan on the western bank of the River Nile (Fig. 1). It covers an area of approximately 30,000 km², extending around Jebel Abu Telaih and Jebel Elbarda. Shendi area is a mid-sized town in the River Nile State, situated approximately 159 km northeast of Khartoum capital city of Sudan. The area has low hills and flat lands (Fig. 1), with some isolated sedimentary hills, especially in the Northern and Eastern parts. The climate is mostly dry with rare rainfall with an average of 75 mm/year (Belal et al., 2024); winter temperatures drop to 10 °C, while summer can reach 45 °C, with some rain from March to mid-July. Seasonal streams flow through the area, mainly heading towards the Nile River, and are shaped by the land's structure. The population of the Shendi area is approximately 1000,000 inhabitants. Population density is generally low, where inhabitants maintain a nomadic lifestyle, relying on camels, sheep, and goats for their livelihood. These communities depend on water from wells and aqueducts during the rainy season. The majority population is concentrated along the River Nile, where access to water and fertile agricultural land supports settled agricultural practices.

The study area is characterized by a complex geological framework primarily composed of sedimentary rocks, which overlie Precambrian basement complex (Fig. 2a). The sedimentary units include Quaternary superficial deposits and the Shendi Formation. The basement rocks are exposed in the Sabaloka igneous complex to the south. The Sabaloka basement complex primarily consists of gray gneisses interspersed with patches of granulite, dating back to the Pan-African orogeny (Dawoud, 1980). Subsequently, the complex was intruded by poorly foliated granitic rocks and younger post-orogenic granites, including ring complexes (Vail, 1985). The Shendi Formation consists of continental fluvial and lacustrine facies, characterized by siliceous conglomeritic sandstone with substantial clay deposits, dating to the Upper Cretaceous period (Eltijani et al., 2024). Kheiralla (1966) and Whiteman (1971) referred to the "Nubian Sandstone" to describe sediments of variegated colors lying unconformably on the basement and Paleozoic formations, composed of conglomerates, sandstones, and mudstones. The superficial deposits include wadi sediments and recent fan deposits (Babkir et al., 2023). These deposits consist of poorly sorted materials, including boulders, fragments, and leached coarse to fine sediments. East of the Shendi area, extensive mobile sand dunes made up of well-sorted, medium- to

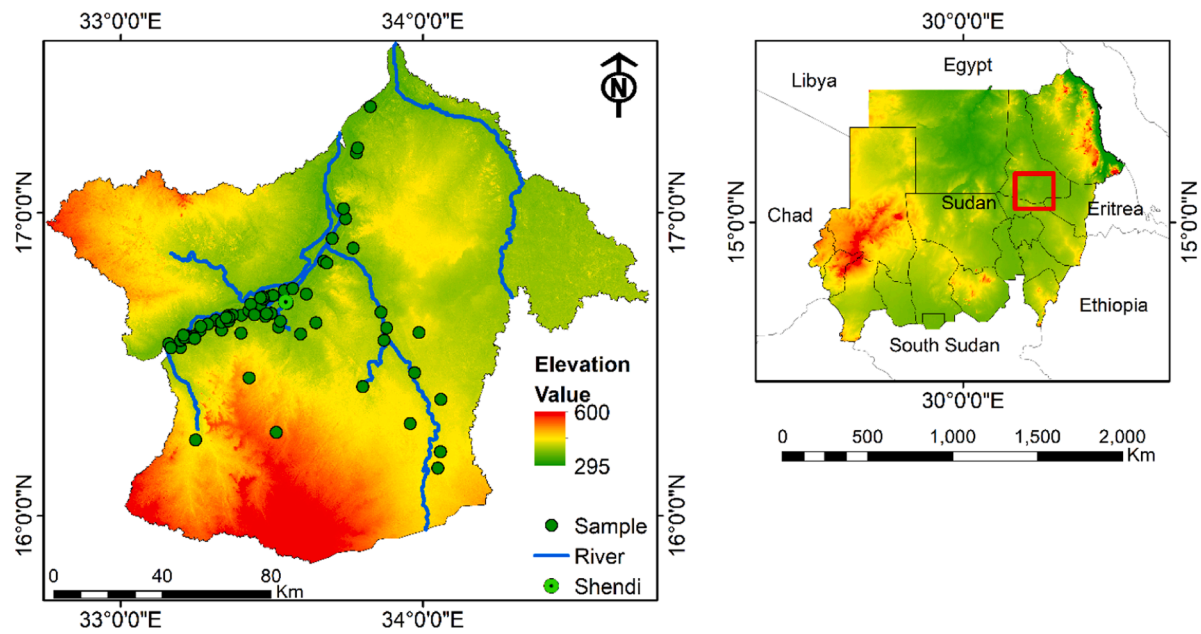


Fig. 1. Geographical location of study area within the Nile River State in Northern Sudan.

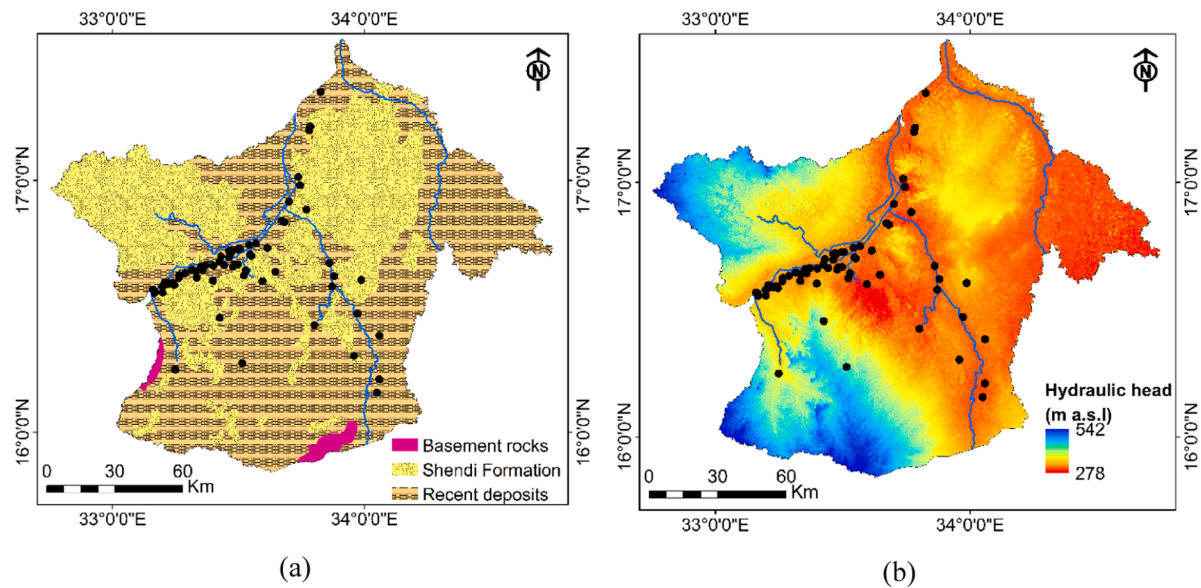


Fig. 2. (a) The surface geology of the study area modified after Whiteman (1971) and (b) The distribution of the hydraulic head in the study area.

fine-grained sand cover the underlying Shendi Formation. Additionally, substantial accumulations of Nile silt are present as superficial deposits along both banks of the River Nile. These deposits consist of very fine sands and silts, deposited annually during seasonal floods.

The hydrogeology of the study area is characterized by two main aquifers in the Cretaceous Shendi Formation: the upper unconfined to semi-confined aquifer and the lower confined aquifer. The upper aquifer varies in thickness from 15 to 66 m, while the lower aquifer has a thickness ranging from 27 to 74 m. The hydraulic head fluctuates between 278 and 542 m above mean sea level (a.m.s.l.) (Fig. 2b). Accordingly, the groundwater flows from the western to eastern part of the study area. The hydraulic conductivity values are assigned as 2.5 m/d for the upper aquifer and 5.58 m/d for the lower aquifer, indicating that the lower aquifer has a higher capacity to transmit water. The specific storage coefficients are 0.000124 for the upper aquifer and 0.00131 for the lower aquifer (Suliman, 2017). The primary sources of

recharge to the aquifers in the area are the Main Nile and seasonal streams. In contrast, direct precipitation, given the semi-arid climate, contributes minimally to recharge (Farah et al., 2000).

3. Materials and methods

3.1. Groundwater sampling and analysis

Groundwater samples were collected using a systematic random sampling approach to ensure representative analysis across the study area. A total of 69 groundwater samples were taken from various locations, predominantly from the lower Nubian sandstone aquifer, at depths ranging from 150 to 180 m. The sampling was conducted during the dry season of 2021, ensuring consistent conditions for hydrochemical analysis. To maintain sample integrity and prevent contamination, all samples were collected in previously cleaned and rinsed

polyethylene bottles following standard protocols outlined by the Sudanese Standards and Metrology Organization (SSMO) guidelines for water sampling and storage. Measures were taken to minimize groundwater degassing by sealing the bottles immediately after collection and avoiding agitation during sampling. Samples were stored at a temperature of 8 °C during transportation to the laboratory to preserve their chemical properties.

Field measurements included pH and total dissolved solids (TDS) using calibrated portable integrated EC-meter. The temperature of the groundwater is measured in the field using thermometers, which typically range from 15 °C to 20 °C. Calcium (Ca) was determined using flame photometry while Chloride (Cl) was analyzed using titration method. Fluoride (F) measured using an ion-selective electrode. Nitrogenous compounds including Ammonia (NH₃), Nitrite (NO₂), and Nitrate (NO₃) were quantified using spectrophotometric methods. To ensure quality control (QC) and quality assurance (QA) of the analysis, duplicate samples and field blanks were collected to check for consistency and potential contamination during sampling and handling. Standardized methods were employed for the analysis of physiochemical parameters, and calibration of instruments was performed. Additionally, analytical precision was monitored, and results were cross verified using ion balance calculations (Appelo and Postma, 2005). Any deviation in ion balance exceeding 10 % was reviewed and corrected where necessary.

3.3. Self organizing map (SOM)

SOM is an unsupervised neural network that is highly effective for clustering and visualizing high-dimensional data by projecting it onto a lower-dimensional space. In this study, SOM analysis is conducted within the MATLAB framework to cluster groundwater samples based on their chemical signature and analyze the relationships between various hydrochemical parameters. The SOM is initially proposed by (Kohonen, 1982), and operates on the principle of competitive learning. In this application, each neuron in the SOM represents a groundwater sample, characterized by a weight vector (W_i) of the same dimension as the input data $X = (x_1, x_2, \dots, x_n)$. The training process involves iteratively presenting meteorological and hydrological parameters to the network, identifying the best matching unit (BMU) which is the neuron whose weight vector is most like the input data, and updating the weights of the BMU and its neighboring neurons based on Euclidean Distance. The BMU can be found using Eq. (2) as

$$BMU = \operatorname{argmin}_i = \|X - W_i(t)\| \quad (2)$$

Where argmin_i is the argument of the minimum over all indices i , $\|\cdot\|$ represents the Euclidean distance $W_i(t)$ is the weight vector of node i at time step t . The BMU is the neuron with the smallest Euclidean distance to the input vector. In other words, it is the neuron that minimizes the equation. Once the BMU is identified, the weight vectors of the BMU and its neighboring nodes are updated to move closer to the input vector. The weight update can be performed using Eq. (3) as

$$W_i(t+1) = W_i(t) + \eta(t) \cdot h_{i, BMU}(t) \cdot X - W_i(t) \quad (3)$$

Where $\eta(t)$ is the learning rate at time step t and $h_{i, BMU}(t)$ is the neighborhood function that determines how much the neighboring nodes of the BMU are adjusted. The neighborhood function, typically a Gaussian function, ensures that the BMU and its surrounding nodes are updated, but nodes farther away are updated less (Eq. (4)) as

$$h_{i, BMU}(t) = \exp\left(-\frac{\|r_i - r_{i, BMU}\|^2}{2\sigma^2(t)}\right) \quad (4)$$

$r_i - r_{i, BMU}$ is the distance between node i and the BMU in the grid and σ is the neighborhood radius, which decreases over time. The choice of SOM architecture, including the number of neurons and grid topology, was guided by a series of preliminary experiments to balance

resolution and generalization. Specifically, the total number of neurons was selected based on the commonly used heuristic of the square root of the number of samples, ensuring sufficient granularity while avoiding overfitting. A rectangular grid topology was adopted to facilitate clearer visualization and interpretation of spatial patterns among the hydrochemical parameters.

3.4. Multivariate statistical method

PCA is used for the exploration of relationships between physiochemical parameters and identifying key contributing factors to the variation in groundwater chemistry in the study area. The analysis was based on the Pearson correlation matrix (Mohammed et al., 2022b), which evaluates the linear relationships between the physiochemical parameters of groundwater. This approach ensures that variables with strong correlations are grouped together, allowing for the reduction of data dimensionality while retaining most of the variance. PCA reduces the dimensionality of the dataset by transforming interrelated variables into a smaller set of uncorrelated components, called principal components (PCs). These PCs are derived as linear combinations of the original variables through an orthogonal transformation process, making them effective tools for uncovering hidden patterns and relationships within the data. The selection of principal components (PCs) was guided by the Kaiser criterion (Kaiser, 1991), where only components with eigenvalues greater than 1 were retained for interpretation, as they represent a variance larger than that of a single original variable. The first few PCs explained a substantial portion of the total variance, capturing the dominant geochemical processes. Loading coefficients were analyzed to determine the contribution of individual variables to each component. For this analysis, a data matrix (X) (Eq. (5)) is constructed where each row corresponds to a groundwater sample and each column represents a specific physicochemical parameter. PCA is then performed on the matrix to reorganize and decompose the data using matrix decomposition (Eq. (6)).

$$X = \begin{bmatrix} pH_1 & TDS_1 & \dots & NO_3 \ 1 \\ pH_2 & TDS_2 & \dots & NO_3 \ 2 \\ \vdots & \vdots & \ddots & \vdots \\ pH_i & TDS_i & \dots & NO_3 \ i \\ \vdots & \vdots & \ddots & \vdots \\ pH_n & TDS_n & \dots & NO_3 \ n \end{bmatrix} \quad (5)$$

$$X = FW^T + E, \quad (6)$$

where F stands for the factor score matrix, W for the factor loading matrix, E for the matrix of residuals, and T is the matrix transpose operator.

3.4. Health risk assessment

Health risk assessment (HRA) in this study evaluates the potential non-carcinogenic health risks associated with exposure to NO₂, NO₃, NH₃, and F based on U.S. Environmental Protection Agency (USEPA) protocol. The assessment includes calculating health quotients (HQs) for each metal through oral and dermal exposure pathways for both adults and children, followed by calculating the health index (HI) to provide an aggregated measure of the risk (USEPA, 2017). HQ is a measure of the risk of health effects due to exposure to each metal. It is calculated for each exposure route and compared to a reference dose (RfD) for the metal in question. The RfD for NO₂ is 0.1, NO₃ 1.6, NH₃ 0.1 and F 0.05 (USEPA, 2019). The oral dose for each metal (D) is calculated using Eq. (7) and subsequently, the HQ is calculated for each age group (Eq. (8)).

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

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

Where C is the concentration of the metal in groundwater (mg/L), IR: ingestion rate of groundwater (L/day), EF: exposure frequency (days/year), ED: exposure duration (years), BW: body weight of the exposed individual (kg), AT: averaging time for non-carcinogenic effects (days), calculated as $AT = ED \times 365$ (Table 1). If $HQ > 1$, the metal concentration in groundwater poses a potential health risk through that exposure route; if $HQ < 1$, the exposure is considered within acceptable limits.

HI represents the cumulative non-carcinogenic risk from multiple parameters via oral exposure. It is the sum of all HQs calculated for each parameter. The HI is computed separately for each population group (adults and children) using Eq. (9). In this study, HI is calculated as the cumulative effect of HQ values for NO_2 , NO_3 , NH_3 , and F.

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

The HI values indicate the combined health risk across parameters. An HI greater than 1 suggests that cumulative exposure may lead to health effects, while an HI < 1 indicates that the cumulative exposure is within acceptable health risk levels.

3.5. Monte carlo and sobol sensitivity analysis

Monte Carlo simulation is a probabilistic method used to assess uncertainty and variability in complex systems based on repeated random sampling. In this study, It is used to assess the uncertainty associated with the health risks due to nitrogen and fluoride consumption (USEPA, 1991). The Monte Carlo simulation begins by defining a mathematical model that expresses health risk calculation. Key input parameters for the model, including concentrations of contaminants, exposure durations, and ingestion rates, are characterized by probability distributions rather than fixed values. The simulation operates by randomly sampling values from these input distributions and using them to compute the outcome which is in this case HQ and HI. This process is repeated thousands of times, generating a large dataset of possible outcomes. The results are then statistically analyzed to provide an understanding of the system's behavior. Monte Carlo simulation allows for the evaluation of potential risks under different scenarios and supports more informed decision-making.

Sobol sensitivity analysis is a variance-based method that decomposes the variance of the model output into contributions from individual input parameters and their interactions (Sobol, 1993). This analysis quantifies the relative influence of each parameter on the final output and identifies which variables most significantly impact the health risk index. Sobol analysis provides two primary metrics including first-order indices (S1) and total-order indices (ST). S1 represents the proportion of output variance attributed to each parameter in isolation, showing how much each parameter alone influences the health risk index. ST measures the total variance in the output due to a given parameter, including both its effect and its interactions with other parameters. High total-order indices suggest that a parameter, either alone or through interaction, plays a critical role in the model outcome. In this study, Sobol sensitivity analysis is applied to assess the influence of

Table 1

Parameters used for calculation of hazard quotients and indices for adults and children based on expert judgment and U.S. Environmental Protection Agency (USEPA) guidelines.

Item	Unit	Children	Adults
Exposure frequency (EF)	Days	365	365
Ingestion rate (IR)	L/day	1	2.5
Body weight (BW)	Kg	15	70
Exposure duration (ED)	Years	12	60
Average exposure time (AT)	Days	5475	25,550

exposure variables and metal concentrations on the HI.

3.6. Geospatial interpolation

The chemical parameters and HI of groundwater in the Shendi area were spatially analyzed using the Topo to Raster (TtR) interpolation method (Hutchinson and Dowling, 1991). It is a hydrologically robust tool designed for generating continuous surfaces from discrete data points. This technique is particularly suited for hydrological studies, ensuring the creation of smooth and hydrologically consistent interpolated surfaces that accurately represent the spatial distribution of groundwater quality parameters. This method employs an iterative finite difference interpolation technique, which prioritizes the preservation of input data values while minimizing interpolation artifacts, making it highly effective for handling unevenly distributed sampling points such as groundwater wells. The process began with the preparation of input data, where the sample locations were georeferenced, and the measured concentrations of each parameter were attributed to their respective points in a GIS environment. To ensure accurate interpolation, the method was configured to account for the distribution and density of the sampling points, maintaining a balance between local variability and computational efficiency. The raster resolution was set to 30 m, providing sufficient spatial detail while optimizing processing time.

4. Results and discussion

4.1. Overview of the chemical parameters

The statistical summary of the analyzed parameters is illustrated in Table 2. The mean pH value of 7.8 suggests that the groundwater is slightly alkaline, with a narrow standard deviation (0.4597) reflecting relatively stable pH levels across the study area. However, the minimum value of 6.74 indicates some localized acidic conditions. TDS shows a high variability, with values ranging from 219.3 mg/L to a maximum of 2118 mg/L and a mean of 648.8 mg/L. These elevated TDS levels indicate high mineral dissolution rates. Based on the TDS, the majority of the samples (87 %) can be classified as freshwater since they exhibited TDS lower than 1000 mg/L (WHO, 2011). The remaining 13 % of the samples can be classified as brackish water (Carroll, 1962). Calcium (Ca), as an essential component of water hardness exhibited variability with a mean value of 43.9 mg/L and a maximum value of 214 mg/L with only one sample exceeded the WHO (2011) standard. Similarly, Cl levels are highly variable, with concentrations ranging from 4.26 mg/L to an alarming 745.5 mg/L with 3 samples having concentration higher than the permissible limit. NO_2 levels are relatively low, with a mean of 0.25 mg/L, while NO_3 concentrations are considerably higher, with a mean of 7.8 mg/L and a maximum value of 42.2 mg/L, indicating areas of nitrate pollution. NH_3 levels are low on average (0.09 mg/L) while F concentrations range from 0.02 mg/L to 2.06 mg/L, with a mean of 0.57 mg/L, highlighting the possibility of F contamination in some areas.

The spatial distribution maps indicated the variation of these parameters across the study area (Fig. 3). The pH distribution (Fig. 3a) indicates that the groundwater is generally slightly alkaline, particularly in areas along the streams, reflecting the influence of recharge water with naturally alkaline characteristics. A few isolated zones with lower pH values are also observed, suggesting localized deviations in hydrochemical conditions. The TDS distribution (Fig. 3b) highlights hotspots of high concentrations, particularly in regions dominated by igneous rocks in the southwestern part of the study area. These elevated TDS values are likely due to enhanced mineral dissolution. In contrast, the lowest TDS values are observed along the streams, indicating the dilution effect of recharge water infiltrating into the system.

The spatial map for NO_2 (Fig. 3c) reveals generally low concentrations across most areas, with localized spikes observed in specific regions. The highest NO_2 values are found in the southwestern part of the

Table 2

The statistical summary of the analyzed physiochemical parameters in the study area.

Parameter	Unit	Mean	Standard Deviation	Minimum	Maximum	WHO (2011) guidelines
pH	pH unit	7.8	0.45	6.74	8.9	8.5
TDS	mg/L	648.8	380.9	219.3	2118	1000
Ca	mg/L	43.9	37.1	4.8	214	200
Cl	mg/L	70.2	116.4	4.26	745.5	250
NO ₂	mg/L	0.25	0.45	0.003	1.3	3
NO ₃	mg/L	7.8	8.2	0.04	42.2	45
NH ₃	mg/L	0.09	0.13	0.007	0.66	1.5
F	mg/L	0.57	0.45	0.02	2.06	1.5

study area, where igneous rocks dominate. Additionally, smaller localized high values are noted along the Nile River, coinciding with small agricultural fields. NO₃ concentrations (Fig. 3d) are significantly elevated in the eastern part of the study area, aligning with the primary agricultural fields. The NH₃ distribution (Fig. 3e) shows relatively uniform but low concentrations throughout the study area, with the highest values observed in the northern regions, which coincide with areas of elevated pH. This suggests a possible link between ammonia stability and alkaline conditions in these regions. Fluoride (F) distribution (Fig. 3f) displays pockets of high concentrations scattered across different locations in the southern, eastern, and northern parts of the study area. This widespread distribution may indicate multiple sources, including both geogenic processes and localized anthropogenic inputs.

4.3. Clustering of the groundwater samples

The SOM analysis for the physicochemical parameters is used to reveal clustering patterns and relationships among the parameters. A 3 by 3 neurons architecture is used considering the number of samples used in this study. The SOM neighbor weight distances (Fig. 4a) visualize the distances between neighboring neurons, where the color gradient represents variations in groundwater quality parameters. Red and dark-colored areas indicate larger distances, suggesting significant hydrochemical differences between neurons, while yellow and light-colored regions denote smaller distances, implying greater similarity in hydrochemical properties. The observed pattern suggests that certain regions within the SOM represent distinct clusters of groundwater samples, with high-distance zones indicating strong hydrochemical contrasts and potential transitions between different water types (Qu et al., 2022).

The hits of the SOM neurons (Fig. 4b) represent the number of groundwater samples assigned to each neuron. The numbers within each hexagon indicate the frequency of samples belonging to that neuron, with high-density neurons (e.g., 17, 11, 10 samples) reflecting dominant hydrochemical characteristics shared by multiple groundwater samples. In contrast, neurons with fewer hits (e.g., 1, 3, 4 samples) represent less common water compositions, possibly linked to outliers, transitional samples, or unique geological and anthropogenic influences. Neurons with a high number of hits are generally located in low-distance zones (yellow/light regions in Fig. 4a), indicating that most groundwater samples share similar hydrochemical properties, forming well-defined clusters. Conversely, neurons located in high-distance zones (red/black regions) correspond to low-hit neurons in Fig. 4b, suggesting that these groundwater samples exhibit distinct chemical characteristics. This clustering pattern indicates that groundwater in the study area follows three major hydrochemical trends, with a few isolated cases exhibiting distinct compositions (Nourani et al., 2016).

The SOM weight planes in Fig. 5 illustrate the clustering patterns of key hydrochemical parameters, revealing their spatial relationships and distribution of groundwater samples (hits) across different neurons. Each hexagonal grid represents a distinct cluster, with yellow and lighter colors indicating higher concentrations, while darker regions signify lower values. The pH weight plane exhibits higher values in the upper and central regions of the SOM grid, confirming that most groundwater samples are slightly alkaline. The hit map shows that many groundwater

samples (e.g., 17 hits) are clustered in areas with moderate pH values, reinforcing that alkaline conditions are dominant in the dataset. The TDS weight plane highlights higher concentrations in the left regions, with lighter zones clustering in areas that overlap with the low pH zones. The lowest TDS values align with moderate pH regions along the upper-right portion of the grid, supporting the notion that recharge alkaline water dilutes dissolved constituents in these areas (Mohammed et al., 2023a). The hit map shows 11 samples influenced by this process.

The NO₂ weight plane primarily features darker colors across most of the SOM grid, indicating low NO₂ concentrations in the majority of groundwater samples. However, a localized high concentration zone appears in the lower-left region. The hit map shows that clusters with elevated NO₂ concentrations contain a moderate number of samples (e.g., 11 hits), suggesting that NO₂ contamination is not widespread but occurs in specific localized zones. In contrast, the NO₃ weight plane shows high concentrations in the lower-central and left regions. The similarity between the NO₃ and TDS distributions suggests a common contamination source. The hit map indicates that NO₃ contamination is concentrated in areas with relatively fewer samples (e.g., 3 hits in certain clusters), suggesting that nitrate contamination is more localized but significant where present. The NH₃ weight plane reveals sporadic high concentrations, particularly in the lower-right region of the SOM grid, which aligns with moderate pH zones. This correlation suggests that NH₃ is more stable in alkaline conditions, as NH₄ shifts toward NH₃ at higher pH levels (Guštin and Marinšek-Logar, 2011). The hit map for NH₃ shows a relatively low number of samples (e.g., 7 hits in certain clusters), suggesting that elevated NH₃ concentrations are limited to specific regions. The F weight plane displays high concentrations in multiple regions, particularly in the upper-left, central, and lower-right zones of the SOM grid. The overlap between high F and TDS regions in some areas indicates that fluoride concentrations may be influenced by water-rock interactions, particularly in areas where fluorine-bearing minerals are present in the aquifer matrix (Xiao et al., 2022). The hit map for fluoride shows that many samples (e.g., 4 and 10 hits) are associated with moderate to high fluoride levels.

4.4. Factors influencing contamination

The factors influencing groundwater contamination in the study area are studied using PCA. Initially, Pearson correlation analysis is conducted to reveal the interrelationships among the physiochemical parameters (Fig. 6). The strongest correlations were observed between TDS and Cl ($r = 0.63$) and TDS and F ($r = 0.57$), suggesting that these ions contribute significantly to the total dissolved solids and may originate from similar geochemical sources. A moderate correlation between TDS and Ca ($r = 0.40$) further supports the role of dissolution of Ca rich minerals in shaping water chemistry. In contrast, pH exhibited weak correlations with most parameters, slightly positive with F ($r = 0.25$) and TDS ($r = 0.17$), and slightly negative with Ca ($r = -0.16$) and NO₂ ($r = -0.16$). These weak associations imply that pH is largely regulated by buffering reactions and is less influenced by ion concentrations directly. The nitrogen species (NH₃, NO₂, NO₃) displayed weak correlations with both each other and other major ions, reflecting their distinct and variable sources. This suggests that nitrogen dynamics operate

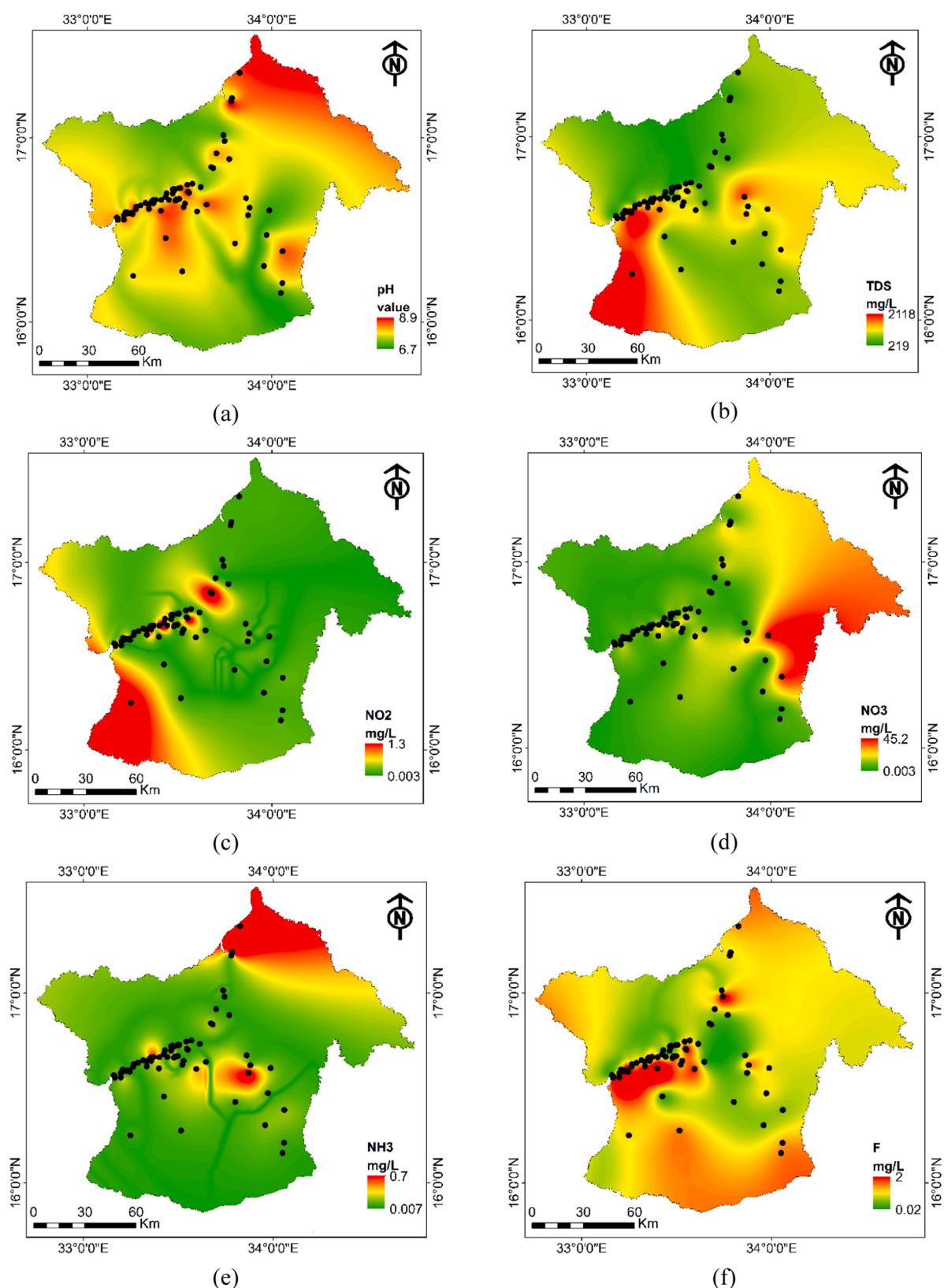


Fig. 3. The spatial distribution of the (a) pH, (b) TDS, (c) NO₂, (d) NO₃, (e) NH₃, and (f) F in the study area.

independently of the dominant geochemical controls influencing other solutes in the karst system.

The PCA highlighted the main factors influencing water quality, with three PCs explaining 62.04 % of the total variance (Table 3). PC1, which

explains 28.5 % of the total variance, is predominantly influenced by strong positive loadings of TDS (0.58), Cl⁻ (0.53), and Ca²⁺ (0.44). This suggests that natural mineralization processes, primarily driven by water-rock interactions, play a significant role in shaping groundwater

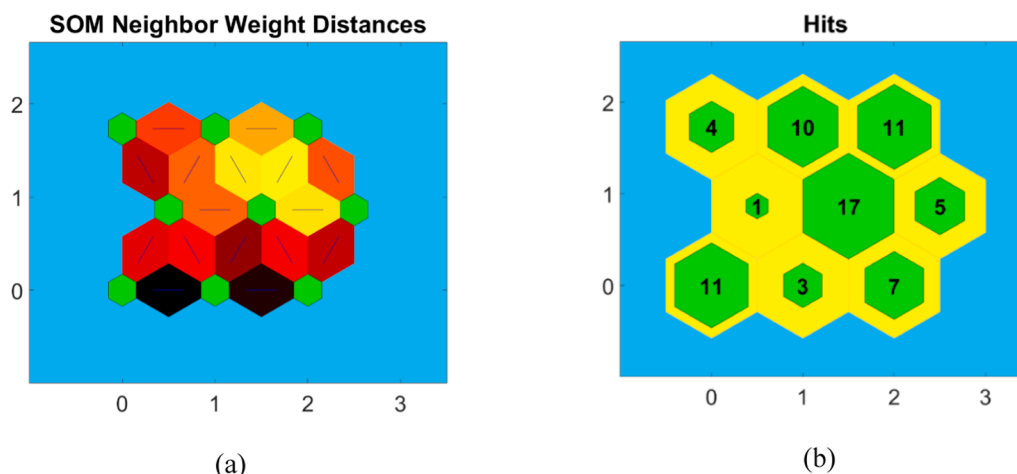


Fig. 4. Features of the SOM showing the (a) neighbor weight distances and (b) hits of the SOM showing the clustering and distribution of the samples in neurons.

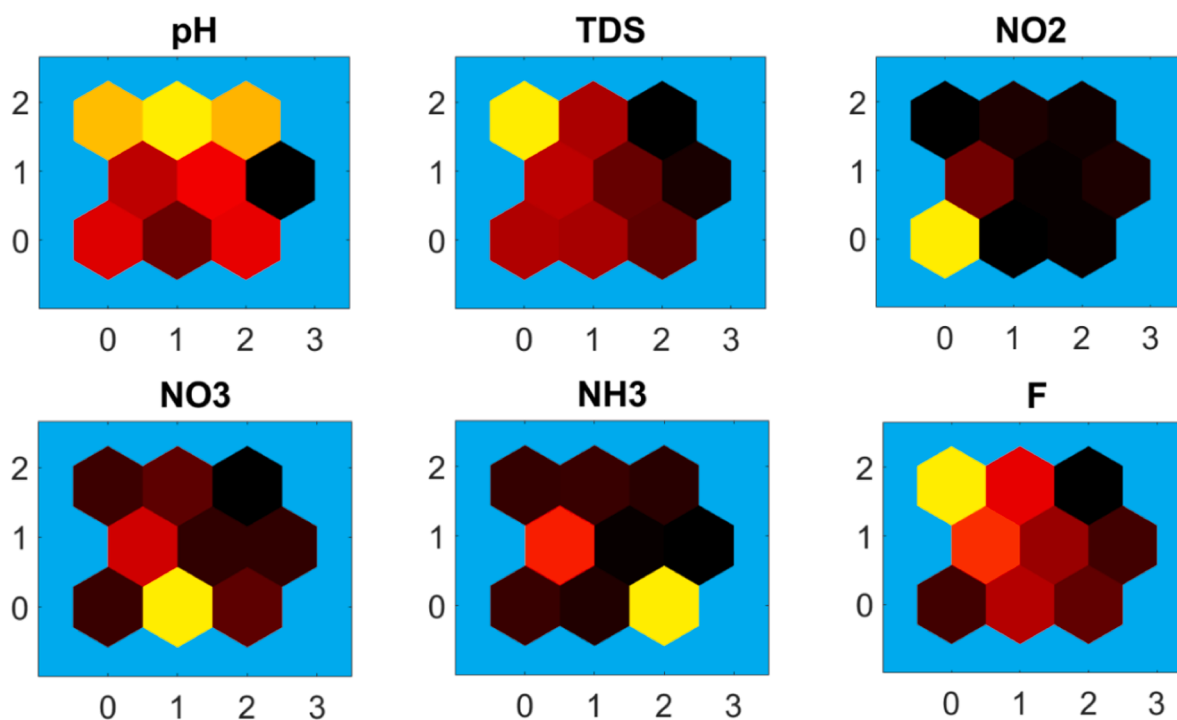


Fig. 5. The weight planes of the SOM considering all the physiochemical parameters showing the effect of each parameter on the clustering of the groundwater samples.

chemistry. The high TDS and Cl^- concentrations indicate evaporative enrichment or dissolution of halite minerals, while elevated Ca^{2+} levels point to carbonate weathering and gypsum dissolution. These findings highlight the dominance of geogenic factors in groundwater composition, with rock-water interactions being the primary source of dissolved ions. PC2, explaining 18.8 % of the variance, is marked by a strong negative correlation with pH (-0.57) and a positive correlation with NO_2 (0.40), reflecting anthropogenic pollution, likely from agricultural activities. The negative correlation with pH is likely attributed to acidification caused by the application of ammonium-based fertilizers, which release protons during nitrification ($\text{NH}_4 + 2\text{O}_2 \rightarrow \text{NO}_3 + 2\text{H} + \text{H}_2\text{O}$) (Johnson et al., 2010). Nitrite levels are elevated as an intermediate in denitrification ($\text{NO}_3 + 2\text{H} + 2\text{e} \rightarrow \text{NO}_2 + \text{H}_2\text{O}$). These processes highlight the significant impact of agricultural runoff on groundwater quality. PC3, which explains 14.8 % of the total variance, is dominated by high loadings of NH_3 (0.69) and NO_3 (0.54), suggesting additional

sources of pollution from urban wastewater and intensive agricultural practices. NH_3 levels are likely elevated due to the decomposition of organic matter in wastewater ($\text{R-NH}_2 + \text{H}_2\text{O} \rightarrow \text{NH}_4 + \text{R-OH}$) (Mohammed et al., 2022b). NO_3 can also result from the complete oxidation of ammonium through nitrification, with intermediate nitrite formation ($\text{NH}_4 \rightarrow \text{NO}_2 \rightarrow \text{NO}_3$).

The biplot of the PCA indicated showed the distribution of groundwater samples (scores) based on their chemical composition (Fig. 7). The first cluster, represented by samples 52 and 59 on the right side of the biplot, shows high TDS and Cl content with strong Ca association, indicating highly mineralized groundwater possibly influenced by carbonate and evaporative dissolution (Bouteraa et al., 2019). The second cluster in the upper quadrant, including samples as 13, 43, and 54, correlate strongly with NO_2 and NH_3 , suggesting significant agricultural pollution impact and vulnerability to surface contamination. The third and largest cluster, located in the central region within the 95 %

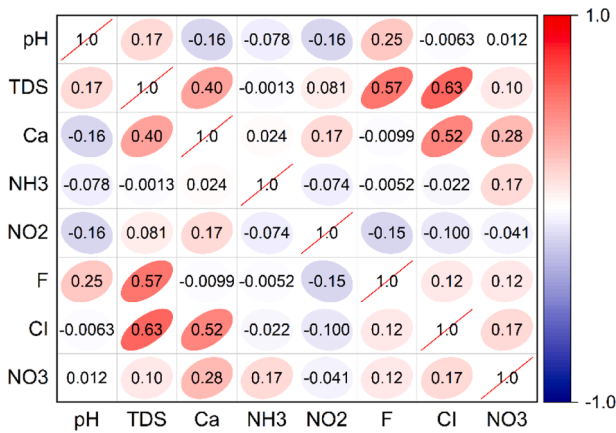


Fig. 6. Pearson correlation matrix showing the correlation coefficients between different parameters.

Table 3

The principal components coefficients of the parameters in each eigenvector.

Parameter	PC1 (28.47 %)	PC2 (18.77 %)	PC3 (14.80 %)
pH	0.08878	-0.56782	-0.04273
TDS	0.58253	-0.13516	-0.18895
Ca	0.43559	0.44536	-0.02155
NH ₃	0.0252	0.09587	0.69407
NO ₂	-0.00864	0.40404	-0.4333
F	0.34776	-0.49898	-0.00352
Cl	0.52836	0.14198	-0.05844
NO ₃	0.24944	0.13925	0.53768

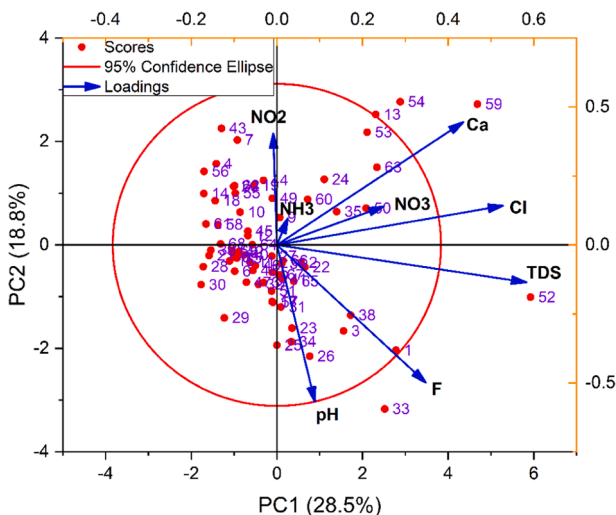


Fig. 7. Biplot showing the distribution of groundwater samples (scores) across the extracted loadings.

confidence ellipse, represents background groundwater chemistry with moderate mineralization. The spatial relationships between parameters confirm the origin of the chemical parameters. The close association of TDS, Cl, and Ca vectors confirms a common source through mineral dissolution, while the clustering of NO₂ and NH₃ vectors confirms agricultural influence (Riedel et al., 2022).

4.5. Health quotients and indices

The health risk of groundwater contaminants, NO₂, NO₃, NH₃, and F, is evaluated using HQs and HIs for adults and children. The statistical

summary of the calculated HQs (Fig. 8) is indicated using box plots for adults (Fig. 8a) and children (Fig. 8b). Nitrate emerges as a significant health concern, with median HQ values of approximately 0.2 for adults and 0.3 for children. The presence of multiple outliers exceeding 0.5 further underscores the moderate non-carcinogenic health risks, particularly in children. In specific cases, outliers for children exceed an HQ of 1, signaling potential health hazards in localized regions. While NO₃ concentrations in the groundwater are within safe limits as per WHO (2011) thresholds, the prolonged exposure to NO₃, even if below the maximum allowable concentrations, can still pose potential health risks (USEPA, 2006). Fluoride demonstrated the highest variability and magnitude of HQ values among the contaminants. The ranges span 0–0.9 for adults and 0–1.6 for children, with larger interquartile ranges. Outliers for children approach HQ values of 1.7, pointing to severe health risks in specific locations (USEPA, 2006). NO₂ and NH₃ show relatively lower HQ values compared to NO₃ and F, with most values below 0.2 for both adults and children.

The results of the ANOVA test confirmed this and showed that there are significant differences in the HQ for both adults and children across the chemicals analyzed (NO₂, NO₃, NH₃, and F). The F-statistic for both groups is 44.98, which indicates that the variance between the group means is much larger than the variance within the groups, suggesting that the chemical exposure significantly impacts the hazard quotients for both adults and children. The p-value for both groups is extremely small (1.23×10^{-23}), well below the threshold of 0.05, which provides strong evidence to reject the null hypothesis, indicating that the hazard quotients for the different chemicals are not equal. This means that the levels of exposure to these chemicals vary significantly and affect health risks differently for both adults and children.

The spatial distribution of HI values for the health risk assessment due to NO₂, NO₃, NH₃, and F is illustrated in Fig. 9. The HI values ranging from 0.003 to 1.53 for adults (Fig. 9a) and 0.004 to 2.07 for children (Fig. 9b). According to (USEPA, 2006), this demonstrate that children face greater overall health risks from groundwater contamination. The northern and northeastern regions stand out as an area of significant concern, with consistently elevated HI values. In contrast, the southern and western parts of the study area exhibit lower HI values, indicating that the groundwater in this region poses a relatively lower risk to human health. Two additional hotspots of concern are observed in the eastern region, where HI values peak for both adults and children. These localized areas show the highest cumulative health risks, with children's HI values exceeding 2.0. The spatial distribution of HI closely mirrors that of F and NO₃, highlighting these contaminants as having the most

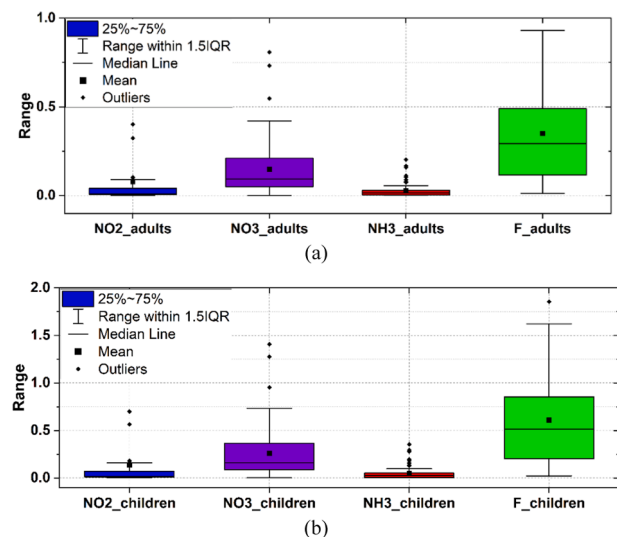


Fig. 8. Box plot showing the statistical summary of the calculated HQs for each chemical parameter for (a) adults and (b) children.

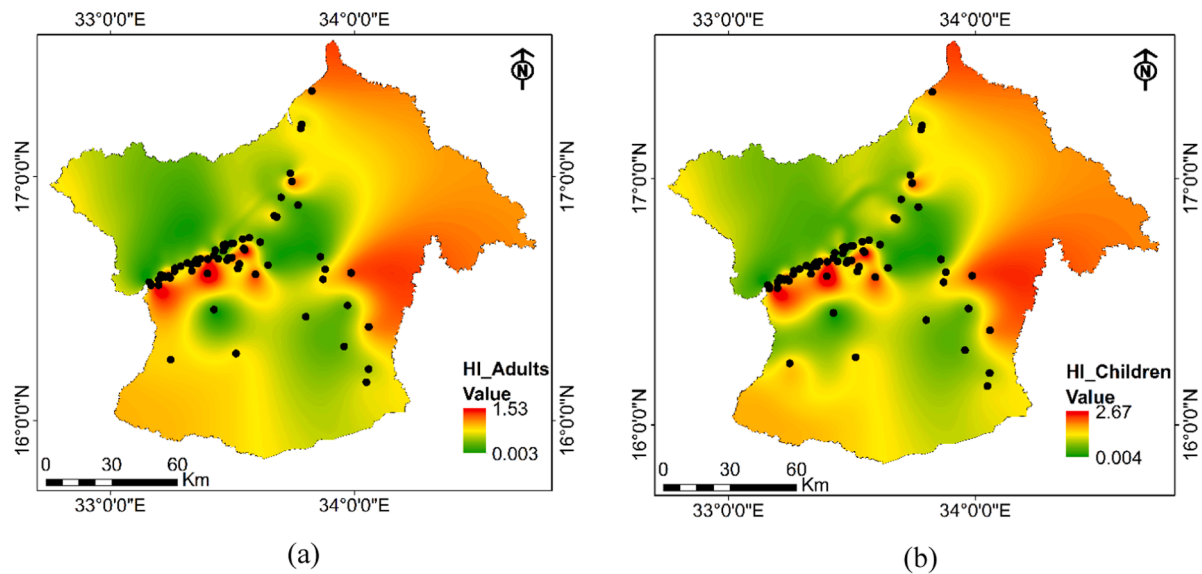


Fig. 9. The spatial distribution of the calculated HIs in the study area for (a) adults and (b) children.

significant influence on health risks across the study area (Duvva et al., 2022). This correlation suggests that agricultural activities, which contribute to elevated nitrate levels using nitrogen-based fertilizers (Baghapour et al., 2016), and rock-water interactions (Enalou et al., 2018), which often lead to high fluoride concentrations, are the primary factors driving health risks.

4.6. Modeling of the health risks

The Monte Carlo simulation modeled the probabilistic health risks associated with fluoride and nitrogen compounds for adults and children. The test is run for 10,000 iterations to examine each possible combination of the health risk parameters and thus verify the results of

health risk calculation. The results reveal distinct probability distributions of HQ for each contaminant (Fig. 10) and their cumulative Hazard Index (HI) (Fig. 11) for both adults and children. These distributions provide critical insights into the uncertainty and variability of potential health risks. For all four contaminants (NO_2 , NO_3 , NH_3 , and F), the probability distributions exhibit clear separation between adult and child populations, with consistently higher exposure risks for children. Nitrate (NO_3) demonstrated the highest probable mean HQ values, with 0.15 for adults (with approximately 95 % of values falling between 0.10–0.20) and 0.30 for children (with approximately 95 % of values between 0.22–0.38). The minimal overlap between these distributions indicates significantly different exposure risks between the two populations. Fluoride (F) presents the second highest exposure concern,

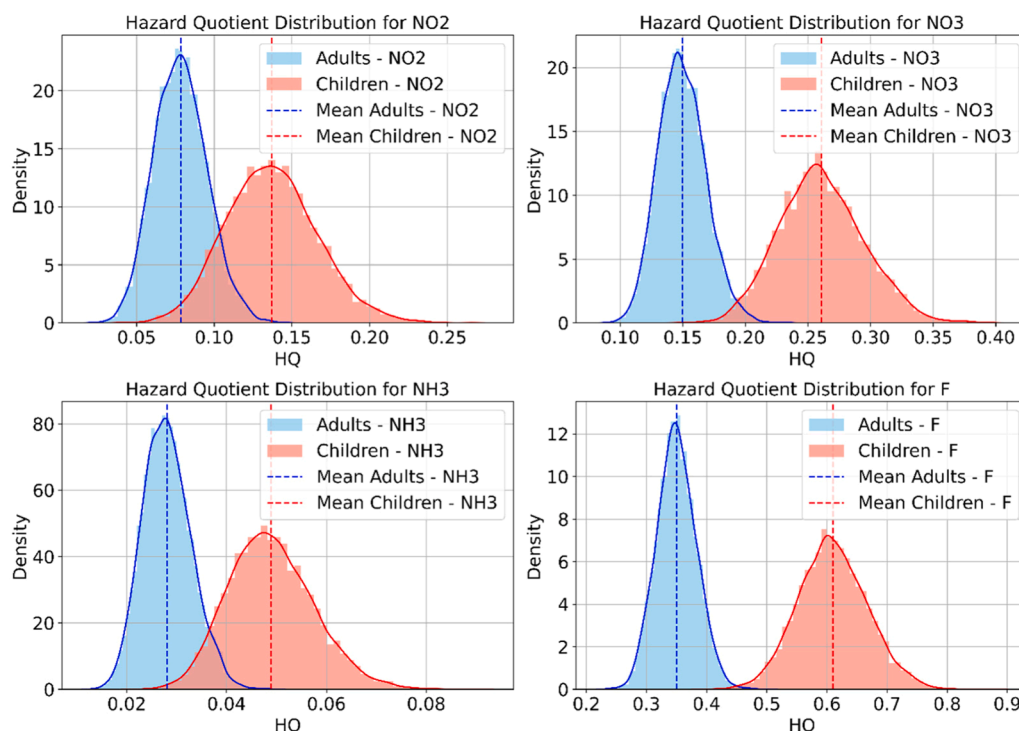


Fig. 10. The probability distribution based on Monte Carlo simulation of the HQs for each parameter for adults and children.

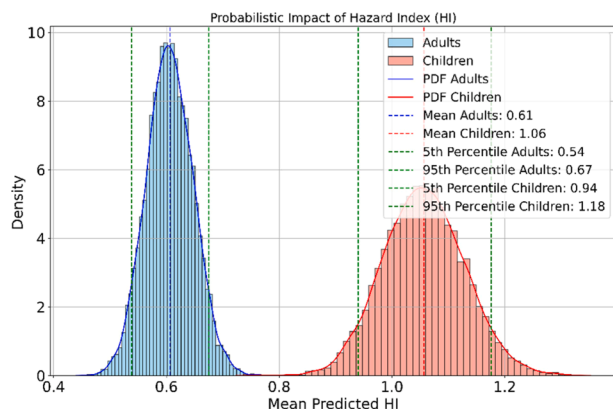


Fig. 11. The probability distribution based on Monte Carlo simulation of the HIs for adults and children.

with mean HQ values of 0.30 for adults and 0.60 for children. The distributions show well-defined peaks with adult values concentrated around 0.30 and children's values around 0.60, with distinct separation between the two populations. Nitrogen dioxide (NO_2) exhibits mean HQ values of approximately 0.08 for adults and 0.14 for children, with moderate distribution overlap but still maintaining distinct probability peaks. Ammonia (NH_3) shows the lowest individual HQ values among the tested contaminants, with mean values of approximately 0.03 for adults and 0.05 for children yet maintains the pattern of higher children's exposure risk.

The probabilistic assessment of the cumulative HI provides a comprehensive view of the total non-carcinogenic health risks posed by all contaminants. The mean HI for adults is 0.61 (5th percentile: 0.54; 95th percentile: 0.67), indicating that the overall risk remains below the threshold of concern ($\text{HI} < 1$) even at the upper confidence bound. In contrast, children exhibit a mean HI of 1.06 (5th percentile: 0.94; 95th percentile: 1.18), suggesting that cumulative exposure exceeds the threshold of concern for a substantial portion of the child population. The probability distributions of HI show minimal overlap between adult and child populations, with clearly separated peaks and confidence intervals. The adult HI distribution is narrower, suggesting less variability in exposure outcomes compared to children. The 5th and 95th percentile values serve as confidence bounds, indicating that even at the lower bound (5th percentile), children's HI (0.94) approaches the threshold of concern, while at the upper bound (95th percentile), it significantly exceeds this threshold (1.18). The Monte Carlo simulation results strongly corroborate the deterministic health risk calculations presented earlier, providing statistical validation of the initial risk estimates. The consistency between these approaches reinforces the reliability of the findings in quantifying the potential non-carcinogenic health risks, particularly for children (Jiang et al., 2021). These probability distributions further elucidate the degree of uncertainty in risk estimation and highlight the importance of considering population-specific exposure factors in environmental health assessments.

The Sobol sensitivity analysis provides critical insights into how different parameters influence health risk outcomes, revealing complex interactions within the risk assessment model (Fig. 12). Average exposure time (AT) emerges as the dominant influential parameter with a total sensitivity index of 0.76, indicating its extensive contribution to risk variance both independently and through interactions with other factors. The substantial difference between AT's first-order index (0.18) and total order index (0.76) demonstrates that nearly 76 % of risk variability can be attributed to AT when considering its interactions with all other parameters, while only 18 % is due to AT's direct effects alone. This suggests AT serves as a critical amplifier of health risks when interacting with other exposure parameters (Pang et al., 2024). Body weight (BW) represents the second most influential factor with a total

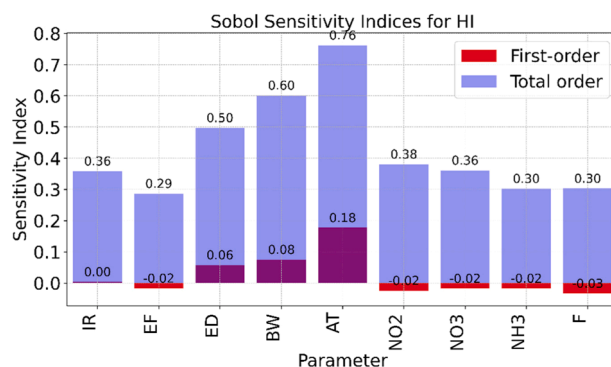


Fig. 12. The features importance of the HRA parameters based on the Sobol sensitivity analysis.

sensitivity index of 0.60, yet its first-order index is only 0.08. This pronounced disparity indicates that BW primarily influences health risks through interactions with other parameters rather than direct effects. This finding aligns with risk assessment fundamentals, where BW appears in the denominator of dose calculations, creating an inverse relationship that particularly affects children due to their lower body weights (Ramesh et al., 2021), thereby amplifying their vulnerability to contaminant exposure.

Exposure duration (ED) similarly shows a significant interaction component, with a total sensitivity index of 0.50 compared to its first-order index of 0.06, highlighting how chronic exposure interacts synergistically with other parameters to enhance health risks over time. Regarding contaminant contributions, the analysis reveals that while NO_2 and NO_3 have higher total sensitivity indices (0.38 and 0.36, respectively) than F and NH_3 (both 0.30), their first-order indices are minimal or slightly negative. This counterintuitive pattern indicates these contaminants primarily contribute to health risks through their interactions with exposure parameters rather than through independent effects. The consistent and substantial disparities between total and first-order indices across nearly all parameters underscores that health risks emerge predominantly from complex parameter interactions rather than from individual parameter influences.

The consistently higher HQ and HI values for children underscore their increased susceptibility to groundwater contaminants. This heightened vulnerability is driven by physiological factors such as lower body weight and developing organ systems. The finding that children's HI values exceed 1.0 suggests the potential for significant health effects, particularly from NO_3 and F exposure. Ingested NO_3 is converted into NO_2 in the body, interfering with the oxygen-carrying capacity of hemoglobin in the blood, leading to potentially life-threatening oxygen deprivation (Fan and Steinberg, 1996; Ward et al., 2018). Excessive F intake during early childhood can lead to dental fluorosis (Nizam et al., 2022), characterized by discoloration and pitting of the teeth, which may affect appearance and function (Dar and Kurella, 2024). Prolonged exposure to high F levels can result in skeletal fluorosis, a more severe condition involving pain, stiffness, and damage to bones and joints (Ayoob and Gupta, 2006).

Given the higher risks faced by children, interventions should prioritize reducing their exposure to contaminants. This may include providing alternative water sources, such as treated or bottled water, and implementing education programs to raise awareness of safe water practices. Regular monitoring of groundwater quality is essential, particularly in areas with known sources of contamination. This will help identify and mitigate risks before they reach critical levels. Based on the findings, targeted interventions in areas with elevated cumulative HI values are essential. These should include advanced water treatment methods to remove contaminants, as well as stricter regulation of pollution sources, such as agricultural runoff. Public awareness campaigns should also be implemented to educate communities on the risks

of consuming untreated groundwater, especially for households with children.

5. Conclusions

This study provided a comprehensive assessment of groundwater quality and associated health risks in the Shendi area, Sudan, focusing on the Nubian Aquifer System (NAS). The study used a combination of SOM, PCA, Monte Carlo simulations, and Sobol sensitivity analysis to identify the contamination sources and assess the health risk of nitrogen compounds (NO_2 , NO_3 , NH_3) and fluoride (F). The conclusions of the study can be summarized in the following

- The SOM and PCA analysis revealed distinct hydrochemical trends in groundwater, with elevated NO_3 concentrations linked to agricultural runoff, while NO_2 and NH_3 were associated with both agricultural and wastewater pollution. High fluoride concentrations were found to be primarily geogenic, resulting from water-rock interactions involving fluorine-bearing minerals.
- Monte Carlo simulations highlighted the probabilistic nature of health risks, revealing significantly higher hazard quotients and hazard indices for children compared to adults. The mean HI for children was found to exceed the safe threshold, indicating a heightened vulnerability to non-carcinogenic health hazards. Sobol sensitivity analysis emphasized the importance of exposure time, body weight, and exposure duration, along with the interactions between these parameters, in shaping health outcomes. Among the contaminants, NO_3 and F were identified as the primary contributors to cumulative HI values.
- The findings of this study underscored the urgent need for targeted interventions in the Shendi area. Key recommendations include the implementation of advanced water treatment technologies, stricter regulation of agricultural runoff and wastewater discharge, and public health awareness programs to reduce exposure to harmful contaminants. Furthermore, the study highlights the importance of regular groundwater monitoring to detect contamination hotspots and mitigate risks before they reach critical levels.
- This research provides valuable insights into the health risks posed by groundwater contamination and offers a robust framework for assessing and managing these risks in arid and semi-arid regions. The findings can inform policy decisions and support sustainable groundwater management practices, both in the Shendi area and in similar regions facing groundwater quality challenges.

CRediT authorship contribution statement

Musaab A.A. Mohammed: Writing – original draft, Methodology, Investigation, Conceptualization. **Norbert P. Szabó:** Writing – review & editing, Validation, Supervision. **Elamin D. Suliman:** Resources, Investigation, Data curation. **Magboul M.S. Siddig:** Writing – review & editing, Validation. **Mohammed N.M. Hassan:** Writing – review & editing, Validation. **Péter Szűcs:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We also claim that this article is original, has not been published before, and is not considered for publication elsewhere.

Acknowledgment

This research is conducted in the framework of the Sustainable Development and Technologies, National Program of the Hungarian

Academy of Sciences (FFT NP FTA).

Data availability

The data that supports the findings of this study are available from the corresponding author upon reasonable request.

References

- Abdallrhman, A.T., Fagier, T.G., Abakar, N.H., Abdelrhman, E.M., Khairy, A.M., Ali, K.A., Khalifa, I., 2022. The Paleodepositional Environments in the Shandi Sub-Basin as part of the Atbara Basin, North Sudan. *Int. J. Earth Sci. Geol.* 4, 111–120.
- Appelo, C.A.J., Postma, D., 2005. *Geochemistry, groundwater and pollution*, 2nd. Ed. Rotterdam, Balkema.
- Ayoob, S., Gupta, A.K., 2006. Fluoride in drinking water: a review on the status and stress effects. *Crit. Rev. Env. Sci. Technol.* 36, 433–487.
- Babkir, M.D.B.H., Nton, M.E., Eisawi, A.A.M., 2023. Depositional environment and hydrocarbon exploration potential based on sedimentary facies and architectural analysis of the upper cretaceous Shendi Formation in Musawwarat-Naga area, Shendi-Atbara Basin, Sudan. *Earth Sci. Res. J.* 27, 109–128. <https://doi.org/10.15446/esrj.v27n2.98795>.
- Baghapour, M.A., Nobandegani, A.F., Talebheydokhti, N., Bagherzadeh, S., Nadiri, A.A., Gharekhan, M., Chitsazan, N., 2016. Optimization of DRASTIC method by artificial neural network, nitrate vulnerability index, and composite DRASTIC models to assess groundwater vulnerability for unconfined aquifer of Shiraz Plain, Iran. *J. Environ. Health Sci. Eng.* 14, 1–16. <https://doi.org/10.1186/s40201-016-0254-y>.
- Belal, A.A.A., Hussein, A.M., Ibrahim, A.D., 2024. Assessment of ground water quality for drinking purposes based on turbidity and water quality index in Shendi City, River Nile State, Sudan. *Int. J. Epidemiol. Public Health Res.* 5, 2810–2836.
- Bouteraa, O., Mebarki, A., Bouaicha, F., Nouaceur, Z., Laignel, B., 2019. Groundwater quality assessment using multivariate analysis, geostatistical modeling, and water quality index (WQI): a case of study in the Boumerzoug-El Khroub valley of Northeast Algeria. *Acta Geochim.* 38, 796–814. <https://doi.org/10.1007/s11631-019-00329-x>.
- Carroll, D., 1962. Rainwater as a chemical agent of geologic processes. A review-geological survey water-supply 1535, 1–16.
- Cauich-Kau, D., Castro-Larragoitia, J., Cardona-Benavides, A., Garc'ia-Arreola, M.E., Garc'ia-Vargas, G.G., 2024. An adapted Groundwater Quality Index including toxicological critical pollutants. *Groundw. Sustain. Dev.*, 101401.
- Dar, F.A., Kurella, S., 2024. Fluoride in drinking water: an in-depth analysis of its prevalence, health effects, advances in detection and treatment. *Mater. Today: Proc.* 102, 349–360.
- Dawoud, A.S., 1980. Structural and metamorphic evolution of the area southwest of Abu Hamed. University of Khartoum. University of Khartoum, Nile Province, Sudan. Ph. D Thesis.
- Dey, S., Singh, S., Raju, N.J., Mall, R.K., 2024. Hydrogeochemical characterization for groundwater quality and risk assessment in part of central gangetic alluvium, India. *Groundw. Sustain. Dev.* 25, 101108.
- Dharmani, A.B., Verma, M., Rani, S., Narang, A., Singh, M.R., Saya, L., Hooda, S., others, 2024. Unravelling groundwater contamination and health-related implications in semi-arid and cold regions of India. *J. Contam. Hydrol.* 261, 104303.
- Duggal, V., Sharma, S., 2022. Fluoride contamination in drinking water and associated health risk assessment in the Malwa Belt of Punjab, India. *Environ. Adv.* 8, 100242. <https://doi.org/10.1016/j.envadv.2022.100242>.
- Duvva, L.K., Panga, K.K., Dhakate, R., Himabindu, V., 2022. Health risk assessment of nitrate and fluoride toxicity in groundwater contamination in the semi-arid area of Medchal, South India. *Appl. Water Sci.* 12, 1–21. <https://doi.org/10.1007/s13201-021-01557-4>.
- Eltijani, A., Mohammed, M.A.A., Abuobida, Y., Yousif, I.M., 2024. Integrating CoDA and PCA for enhanced characterization of fluvial depositional processes: a case study of the Shendi formation, Sudan. *Discov. Geosci.* 2. <https://doi.org/10.1007/s44288-024-00011-7>.
- Enalou, H.B., Moore, F., Keshavarzi, B., Zarei, M., 2018. Source apportionment and health risk assessment of fluoride in water resources, south of Fars province, Iran: stable isotopes ($\delta^{18}\text{O}$ and δ_{F}) and geochemical modeling approaches. *Appl. Geochem.* 98, 197–205. <https://doi.org/10.1016/j.apgeochem.2018.09.019>.
- Famiglietti, J.S., 2014. The global groundwater crisis. *Nat. Clim. Chang.* 4, 945–948.
- Fan, A.M., Steinberg, V.E., 1996. Health implications of nitrate and nitrite in drinking water: an update on methemoglobinemia occurrence and reproductive and developmental toxicity. *Regul. Toxicol. Pharmacol.* 23, 35–43.
- Farah, E.A., Mustafa, E.M.A., Kumai, H., 2000. Sources of groundwater recharge at the confluence of the Nile, Sudan. *Environ. Geol.* 39, 667–672. <https://doi.org/10.1007/s002540050479>.
- Foster, S., Chilton, J., Nijsten, G.-J., Richts, A., 2013. Groundwater—a global focus on the 'local resource'. *Curr. Opin. Env. Sustain.* 5, 685–695.
- Griebler, C., Avramov, M., 2015. Groundwater ecosystem services: a review. *Freshw. Sci.* 34, 355–367.
- Gupta, S., Mishra, N., Kumar, A., Yadav, A.K., 2024. Fluoride pollution in subsurface water: challenges and opportunities. *Adv. Treat. Technol. fluoride Remov. Water: water Purif.* 19–39.
- Gustín, S., Marínsek-Logar, R., 2011. Effect of pH, temperature and air flow rate on the continuous ammonia stripping of the anaerobic digestion effluent. *Process Saf. Environ. Prot.* 89, 61–66.

- Hutchinson, M.F., Dowling, T.I., 1991. A continental hydrological assessment of a new grid-based digital elevation model of Australia. *Hydrol. Process* 5, 45–58.
- Islam, A.R.M.T., Abdullah-Al Mamun, M., Hasan, M., Aktar, M.N., Uddin, M.N., Siddique, M.A.B., Chowdhury, M.H., Islam, M.S., Bari, A.B.M.M., Idris, A.M., others, 2025. Optimizing coastal groundwater quality predictions: a novel data mining framework with cross-validation, bootstrapping, and entropy analysis. *J. Contam. Hydrol.* 269, 104480.
- Jannat, J.N., Khan, M.S.I., Islam, H.M.T., Islam, M.S., Khan, R., Siddique, M.A.B., Varol, M., Tokatli, C., Pal, S.C., Islam, A., others, 2022. Hydro-chemical assessment of fluoride and nitrate in groundwater from east and west coasts of Bangladesh and India. *J. Clean Prod.* 372, 133675.
- Jasechko, S., Perrone, D., 2021. Global groundwater wells at risk of running dry. *Science* 372, 418–421.
- Jiang, C., Zhao, Q., Zheng, L., Chen, X., Li, C., Ren, M., 2021. Distribution, source and health risk assessment based on the Monte Carlo method of heavy metals in shallow groundwater in an area affected by mining activities, China. *Ecotoxicol. Env. Saf.* 224, 112679.
- Johnson, C.N., Fisher, P.R., Huang, J., Vetanovetz, R.P., Argo, W.R., 2010. Quantifying the acidity of an ammonium-based fertilizer in containerized plant production microbial conversion of NH 4-N to NO 3-N (nitrification) also produces H +. *Hortscience* 45, 1099–1105.
- Kaiser, H.F., 1991. Coefficient alpha for a principal component and the Kaiser-Guttman rule. *Psychol. Rep.* 68, 855–858.
- Karadeniz, S., Ustaoglu, F., Aydin, H., Yüksel, B., 2024. Toxicological risk assessment using spring water quality indices in plateaus of Giresun Province/Türkiye: a holistic hydrogeochemical data analysis. *Env. Geochem. Health* 46, 285. <https://doi.org/10.1007/s10653-024-02054-8>.
- Karunanidhi, D., Subramani, T., Roy, P.D., Li, H., 2021. Impact of groundwater contamination on human health. *Env. Geochem. Health*.
- Kheiralla, M.K., 1966. Study of the Nubian Sand stone Formation of the Nile Vally between 14 N and 17 42 N, with reference to groundwater geology. University of Khartoum.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59–69.
- Kumar, V., Singh, E., Singh, S., Pandey, A., Bhargava, P.C., 2023. Micro-and nano-plastics (MNPs) as emerging pollutant in ground water: environmental impact, potential risks, limitations and way forward towards sustainable management. *Chem. Eng. J.* 459, 141568.
- Kumari, S., V. A., Sengupta, S., 2021. Adsorptive mitigation of fluoride ions using aluminosilicate adsorbents: a state-of-the-art review. *Environ. Chall.* 5, 100329. <https://doi.org/10.1016/j.envc.2021.100329>.
- Li, R., Yan, Y., Xu, J., Yang, C., Chen, S., Wang, Y., Zhang, Y., 2024. Evaluate the groundwater quality and human health risks for sustainable drinking and irrigation purposes in mountainous region of Chongqing, Southwest China. *J. Contam. Hydrol.* 264, 104344.
- Mia, M.Y., Haque, M.E., Islam, A.R.M.T., Jannat, J.N., Jion, M.M.M.F., Islam, M.S., Siddique, M.A.B., Idris, A.M., Senapathi, V., Talukdar, S., others, 2023. Analysis of self-organizing maps and explainable artificial intelligence to identify hydrochemical factors that drive drinking water quality in Haor region. *Sci. Total Environ.* 904, 166927.
- Ming, D., Wang, L., Dai, L., Zhang, Q., Liang, T., 2024. Evaluation of water environment quality in a typical wetland on the Qinghai-Tibet Plateau using positive matrix factorization and self-organizing map. *J. Clean Prod.* 479, 144069.
- Moeini, Z., Azhdarpoor, A., 2021. Health risk assessment of nitrate in drinking water in Shiraz using probabilistic and deterministic approaches and impact of water supply. *Environ. Chall.* 5, 100326. <https://doi.org/10.1016/j.envc.2021.100326>.
- Mohammed, M.A.A., Eltiqani, A., Szabó, N.P., Szűcs, P., 2023a. Hydro-chemometrics of the Nubian Aquifer in Sudan: an integration of groundwater quality index, multivariate statistics, and human health risk assessment. *Discov. Water* 3, 15.
- Mohammed, M.A.A., Khleel, N.A.A., Szabó, N.P., Szűcs, P., 2022a. Development of artificial intelligence model with aid of statistical methods for simulation of water quality indices in north Khartoum area, Sudan. *Reas. sq.* <https://doi.org/10.21203/rs.3.rs-1679487/v1>.
- Mohammed, M.A.A., Mohamed, A., Szabó, N.P., Szűcs, P., 2023b. Development of machine learning- based models for identifying the sources of nitrate and fluoride in groundwater and predicting their human health risks. *Int. J. Energy Water Resour.* <https://doi.org/10.1007/s42108-023-00271-y>.
- Mohammed, M.A.A., Szabó, N.P., Mikita, V., Szűcs, P., 2025. Tracking the spatiotemporal evolution of groundwater chemistry in the quaternary aquifer system of Debrecen area, Hungary: integration of classical and unsupervised learning methods. *Environ. Sci. Pollut. Res.* 1–20.
- Mohammed, M.A.A., Szabó, N.P., Szűcs, P., 2022b. Multivariate statistical and hydrochemical approaches for evaluation of groundwater quality in north Bahri city- Sudan. *Heliyon* 8. <https://doi.org/10.1016/j.heliyon.2022.e11308>.
- Nizam, S., Virk, H.S., Sen, I.S., 2022. High levels of fluoride in groundwater from Northern parts of Indo-gangetic plains reveals detrimental fluorosis health risks. *Environ. Adv.* 8, 100200. <https://doi.org/10.1016/j.envadv.2022.100200>.
- Nourani, V., Alami, M.T., Voutsoughi, F.D., 2016. Self-organizing map clustering technique for ANN-based spatiotemporal modeling of groundwater quality parameters. *J. Hydroinform.* 18, 288–309. <https://doi.org/10.2166/hydro.2015.143>.
- Nti, E.K., Kranjac-Berisavljevic, G., Doka, D.A., 2024. Assessing the impact of artisanal gold mining on the environmental sustainability of groundwater resource for water security in southwestern Ghana. *Environ. Chall.* 14, 100804. <https://doi.org/10.1016/j.envc.2023.100804>.
- Oki, T., Kanae, S., 2006. Global hydrological cycles and world water resources. *Science* 313, 1068–1072.
- Omer, E.O.M., Alzahane, A.A., 2013. Nitrate contamination of public drinking water sources in Shendi locality, River Nile state, Sudan. *Greener. J. Epidemiol. Public Health* 1, 6–9.
- Pang, K., Luo, K., Zhang, S., Hao, L., 2024. Source-oriented health risk assessment of groundwater based on hydrochemistry and two-dimensional Monte Carlo simulation. *J. Hazard. Mater.* 479, 135666.
- Patil, S., Patil, B., Kadam, A., Wagh, V., Patil, A., Pimparkar, A., Sahu, U., others, 2024. Nitrate and fluoride contamination in the groundwater in a tribal region of north Maharashtra, India: an account of health risks and anthropogenic influence. *Groundw. Sustain. Dev.* 25, 101107.
- Qu, S., Duan, L., Shi, Z., Mao, H., Wang, G., Liu, T., Yu, R., Peng, X., 2022. Identifying the spatial pattern, driving factors and potential human health risks of nitrate and fluoride enriched groundwater of Ordos Basin, Northwest China. *J. Clean Prod.* 376, 134289. <https://doi.org/10.1016/j.jclepro.2022.134289>.
- Rajan, M., Karunanidhi, D., Jaya, J., Preethi, B., Subramani, T., Aravinthasamy, P., 2024. A comprehensive review of human health hazards exposure due to groundwater contamination: a global perspective. *Phys. Chem. Earth Parts A/B/C*, 103637.
- Ramesh, R., Subramanian, M., Lakshmanan, E., Subramaniyan, A., Ganesan, G., 2021. Human health risk assessment using Monte Carlo simulations for groundwater with uranium in southern India. *Ecotoxicol. Env. Saf.* 226, 112781.
- Riedel, T., Kübeck, C., Quirin, M., 2022. Legacy nitrate and trace metal (Mn, Ni, As, Cd, U) pollution in anaerobic groundwater: quantifying potential health risk from the other nitrate problem. *Appl. Geochem.* 139, 105254. <https://doi.org/10.1016/j.apgeochem.2022.105254>.
- Saito, L., Christian, B., Diffley, J., Richter, H., Rohde, M.M., Morrison, S.A., 2021. Managing groundwater to ensure ecosystem function. *Groundwater* 59, 322–333.
- Seben, D., Toebe, M., Wastowski, A.D., Hofstätter, K., Volpatto, F., Zanella, R., Prestes, O. D., Golombieski, J.I., 2021. Water quality variables and emerging environmental contaminant in water for human consumption in Rio Grande do Sul, Brazil. *Environ. Chall.* 5, 100266. <https://doi.org/10.1016/j.envc.2021.100266>.
- Shang, Y., Fu, C., Zhang, W., Li, Xiang, Li, Xiangquan, 2024. Groundwater hydrochemistry, source identification and health assessment based on self-organizing map in an intensive mining area in Shanxi, China. *Env. Res.* 252, 118934.
- Sobol, I.M., 1993. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* 1, 407.
- Su, H., Kang, W., Li, Y., Li, Z., 2021. Fluoride and nitrate contamination of groundwater in the Loess Plateau, China: sources and related human health risks. *Environ. Pollut.* 286. <https://doi.org/10.1016/j.envpol.2021.117287>.
- Suliman, E.D., 2017. Qualitative and Quantitative Assessment of Groundwater Resources and Numerical Simulation in Shendi Sedimentary Subbasin River Nile state - Sudan. Al Neelain University.
- Urme, O., Reza, A.H.M.S., Adham, M.I., Sattar, G.S., 2025. Arsenic, manganese, and iron concentration in groundwater of northwestern part of Bangladesh using self-organizing maps: implication for health risk assessment. *Heliyon*.
- USEPA, 2019. Guidelines for human exposure assessment. Risk Assessment Forum, Washington, DC, p. 223. U.
- USEPA, 2017. Human Health Evaluation Manual, supplemental guidance: update of Standard Default Exposure Factors OSWER Directive 9200. 1–120.
- USEPA, 2006. USEPA region III risk-based concentration table: technical background information.
- USEPA, 1991. Risk-assessment guidance for superfund. Volume 1. Human health evaluation manual. Part A. Interim report (Final).
- Vail, J.R., 1985. Alkaline ring complexes in Sudan. *J. Afr. Earth Sci.* 3, 51–59 (1983).
- Ward, M.H., Jones, R.R., Brender, J.D., De Kok, T.M., Weyer, P.J., Nolan, B.T., Villanueva, C.M., Van Breda, S.G., 2018. Drinking water nitrate and human health: an updated review. *Int. J. Env. Res. Public Health* 15, 1557.
- Whiteman, A.J., 1971. Geology of the Sudan Republic.
- WHO, 2011. Guidelines for drinking-water quality. *WHO Chron.* 38, 104–108.
- Xiao, Y., Hao, Q., Zhang, Y., Zhu, Y., Yin, S., Qin, L., Li, X., 2022. Investigating sources, driving forces and potential health risks of nitrate and fluoride in groundwater of a typical alluvial fan plain. *Sci. Total Environ.* 802, 149909. <https://doi.org/10.1016/j.scitotenv.2021.149909>.
- Yang, J., Wang, J., Xu, H., Xu, Z., Zhang, Y., Chen, J., 2024. Fluoride and nitrogen contamination and potential health risks in the groundwater of a typical agricultural region. *J. Water Health* 22, 2331–2345.
- Yang, X., Du, J., Jia, C., Yang, T., Shao, S., 2025. Unravelling integrated groundwater management in pollution-prone agricultural cities: a synergistic approach combining probabilistic risk, source apportionment and artificial intelligence. *J. Hazard. Mater.* 481, 136514.
- Yazman, M.M., Yüksel, B., Ustaoglu, F., Şen, N., Tepe, Y., Tokatli, C., 2024. Investigation of groundwater quality in the Southern Coast of the Black Sea: application of computational health risk assessment in Giresun. *Türk. Environ. Sci. Pollut. Res.* 31, 52306–52325. <https://doi.org/10.1007/s11356-024-34712-w>.
- Yu, D., Deng, J., Jiang, Q., Liu, H., Yu, C., Ma, H., Pu, S., 2024. Evaluation of groundwater quality with multi-source pollution based on source identification and health risks. *Sci. Total Environ.* 949, 175064.
- Yüksel, B., Ustaoglu, F., Arica, E., 2021. Impacts of a garbage disposal facility on the water quality of Çavuşlu Stream in Giresun, Turkey: a health risk assessment study by a validated ICP-MS assay. *Aquat. Sci. Eng.* 36, 181–192.