Pol. J. Environ. Stud. Vol. XX, No. X (XXXX), 1-13
DOI: *10.15244/pjoes/204425* ONLINE PUBLICATION DATE:

Original Research

Study on the Evaluation of Shallow Groundwater Quality and Health Risk in Luannan County Based on Random Forest Approach

Yue Wang[1,2,3], Zhihui Qu[1,2,3]*, Zhiqiang Gong[4,5], Yang Wang[4,5], Sihui Shao[4,5], Wei Du[6], Xiaoyu Fang[1,2], Ze Liu[1,2]

[1]School of Ecology and Environment, Institute of Disaster Prevention, Sanhe, Hebei, China
[2]Hebei Key Laboratory of Resource and Environmental Disaster Mechanism and Risk Monitoring, Sanhe, Hebei, China
[3]Beijing Disaster Prevention Science and Technology CO.LTD
[4]Hebei Geological Environment Monitoring Institute, Shijiazhuang, Hebei, China [5]Hebei Key Laboratory of Geological Resources and Environment Monitoring and Protection, Shijiazhuang, Hebei, China
[6]Suzhou Qianxing Environmental Engineering Co., Ltd., Taicang 215431, China

Received: 15 February 2025
Accepted: 27 April 2025

Abstract

This paper collected 134 groups of shallow groundwater samples from Luannan County from 2015 to 2023, constructed the groundwater quality evaluation model by the Random Forest method, and used the USEPA model to evaluate human health. The results showed that: (1) the groundwater quality in the study area was generally good, and only the samples of superwater accounted for 10.4%. The areas with poor water quality are concentrated in the southwest and north, and the central and eastern regions are good; (2) the importance evaluation of Random Forest shows that Mn^{2+} is the most important index affecting groundwater quality, mainly derived from mine drainage and unreasonable use of pesticides and fertilizers; (3) the proportion of children with H_{Itotal} greater than 1 reaches 64%, indicating that children are the most vulnerable population. Measures such as controlling pollution sources, regular monitoring, and health risk assessment are recommended to reduce risk.

Keywords: random forest approach, shallow groundwater, groundwater quality assessment, health risk assessment

Introduction

Groundwater is one of the most important water resources for nature and human societies [1-5] and is facing a serious challenge of deterioration in quality due to the impacts of climate change and human activities [6-8]. Groundwater quality evaluation, as an effective means of water body pollution assessment, can quantitatively and qualitatively assess the pollution status of groundwater bodies and is also an important basic task for groundwater environmental risk analysis, pollution source determination, and water resource protection [9]. For the evaluation of groundwater quality, scholars at home and abroad have explored

*e-mail: qzh-19831001@163.com

various methods: Liang et al. [10] established a fuzzy comprehensive evaluation model of groundwater quality based on Geographic Information System (GIS) technology by combining GIS technology with the improved fuzzy comprehensive evaluation model; Li et al. [11] evaluated the quality of deeply buried groundwater in Kaifeng City using genetic algorithm–BP neural network method; Egbueri et al. [12] used groundwater pollution index (PIG), ecological risk index (ERI), and hierarchical clustering algorithm to evaluate groundwater quality in the study area; Su et al. [13] used a set-pair analysis (SPA)-Markov chain model to evaluate groundwater quality in Xi'an city.

Although the above studies have explored the groundwater quality evaluation methods from different perspectives and achieved many useful results, there are some limitations: the evaluation results of the fuzzy mathematical method are more in line with the actual situation, but the determination of the relevant parameters such as weights and affiliation degrees mainly depends on the subjective experience of the human being [14]; the evaluation results of the evaluation methods, such as neural networks, are not intuitive enough and the construction of the model and the calculation process are relatively complicated [15]. In recent years, with the development of machine learning models, the Random Forest method has been widely used due to its simple operation, high prediction accuracy, and ability to identify the importance of evaluation indicators [16, 17]. In the field of hydrogeology, [18] used the Random Forest method to analyze the change in groundwater burial depth and its causes in the middle reaches of the Heihe River; Band et al. [19] compared the performance of 4 artificial intelligence models in simulating nitrate concentration of groundwater: cubic regression, support vector machine, Random Forest, and Bayesian artificial neural network; and Zhang Ying et al. used the Random Forest classification method to evaluate the water quality of Chaohu Lake [20].

Currently, studies based on the Random Forest method and evaluating health risks in human beings, although there are some results, are still relatively few, especially in this study area. Chu et al. [21] conducted a health risk assessment and groundwater quality classification of arsenic in the Salt Lake area based on Random Forest and found that 33% of the groundwater samples had HQ values >1, and all groundwater had CR values $>1.00\times10^{-6}$ for children. which implies a serious health risk. 21% of groundwater samples had a health risk for adults. Muhammad et al. [22] and Sultana et al. [23] evaluated the health risk of arsenic (As) in Pakistan and found that more than 85% of drinking water had Hazard Quotient (HQ) values >1, indicating a significant health risk. In the semi-arid region of northwestern China, Li et al. [24] revealed the impact of groundwater contaminants on human health, stating that women and children face higher non-carcinogenic risks than men due to industrial and agricultural activities. In addition, other scholars have conducted health risk assessments of drinking water to check the adverse effects of groundwater contaminants on human health in certain areas.

Therefore, this study adopts the Random Forest method combined with health risk assessment to conduct a systematic study of shallow groundwater in Luannan County. The core objective of the study is to accurately identify the most influential indicators on groundwater quality in Luannan County from the 9 selected indicators using the Random Forest algorithm. Based on this, the health risk assessment methodology was further applied to scientifically and systematically assess the health risks of the screened key indicators. Through this innovative research methodology, the study aims to provide scientific evidence and technical support for the management and protection of groundwater quality in Luannan County.

Materials and Methods

Study Area

Luannan County is located in the eastern part of Hebei Province, on the Jidong Plain, and covers a total land area of 117,000 km^2 (see Fig. 1). The Quaternary loose accumulations in Luannan County are vertically divided into four aquifer groups, from top to bottom [25]. The I and II aquifer groups are subsurface water-bearing strata, located in the surface and shallow zones. They receive recharge from atmospheric precipitation and discharge through evaporation. Water circulation in this zone is active, making it a vertically strong circulation alternation zone. The III and IV aquifer groups are deep water-bearing strata characterized by pressurized conditions, limited runoff, and weak circulation [26].

Vertically, each aquifer group is separated by layers of powdery clay or clay thicker than 5 m, with no obvious hydraulic connection. However, macro-scale analysis shows that between the I and II aquifer groups, as well as the II and III groups, particularly across the freshwater zones, mixed use of aquifers and varying extraction well depths have led to long-term interconnection due to exploitation, resulting in different degrees of hydraulic connectivity [27].

The lithology of each aquifer group is dominated by gravel pebbles and sand-bearing gravel pebbles (see Fig. 2), and their distribution ranges from old to new, with a tendency to decrease gradually. The thickness of a single layer of aquifer is 15-20 m, and the maximum thickness is more than 40 m [28].

Groundwater Sample Collection and Analysis

In this study, a total of 134 shallow groundwater samples were collected from 2015 to 2023, and a total of nine indicators, pH, total dissolved solids (TDS), total hardness (TH), SO_4^{2-}, NO_3^-, F^-, Cl^-, chemical oxygen

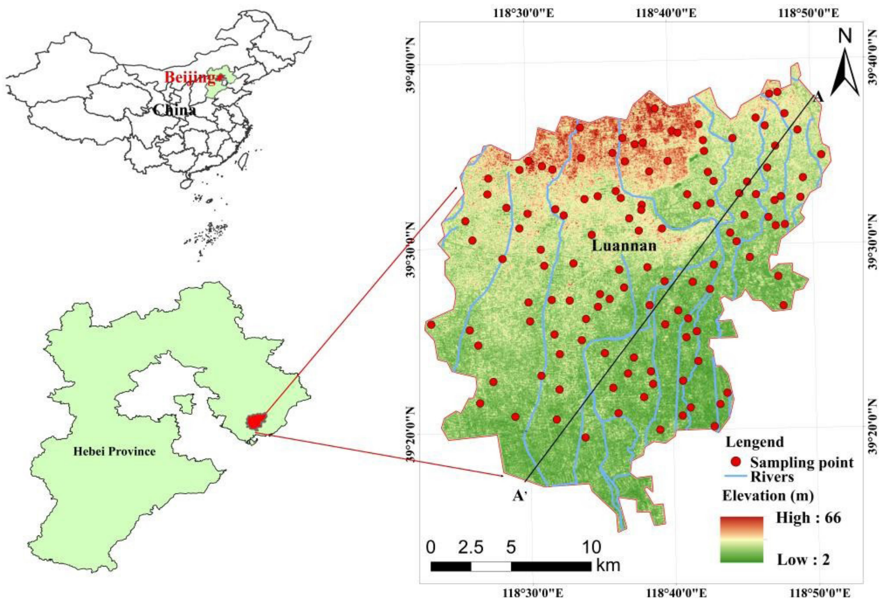

Fig. 1. Comprehensive map of the study area.

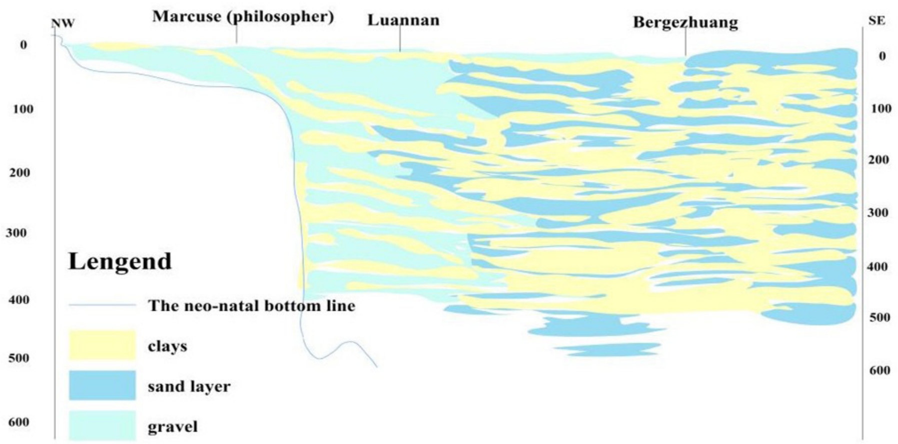

Fig. 2. A-A geological profile.

demand (COD), and Mn^{2+}, which mainly affect the local water quality, were selected as evaluation factors. The specific distribution is shown in Fig. 1. Groundwater samples were collected in accordance with the Technical Specification for Groundwater Environmental Monitoring (HJ164-2020), and the determination was carried out by the Experimental Centre for Water Environment Monitoring of Hebei Province.

Inductively coupled plasma mass spectrometry for the determination of 65 elements in water quality (HJ700-2014) was used to determine the concentration of dissolved divalent manganese (Mn^{2+}) in the samples, and ion chromatography for the determination of inorganic anions in water quality (F^-, Cl^-, NO_2^-, Br^-, NO_3^-, PO_4^{3-}, SO_3^{2-}, SO_4^{2-}) (HJ84-2016) was used for the determination of SO_4^{2-}, NO_3^-, F^-, and Cl^- content in groundwater; TH was determined by EDTA titration, TDS by gravimetric method, and pH by glass electrode method.

Random Forest Approach

Random Forest Modelling

The Random Forest algorithm consists of a series of decision tree single learners, which individually vote on the input samples and judge their classification; the results of each single learner are then pooled to arrive at the final result of the Random Forest [29]. The flow of the Random Forest algorithm is shown in Fig. 3.

The training samples directly determine the accuracy of the Random Forest algorithm for groundwater quality evaluation. This study used the EWQI evaluation method to generate training samples. The 134 groups of water quality samples in the study area were classified

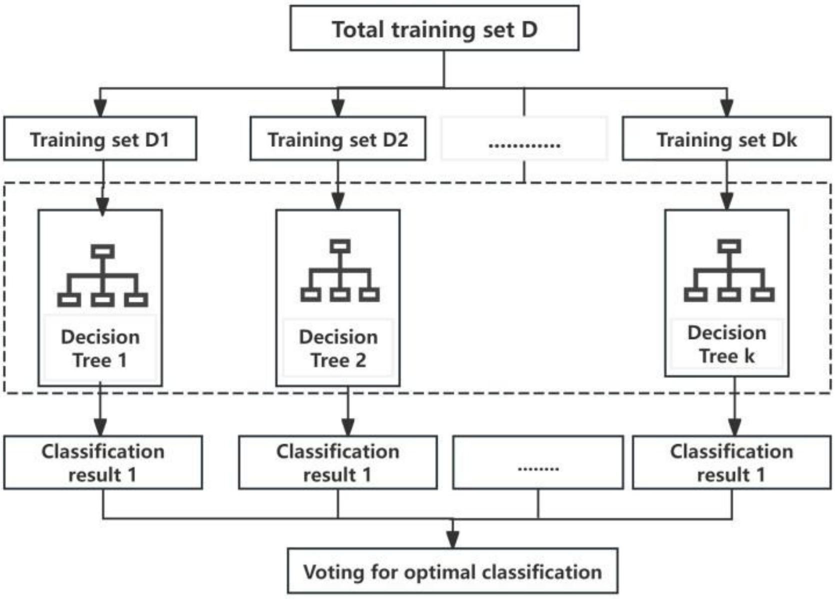

Fig. 3. Schematic diagram of the Random Forest algorithm.

into groundwater quality classes using the EWQI evaluation method. The entropy-weighted water quality index (EWQI) is a method proposed by [30] for the comprehensive evaluation of groundwater quality. The calculation process of the entropy weight method is similar to other comprehensive index methods, and the calculation steps are as follows:

(1) Step 1: Establish the initial matrix. Assume that there are m (i=1, 2, ... , m) water samples, each water sample contains n (j=1, 2, ... , n) water quality parameters, then the initial matrix X is as follows:

$$X = \begin{bmatrix} X_{11} X_{12} \cdots X_{1n} \\ X_{21} X_{22} \cdots X_{2n} \\ \vdots \vdots \ddots \vdots \\ X_{m1} X_{m2} \dots X_{mn} \end{bmatrix} \tag{1}$$

(2) Step 2: Normalize the initial matrix. The matrix X is standardized to eliminate the effect of dimensionality through Equation (2), where $min_j(x_{ij})$ and $max_j(x_{ij})$ are the minimum and maximum values of the j th indicator in the matrix X, respectively. y_{ij} is the normalized value. Then, the standardized assessment matrix is written as $Y=(y_{ij})_{(m\times n)}$

$$y_{ij} = \begin{cases} \dfrac{x_{ij}-min_j(x_{ij})}{max_j(x_{ij})-min_j(x_{ij})} & (efficiency\ \ type) \\ \dfrac{max_j(x_{ij})-x_{ij}}{max_j(x_{ij})-min_j(x_{ij})} & (cost\ \ type) \end{cases} \tag{2}$$

(3) Step 3: Calculate the information entropy of each indicator by Equation (4). The ratio of the value of the j-th indicator in the i-th sample is represented by Equation (3) and 10^{-4} is a correction parameter used to make sense of the Equation when y_{ij} is zero.

$$p_{ij} = \frac{y_{ij}+10^{-4}}{\sum_{i=1}^{m}(y_{ij}+10^{-4})} \tag{3}$$

$$e_j = -\frac{1}{\ln m}\sum_{i=1}^{m} p_{ij} \ln p_{ij} \tag{4}$$

(4) Step 4: According to Equation (5), the entropy weight (w_j) can be calculated:

$$w_j = \frac{1-e_j}{\sum_{j=1}^{n}(1-e_j)} \tag{5}$$

(5) Step 5: The EWQI value for each sample is calculated by Equations (6) and (7).

$$q_j = \frac{C_j}{S_j} \times 100 \tag{6}$$

$$\mathrm{EWQI} = \sum_{j=1}^{n} w_j q_j \tag{7}$$

where q_j is the concentration ratio of indicator j in each sample, C_j is the measured concentration (mg/L) in each sample, and S_j is the standard limit value of Class III water according to the Chinese groundwater quality standard in mg/L [31].

The accuracy and stability of the Random Forest water quality classification model were examined by randomly selecting 70% (93 groups) of the 134 sets of water quality samples as the training dataset and the remaining 30% (40 groups) as the test dataset. During the model training process, nine water quality evaluation factors, pH, TDS, TH, SO_4^{2-}, NO_3^-, F^-, Cl^-, COD, and Mn^{2+}, were used as independent variables, and the EWQI water quality evaluation classification results were used as dependent variables. By continuously adjusting and optimizing the model parameters, the simulated values of water quality classification of the model and the classification results of EWQI water quality evaluation were matched as much as possible to complete the training and optimization of the Random Forest classification model.

Hyperparametric Optimization of Random Forest Models

In order to prevent the random selection of hyperparameters and overfitting in the decision tree construction process of the Random Forest model, this study first uses the lattice search algorithm in the sklearn machine learning library to optimize the parameters of relevant hyperparameters (the number of decision trees, the maximum depth of the decision tree, and the maximum number of features, etc.) of the model as a means to improve the classification accuracy and efficiency of the Random Forest model. The lattice search algorithm is a model hyperparameter optimization technique that optimizes model performance by traversing a given combination of parameters. The number of decision trees in the Random Forest algorithm is usually the more the better, but if the number of decision trees is too large, it will bring a larger computational burden, and the computation time will increase accordingly; at the same time, when the number of decision trees reaches a critical value, further increasing the number of trees will not significantly improve the model's classification performance. The optimization result of the grid search algorithm for the number of decision trees is shown in Fig. 4.

Random Forest Classification Indicator Importance Assessment

The groundwater quality evaluation can identify the groundwater pollution status in the study area, but it cannot reveal the relative importance between different evaluation indicators. One obvious advantage of the Random Forest model is that the importance of each indicator to the water quality classification results can be assessed by the Gini index, and the larger the Gini index, the higher the relative importance of the evaluation indicators. The relative importance of the evaluation indicators of groundwater quality in the study area is shown in Fig. 5.

Health Risk Assessment

Health risk assessment helps to clarify priorities for pollution control and integrated groundwater quality assessment [32]. According to the International Agency for Research on Cancer (IARC), Mn^{2+} is a non-carcinogenic pollutant. Therefore, this paper used the model recommended by the Ministry of Environmental Protection of the People's Republic of China (2014) to estimate the non-carcinogenic health risk caused by Mn^{2+} [33]. Drinking and dermal contact are two common modes of exposure to contaminated water affecting people's health [34], and the non-carcinogenic risk calculations through drinking and dermal exposure were calculated by the following equations [35, 36]. In the present study, the non-carcinogenic risk through alcohol consumption (oral route) and dermal intake was estimated with the following formula [37-39]. The meaning of each parameter is presented in Table 1.

$$Intake_{oral} = \frac{C \times IR \times EF \times ED}{BW \times AT} \tag{8}$$

$$Intake_{dermal} = \frac{K \times C \times t \times CF \times SA \times EV \times EF \times ED}{BW \times AT} \tag{9}$$

$$\mathrm{SA} = 239 \times \mathrm{H}^{0.417} \times \mathrm{BW}^{0.517} \tag{10}$$

The hazard quotient (HQ) of the non-carcinogenic risk of Mn^{2+} through consumption (HQ_{oral}) and dermal contact (HQ_{dermal}) can be expressed as at least one of the Equations (11) and (12). See Table 1 for parameter values.

$$HQ_{oral} = \frac{Intake_{oral}}{RfD_{oral}} \tag{11}$$

$$HQ_{dermal} = \frac{Intake_{dermal}}{RfD_{dermal}} \tag{12}$$

The total non-carcinogenic risk in this study was quantified by the hazard index (HI), which can be calculated by Equation (13). HQ and HI values less than 1 indicate acceptable non-carcinogenic risk, and vice versa when HQ and HI are greater than 1.

$$\mathrm{HI} = HQ_{oral} + HQ_{dermal} \tag{13}$$

Results and Discussion

Web Search Optimization

It can be seen from Fig. 4 that with the increase in the number of decision trees, the error decreases. When the number of decision trees reaches 60, the error stabilizes, and this point also corresponds to the lowest error.

The other hyperparameters are further optimized using the grid search algorithm, and the optimization results are shown in Table 2.

Relative Importance

The relative importance of groundwater quality evaluation indicators in the study area is shown in Fig. 5. From Fig. 5, it can be seen that the order of importance

Table 1. Parameter values for different exposure pathways in the health risk model.

Pathway	Parameters	Unit	Children	Female	Males
Oral intake	$Intake_{oral}$ (chronic daily intake)	mg/(kg day)	/	/	/
	C (pollutant concentration)	mg/L	/	/	/
	IR (intake rate)	L/day	0.7	1.5	1.5
	EF (exposure frequency)	day/year	365	365	365
	ED (exposure duration)	year	12	30	30
	BW (body weight)	kg	15	55	70
	AT (average exposure time)	day	4380	10950	10950
	HQ_{oral} (hazard quotient)	/	/	/	/
	RfD_{oral} (reference dosage for Mn)	mg/(kg day)	0.046	0.046	0.046
Dermal contact	$Intake_{dermal}$ (chronic daily intake)	mg/(kg day)	/	/	/
	K (dermal permeability coefficient)	cm/h	0.001	0.001	0.001
	t (contact duration)	h/day	0.4	0.4	0.4
	CF (units conversion factor)	/	0.001	0.001	0.001
	EV (daily exposure rate)	/	1	1	1
	SA (exposed skin area)	cm^2	/	/	/
	H (height of a person)	cm	99.4	153.4	165.3
	HQ_{dermal} (hazard quotient)	/	/	/	/
	RfD_{dermal} (reference dosage for Mn)	mg/(kg day)	0.0018	0.0018	0.0018

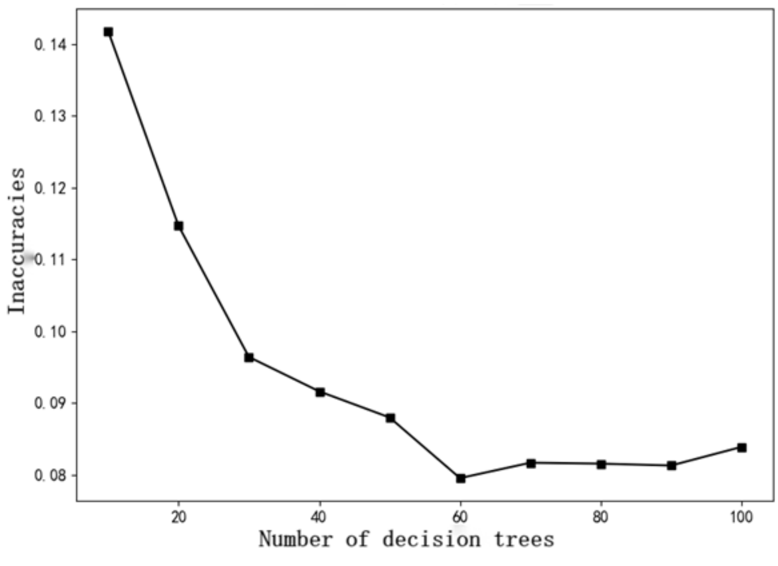

Fig. 4. Relationship between the number of decision trees and the error.

of groundwater quality evaluation indicators in the study area is Mn^{2+} > TH > SO_4^{2-} > TDS > COD > Cl^- > NO_3^- > F^- > pH. Among these indicators, the importance of Mn^{2+} and TH is more prominent, and both of them have Gini indexes exceeding 20%, which makes them the main controlling factors for groundwater quality in the study area. The Gini indexes of NO_3^-, F^-, and pH were less than 5%, indicating that they were of low importance, while the Gini index of Mn^{2+} was more than 30%, making it the most important.

The reason for this may be that, on the one hand, there are some industrial enterprises in Luannan County. The wastewater discharged by these industrial enterprises contains Mn^{2+} and other pollutants, and this wastewater

Table 2. Random Forest hyperparameter optimization results.

Hyperparameterization	Parametric
Number of decision trees	60
Maximum number of features	2
Minimum number of samples for leaf nodes	2
Maximum depth of the decision tree	3
Minimum number of samples required for internal node repartitioning	2

is discharged directly without effective treatment. It then penetrates into the underground aquifer through soil pore spaces and other pathways, leading to elevated Mn^{2+} levels in the groundwater. For example, Sijiaying Iron Ore Mine South and Macheng Iron Ore Mine are important iron ore resources in Luannan County. Sijiaying Iron Ore Mine South is a super-large-scale low-grade iron ore deposit in China, with B+C+D grade iron ore resource reserves amounting to 14.50×10^8 t, while Macheng Iron Ore Mine is the largest single iron ore deposit discovered in China since the 1980s, with total reserves of 10.44×10^8 t. Iron ore mining activities can lead to an increase in the Mn^{2+} content of groundwater through a combination of acidic wastewater formation, leaching of waste rock and tailings, mine drainage, and the regional geological background.

On the other hand, Luannan County has a large amount of agricultural land, where farmers may overuse fertilizers and pesticides. These substances seep into the underground aquifer with irrigation water or rainwater, and the Mn^{2+} they contain gradually accumulates, resulting in increased Mn^{2+} concentrations in the groundwater [40].

Water Quality Assessment

The 134 groups of water quality samples in the study area were classified into groundwater quality classes using the EWQI evaluation method, and the results of the classification are shown in Table 3. It shows that there is no Class I in the study area, and the numbers of Class II, Class III, Class IV, and Class V water quality are 25, 95, 13, and 1 groups, respectively, with Class III water accounting for the highest percentage. This indicates that the groundwater quality in the study area is better as a whole.

As can be seen from Fig. 6, the proportion of Class II water is 19%, Class III water is 71%, Class IV water and Class V water is 10%, and the poorer water quality areas are mainly concentrated in the southwest and north. Meanwhile, the better water quality areas are distributed in the center and the east, and on the whole, the water quality in Luannan County is better. The points with numbers in the figure indicate poor water quality. The poor water quality in the northern region may be influenced by mine pit drainage, while the poor water quality in the southern region is mainly due to the fact that it is predominantly agricultural land. This suggests that the southern region's groundwater deterioration is primarily caused by the excessive use of pesticides and chemical fertilizers, which leads to declining groundwater quality [41]. According to statistics, the amount of fertilizer applied in Luannan County is much higher than the safety limit of 225 kg/hm^2 set by developed countries to prevent fertilizer pollution, and the impact of agricultural surface pollution on river water quality and groundwater can no longer be ignored [42].

The rapid development of the agricultural economy in Luannan County has put forward higher requirements for groundwater extraction and development. Changes in land use types can reflect changes in human activities

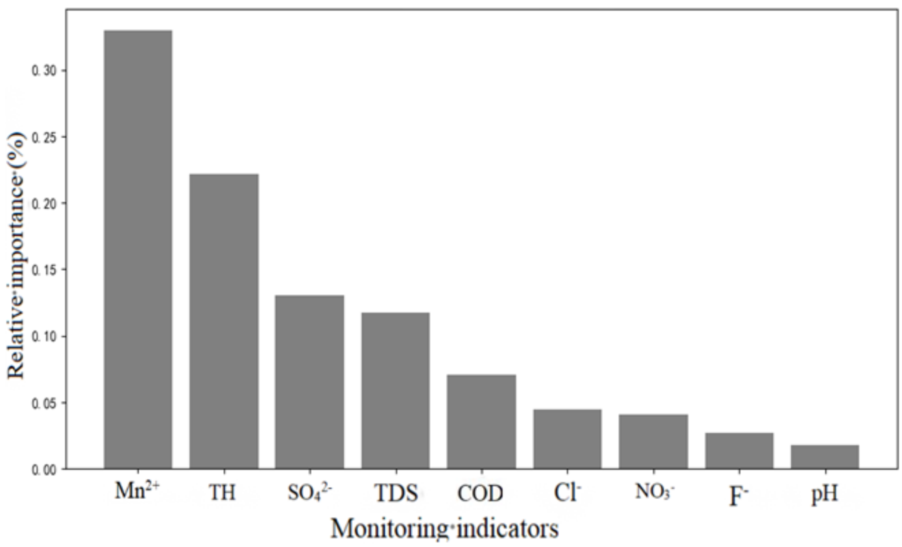

Fig. 5. Ranking of relative importance of groundwater quality indicators.

Table 3. EWQI-based groundwater quality classification and the number of samples of different classes in the study area.

EWQI	Rank	Water quality	Number of samples
<25	I	Excellent	0
[25, 50]	II	Good	25
[50, 100]	III	Medium	95
[100, 150]	IV	Poor	13
>150	V	Very poor	1

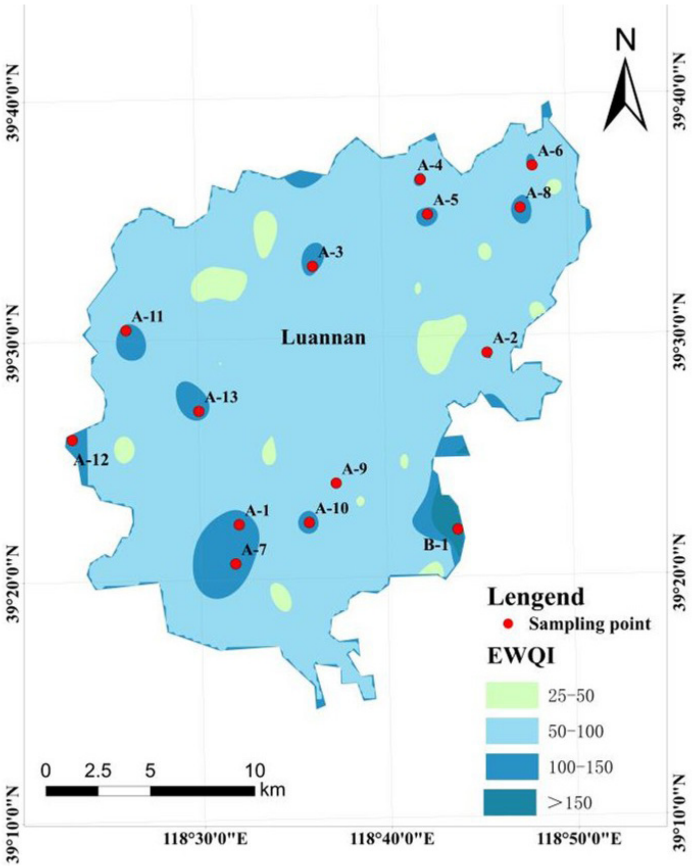

Fig. 6. Groundwater quality map based on EWQI. Higher value indicates worse water quality.

and are often used to study human activities. The population growth rate in the area is also a reflection of human activities. The increase in pollution puts a lot of pressure on groundwater, and the type of land use affects the groundwater cycle [43, 44].

Health Risk Assessment

In the evaluation of groundwater quality, the spatial distribution of physico-chemical parameters is crucial for understanding the water quality conditions. Fig. 7 shows that the spatial distribution of physicochemical parameters within the study area shows some similarity, with high concentration values concentrated in the southwest and northern regions. This distribution pattern is consistent with the overall groundwater quality map (Fig. 4), further confirming that these regions may be high-risk areas for groundwater contamination. In particular, the distribution map of Mn^{2+} ions showed that all points with high concentration values exceeded the standard of Class III water, which verified that Mn^{2+} ions are one of the most important indicators affecting the quality of groundwater, which is consistent with the previous results obtained through the Random Forest method.

Mn^{2+} is an important non-carcinogenic factor affecting human health. Therefore, Mn^{2+} was selected for non-carcinogenic health risk assessment in this study. Non-carcinogenic risk assessment was performed separately for men, women, and children [45-47], and the results are shown in Table 4.

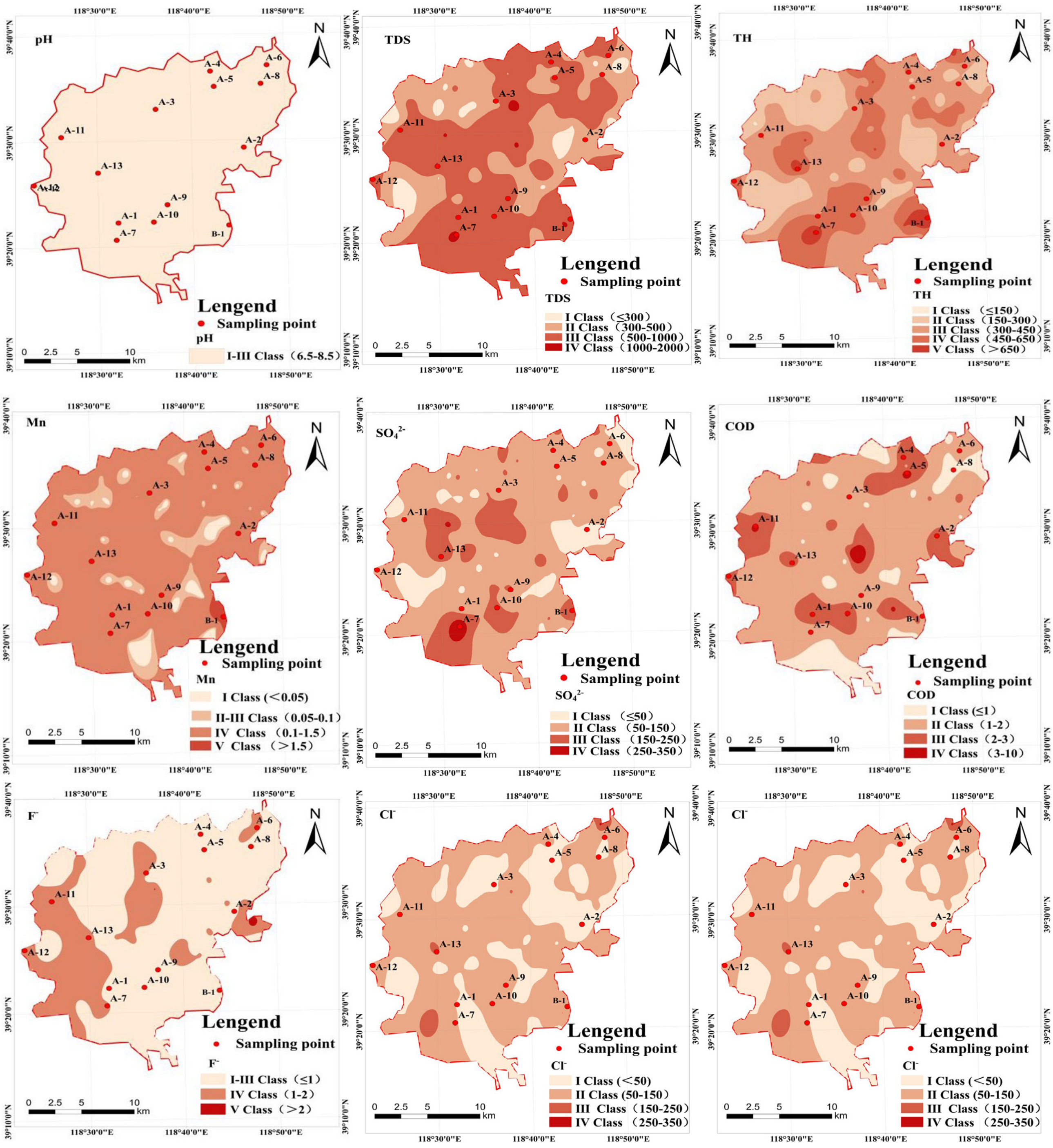

Fig. 7. Spatial distribution of physico-chemical parameters.

Table 4 shows that the HQ_{oral} for children ranged from 0.43 to 6.51 with a mean value of 1.33. 43% of the samples had HQ_{oral} greater than 1, whereas the HQ_{dermal} for children were all less than 1, suggesting a lower risk due to dermal exposure. For adults, the maximum risk due to dermal exposure (HQ_{dermal}) was 2.99 and 3.81 for males and females, respectively, while the maximum risk due to oral ingestion exposure (HQ_{oral}) was 3.36 and 4.21 for males and females, respectively, suggesting that the risk due to oral ingestion is greater than that due to dermal exposure.

As for the total risk (HI_{total}), 64%, 1%, and 1% of the samples had a HI_{total} greater than 1 for children, females, and males, respectively. The results suggest that children are the most vulnerable population because of their lower body weight [48, 49].

In this study, a deterministic approach was used for human risk assessment, and the values of BW, AT, and H are the results of statistical investigations. They

Table 4. Non-Carcinogenic Risks through Drinking Water Intake and Dermal Exposure in Different Populations.

Sample ID	HQ_{oral}(Children)	HQ_{oral}(Females)	HQ_{oral}(Males)	HQ_{dermal}(Children)	HQ_{dermal}(Females)	HQ_{dermal}(Males)	HI (Children)	HI (Females)	HI (Males)	Children	Females	Males	Children	Females	Males
A-1	1.01	0.68	0.53	0.11	0.07	0.07	1.13	0.75	0.60	0.05	0.03	0.02	0.00	0.00	0.00
A-2	0.93	0.55	0.43	0.09	0.06	0.06	1.02	0.60	0.48	0.04	0.03	0.02	0.00	0.000	0.00
A-3	1.30	0.76	0.60	0.13	0.08	0.07	1.42	0.84	0.67	0.06	0.04	0.03	0.00	0.00	0.00
A-4	0.94	0.55	0.43	0.09	0.06	0.05	1.03	0.61	0.49	0.04	0.03	0.02	0.00	0.00	0.00
A-5	1.25	0.73	0.57	0.12	0.08	0.07	1.37	0.81	0.64	0.06	0.03	0.03	0.00	0.00	0.00
A-6	0.96	0.56	0.44	0.09	0.06	0.05	1.06	0.62	0.50	0.04	0.03	0.02	0.00	0.00	0.00
A-7	1.07	0.62	0.49	0.10	0.07	0.06	1.17	0.69	0.55	0.05	0.03	0.02	0.00	0.00	0.00
A-8	1.40	0.82	0.64	0.14	0.09	0.08	1.53	0.90	0.72	0.06	0.04	0.03	0.00	0.00	0.00
A-9	0.43	0.25	0.20	0.04	0.03	0.02	0.47	0.28	0.22	0.02	0.01	0.01	0.00	0.00	0.00
A-10	0.83	0.49	0.38	0.08	0.05	0.05	0.91	0.54	0.43	0.04	0.02	0.02	0.00	0.00	0.00
A-11	0.74	0.43	0.34	0.07	0.05	0.04	0.81	0.48	0.38	0.03	0.02	0.02	0.00	0.00	0.00
A-12	0.44	0.25	0.20	0.04	0.03	0.02	0.48	0.28	0.22	0.02	0.01	0.01	0.00	0.00	0.00
A-13	0.82	0.48	0.38	0.08	0.05	0.05	0.90	0.53	0.42	0.04	0.02	0.02	0.00	0.00	0.00
B-1	6.51	3.81	2.99	0.63	0.40	0.4	7.14	4.21	3.36	0.30	0.18	0.14	0.00	0.00	0.00

represent only the average level, and the results of the study can provide general risk information for decision makers. Based on the results of the water quality assessment, policymakers are recommended to take the following measures to protect and improve water quality: First, residents' awareness of safe drinking water should be raised, as many residents do not have sufficient knowledge of safe and healthy drinking water and are not fully aware of the possible health risks associated with drinking contaminated groundwater. Therefore, managers need to urge villagers to change their habit of drinking groundwater directly, which can be achieved through education and publicity campaigns. Secondly, the water supply sector needs to strengthen the construction of centralized water supply systems to ensure the quantity and quality of water supply, so as to provide basic protection for residents' daily use of water and to reduce the direct access of residents to poor-quality groundwater. In addition, local authorities need to strengthen groundwater monitoring to ensure the health of the groundwater environment, which can also ensure the safety of residents using groundwater. Through these measures, groundwater quality can be effectively protected and improved to ensure the safety of residents' drinking water.

Conclusions

In this paper, a Random Forest classification model for conducting groundwater quality assessment was constructed, and 70% (93 groups) of the 134 groups of groundwater quality samples in Luannan County were randomly selected as the training dataset. The remaining 30% (40 groups) of data were used as the test dataset; the hyperparameters of the Random Forest model were optimized using the grid search algorithm, and the following conclusions were drawn:

(1) The overall groundwater quality condition in the study area is good, and the indicators of exceeding Class III water only account for 10%. The areas with poor water quality are mainly concentrated in the southwest and north, while the areas with good water quality are distributed in the center and east. Evaluation of the importance of categorical indicators of Random Forests shows that the most important influence indicator of groundwater quality in the study area is Mn^{2+}. Among them, the main sources of Mn^{2+} are mine pit drainage and excessive use of pesticides and fertilizers.

(2) In this study, Mn^{2+} was selected for non-carcinogenic risk assessment for men, women, and children, respectively. The results showed that the risk due to oral intake was greater than that due to dermal contact, and the percentage of samples with $HI_{total} > 1$ was 1% for both men and women, while the percentage of samples with $HI_{total} > 1$ for children was 64%. This indicates that children are the most vulnerable group, and Mn^{2+} poses a health risk to human beings. Certain measures, such as controlling pollution and reducing Mn^{2+} levels in groundwater, must be taken to minimize health risks.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (ZY20210304) and the Hebei Provincial Key Laboratory of Geological Resources and Environment Monitoring and Protection Funding Project (JCYKT202408).

Conflict of Interest

The authors declare no conflict of interest.

References

1. LI P., QIAN H. Water resources research to support a sustainable China. International Journal of Water Resources Development. **34** (3), 327, **2018**.
2. RAHMAN M.M., BODRUD-DOZA M., SIDDIQUA M.T., ZAHID A., ISLAM A. Spatiotemporal distribution of fluoride in drinking water and associated probabilistic human health risk appraisal in the coastal region, Bangladesh. Science of the Total Environment. **724**, 138316, **2020**.
3. SHAMSUDDUHA M., JOSEPH G., HAQUE S.S., KHAN M.R., ZAHID A., AHMED K.M.U. Multi-hazard Groundwater Risks to Water Supply from Shallow Depths: Challenges to Achieving the Sustainable Development Goals in Bangladesh. Exposure and Health. **12** (4), 657, **2020**.
4. LI W., WU J., ZHOU C., NSABIMANA A. Groundwater Pollution Source Identification and Apportionment Using PMF and PCA-APCS-MLR Receptor Models in Tongchuan City, China. Archives of Environmental Contamination and Toxicology. **81** (3), 397, **2021**.
5. SNOUSY M.G., WU J., SU F., ABDELHALIM A., ISMAIL E. Groundwater Quality and Its Regulating Geochemical Processes in Assiut Province, Egypt. Exposure and Health. **14**, 305, **2021**.
6. CUTHBERT M.O., GLEESON T., MOOSDORF N., BEFUS K.M., SCHNEIDER A., HARTMANN J., LEHNER B. Global patterns and dynamics of climate–groundwater interactions. Nature Climate Change. **9** (2), 137, **2019**.
7. KANAGARAJ G., ELANGO L. Chromium and fluoride contamination in groundwater around leather tanning industries in southern India: Implications from stable isotopic ratio δ53Cr/δ52Cr, geochemical and geostatistical modelling. Chemosphere. **220**, 943, **2019**.
8. RAO N.S., DINAKAR A., SRAVANTHI M., KUMARI B.K. Geochemical characteristics and quality of groundwater evaluation for drinking, irrigation, and industrial purposes from a part of hard rock aquifer of South India. Environmental Science and Pollution Research. **28** (24), 31941, **2021**.
9. FANG Y.H., ZHENG X.L., PENG H., WANG H., ZHANG B. Groundwater quality assessment based on optimization of fuzzy syntheticevaluation. Earth Science Frontiers. **26**

(4), 301, **2019**.
10. LIANG N.S., QIAN C., MU W.P., DUAN Y., ZHU G., ZHANG R.S., WU X. Fuzzy comprehensive evaluation of groundwater quality of the Daniudi gas field area. Hydrogeology & Engineering. **47** (3), 52, **2020**.
11. LI S.Q., WANG X.Y., JI H.Y., ZHAO W., LIU X.M. Evaluation of Deep Buried Groundwater Based onGenetic Algorithm and BP Neural Network. Water Resources and Power. **37** (1), 49, **2019**.
12. EGBUERI J.C. Groundwater quality assessment using pollution index of groundwater (PIG), ecological risk index (ERI) and hierarchical cluster analysis (HCA): A case study. Groundwater for Sustainable Development. **10**, 100292, **2020**.
13. SU F., WU J., HE S. Set pair analysis-Markov chain model for groundwater quality assessment and prediction: A case study of Xi'an city, China. Human and Ecological Risk Assessment: An International Journal. **25**, 158, **2019**.
14. KANG X.B., LI K., ZHU Z.Q., LIU Q.H., LIU X. Application of Cloud Model Based on Fusion Weight in Groundwater Quality Evaluation in Xichang Area. Water-saving Irrigation. (7), 62, **2019**.
15. YAN B., JIANG X.H., ZHONG Z.H., LIU Y.T., WANG T.H. Water Quality Evaluation of the Upstream of Dahuofang ReservoirBased on Comprehensive Water Quality Identification Index Methodof Improved Weight. Journal of Shenyang Agricultural University. **50** (3), 314, **2019**.
16. CHENG L., CHEN X., DE VOS J., LAI X., WITLOX F. Applying a Random Forest method approach to model travel mode choice behavior. Travel Behaviour and Society. **14**, 1, **2019**.
17. PAUL A., MUKHERJEE D.P., DAS P., GANGOPADHYAY A., CHINTHA A.R., KUNDU S. Improved Random Forest for Classification. IEEE Trans Image Process. **27** (8), 4012, **2018**.
18. YANG G., SU X.L. Change of Groundwater Depth and Its Causes in Middle Stream of theHeihe River Basin Based on the Random Forest. Research of Soil and Water Conservation. **24** (1), 109, **2017**.
19. BAND S.S., JANIZADEH S., PAL S.C., CHOWDHURI I., SIABI Z., NOROUZI A., MELESSE A.M., SHOKRI M., MOSAVI A. Comparative Analysis of Artificial Intelligence Models for Accurate Estimation of Groundwater Nitrate Concentration. Sensors (Basel). **20** (20), **2020**.
20. ZHANG Y., GAO Q.Q. Water quality evaluation of Chaohu Lake basedon Random Forest method. Chinese Journal of Environmental Engineering. **10** (2), 992, **2016**.
21. WU C., FANG C., WU X., ZHU G. Health-Risk Assessment of Arsenic and Groundwater Quality Classification Using Random Forest in the Yanchi Region of Northwest China. Exposure and Health. **12** (4), 761, **2020**.
22. MUHAMMAD S., TAHIR SHAH M., KHAN S. Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques in Kohistan region, northern Pakistan. Food and Chemical Toxicology. **48** (10), 2855, **2010**.
23. SULTANA J., FAROOQI A., ALI U. Arsenic concentration variability, health risk assessment, and source identification using multivariate analysis in selected villages of public water system, Lahore, Pakistan. Environmental Monitoring and Assessment. **186** (2), 1241, **2014**.
24. LI P., LI X., MENG X., LI M., ZHANG Y. Appraising Groundwater Quality and Health Risks from Contamination in a Semiarid Region of Northwest China. Exposure and Health. **8** (3), 361, **2016**.
25. WANG X.Y. Soil nutrient status and countermeasures for improvement and utilisation in Luannan County. China Agricultural Technology Extension. **28** (9), 41, **2012**.
26. LI M., GE D.Q., ZHANG L., LIU B., GUO X.F., WANG Y. Land Subsidence of coastal area in southern tangshanusing PSInSAR techniqueTECHNIQUE. Journal of Engineering Geology. **24** (4), 704, **2016**.
27. HOU G.H., GAO M.S., DANG X.Z. Hydrochemical characteristics and salinization causes of shallow groundwater in Caofeidian, Tangshan City. Earth Science Frontiers. **26** (6), 49, **2019**.
28. ZHANG L.F., HE M.F. Assessment of Groundwater Quality in Tangshan City Based on Weber-Fechner's Law. Journal of China Hydrology. **31** (2), 75, **2011**.
29. AVILA R.G., HORN B., MORIARTY E.M., HODSON R., MOLTCHANOVA E. Evaluating statistical model performance in water quality prediction. Journal of Environmental Management. **206**, 910, **2018**.
30. LI P., HUI Q., WU J. Groundwater Quality Assessment Based on Improved Water Quality Index in Pengyang County, Ningxia, Northwest China. E-journal of Chemistry. **7**, S209, **2010**.
31. General Administration of Quality Supervision Inspection and Quarantine of China, Standardization Administration of China. Standards for groundwater quality (GB/T14848-2017). Beijing: Standards Press of China. **2017** [in Chinese].
32. LI P., HE X., GUO W. Spatial groundwater quality and potential health risks due to nitrate ingestion through drinking water: A case study in Yan'an City on the Loess Plateau of northwest China. Human and Ecological Risk Assessment: An International Journal. **25** (1-2), 11, **2019**.
33. ZHOU J.M., JIANG Z.C., XU G.L., QIN X.Q., HUANG Q.B., ZHANG L.K. Distribution and health risk assessment of metals in groundwater around iron mine. China Environmental Science. **39** (5), 1934, **2019**.
34. LI P., QIAN H. Human health risk assessment for chemical pollutants in drinking water source in Shizuishan city, northwest China. Iranian Journal of Environmental Health Science & Engineering. **8**, 41, **2011**.
35. ADIMALLA N., WU J. Groundwater quality and associated health risks in a semi-arid region of south India: Implication to sustainable groundwater management. Human and Ecological Risk Assessment: An International Journal. **25** (1-2), 191, **2019**.
36. HE S., WU J. Hydrogeochemical Characteristics, Groundwater Quality, and Health Risks from Hexavalent Chromium and Nitrate in Groundwater of Huanhe Formation in Wuqi County, Northwest China. Exposure and Health. **11** (2), 125, **2019**.
37. WU J., ZHANG Y., ZHOU H. Groundwater chemistry and groundwater quality index incorporating health risk weighting in Dingbian County, Ordos basin of northwest China. Geochemistry. **80** (4), 125607, **2020**.
38. WEI M., WU J., LI W., ZHANG Q., SU F., WANG Y. Groundwater Geochemistry and its Impacts on Groundwater Arsenic Enrichment, Variation, and Health Risks in Yongning County, Yinchuan Plain of Northwest China. Exposure and Health. **14** (2), 219, **2022**.
39. WANG Z.W., LIU W., XU Y.W., ZHOU S.Y., YANG F. Heavy metal pollution sources analysis and health risk assessment of groundwater in Lingjiang Basin. Resources Environment & Engineering. 1, **2025**.

40. YANG Q.H., ZUO W.Z., ZHU B. Analysis on the hydrogeologial characteristicsof luanhealluvial plain in luanxian-luannan section, Hebei Province. Geology and resources. **29** (4), 374, **2020**.
41. CHENG S.S. Evaluation of Water Resources Carrying Capacity in Plain Area of Tangshan City. Hebei GEO University, **2017**.
42. DU W.J. Development, utilisation and protection of Water resources in Luannan County. Science and Technology Innovation Herald. (6), 137, **2012**.
43. MOUSAZADEH R., GHAFFARZADEH H., NOURI J., GHARAGOZLOU A., FARAHPOUR M. Land use change detection and impact assessment in Anzali international coastal wetland using multi-temporal satellite images. Environmental Monitoring and Assessment. **187** (12), 776, **2015**.
44. HU J., WU Y., WANG L., SUN P., ZHAO F., JIN Z., WANG Y., QIU L., LIAN Y. Impacts of land-use conversions on the water cycle in a typical watershed in the southern Chinese Loess Plateau. Journal of Hydrology. **593**, 125741, **2021**.
45. FAKHRI Y., JAFARZADEH S., MORADI B., ZANDSALIMI Y., LANGARIZADEH G., AMIRHAJELOO L.R., MIRZAEI M. The Non-carcinogenic Risk of Cadmium in Bottled Water in Different Age Groups Humans: Bandar Abbas City, Iran. Mater Sociomed. **27** (1), 52, **2015**.
46. LI Y., LI P., CUI X., HE S. Groundwater quality, health risk, and major influencing factors in the lower Beiluo River watershed of northwest China. Human and Ecological Risk Assessment: An International Journal. **27** (7), 1987, **2021**.
47. SHAHRBABKI P.E., HAJIMOHAMMADI B., SHOEIBI S., ELMI M., YOUSEFZADEH A., CONTI G.O., FERRANTE M., AMIRAHMADI M., FAKHRI Y., MOUSAVI KHANEGHAH A. Probabilistic non-carcinogenic and carcinogenic risk assessments (Monte Carlo simulation method) of the measured acrylamide content in Tah-dig using QuEChERS extraction and UHPLC-MS/MS. Food and Chemical Toxicology. **118**, 361, **2018**.
48. LI P., QIAN H., HOWARD K.W.F., WU J., LYU X. Anthropogenic pollution and variability of manganese in alluvial sediments of the Yellow River, Ningxia, northwest China. Environmental Monitoring and Assessment. **186** (3), 1385, **2014**.
49. HUANG B., LI Z., CHEN Z., CHEN G., ZHANG C., HUANG J., NIE X., XIONG W., ZENG G. Study and health risk assessment of the occurrence of iron and manganese in groundwater at the terminal of the Xiangjiang River. Environmental Science and Pollution Research. **22** (24), 19912, **2015**.