ORIGINAL PAPER

Health-Risk Assessment of Arsenic and Groundwater Quality Classification Using Random Forest in the Yanchi Region of Northwest China

Chu Wu[1] · Chen Fang[2] · Xiong Wu[1] · Ge Zhu[1]

Abstract

High-arsenic (As) groundwater was first discovered in the Yanchi region, Northwest China, which is an arid or semiarid area, and the groundwater quality seriously affects the health of local residents. A comprehensive understanding of the spatiotemporal distribution characteristics, water quality, and health risk of high-As groundwater is indispensable for the sustainable utilization of groundwater sources and resident health. Seventy-nine groundwater samples were collected from different aquifers and seasons. The hazard quotient (HQ) and carcinogenic risk (CR) of As for adults and children were assessed. Moreover, the effects of groundwater sampling site and seasonal change on As concentration were investigated. Then, the random forest method was used to evaluate the importance of the indicators and the influence of these important indicators on groundwater classification. Thirty-three percent of the groundwater samples had HQ values > 1, and the CR values of all groundwater > 1.00×10^{-6} for children, representing a serious health risk. Twenty-one percent of the groundwater samples had health risk for adult. High-As groundwater is present at depths less than 60 m, and groundwater As concentrations are slightly affected by seasonal changes. The random forest shows that the most important indicators that affect groundwater quality are Na, TDS, TH, and F, and the least important is As. Furthermore, the optimal set of indicators contained all four of the most important indicators obtained by the random forest model, which achieved a classification accuracy of 88.21% for groundwater quality.

Keywords High arsenic · Health risk · Random forest · Groundwater quality

Introduction

Water resources in China are unevenly distributed over space and time, and the surface water is of poor quality and scarce in some arid or semiarid areas (Li et al. 2014; Wu et al. 2018a; Yan and Chen 2013). Therefore, groundwater plays an important role in irrigation and domestic purposes and is considered to be a source of uncontaminated water (Li et al. 2018; Qian et al. 2016; Wu et al. 2018b). However, extensive contamination has posed threats to groundwater quality, and the most common groundwater contaminants include pathogens (e.g., bacteria, viruses and protozoans) and some common solutes (e.g., nitrate, sulfate, Ca^{2+}, and Mg^{2+}) (Azizullah et al. 2011; Memon et al. 2011). Particularly, the highlighted reports with respect to arsenic (As) and fluoride (F) groundwater contamination represent one of the most serious and emerging human health challenges (Niazi et al. 2018; Nickson et al. 2004). Groundwater As and F concentrations are quite patchy, varying from drinking water standards to several hundred μg/L in wells that are several meters apart (Guo et al. 2012). Therefore, understanding the spatial variability of As is crucial not only for understanding the biogeochemical–hydrological processes that control the groundwater concentrations of As but also for developing strategies for the sustainable use of shallow groundwater that will reduce exposure to As (Farooqi 2015; Rasool et al. 2016). On the other hand, As ingestion from groundwater used as drinking water is a major source of As exposure to humans in arid or semiarid areas, including Northwest

✉ Xiong Wu
wuxiong@cugb.edu.cn

1 School of Water Resources and Environment, China University of Geosciences, Beijing 100083, People's Republic of China

2 Center for Advanced Technology and Education (CATE), Florida International University, Miami, FL 33174, USA

China (Wang et al. 2009, 2018; Guo et al. 2014). The long-term intake of high-As groundwater has negative effects on human health, resulting in residents suffering from chronic intoxication, including hematological, cardiovascular, neurological, respiratory, and renal diseases (Chappells et al. 2014; Shahid et al. 2018).

Hence, health-risk assessment of As in groundwater used as drinking water will be of great help to ensure the basic safety of drinking water for the increasing populations. The health-risk assessment of As was evaluated by Muhammad et al. (2010) and Sultana et al. (2014) in Pakistan, and the authors reported hazard quotient (HQ) values > 1, and more than 85% of drinking water had HQ values > 1, respectively. In the semiarid region of Northwest China, Li et al. (2016) revealed the groundwater's effect on human health, and the health-risk assessment suggested that females and children faced higher non-carcinogenic risks than males due to groundwater contaminants impacted by industrial and agricultural activities. Similarly, other scholars conducted health-risk assessments of drinking water to examine the adverse impacts of groundwater contaminants on human health in some areas (He and Wu 2019; He et al. 2019; Li and Qian 2011; Radfard et al. 2018; Ravindra and Mor 2019; Shahid et al. 2018; Sharifi et al. 2018; Wang et al. 2015; Wu et al. 2019). However, very limited research has been undertaken regarding the status of health-risk assessments for As in drinking water in the Yanchi region.

The Yanchi region is in an arid or semiarid region, and groundwater has been a major source of the water supply in this area, where there is a great shortage of water resources and almost no surface water resources are present. High-F groundwater has been found in the Yanchi area, and the sources of high F⁻ are the dissolution of fluorite and the desorption of exchangeable F⁻; additionally, the groundwater environment provides favorable conditions for F enrichment (Wu et al. 2018b). Moreover, Duan et al. (2016) revealed that most of the groundwater is of poor quality and is high in fluoride and salinity. Most research has focused on high-F groundwater and correlative groundwater quality issues. However, there is almost no research on high-As groundwater, groundwater quality assessments related to As or pollution source identification.

Random forest (RF) generates a large number of trees, which reduces generalization error due to the ensemble of permutations that can overcome the classification error of one permutation, and RF does not overfit the model to the dataset (Ghimire et al. 2010; Prasad et al. 2006). Therefore, the RF method can offer substantial improvements in classification accuracy and can be easily implemented for parallel computing, making it computationally efficient and a popular choice for data classification (Cutler et al. 2007; Hastie et al. 2009). Bindal and Singh (2019) used random forest to predict the regions in Uttar Pradesh at risk due to As contamination; the model showed an overall accuracy of 84.67% and performed better than other widely methods for As prediction. Tesoriero et al. (2017) used random forest classification to determine the probabilities of detecting elevated concentrations of nitrate, iron, and arsenic, and elevated nitrate, iron, and arsenic concentrations are a major risk to groundwater quality in the Fox-Wolf-Peshtigo study area. Similarly, random forest has been used extensively in hydrology for applications, such as identifying contaminant sources (Lee et al. 2018), classifying surface–groundwater interaction (Yang et al. 2019), modeling groundwater potential (Naghibi et al. 2017), and modeling groundwater level (Sahoo et al. 2017). Moreover, some studies have reported that RF produces very good results compared to other machine learning-based classification systems, such as support vector and neural networks or other decision tree algorithms (Baudron et al. 2013; Liaw and Wiener 2002).

In the study area, very limited research has been undertaken on the status of the spatial variability of As, the health-risk assessment of As and the effects on groundwater quality. In such cases, the presence of As in groundwater that is used as drinking water is applied to assess the potential health risks, including the HQ and CR for adults and children. Then, the RF method is used to classify groundwater quality and pollution source identification. Ultimately, the objectives of this study are to (1) investigate the distribution of As concentrations from shallow groundwater in time and space and assess the health risk posed by As to adults and children and (2) evaluate the importance of indicators on the groundwater classification task and further investigate the influence of these important indicators on the groundwater classification performance.

Materials and Methods

Study Area

The study area is located in the east of Yanchi County, Ningxia Hui Autonomous, and the Midwest of the Ordos Basin, northwest China (Fig. 1). It belongs to the transitional belt from the Ordos Platform to the north of Loess Plateau. The Yanchi County has many natural salt lakes, so it is also named 'Salt pond,' and there are two lakes—Beidachi and Gouchi, in the study area. Moreover, it is an extensive steppe and desert grassland from southeast to northwest, and it enjoys the well-known ningxia tan-sheep. In terms of the terrain topography, the terrain is high in the south and lows in the north, and the northern region is adjacent to the Mu Us Desert. The study area is a typical continental monsoon climate, cold in winter, and hot in summer with average temperature of 22.4 °C, temperature difference between day and night can reach 20 °C in autumn and winter handover. It
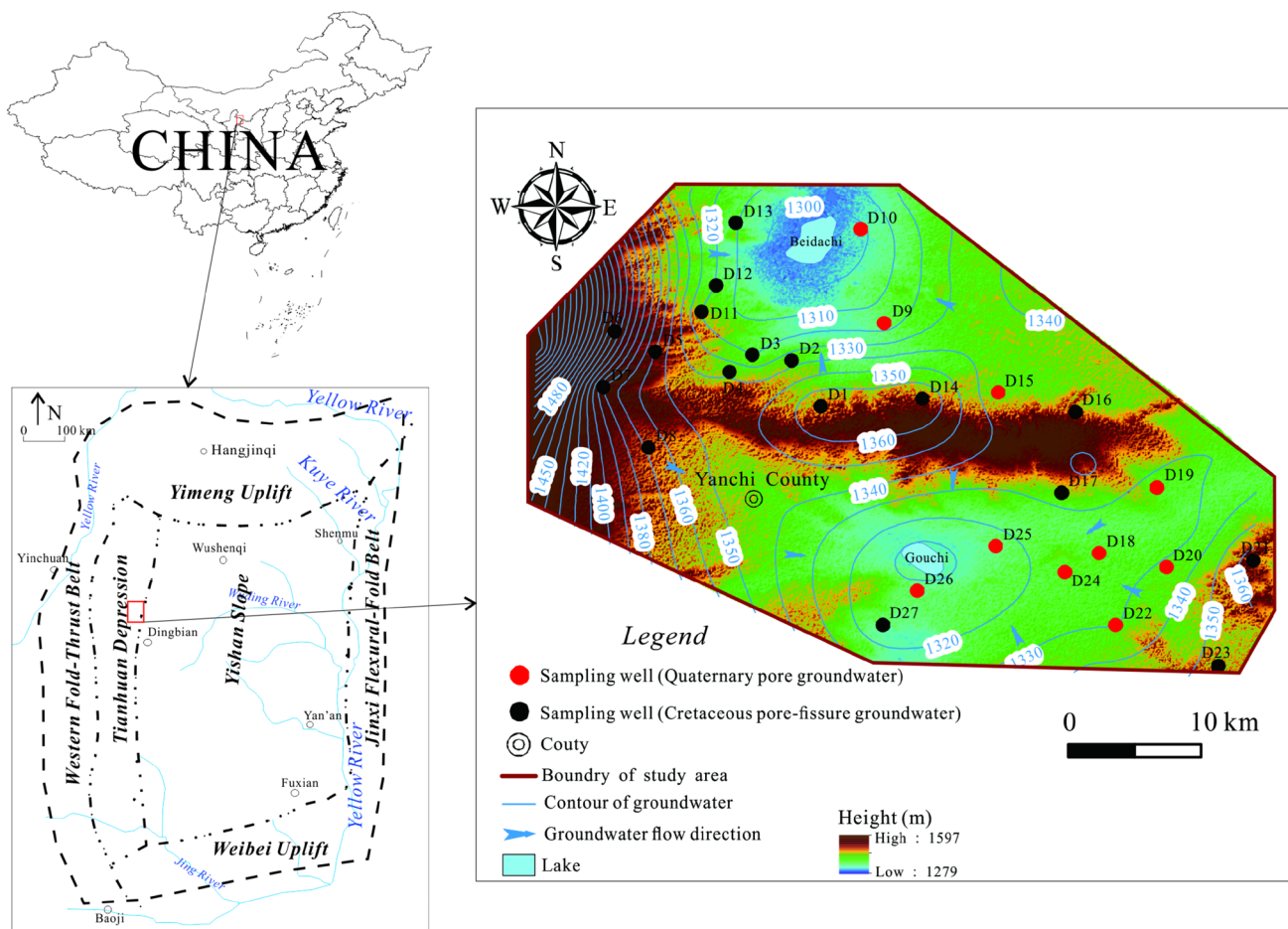
Fig. 1 Location and hydrogeological condition of the study area and sampling locations

belongs to an arid or semiarid climate with little precipitation and sufficient sunshine, the average annual evaporation of 2505 mm that is almost seven times the average annual precipitation of 375 mm.

In terms of the regional geology, the study area is located in the east of the Tianhuan Depression, Ordos Basin. It is mainly composed of Cretaceous strata-Huanhe and Luohandong formation, which consists of fine sandstones, mudstones, and sandstones and Quaternary unconsolidated alluvial and eluvium cover widely that comprises of sandy silt, clayey silt, silty sand, and silty fine sands (Duan et al. 2016; Wu et al. 2018b). The phreatic aquifer consists mainly of the lower Cretaceous and Quaternary aquifers, water-bearing formation of the Cretaceous-Luohe and Luohandong formation are extremely thick and widely distributed. However, Quaternary pore water is formed by an alluvial and eluvium stratum with variations of the thickness that has close hydraulic connection with underlying Cretaceous aquifer.

As shown in Fig. 1, it is observed that the terrain is high in the middle and low in the north and south, with an elevation ranging from 1279 to 1597 m. There is no perennial river in this study area, and the seasonal valley is confined to transient streams that run only in the rainy season. The drinkable surface water is scarce, Beidachi and Gouchi as salt water lake can't be directly drink. Hence, groundwater has played an important role in local inhabitants and livestock for various uses. In the study area, groundwater is mainly recharged by rainfall, accounting for 70% of the total recharge, while the rest comes from the lateral recharge of the high-lying region (Duan et al. 2016). According to the direction of groundwater flow (Fig. 1), the groundwater flows along the topography and drains into the Beidachi and Gouchi lakes, and groundwater is mainly discharged through evaporation, overflow to the surface water body, and artificial exploitation (Wu et al. 2018b).

Sample Collection and Analysis

As shown in Fig. 1, the location of the sampling wells, which were evenly distributed in the study area, is observed. Twenty-seven groundwater samples were collected from hand-pumped tube wells and motor-pumped wells in three

seasons, including dry season (January), normal season (April) and wet season (July). Among, eleven samples (*red circle* in Fig. 1) and sixteen samples (*black circle* in Fig. 1) were taken from Quaternary pore water (depths of approximately 10–50 m) and from Cretaceous pore-fissure water (depths of approximately 50–300 m), respectively. As a result, a total of 77 samples through 0.45-μm membrane filters were collected in clean and dry polyethylene plastic bottles, and one well was not sampled for some reasons in normal and wet seasons. Sample collection, handling, and storage followed the standard procedures recommended by the Chinese Ministry of Water Resources (Wu et al. 2018b). These groundwater samples used for major and trace element analysis were collected in 100 mL HNO_3 washed polyethylene plastic bottles, followed by the addition of 6 M HNO_3 to pH < 2, and the samples for the analysis of As species were preserved with 0.25 M EDTA (10%). We sent these samples to the Pony Testing International Group for the measurement of 8 parameters, such as Na^+ was determined using a flame photometer; SO_4^{2-} and Cl^- were analyzed using the ion chromatography method; and F^- and NO_3^- were determined by spectrophotometry. The concentrations of total dissolved As were determined by atomic fluorescence spectrometry with hydride generation (Guo et al. 2014). The accuracy of the water quality test was controlled by blank samples, parallel samples, and internal standard analysis, and the percentage of charge balance error (%CBE) was calculated to be less than 5%, suggesting that the accuracy of each index was within the quality requirements.

Risk Assessment Model

The models recommended by the Ministry of Environmental Protection of the P.R. China (MEPC 2014) are based on United States Environmental Protection Agency (USEPA) models. The models for non-carcinogenic risk through drinking water intake and dermal contact are as follows (MEPC 2014; Li et al. 2016). The non-carcinogenic risk through drinking water intake is expressed as (Li et al. 2019a, b):

$$\text{Intake}_{\text{oral}} = \frac{C \times IR \times EF \times ED}{BW \times AT}$$

$$HQ_{\text{oral}} = \frac{\text{Intake}_{\text{oral}}}{RfD_{\text{oral}}}$$

where $\text{Intake}_{\text{oral}}$ is the daily average exposure dosage through drinking water intake (oral pathway) per unit weight (ug/kg day). *C* is the contaminant concentration in groundwater (mg/L), as determined by laboratory analyses. *IR* is the ingestion rate of water (L/day), 1.5 L/day of water is selected for adults and 0.7 L/day is selected for children in this study (Li et al. 2016; Wu and Sun. 2016). *EF* and *ED* represent the exposure frequency (day/year) and exposure duration (average life expectancy of the P.R. China, year), respectively, for the non-carcinogenic risk assessment, EF is 365 days per year and ED is 30 years for adults and 12 years for children (Li et al. 2019a, b). *BW* and *AT* are the body weight (average body weight of the P.R. China, kg) and average time of life for non-carcinogenic pollutants (day), respectively. For children, the *BW* is 15 kg; for adults, the *BW* is 76 kg. The *AT* values for adults and children are 10,950 and 4380 days, respectively. On the other hand, HQ_{oral} and RfD_{oral} represent the hazard quotient and reference dosage, respectively. For non-carcinogenic pollutants introduced through the oral exposure pathway, the RfD_{oral} values for F and As are 0.04 and 0.0003 mg/kg day, respectively (MEPC 2014; Li et al. 2016; Wu and Sun 2016).

If the values of HQ > 1.00, the substance is considered unsafe for human health (MEPC 2014), and residents are exposed to a non-carcinogenic risk. As can also create a CR for humans, and the CR of As is calculated by using the following equation (MEPC 2014):

$$CR_{\text{oral}} = \text{Intake}_{\text{oral}} \times SF_{\text{oral}}$$

where CR denotes the carcinogenic risk, according to the Ministry of Environmental Protection of the P.R. China, the acceptable limit for CR is 1×10^{-6}. SF is the slope factor for the carcinogenic contaminants (mg/ kg day). The Chinese technical guidelines for risk assessments of contaminated sites (MEPC 2014) set the SF values of As at 1.5 mg/kg day. The exposure duration (*ED*) for CR assessments is set at 25,550 days for both adults and children, indicating that the adverse effects from As on human health will last a lifetime.

Results and Discussion

Human Health-Risk Assessment

We calculated HQ_{oral} and CR_{oral} based on the As concentrations of groundwater in the study area. Table 1 shows the calculated health risk to adults and children when they are exposed through drinking water intake in three seasons. For adults, the HQ from drinking water intake (HQ_{oral}) ranged from 0.01 to 4.83, with a mean of 0.92 in the wet season. In the normal and dry seasons, the HQ_{oral} ranged from 0.01 to 5.20 and 0.18 to 7.33, with mean values of 0.68 and 1.19, respectively. Meanwhile, the mean values of HQ_{oral} decreased slightly and then increased by 0.51 from the wet season to the dry season, indicating that the HQ_{oral} values were obviously affected by seasonal changes.

Based on the Ministry of Environmental Protection of the P.R. China, 6/27, 5/26, and 6/26 of groundwater samples had HQ_{oral} values exceeding 1 in three seasons, suggesting

Table 1 The HQ and CR values for adults and children in three seasons

Sample	Dry season				Normal season				Wet season			
	HQ_{oral}		CR_{oral}		HQ_{oral}		CR_{oral}		HQ_{oral}		CR_{oral}	
	Adult	Child	Adult	Child	Adult	Child	Adult	Child	Adult	Child	Adult	Child
A1	2.25	4.20	1.05E−03	4.90E−03	1.66	3.10	7.74E−04	3.61E−03	1.67	3.11	7.78E−04	3.63E−03
A2	4.83	9.02	2.26E−03	1.05E−02	5.20	9.71	2.43E−03	1.13E−02	4.58	8.56	2.14E−03	9.98E−03
A3	0.58	1.09	2.72E−04	1.27E−03	–	–	–	–	0.67	1.24	3.11E−04	1.45E−03
A4	2.50	4.67	1.17E−03	5.44E−03	0.02	0.03	7.78E−06	3.63E−05	7.33	13.69	3.42E−03	1.60E−02
A5	0.01	0.02	3.89E−06	1.81E−05	0.01	0.02	3.89E−06	1.81E−05	0.65	1.21	3.03E−04	1.42E−03
A6	0.01	0.02	3.89E−06	1.81E−05	0.01	0.02	3.89E−06	1.81E−05	–	–	–	–
A7	0.01	0.02	3.89E−06	1.81E−05	0.01	0.02	3.89E−06	1.81E−05	0.20	0.37	9.33E−05	4.36E−04
A8	2.58	4.82	1.21E−03	5.63E−03	0.01	0.02	3.89E−06	1.81E−05	0.22	0.40	1.01E−04	4.72E−04
A9	0.01	0.02	3.89E−06	1.81E−05	0.01	0.02	3.89E−06	1.81E−05	0.18	0.33	8.17E−05	3.81E−04
A10	3.67	6.84	1.71E−03	7.99E−03	3.38	6.30	1.58E−03	7.35E−03	0.32	0.59	1.48E−04	6.90E−04
A11	0.45	0.84	2.10E−04	9.80E−04	0.03	0.06	1.56E−05	7.26E−05	0.38	0.72	1.79E−04	8.35E−04
A12	0.45	0.84	2.10E−04	9.80E−04	0.06	0.11	2.72E−05	1.27E−04	0.20	0.37	9.33E−05	4.36E−04
A13	0.67	1.24	3.11E−04	1.45E−03	0.05	0.09	2.33E−05	1.09E−04	0.53	0.98	2.45E−04	1.14E−03
A14	0.53	1.00	2.49E−04	1.16E−03	4.72	8.80	2.20E−03	1.03E−02	5.08	9.49	2.37E−03	1.11E−02
A15	0.42	0.78	1.94E−04	9.07E−04	0.13	0.25	6.22E−05	2.90E−04	0.45	0.84	2.10E−04	9.80E−04
A16	0.23	0.44	1.09E−04	5.08E−04	0.01	0.02	3.89E−06	1.81E−05	0.83	1.56	3.89E−04	1.81E−03
A17	0.41	0.76	1.91E−04	8.89E−04	0.01	0.02	3.89E−06	1.81E−05	0.43	0.81	2.02E−04	9.44E−04
A18	0.38	0.70	1.75E−04	8.17E−04	0.01	0.02	3.89E−06	1.81E−05	0.42	0.78	1.94E−04	9.07E−04
A19	0.22	0.40	1.01E−04	4.72E−04	0.01	0.02	3.89E−06	1.81E−05	0.32	0.59	1.48E−04	6.90E−04
A20	0.18	0.34	8.56E−05	3.99E−04	0.07	0.12	3.11E−05	1.45E−04	0.53	1.00	2.49E−04	1.16E−03
A21	0.77	1.43	3.58E−04	1.67E−03	0.01	0.02	3.89E−06	1.81E−05	0.92	1.71	4.28E−04	2.00E−03
A22	0.28	0.53	1.32E−04	6.17E−04	0.08	0.16	3.89E−05	1.81E−04	0.33	0.62	1.56E−04	7.26E−04
A23	0.17	0.31	7.78E−05	3.63E−04	0.01	0.02	3.89E−06	1.81E−05	1.25	2.33	5.83E−04	2.72E−03
A24	0.20	0.37	9.33E−05	4.36E−04	0.01	0.02	3.89E−06	1.81E−05	0.38	0.72	1.79E−04	8.35E−04
A25	0.27	0.50	1.24E−04	5.81E−04	0.01	0.02	3.89E−06	1.81E−05	0.83	1.56	3.89E−04	1.81E−03
A26	2.33	4.36	1.09E−03	5.08E−03	2.10	3.92	9.80E−04	4.57E−03	1.83	3.42	8.56E−04	3.99E−03
A27	0.43	0.79	1.98E−04	9.26E−04	0.01	0.02	3.89E−06	1.81E−05	0.52	0.96	2.41E−04	1.13E−03
Max	4.83	9.02	2.26E−03	1.05E−02	5.20	9.71	2.43E−03	1.13E−02	7.33	13.69	3.42E−03	1.60E−02
Min	0.01	0.02	3.89E−06	1.81E−05	0.01	0.02	3.89E−06	1.81E−05	0.18	0.33	8.17E−05	3.81E−04
Mean	0.92	1.72	4.29E−04	2.00E−03	0.68	1.26	3.16E−04	1.47E−03	1.19	2.23	5.57E−04	2.60E-03

that some samples may induce non-carcinogenic risks to adults if they drink the contaminated groundwater. Similarly, the HQ_{oral} values for children ranged from 0.02 to 9.02, 0.02 to 0.91, and 0.33 to 13.69, with mean values of 1.72, 1.26, and 2.23 in the three seasons, respectively. As a result, children have smaller body weights and lower average life spans than adults due to non-carcinogenic risks; the HQ_{oral} values of all samples for children exceeded those of adults in three seasons; children in the study area faced higher non-carcinogenic risks.

Compared with the non-carcinogenic risk, as shown in Table 1, the CR created by drinking water (CR_{oral}) is significantly lower for adults and children. In the wet season, the CR_{oral} values of adults and children ranged from 3.89×10^{-6} to 2.26×10^{-3} and 1.81×10^{-5} to 1.05×10^{-2}, with averages

of 4.29×10^{-4} and 2.00×10^{-3}, respectively. The CR_{oral} values changed with seasons in ways that were similar to those of the HQ_{oral} values; the mean CR_{oral} values were 3.16×10^{-4} and 5.57×10^{-4} for adults and 1.47×10^{-3} and 2.60×10^{-3} for children in the normal and dry seasons, respectively. In the study area, these values were higher than the acceptable limit (1.00×10^{-6}) recommended by the MEPC (2014), and these results are consistent with those of other studies conducted in homogeneous areas in China (Li and Qian 2011; Li et al. 2014) Li et al (2016). revealed the carcinogenic risk created by drinking the water (CRoral) were higher than the acceptable limit with an average of $5.93 \times 10-5$ in a is a part of the Weining Plain in northwest China.

Coal and oil were important resource in the study area, it focused on the development of petrochemical industry, and

several gas fields had been put into production. To reduce the health risks induced by ingestion of the contaminated groundwater, local Chinese governments have to take action to control and eliminate groundwater contamination; it is wise to pursue additional water for domestic uses. The As concentration of groundwater with burial depth > 100 m was less than 10 µg/L, local government established new water collection and transition networks to supply deep-layer groundwater to local residents for drinking purpose. Furthermore, human activities such as the disposal of sewage and animal wastes should also be regulated, preventing further deterioration of shallow groundwater (depths of approximately < 60 m). The capital construction of gas field should be strictly implemented in accordance with national standards and industrial effluents from it must be purified to national standards, groundwater pollution in this area can be effectively controlled. In the study area, groundwater had been contaminated may be due to the lack of public awareness on groundwater protection and ignorance of groundwater research, it is necessary to first establish groundwater-monitoring networks, and cleaning shallow groundwater was also recommended.

Effect of Groundwater Sampling Depth on As Concentration.

Groundwater sampling depth is a crucial factor affecting the As concentration in groundwater (Ahmad et al. 2013; Shahid et al. 2018). Guo et al. (2012) observed that high-As groundwater is present at depths greater than 15 m and is found in deep aquifers (40–100 m), which are currently used as domestic water supplies and for irrigation in the Hetao Basin of China. Furthermore, the mean As concentration (5.2 µg/L) of shallow wells (9–40 m) does not exceed the WHO safe limit of As in Punjab, Pakistan, indicating a safe aquifer zone for pumping groundwater; in contrast, the deeper wells at 41–90 m and > 90 m depth had mean As concentrations of 51 and 123 µg/L, respectively (Shakoor et al. 2018). Wang et al. (2018) reported that the As concentrations of unconfined groundwater samples were all < 10 µg/L (WHO), with an average of 1.99 µg/L. The As concentrations in confined groundwater ranged between 9.9 and 377 µg/L (average 109 µg/L), with the highest concentration (377 µg/L) reported at a depth of 221 m in the Guide basin, Northwest China.

Figure 2 shows that the groundwater As concentration did not obviously vary with sampling depth in the three seasons. In the study area, 12 samples were collected at water table depths less than 20 m, 5 samples were collected at depths between 20 and 60 m, and 10 samples were collected at depths of more than 60 m. Generally, As concentration was dependent on depth. For the samples collected at a depth of more than 60 m, the As
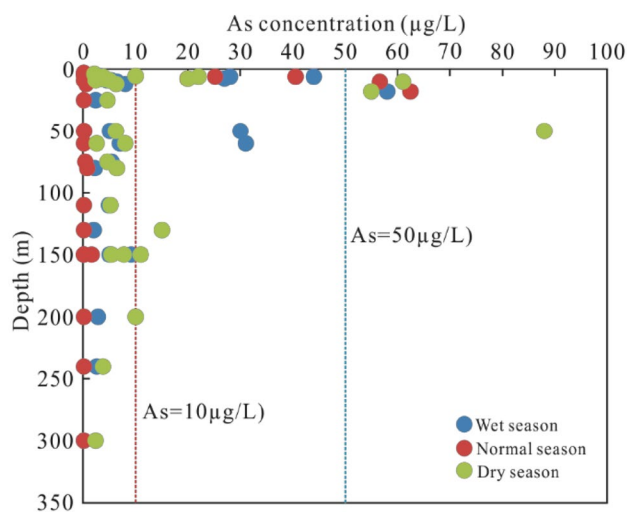

Fig. 2 The relationship between As concentration and sampling depth

concentration was ≤ 10 µg/L in the three seasons, except for one sample (A23) at a depth of 130 m, which had an As concentration of 15 µg/L. Of the samples collected from depths of less than 60 m, the As concentrations were greater than 10 µg/L: the As concentrations of 11 samples were between 10 and 50 µg/L. Moreover, one sample (A2) from a depth of 18 m had an As concentration of 58 µg/L in the dry season, and two and three samples had an As concentration greater than 50 µg/L in the normal and dry seasons, respectively.

The As concentrations of six samples (A1, A2, A4, A8, A10, and A26) were > 10 µg/L, and the depth of the water table was < 60 m. Among the samples, the A10 and A26 sampling wells belonged to the Quaternary pore groundwater and were located in the discharge area near Beida and Gouchi lakes, respectively. The other four groundwater samples came from Cretaceous pore-fissure groundwater, and the sampling wells were located in the mountain front with higher altitudes. Similarly, Sanaullah et al. (2016) and Toor and Tahir (2009) revealed that there was no significant relationship between As concentration and sampling depth. In the study area, it is found that there is no such relation exists; it may be resulted due to the dissolution of solutes from different sources and later its leakage towards ground water.

Moreover, Guo et al. (2012) reported that large variations in As concentration are common over short distances. Similarly, there was a low As concentration in well A3 from a depth of 60 m (7.0 µg/L), while the neighboring A4 groundwater samples from a depth of 50 m contained 30.0 µg/L As in the study area. Overall, in the normal and dry seasons, the relationship between As contents and sampling depths was similar to that in the wet season. Therefore, it is likely that high-As groundwater is present at depths less than 60 m, and this range is currently used for domestic water supplies

and for irrigation; thus, additional research must focus on these depths.

Effect of Seasonal Variation on As Concentrations

The groundwater As concentration might also vary with season; relatively higher As contents are detected in groundwater during the pre-monsoon period compared to those during the post-monsoon season; which was attributed to increases in groundwater recharge due to rainfall (Zubair et al. 2014). Similarly, Savarimuthu et al. (2006) observed seasonal variation in As concentration in West Bengal, with the maximum As concentration in the monsoon season and the minimum As concentration in the dry season. However, some authors reported mixed variation in As content during different seasons. In a shallow aquifer of the Bengal Delta, the groundwater As concentration increased and decreased over an 8-year (2002–2010) observation period (McArthur et al. 2010). Moreover, in Bangladesh, one out of 20 wells showed higher As concentrations during the rainy season (Savarimuthu et al. 2006).

Figure 3 shows the spatial distribution of As concentration in the three seasons. D10 and D26 were located in groundwater discharge area and had higher As concentration in the wet season, however, their As concentration were less than 10 μg/L in the normal and dry seasons. Correspondingly, As concentration of D13 and D25 near the lakes became higher from wet season to dry season. Other groundwater samples with higher As concentration were basically located in recharge area. In the dry season, D20 and D22 were located in groundwater runoff area and had higher As concentration. There weren't obviously relationship between As concentration and sampling point for all groundwater samples, suggesting that higher As concentration may be due to human activities such as the disposal of sewage and animal wastes (Wu and Sun 2016).

Figure 4 shows that As concentration varied with season, 77% of all samples had a positive difference from the dry to normal seasons, indicating that the groundwater As concentration decreased. Particularly, the As concentration of one sample (A14) ranged from 6.4 to 56.6 μg/L, and that sample was located in a recharge area with a higher altitude (Fig. 1),
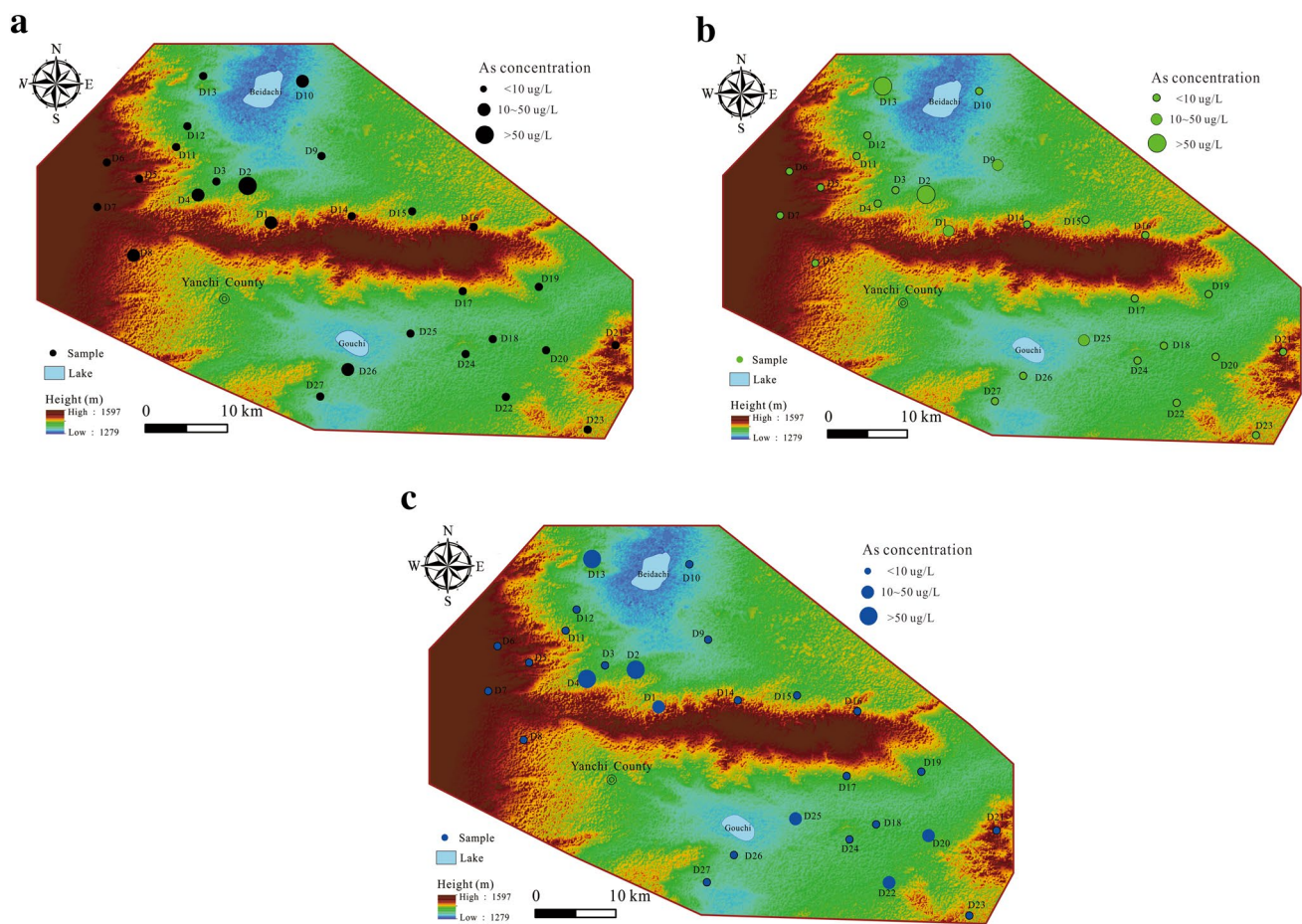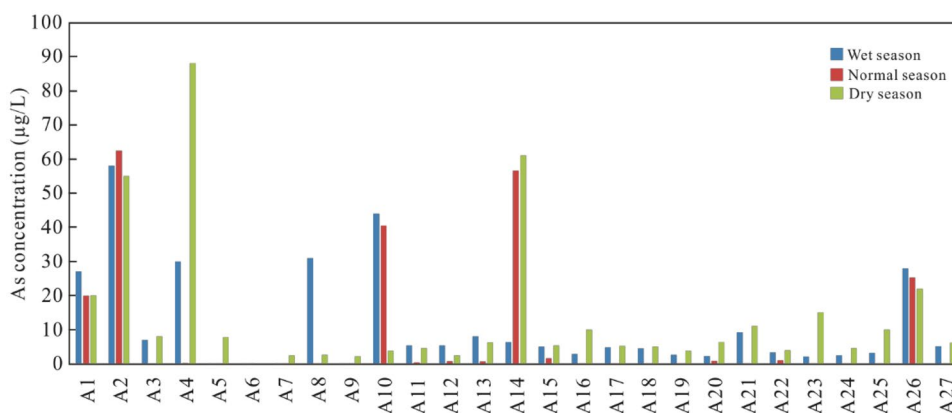

Fig. 3 Spatial distribution of As concentration in three seasons (**a** wet season, **b** normal season, **c** dry season)

Fig. 4 Groundwater As concentration in three seasons

and the buried depth of the sampling well was approximately 10 m. Moreover, the groundwater As concentration was reduced by 29.8 and 30.9 µg/L for the two samples of A4 and A8, respectively. The difference in As concentration in other samples was less than 10 µg/L, suggesting that seasonal variation was not obvious.

On the other hand, the As concentration of most samples increased from the normal season to the wet season, except for three samples with negative differences in As concentration. The A4 samples were located in the groundwater runoff area, and the buried depth of that sampling well was approximately 50 m. The concentration of groundwater As increased by 87.8 µg/L. Conversely, the As concentration decreased by 36.7 µg/L for the A10 samples, which were located in the discharge area near Beida Lake. The concentrations of other samples varied within 10 µg/L from the normal season to the wet season. As a result, seasonal variation had a slight effect on As concentration, and there was mixed variation in the As content from the dry season to the normal season to the wet season due to the local dissolution of Fe oxy-hydroxides (Nicolli et al. 2010).

Some studies have reported that the high concentration of bicarbonate in most groundwater samples can trigger As releases in groundwater and mobilize As from sediments in both oxic and anoxic conditions (Anawar et al. 2004; Farooqi 2015; Rasool et al. 2016). However, it has been reported that As releases are not always closely correlated with bicarbonate content and vary greatly with the presence of cations and redox conditions; furthermore, not all studies agree with

the bicarbonate-mediated release of As in groundwater (Birkle and Alvarado 2010). Moreover, the combined role of pH and bicarbonate content may exert a significant effect on As leaching from the sediments (Zubair et al. 2014). In the study area, the groundwater As concentrations of the three samples (A4, A10 and A14) had a larger change, which did not occur in the corresponding concentrations of bicarbonate and pH values with seasonal variation (Table 3).

Significantly, groundwater contamination by leaching from pesticides and fertilizers is unlikely to occur in many areas with low rainfall, high evapotranspiration and strong As binding with soil minerals in alkaline soils, and the use of agrochemicals may result in As releases to surface and/or shallow groundwater (Shahid et al. 2018). Thus, the As concentrations of the three samples varied abnormally with seasonal variation, which might be related to the leaching of the pesticides and fertilizers from soils to groundwater in the study area.

Random Forest Model

The comprehensive evaluation of groundwater quality was determined by the highest classification for the single index evaluation, as recommended by the Quality Standard of Groundwater (QSGW 2015), and the groundwater quality was divided into five categories (Table 2). Moreover, if the Cl^- and F^- concentrations in groundwater are 400 mg/L and 2.5 mg/L, respectively, these two indicators fall into five categories, and the other indicators have fewer than five

Table 2 Classification of the QSGW

TH (mg/L)	TDS (mg/L)	SO$_4$$^{2-}$ (mg/L)	Cl$^-$ (mg/L)	Na$^+$ (mg/L)	NO$_3$$^-$ (mg/L)	F$^-$ (mg/L)	As (mg/L)	Class
≤150	≤300	≤50	≤50	≤100	≤2	≤1	≤0.001	1
150–300	300–500	50–150	50–150	100–150	2–5	≤1	≤0.001	2
300–450	500–1000	150–250	150–250	150–200	2–5	≤1	≤0.01	3
450–650	1000–2000	250–350	250–350	200–400	20–30	1–2	0.01–0.05	4
>650	>2000	>350	>350	>400	>30	>2	>0.05	5

Table 3 Classification results of the QSGW and random forest

Sample	TH (mg/L)	TDS (mg/L)	SO_4^{2-} (mg/L)	Cl^- (mg/L)	Na^+ (mg/L)	NO_3^- (mg/L)	F^- (mg/L)	As (mg/L)	pH (mg/L)	Class of QSGW	Predicted class
D1-D	474	2700	543	714	889	91.2	3.2	0.027	8.49	4	5
D2-D	192	1800	471	283	618	22.8	7.36	0.058	8.44	5	5
D3-D	673	2670	760	595	681	4.26	3.11	0.007	7.91	5	5
D4-D	135	529	114	67.9	135	9.82	1.45	0.03	8.3	4	4
D5-D	282	1300	344	300	377	7.01	2.7	0.0001	8.25	4	5
D6-D	367	1200	436	277	309	3.54	3.12	0.0001	8.14	4	4
D7-D	516	1800	570	451	444	5.04	2.01	0.0001	8.31	4	5
D8-D	1270	3700	1680	746	711	5.85	2.12	0.031	8	5	5
D9-D	666	1300	213	298	265	27.3	1.01	0.0001	8.28	4	4
D10-D	26.2	700	259	71.8	220	0.04	1.59	0.044	8.48	4	4
D11-D	724	2590	766	725	631	19.7	2.18	0.0054	8.49	5	5
D12-D	270	1200	337	225	338	13.5	4.84	0.0054	8.14	4	4
D13-D	818	2500	882	621	541	15.9	2.41	0.008	8.29	5	5
D14-D	201	2000	578	290	672	12.8	6.42	0.0064	8.42	5	5
D15-D	248	1900	509	366	604	9.24	2.02	0.005	8.22	5	5
D16-D	205	990	278	182	286	2.88	1.54	0.0028	8.29	4	4
D17-D	401	2500	824	585	744	7.64	3.05	0.0049	8.28	5	5
D18-D	147	1500	286	253	545	19.9	9.41	0.0045	8.42	5	5
D19-D	319	1440	450	293	389	1.31	2.65	0.0026	8.17	5	5
D20-D	327	1260	259	275	363	7.65	4.77	0.0022	8.1	4	4
D21-D	183	1460	348	301	487	7.35	4.91	0.0092	8.29	5	5
D22-D	355	2200	410	456	721	14.7	4.42	0.0034	8.03	5	5
D23-D	135	640	92.6	52.6	190	21.7	3.63	0.002	8.22	4	5
D24-D	289	1200	244	223	367	6.4	4.23	0.0024	8.14	4	4
D25-D	972	4100	1140	1180	1070	37.4	2.99	0.0032	8.21	5	5
D26-D	301	1200	167	329	346	9.22	2.57	0.028	8.22	4	4
D27-D	170	1060	261	149	364	13.3	3.91	0.0051	8.07	5	5
D1-N	416	3060	504	664	829	88.9	2.98	0.0199	8.21	5	5
D2-N	163	1940	466	280	600	25	5.87	0.0624	8.59	5	5
D4-N	867	3170	760	732	733	32.2	1.5	0.0002	7.91	5	5
D5-N	245	1400	331	286	382	10.7	2.49	0.0001	8.05	5	5
D6-N	406	1820	586	355	434	5.38	1.76	0.0001	8.16	5	5
D7-N	449	2010	592	461	471	6.31	2.26	0.0001	8.03	5	4
D8-N	1330	3810	1550	703	698	5.58	2.29	0.0001	7.82	5	5

Table 3 (continued)

Sample	TH (mg/L)	TDS (mg/L)	SO_4^{2-} (mg/L)	Cl^- (mg/L)	Na^+ (mg/L)	NO_3^- (mg/L)	F^- (mg/L)	As (mg/L)	pH (mg/L)	Class of QSGW	Predicted class
D9-N	509	1380	201	266	260	27.4	1.28	0.0001	7.85	4	4
D10-N	22.3	782	259	66.2	268	1	1.69	0.0405	8.77	4	4
D11-N	1300	4180	1200	1110	844	30.6	2.91	0.0004	8.14	5	5
D12-N	244	1360	346	219	369	15.1	3.94	0.0007	8.1	5	5
D13-N	716	2860	926	648	583	18.8	2.06	0.0006	8.14	5	5
D14-N	180	2270	590	294	709	14.2	5.66	0.0566	8.32	5	5
D15-N	214	2040	511	390	593	9.08	2.31	0.0016	8.13	5	5
D16-N	131	685	221	90	148	2.29	1	0.0001	8.3	3	–
D17-N	389	3060	945	658	820	9.58	3.22	0.0001	8.16	5	5
D18-N	143	1650	286	250	522	22.3	8.84	0.0001	8.57	5	5
D19-N	268	2510	401	371	876	10.8	9.01	0.0001	8.24	5	5
D20-N	209	1290	247	236	363	16.8	2.66	0.0008	8.06	5	5
D21-N	171	1680	368	320	524	8.75	4.37	0.0001	8.35	5	5
D22-N	270	2500	410	404	805	18	3.88	0.001	8.2	5	5
D23-N	94.4	485	51	20.5	127	12.5	1.56	0.0001	8.21	3	–
D24-N	279	1550	288	274	443	5.85	4.56	0.0001	8.12	5	5
D25-N	861	4790	1230	1300	1130	39.9	2.56	0.0001	8.16	5	5
D26-N	246	1220	172	270	322	8.92	2.11	0.0252	8.09	5	4
D27-N	154	1270	290	163	381	15.2	3.27	0.0001	8.03	5	5
D1-W	114	996	184	78.6	331	9.9	3.81	0.02	8.45	5	4
D2-W	187	2020	517	298	628	27.7	6.49	0.055	8.44	5	5
D3-W	683	2890	730	519	749	6.48	2.12	0.008	8.45	5	5
D4-W	829	3050	732	639	714	29.1	1.17	0.088	8.16	5	5
D5-W	284	1430	339	272	382	6.98	2.51	0.0078	8.38	5	5
D7-W	449	1740	497	333	396	4.93	1.73	0.0024	8.27	4	5
D8-W	1850	4980	2300	853	1080	12.4	1.77	0.0026	8.67	5	5
D9-W	695	1630	256	311	273	30.2	0.48	0.0021	8.1	4	4
D10-W	25.9	791	267	76.8	276	0.04	0.98	0.0038	8.64	4	4
D11-W	688	2570	651	564	581	19.4	2.59	0.0046	8.26	5	5
D12-W	139	346	59	33.5	59.8	10.9	0.45	0.0024	8.13	2	–
D13-W	860	2710	844	544	577	16.2	1.54	0.0063	8.22	5	5
D14-W	173	2710	561	256	696	15.9	6.29	0.061	8.43	5	5
D15-W	240	2100	524	384	622	11.1	2.3	0.0054	8.23	5	5
D16-W	210	1040	291	168	280	16.7	1.34	0.01	8.44	4	4

Table 3 (continued)

Sample	TH (mg/L)	TDS (mg/L)	SO_4^{2-} (mg/L)	Cl^- (mg/L)	Na^+ (mg/L)	NO_3^- (mg/L)	F^- (mg/L)	As (mg/L)	pH (mg/L)	Class of QSGW	Predicted class
D17-W	280	1850	564	347	531	7.48	1.98	0.0052	8.37	5	5
D18-W	173	1590	285	245	495	22.8	9.06	0.005	8.42	5	5
D19-W	351	2980	962	415	914	11	9.73	0.0038	8.37	5	5
D20-W	239	1310	265	242	347	17.7	1.72	0.0064	8.43	4	4
D21-W	188	1700	389	330	516	9.35	3.3	0.011	8.56	5	5
D22-W	252	2280	381	374	723	22.7	3.86	0.004	8.32	5	5
D23-W	89.2	571	71.1	34.6	171	14	2.48	0.015	8.49	5	4
D24-W	775	2910	758	654	673	5.16	1.9	0.0046	8.14	5	5
D25-W	1360	5340	1460	1470	1310	27.6	1.26	0.01	8.33	5	5
D26-W	405	1830	210	462	419	12.8	1.25	0.022	8.11	5	5
D27-W	166	1230	306	132	364	15.7	2.54	0.0062	8.45	5	5

A1-D represent sample in dry season, A1-N is in normal season and A1-D is in wet season

categories. The comprehensive evaluation of groundwater quality has five categories according to the QSGW. Based on the above method, 79 groundwater samples were classified in the study area, as shown in Table 3. The groundwater quality was very poor, possibly due to low rainfall, high evapotranspiration, and the underground environment.

To determine which indicators are the most important in influencing the groundwater quality of the study area, a random forest algorithm is utilized to classify the samples and to evaluate the importance of features/indicators. Scikit-learn was used to implement the random forest algorithm with cross-validation procedure and to build the experiment pipeline (Pedregosa et al. 2012). It should be noted that, since the goal of this study is to take advantage of the random forest algorithm for the feature importance evaluation of groundwater quality indicators, the detailed formulation is not included, and additionally, in contrast to the original method (Breiman 2001), which lets each classifier vote for a single class, the scikit-learn implementation used in this study combines classifiers by averaging their probabilistic prediction instead. Random forest is one of the most widely used ensemble learning methods for classification, regression, and other tasks and is based on randomized decision trees. Moreover, because this approach is based on tree structures, the depth of a feature used as a decision node in a tree can be processed as the relative rank to assess the importance of the corresponding indicator with respect to the predictability of the target variable. As shown in Table 3, overall, there are eight indicators involved in the groundwater samples, and among all samples, only three of the samples (A16-N, A23-N, and A12-D) are not in category 4 or category 5; therefore, these three samples were removed from the dataset to simplify the groundwater classification task as a binary classification problem.

Indicator Importance Analysis

To evaluate the importance of indicators on the groundwater classification task, as mentioned above, the remaining 76 samples were used to fit a random forest model. The results are demonstrated in Fig. 5, where the red bars represent the feature importance of the forest as well as their inter-tree variability. As shown in Fig. 5, the four indicators of Na, TDS, TH, and F have an importance in the forest greater than 0.1. These indicators are deemed the most important indicators that affect the groundwater quality of the study area. Moreover, the effects of the four indicators on the overall groundwater quality decrease in the following order: Na > TDS > TH > F.

On the other hand, the importance of the other indicators was less than 0.1, and As was the least important indicator that affected the groundwater quality. In the study area, the poor groundwater quality is not caused only by the high
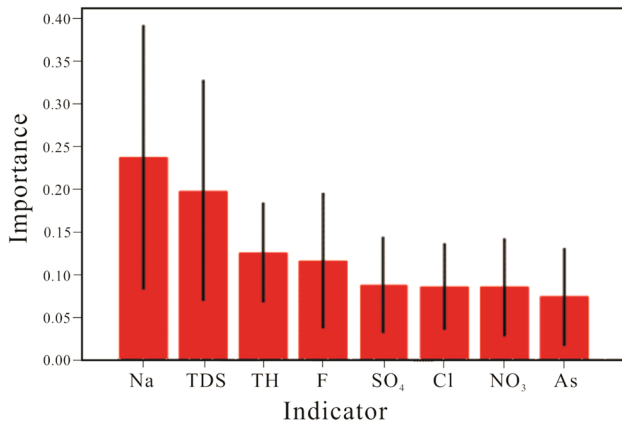
Fig. 5 The indicator importance analysis with random forest

Fig. 6 The optimal indicators on the groundwater classification

content of some conventional indicators in groundwater; rather, fluorine had a great contribution. The long-term intake of this poor-quality groundwater has negative effects on resident health, such as suffering from chronic intoxication due to fluorine, and local residents and management departments must give this problem appropriate attention. In addition, the other indicator causing chronic intoxication is As; although the overall impact on groundwater quality is small, it should be taken seriously in some areas with high-As groundwater.

Given the relative rank of all indicators, we can further investigate the influence of these important indicators on the groundwater classification performance. The goal is to select features by incrementally considering larger and larger sets of features. First, the estimator is trained on the most important feature (i.e., Na). Then, based on the importance of each feature obtained by the random forest model, the second most important feature is added to the set of features. This procedure is repeated until all features are selected and the optimal set of features with the best classification performance is eventually found.

The classification performance is measured by using the accuracy based on a ten-fold cross-validation process. As mentioned above, a binary classification (i.e., category 4 vs. category 5) experiment is carried out to select the optimal set of indicators. In the ten-fold cross-validation process, the sample water data points are randomly assigned to 10 sets, d_0, $d1$,…, $d9$, so that all sets are roughly equal in size. Then, each of the ten sets is retained as the validation data, while the remaining 9 sets are used for training; thus, every data point was used for both training and validation on each fold. It should be noted that for every single validation, the random forest classifier is trained separately to ensure the reliability of the experimental results. As shown in Fig. 6, the optimal set of indicators contained all four of the most important indicators (i.e., Na, TDS, TH, and F) obtained by
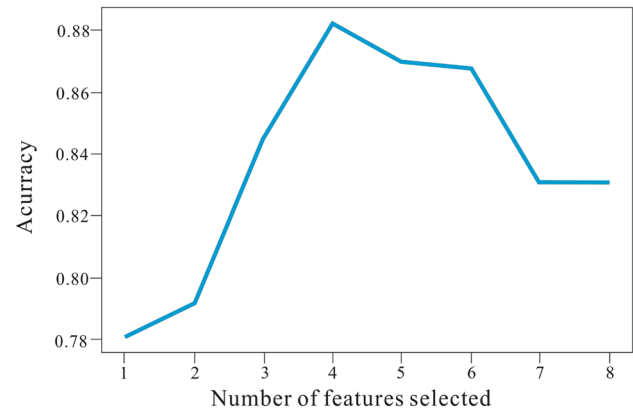
the random forest model, which achieved a classification accuracy of 88.21%. This result indicates that the groundwater quality was primarily determined by Na, TDS, TH, and F based on the collected samples, and these four indicators significantly reduced the groundwater quality in the study area.

Conclusions

In this study, we investigated the distribution characteristics of high-As groundwater in time and space, and an attempt was made to assess the health risk of As. The random forest model was used to simplify the groundwater classification task as a binary classification problem, to evaluate the importance of indicators on the groundwater classification task, and to investigate the influence of these important indicators on the groundwater classification performance. We obtained the following main conclusions from this study:

(1) Groundwater As concentration did not obviously vary with sampling depth in three seasons, and high-As groundwater is present at depths less than 60 m.

(2) Seasonal variation had a slight effect on the As concentration, and the groundwater As concentration varied abnormally with seasonal variation, which might be related to human activity.

(3) The HQ and CR values of all groundwater samples for children exceeded those of adults in three seasons: 33% samples had HQ > 1, and the CR values of all samples were higher than the acceptable limit, indicating a higher non-carcinogenic risk and a serious carcinogenic health risk.

(4) The most important indicators that affected the groundwater quality were Na, TDS, TH, and F; when combined, this optimal set of indicators achieved a classification accuracy of 88.21% for groundwater quality.

References

Ahmad SA, Gulzar A, Rehman HU, Soomro ZA, Hussain M, Rehman M, Qadir MA (2013) Study of arsenic in drinking water of Distric Kasur Pakistan. World Appl Sci J 24(5): 634–640. https://doi.org/10.58129/idosi.wasj.2013.24.05.13228

Anawar HM, Akai J, Sakugawa H (2004) Mobilization of arsenic from subsurface sediments by effect of bicarbonate ions in groundwater. Chemosphere 54(6):753–762. https://doi.org/10.1016/j.chemosphere.2003.08.030

Azizullah A, Khattak MNK, Richter P, Hader DP (2011) Water pollution in Pakistan and its impact on public health-a review. Environ Int 37(2):479–497. https://doi.org/10.1016/j.envint.2010.10.007

Baudron P, Alonso-Sarría F, García-Aróstegui JL, Cánovas-García F, Martínez-Vicente D, Moreno-Brotóns J (2013) Identifying the origin of groundwater samples in a multi-layer aquifer system with random forest classification. J Hydrol 499:303–315. https://doi.org/10.1016/j.jhydrol.2013.07.009

Bindal S, Singh CK (2019) Predicting groundwater arsenic contamination: regions at risk in highest populated state of India. Water Res 159:65–76. https://doi.org/10.1016/j.watres.2019.04.054

Birkle P, Alvarado IST (2010) Water-rock interaction XIII. CRC Press, Boca Raton

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Chappells H, Parker L, Fernandez CV, Conrad C, Drage JO, Toole G, Campbell N, Dummer TJ (2014) Arsenic in private drinking water wells: an assessment of jurisdictional regulations and guidelines for risk remediation in North America. J Water Health. 12(3):372–392. https://doi.org/10.2166/wh.2014.054

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JL (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

Duan L, Wang W, Zhou L, Cheng Z (2016) The formation of shallow fresh groundwater in the north of Yanchi County, Ningxia, China: main influencing factors and mechanism. Environ Earth Sci 75(6):461. https://doi.org/10.1007/s12665-016-5333-y

Farooqi A (2015) Arsenic and Fluoride contamination: a Pakistan perspective. Springer, New York

Ghimire B, Rogan J, Miller J (2010) Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. Remote Sens Lett 1(1):45–54. https://doi.org/10.1080/01431160903252327

Guo HM, Zhang Y, Xing LN, Jia YF (2012) Spatial variation in arsenic and fluoride concentrations of shallow groundwater from the town of Shahai in the Hetao basin, Inner Mongolia. Appl Geochem 27(11):2187–2196. https://doi.org/10.1016/j.apgeochem.2012.01.016

Guo HM, Zhang D, Wen DG, Wu Y, Ni P, Jiang YX, Guo Q, Li FL, Zheng H, Zhou YZ (2014) Arsenic mobilization in aquifers of the southwest Songnen basin, P.R. China: evidences from chemical and isotopic characteristics. Sci Total Environ 490:590–602. https://doi.org/10.1016/j.scitotenv.2014.05.050

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer, p 745, ISBN: 978-0-387-84857-0

He S, Wu J (2019) Hydrogeochemical characteristics, groundwater quality and health risks from hexavalent chromium and nitrate in groundwater of Huanhe Formation in Wuqi County, northwest China. Expo Health 11(2):125–137. https://doi.org/10.1007/s12403-018-0289-7

He X, Wu J, He S (2019) Hydrochemical characteristics and quality evaluation of groundwater in terms of health risks in Luohe aquifer in Wuqi County of the Chinese Loess Plateau, northwest China. Hum Ecol Risk Assess 25(1–2):32–51. https://doi.org/10.1080/10807039.2018.1531693

Lee YJ, Park C, Lee ML (2018) Identification of a contaminant source location in a river system using random forest models. Water 10(4):391. https://doi.org/10.3390/w10040391

Liaw A, Wiener M (2002) Classification and regression by random Forest. R News. 2(3):18–22

Li P, Qian H (2011) Human health risk assessment for chemical pollutants in drinking water source in Shizuishan City, Northwest China. Iran J Environ Health Sci Eng 8(1):41–48

Li P, Wu J, Qian H (2014) Hydrogeochemistry and quality assessment of shallow groundwater in the southern part of the Yellow River alluvial plain (Zhongwei section), Northwest China. Earth Sci Res J 18(1), 27–38. https://doi.org/10.15446/esrj.v18n1.34048

Li P, Li X, Meng X, Li M, Zhang Y (2016) Appraising groundwater quality and health risks from contamination in a semiarid region of northwest China. Expo Health. 8(3):361–379. https://doi.org/10.1007/s12403-016-0205-y

Li P, Qian H, Wu J (2018) Conjunctive use of groundwater and surface water to reduce soil salinization in the Yinchuan Plain, North-West China. Int J Water Resour Dev 34(3):337–353. https://doi.org/10.1080/07900627.2018.1443059

Li P, He X, Guo W (2019a) 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. Hum Ecol Risk Assess 25(1–2):11–31. https://doi.org/10.1080/10807039.2018.1553612

Li P, He X, Li Y, Xiang G (2019b) Occurrence and health implication of fluoride in groundwater of loess aquifer in the Chinese loess plateau: a case study of Tongchuan Northwest China. Expo Health 11(2):95–107. https://doi.org/10.1007/s12403-018-0278-x

McArthur J, Banerjee D, Sengupta S, Ravenscroft P, Klump S, Sarkar A, Disch B, Kipfer R (2010) Migration of As, and ^{3}H/^{3}He ages, in groundwater from West Bengal: implications for monitoring. Water Res 44(4):4171–4185. https://doi.org/10.1016/j.watres.2010.05.010

Memon M, Soomro MS, Akhtar MS, Memon KS (2011) Drinking water quality assessment in Southern Sindh (Pakistan). Environ Monit Assess. 177(1–4):39–50. https://doi.org/10.1007/s10661-010-1616-z

Ministry of Environmental Protection of the P.R. China (2014) Technical guidelines for risk assessment of contaminated sites, (HJ 25.3–2014). China Environmental Science Press, Beijing (**in Chinese**)

Ministry of Land and Resources of the P.R. China (2015) Standard for groundwater quality (DZ/T 0290–2015). National standard for geological exploration of mineral resources (**in Chinese**)

Muhammad S, Shah MT, Khan S (2010) Arsenic health risk assessment in drinking water and source apportionment using multivariate statistical techniques in Kohistan region, northern Pakistan. Food Chem Toxicol. 48(10):0278–2864. https://doi.org/10.1016/j.fct.2010.07.018

Naghibi SA, Ahmadi K, Daneshi A (2017) Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour Manag 31(9):2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Niazi NK, Bibi I, Shahid M, Ok YS, Burton ED, Wang H, Shaheen SM, Rinklebe J, Lüttge A (2018) Arsenic removal by perilla leaf biochar in aqueous solutions and groundwater: an integrated spectroscopic and microscopic examination. Environ Pollut 232:31–41. https://doi.org/10.1016/j.envpol.2017.09.051

Nickson RT, Mcarthur JM, Shrestha B, Kyaw-Myint TO, Lowry D (2004) Arsenic and other drinking water quality issues, Muzaffargarh District, Pakistan. Appl Geochem 20(1):55–68. https://doi.org/10.1016/j.apgeochem.2004.06.004

Nicolli HB, Bundschuh J, García JW, Falcón CM, Jean JS (2010) Sources and controls for the mobility of arsenic in oxidizing groundwaters from loess-type sediments in arid/semi-arid dry climates-evidence from the Chaco-Pampean plain (Argentina). Water Res 44(19):5589–5604. https://doi.org/10.1016/j.watres.2010.09.029

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O (2012) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181–199. https://doi.org/10.1007/s10021-005-0054-1

Qian C, Wu X, Mu WP, Fu RZ, Zhu G, Wang ZR, Wang DD (2016) Hydrogeochemical characterization and suitability assessment of groundwater in an agro-pastoral area, Ordos Basin NW China. Environ Earth Sci. 75(20):1356. https://doi.org/10.1007/s12665-016-6123-2

Radfard MM, Yunesian M, Nodehi RN, Biglari H, Nazmara S, Hadi M (2018) Drinking water quality and arsenic health risk assessment in Sistan and Baluchestan, Southeastern Province, Iran. Hum Ecol Risk Assess 25(4):949–965. https://doi.org/10.1080/10807039.2018.1458210

Rasool A, Farooqi A, Xiao T, Masood S, Kamran MA (2016) Elevated levels of arsenic and trace metals in drinking water of Tehsil Mailsi, Punjab Pakistan. J Geochem Explor 169:89–99. https://doi.org/10.1016/j.gexplo.2016.07.013

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

Sahoo S, Russo TA, Elliott J, Foster I (2017) Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US. Water Resour Res 53(5):3878–3895. https://doi.org/10.1002/2016WR019933

Sanaullah M, Mehmood Q, Ahmad SR, Rehman HU (2016) Arsenic contamination trends of abandoned river banks: a case study at the left bank of river Ravi, Lahore. Int J Econ Environ Geol 6(1):21–24

Savarimuthu X, Hira-Smith MM, Yuan Y, von Ehrenstein OS, Das S, Ghosh N, Mazumder DG, Smith AH (2006) Seasonal variation of arsenic concentrations in tubewells in West Bengal, India. J Health Popul Nutr 24(3):277–281

Shahid M, Khalid M, Dumat C, Khalid S, Niazi NK, Imran M, Bibi I, Ahmad I, Hammad HM, Tabassum RA (2018) Arsenic level and risk assessment of groundwater in Vehari, Punjab Province, Pakistan. Expo Health. 10(4):229–239. https://doi.org/10.1007/s12403-017-0257-7

Shakoor MB, Bibi I, Niazi NK, Shahid M, Nawaz MF, Farooqi A, Naidu R, Rahman MM, Murtaza G, Lüttge A (2018) The evaluation of arsenic contamination potential, speciation and hydrogeochemical behaviour in aquifers of Punjab, Pakistan. Chemosphere 199:737–746. https://doi.org/10.1016/j.chemosphere.2018.02.002

Sharifi R, Moore F, Keshavarzi B, Badiei S (2018) Assessment of health risks of arsenic exposure via consumption of crops. Water Qual Expos Health 10(2):129–143. https://doi.org/10.1007/s12403-017-0250-1

Sultana J, Farooqi A, Ali U (2014) Arsenic concentration variability, health risk assessment, and source identification using multivariate analysis in selected villages of public water system, Lahore Pakistan. Environ Monit Assess 186(2):1241–1251. https://doi.org/10.1007/s10661-013-3453-3

Tesoriero AJ, Gronberg JA, Juckem PF, Miller MP, Austin BP (2017) Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. Water Resour Res 53(8):7316–7331. https://doi.org/10.1002/2016WR020197

Toor I, Tahir S (2009) Study of arsenic concentration levels in Pakistani drinking water. Pol J Environ Stud 18(5):907–912

Wang YX, Shvartsev SL, Su CL (2009) Genesis of arsenic/fluoride-enriched soda water: a case study at Datong Northern China. Appl Geochem 24(4):641–649. https://doi.org/10.1016/j.apgeochem.2008.12.015

Wang WL, Wu QY, Wang C, He T, Hu HY (2015) Health risk assessment of phthalate esters (paes) in drinking water sources of china. Environ Sci Pollut Res 22(5):3620–3630. https://doi.org/10.1007/s11356-014-3615-z

Wang Z, Guo HM, Xiu W, Wang J, Shen MM (2018) High arsenic groundwater in the Guide basin, northwestern China: distribution and genesis mechanisms. Sci Total Environ 640–641:194–206. https://doi.org/10.1016/j.scitotenv.2018.05.255

Wu J, Sun Z (2016) Evaluation of shallow groundwater contamination and associated human health risk in an alluvial plain impacted by agricultural and industrial activities, mid-west China. Expo Health 8(3):311–329. https://doi.org/10.1007/s12403-015-0170-x

Wu C, Wu X, Zhu G, Qian C, Mu WP, Zhang YZ (2018a) Influence of a power plant in Ezhou city on the groundwater environment in the nearby area. Environ Earth Sci 77(13):503. https://doi.org/10.1007/s12665-018-7674-1

Wu C, Wu X, Qian C, Zhu G (2018b) Hydrogeochemistry and groundwater quality assessment of high fluoride levels in the Yanchi endorheic region, northwest China. Appl Geochem 98:404–417. https://doi.org/10.1016/j.apgeochem.2018.10.016

Wu J, Zhou H, He S, Zhang Y (2019) Comprehensive understanding of groundwater quality for domestic and agricultural purposes in terms of health risks in a coal mine area of the Ordos basin, north of the Chinese Loess Plateau. Environ Earth Sci 78(15):446. https://doi.org/10.1007/s12665-019-8471-1

Yan BW, Chen L (2013) Coincidence probability of precipitation for the middle route of south-to-north water transfer project in China. J Hydrol 499(9):19–26. https://doi.org/10.1016/j.jhydrol.2013.06.040

Yang J, Griffiths J, Zammit C (2019) National classification of surface–groundwater interaction using random forest machine learning technique. River Res Appl 35(7):932–943. https://doi.org/10.1002/rra.3449

Zubair A, Khan A, Khan M (2014) Incidence and allocation of arsenic in the coastal area of Sindh, Pakistan. World Appl Sci J 32(5), 945–951. https://doi.org/10.58129/idosi.wasj.2014.32.05.34