Chemie der Erde 75 (2015) 197–205

Contents lists available at ScienceDirect

Chemie der Erde

journal homepage: www.elsevier.de/chemer

Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia

Yuxiao Jiang[a,b], Huaming Guo[a,b,*], Yongfeng Jia[a,b], Yongsheng Cao[a,b], Chao Hu[b]

[a] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing 100083, PR China
[b] School of Water Resources and Environment, China University of Geosciences, Beijing 100083, PR China

ARTICLE INFO

Article history:
Received 23 August 2014
Accepted 26 December 2014
Editorial handling – J.C.J. Petit

Keywords:
Quantitative investigations
Deep aquifers
Principal component analysis
Hierarchical cluster analysis
As species
Dominant factor

ABSTRACT

Although high As groundwater has been observed in shallow groundwater of the Hetao basin, little is known about As distribution in deep groundwater. Quantitative investigations into relationships among chemical properties and among samples in different areas were carried out. Ninety groundwater samples were collected from deep aquifers of the northwest of the basin. Twenty-two physicochemical parameters were obtained for each sample. Statistical methods, including principal component analysis (PCA) and hierarchical cluster analysis (HCA), were used to analyze those data. Results show that As species were highly correlated with Fe species, NH_4-N and pH. Furthermore, result of PCA indicates that high As groundwater was controlled by geological, reducing and oxic factors. The samples are classified into three clusters in HCA, which corresponded to the alluvial fans, the distal zone and the flat plain. Moreover, the combination of PCA with HCA shows the different dominant factors in different areas. In the alluvial fans, groundwater is influenced by oxic factors, and low As concentrations are observed. In the distal zone, groundwater is under suboxic conditions, which is dominated by reducing and geological factors. In the flat plain, groundwater is characterized by reducing conditions and high As concentrations, which is dominated by the reducing factor. This investigations indicate that deep groundwater in the alluvial fans mostly contains low As concentrations but high NO_3 and U concentrations, and needs to be carefully checked prior to being used for drinking water sources.

1. Introduction

In China, about 15 million people have been affected by high As groundwater with As concentration >10 μg/L, especially in Xinjiang, Inner Mongolia and Shanxi provinces (Guo et al., 2014a). The Hetao plain is one of the well-known high As groundwater areas, where local residents were exposed to drinking groundwater with high As concentration (Guo et al., 2014a; Luo et al., 2012).

High As concentration was frequently observed in shallow groundwater in the flat plain of the Hetao basin (Guo et al., 2008; Deng et al., 2009). Previous investigations indicated that distribution of high As groundwater exhibited a distinct spatial zonation, including Shuangmiao-Sandaoqiao, Shahai-Manhui, Bainaobao-Langshan, Taerhu, and Shengfeng (Yang et al., 2008), which were coincident with the subsidence center of the basin. Locally, biogeochemical and hydrogeological conditions also affected groundwater As concentrations (Guo et al., 2011). High As groundwater mainly occurred at depths between 20 and 30 m below land surface. However, few studies were carried out along the piedmont areas of the Langshan Mountains, which are located in the north of the Hetao basin. In the areas, many wells with various screening depths (20–100 m) are used for irrigation and domestic water supply. Jia et al. (2014) showed that low As concentrations were observed in deep groundwater near the mountains, while high As concentrations in deep groundwater of the flat plain. Since groundwaters are used for both drinking water resources and agricultural irrigation in this region, As concentrations of these waters would not only affect quality of food products, but also determine drinking water quality.

In recent years, multivariate statistical methods, including principal component analysis (PCA) and hierarchical cluster analysis (HCA), have been applied to investigate the factors controlling As mobilization from large sets of groundwater chemistry data and to classify the groundwater areas (Andrade and Stigter, 2013; Mukherjee et al., 2012). Based on PCA and HCA methods Andrade and Stigter (2013) established FRA explanatory models, which

* Corresponding author at: School of Water Resources and Environment, China University of Geosciences (Beijing), Beijing 100083, PR China.
Tel.: +86 10 8232 1366; fax: +86 10 8232 1081.
E-mail address: hmguo@cugb.edu.cn (H. Guo).

http://dx.doi.org/10.1016/j.chemer.2014.12.002
0009-2819/

improved the understanding of the role of land use and intrinsic factors, such as aquifer lithology and water depth, on As contamination. Mukherjee et al. (2012) studied the relationships among the samples in the areas between the Himalayan foothills and Indian craton (including central Gangetic plain) using HCA and PCA, which finally explained the influence of geology and geomorphology on the As fate in aquifers. Another study in Taiwan, which also used the PCA and HCA, indicated that the reductive dissolution of Fe minerals was prerequisite for mobilization of As, and the shift of redox conditions from Fe- to As-reducing led to the accumulation of dissolved As in aquifers of Choushui River alluvial fan and Chianan Plain (Lu et al., 2012). Moreover, an investigation on the sources of As by PCA and HCA showed that groundwater As might be from geologic and climatic origins, instead of anthropogenic sources (Andrade and Stigter, 2013).

In this study, PCA and HCA methods are applied to investigate groundwater As distribution and geochemical processes in the piedmont areas. Objectives are to (1) investigate distribution of groundwater As along the piedmont areas of the Langshan Mountains; (2) evaluate factors contributing to high As concentrations in groundwater by means of PCA; (3) locate low As groundwater areas by means of HCA in combination with PCA; (4) assess geochemical processes controlling occurrence and mobility of As in groundwater.

2. Study area

2.1. Location and climate

The study area is located in the northwest of the Hetao basin. It covers the piedmont of the Langshan Mountains, and extends from the front of the mountain ranges in the north to the plain in the south (Fig. 1).

The climate in the region is arid, with the annual average precipitation between 130 and 220 mm, annual average evaporation between 2000 and 2500 mm; and the average temperature between 5.6 and 7.8 °C.

2.2. Geology and hydrogeology

The Langshan mountain ranges are mainly composed of Jurassic to Cretaceous metamorphic rocks (slate, gneiss and marble) (Guo et al., 2008). Basement bedrocks in the Hetao basin are characterized by Early Archean and Proterozoic metamorphic complexes (gneiss, schist and slate), granite, and quartzite (Hu et al., 2013). Inland lacustrine sediments with fine clast have locally been deposited in the Quaternary period and thick Mid-Cenozoic sedimentary formation has developed (Deng et al., 2011; Lin et al., 2002). The thickness of the sediment in the southeast of the basin ranges from 500 to 1500 m, and in the northwest from 7000 to 8000 m (Guo et al., 2008, 2011). The sediments are mainly composed of alluvial–pluvial sand, sandy silt, lacustrine and fluvial–lacustrine sandy silt, silty clay and clay rich in organic matter in the central part of the basin, fluvial sand and fine sand on the banks of rivers, and alluvial sand in the fan areas (Guo et al., 2008, 2011).

There are evident zonation characteristics in geological and hydrogeological conditions in the study area. From the foothills to the plain, grain-size of sediments decreases, with the decreasing in permeability and hydraulic conductivity. Hydraulic conductivity decreases along the flow path from the alluvial fans to the flat plain, ranging from ~2.0 m/d along the mountain front to ~0.2 m/d in the downdip region (Inner Mongolia Institute of Hydrogeology, 1982). According to hydrogeological conditions, groundwater at depths >40 m is usually considered as deep groundwater (Jia et al., 2014). Shallow groundwater usually occurs at depths <10 m. Deep groundwater is mainly recharged by lateral flow from mountains and vertical infiltration from shallow groundwater, and discharged by artificial extraction. Shallow groundwater is mainly recharged from precipitation, irrigation water, and surface water (from lakes and water channels), which is mainly discharged by artificial extraction, evaporation, and vertical flow into deep groundwater. Surface topography strongly affects directions of groundwater flow. The direction of groundwater flow is from north to south (Fig. 1) with the flow rate generally higher in the alluvial fans than in the flat plain. Due to the relatively flat terrain, fine grain of the aquifer, arid climate and low hydraulic conductivity, the groundwater flow rate in the plain is very low (Guo et al., 2012).

3. Materials and methods

3.1. Sample collection and analysis procedures

Ninety groundwater samples were collected from tube wells in the study area in the years of 2011 and 2012 (Fig. 1). Depths of sampling wells varied from 50 to 110 m below ground level (bgl), thus essentially representing deep groundwaters.

In the field, groundwater was sampled after pumping sufficiently until the flowing water showed a stabile temperature, pH, EC (specific conductance), and ORP. Parameters (including ORP, temperature, pH, and EC) were measured in an inline flow cell under minimal atmospheric contact using a multiparameter portable meter (HANNA, HI 9828), which was calibrated using standard solution before use. Redox-sensitive parameters (including Fe(II), NH_4-N, H_2S) were determined using a portable spectrophotometer (HACH, DR2800). Alkalinity was determined by titration with 0.8 M H_2SO_4 using a Model 16900 digital titrator (HACH) with phenolphthalein and methyl-orange indicators.

All groundwater samples were filtered through 0.22 μm filters (Sartorius). Samples for cation and trace element analysis were collected in HDPE bottles, and acidified to pH <2 using ultrapure HNO_3. Those for anion analysis were not acidified. Samples for dissolved organic carbon (DOC) analysis were collected in amber glass bottles and acidified with H_2SO_4 to pH <2. Sub-samples for As species analysis were preserved with 0.25 M EDTA (10%) in 2.0 mL amber glass bottles. All groundwater samples were stored in an ice box, and delivered to the laboratory, where they were immediately preserved in a refrigerator at 4 °C until analysis within two weeks.

Major cations and trace metal elements were measured using an inductively coupled plasma atomic emission spectrometer (iCAP6300, Thermo) and ICP-MS (7500C, Agilent), respectively. The analysis precision of ICP-AES and ICP-MS was 0.5%. The detection limit for As was 0.01 $\mu g\,L^{-1}$. Concentrations of Cl, NO_3, and SO_4 were determined using an ion chromatography system (ICS2000, Dionex), with the analysis precision less than 3.0%. Arsenic species (including As(III), As(V), MMA, and DMA) were analyzed by high-performance liquid chromatography-atomic fluorescence spectrometry (HPLC-AFS) (AFS9130, Titan), with the relative standard deviation (RDS) <±5% and the analysis precision of 5.0% (Guo et al., 2014b). Dissolved organic carbon (DOC) was determined by total carbon analyzer (TOC-Vwp, Shimadzu). For most samples, ion charge imbalances were less than 5%.

3.2. Statistical methods

Multivariate statistics (including principal component analyses (PCA) and hierarchical cluster analyses (HCA)) were employed to quantitatively investigate relationships among the dataset of 1980 values of the investigated samples (22 parameters determined in 90 samples).

Principal component analysis (PCA) is one of the most commonly used multivariate statistical methods in natural sciences, which was developed by Hotelling (1933) in the thirties from original work of Pearson. The main objective of this method is to simplify data structure by reducing the dimension of the data. The original parameters would be rearranged into several new uncorrelated comprehensive components (or factors) without losing significant information (Brown and Brown, 1998; Pereira and Sousa, 1988). Every new component is the linear combination of the original variables and unrelated, which makes it accurate to describe the characteristics of the analyzed data. The calculated factors are rotated with the method of Varimax rotation, thus making the loadings of closely related variables in each factor strengthened. Each component describes a certain amount of the statistical variance of the analyzed data and is interpreted according to the inter-correlated variables. Variable loadings are defined by the orthogonal projection of the variables on each of the factors. The selection of the factors is based on both the significance (eigenvalues >1) of the factor and the cumulative percentage of data variance explained. The last and significant step is to interpret each factor in association with the studied issue.

In this study, after the interpretation of each factor, the contribution of each factor (factor scores) at each monitoring well was computed. The factor scores showed the relation between samples and the components quantitatively, which indicated