Hydrogeochemical evolution of groundwater in a Quaternary sediment and Cretaceous sandstone unconfined aquifer in Northwestern China

Zhi Qiang Yu · Hiroki Amano · Kei Nakagawa*· Ronny Berndtsson

Zhi Qiang Yu・Hiroki Amano・Kei Nakagawa (✉)

Graduate School of Fisheries and Environmental Sciences, Nagasaki University,

1-14 Bunkyo-machi, Nagasaki 852-8521, Japan

e-mail: kei-naka@nagasaki-u.ac.jp

Tel.: +81 95 819 2763; fax: +81 95 819 2763

* corresponding author

Ronny Berndtsson

Division of Water Resources Engineering & Center for Middle Eastern Studies, Lund University,

Box 118, SE-221 00 Lund, Sweden

Abstract

A better understanding of the hydrogeochemical evolution of groundwater in vulnerable aquifers is important for the protection of water resources. To assess groundwater chemistry, groundwater sampling was performed from different representative aquifers in 2012-13. A Piper trilinear diagram showed that the groundwater types can be classified into Na-SO_4 and Na-Cl types. Only one groundwater sample was Na-HCO_3 type. The dominant cations for all samples were Na^+. However,

the dominant anions varied from HCO_3^- to SO_4^{2-}, and as well Cl^-. The mean total dissolved solid (TDS) content of groundwater in the region was 1,889 mg/L. Thus, only 20% of groundwater samples meet Chinese drinking water standards (<1,000 mg/L). Principal component analysis (PCA) combined with hierarchical cluster analysis (HCA) and self-organizing maps (SOM) were applied for the classification of the groundwater geochemistry. The three first principal components explained 58, 20, and 16% of the variance, respectively. The first component reflects sulfate minerals (gypsum, anhydrite) and halite dissolution, and/or evaporation in the shallow aquifer. The second and third components are interpreted as carbonate rock dissolution. The reason for two factors is that the different aquifers give rise to different degree of hydrogeochemical evolution (different travel distances and travel times). Identified clusters for evolution characteristic and influencing factors were confirmed by the PCA-HCA methods. Using information from eight ion components and SOM, formation mechanisms and influencing factors for the present groundwater quality were determined.

Keywords

Self-organizing maps, Hydrogeochemical characteristics, The Dosit River, Principal component analysis, Sulfate minerals

Introduction

The Ordos Basin is part of a large-scale sedimentary geological area in northwestern China. Abundant organic minerals such as coal, petroleum, natural gas, and halite have been deposited in the basin, which makes it one of the largest sources for energy and petrochemical production in

China (Jiang et al., 2012). During recent decades, the economy and mining industry have boomed in the area, resulting in a dramatic increase in water demand. However, due to the lack of surface water, groundwater is the main water resource for local development needs. The Dosit River is one of a number of the sub-basins of the Ordos Basin with similar resources and water problems. In order to meet the water resources demands for drinking water and industrial water supply in this region, research has focused on determining: (1) the groundwater hydrogeochemical characteristics; (2) recharge and evaporation amount estimates; and (3) assessment of the characteristics of the local groundwater flow systems (Yin et al., 2011; Jiang et al., 2014; Wang et al., 2015, Jiang et al., 2018). As water scarcity is a serious issue in the region, determining the key aspects of the hydrogeology of the region is important for the local economic development. This is particularly the case for understanding the quality of groundwater and its availability for extractive uses, the characteristics of surface-groundwater interactions, and for an improved understanding of the hydrogeochemistry and groundwater circulation characteristics. To improve this understanding, groundwater samples were collected in the downstream region of the Dosit River for assessing water quality and its chemical composition. Piper trilinear diagrams, multivariate statistical analyses such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to investigate collected groundwater samples (e.g., Brown, 1998; Cloutier et al., 2008; Nakagawa et al., 2016). Recently, self-organizing maps (SOM) have been shown to efficiently classify groundwater chemistry (Choi et al., 2014; Nguyen et al., 2015). Consequently, we combined SOM with PCA-HCA methods to investigate groundwater chemistry for the downstream Dosit River area.

In view of the above, the main objectives of this study were: (1) to assess groundwater quality, hydrogeochemistry, and evolution characteristics of the groundwater by use of the

aforementioned techniques; and (2) to determine usefulness of these methods for hydrogeochemical investigations in regions like the Dosit River area.

Materials and methods

Study area

The Dosit River Basin is located between 106°54'28'' and 108°16'16'' E, and 38°18'21'' and 39° 36'06'' N (Fig. 1). It occupies an area of 10,924 km^2 and constitutes a sub-basin of the Ordos Basin in Northwestern China. In turn, the Dosit River discharges into the important Yellow River. The altitude of the Dosit River Basin varies from 1,080 to 1,500 m. The basin is mostly surrounded by mountains with the exception of the west and water flows in a westerly direction in the basin. The average monthly temperatures range from -10.5 ^{o}C in January to 22.4 ^{o}C in July and the mean annual temperature is about 6.9 ^{o}C (1955-2007). The mean annual rainfall of the region is about 267 mm (1955-2007) and 60-80% of rainfall takes place from June to September. The mean annual evaporation is about 2,465 mm (Sun, 2010; Jiang et al., 2014). The basin climate is thus characterized as arid to semiarid.

The aquifers of the Dosit River watershed can be classified into two groups. The first, is the uppermost aeolian and alluvial-lacustrine pore aquifer system in Quaternary sediments that is a thin and uneven unconfined aquifer. The second, deeper aquifer system is constituted by a poorly consolidated, pore-fissure aquifer system of Cretaceous sandstone with a thickness of 700-1,000 m which is the main aquifer system of the Dosit River Basin. The Jurassic mudstone with coal layers, which underlies the Cretaceous sandstone, is generally assumed to be an aquiclude. The main minerals of the Cretaceous sandstone aquifer contain quartz, albite, and feldspar, as well as some

minor minerals such as gypsum, mirabilite, halite, calcite, and dolomite (Sun, 2010; Wang et al., 2015). According to hydrogeochemical and isotope studies, a three-part structure with boundaries at depths of 200 m and 600-750 m, is consistent with local, intermediate, and regional groundwater flow systems (Wang et al., 2015). The principle recharge and discharge areas of local flow systems are mainly distributed on sides of valley slopes. Groundwater discharge from the deeper aquifers mostly takes place in river channels and recharge to these aquifers mostly takes place at a distance from major rivers. The groundwater of these flow systems contributes to baseflow in the Dosit River (Wang et al., 2016).

Sampling and analysis

Groundwater samples were collected by pumping at 20 locations from existing water supply wells in the downstream part of the Dosit River in 2012-2013 (Fig. 1). Water samples were collected in pre-washed bottles. Temperature, pH, and electrical conductivity (EC) were analyzed at the sampling sites, CO_3^{2-} and HCO_3^- concentrations were identified with titration by use of HCl. Cl^- and SO_4^{2-} concentration were analyzed by ion chromatography (ICP-900, Dionex), while Na^+, K^+, Mg^{2+}, and Ca^{2+} concentrations were measured by inductively coupled plasma (ICP-900, Thermo) (Wang et al., 2015). Charge balance errors (CBE) were checked for all groundwater samples according to:

$$\mathrm{CBE} = \frac{\sum \text{cations} - \sum \text{anions}}{\sum \text{cations} + \sum \text{anions}} \times 100$$

where all ions concentrations are expressed in $mmol_c/L$. We confirmed that all CBEs of samples were less than 10%.

Multivariate statistical analysis

Principal component analysis (PCA) is a well-tested multivariate statistical method that has been widely used to analyze hydrochemical groundwater data. Its key feature is data reduction from high- to low-dimensional space (Morell et al., 1996; Choi et al., 2014). It extracts synthetic variables with minimal information loss (Aiuppa et al., 2003; Nakagawa et al., 2017). HCA is an effective method for data classification. It has been used to identify clusters of groundwater samples based on similarity of hydrogeochemical components (Cloutier et al., 2008; Montcoudiol et al., 2015). The classification result is expressed in a dendrogram, using the Euclidean distance as a distance measure (Ward's method) (Güler et al., 2002). In this study, the statistical software JMP Pro 13 (SAS Institute Inc.) was used when using PCA and HCA.

The Self-organizing map (SOM) technique is also a powerful and effective tool for data classification (Nguyen et al., 2015; Nakagawa et al., 2017). It has been used in different research fields such as hydrology (Kalteh and Berndtsson, 2007), wastewater treatment (Yu et al., 2014), and meteorology (Nishiyama et al., 2007). As well, SOM has been applied to investigating water chemistry of river and groundwater (Choi et al., 2014; Nguyen et al., 2015). SOM is a type of artificial neural network technique, which is distinguished by unsupervised training (Kohonen, 2001). It can project high-dimensional data onto a low-dimensional array, and let complex target data simplify into a regular arranged map based on the degree of similarity (Jin et al., 2011). In general, the purpose of the SOM application is to acquire useful and informative reference vectors. These vectors are obtained by iterative updates in the training phase of SOM that is made up from three main steps; competition between nodes, selection of a winner nodes, and updating of the reference vectors (Nguyen et al., 2015). When applying SOM methodology, the selection of an

appropriate initialization and data transformation method is important. On the basis of SOM properties, larger map sizes will obtain a higher resolution for pattern recognition. The optimal number of map nodes is determined by heuristic rules according to $m = 5\sqrt{n}$, where *m* represents the number of map nodes and *n* represents the number of input data (Hentati et al., 2010). The number of rows and columns is dependent on square root of the ratio between the two largest eigenvalues of the transformed data (García and González, 2004). The eigenvalues are calculated by PCA.

Following the above rules, the SOM structure was organized. Using a linear initialization technique, each node is set with a reference vector. Under the proper limited data condition, a linear initialization technique is better for the pattern classification as compared to random initialization due to small data sets and boundary effects (Nguyen et al., 2015). More details of SOM are discussed by Kohonen, (1982, 2001) and Vesanto et al. (2000). Results of SOM analysis are achieved at the end of the training process, which is fine-tuned using cluster analysis. K-means algorithms, which is a partitioned algorithm, is frequently used in SOM (Jin et al., 2011). Davies-Bouldin Index (DBI) applying k-means algorithms determine the optimal number of clusters (García and González, 2004). According to the principle of "similarity within a cluster" and "dissimilarity between clusters", the DBI values were calculated from the minimum to maximum number of clusters. The smaller DBI values show that the dissimilarity of each cluster becomes larger (Nakagawa et al., 2017). In other words, the minimum DBI corresponds to the optimal number of clusters in SOM application. These processes were put into practice using a modified version of SOM Toolbox 2.0 (Vesanto et al., 2000).

Results and discussion

Hydrogeochemical characteristics

Table 1 shows the statistical summary of hydrogeochemical variables (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, CO_3^{2-}, HCO_3^-, TDS, pH) for the 20 samples. Mean TDS corresponded to 1,889 mg/L, with a minimum of 455 mg/L and a maximum of 3,974 mg/L. TDS concentration of surface water is generally higher in the downstream of Dosit River, due to evaporation and the effects of groundwater evolution. pH varied from 8.0 to 9.0, indicating that the water environment is slightly alkaline. Figure 2 shows results plotted in a Piper trilinear diagram (Piper, 1944). All samples are located in area IV. Only one sample is classified as Na-HCO_3 type. Seven samples showed Na-SO_4 type. The other samples represented Na-Cl type. Cations were dominated by Na^+ in all samples. Principal anions changed from HCO_3^- to SO_4^{2-}, and then to Cl^- along groundwater flow direction. This process corresponds to anion evolution according to $HCO_3^- \rightarrow HCO_3^- + SO_4^{2-} \rightarrow SO_4^{2-} + Cl^- \rightarrow Cl^- + SO_4^{2-} \rightarrow Cl^-$ from recharge to discharge areas in aquifers of large sedimentary basins (Chebotarev, 1955; Singhal and Gupta 2010).

Principal component and hierarchical cluster analysis

Eight parameters (Na^+, K^+, Mg^{2+}, Ca^{2+}, Cl^-, HCO_3^-, CO_3^{2-}, and SO_4^{2-}) were selected for analysis by PCA and HCA. Firstly, input data were standardized, and then eigenvalues, factor loadings, and principal component scores were calculated using a correlation matrix (Nakagawa et al., 2016). The total number of common factors in the PCA was selected on the basis of the Kaiser Criterion (Cloutier et al., 2008). In this criterion, if an eigenvalue is greater than 1, it is retained. According to this rule, the first three components (Factors 1, 2, and 3) were extracted. Ward's

method with Euclidean distances was selected for HCA. For more distinct understanding of the eight ionic correlations, TDS, pH, and water table depth were not considered in the PCA-HCA analysis. The result of PCA is shown in Table 2. The first three principal components accounted for a total of 94% of the variance. The three components explained 57.8, 20.1, and 16.1% of the variance, respectively. Hereafter, the first three principal components are called Factors 1, 2 and 3, respectively. Factor 1 had positive loadings for Ca^{2+}, Mg^{2+}, SO_4^{2-}, Na^+, Cl^-, and K^+ (r = 0.80~0.94). Factor 1 is interpreted as sulfate minerals (gypsum and anhydrite) and halite dissolution, and/or evaporation in the shallow aquifer. Factor 2 has positive high loadings for HCO_3^-, and CO_3^{2-} (r = 0.66 and 0.93). Factor 2 is mainly affected by carbonate rock dissolution. Factor 3 has moderately positive loadings for Mg^{2+} and Ca^{2+} (r = 0.49 and 0.50). Factor 3 is also mainly affected by carbonate rock dissolution. The reason may be a different degree of carbonate rock dissolution due to different travel distances and travel times in the different aquifers.

The scatter plot of the 20 samples described by principal components (Factors 1 and 2; Factors 2 and 3) and classified into four clusters based on HCA is shown in Fig. 3. If a factor score is greater than 0, it means that the component is influenced by the water chemistry characteristic at the site. Conversely, if a factor score is less than 0, it means that the component was not significantly affected by the water chemistry at the site (Banoeng-Yakubo et al., 2009). In Fig. 3(a) Cluster A is to some extent influenced by Factor 2. Cluster C is to some extent affected by Factor 1. Clusters B and D are influenced by both Factors 1 and 2. As well, Fig. 3(b) indicates that Clusters A and B are affected by Factor 3. Factor 3 has less influence on Clusters C and D. According to the local and regional groundwater flow systems of the Dosit River Watershed (Wang et al., 2016) and evolution, samples of Clusters A and B are located at different sites but in the same aquifer (depth less than

200 m; local flow system). Samples of Cluster A are located in the upstream area as opposed to that of Cluster B (Fig. 4). Similarly, samples of Clusters C and D are located at different sites but in the same aquifer (depth between 200 and 600 m; regional flow system). The location of samples classified into Cluster D is closer to the river than that of Cluster C (Fig. 4).

Water samples were classified into Clusters A, B, C, and D in a Piper trilinear diagram (Fig. 2). Water composition of samples in Cluster A was modified to that of Cluster B due to sulfate minerals and halite dissolution with groundwater flow from upstream to downstream (Figs. 2 and 4). Because of the groundwater discharge from the peripheral aquifer to the river (Jiang et al., 2018), carbonate dissolution changed water composition of samples in Cluster C to that in Cluster D (Figs. 2 and 4). Thus, groundwater samples of Cluster A affected by Factor 1 appear to evolve into groundwater samples of Cluster B. As well, water samples of Cluster C affected by Factor 2 appear to evolve into water samples of Cluster D.

Self-organizing map analysis

Based on the methodology described above, concentrations of the eight chemical variables (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, CO_3^{2-}, and HCO_3^-) for the 20 samples were used as input to the SOM application. Figure 5 shows visual SOMs for each parameter after the training process. The SOMs are characterized by 24 nodes (number of rows and columns are 6 and 4, respectively) and display concentration characteristics of each variable. If the node color is dark gray, it represents variables that have high values. On the contrary, nodes color that are light gray, means that variables have low values. From Fig. 5 it can be seen that maps for Ca^{2+}, Mg^{2+}, and SO_4^{2-} have similar gray gradients, indicating that there is a strong positive correlation among these ions. Maps for Na^+ and

Cl^- also have a similar gray gradient. In contrast, HCO_3^- displays negative correlation with inverse gray gradients for the SOM maps. The main sources of HCO_3^- is dissolution of carbonate rock (calcite, dolomite), which commonly occurs in the area (Sun, 2010). However, there is no significant correlation between HCO_3^- and cations, which suggests that ion exchange of cations and carbonate precipitation occurs. Chloro-alkaline indices (CAI-I and CAI-II) suggested by Schoeller (1972) were calculated to assess the ion exchange. Both CAI-I and CAI-II for all samples were negative from -2.99 to -0.15, and -1.02 to -0.13, respectively, indicating that there is exchange Na^+ and K^+ in the soil matrix with Ca^{2+} and Mg^{2+} in groundwater. Saturation Indices (SI) with respect to calcite, dolomite, and aragonite were calculated by using PhreeqcI version 3.20-9820 in this study (Parkhurst and Appelo 2013). The SI values of calcite, dolomite, and aragonite ranged from -0.33 to 1.38, -1.49 to 2.93, and -0.47 to 1.24, respectively. Except for the sampling location CBS3 that is located in the upstream reaches, most of the samples showed positive SI values, which explain that groundwater is saturated to oversaturated with these carbonate minerals in the lower reaches of the watershed. The main sources of Ca^{2+} and SO_4^{2-} are likely to be from the dissolution of gypsum and anhydrite. Thus, there is a positive correlation between Ca^{2+} and SO_4^{2-}. However, Ca^{2+} concentrations for the groundwater samples are generally lower than SO_4^{2-} concentration, suggesting that carbonate deposition is taking place (Sun, 2010). Similarly, Na^+ and Cl^- come mainly from halite dissolution and/or evaporation. To further confirm quantitative relations between the eight main ion components, correlation coefficients for all variables were calculated using the reference vectors (Table 3). Positive correlations exist between Ca^{2+} and Mg^{2+} (r = 0.99), Ca^{2+} and SO_4^{2-} (r = 0.97), and Mg^{2+} and SO_4^{2-} (r = 0.95), respectively. Moreover, there is a high correlation between Na^+ and Cl^- (r = 0.96). These results are consistent with the SOM maps.

The variation of DBI from minimum to maximum cluster is shown in Fig. 6. The minimum DBI was obtained for a number of groups equal to 4. After determining the optimal number of clusters, the HCA with Ward's method was implemented. Figure 7 shows the hierarchical cluster dendrogram for the SOM nodes. According to the figure, 24 SOM nodes were divided into four different groups.

The pattern classification maps for the four groups are shown in Fig. 8. All samples were classified into these groups (nodes). Group 2 (lower right part) of the pattern classification map is linked to high concentration of Na^+, Cl^-, and SO_4^{2-}. These features are also observed for the same location in respective component plane (Fig. 5). On the other hand, all ions (except for HCO_3^-) in Group 1 located at the upper left part have lower concentrations than corresponding ones for the other groups (Fig. 5).

The groundwater chemistry characteristics for respective groups using standardized reference vectors are shown in Fig. 9. Group 1 is characterized by relatively high concentration of HCO_3^-. Group 2 displays high concentrations for all cations, SO_4^{2-} and Cl^-. Group 3 is characterized by high Na^+ and Cl^- concentrations. Group 4 is identified by high Na^+, K^+, Cl^-, HCO_3^-, and CO_3^{2-} concentrations. In order to further understand group characteristics, average ion concentrations for each group were calculated using raw data (Table 4). The results of this analysis showed that Group 1 can be identified by relatively high contents of SO_4^{2-} and HCO_3^-. Other ions have lower concentrations. On the other hand, Groups 2, 3, and 4 are characterized by high concentrations of Cl^- and SO_4^{2-}. Compared to Chinese drinking water standards, concentrations of Cl^- and SO_4^{2-} for Groups 2, 3, and 4 greatly exceed these standards (250 mg/L), respectively. Besides, SOM classification results are consistent with the HCA category from Groups 1, 2, 3, and 4 to Clusters

A, B, C, and D, respectively.

Conclusions

In this study, to improve the understanding of hydrogeochemical evolution characteristics and assessing water quality, groundwater samples were collected in the lower part of the Dosit River in the Ordos Basin. By using a Piper trilinear diagram, PCA-HCA and SOM methods, the main conclusions were obtained as follows: (1) TDS for 80% of samples exceeded the Chinese drinking water standard (TDS<1,000 mg/L). The hydrogeochemical facies included Na-HCO_3, Na-SO_4 and Na-Cl types. Anions evolve from HCO_3^-, SO_4^{2-}, and Cl^- along the groundwater flow. This process is in accordance with anion evolution characteristics from recharge to discharge areas for aquifers in large sedimentary basins. (2) By applying PCA, three main influencing factors were identified. These correspond to gypsum, halite as well as carbonate dissolution, evaporation, and carbonate precipitation. (3) Using PCA and HCA, water samples were divided into four clusters. In addition, evolution features among these clusters could be identified. (4) Using SOM, we obtained component maps for each variable, which is a readily understandable and visualized map for the strong relationships between Ca^{2+} - Mg^{2+} (r = 0.99), Ca^{2+} - SO_4^{2-} (r = 0.97), Mg^{2+} - SO_4^{2-} (r = 0.95), and Na^+ - Cl^- (r = 0.96). (5) SOM classification results are consistent with the HCA category from Groups 1, 2, 3, and 4 to Clusters A, B, C, and D, respectively.

References

Aiuppa A, Bellomo S, Brusca L, D'Alessandro W, Federico C (2003) Natural and anthropogenic factors affecting groundwater quality of an active volcano (Mt. Etna, Italy). Applied

geochemistry 18(6): 863-882. https://doi.org/10.1016/S0883-2927(02)00182-8

Banoeng-Yakubo B, Yidana SM, Nti E (2009) Hydrochemical analysis of groundwater using multivariate statistical methods -the Volta region, Ghana. KSCE Journal of Civil Engineering 13(1): 55-63. https://doi.org/10.1007/s12205-009-0055-2

Brown C (1998) Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, Berlin

Cloutier V, Lefebvre R, Therrien R, Savard MM (2008) Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. Journal of Hydrology 353(3-4): 294-313. https://doi.org/10.1016/j.jhydrol.2008.02.015

Choi BY, Yun ST, Kim KH, Kim JW, Kim HM, Koh YK (2014) Hydrogeochmical interpretation of South Korean groundwater monitoring data using self-organizing maps. Journal of Geochemical Exploration 137: 73-84. https://doi.org/10.1016/j.gexplo.2013.12.001

Chebotarev II (1955) Metamorphism of natural waters in the crust of weathering-1. Geochimica et Cosmochimica Acta 8(1-2): 22-48. https://doi.org/10.1016/0016-7037(55)90015-6

Güler C, Thyne GD, McCray JE, Turner KA (2002) Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeology Journal 10(4): 455-474. https://doi.org/10.1007/s10040-002-0196-6

García HL, González IM (2004) Self-organizing map and clustering for wastewater treatment monitoring. Engineering Applications of Artificial Intelligence 17(3): 215-225. https://doi.org/10.1016/j.engappai.2004.03.004

Hentati A, Kawamura A, Amaguchi H, Iseri Y (2010) Evaluation of sedimentation vulnerability at

small hillside reservoirs in the semi-arid region of Tunisia using the Self-Organizing Map. Geomorphology 122(1-2): 56-64. https://doi.org/10.1016/j.geomorph.2010.05.013

Jiang XW, Wan L, Ge S, Cao GL, Hou GC, Hu FS, Wang XS, Li H, Liang SH (2012) A quantitative study on accumulation of age mass around stagnation points in nested flow systems. Water Resource Research 48(12): W12502. https://doi.org/10.1029/2012WR012509

Jiang XW, Wan L, Wang JZ, Yin BX, Fu WX, Lin CH (2014) Field identification of groundwater flow systems and hydraulic traps in drainage basins using a geophysical method. Geophysical Research Letter 41(8): 2812-2819. https://doi.org/10.1002/2014GL059579

Jiang XW, Wan L, Wang XS, Wang D, Wang H, Wang JZ, Zhang H, Zhang ZY, Zhao KY (2018) A multi-method study of regional groundwater circulation in the Ordos Plateau, NW China. Hydrogeology Journal. https://doi.org/10.1007/s10040-018-1731-4

Jin YH, Kawamura A, Park SC, Nakagawa N, Amaguchi H, Olsson J (2011) Spatiotemporal classification of environmental monitoring data in the Yeongsan River basin, Korea, using self-organizing maps. Journal of Environmental Monitoring 13(10): 2886-2894. https://doi.org/10.1039/C1EM10132C

Kalteh AM, Berndtsson R (2007) Interpolating monthly precipitation by self-organizing map (SOM) and multilayer perceptron (MLP). Hydrological Sciences Journal 52(2):305-317. https://doi.org/10.1623/hysj.52.2.305

Kohonen T (2001) Self-organizing maps, third ed. Springer, Berlin

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biological Cybernetics 43(1): 59-69. https://doi.org/10.1007/BF00337288

Morell L, Giménez E, Esteller MV (1996) Application of principal component analysis to the study

of salinization on the Castellon Plain (Spain). Science of the Total Environment 177(1-3): 161-171. https://doi.org/10.1016/0048-9697(95)04893-6

Montcoudiol N, Molson J, Lemieux JM (2015) Groundwater geochemistry of the Outaouais Region (Québec Canada): a regional-scale study. Hydrogeololgy Journal 23(2): 377-396. https://doi.org/10.1007/s10040-014-1190-5

Nakagawa K, Amano H, Asakura H, Berndtsson R (2016) Spatial trends of nitrate pollution and groundwater chemistry in Shimabara, Nagasaki, Japan. Environmental Earth Sciences 75: 234. https://doi.org/10.1007/s12665-015-4971-9

Nakagawa K, Amano H, Kawamura A, Berndtsson R (2017) Classification of groundwater chemistry in Shimabara, using self-organizing maps. Hydrology Research 48(3): 840-850. https://doi.org/10.2166/nh.2016.072

Nguyen TT, Kawamura A, Tong TN, Nakagawa N, Amaguchi H, Gilbuena R (2015) Clustering spatio-seasonal hydrogeochemical data using self-organizing maps for groundwater quality assessment in the Red River Delta, Vietnam. Journal of Hydrology 522: 661-673. https://doi.org/10.1016/j.jhydrol.2015.01.023

Nishiyama K, Endo S, Jinno K, Betracchi Uvo C, Olsson J, Berndtsson R (2007) Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self-Organizing Map. Atmospheric Research 83 (2-4): 185-200. https://doi.org/10.1016/j.atmosres.2005.10.015

Parkhurst DL, Appelo CAJ (2013) Description of input and examples for PHREEQC version 3-A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculation. U.S. Geological Survey Techniques and Methods, book 6, chap.

A43. http://pubs.usgs.gov/tm/06/a43. Accessed 8 July 2018

Piper AM (1944) A graphic procedure in the geo chemical interpretation of water analysis. Trans Am Geophys Union 25(6): 914-928. https://doi.org/10.1029/TR025i006p00914

Schoeller H (1972) Geochemistry of ground water. In: Brown RH (ed) Ground-water studies: an international guide for research and practice. UNESCO, Paris, pp 457-498

Singhal BBS, Gupta RP (2010) Applied hydrogeology of fractured rocks, Second Edition. Springer Netherland, Dordrecht

Sun FQ (2010) Research on groundwater circulation and environment effect of Dosit River in Ordos Basin. Changan University (in Chinese with English abstract)

Vesanto J, Himberg J, Alhoniemi E, Parahankangas J (2000) SOM Toolbox for Matlab 5, Helsinki University of Technology Report A57

Wang H, Jiang XW, Wan L, Han G, Guo H (2015) Hydrogeochemical characterization of groundwater flow systems in the discharge areas of a river basin. Journal of Hydrology 527: 433-441. https://doi.org/10.1016/j.jhydrol.2015.04.063

Wang JZ, Wörman A, Bresciani E, Wan L, Wang XS, Jiang XW (2016) On the use of late-time peaks of residence time distributions for the characterization of hierarchically nested groundwater flow systems. Journal of Hydrology 543: 47-58. https://doi.org/10.1016/j.jhydrol.2016.04.034

Yin L, Hou GC, Su XS, Wang D, Dong J, Hao Y, Wang X (2011) Isotopes (δD and $\delta^{18}O$) in precipitation, groundwater and surface water in the Ordos Plateau, China: implications with respect to groundwater recharge and circulation. Hydrogeology Journal 19(2): 429-443. https://doi.org/10.1007/s10040-010-0671-4

Yu H, Song Y, Liu R, Pan H, Xiang L, Qian F (2014) Identifying changes in dissolved organic matter content and characteristics by fluorescence spectroscopy coupled with self-organizing map and classification and regression tree analysis during wastewater treatment. Chemosphere 113: 79-86. https://doi.org/10.1016/j.chemosphere.2014.04.020

Figure Captions

Fig. 1 Topography and groundwater sample points for the research area

Fig. 2 Piper trilinear diagram for 20 samples

Fig. 3 Scatter plot for two principal components with respect to clusters (a) Factors 1 and 2; (b) Factors 2 and 3

Fig. 4 Spatial distribution of respective cluster in the study area

Fig. 5 Component planes for (Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, HCO_3^-, and CO_3^{2-})

Fig. 6 Variation of DBI values with optimal number of groups

Fig. 7 Dendrogram for respective group using node numbers

Fig. 8 Pattern classification map of the four groups by the SOM

Fig. 9 Radar charts for respective groups using standardized data

Table captions

Table 1 Statistical summary of water samples (eight parameters, pH, TDS)

Table 2 Results of principal component analysis

Table 3 Correlation coefficients among eight physicochemical variables using standardized data

Table 4 Mean of the eight variables for four groups and all data

Figure 1

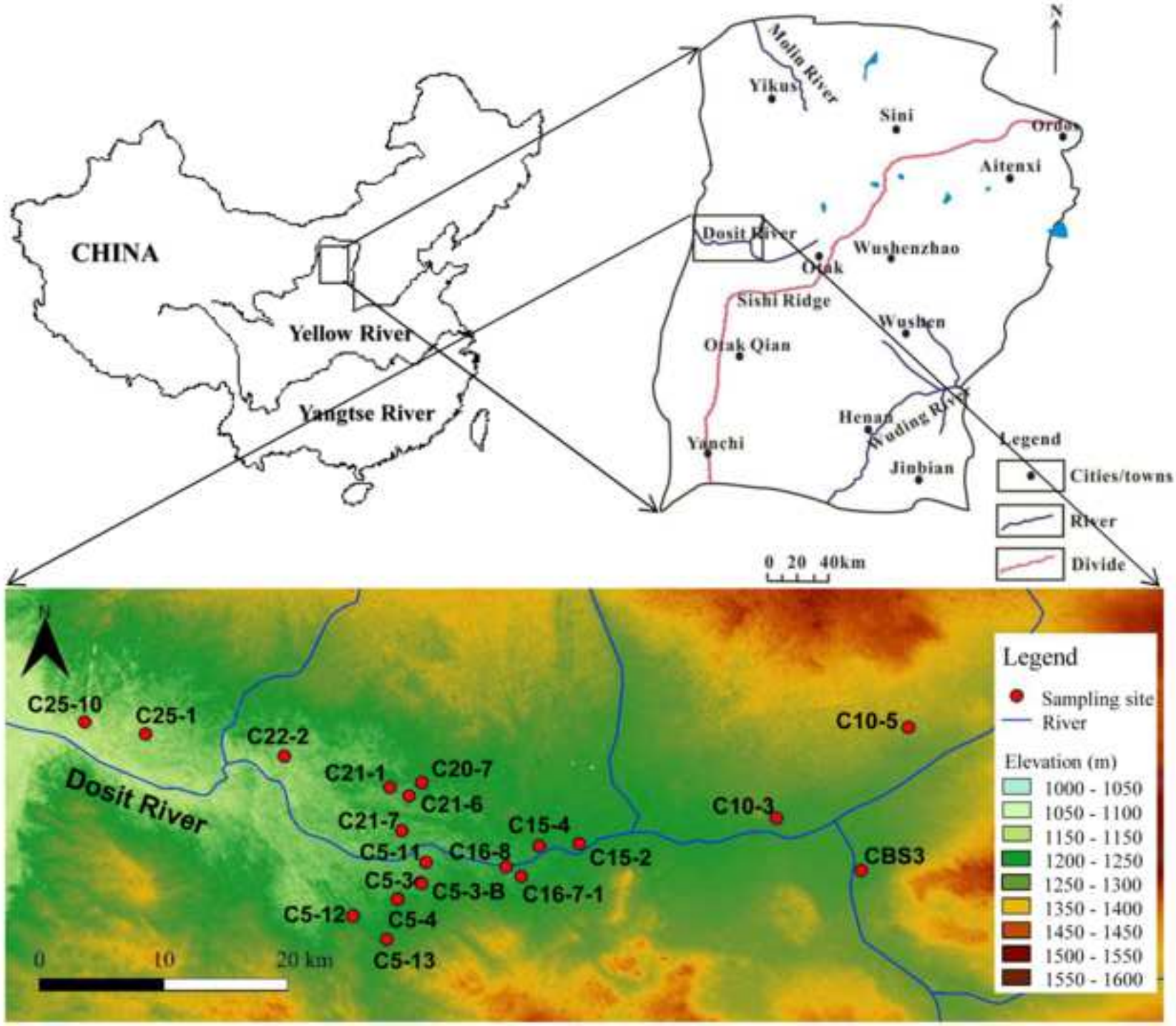

CHINA
Yellow River
Yangtse River
Molin River
Yikus
Sini
Aitenxi
Dosit River
Otak
Wushenzhao
Sishi Ridge
Otak Qian
Wushen
Yanchi
Henan
Wuding River
Jinbian
Legend
Cities/towns
River
Divide
0 20 40km
N
C25-10
C25-1
C22-2
C21-1
C20-7
C21-6
C21-7
C15-4
C5-11
C16-8
C15-2
C5-3
C5-3-B
C16-7-1
C5-12
C5-4
C5-13
C10-5
C10-3
CBS3
Dosit River
0 10 20 km
Legend
Sampling site
River
Elevation (m)
1000 - 1050
1050 - 1100
1150 - 1150
1200 - 1250
1250 - 1300
1350 - 1400
1450 - 1450
1500 - 1550
1550 - 1600

Figure 2

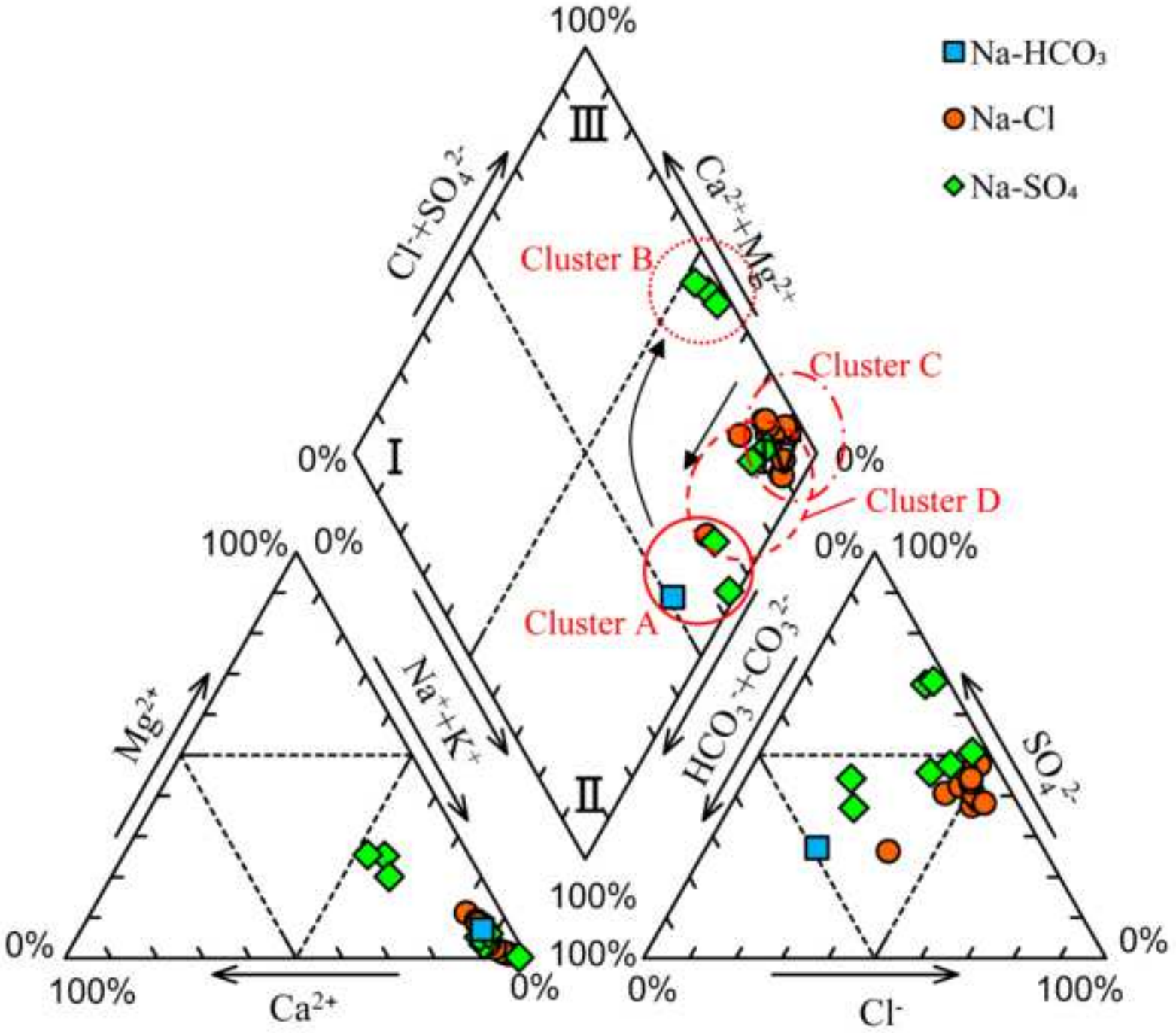

100%
Na-HCO3
Na-Cl
Na-SO4
III
Cl-+SO4 2-
Ca2++Mg2+
Cluster B
Cluster C
Cluster D
Cluster A
0%
I
II
100%
0%
Mg2+
Na++K+
HCO3-+CO3 2-
SO4 2-
Ca2+
Cl-

Figure 3

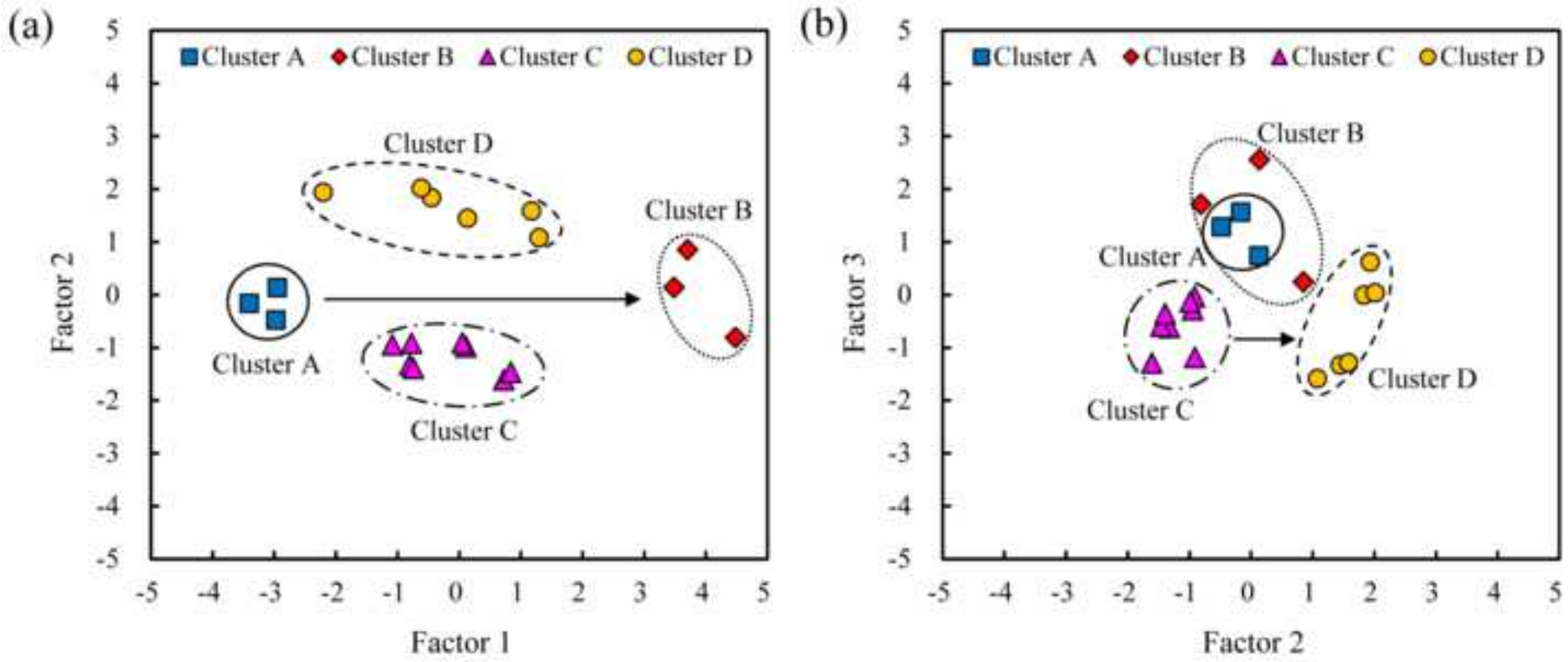

(a)
Cluster A
Cluster B
Cluster C
Cluster D
Factor 2
Factor 1
(b)
Cluster A
Cluster B
Cluster C
Cluster D
Factor 3
Factor 2

Figure 4

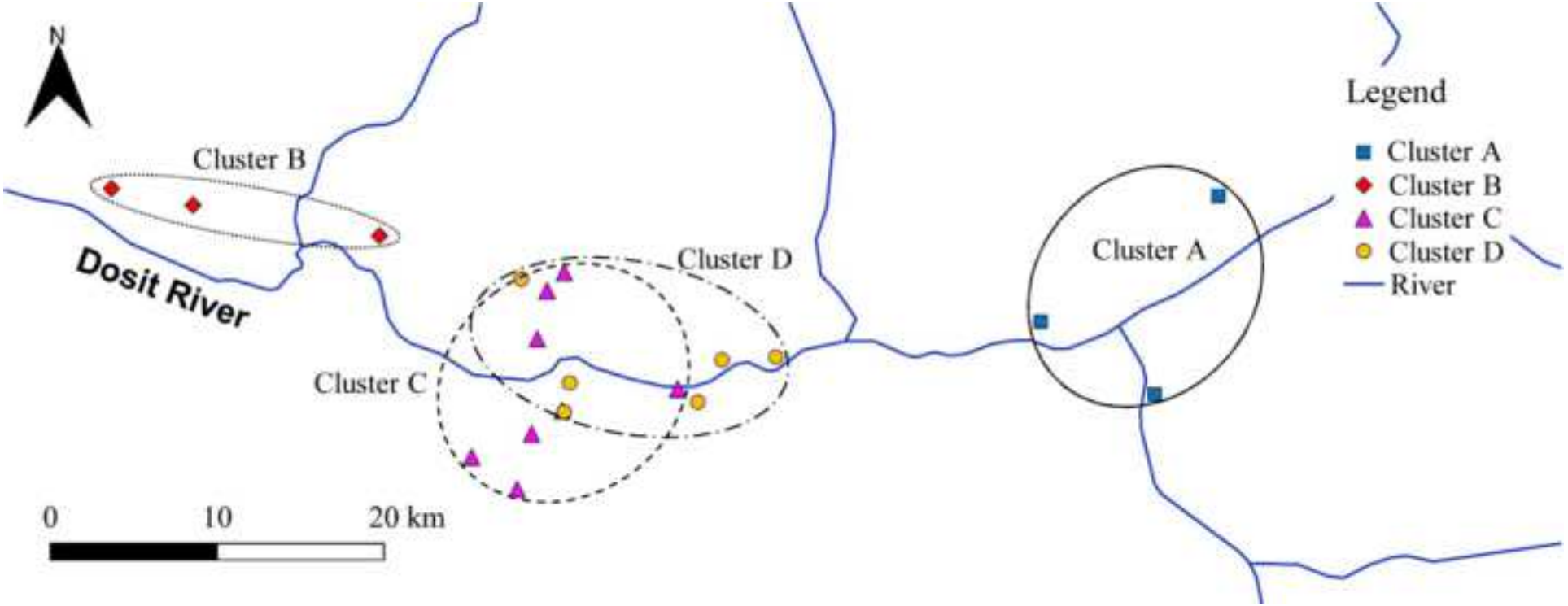

N
Cluster B
Dosit River
Cluster D
Cluster C
Cluster A
0
10
20 km
Legend
Cluster A
Cluster B
Cluster C
Cluster D
River

Figure 5

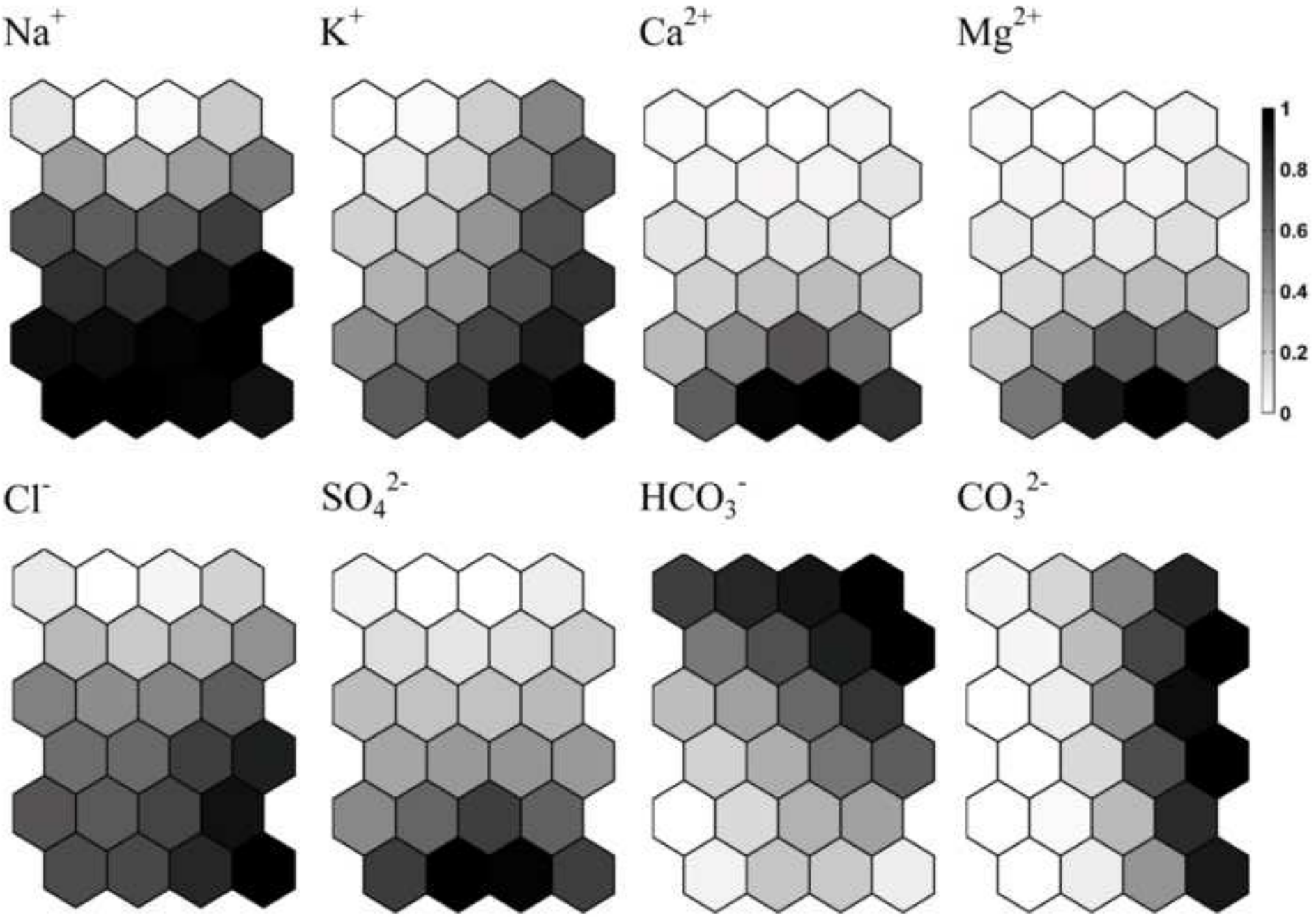


Figure 6

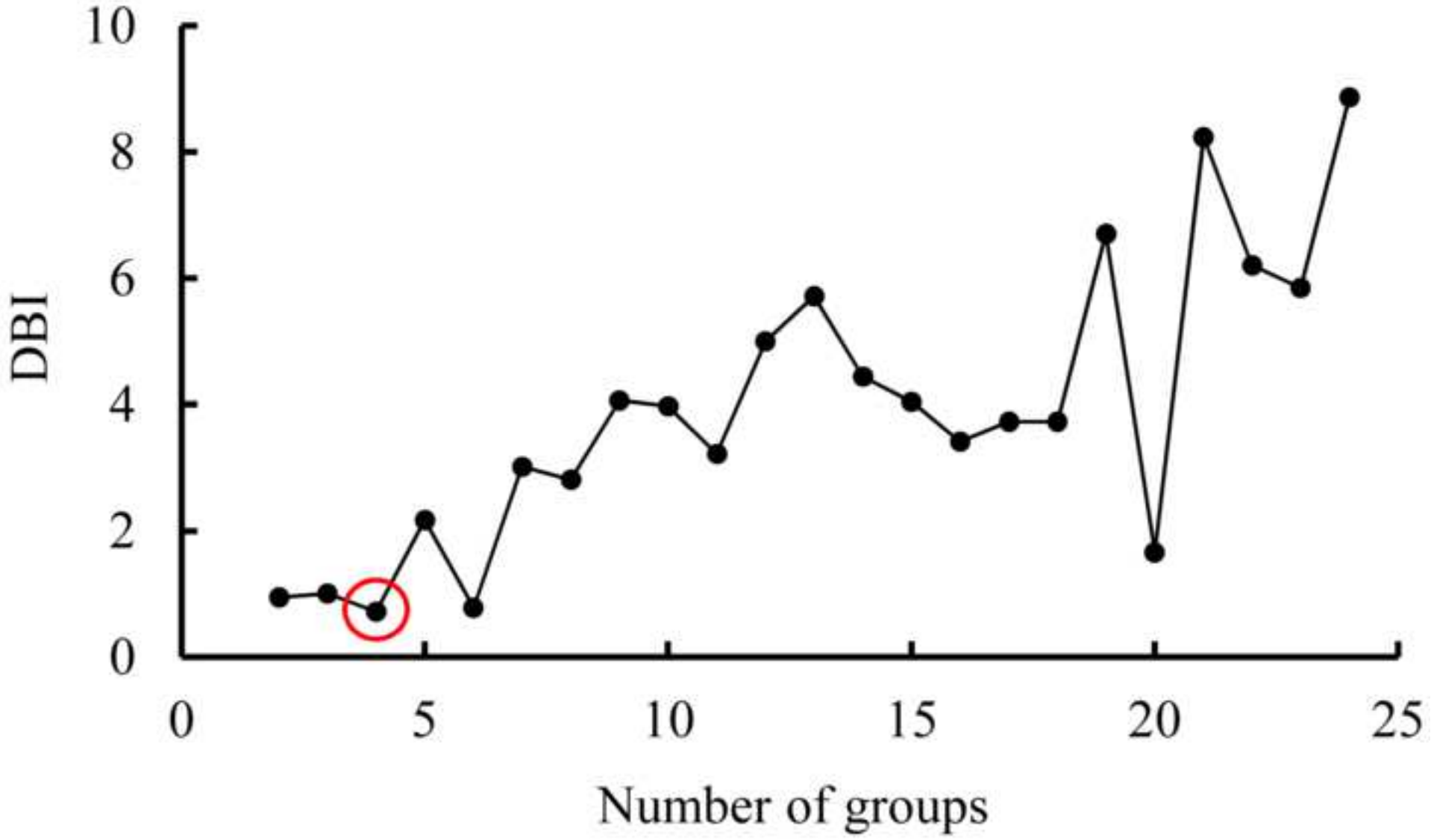
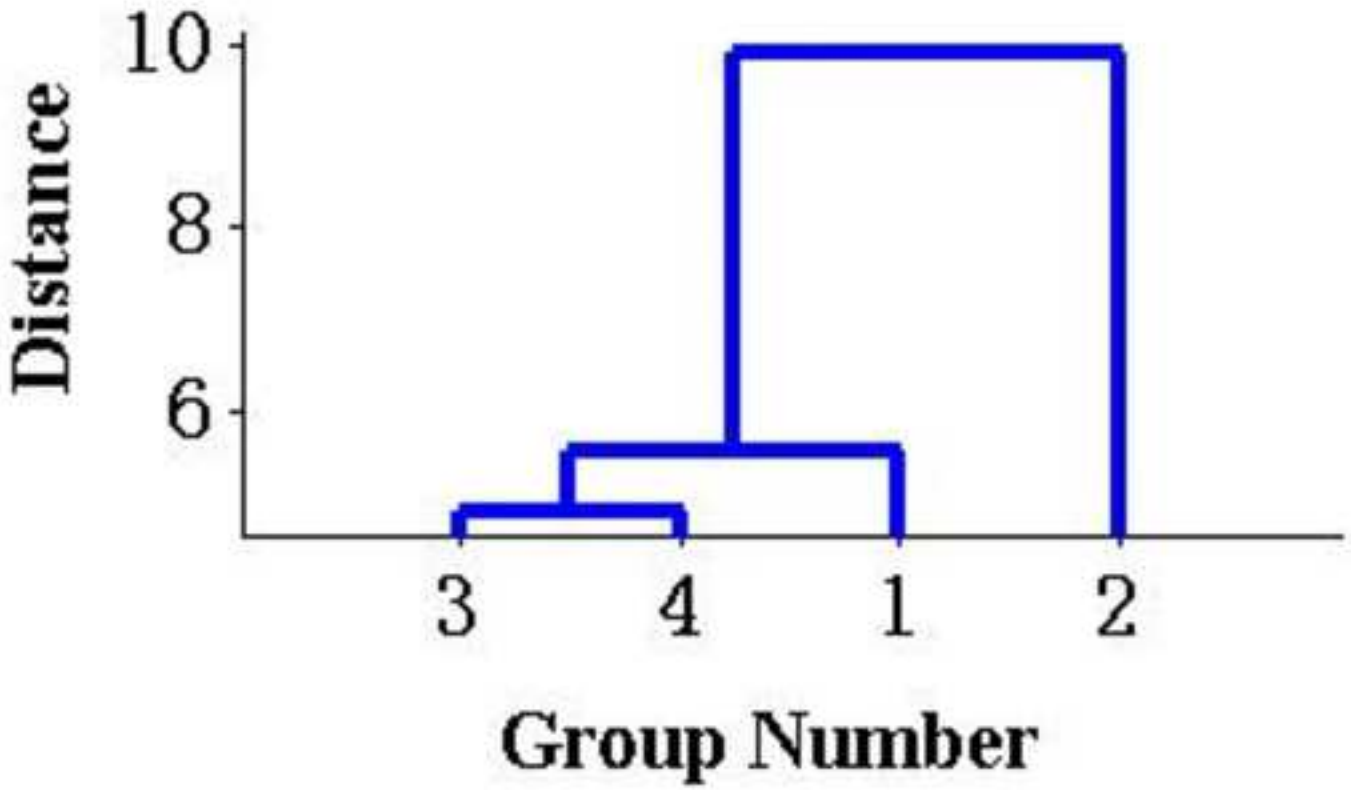


Figure 7



3	14	1	6
4	16	2	11
5	19	7	12
9	20	8	17
10	21	13	18
15	22		23
			24

Figure 8

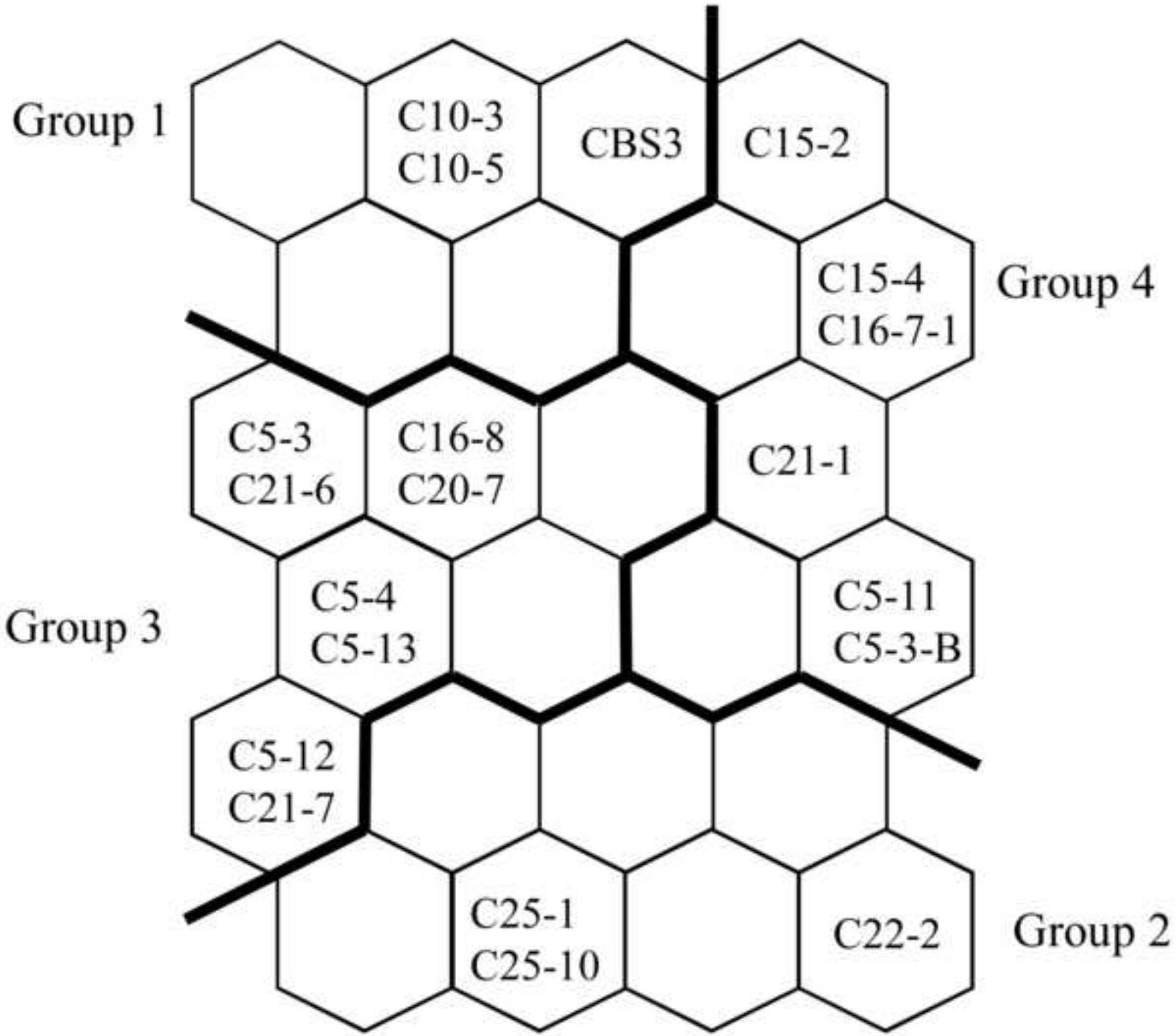
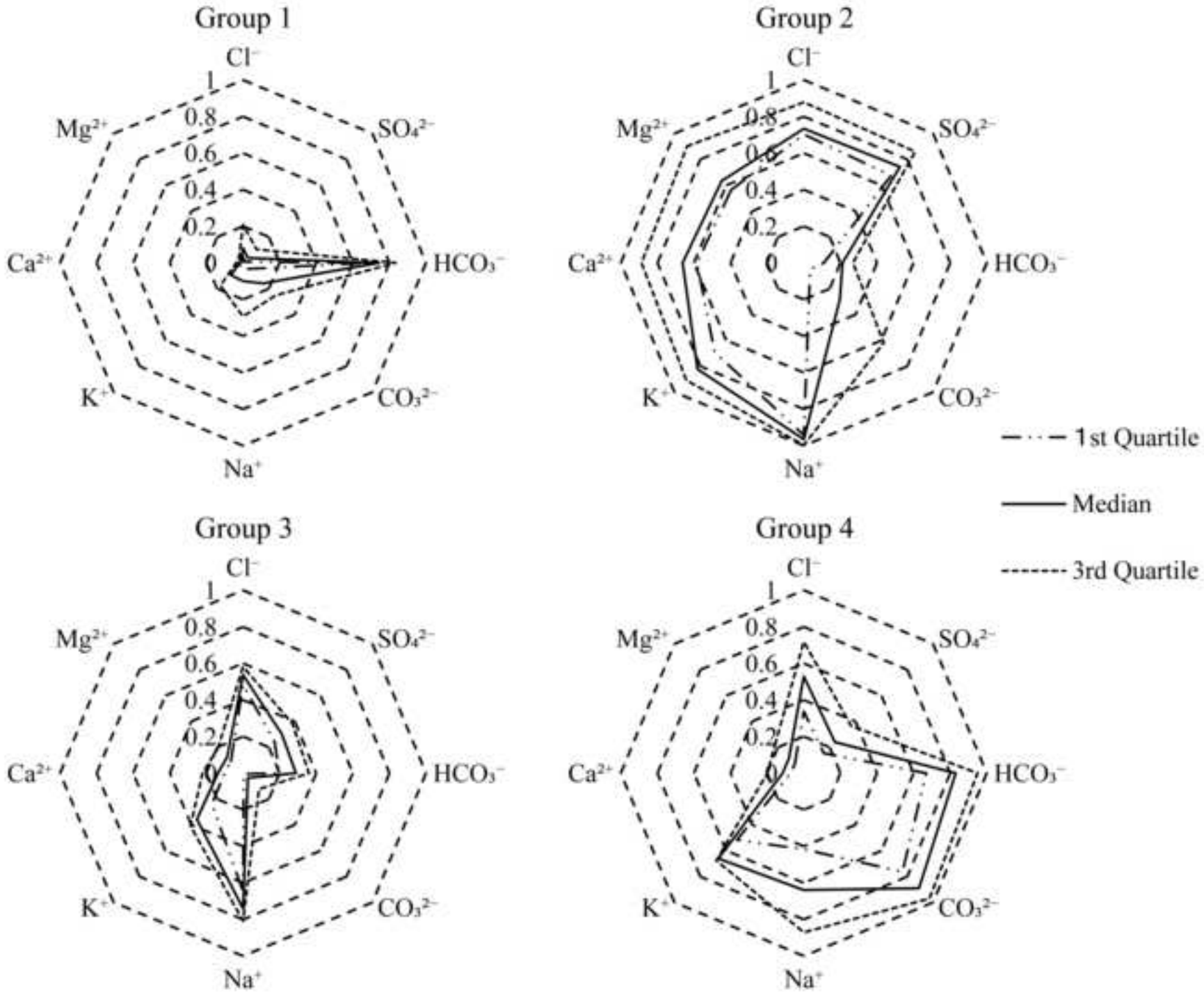

Group 1
C10-3
C10-5
CBS3
C15-2
C15-4
C16-7-1
Group 4
C5-3
C21-6
C16-8
C20-7
C21-1
Group 3
C5-4
C5-13
C5-11
C5-3-B
C5-12
C21-7
C25-1
C25-10
C22-2
Group 2

Figure 9


Group 1
Group 2
Group 3
Group 4
Cl⁻
SO₄²⁻
HCO₃⁻
CO₃²⁻
Na⁺
K⁺
Ca²⁺
Mg²⁺
1
0.8
0.6
0.4
0.2
0
1st Quartile
Median
3rd Quartile

Table

Table 1 Statistical summary of water samples (eight parameters, pH, TDS)

	Na^+	K^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	HCO_3^-	CO_3^{2-}	TDS	pH
	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	
Maximum	812.8	7.0	235.0	154.0	813.3	2072.3	243.5	12.6	3973.9	9.0
Minimum	156.7	0.9	3.0	0.2	65.4	100.4	73.4	0.0	454.7	8.0
Mean	543.7	3.8	56.5	36.4	466.8	698.8	157.1	4.7	1889.3	8.4
SD	200.6	2.0	67.2	48.9	226.1	491.2	48.8	6.1	913.6	0.3

SD means standard deviation.

n = 20

Table 2 Results of principal component analysis

	Components		
	Factor 1	Factor 2	Factor 3
Na^+	0.80	-0.14	-0.51
K^+	0.81	0.53	-0.05
Ca^{2+}	0.86	-0.01	0.50
Mg^{2+}	0.85	0.06	0.49
Cl^-	0.83	0.02	-0.52
SO_4^{2-}	0.94	-0.1	0.25
HCO_3^-	-0.53	0.66	0.35
CO_3^{2-}	0.07	0.93	-0.28
Eigen values	4.62	1.61	1.29
% of variance	57.8	20.1	16.1
Cumulative %	57.8	77.9	94.0

Table 3 Correlation coefficients among 8 physicochemical parameters

	K^+	Ca^{2+}	Mg^{2+}	Cl^-	SO_4^{2-}	HCO_3^-	CO_3^{2-}
Na^+	0.76**	0.72**	0.7**	0.96**	0.85**	-0.76**	0.03
K^+		0.78**	0.81**	0.86**	0.79**	-0.34	0.54**
Ca^{2+}			0.99**	0.74**	0.97**	-0.66**	-0.03
Mg^{2+}				0.76**	0.95**	-0.63**	0.05
Cl^-					0.83**	-0.68**	0.23
SO_4^{2-}						-0.75**	-0.07
HCO_3^-							0.51*

* Correlations significant at $p = 0.05$

** Correlations significant at $p = 0.01$

$n = 20$

Table 4 Mean values of 8 parameters for 4 groups and whole data

	Na^+ mg/L	K^+ mg/L	Ca^{2+} mg/L	Mg^{2+} mg/L	Cl^- mg/L	SO_4^{2-} mg/L	HCO_3^- mg/L	CO_3^{2-} mg/L
Group 1	184.9	1.2	9.5	3.7	75.8	154.2	200.9	2.2
Group 2	670.5	6.5	207.7	146.6	653.6	1664.6	125.3	4.2
Group 3	609.6	2.8	33.7	17.6	500.9	634.6	122.4	0
Group 4	571.8	5.1	34.7	22.7	523.3	573.9	197.2	12.6
Whole data	543.7	3.8	56.5	36.4	466.8	698.8	157.1	4.8