Clustering spatio–seasonal hydrogeochemical data using self-organizing maps for groundwater quality assessment in the Red River Delta, Vietnam

Thuy Thanh Nguyen [a,*], Akira Kawamura [a], Thanh Ngoc Tong [b], Naoko Nakagawa [a], Hideo Amaguchi [a], Romeo Gilbuena Jr. [a]

[a] Department of Civil and Environmental Engineering, Tokyo Metropolitan University, 1-1 Minami-Oshawa, Hachioji, Tokyo 192-0397, Japan
[b] National Center for Water Resource Planning and Investigation, Viet Nam

ARTICLE INFO

SUMMARY

The Red River Delta (RRD) is the second largest delta in Vietnam, and its local communities depend on groundwater sources for water supply. A clear understanding of the groundwater hydrogeochemical properties, particularly their changes from the dry to rainy seasons and spatial characteristics, is invaluable and indispensable for the management and protection of this important water resource. In this study, self-organizing maps was systematically applied for the first time to investigate the seasonal and spatial hydrogeochemical characteristics of groundwater in the Pleistocene confined aquifer of the RRD. The hydrogeochemical characteristics clustered by SOM were further examined using the Gibbs Diagrams. The groundwater chemistry dataset used in the analysis comprised eight major dissolved ions (i.e., Ca^{2+}, Mg^{2+}, Na^+, K^+, HCO_3^-, Cl^-, SO_4^{2-}, and CO_3^{2-}) and total dissolved solids that were collected from 52 groundwater monitoring wells within the study area during the dry and rainy seasons. Based on the results, the hydrogeochemical groundwater data of the confined aquifer monitoring wells for the delta were classified into 8 clusters, which revealed three basic representative water types: high salinity (2 clusters), low salinity (3 clusters), and freshwater (3 clusters). The high-salinity types were located in the middle-stream and coastal areas of the RRD, while the low-salinity types were observed near the western and northeastern boundaries of the delta. Cluster changes from the dry to rainy seasons were detected in approximately one-third of the observation wells. The increase in groundwater recharge during the rainy season is the main reason for these changes. Based on Gibbs diagrams, the source of soluble ions in the groundwater of the freshwater types was found to be the weathering of rock-forming minerals, while evaporation and marine activities (leaching from salty paleowater and salt water intrusion) were found to be the main factors affecting the chemistry of the groundwater characterized by the low- and high- salinity types, respectively.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Clustering is an unsupervised method of data grouping using a given measure of similarity. Clustering algorithms attempt to organize unlabeled feature vectors into clusters (natural groups) such that samples within a cluster are similar to each other but differ from those in other clusters (Hilario and Ivan, 2004). Clustering analysis is an important and useful tool for analyzing large datasets that contain many variables and experimental units. Therefore, the application of cluster analysis to complex datasets has attracted a high level of scientific interest in various aspects of water research, such as surface water (Hall and Minns, 1999), rainfall (Astel et al., 2004), and water quality (Alberto et al., 2001; Vialle et al., 2011).

In hydrogeochemical studies, cluster analysis serves the purpose of isolating a group of representative clusters (also known as water type or hydrogeochemical facies) that reflect the processes generating the natural variability found in hydrogeochemical parameters. These representative clusters, which help define the major chemical trends, can provide insight into aquifer heterogeneity and connectivity, as well as the physical and chemical processes controlling water chemistry (Güler and Thyne, 2004). A number of studies have been published during the past few decades that investigate hydrogeochemical characteristics of

groundwater by applying cluster analysis, e.g., from Europe (Lambrakis et al., 2004), Africa (Belkhiri et al., 2011; Hussein, 2004), and Asia (Zhang et al., 2012; Reghunath et al., 2002). Our study area is the Red River Delta (RRD) in Vietnam, where hydrogeochemical facies, an important diagnostic chemical aspect of groundwater solutions occurring in hydrologic systems, has not been examined adequately by cluster analysis in the assessment of groundwater quality.

The RRD is the second largest delta in Vietnam with an area of about 13,000 km^2 and a population of around 20 million people in 2012 (23% of Vietnam's total population), which makes it one of Vietnam's most densely populated regions (Vietnam General Statistic Office, 2013). All its residents depend entirely on groundwater for their domestic water supply. Due to the importance of groundwater in the RRD as well as the region's importance in the development of Vietnam, in recent years several studies on groundwater have been carried out. For example, Tran et al. (2012) investigated the origin and extent of fresh groundwater, salty paleowaters, and saltwater from recent seawater intrusions in the RRD using geological observations, geophysical borehole logging, and transient electromagnetic methods. Arsenic pollution of groundwater in the entire RRD has been studied by Winkel et al. (2011) on the basis of a complete geo-referenced database with 37 chemical parameters from several hundred wells. In our earlier studies, we investigated the spatial characteristics of the aquifer system (Bui et al., 2011) as well as groundwater level trends in the entire RRD (Bui et al., 2012). The hydrogeochemical characteristics of groundwater in the two main aquifers of the RRD were also investigated by analyzing the physicochemical data from 31 conjunctive wells using classical hydrogeological and hydrochemical approaches (Piper and Gibbs diagrams) (Nguyen et al., 2014). Piper diagram was valuable in pointing out features of analyses of the hydrogeochemical data, but did not suffice to investigate the intrinsic relationships of the data in the RRD. Therefore, it is necessary to apply the methods of clustering analysis for the hydrogeochemical data in order to achieve a better understanding of the physical and chemical properties of the groundwater system in space as well as in time (Subyani and Al Ahmadi, 2009).

The hydrogeochemical characteristics in the RRD can be affected by the change in seasons; hence, investigation of the changes in the hydrogeochemical properties from the dry to the rainy seasons (or vice versa) may reflect the groundwater hydrodynamics and circulation (Nguyen et al., 2014).

In order to investigate the spatial–seasonal hydrogeochemical characteristics of groundwater, it is essential for a robust classification scheme to cluster water chemistry samples into homogeneous groups (Güler and Thyne, 2004). Several common clustering techniques have been utilized to divide groundwater samples into similar homogeneous groups (each representing a hydrogeochemical facies) with the ultimate objective of characterizing the quality of groundwater. For example, Belkhiri et al. (2011) adopted principal component analysis and Q-mode hierarchical cluster analysis to assess the chemistry of groundwater and identify the geological factors that affect the water chemistry in the east of Algeria. Güler and Thyne (2004) applied the fuzzy c-means clustering technique to a large hydrochemical dataset from the Indian Wells-Owens Valley area of southeastern California to delineate clusters of water samples with similar characteristics. Reghunath et al. (2002) applied Q- and R-mode factor and cluster analysis to improve the understanding of groundwater systems in Karnataka, India. These methods are efficient at grouping water samples by chemical similarities, but are not useful for the visual assessment of the results and presentation of maps showing hydrogeochemical facies (Güler et al., 2002). The recently proposed method of the self-organizing maps (SOM) is likely to become a complementary or alternative tool to the clustering methods (Kalteh et al., 2008; Iseri et al., 2009).

The SOM is based on an unsupervised learning algorithm, and has excellent visualization capabilities, including techniques that use the reference vectors of the SOM to give an informative picture of the data (Hong et al., 2003). The SOM has been implemented in various aspects of hydrology, e.g., identification of homogeneous regions for regionalization of watersheds (Farsadnia et al., 2014), regional flood frequency analysis (Srinivas et al., 2008), and regionalization of hydrological model parameters (Wallner et al., 2013). The SOM has also proven to be a powerful and effective data analysis tool in meteorological analysis and detection of long-term changes in climate (Nishiyama et al., 2007; Leloup et al., 2007). However, the SOM has not yet been systematically applied for the classification of groundwater quality samples in order to investigate hydrogeochemical characteristics. This study is the first attempt to apply the SOM in combination with a hierarchical cluster analysis for clustering hydrogeochemical groundwater data.

Through the initiative of the national government (National Hydrogeological Database Project), hydrogeochemical data of the Pleistocene confined aquifer in the RRD were collected in 2011 during the dry and rainy seasons. The objective of this study is to cluster spatial–seasonal hydrogeochemical data to assess the groundwater quality of the confined aquifer in the RRD using SOM and Gibbs diagrams. In this study, Gibbs diagrams were aptly used to elucidate the cause and significance of the hydrogeochemical characteristics clustered by the SOM. Gibbs (1970) proposed chemical diagrams for the assessment of functional sources of dissolved chemical constituents and for inferring the mechanism controlling the chemistry of surface water. Various researchers have already demonstrated the usefulness of Gibbs diagrams for groundwater (Raju et al., 2011; Marghade et al., 2012; Yidana et al., 2010). The findings from this study will provide valuable insights into the spatial–seasonal hydrogeochemical characteristics of groundwater in the Pleistocene confined aquifer of the RRD.

2. Materials and methods

2.1. Study area

Fig. 1 shows the geographical location of the study area (the entire RRD) and the 52 groundwater observation wells for the confined aquifer. In order to facilitate investigation of spatial hydrogeochemical characteristics, the RRD was divided into three zones: upstream, middle-stream, and downstream by two lines, AA′ and BB′, as shown in Fig. 1. The two lines are the lines connecting boreholes of two typical hydrogeological cross-sections, which were created in our previous study (Bui et al., 2011). Well Nos. 1–15 and 32–50 were in the upstream area; Well Nos. 16–24, 51, and 52 were in the middle-stream area; and Well Nos. 25–31 were in the downstream area. The RRD is the most-developed region in Vietnam and comprises 11 provinces and cities (Fig. 1). It has a population density of more than five times the national average. Two of Vietnam's major economic centers, Hanoi and Hai Phong, are located in the RRD (Bui et al., 2012).

The RRD is situated in the tropical monsoonal region with two distinct seasons: rainy (May–October) and dry (November–April). The annual average rainfall is about 1600 mm, 75% of which occurs during the rainy season. The annual average humidity is about 80%, and the average temperature is 24 °C. The annual evaporation average is approximately 900 mm. The river network is quite extensive, with a network density of about 0.7 km/km^2 (Bui et al., 2011). The average discharge of the Red River at the Hanoi station is 1160 m^3/s during the dry season and 3970 m^3/s during the rainy season
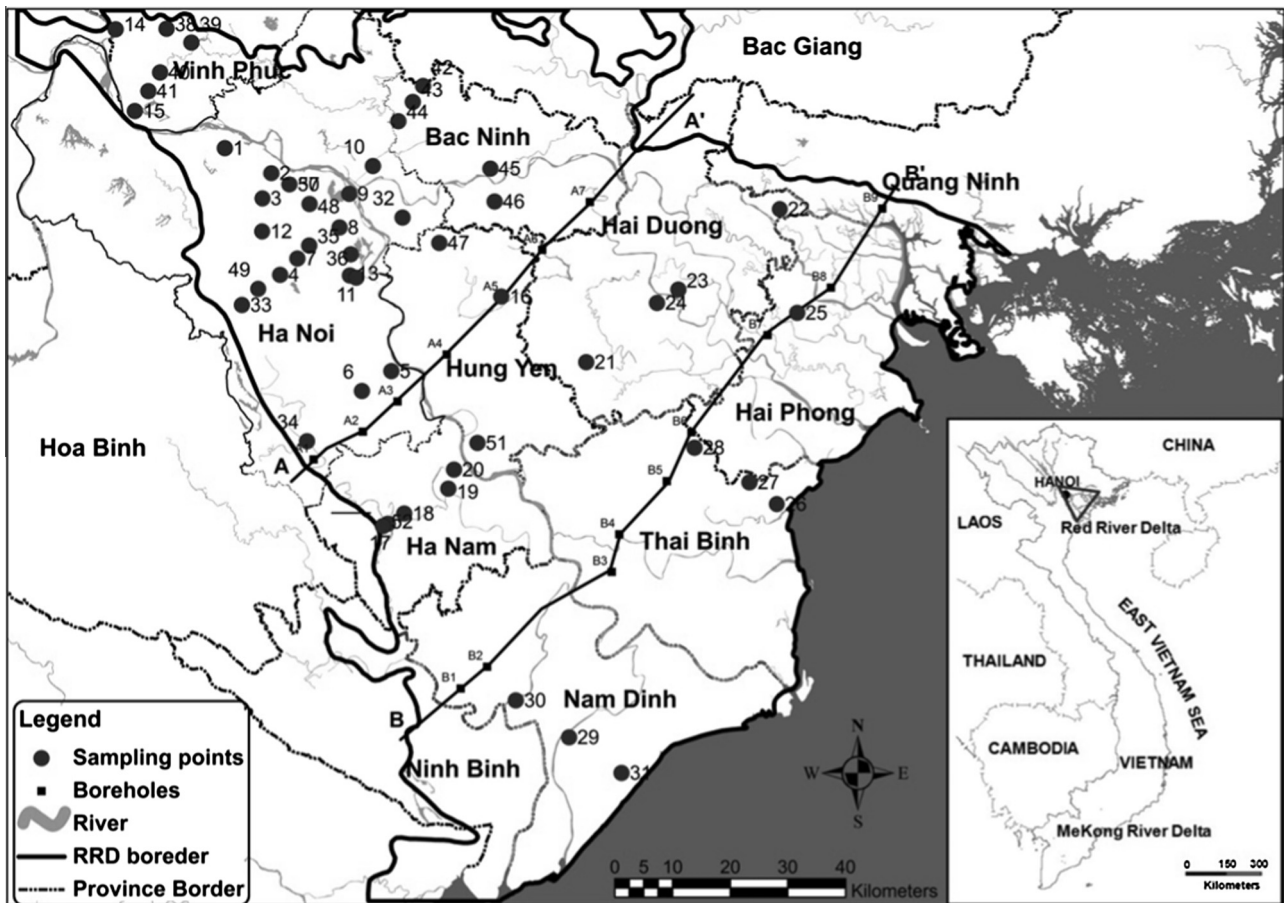
Fig. 1. Study area and location of sampling points.

(IMHE-MONRE, 2011). In the Red River, a high concentration of suspended solids is always present, which gives it its "reddish" color. The tidal range along the coast is approximately 4 m. The lakes, ponds, and canals in highly urbanized areas are seriously polluted with untreated domestic and industrial wastewater. The groundwater, being relatively cleaner and generally unaffected by the surface environmental problems, has become the most-trusted freshwater source in the RRD (Bui et al., 2011).

In terms of regional geology, the RRD is composed of Quaternary-aged unconsolidated sediments with the thickness ranging from a few meters in the northwest to 150–200 m at the coastline in the southeast (Tran et al., 2012). In our previous study (Bui et al., 2011), five hydrogeological cross-sections were identified by hydrostratigraphically interpolating strata data from a number of well logs to demonstrate the vertical framework of the aquifer system. From these cross-sections, we found that the groundwater mostly exists as porous water that forms the topmost Holocene unconfined aquifer and the Pleistocene confined aquifer, with the latter serving as the highest groundwater potential and most important aquifer for water supply. Thus, the Pleistocene confined aquifer was the focus of this study. The confined aquifer consists of sands mixed with cobbles and pebbles, and is situated below the Holocene unconfined aquifer in the stratigraphic sequence. The thickness of this aquifer fluctuates over a large range with an average of about 80 m, and gradually increases from the northwest to southeast of the delta. The transmissivity ranges from 700 to 3000 m²/day and indicates a very high potential of groundwater resources (Bui et al., 2011).

2.2. Data used

The RRD has the most extensive hydrogeochemical database in Vietnam with a large number of data owners such as the Vietnamese geological survey departments, local or national environmental agencies, public and private institutions, consultant firms and many others. However, the record lengths and intervals vary greatly depending on the completion time and the intended usage of the observation wells, as well as the aquifers and variables that are being monitored. In this study, we used the most recent groundwater chemical data from the National Hydrogeological Database Project (Tong, 2004), which were collected from 52 observation wells in the confined aquifer (Fig. 1) in the months of February (dry season) and August (rainy season) in 2011 to investigate the hydrogeochemical characteristics of groundwater in the RRD.

Sampling was done in accordance to the guidance on the sampling, preservation and handling of groundwater samples of Ministry of Natural Resources and Environment (MONRE, 2008). All samples were filtered with 0.45-μm filter membranes and collected in clean and dry Polyethylene or Polytetrafluoroethylene plastic bottles. To take account of any physicochemical change that might take place, all field-based water parameters such as temperature and pH were measured in situ. Chemical analyses were undertaken at the laboratory of Analytical Chemistry Department, Vietnam Academy of Science and Technology, following the national technical regulation on underground water quality of MONRE (2008).

In our previous study (Nguyen et al., 2014), the Piper diagram was used to classify the major ions in the groundwater into various hydrogeochemical types to investigate and identify the hydrogeochemical facies of groundwater in the RRD. In this study, we also used the same kind of the chemical data, consisting of major cations (Ca^{2+}, Mg^{2+}, Na^+, and K^+) and major anions (HCO_3^-, Cl^-, SO_4^{2-}, and CO_3^{2-}) to cluster and characterize the groundwater quality. Total dissolved solids (TDS) data were used for Gibbs diagram to validate the significance of the hydrogeochemical characteristics clustered by the SOM.

Standardization of the data was necessary prior to the application of the SOM to ensure that all values of the chemical parameters were given the same or similar importance. The results of the SOM application were sensitive to the data pre-processing method used, as the SOM is trained to be organized according to the Euclidean distances between input data (Jin et al., 2011). In this study, the range of the standardized values of the hydrogeochemical data for all parameters was 0–1.

2.3. Methods

The SOM, developed by Kohonen (2001), is one kind of artificial neural network that is characterized by unsupervised training. It can project high-dimensional, complex target data onto a two-dimensional, regularly arranged map in proportion to the degree of similarity (Jin et al., 2011). In other words, the SOM accomplishes two objectives simultaneously: reducing dimensions and displaying similarities. Therefore, it is an effective tool to visualize and explore data properties. In general, the objective of the SOM application is to obtain useful and informative reference vectors (also referred to as weight vectors, prototype vectors, and codebook vectors (Hilario and Ivan, 2004)). These vectors can be acquired after iterative updates through the training of the SOM, which is composed of three main procedures: competition between nodes, selection of a winner node, and update of the reference vector of each node.

Design of the SOM structure (calculation of the total number of nodes, side lengths), selection of a proper initialization method, and data transformation methods are very important features in the SOM application. The number of map nodes determines the accuracy and generalization capability of the SOM. According to the properties of the SOM, the bigger the map size is, the higher the resolution for pattern recognition, while the topographical adjacency is further among the clusters. A reasonable optimum solution of the compromise among the accuracy of pattern classification and topographical proximity of clusters to determine the number of the SOM nodes is the heuristic rule of $m = 5\sqrt{n}$, with m denoting the number of the SOM nodes and n representing the number of input data (Vesanto et al., 2000; Jeong et al., 2010; Hentati et al., 2010; Jin et al., 2011). In this study, this heuristic formula was used to determine the total number of nodes in the SOM. The ratio of the number of rows and columns was calculated by the square root of the ratio between the two biggest eigenvalues of the transformed data (Hilario and Ivan, 2004).

After establishing the SOM structure, reference vectors for the SOM with the commonly used hexagonal array are initially set using the linear initialization method. The method first determines the two eigenvectors of the autocorrelation matrix of input vectors that have the largest eigenvalues, and then to let these eigenvectors span a two-dimensional linear subspaces. A rectangular array is then defined along this subspace, in which its center coincides with that of the mean of the input vectors, and the dimensions are the same as the two largest eigenvalues. The initial values of the weight vectors are then identified with the array points (Su et al., 2002). In this study, due to limited data, the linear initialization method was used, as it is more suitable for the pattern

classification than the random initialization. The random initialization requires a large dataset and might cause boundary effects near the edges of the map (Vesanto et al., 2000; Hentati et al., 2010; Jin et al., 2011). In addition, the linear initialization approach can use eigenvalues and eigen vectors of the input data to set the initial reference vectors on the structured SOM. This means that the initial reference vectors already include prior information about the input data, resulting in an acceleration of the training phase (Vesanto et al., 2000; Jin et al., 2011). In this study, each reference vector was updated through the SOM training process using the batch mode. The reference vectors obtained at the end of the training process can be fine-tuned using cluster analysis methods.

Various clustering algorithms are available in literature. These algorithms are generally classified into two types: hierarchical clustering and partitional clustering. These two clustering types can be integrated such that a result given by a hierarchical method can be improved via a partitional step, which refines via interactive relocation of points (Hilario and Ivan, 2004). In this study, both clustering algorithms were applied for the fine-tuning of the reference vectors. For the partitional clustering methods, the k-means algorithm is the most frequently used method for the SOM (Jin et al., 2011; Hentati et al., 2010; Nishiyama et al., 2007; Hilario and Ivan, 2004). The optimal number of clusters was selected by the Davies–Bouldin index (DBI) using the k-means algorithm. The DBI values were calculated from a minimum of 2 clusters to the total number of nodes. The calculation was based on the "similarity within a cluster" and "dissimilarity between clusters." Therefore, the number of clusters showing the minimum DBI was optimal for the trained SOM (Hilario and Ivan, 2004; Nishiyama et al., 2007). For the hierarchical method, Ward's linkage method is the most commonly used approach (Jin et al., 2011; Hentati et al., 2010). In this study, a final fine-tuning cluster analysis was carried out by Ward's method using the optimal number of clusters.

To investigate the mechanisms governing the groundwater chemistry of the RRD, the chemical diagrams that were proposed by Gibbs (1970) were used to further evaluate the clustered data from the confined aquifer wells. The weight ratios $Na/(Na + Ca)$ and $Cl/(Cl + HCO_3)$ were plotted against the TDS separately on a logarithmic axis to represent the Gibbs cation and anion diagrams, respectively. The Gibbs diagrams were originally used to evaluate surface waters, but recent groundwater quality studies have used these diagrams to assess the sources of dissolved chemical constituents of groundwater (Raju et al., 2011; Marghade et al., 2012; Yidana et al., 2010).

3. Results and discussion

3.1. SOM and clustering results

The input data for the SOM application were concentrations of 8 chemical parameters (major ions as described in Section 2.2) of 104 samples, which were observed in 52 confined aquifer wells during the dry and rainy seasons. Based on the methodology described above, the number of the SOM nodes was calculated as 56, and the numbers of rows and columns were 8 and 7, respectively. This SOM was used for the cluster analysis of the standardized groundwater chemistry monitoring data.

Fig. 2 shows the 8 component SOM maps finally obtained after the training process. Each map represents the component value of the reference vectors for the 56 SOM nodes, in which the reference vectors were standardized to range 0–1, using shades of gray. The nodes that represent the high values are in dark gray and the low values are colored light gray. A comparison between the component SOM maps, by means of a gradient of gray shades, can indicate informative and qualitative relations (or correlations) among
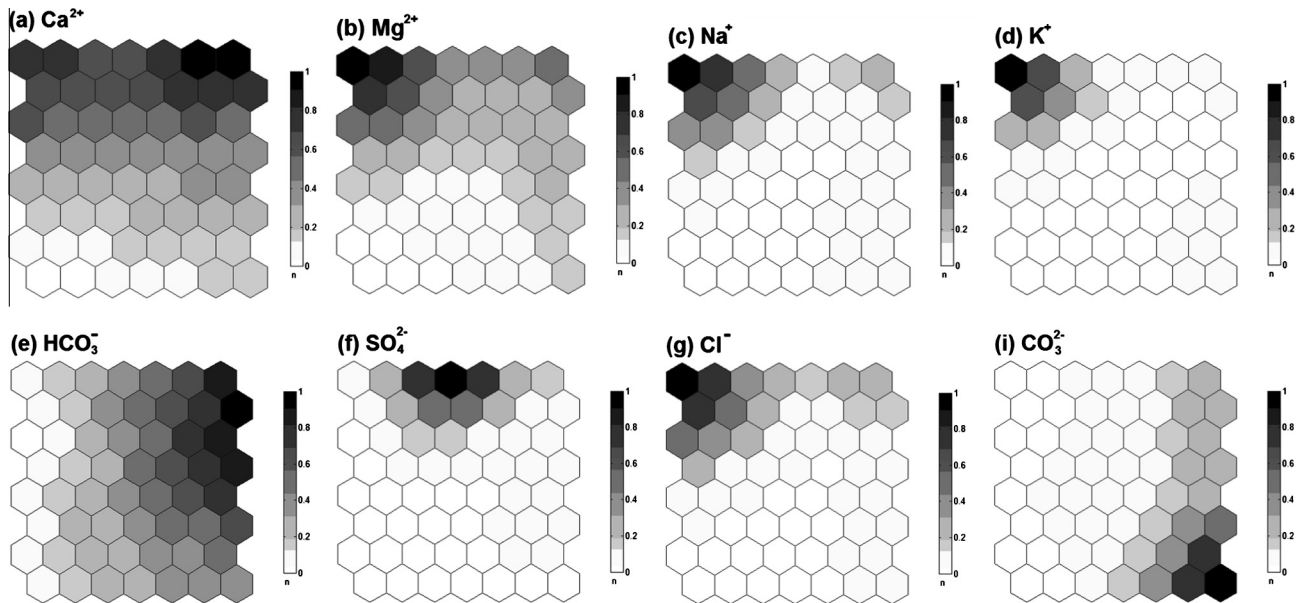
Fig. 2. Component planes for (a) Ca^{2+}, (b) Mg^{2+}, (c) Na^+, (d) K^+, (e) HCO_3^-, (f) SO_4^{2-}, (g) Cl^-, (i) CO_3^{2-}.

the studied parameters. Through visual investigation of SOM maps in Fig. 2, Mg^{2+}, Na^+, K^+, and Cl^- have similar gray gradients. This means that there is strong positive correlation among these 4 parameters. In contrast, CO_3^{2-} showed a negative correlation with these parameters by the inverse gray gradient of the SOM maps. The component maps of HCO_3^- and SO_4^{2-} show weak positive correlation with other parameters, except for Ca^{2+}, as can be seen in Fig. 2. HCO_3^- is mainly dissolved from carbonate rock, which wide spread in the RRD, and combines with any cations by ion exchange reactions. Therefore, there is no specific correlation between HCO_3^- and other ions. The source of Ca^{2+} and SO_4^{2-} are mainly from the dissolution of sulfate minerals (gypsum and anhydrite). Hence, SO_4^{2-} shows higher correlation with Ca^{2+} than other parameters. To quantitatively confirm the strength of relations between these parameters which are shown in Fig. 2, the correlation coefficients among 8 physicochemical parameters were calculate by using the standardized reference vectors, as shown in Table 1. The results show that the relationships between those parameters are consistent with the findings in the qualitative correlations from Fig. 2 as mentioned above. In order to select the optimal number of clusters, the DBI values based on the k-means clustering algorithm were calculated for the minimum (2 clusters) to the maximum (56 clusters) number of possible clusters. Fig. 3 shows the variation of the DBI values after being applied to the data and the front part between 2 and 20 clusters was magnified to show the minimum DBI visibly. The most appropriate number of clusters corresponding to the minimum DBI was 8. Once the optimum cluster number had been selected, the hierarchical clustering algorithm using Ward's method was carried out to obtain the 8 clusters for fine-tuning the pattern classification.

Fig. 4 shows the hierarchical cluster tree with the nodes of the SOM classified into 8 different clusters. The nodes in the SOM map are numbered from top to bottom and from left to right. As shown in this figure, Clusters 4 and 6 have the smallest distance or highest similarity between clusters. This means that Cluster 4 has similar hydrogeochemical characteristics to Cluster 6. In the same way, Clusters 1 and 3 have similar characteristics, as well as Clusters 5 and 8. In addition, Clusters 2 and 7 have higher similarity with Clusters 5 and 8 and Clusters 4 and 6, respectively, than the other clusters, as shown in Fig. 4.

Fig. 5 shows the pattern classification map of the 8 clusters, in which the numbers in the nodes represent Well Nos; the characters D and R correspond to the dry and rainy seasons; and the last characters u, m, and d denote upstream, middle-stream, and downstream areas, respectively. Simultaneous analysis of Fig. 2 (component SOM maps) and Fig. 5 reveals what kind of data the respective clusters include. For example, Cluster 1 (upper left part of Fig. 5) is associated with high-salinity water characterized by high Na^+, K^+, Mg^{2+}, and Cl^-, which is observed in the same location of the respective component SOM maps as shown in Fig. 2. On the other hand, the groundwater samples in nodes with extremely low concentrations of all ions are located at the lower left part of each SOM map (classified as Cluster 6), as shown in Fig. 2.

3.2. Fundamental characteristics of the respective clusters

The reference vector values of each node obtained from the SOM can provide quantitative information. In order to numerically characterize the classified data, the first quartile, median, and third quartile of the reference vector values for the 8 clusters were calculated. Fig. 6 displays the radar charts of the 8 parameters for the 8 clusters with the first quartile, median, and third quartile plotted. As shown in this figure, the visible patterns of Clusters 1 and 3 are similar (as mentioned above for Fig. 4). Both the clusters have the pattern of significantly high values of all cations and Cl^- and very low values of HCO_3^- and CO_3^{2-}. In particular, Cluster 1 with the highest Na^+ and Cl^- values represents the most saline water type of all the clusters. Clusters 4 and 6 have low concentrations of all the major ions, and it can be assumed that the wells in these clusters are of freshwater type. In particular, Cluster 6 with the lowest concentrations of all ions represents the freshest water
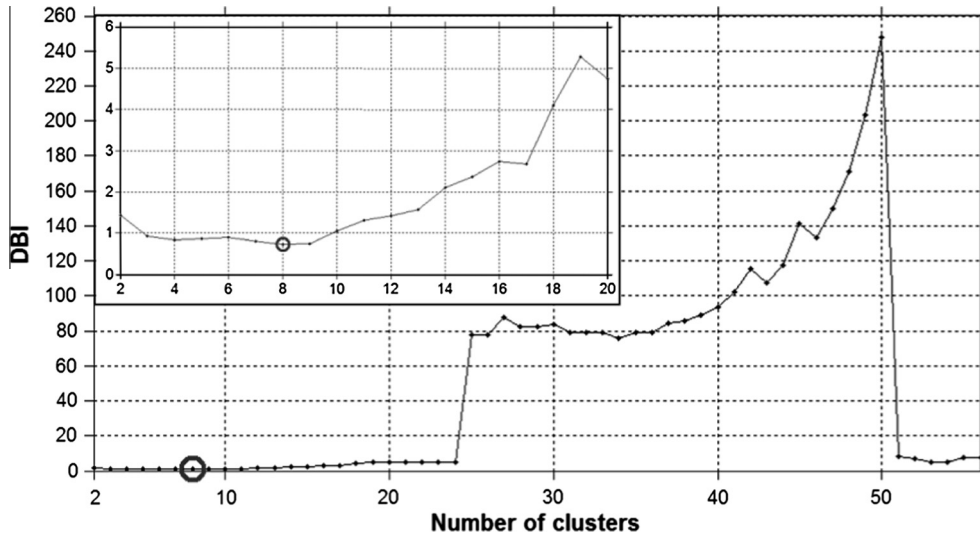
Table 1
Correlation coefficients among 8 physicochemical parameters.

	Ca^{2+}	Mg^{2+}	K^+	Na^+	HCO_3^-	SO_4^{2-}	Cl^-
Mg^{2+}	0.77						
K^+	0.43	0.88					
Na^+	0.56	0.95	0.98				
HCO_3^-	0.37	−0.09	−0.37	−0.30			
SO_4^{2-}	0.53	0.42	0.17	0.27	0.07		
Cl^-	0.59	0.96	0.97	0.99	−0.32	0.28	
CO_3^{2-}	−0.16	−0.19	−0.18	−0.18	0.44	−0.16	−0.23

Fig. 3. Variation of DBI values with the optimal number of clusters marked by the circle on the figure.
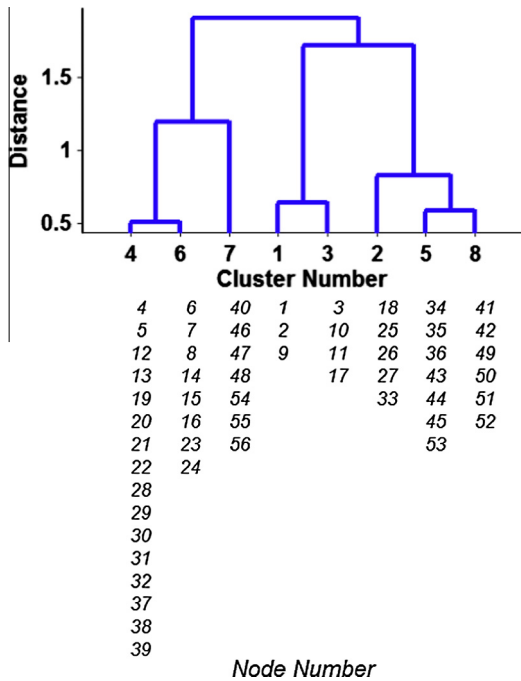

Fig. 4. Dendrogram with node numbers classified into the respective clusters.

type. Clusters 5 and 8 have similar ion patterns with significantly high values of Ca^{2+} and HCO_3^-, in which Cluster 8 has higher concentrations of all major ions than Cluster 5. Cluster 2 is characterized by high Ca^{2+}, SO_4^{2-}, Mg^{2+}, and HCO_3^- and low Na^+, Cl^-, K^+, and CO_3^{2-}, which according to Fig. 4, is close to the water type of Clusters 5 and 8. Cluster 7 shows a pattern where almost all ions have low values except CO_3^{2-} and HCO_3^-, which is close to the freshwater type (Clusters 4 and 6).

As mentioned above, groundwater in the RRD is mainly used for domestic water supply, drinking water and agriculture purposes. Therefore, from the practical point of view the 8 classified clusters should be divided into three main water types based on the similarity observed from the 8 radar charts in Fig. 6, in order to be more generalized and understandable for groundwater users and authorities in Vietnam. The freshwater type was associated with Clusters 4, 6, and 7 due to the low values of Na^+ and Cl^-, as

indicated in the lower part of Fig. 5. Clusters 1 and 3 were characterized by high concentrations of all cations and Cl^- ion, representing the high-salinity type, as seen in the upper left part of Fig. 5. The remaining 3 clusters (Clusters 2, 5, and 8) were characteristic of the low-salinity type, as shown in the upper right part of Fig. 5. In fact, to make the water users and authorities understand more easily, grouping clusters obtained from SOM has also been carried out in other studies (Jin et al., 2011).

In order to confirm and give more quantitative information of each cluster, the mean values of each parameter for the whole and the classified data were calculated from the raw data, as shown in Table 2. Clusters 1 and 3 show considerable higher mean values of cations and Cl^- than those for the whole data. It confirms that the two clusters represent the most saline groundwater in the study area, as mentioned above. Similarly, clusters 4, 6, and 7 have lower mean values of almost all ions, which confirms that these clusters represent the freshwater type. Cluster 8 shows relatively higher values of almost all ions except K^+ than those for the whole data, while Cluster 5 shows slightly lower values. Cluster 2 indicates the highest SO_4^{2-} mean value and slightly higher mean values of Ca^{2+}, Mg^{2+}, HCO_3^-, and Cl^- than those for the whole data.

3.3. Seasonal changes in the respective clusters

Fig. 7 displays the SOM map, in which all the observation wells showing cluster changes from the dry to rainy seasons are indicated. From this figure, it is observed that 16 out of the 52 observation wells exhibited seasonal changes, in which 6 wells (Well Nos. 3, 7, 16, 20, 31, and 33) showed changes of water types and the other 10 wells showed changes within the same water type.

With regard to changes in water types, it is noted that samples from Well No. 31 changed from the high-salinity type (Cluster 1) to the freshwater type (Cluster 7). In fact, Well No. 31 is located in the southern coastal area, as shown in Fig. 1. According to Wagner et al. (2012), in this region there is a constant influx of fresh groundwater from the adjacent mountain coming to the aquifer. The increase of the groundwater recharge during the rainy season may be the reason causing the change in the water type of this well. Furthermore, samples from Well Nos. 16 and 7 changed from Clusters 2 and 8 (the low-salinity type) to Clusters 6 and 4 (the freshwater type), respectively, and Well No. 20 changed from Cluster 3 (the high-salinity type) to Cluster 8 (the low-salinity type). These changes imply that water infiltration from the Holocene
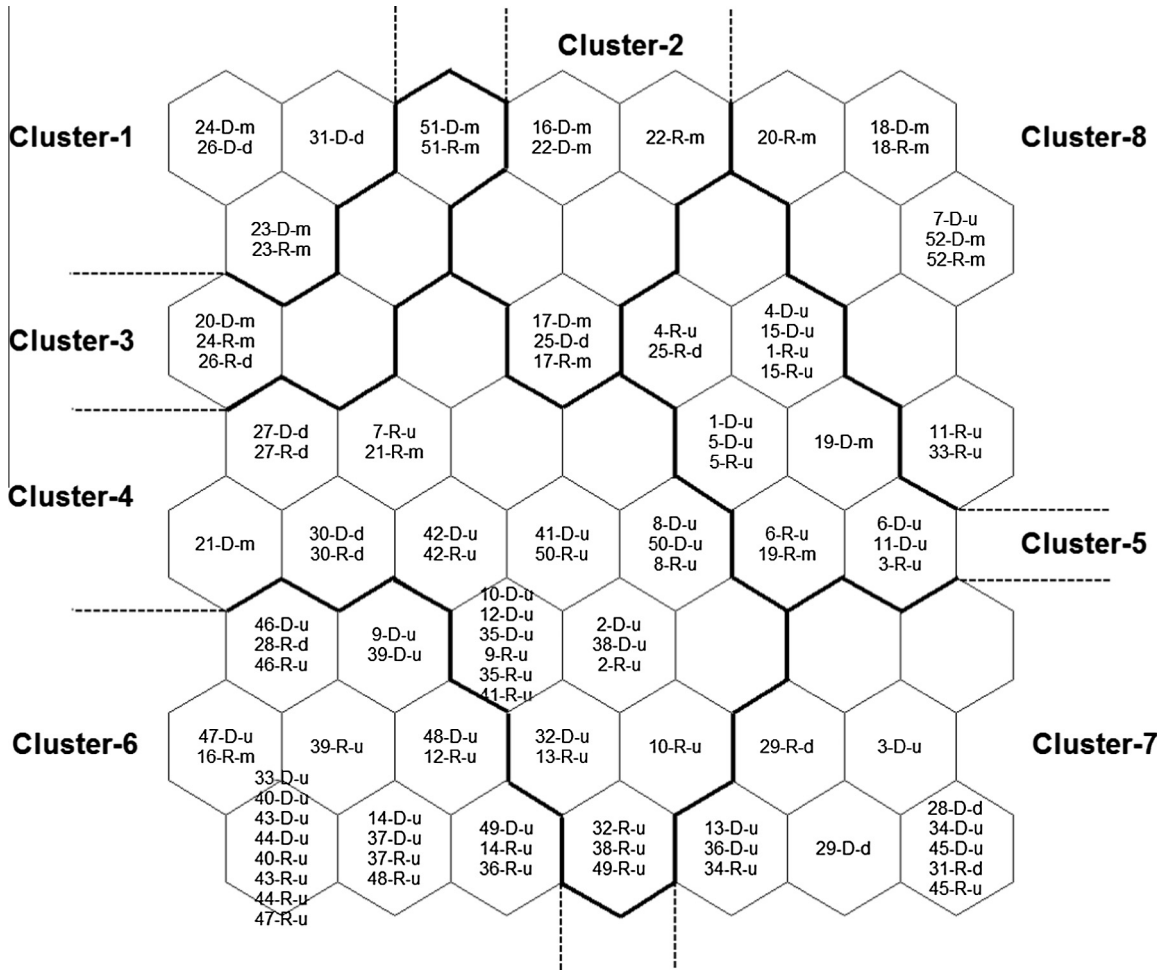
Fig. 5. Pattern classification map of the eight clusters by the SOM. Numbers represent the name of sampling wells in Fig. 1. The characters D and R correspond to the dry and rainy season. The last characters u, m, and d denote upstream, middle-stream, and downstream areas.
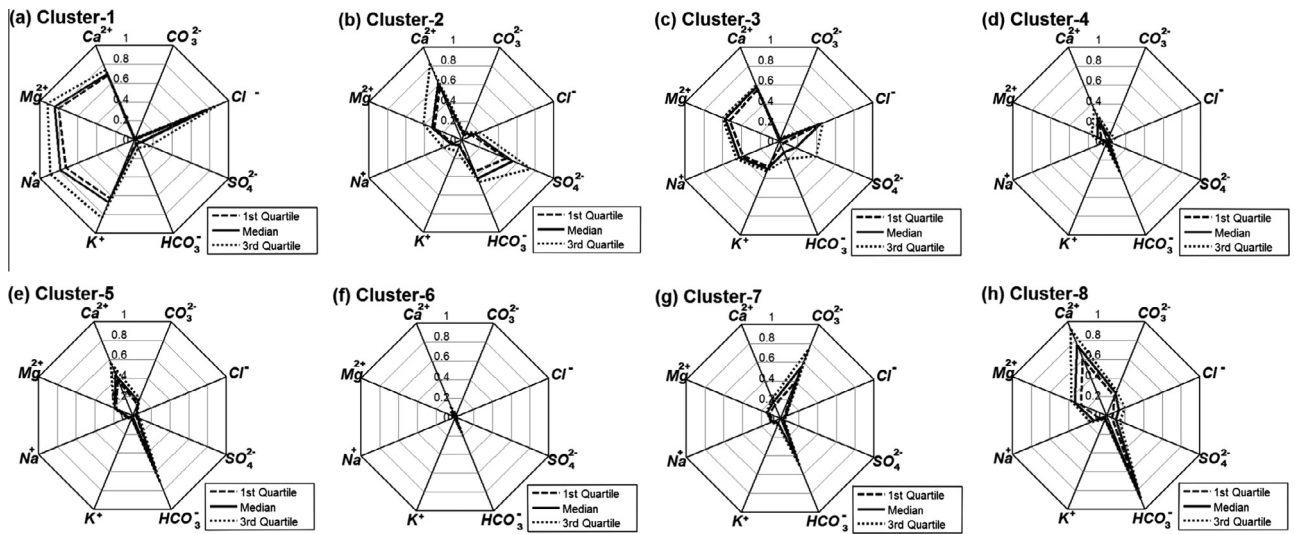

Fig. 6. Radar charts for the respective clusters with the first quartile (dashed lines), median (solid lines) and the third quartile (dotted lines) by obtained reference vectors.

unconfined aquifer may affect the concentrations of chemical constituents of the confined aquifer groundwater during the rainy season through hydrogeological windows, where the aquitard sandwiched by the aquifers is completely missing that directly connect the two aquifer systems (Bui et al., 2012). Interestingly, samples from Well Nos. 3 and 33 located near the western boundary in the upstream area (as shown in Fig. 1) changed from Clusters 7 and 6 (the freshwater type) to Clusters 5 and 8 (the low-salinity

Table 2
Mean values of eight parameters for the five clusters and whole data.

| Mean	Ca^{2+} (mg/L)	Mg^{2+} (mg/L)	Na^+ (mg/L)	K^+ (mg/L)	HCO_3^- (mg/L)	SO_4^{2-} (mg/L)	Cl^- (mg/L)	CO_3^{2-} (mg/L)
Cluster 1	113.57	146.19	1387.54	81.20	86.58	10.81	2860.21	0.00267
Cluster 2	93.02	51.17	138.71	7.58	266.71	157.77	338.84	0.00555
Cluster 3	87.35	85.26	474.99	13.22	91.56	34.17	1111.71	0.00315
Cluster 4	41.50	20.01	51.86	3.89	203.35	6.78	99.51	0.00760
Cluster 5	72.18	28.45	73.37	4.13	414.33	10.40	114.96	0.01798
Cluster 6	18.21	8.83	29.08	2.80	117.77	3.87	45.52	0.00448
Cluster 7	31.23	23.53	137.85	7.94	272.65	8.15	192.61	0.07371
Cluster 8	113.67	66.63	254.66	5.73	599.52	20.08	533.53	0.02836
Whole data	53.43	33.50	163.75	8.61	243.68	17.99	327.19	0.01637
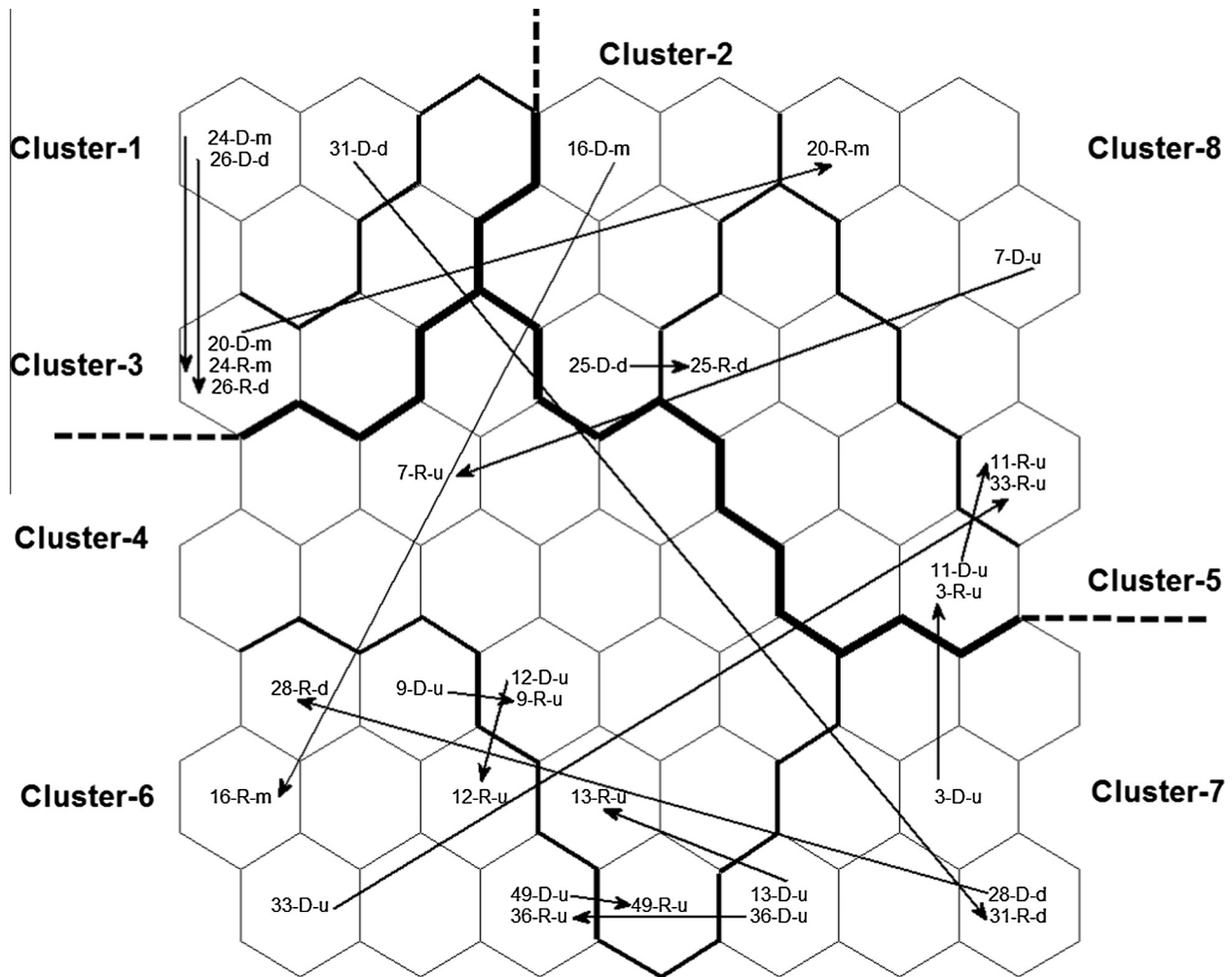
Fig. 7. Representation of the sampling points showing the changes in clusters from the dry to the rainy seasons.

type), respectively, which were characterized by high values of Ca^{2+} and HCO_3^-. According to Tran et al. (2012), the Pleistocene confined aquifer in the RRD is recharged mainly from the surrounding mountains, in which the western mountains are carbonate rock formations consisting of marble, limestone, and dolomite (Drogue et al., 2000). This suggests that the increase of groundwater recharge from the western mountains during the rainy season, which causes the increase of dissolution of carbonate minerals, is the reason for these changes.

Regarding changes in clusters within the same water type, most observation wells exhibited the changes from the cluster with higher concentrations of most major ions in the dry season to the cluster with lower concentrations in the rainy season. For

example, within the high-salinity type, samples from Well Nos. 24 and 26 changed from Cluster 1 to Cluster 3; within the freshwater type, Well No. 12 changed from Cluster 4 to Cluster 6; and within the low-salinity type, Well No. 25 changed from Cluster 2 to Cluster 5. The increase of groundwater recharge during the rainy season may create a dilution effect, which could explain the downward trends in the ion concentrations during the rainy season.

In our former study (Nguyen et al., 2014), we used Piper diagram to investigate the seasonal changes in hydrogeochemical facies. However, some of the seasonal changes in water types as well as clusters within the same water type, which were obtained by using SOM, were not detected by using Piper diagram. For example, samples in Well Nos. 31 and 20 showed significant sea-

sonal changes as mentioned above, but Piper diagram showed that groundwater in these wells was the [Na⁺–Cl⁻] ion type in both seasons. Similarly, samples in Well Nos. 24 and 26, which showed the seasonal changes in clusters within the same water type, were the [Na⁺–Cl⁻] ion types and in Well No. 12 was the [Ca²⁺–HCO₃⁻] ion type in both seasons.

Furthermore, it is difficult for Piper diagram to visually detect the seasonal changes in ion types because of many overlapped samples plotted in Piper diagram, especially when the amount of data increases. On the other hand, SOM can distinctly visualize the seasonal changes as plotting in the different nodes, as well as how big the changes were in a manner more easily by distances between the nodes of samples in both seasons.

3.4. Spatial distribution of the respective clusters

Figs. 8 and 9 show the spatial distribution of the 8 clusters classified by the SOM in the RRD in the dry and rainy seasons, respectively. The symbols star, triangle, asterisk, circle, diamond, rectangular, cross, and inverse triangle represent Clusters 1–8, respectively. As seen in this figure, observation well locations are unevenly distributed across the study area. Due to the region's importance, the wells are denser in urbanized areas, especially around Hanoi. Therefore, well density should be taken into consideration while discussing the spatial distribution of the clusters.

Besides the fact that Clusters 1 and 3 (the high-salinity type) are observed in the coastal area, such as Well Nos. 26 and 31, they are also found in the middle-stream area (Well Nos. 20, 23, 24, and 51), as shown in Figs. 8 and 9. This is consistent with our previous study (Nguyen et al., 2014), the [Na⁺–Cl⁻] ion type, typical of saline

water, was also found in these observation wells. Saltwater intrusion could be the reason for the presence of the high-salinity type in the coastal area. However, high salinity in the middle-stream area could be due to leaching of salty paleowater. According to Tanabe et al. (2003), during the Holocene, the sea transgressed the flood plain as far inland as the present location of Hanoi. The transgression during the Holocene, induced by sea-level rise, must have caused an intrusion of seawater into the underlying highly permeable Pleistocene sediment and the salty porewater may still be present in the middle-stream area.

It is noted that the low salinity type clusters (Clusters 2, 5, and 8) were found near the western and northeastern boundaries of the RRD. Clusters 5 and 8 characterized by high Ca²⁺ and HCO₃⁻ were distributed near the western boundary, while Cluster 2 with significantly high Ca²⁺ and SO₄²⁻ is found near the northeastern boundary. Ca²⁺ and HCO₃⁻ are released by the dissolution of carbonate rock (Jalali, 2009). In fact, the mountains near the western boundary of the delta, which are the main recharge zones for the confined aquifer, are carbonate rock formations comprising marble, limestone, and dolomite, as mentioned in Section 3.3. This suggests that dissolution of these minerals will add significant amounts of Ca²⁺ and HCO₃⁻ to the groundwater near the western boundary of the RRD. On the other hand, high Ca²⁺ and SO₄²⁻ (Cluster 2) probably resulted from the dissolution of sulfate minerals (gypsum and anhydrite), which are commonly found in the Quaternary aquifer system (El-Fiky, 2009). Piper diagram in our previous study (Nguyen et al., 2014) could not detect the samples in the western and northeastern boundaries that showed significantly high Ca²⁺ and HCO₃⁻ (Clusters 5 and 8) and Ca²⁺ and SO₄²⁻ (Cluster 2) respectively. It only revealed that groundwater in the upstream area is
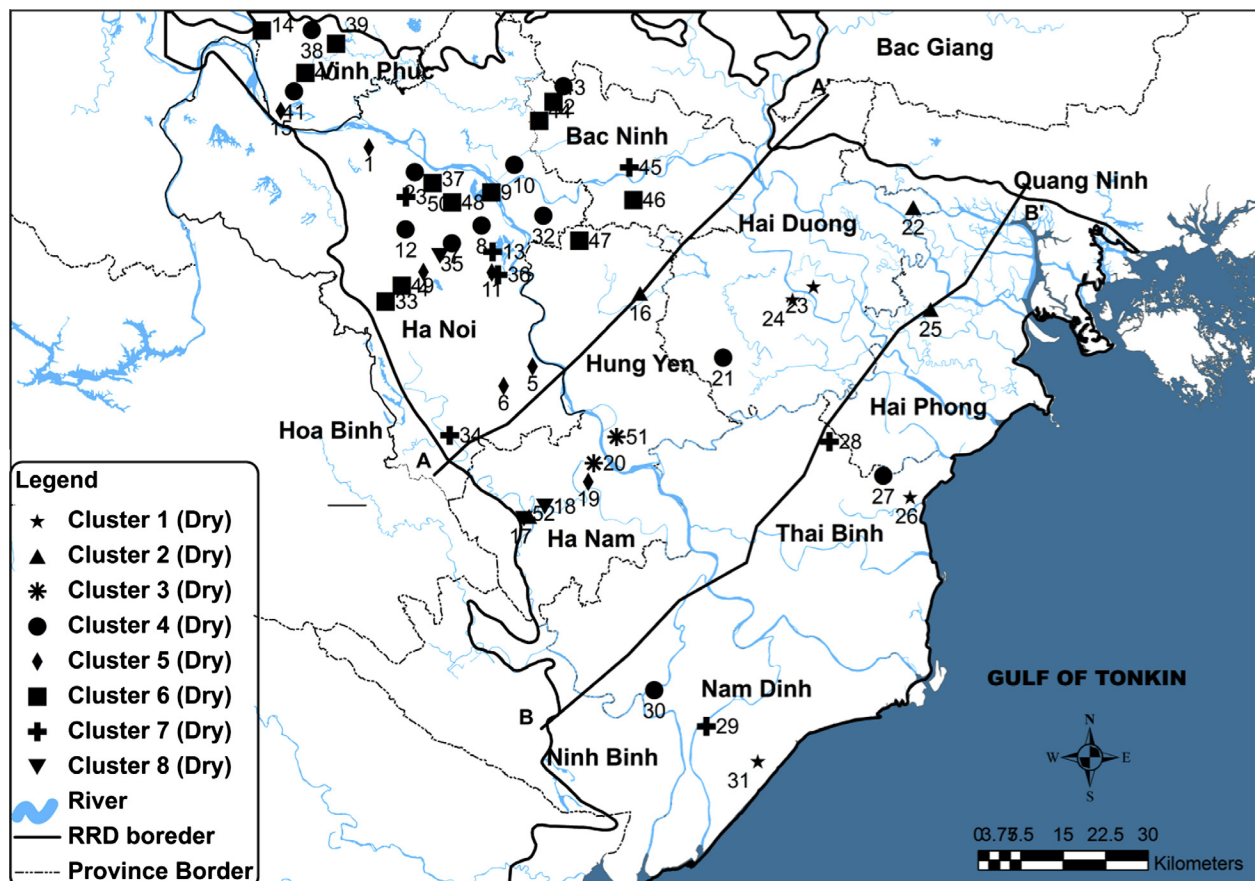

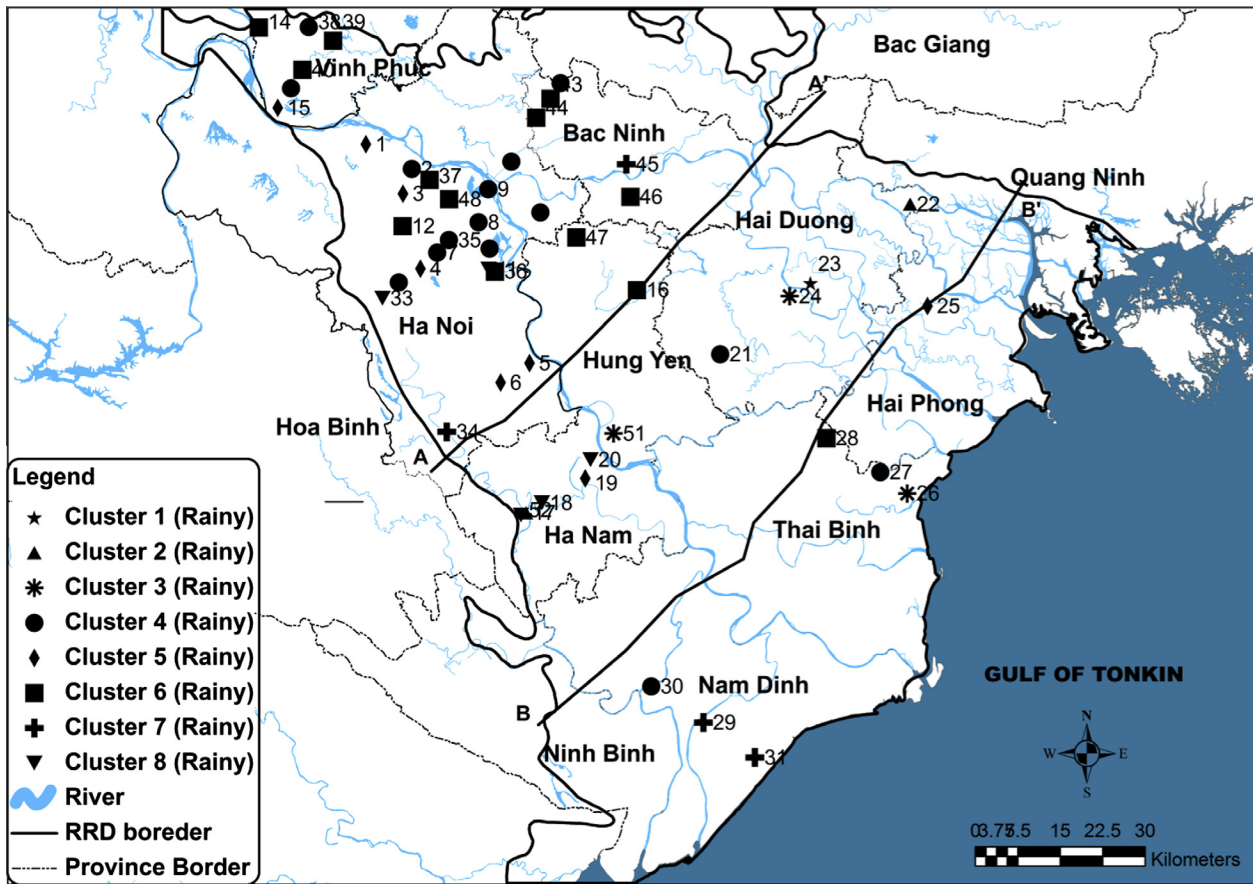**Fig. 8.** Spatial distribution of the respective clusters in dry season.

Fig. 9. Spatial distribution of the respective clusters in rainy season.

dominated by the [Ca^{2+}-HCO_3^-] ion type, whereas the [Na^+-SO_4^{2-}] ion type was found only in the middle area of RRD (Well No. 16). These findings were neither found in the previous study of Tran et al. (2012), which were based on the geophysical investigations.

The freshwater type clusters (Clusters 4, 6, and 7) were distributed mostly in the upstream and downstream areas except the coastal area, in which the freshest water type cluster (Cluster 6) was found in the upstream area, as shown in Figs. 8 and 9. The study of Tran et al. (2012), which investigated the distribution of groundwater types in the RRD by using the different kind of data (formation electrical conductivities, the transient electromagnetic soundings) and different methods (geophysics and electromagnetic methods), also revealed that in the Pleistocene aquifer, fresh groundwater is present in those areas. A local lens of freshwater existing in the Pleistocene aquifer in the downstream area of RRD, which was identified by Wagner et al. (2012), could be the reason for the presence of the freshwater type found in the downstream area. The existence of freshwater type in the downstream area was not obtained by using Piper diagram as in our previous study (Nguyen et al., 2014).

The objective of using Piper diagram is neither to cluster the hydrogeochemical data to discover absorbing characteristics nor to find a new set of water types, but rather to decide how the hydrogeochemical data should be classified into different ion types. Basically in Piper diagram, cations of Na^+ and K^+ as well as anions of HCO_3^- and CO_3^{2-} are plotted as their sum values, not separately treated like in SOM. Furthermore, Piper diagram just classifies the water ion types into the combination of one of three cation types and one of three antion types, whereas application of SOM can flexibly cluster the hydrogeochemical data into any combination among 8 cation and anion parameters by the automated

process. Therefore, application of Piper diagram did not suffice to investigate the intrinsic relationship between each parameter of the hydrogechemical data.

In conclusion, in this study the SOM provided understandable and visualized results for clustering the hydrogeochemical groundwater data into exclusively distinguishable water types. These results are not easily obtained from the traditional graphical method like Piper diagram. In addition, the seasonal changes in clusters and water types are easy to be detected. In other words, by using SOM, the hydrogeochemical data is easily interpreted and understood. It does not need great knowledge about the geology (e.g., formation factor of the sediments) and geophysics (e.g., electromagnetic properties) in the RRD (Tran et al., 2012), nor how to interpret Piper diagram (Nguyen et al., 2014).

3.5. Factors governing chemistry of groundwater in each cluster

Fig. 10 shows the Gibbs diagrams for the 8 clusters classified by the SOM. The symbols for expressing the Clusters 1–8 in this figure are the same as that in Figs. 8 and 9. Gibbs (1970) found that most of the world's surface water falls within the boomerang-shaped boundaries. Based on analytical chemical data for numerous surface samples, Gibbs theorized the three major mechanisms controlling world surface water chemistry, which are presented in three domains: precipitation dominance (lower part), rock dominance (middle part), and evaporation dominance (upper part), as shown in Fig. 10. In addition, Gibbs diagrams have been also used for the functional sources assessment of dissolved ions in groundwater in various studies with careful interpretations. For example, Marghade et al. (2012) used Gibbs diagram to assess the functional sources of dissolved chemical constituents of groundwater in Neg-
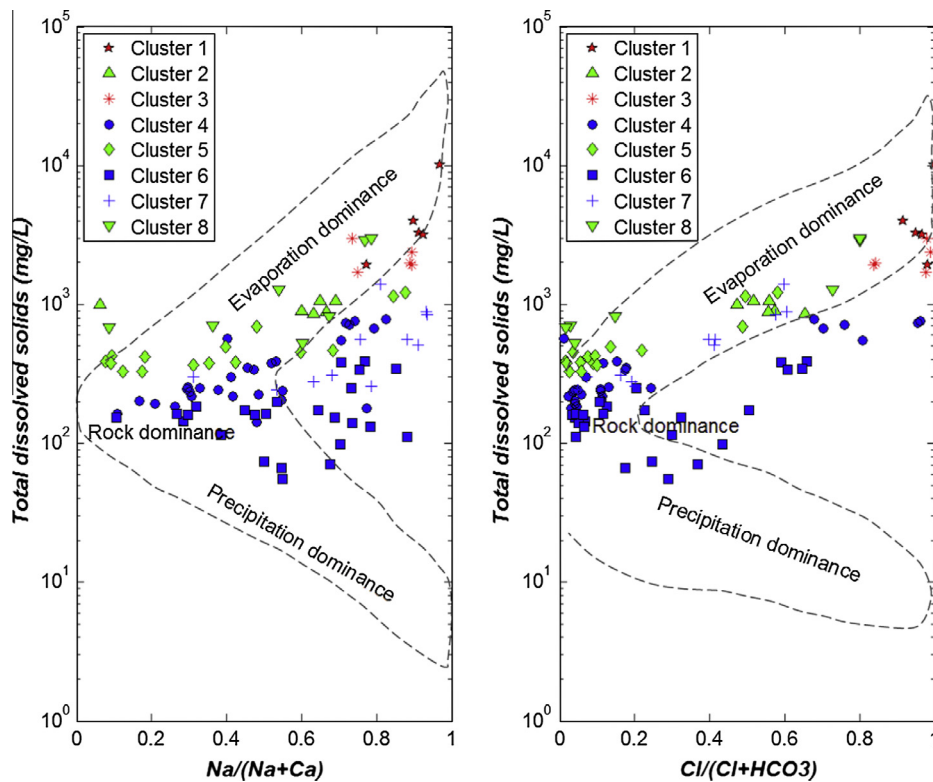
Fig. 10. Gibbs diagrams for the classified data into the respective cluster.

pur, Central India, with an interpretation that most of the samples falling in the evaporation zone were influenced by anthropogenic activities. Yidana et al. (2010) used Gibbs diagram for groundwater in the Keta basin, Ghana, with an interpretation that samples plot within the evaporation dominance domain due to elevated TDS and sodium concentration arising from saline seawater intrusion. Since this study has focused on the Pleistocene confined aquifer, the evaporation mechanism probably does not affect directly the water chemistry. Gibbs diagram has been already used in our previous study (Nguyen et al., 2014) to investigate the main mechanism controlling groundwater chemistry of RRD, Vietnam. The results showed that the source of the dissolved ions in the groundwater is rock-water interaction in the upstream and the southwestern part of the downstream area, but salty paleowater and salt water from the sea are the main factors influencing groundwater chemistry in the middle-stream and northern downstream areas. In this study, we used Gibbs diagram with more samples in order to investigate factors governing chemistry of groundwater in each cluster which was classified by the SOM. Therefore, in this study the samples plotted in Gibbs diagram were represented in accordance with each cluster, which was not considered in the previous study.

As shown in the Fig. 10, Clusters 1 and 3 (the high-salinity type) were plotted toward the evaporation dominance domain as the result of high TDS and weight ratios (Na/Ca + Na and Cl/Cl + HCO$_3$). It can be inferred that the occurrence of the high TDS is mostly due to high concentrations of Na$^+$ and Cl$^-$ (as shown in Table 2). According to Tran et al. (2012), the source of sodium and chloride in groundwater of the Pleistocene confined aquifer in the middle-stream area is mixing with salty paleowater, and in the coastal area is saltwater intrusion from rivers and the sea. This suggests that saltwater intrusion is the main factor affecting the groundwater chemistry of these clusters in the coastal area, while high-salinity

type (clusters 1 and 3) found in the middle-stream area could be influenced by mixing with salty paleowater.

On the other hand, groundwater samples belonging to Clusters 4, 6, and 7 (the freshwater type) fall toward the domain of rock dominance due to low TDS. Some samples belonging to these clusters fall outside the boomerang-shaped boundaries. It is common for groundwater that the domain of rock dominance extends further toward higher weight ratios (Raju et al., 2011; Marghade et al., 2012; Yidana et al., 2010). The fall in the rock dominance domain suggests that these clusters are dominated by the processes of mineral dissolution. In particular, Cluster 6 has the lowest TDS with some samples tending to fall in the domain of precipitation dominance. This explains why this cluster has the lowest concentrations of all ions (Table 2). In our previous study (Nguyen et al., 2014), there is no sample falling in the domain of precipitation dominance due to the fewer set of hydrgeochemical data of groundwater in the confined aquifer.

In the low-salinity type, Clusters 2 and 8 have relatively high TDS as a result of the high concentrations of Ca^{2+} and SO$_4^{2-}$ for Cluster 2 and Ca^{2+} and HCO$_3^-$ for Cluster 8, and thus, falling in the evaporation dominance domain. This could be due to evaporation, which increases salinity and precipitation of CaCO$_3$ from solution, which increases the relative proportion of Na$^+$ to Ca^{2+} and Cl$^-$ to HCO$_3^-$ (Gibbs, 1970). Cluster 5 falls in the rock dominance domain, which suggests that rock–water interaction is the natural mechanism controlling the dissolved ions in this cluster.

4. Conclusions

In this study, the SOM in combination with a hierarchical cluster analysis was systematically applied for clustering hydrogeochemical groundwater data comprising major ions from 52 observation wells to investigate the seasonal and spatial hydrogeo-

chemical characteristics of groundwater in the Pleistocene confined aquifer of the RRD. In addition, Gibbs diagrams were also created to elucidate the hydrogeochemical characteristics classified by the SOM. The main conclusions drawn from this study are as follows:

- From the results of the SOM application, the major ion chemistry data were divided into eight clusters. From the practical point of view, these eight clusters revealed three basic representative water types characterized by the high salinity (Clusters 1 and 3), low salinity (Clusters 2, 5, and 8), and freshwater (Clusters 4, 6, and 7). The high-salinity water type is distributed in the middle-stream and coastal areas, while the low-salinity water type is found near the western and northeast boundaries of the RRD.
- Changes in the water types from the dry to rainy seasons were detected in more than 10% of the observation wells, while cluster changes within the same water type was about 20%. The increase in groundwater recharge during the rainy season could be the main reason for these changes.
- The results of the Gibbs diagram suggest that the source of water-soluble ions in the groundwater characterized by the freshwater type and the low-salinity type are the chemical weathering of the rock-forming minerals and the chemical interaction between aquifer rocks and groundwater, while salty paleowater and saltwater intrusion are the main source of the dissolved solids in the groundwater characterized by the high-salinity water type.
- The SOM provided readily understandable and visualized results for classifying the hydrogeochemical groundwater data into exclusively distinguishable hydrogeochemical types. Therefore, the SOM was found to be a very effective tool for the assessment of groundwater quality in terms of the seasonal and spatial hydrogeochemical characteristics.

Acknowledgements

References

Astel, A., Mazerski, J., Polkowska, Z., Namiesnik, J., 2004. Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). Adv. Environ. Res. 8 (3–4), 337–349.

Belkhiri, L., Boudoukha, A., Mouni, L., 2011. A multivariate statistical analysis of groundwater chemistry data. Int. J. Environ. Res. 5 (2), 537–544.

Bui, D.D., Kawamura, A., Tong, T.N., Amaguchi, H., Nakagawa, N., Iseri, Y., 2011. Identification of aquifer system in the whole Red River Delta, Vietnam. Geosci. J. 15 (3), 323–338.

Bui, D.D., Kawamura, A., Tong, T.N., Amaguchi, H., Nakagawa, N., 2012. Spatiotemporal analysis of recent groundwater-level trends in the Red River Delta, Vietnam. Hydrogeol. J. 20, 1635–1650.

Drogue, C., Cat, N.N., Dazy, J., 2000. Geological factors affecting the chemical characteristics of the thermal waters of the carbonate karstified aquifers of Northern Vietnam. Hydrol. Earth Syst. Sci. 4 (2), 332–340.

El-Fiky, A.A., 2009. Hydrogeochemical characteristics and evolution of groundwater at the Ras Sudr-Abu Zenima Area, Southwest Sinai, Egypt. Earth Sci. 21 (1), 79–109.

Farsadnia, E., Rostami Kamrood, M., Moghaddam Nia, A., Modarres, R., Bray, M.T., Han, D., Sadatinejad, J., 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. J. Hydrol. 509, 387–397.

Gibbs, R.J., 1970. Mechanisms controlling world water chemistry. Science 17 (1088–1090), 1970.

Güler, C., Thyne, G.D., 2004. Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering. Water Resour. Res. 40. http://dx.doi.org/10.1029/2004WR003299.

Güler, C., Thyne, G.D., McCray, J.E., Turner, A.K., 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol. J. 10, 455–474.

Hall, M.J., Minns, A.W., 1999. The classification of hydrologically homogeneous regions. Hydrol. Sci. J. 44 (5), 693–704.

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, 56–64.

Hilario, L.G., Ivan, M.G., 2004. Self-organizing map and clustering for wastewater treatment monitoring. Eng. Appl. Artif. Intell. 17, 215–225.

Hong, Y.T., Rosen, M.R., Bhamidimarri, R., 2003. Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis. Water Res. 37, 1608–1618.

Hussein, M.T., 2004. Hydrochemical evaluation of groundwater in the Blue Nile Basin, eastern Sudan, suing conventional and multivariate teachniques. Hydrogeol. J. 12, 144–158.

IMHE-MONRE, 2011. Annual Report on Hydrological Observation in Vietnam. Vietnamese Ministry of Environment and Natural Resources.

Iseri, E., Matsuura, T., Iizuka, S., Nishiyama, K., Jinno, K., 2009. Comparison of pattern extraction capability between self-organizing maps and principal component analysis. Mem. Fac. Eng. Kyushu Univ. 69 (2), 37–47.

Jalali, M., 2009. Geochemistry characterization of groundwater in an agricultural area of Razan, Hamadan, Iran. Environ. Geol. 56, 1479–1488.

Jeong, K.S., Hong, D.G., Byeon, M.S., Jeong, J.C., Kim, H.G., 2010. Stream modification patterns in a river basin: field survey and self-organizing map (SOM) application. Ecol. Inform. 5, 293–303.

Jin, Y.H., Kawamura, A., Park, S.C., Nakagawa, N., Amaguchi, H., Olsson, J., 2011. Spatiotemporal classification of environmental monitoring data in the Yeongsan River basin, Korea, using self-organizing maps. J. Environ. Monit. 13, 2886–2894.

Kalteh, A.M., Hjorth, P., Berndtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: analysis, modeling and application. Environ. Modell. Software 23, 835–845.

Kohonen, T., 2001. Self-Organizing Maps, third ed. Springer.

Lambrakis, N., Antonakos, A., Panagopoulos, G., 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. Water Res. 38, 1862–1872.

Leloup, J.A., Lachkar, Z., Boulanger, J.P., Thiria, S., 2007. Detecting decadal changes in ENSO using neural networks. Clim. Dyn. 28, 147–162.

Marghade, D., Malpe, D.B., Zade, A.B., 2012. Major ion chemistry of shallow groundwater of a fast growing city of Central India. Environ. Monit. Assess. 184, 2405–2418.

MONRE, 2008. National technical regulation on underground water quality. Vietnamese Ministry of Environment and Natural Resources (in Vietnamese).

Nguyen, T.T., Kawamura, A., Tong, N.T., Nakagawa, N., Amaguchi, H., Gilbuena, R., 2014. Hydrogeochemical characteristics of groundwater from the two main aquifers in the Red River Delta, Vietnam. J. Asian Earth Sci. 93, 180–192.

Nishiyama, K., Endo, S., Jinno, K., Uvo, C.B., Olsson, J., Berndtsson, R., 2007. Identification of typical synoptic patterns causing heavy rainfall in the rainy season in Japan by a Self-Organizing Map. Atmos. Res. 83, 185–200.

Alberto, W.D., Pilar, D.M.D., Valeria, A.M., Fabiana, P.S., Cecilia, H.A., Angeles, B.M.D.L., 2001. Pettern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía river basin (Córdoba-Argentina). Water Res. 35 (12), 2881–2894.

Raju, N.J., Shukla, U.K., Ram, P., 2011. Hydrogeochemistry for the assessment of groundwater quality in Varanasi: a fast-urbanizing center in Uttar Pradesh, India. Environ. Monit. Assess. 173, 279–300.

Reghunath, R., Sreedhara Murthy, T.R., Raghavan, B.R., 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: an examples from Karnataka, India. Water Res. 36, 2437–2442.

Srinivas, V.V., Tripathi, S., Ramachandra Rao, A., Govindaraju, R.S., 2008. Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering. J. Hydrol. 348, 148–166.

Su, M.C., Liu, T.K., Chang, H.T., 2002. Improving the self-organizing feature map algorithm using an efficient initialization scheme. Tamkang J. Sci. Eng. 5 (1), 35–48.

Subyani, A.M., Al Ahmadi, M.E., 2009. Multivariate statistical analysis of groundwater quality in Wadi Ranyah, Saudi Arabia. Earth Sci. 21 (2), 29–46.

Tanabe, S., Hori, K., Saito, Y., Haruyama, S., Vu, P.V., Kitamura, A., 2003. Song Hong (Red River) delta evolution related to millennium-scale Holocene sea-level changes. Quatern. Sci. Rev. 22, 2345–2361.

Tong, T.N., 2004. National Hydrogeology Database Program. Final Project Report. Department of Geology and Minerals of Vietnam, 120p (in Vietnamese).

Tran, T.L., Larsen, F., Pham, Q.N., Christiansen, A.V., Tran, N., Vu, V.H., Tran, V.L., Hoang, V.H., Hinsby, K., 2012. Origin and extent of fresh groundwater, salty paleowaters and recent saltwater intrusions in Red River flood plain aquifers, Vietnam. Hydrogeol. J. 20, 1295–1313.

Vesanto, J., Himberg, J., Alhoniemi, E., Parahankangas, J., 2000. SOM Toolbox for Matlab 5, Helsinki Univeriity Report A57.

Vialle, C., Sablayrolles, C., Lovera, M., Jacob, S., Huau, M.-C., Montrejoud-Vignoles, M., 2011. Monitoring of water quality from roof runoff: interpretation using multivariate analysis. Water Res. 45, 3765–3775.

Vietnam General Statics Office, 2013. Statistical Handbook of Vietnam. General Statics Office, Hanoi. <http://www.gso.gov.vn> (accessed 08.01.14).

Wagner, F., Ludwig, R.R., Noell, U., Hoang, H.V., Pham, Q.N., Larsen, F., Lindenmaier, F., 2012. Genesis of economic relevant fresh groundwater resources in Pleistocene/Neogene aquifers in Nam Dinh (Red River Delta, Vietnam). In: EGU Ceneral Assembly, vol. 14, p. 2273.

Wallner, M., Haberlandt, U., Dietrich, J., 2013. A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. J. Hydrol. 494, 59–71.

Winkel, L.E., Pham, T.K.T., Vi, M.L., Stengel, C., Amini, M., Nguyen, T.H., Pham, H.V., Berg, M., 2011. Arsenic Pollution of Groundwater in Vietnam Exacerbated by Deep Aquifer Exploitation for more than a Century. PNAS Early Edition 1–6, PNAS, Washington, DC.

Yidana, S.M., Bruce, B.Y., Thomas, M.A., 2010. Analysis of groundwater quality using multivariate and spatial analyses in the Keta basin, Ghana. J. Afr. Earth Sc. 58, 220–234.

Zhang, B., Song, X., Zhang, Y., Han, D., Tang, C., Yu, Y., Ma, Y., 2012. Hydrochemical characteristics and water quality assessment of surface water and groundwater in Songen plain, Northeast China. Water Res. 46, 2737–2748.