Journal of Hydrology 522 (2015) 661–673

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

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

Article history:
Received 16 September 2014
Received in revised form 6 January 2015
Accepted 7 January 2015
Available online 21 January 2015
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Renduo Zhang, Associate Editor

Keywords:
SOM
Gibbs diagrams
Pleistocene confined aquifer
Major ions

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.

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

* Corresponding author. Tel.: +81 42 677 4542; fax: +81 42 677 2772.
E-mail address: nguyen-thanhthuy@ed.tmu.ac.jp (T.T. Nguyen).

http://dx.doi.org/10.1016/j.jhydrol.2015.01.023
0022-1694/

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