Quantification of hydrochemical facies of groundwater using model-based and fuzzy c-means clustering

SUBIN CHOI[1], KYOUNG-HO KIM[1], SEONG-TAEK YUN[1*]

[1]Department of Earth and Environmental Sciences, Korea University, Seoul 136-701, Republic of Korea
Email : styun@korea.ac.kr (*correspondence)

Graphical methods (e.g., Piper diagram) and statistical cluster analysis (e.g., hierarchical clustering) have been used to cluster water samples as an initial step to interpret hydrochemical processes (natural versus anthropogenic) controlling groundwater quality. However, these conventional approaches can be unsatisfactory for reasonable interpretation because 1) groundwater quality can be too complex to successfully analyze using conventional graphical methods, 2) hydrochemical properties of the natural groundwater system can vary continuously, and 3) statistically, the geometrical space of compositional data makes up the simplex and therefore the distance between two compositional observations should be measured by the Aitchison distance, whereas most of statistical clustering methods are based on Euclidean space. This study aims to evaluate various clustering methods to better understand hydrochemical processes using hydrochemical dataset (a kind of compositional data) and to effectively visualize the results of clustering. For this study, hydrochemical data of 80 groundwater samples were collected in an agricultural area of Anseong, South Korea. The Ilr (isometric log-ratio)-transformed data of the concentrations of major solutes (Na, K, Ca, Mg, HCO_3, SO_4, NO_3, and Cl) were interpreted using the soft clustering method (i.e., Model-based and Fuzzy C-means clustering). Then, the fuzziness/probability of the samples belonging to specific groups was plotted on Piper diagram for the visualization of quantitative clustering results. In addition, the principal component analysis (PCA) was conducted to understand the factors determining hydrochemical facies. The results yielded three well-defined water types: 1) Ca-HCO_3, 2) Ca-Na-HCO_3, and 3) Ca-Cl-NO_3 types. Each group is interpreted to represent distinct groundwater masses: 1) Ca-HCO_3 type represents a recently recharged groundwater, 2) Ca-Na-HCO_3 type represents a highly evolved groundwater through water-rock interaction, and 3) Ca-Cl-NO_3 type represents highly contaminated (esp., nitrate-rich) groundwater. The PCA results and kriging maps also supported the results of the model-based and Fuzzy C-means clustering. Therefore, this study shows the successful application of the model-based and Fuzzy C-means clustering method to compositional hydrochemical data.