Assessment of Groundwater with Combination of Support Vector Machine and Principal Component Analysis in an Irrigation District of the Yellow River Basin

Shuai Song\textsuperscript{1, 3}, Fadong Li\textsuperscript{*}, Qiang Liu\textsuperscript{2}, Xiangdong Wu\textsuperscript{2}, Guangshuai Zhao\textsuperscript{1,3}

\textsuperscript{1} Yucheng Comprehensive Experiment Station, Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, 100101, China
\textsuperscript{2} Chang’an University, Xi’an 710064, China
\textsuperscript{3} Graduate University of Chinese Academy of Science, Beijing, 100049, China

The area of irrigation districts in the lower reach of the Yellow River is over 25,800 km\textsuperscript{2}, accounting for 10.18\% of arable land in North China Plain. Groundwater quality in these irrigation districts are affected by anthropogenic interventions and natural sources, especially the industrial and agricultural pollution and irrigation water supplied by artificial channels connecting to the Yellow River. To assess the groundwater quality in these districts, 30 groundwater samples were investigated in Xinxiang City, Henan province, north China from 2002 to 2004. Using 13 physical and chemical parameters for water quality indexes, an improved support vector machine (SVM) model that combines the principal component analysis (PCA) technique was used to assess the groundwater quality and their annual changes. The PCA method can transform input variables into a number of uncorrelated variables by an orthogonal transformation, and reduce the original input variables. Then, these five selected principal components as new input variables were used in the process of constructing SVM model via the bintree structure. Results of this improved model were compared with grading method recommended by the current groundwater quality criteria (GB/T 14848-93) in China. It is shown that only 28.4\% of the water samples in the well field were better than the quality standard III level. Up to 70\% of the groundwater samples were worse than the groundwater standard rated level. It indicates that the improved model can provide reliable assessment and is expected to be used widely in the future.

Keywords: principal component analysis; support vector machine; groundwater quality assessment; irrigation district; Yellow River Basin