Journal of Contaminant Hydrology 180 (2015) 25–33

Contents lists available at ScienceDirect

Journal of Contaminant Hydrology

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

Decision-tree-model identification of nitrate pollution activities in groundwater: A combination of a dual isotope approach and chemical ions

Dongmei Xue [a], Fengmei Pang [b], Fanqiao Meng [b,*], Zhongliang Wang [a], Wenliang Wu [b]

[a] Tianjin Key Laboratory of Water Resources and Environment, Tianjin Normal University, Tianjin 300387, China
[b] College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China

ARTICLE INFO

Article history:
Received 20 November 2014
Received in revised form 16 July 2015
Accepted 17 July 2015
Available online 23 July 2015

Keywords:
NO_3^- pollution
NO_3^- source
Groundwater
$\delta^{15}N$- and $\delta^{18}O$-NO_3^-
k-means clustering
Decision tree model

ABSTRACT

To develop management practices for agricultural crops to protect against NO_3^- contamination in groundwater, dominant pollution activities require reliable classification. In this study, we (1) classified potential NO_3^- pollution activities via an unsupervised learning algorithm based on $\delta^{15}N$- and $\delta^{18}O$-NO_3^- and physico-chemical properties of groundwater at 55 sampling locations; and (2) determined which water quality parameters could be used to identify the sources of NO_3^- contamination via a decision tree model. When a combination of $\delta^{15}N$-, $\delta^{18}O$-NO_3^- and physico-chemical properties of groundwater was used as an input for the k-means clustering algorithm, it allowed for a reliable clustering of the 55 sampling locations into 4 corresponding agricultural activities: well irrigated agriculture (28 sampling locations), sewage irrigated agriculture (16 sampling locations), a combination of sewage irrigated agriculture, farm and industry (5 sampling locations) and a combination of well irrigated agriculture and farm (6 sampling locations). A decision tree model with 97.5% classification success was developed based on SO_4^{2-} and Cl^- variables. The NO_3^- and the $\delta^{15}N$- and $\delta^{18}O$-NO_3^- variables demonstrated limitation in developing a decision tree model as multiple N sources and fractionation processes both resulted in difficulties of discriminating NO_3^- concentrations and isotopic values. Although only the SO_4^{2-} and Cl^- were selected as important discriminating variables, concentration data alone could not identify the specific NO_3^- sources responsible for groundwater contamination. This is a result of comprehensive analysis. To further reduce NO_3^- contamination, an integrated approach should be set-up by combining N and O isotopes of NO_3^- with land-uses and physico-chemical properties, especially in areas with complex agricultural activities.

1. Introduction

Nitrate (NO_3^-) is one of the most common groundwater pollutants, especially in intensive agricultural regions (Strebel et al., 1989; Chen et al., 2008; Ju et al., 2009). Sources of NO_3^- include the use of N-containing organic and inorganic fertilizers (Smil, 1999), animal manure and elevated atmospheric N deposition (Benkovitz et al., 1996). Nitrates enter groundwater during recharge events, mostly from leached surface or subsurface soils (Landon et al., 2000). To control contamination in groundwater, the NO_3^- sources must be identified before efficient management strategies can be executed.

Traditionally, NO_3^- concentrations, land-uses and farming practices were applied to identify NO_3^- sources in groundwater (Liu et al., 2006). However, these traditional methods were restricted by complex environmental factors, e.g., unbalanced spatial variabilities of NO_3^- contamination, mixing of different NO_3^- sources and the complexity of physico-chemical effects in the N cycle (Hall and Risser, 1993). Recent research (Karr et al.,

* Corresponding author.
E-mail address: mengfq@cau.edu.cn (F. Meng).

http://dx.doi.org/10.1016/j.jconhyd.2015.07.003
0169-7722/

2001; Mayer et al., 2002; Mitchell et al., 2003; Kaushal et al., 2006) has proved that a dual isotope approach (using N and O) is a useful tool to discriminate pollution sources, as different NO_3^- sources have unique isotope ratios of N ($^{15}N/^{14}N$) and O ($^{18}O/^{16}O$). Thus, it is possible to discriminate inorganic fertilizer from organic fertilizer (Kendall and Aravena, 1999); forested catchment from agricultural land (Mayer et al., 2002); and atmospheric deposition from microbial nitrification (Pardo et al., 2004; Barnes et al., 2008). However, mixing of multiple NO_3^- sources will result in intermediate values that affect NO_3^- source identification. Moreover, various fractionation processes may alter the initial composition of NO_3^- sources before or after mixing, giving rise to inaccurate results for NO_3^- source identification (Kendall, 1998; Panno et al., 2006). Linking $\delta^{15}N$- and $\delta^{18}O$-NO_3^- values with land-uses and physico-chemical properties of water provided essential information for identifying NO_3^- sources, as chemical constituents from anthropogenic sources are inclined to dominate those of natural sources (Karr et al., 2001; Spruill et al., 2002; Kaushal et al., 2006). Spruill et al. (2002) successfully classified five NO_3^- sources (fertilizer on crops, fertilizer on golf courses, irrigation spray from hog wastes, and leachate from poultry litter and septic systems) in groundwater based on a classification tree model using N isotopes, anions and cations. The basis of a classification/decision tree model is the creation of a hierarchical tree based on a dataset of known classes, which is then used to predict classes from another independent dataset with the same parameters but unknown classes (Xue et al., 2013).

The sites investigated in this study are located in an intensive agricultural region, where 66.5% of irrigation water comes from groundwater (Liu et al., 2005). Contamination of groundwater in this region is a serious environmental issue, as more than half of the area has been suffering various levels of NO_3^- pollution. Thus, the aims of this work were to (1) classify potential NO_3^- polluting activities via an unsupervised learning algorithm based on $\delta^{15}N$- and $\delta^{18}O$-NO_3^- and groundwater physico-chemical properties; and (2) determine which water quality parameters could be used to identify the sources of NO_3^- contamination via a decision tree model. Through establishment of classification tree models, researchers and regulators can use or develop decision tree models to find the predominant agricultural activities responsible for NO_3^- contamination in groundwater.

2. Materials and methods

2.1. Study sites

The study sites are located in Huantai County, Shandong Province, North China Plain (Fig. 1). The sampling campaign was conducted in November 2011 as irrigation and fertilization could affect groundwater during the rainy season. Before the rainy season (December to next May), groundwater level is too low to reflect the annual average in this area. In November, the nature and quantity of groundwater remained relatively balanced and stable after a summer rain collection, reflecting the effects of anthropogenic activities on groundwater, in particular the influence of pumping and drainage. The groundwater in this region is late tertiary and quaternary sediment pore water, with aquifer characteristics as shown in Table 1. The groundwater system is made up of a series of confined aquifers located in bicarbonate and bicarbonate-sulfate layers, whose porosity is mainly related to the fracture network. Currently, the crop production in Huantai County is based on shallow groundwater, flowing from south to north consistent with surface water flow (Jiang, 1982). Groundwater samples were collected from 55 irrigation wells (Fig. 1), distributed across with light (5 wells in GL and ZJ), medium (12 wells in MQ, JJ and QF), and severe pollution (38 wells in CZ, TZ, XJ, XC, TS, SZ) according to the World Health Organization (WHO) drinking water standard. Previous investigations in Huantai County (Liu and Wu, 2003; Chen et al., 2008) demonstrated that mean NO_3^-–N concentration doubled over 5 years, from 8.1 mg L^{-1} in 2002 to 14.7 mg L^{-1} in 2007. The irrigation wells were located either in the vegetable or in wheat/maize growing areas.

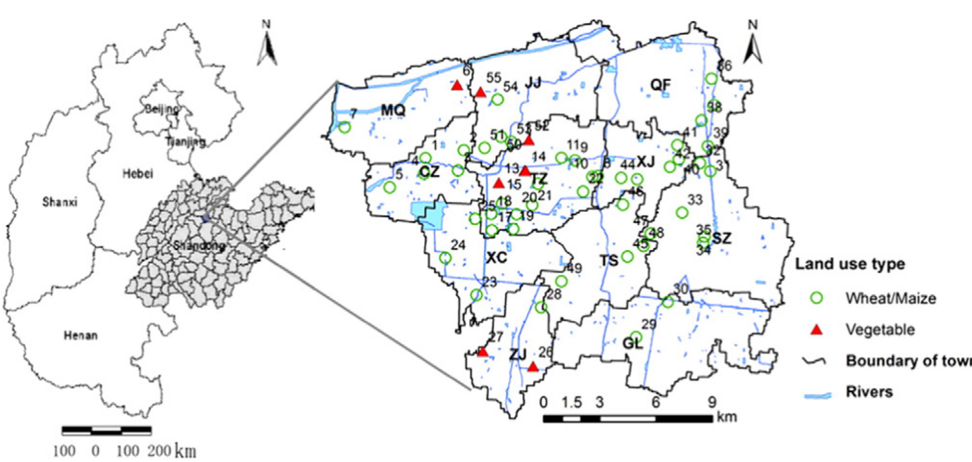

Fig. 1. Sampling locations and the corresponding land use types of groundwater. CZ represents Chenzhuang Town; GL represents Guoli Town; JJ represents Jingjia Town; MQ represents Maqiao Town; QF represents Qifeng Town; SZ represents Suozhen Town; TS represents Tangshan Town; TZ represents Tianzhuang Town; XC represents Xincheng Town; XJ represents Xingjia Town; and ZJ represents Zhoujia Town.