Physics and Chemistry of the Earth 36 (2011) 379–386

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

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

Spatio-temporal patterns and source apportionment of pollution in Qiantang River (China) using neural-based modeling and multivariate statistical techniques

Shiliang Su [a], Junjun Zhi [a], Liping Lou [b], Fang Huang [a], Xia Chen [a], Jiaping Wu [a,*]

[a] *Zhejiang University – The University of Western Australia Joint Centre in Integrated Water Management and Protection, College of Environment and Natural Resources, China*
[b] *Ministry of Agriculture Key Laboratory for Non-Point Source Pollution Control, College of Environment and Natural Resources, Zhejiang University, China*

ARTICLE INFO

Article history:
Received 28 May 2009
Received in revised form 23 November 2009
Accepted 9 March 2010
Available online 12 March 2010

Keywords:
Water quality
Source apportionment
Multivariate methods
Neural-based modeling
Spatial pattern
Temporal variation

ABSTRACT

Characterizing the spatio-temporal patterns and apportioning the pollution sources of water bodies are important for the management and protection of water resources. The main objective of this study is to describe the dynamics of water quality and provide references for improving river pollution control practices. Comprehensive application of neural-based modeling and different multivariate methods was used to evaluate the spatio-temporal patterns and source apportionment of pollution in Qiantang River, China. Measurement data were obtained and pretreated for 13 variables from 41 monitoring sites for the period of 2001–2004. A self-organizing map classified the 41 monitoring sites into three groups (Group A, B and C), representing different pollution characteristics. Four significant parameters (dissolved oxygen, biochemical oxygen demand, total phosphorus and total lead) were identified by discriminant analysis for distinguishing variations of different years, with about 80% correct assignment for temporal variation. Rotated principal component analysis (PCA) identified four potential pollution sources for Group A (domestic sewage and agricultural pollution, industrial wastewater pollution, mineral weathering, vehicle exhaust and sand mining), five for Group B (heavy metal pollution, agricultural runoff, vehicle exhaust and sand mining, mineral weathering, chemical plants discharge) and another five for Group C (vehicle exhaust and sand mining, chemical plants discharge, soil weathering, biochemical pollution, mineral weathering). The identified potential pollution sources explained 75.6% of the total variances for Group A, 75.0% for Group B and 80.0% for Group C, respectively. Receptor-based source apportionment was applied to further estimate source contributions for each pollution variable in the three groups, which facilitated and supported the PCA results. These results could assist managers to develop optimal strategies and determine priorities for river pollution control and effective water resources management.

1. Introduction

Water pollution, from which approximately 25 million persons die every year, has become a major problem in many countries (Pimpunchat et al., 2008). As the main inland water resources for domestic, industrial and irrigation purposes, rivers are among the most vulnerable water bodies to pollution due to their important part in transporting industrial and sewage discharges, and run-off from farmland in their vast drainage basins (Singh et al., 2005). Increasing industrialization leads to ever increasing pollution of rivers in developing countries (Jonnalagadda and Mhere, 2001). One critical step to effectively control river pollution is the development of water quality monitoring programs to follow environmental conditions and spatio-temporal trends. In China, the government has established environmental monitoring system and carried out various water quality monitoring programs for decades, but such monitoring system results in voluminous data, including physical properties, nutrient, inorganic and biological parameters, which are difficult to analyse and interpret because of the latent interrelationships between parameters and monitoring sites (Shin and Fong, 1999; Zhou et al., 2007b; Zhang et al., 2009). Thus, it is a basic requirement to extract meaningful information from these data, using advanced mathematical methods to interpret spatio-temporal patterns, identify significant parameters, and characterize the pollution sources as well as their quantitative contributions to water quality issues (Zhou et al., 2007c; Zhang et al., 2009).

Application of multivariable statistical methods serves as a valuable tool to better understand and interpret complicated data sets. Parallel factor analysis, neural-based modeling and *N*-way principal component analysis have recently been used. Compared with these relatively more sophisticated methods, cluster analysis (CA), discriminant analysis (DA), principal component analysis (PCA), factor analysis (FA), absolute principle component score–multiple linear regression (APCS–MLR), and factor analysis–multiple regression (FA–MR) analysis are the commonly accepted

* Corresponding author. Tel.: +86 571 86971813; fax: +86 57186971359.
E-mail address: jw67@zju.edu.cn (J. Wu).

1474-7065/$ - see front matter
doi:10.1016/j.pce.2010.03.021

traditional multivariate methods to evaluate spatio-temporal variations in river pollution as well as identify the related responsible sources (Shin and Fong, 1999; Love et al., 2004; Singh et al., 2004; Zhou et al., 2007a; Shrestha and Kazama, 2007; Omo-Irabor et al., 2008; Huang et al., 2009). Multivariate analyses are sensitive to outliers and the non-normal distributions of geochemical data sets; thus, appropriate data pretreatment, including estimation of missing data, examination of normal distributions and data transformation should be taken into consideration, a principle ignored in most environmental studies (Peré-Trepat et al., 2006; Zhou et al., 2007c). Meanwhile, a multivariate study requires the integrated application of different multivariate methods to reciprocally validate the models obtained (Lattin et al., 2003; Zhou et al., 2007c). However, comprehensive application of different multivariable statistical methods and sources apportionment has not been fully explored in river studies in China (Zhang et al., 2009).

Qiantang River, located in Eastern China (Fig. 1), is the largest river in Zhejiang Province. It plays a critical role in the sustainable development of Yangtze River Delta, due to its multiple functions: water supply, electricity generation, irrigation, tourism, fishery and shipping. Qiantang watershed covers around 40,000 km^2 and has a population of 20 million. With the opening-up policies started in late of 1970', this area has followed its own development model, dubbed the "Zhejiang model". This is based on prioritizing and encouraging entrepreneurship, and an emphasis on small businesses responsive to the whims of the market, thus making rapid strides in its economic and social undertakings. However, with the rapid development of economy and explosion of population in recent decades, the river has been polluted by domestic and industrial effluents. The quality of water from the river is thus deteriorating (Han et al., 2007), but few detailed studies have been carried out to distinguish the contributing sources in Qiantang River.

This paper addresses the above concern and attempts to understand the spatio-temporal patterns of water quality in Qiantang River as well as the pollution sources. This paper is a further study of Huang et al. (2009). Huang et al. (2009) studied the spatial variations of water quality in Qiantang River for the year 2004, but the temporal dynamics of water quality remained unknown. We therefore extend the study through analyzing a 4-year (2001–2004) data set subjected to neural-based modeling and different multivariate methods (DA, PCA, APCS–MLR).

More specifically, this study aims at evaluating spatio-temporal patterns in water quality and extracting information about the apportionment of pollution in Qiantang River. First, after data pretreatment, a self-organizing map and DA were used to determine spatial patterns, temporal variations and the corresponding significant variables in water quality; second, PCA was used to identify underlying pollution sources; and finally, APCS–MLR was applied to further estimate source contributions for each pollution variable.

2. Materials and methods

2.1. Data

Thirteen parameters [ammonia nitrogen (NH_3–N), fluoride (F^-), permanganate index (COD_{Mn}), hexavalent chromium (Cr^{6+}), vola-

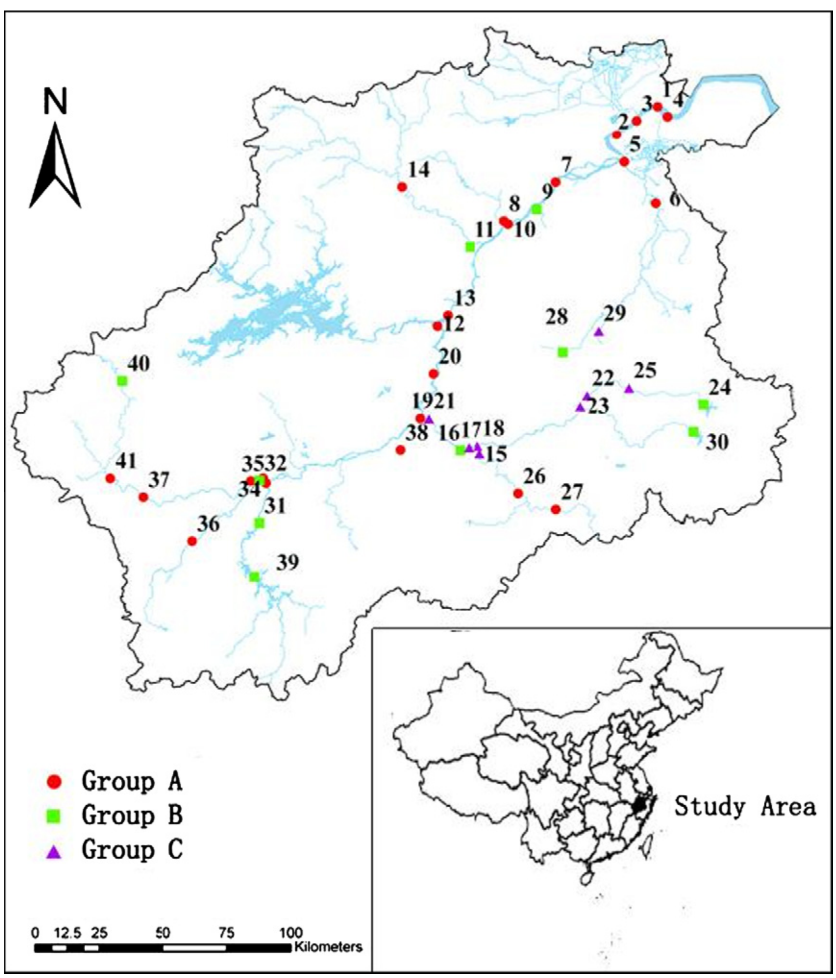

Fig. 1. Location of Qiantang River and spatial pattern of the three classified groups derived by self-organizing map.