Marine Pollution Bulletin 64 (2012) 688–698

Contents lists available at SciVerse ScienceDirect

Marine Pollution Bulletin

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

Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques

Nabeel M. Gazzaz *, Mohd Kamil Yusoff, Mohammad Firuz Ramli, Ahmad Zaharin Aris, Hafizan Juahir

Department of Environmental Science, Faculty of Environmental Studies, University Putra Malaysia, 43400 Serdang, Selangur Darul Ehsan, Malaysia

ARTICLE INFO

Keywords:
Surface water quality
Kinta River
Chemometrics
Factor analysis
Cluster analysis
Discriminant function analysis

ABSTRACT

This study employed three chemometric data mining techniques (factor analysis (FA), cluster analysis (CA), and discriminant analysis (DA)) to identify the latent structure of a water quality (WQ) dataset pertaining to Kinta River (Malaysia) and to classify eight WQ monitoring stations along the river into groups of similar WQ characteristics. FA identified the WQ parameters responsible for variations in Kinta River's WQ and accentuated the roles of weathering and surface runoff in determining the river's WQ. CA grouped the monitoring locations into a cluster of low levels of water pollution (the two uppermost monitoring stations) and another of relatively high levels of river pollution (the mid-, and down-stream stations). DA confirmed these clusters and produced a discriminant function which can predict the cluster membership of new and/or unknown samples. These chemometric techniques highlight the potential for reasonably reducing the number of WQVs and monitoring stations for long-term monitoring purposes.

1. Introduction

Anthropogenic disturbance of the land surface and modifications to river systems besides the ever-increasing rates of water consumption are adversely impacting the quality of freshwater resources throughout the world (Singh et al., 2009). Domestic, agricultural, and industrial waste discharges continue to pollute the rivers, directly through surface runoff and indirectly through drains and river tributaries. On the other hand, the negative impacts of anthropogenic activities on the river systems are cumulative in nature. They are caused by processes and activities which accumulate over time and space (Ren et al., 2003).

The impacts of point and non-point sources of pollution on river water quality (RWQ) are vital issues in many parts of the world; especially in areas like the current study area (Kinta River basin, the State of Perak (Malaysia)) where urban development is continuously transforming forested watersheds into urban and mixed-use ones. The dramatic changes in land uses (LUs), mainly extended housing and the associated commercial and industrial development, in Kinta River basin have resulted in deleterious environmental impacts on the local water resources because of increased demand on water and high inputs of sediments, nutrients, organics, and microbes to the river.

Monitoring of river WQ and analysis of its spatial patterns and temporal trends are issues of great environmental concern. In light of the temporal and spatial variations in the biological and physico-chemical characteristics of streams and rivers, regular WQ monitoring programs are needed for reliable evaluation of the surface WQ. Currently, many countries conduct regular monitoring of the WQ of their important water bodies and spatio-temporal analysis of RWQ has been practiced as an efficient tool for WQ assessment. However, these programs usually generate bulky, complex datasets constituting large numbers of samples and WQ parameters whose analysis and interpretation using traditional monovariate and bivariate statistical methods can be far from complete. Within this context, previous research demonstrated that chemometric methods are useful tools for extraction of considerable, meaningful information from environmental data. For instance, the chemometric techniques of principal factor analysis (PFA), hierarchical cluster analysis (HCA), and discriminant function analysis (DFA) have been frequently employed to evaluate, and to examine the spatial patterns and temporal trends in, the WQ of rivers (Kowalkowski et al., 2006; Shrestha and Kazama, 2007; Simeonov et al., 2002; Singh et al., 2004; Su et al., 2011), groundwater (Lambrakis et al., 2004; Singh et al., 2005; Vončina et al., 2007), and coastal water (Kuppusamy and Giridhar, 2006; Zhou et al., 2007b), amongst other water systems. Thereupon, this

Abbreviations: asl, above sea level; CA, cluster analysis; DF, discriminant function; DFA, discriminant function analysis; DoE, Department of Environment, Malaysia; FA, factor analysis; HCA, hierarchical cluster analysis; HACA, hierarchical agglomerative cluster analysis; EHR, expected hit ratio; HR, hit ratio; KMO, Kaiser–Meyer–Olkin; LU, land use; max., maximum; min., minimum; NE, northeast; PCA, principal component analysis; PCCs, percent of correct classifications; PFA, principal factor analysis; R^2, coefficient of determination; RWQ, river water quality; SEM, standard error of the mean; SW, southwest; WQ, water quality; WQV, water quality variable.

* Corresponding author. Tel.: +60 123 579 864; fax: +60 3 8946 7408.
E-mail address: NabeelMGazzaz@Yahoo.com (N.M. Gazzaz).

0025-326X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.marpolbul.2012.01.032