Gazzaz, N.M., Kamil, M., Hafizan, J., Firuz, M., and Zaharin, A. 2013. Water quality assessment and analysis of spatial patterns and temporal trends. Water Environment Research 85 (8): 751-767.

Water Quality Assessment and Analysis of Spatial Patterns and Temporal Trends

Nabeel M. Gazzaz[1*], Mohd Kamil Yusoff[1], Hafizan Juahir[2], Mohammad Firuz Ramli[1], Ahmad Zaharin Aris[2]

ABSTRACT: This study investigated relationships of a water quality index (WQI) with multiple water quality variables (WQVs), explored variability in water quality over time and space, and established linear and non-linear models predictive of WQI from raw WQVs. Data were processed using Spearman's rank correlation analysis, multiple linear regression, and artificial neural network modeling. Correlation analysis indicated that from a temporal perspective, the WQI, temperature, and zinc, arsenic, chemical oxygen demand, sodium, and dissolved oxygen concentrations increased, whereas turbidity and suspended solids, total solids, nitrate nitrogen (NO_3-N), and biochemical oxygen demand concentrations decreased with year. From a spatial perspective, an increase with distance of the sampling station from the headwater was exhibited by 10 WQVs: magnesium, calcium, dissolved solids, electrical conductivity, temperature, NO_3-N, arsenic, chloride, potassium, and sodium. At the same time, the WQI; *Escherichia coli* bacteria counts; and suspended solids, total solids, and dissolved oxygen concentrations decreased with distance from the headwater. Lastly, regression and artificial neural network models with high prediction powers (81.2% and 91.4%, respectively) were developed and are discussed. *Water Environ. Res.*, **85,** 751 (2013).

KEYWORDS: Kinta River, water quality, archived data, statistical analysis, screening and pretreatment, spatial patterns, temporal trends, linear regression, artificial neural network.

doi:10.2175/106143013X13596524516347

Introduction

Rivers represent the primary inland water resources for domestic, irrigation, and industrial uses. Because they exhibit wide temporal and spatial differences in hydrochemistry, it is vital to prevent and control river pollution and to obtain reliable information on water quality for effective water management. However, as a result of historical and ongoing anthropogenic activities on surface water systems, tracking water quality status and trends requires a systematic and well-planned monitoring program to both identify problems and provide strategies for their resolution.

The term *water quality* describes biological, chemical, and physical characteristics of water in connection with intended use and a set of standards (Liou et al., 2004). Further, a *water quality assessment* of these characteristics includes consideration of human health effects and planned water uses—typically through measurement of a variety of parameters from several monitoring stations at different times (Fernández et al., 2004; Pesce and Wunderlin, 2000). Although long-term surveys and monitoring programs of water quality provide an adequate approach to develop detailed knowledge of river hydrochemistry and pollution, they produce huge and complex data matrices of physico-chemical parameters that are often difficult to interpret (Astel et al., 2006).

Problems of data reduction and interpretation of multi-constituent physical and chemical measurements can be approached by application of a variety of statistical methods, including exploratory data analysis (Tuppad et al., 2010) that can support an assessment of the underlying mechanisms of water quality. For example, correlation and regression statistical techniques can be used to describe natural associations between variables or samples, and reveal information and relationships that otherwise might not be identified.

Spatial and temporal analysis of river water quality has been, in some cases, successfully performed through simple comparison of descriptive statistics of water quality variables between sampling locations and over time (e.g., Hall and Killen, 2005; Maane-Messai et al., 2010). In addition, the multivariate statistical (or chemometric) technique of cluster analysis has been used widely for similar purposes. The objective of cluster analysis is to assign observations to clusters (groups) such that observations within one group are similar to each other with respect to the attributes or variables of interest, but are different from observations in the other groups (i.e., cluster analysis categorizes observations into distinct, homogeneous groups). To this end, cluster analysis has been widely used in surface water quality studies for recognition of patterns (for example combining water systems or water quality monitoring stations into groups of similar water quality characteristics (e.g., Desai et al., 2010; Dong et al., 2010; Iscen et al., 2008; Zhang et al., 2009; Zhou et al., 2007)) and analysis of trends, i.e., categorizing sampling times or dates in terms of months, seasons, or years (e.g., (Dong et al., 2010; Wunderlin et al., 2001; Zhou et al., 2007; Zhang et al., 2009). These types of water quality analyses have also been addressed, albeit to a much lesser extent, using self-organizing feature maps (e.g., Astel et al., 2007; Sánchez-Martos et al., 2002; Su et al., 2011).

With regard to assessing underlying regularities in the water quality variables themselves, two types of analyses are typically used—factor analysis and correlation analysis. Factor analysis is a powerful pattern recognition technique that attempts to explain the variance of a large set of inter-correlated variables through transformation into a small set of independent

[1*] Department of Environmental Sciences, Faculty of Environmental Studies, Universiti Putra Malaysia, 43400 Serdang, Selangor Darul Ehsan, Malaysia; e-mail: NabeelMGazzaz@Yahoo.com.

[2] Center of Excellence for Environmental Forensics, Faculty of Environmental Studies, Universiti Putra Malaysia, Selangor Darul Ehsan, Malaysia.