Water Research 39 (2005) 2621–2635

WATER RESEARCH

www.elsevier.com/locate/watres

Evaluation of river water quality monitoring stations by principal component analysis

Ying Ouyang*

Department of Water Resources, St. Johns River Water Management District, Reid Street, Palatka, Florida 32178-1429, USA

Received 10 August 2004; received in revised form 18 March 2005; accepted 4 April 2005

Abstract

The development of a surface water monitoring network is a critical element in the assessment, restoration, and protection of stream water quality. This study applied principal component analysis (PCA) and principal factor analysis (PFA) techniques to evaluate the effectiveness of the surface water quality-monitoring network in a river where the evaluated variables are monitoring stations. The objective was to identify monitoring stations that are important in assessing annual variations of river water quality. Twenty-two stations used for monitoring physical, chemical, and biological parameters, located at the main stem of the lower St. Johns River in Florida, USA, were selected for the purpose of this study. Results show that 3 monitoring stations were identified as less important in explaining the annual variance of the data set, and therefore could be the non-principal stations. In addition, the PFA technique was also employed to identify important water quality parameters. Results reveal that total organic carbon, dissolved organic carbon, total nitrogen, dissolved nitrate and nitrite, orthophosphate, alkalinity, salinity, Mg, and Ca were the parameters that are most important in assessing variations of water quality in the river. This study suggests that PCA and PFA techniques are useful tools for identification of important surface water quality monitoring stations and parameters.

Keywords: Monitoring network; Principal component analysis; Surface water quality

1. Introduction

Pollution of surface water with toxic chemicals and excess nutrients, resulting from storm water runoff, vadose zone leaching, and groundwater discharges, has been an issue of worldwide environmental concern. With an increased understanding of the importance of drinking water quality to public health and raw water quality to aquatic life, there is a great need to assess surface water quality. This is true for the lower St. Johns River (LSJR), located in Florida, USA. Pollution of the LSJR with contaminants such as nutrients, hydrocarbons, pesticides, and heavy metals comes from both point and non-point sources. These sources are the results of surface runoff generated from urban, rural, and agricultural lands; discharge from ditches and creeks; groundwater seepage from malfunctioning septic tank systems; aquatic weed control and naturally occurring organic inputs; and atmospheric deposition. The degradation of water quality due to these contaminants has resulted in altered species composition and decreased the overall health of aquatic communities within the river basin (Campbell et al., 1993; Durell et al., 2001; Ouyang et al., 2002).

*Tel.: +1 386 312 2320; fax: +1 386 329 4585.
E-mail address: youyang@sjrwmd.com.

0043-1354/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.watres.2005.04.024

In 1987, the Florida State Legislature enacted the Surface Water Improvement and Management (SWIM) Act, which identified the LSJR as a water body of regional significance and in need of restoration and management. Since then, many efforts have been devoted to restoring and protecting the LSJR. While the control of surface water runoff, ditch discharge, vadose zone leaching, groundwater seepage, and atmospheric deposition is necessary to reduce point and non-point source pollution of the river, development of an effective surface water monitoring network is a critical element in these restoration and protection efforts. Surface water quality monitoring within the LSJR has been conducted by various agencies at varying levels of intensity since 1956. The primary objectives are to identify water quality problems, describe seasonal and spatial trends for developing qualitative and quantitative models of the riverine ecosystem, and determine permit compliance. Since its inception, the monitoring network has become one of the most critical efforts in assessment of surface water pollution in the LSJR and has been a significant resource for others working to prevent pollution of the river (Campbell et al., 1993). However, efforts to determine the effectiveness and efficiency of the monitoring network are still warranted. To this end, the principal component analysis (PCA) and principal factor analysis (PFA) techniques were employed in this study.

PCA and PFA are multivariate statistical techniques used to identify important components or factors that explain most of the variances of a system. They are designed to reduce the number of variables to a small number of indices (i.e., principal components or factors) while attempting to preserve the relationships present in the original data. The problems of indicator parameter or import monitoring station identification, data reduction and interpretation, and characteristic change in water quality parameters can be approached through the use of the PCA and PFA. Details for mastering the arts of PCA and PFA are published elsewhere (Manly, 1986; Davis, 1986; Wackernagel, 1995; Tabachnick and Fidell, 2001).

In recent years, the PCA and PCF techniques have been applied to a variety of environmental issues, including evaluation of ground water monitoring wells, interpretation of groundwater hydrographs, examination of spatial and temporal patterns of heavy metal contamination and identification of herbicide species related to hydrological conditions. Some examples of PCA and PCF applications in environmental practices are described below.

Measurements of water level in wells are a routine part of groundwater studies. Recently, Winter et al. (2000) applied the PCA and PCF techniques to investigate the areal distribution of various types of water level fluctuation patterns within a study area and to determine if fewer wells could be measured while still achieving effective long-term monitoring goals at four small, lake-watershed research sites in the USA. These authors found that the PCA technique was very useful in summarizing information from large data sets to select long-term monitoring wells, which would greatly reduce the cost of monitoring programs.

Gangopadhyay et al. (2001) applied the PCA and PCF techniques to identify monitoring wells important in predicting the dynamic variations in potentiometric head at a location in Bangkok, Thailand. Through the years, the groundwater monitoring networks in the area have expanded tremendously, and many networks today consist of dozens, if not hundreds, of sampling wells. These authors argued that at a certain stage, municipalities have to justify their groundwater monitoring networks and ask questions such as how sampling from a particular well can help explain the dynamic variations of potentiometric head in the aquifer, and which subset of observation wells should be selected to continue monitoring in the near future for a municipality facing budget constraints. To answer these questions, the authors performed PCA on all of the monitoring wells, and developed a ranking scheme based on the frequency of occurrence of a particular well as a principal well. Based on the study results, the decision maker with budget constraints can now opt to monitor only the principal wells and still adequately capture the potentiometric head variations in the aquifer.

Additionally, the PCA technique has been used to estimate spatial and temporal patterns of heavy metal contamination (Shine et al., 1995); to investigate nutrient gradients within a eutrophic reservoir (Perkins and Underwood, 2000); and to identify the major herbicide compositions causing the observed data variations (Tauler et al., 2000). These studies have provided good examples of the effective application of PCA. However, there are few documented examples of the evaluation of the highly dynamic and complex surface water quality monitoring networks in river systems using the PCA or PFA technique.

The aims of this study are to demonstrate the application of these novel data reduction techniques (i.e., the PCA and PFA techniques) to evaluate the potential for reducing the number of ambient water quality monitoring stations located in the main stem of the LSJR for long-term monitoring purposes and to evaluate the importance of various water quality parameters. The specific objectives are to: (1) present detailed procedures on how to interpret PCA and PFA results, (2) identify the non-principal surface water quality monitoring stations, and (3) extract the parameters that are most important in assessing variations in river water quality.