Journal of Environmental Management 151 (2015) 87–96

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Modeling water quality in an urban river using hydrological factors – Data driven approaches

Fi-John Chang [a,*], Yu-Hsuan Tsai [a], Pin-An Chen [a], Alexandra Coynel [b], Georges Vachaud [c]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC
[b] Laboratoire d'Environnements et Paléoenvironnements Océaniques et Continentaux, University Bordeaux 1, UMR EPOC, France
[c] Laboratoire Transferts en Hydrologie et Environnement, LTHE, UMR 5564 CNRS-IRD-UJF, Grenoble, France

A R T I C L E   I N F O

Article history:
Received 9 May 2014
Received in revised form
10 November 2014
Accepted 4 December 2014
Available online

Keywords:
Artificial neural network (ANN)
Water quality
River basin management
Ammonia nitrogen (NH_3–N)
Nonlinear autoregressive with exogenous input (NARX) network
Gamma test

A B S T R A C T

Contrasting seasonal variations occur in river flow and water quality as a result of short duration, severe intensity storms and typhoons in Taiwan. Sudden changes in river flow caused by impending extreme events may impose serious degradation on river water quality and fateful impacts on ecosystems. Water quality is measured in a monthly/quarterly scale, and therefore an estimation of water quality in a daily scale would be of good help for timely river pollution management. This study proposes a systematic analysis scheme (SAS) to assess the spatio-temporal interrelation of water quality in an urban river and construct water quality estimation models using two static and one dynamic artificial neural networks (ANNs) coupled with the Gamma test (GT) based on water quality, hydrological and economic data. The Dahan River basin in Taiwan is the study area. Ammonia nitrogen (NH_3–N) is considered as the representative parameter, a correlative indicator in judging the contamination level over the study. Key factors the most closely related to the representative parameter (NH_3–N) are extracted by the Gamma test for modeling NH_3–N concentration, and as a result, four hydrological factors (discharge, days w/o discharge, water temperature and rainfall) are identified as model inputs. The modeling results demonstrate that the nonlinear autoregressive with exogenous input (NARX) network furnished with recurrent connections can accurately estimate NH_3–N concentration with a very high coefficient of efficiency value (0.926) and a low RMSE value (0.386 mg/l). Besides, the NARX network can suitably catch peak values that mainly occur in dry periods (September–April in the study area), which is particularly important to water pollution treatment. The proposed SAS suggests a promising approach to reliably modeling the spatio-temporal NH_3–N concentration based solely on hydrological data, without using water quality sampling data. It is worth noticing that such estimation can be made in a much shorter time interval of interest (span from a monthly scale to a daily scale) because hydrological data are long-term collected in a daily scale. The proposed SAS favorably makes NH_3–N concentration estimation much easier (with only hydrological field sampling) and more efficient (in shorter time intervals), which can substantially help river managers interpret and estimate water quality responses to natural and/or manmade pollution in a more effective and timely way for river pollution management.

1. Introduction

Water quality has deteriorated in most of the major rivers in western Taiwan over the past decades in consequence of urbanization, industrialization and population growth along rivers. Seasonal variations of river flows have also undergone drastic changes due to hydrological and geological characteristics of river basins. Pollutants leak from various sources and accumulate with sediments in river beds. In drought seasons, water levels in general are very low and flows barely occur in river channels. As a consequence, river pollution becomes even worse. On the other hand, during wet seasons, sudden changes in river flow may rinse river beds, deposit particles of sediments and contaminants, and thus river water quality seriously deteriorates and ecosystems encounter huge impacts (Ko et al., 2010). Hydrological characteristics significantly influence the ecological sustainability of aquatic river systems and cause heavy casualties on fishes, shellfishes and mollusks in downstream industries and coastal cultivation in

* Corresponding author. Tel.: +886 2 33663452; fax: +886 2 23635854.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

http://dx.doi.org/10.1016/j.jenvman.2014.12.014
0301-4797/

Taiwan (Chang et al., 2013). Water pollution is a crucial problem in the Dahan River (our study area) because many industrial facilities as well as densely populated cities are located along the river in recent decades and uneven dilution as well as transportation of pollutants frequently occurs in the water body of the river. The incessant accumulation of pollutants and nutrients deteriorates river water quality and exhausts dissolved oxygen in the river (Chiu, 2011). Yang et al. (2009) applied the WASP/EUTOR model to evaluating a number of alternatives on wastewater management in the downstream of the Dahan River for river restoration with an improvement on the assimilative capacity of biochemical oxygen demand and dissolved oxygen.

Water quality models are useful tools for estimating the impacts and risks of chemical pollutants in a water body (Chapra, 2008; Feng et al., 2013). Water quality models can be classified into physically-based or statistical approaches. Physically-based approaches have progressively proved successful and useful in learning the mechanisms of underlying processes; nevertheless, they are usually site-specific and require substantially detailed water quality measurements and/or extensive surveys for calibration, which bear certain time and budget limitation. With the development of model theory and the fast-updated computer techniques, more water quality models have been explored with various statistical methods to overcome data scarcity and simultaneously increase model reliability. Either statistical approaches, such as linear regression (Rothwell et al., 2010); factor analysis technique (Ouyang et al., 2000), or artificial neural networks (ANNs) models based on data driven techniques (Unwin et al., 2010; He et al., 2011) can be applied to the monitored time-series of hydrological and water quality measurements for simulation and/or prediction purposes. ANNs are computational techniques inspired by the brain and nerve systems in biological organisms, and they can tackle large-scale complex problems. ANNs have also been applied with success to diverse fields of environment sciences (Coz et al., 2009; McNamara et al., 2008; Yesilnacar et al., 2008; Singh et al., 2009; Chang et al., 2010; Giri et al., 2011; Hattab et al., 2013; Jiang et al., 2013; Tsai. et al., 2014). A majority of studies were dedicated to exploring the applicability of static ANNs, such as the back propagation neural network (BPNN) and the adaptive network-based fuzzy inference system (ANFIS). Nevertheless, the natural characteristics of hydrogeological processes are complex and dynamic. Static neural networks might fail to properly predict the dynamical features of hydrogeological processes, such that the delivered relationship might be simply the possible impacts of factors on the temporal characteristics of local environments (Chang et al., 2013; Chen et al., 2013). Consequently, the comprehensive analysis on the dynamic features of hydrogeological processes and the modeling of tempo-spatial water quality variations remain a great challenge that needs to be overcome.

Environmental sampling is very complicated, laborious, costly and time-consuming. It is unlikely to have continuous long-term water-quality time series data with complete properties at all sampling locations in a river system. Another great challenge for river managers in pollution assessment is the investigation of pollution patterns with high complexity, dynamism and non-linearity in both spatial and temporal scales (Carafa et al., 2007). Management tools require predictive methods or models to relate water quality with hydrological responses, catchment characteristics and/or human activities. We attempt to estimate water quality concentration in shorter time intervals of interest, without conducting water quality field sampling. The proposed scheme comprises artificial neural networks, factor selection and statistics techniques for a comprehensive assessment of river water quality in responses to natural and human activities over the study basin.

2. Methods

This study formulates a systematic analysis scheme (SAS) for assessing the spatio-temporal water quality by artificial intelligence and statistics techniques based on water quality, hydrological and economic data (Fig. 1). A preliminary analysis is conducted through the correlation coefficient analysis to identify the representative water quality parameter (i.e., a correlative indicator of the contamination level over the study area). Three ANNs (BPNN-a classical ANN; ANFIS-a neuro-fuzzy network; and NARX-a dynamic ANN) coupled with the Gamma test (for factor selection) and cross-validation (for data scarcity) are used for modeling the concentration of the representative water quality parameter. The merits of the main methodologies are briefly addressed as follows:

2.1. Gamma test (GT) - factor selection

The use of input selection methods assist in selecting the combination of explanatory variables best suit a model. The proper selection of input variables can improve prediction performance and help understand the processes that resulted in the observed data (Guyon and Elisseeff, 2003). The current study involves water quality, hydrological and economic factors with data sets limited in size, and therefore there is a need to utilize effective input selection methods to characterize the appropriate input–output relationships. The GT, presented by Agalbjorn et al. (1997), is used to estimate the noise level in a data set without assuming any parametric form of equations that govern the system. The only requirement is that the system should be governed by a smooth function because the GT will exploit the hypothesized continuity of this governing function. Performing a single GT is a fast procedure, which can provide the noise estimate for each subset (combination) of input variables. If a subset's associated noise estimate (Γ value) is the closest to zero, it can be considered as the “best combination” of inputs. Recent applications noted that ANNs combined with the GT can obtain accurate estimation based on the identified non-trivial

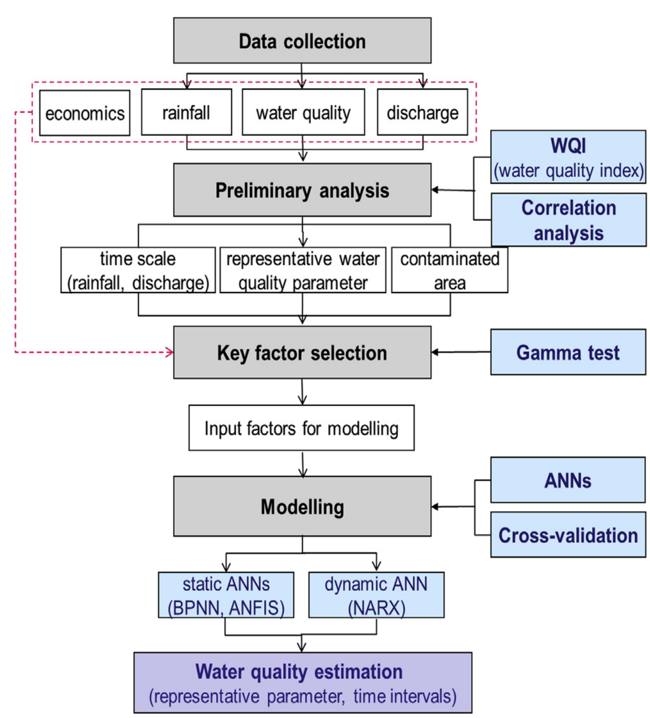

Fig. 1. Study flow of the proposed systematic analysis scheme (SAS).