Science of the Total Environment 562 (2016) 228–236

Contents lists available at ScienceDirect

Science of the Total Environment

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

Estimating spatio-temporal dynamics of stream total phosphate concentration by soft computing techniques

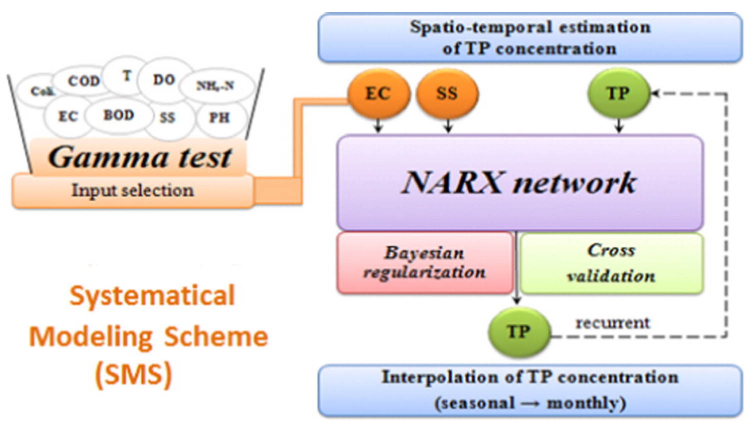
Fi-John Chang [a,*], Pin-An Chen [a], Li-Chiu Chang [b], Yu-Hsuan Tsai [a]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Department of Water Resources and Environmental Engineering, Tamkang University, New Taipei City 25137, Taiwan, ROC

HIGHLIGHTS

- Systematical modeling scheme (SMS) is proposed to make spatio-temporal TP estimation.
- SMS comprises a dynamic neural network and three refined statistical methods.
- SMS can select key input factors, solve data scarcity, and avoid over-fitting.
- SMS suitably estimates site-specific TP conc. at 7 stations along a river simultaneously.
- SMS adequately reconstructs seasonal TP data into a monthly scale in a river system.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 9 October 2015
Received in revised form 24 March 2016
Accepted 28 March 2016
Available online xxxx

Editor: Simon Pollard

Keywords:
Total phosphate (TP)
Water quality
Artificial neural network (ANN)
Nonlinear autoregressive with eXogenous input (NARX) network
Gamma test

ABSTRACT

This study attempts to model the spatio-temporal dynamics of total phosphate (TP) concentrations along a river for effective hydro-environmental management. We propose a systematical modeling scheme (SMS), which is an ingenious modeling process equipped with a dynamic neural network and three refined statistical methods, for reliably predicting the TP concentrations along a river simultaneously. Two different types of artificial neural network (BPNN—static neural network; NARX network—dynamic neural network) are constructed in modeling the dynamic system. The Dahan River in Taiwan is used as a study case, where ten-year seasonal water quality data collected at seven monitoring stations along the river are used for model training and validation. Results demonstrate that the NARX network can suitably capture the important dynamic features and remarkably outperforms the BPNN model, and the SMS can effectively identify key input factors, suitably overcome data scarcity, significantly increase model reliability, satisfactorily estimate site-specific TP concentration at seven monitoring stations simultaneously, and adequately reconstruct seasonal TP data into a monthly scale. The proposed SMS can reliably model the dynamic spatio-temporal water pollution variation in a river system for missing, hazardous or costly data of interest.

1. Introduction

Total phosphate (TP) is a combination of orthophosphate, polyphosphate and organic phosphate and is considered as an index

* Corresponding author.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

http://dx.doi.org/10.1016/j.scitotenv.2016.03.219
0048-9697/

in the representation of the phosphorus quantity in river water. Phosphorus are heavily used in fertilizers and are often introduced to surface water through runoff resulting high nutrient concentrations in runoff water (Allen and Mallarino, 2008; Smith et al., 2007; Schoumans et al., 2014). Additional phosphate transported in runoff can also result from the wash-off of burned or managed forest soil (Santos et al., 2015), and agricultural nonpoint phosphorus sources are becoming the major source of phosphorus in water (Li et al., 2015). Excessive phosphorus would degrade water quality, harm ecosystems and result in algae bloom and economic damages (Davis and Koop, 2006; Dodds et al., 2009; Santos et al., 2015). The enrichment of TP in natural water can evoke declines in water quality, changes in biotic population structures and low dissolved oxygen concentrations in rivers. Even though the pivotal role of phosphorus for eutrophication has been recognized for decades, excessive phosphorus inputs in water bodies still make it difficult to maintain a good ecological status. The Water Quality Index (WQI) adopted by the Environmental Protection Administration (EPA) in Taiwan is used to assess the general conditions of water bodies in rivers and reservoirs. The WQI numerically summarizes the information of multiple water quality parameters into a single value, including dissolved oxygen (DO saturation), coliform group (fecal), acidity (pH), five-day biochemical oxygen demand (BOD_5), ammoniacal nitrogen (NH_3-N), suspended solids (SS) and TP. The calculation of the WQI can be found on the governmental website http://www.tydep.gov.tw/TYDEP/Static/river/main4.html. It is noted that the water quality parameters incorporated in the WQI are measured monthly in Taiwan, except for TP (measured quarterly). Thus, there is a request to incorporate TP into the monthly WQI for a more comprehensive assessment on the irregular variation of river water quality induced by natural processes and/or human activities. Therefore, a model for generating monthly TP estimates is desired.

Water quality management commonly involves a series of complex inter-disciplinary decisions based on speculated responses of water quality to changing controls (McIntyre and Wheater, 2004). Environmental data characterize complex spatio-temporal dynamics. Various water quality models have been developed to explore the complex dynamics and to estimate the levels and risks of pollutants in a water body and widely used for in environmental science and management research (Liu et al., 2011; Papadakis et al., 2015; Park et al., 2014; Santos et al., 2015; Visser et al., 2012; Wang et al., 2013; Jalba et al., 2014; Wellen et al., 2015). Building a model usually requires long sufficient field data to maintain model applicability and reliability. Water quality models, in general, can be classified into mechanistic (process-based) models and data-driven (empirical) models. Mechanistic models usually require defining general frameworks, and they could explicitly explore and/or identify the relationships between waste loads from different sources and the resulting water qualities of the receiving waters (as stated in Santos et al., 2015). Process-based models are considered useful and beneficial in a wide variety of applications, while they are usually data-intensive. As known, water quality monitoring programs are expensive and time-consuming such that it often encounters field data scarcity during modeling, which increases the difficulty in producing reliable water quality estimates within acceptable error ranges and hinders the efforts of modeling, and therefore affects the process of water quality management. Data-driven models are usually inferred from raw or processed data, and their formulation may not be conceptually supported by the mechanism of the phenomenon under consideration. In many situations when availability of water quality data is limited, empirical approaches can be useful and essential to characterize the complex dynamics of pollution loadings. Thus, the development of adequate models in data-poor environments is a current necessity. Being ascribed to the fast-updating soft computing techniques, artificial neural networks (ANNs) are effective data-driven techniques to validly model input–output patterns of real-time and/or collected historical data sets without knowing a detailed mathematical function of the dynamic system. ANNs provide viable tools for modeling complex and nonlinear dynamical systems with great flexibility and capability. In the last decades, ANNs have been widely applied with success in the field of environment and water resources modeling (Bertin et al., 2013; Chang et al., 2013; Chang et al., 2014a, 2014b; Chen et al., 2013; Chen et al., 2014; Fontes et al., 2014; Gevrey et al., 2010; King et al., 2014; Park et al., 2015; Noori et al., 2013; Subida et al., 2013; Tsai et al., 2014; Yang et al., 2012). Due to dynamic features, great uncertainty, complex relation, and/or lack of monitoring data in most of environmental issues, a dynamic neural network, which generally could be more adaption with system dynamic and produce more stable forecast than the static neural network (Chiang, et al., 2004; Noori et al., 2015 & Noori et al., 2016; Tayarani-Bathaie et al., 2014; Yazdizadeh and Khorasani, 2002), for tackling environmental management is desirable and beneficial. The occurrence of TP in surface water is related to various on-ground sources, and its physical and chemical processes are highly nonlinear dynamic and very hard to exhibiting behaved relationships. To the best of our knowledge, the applicability of soft computing techniques in estimating the dynamic spatial–temporal TP concentrations along a river basin has not been fully explored, which should be beneficial for future use.

This study presents an ingenious systematical modeling scheme (SMS) through soft computing techniques for estimating the spatial distribution of TP concentrations along the Dahan River in Taiwan. This river has suffered from a high level of phosphates in downstream due to the large populated cities, agricultural activities, animal husbandry farms and industrial facilities located along the river and its tributaries. The SMS is equipped with a dynamic neural network and three advanced statistical methods, which can effectively deliver non-linear solution for the spatial non-stationary predictions. The advanced statistical methods comprise: the Gamma test (GT) for extracting key input factors; the Cross-Validation technique for overcoming data scarcity; and the Bayesian regularization method for improving model generalizability. A static-ANN and a Multiple Linear Regression (MLR) are comparatively constructed to evaluate the effort of the recurrent output information in modeling the spatio-temporal dynamics. Besides, the constructed models can be favorably considered as an interpolation tool for reconstructing the seasonal TP data into monthly data, and thus monthly WQI (including TP) can be produced for environmental management.

2. Materials

2.1. Study area

Due to the geomorphological characteristics of rivers, the seasonal variation of steamflow in Taiwan is very high such that long-lasting low flows in drought seasons could bring a drastic increase in river pollution levels. Besides, industrial facilities and large populated cities along rivers would easily cause downstream pollution in rivers. In the last decades, the water quality of the Dahan River in Taiwan has deteriorated rapidly due to heavy pollutant loads from surrounding urban areas, and thus the Dahan River becomes one of the most polluted rivers in northern Taiwan. Fig. 1 shows the location of the Dahan River basin, which is divided into two zones based on land use morphology: the upstream zone (from Shihmen Dam to Yuanshan Weir: water quality monitoring stations S1–S3); and the downstream zone (from Yuanshan Weir to the confluence point of the Dahan River and the Xintian River: stations S4–S7). We notice that the intensive human activities in the downstream zone have generated contamination from agricultural, municipal and industrial sectors, which has introduced significant amount of TP and organic materials into the river and its tributaries. Fig. 2 illustrates the observed seasonal TP concentration of the downstream zone (S4–S7) in March of 2011, which shows the TP concentration increases as the density of farms and industrial facilities increase. The WQI indicates that the water quality of the upstream zone is better