Journal of Hydrology 499 (2013) 265–274

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Regional estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling

Fi-John Chang *, Pin-An Chen, Chen-Wuing Liu, Vivian Hsiu-Chuan Liao, Chung-Min Liao

Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 6 February 2013
Received in revised form 23 May 2013
Accepted 3 July 2013
Available online 15 July 2013
This manuscript was handled by Laurent Charlet, Editor-in-Chief, with the assistance of Jose Daniel Salas, Associate Editor

Keywords:
Arsenic (As)
Groundwater quality
NARX network
Gamma test
Bayesian regularization
Indicator kriging

SUMMARY

Arsenic (As) is an odorless semi-metal that occurs naturally in rock and soil, and As contamination in groundwater resources has become a serious threat to human health. Thus, assessing the spatial and temporal variability of As concentration is highly desirable, particularly in heavily As-contaminated areas. However, various difficulties may be encountered in the regional estimation of As concentration such as cost-intensive field monitoring, scarcity of field data, identification of important factors affecting As, over-fitting or poor estimation accuracy. This study develops a novel systematical dynamic-neural modeling (SDM) for effectively estimating regional As-contaminated water quality by using easily-measured water quality variables. To tackle the difficulties commonly encountered in regional estimation, the SDM comprises of a neural network and four statistical techniques: the Nonlinear Autoregressive with eXogenous input (NARX) network, Gamma test, cross-validation, Bayesian regularization method and indicator kriging (IK). For practical application, this study investigated a heavily As-contaminated area in Taiwan. The backpropagation neural network (BPNN) is adopted for comparison purpose. The results demonstrate that the NARX network (Root mean square error (RMSE): 95.11 $\mu g\, l^{-1}$ for training; 106.13 $\mu g\, l^{-1}$ for validation) outperforms the BPNN (RMSE: 121.54 $\mu g\, l^{-1}$ for training; 143.37 $\mu g\, l^{-1}$ for validation). The constructed SDM can provide reliable estimation ($R^2 > 0.89$) of As concentration at ungauged sites based merely on three easily-measured water quality variables (Alk, Ca^{2+} and pH). In addition, risk maps under the threshold of the WHO drinking water standard (10 $\mu g\, l^{-1}$) are derived by the IK to visually display the spatial and temporal variation of the As concentration in the whole study area at different time spans. The proposed SDM can be practically applied with satisfaction to the regional estimation in study areas of interest and the estimation of missing, hazardous or costly data to facilitate water resources management.

1. Introduction

Arsenic (As) contamination in groundwater has been reported and resulted in a massive epidemic of As toxication in several countries such as Bangladesh, Vietnam, Cambodia, China and Taiwan. It is estimated that approximately 57 million people drink As-contaminated groundwater with concentrations exceeding the drinking water standard recommended by the WHO (World Health Organization) (BGS-DPHE, 2001; Chakraborti et al., 2010). As pollution affects not only crop productivity and water quality but also the quality of water bodies, which threatens the health of animals and human beings by way of food chains. Long-term exposure to As in drinking water has been implicated in a variety of health concerns including cancers, cardiovascular diseases, diabetes and neurological effects (National Research Council, 1999). Blackfoot disease as well as cancers of the skin, bladder, lung and liver have been associated with drinking As-contaminated groundwater (Chiou et al., 1997; Rahman et al., 1999). As-contaminated groundwater is derived naturally from As-rich aquifer sediments, and the geochemistry of As can be rather complex (Stollenwerk, 2003). Various hydrogeological and biogeochemical factors affecting As concentration in groundwater have been detected, such as sediment mineralogy, microbial oxidation or reduction of As, groundwater recharge, groundwater flow paths (Ford et al., 2006; Wang et al., 2007, 2011; Xie et al., 2013), and the presence of fractures in bedrock formations (Ayotte et al., 2003; Liao et al., 2011). Even though the processes controlling the release of As into groundwater systems have been extensively discussed over the past decade, the exact chemical conditions and reactions leading to As mobilization still remain a subject of intense debate (Goovaerts et al., 2005; Polizzotto et al., 2006; Winkel et al., 2008). Moreover, the high variability of arsenic concentration can occur within a short distance and/or in different well depths due to the diversity in geology and geomorphology (Serre et al., 2003; Yu et al., 2003). Besides, the detection of As contamination in groundwater by using

* Corresponding author. Tel.: +886 2 23639461; fax: +886 2 23635854.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.07.008

graphite atomic absorption spectrophotometry or inductively coupled plasma mass spectroscopy can be manpower and cost intensive. Consequently, how to adequately estimate As concentrations in complex hydro-geological systems is a crucial and challenging issue.

Artificial neural networks (ANNs) are a biologically motivated method and are considered as powerful alternative computational approaches to modeling complex systems. In the last decades, ANNs have been widely applied with success to various water resources problems, such as rainfall–runoff modeling (Antar et al., 2006; Chiang et al., 2007), groundwater (Krishna et al., 2008; Nikolos et al., 2008), and water quality (Khalil et al., 2011; McNamara et al., 2008; Sahoo et al., 2006). Recurrent neural networks (RNNs) are powerful nonlinear models capable of extracting dynamic behaviors from complex systems through internal recurrence and have attracted much attention for years (Assaad et al., 2005; Chang et al., 2002; Chiang et al., 2004, 2010; Ma et al., 2008). The Nonlinear Autoregressive with eXogenous input (NARX) network (Lin et al., 1996), a subclass of RNNs, can suitably build the temporal relationship between input and output patterns because the network's input vector is cleverly built through two tapped-delay elements: one from the input signal and the other from the network's output (Menezes and Barreto, 2008). NARX networks were applied to various nonlinear systems (Ali, 2009; Ardalani-Farsa and Zolfaghari, 2010; Hong, 2012). However, its feasibility as a nonlinear tool for time series modeling and prediction of different disciplines such as hydrological systems and water quality assessment has not been fully explored yet. Therefore, the practical meaning and importance of recurrent connections from the NARX network's output when dealing with regional estimation problems will be explored in this study.

Groundwater quality parameters exhibit considerable spatial variability. Geostatistical methods are generally based on the regionalized variable theory that delineates the variation behavior in an area and exhibits both randomly and spatially structure properties (Matheron, 1963; Shin and Salas, 2000). One of the most important geostatistical methods is the kriging method, which is an interpolation method for deriving data at unsampled locations by considering the spatial dependence of samples. The kriging method has been applied to the modeling of spatiotemporal distributions in many disciplines such as hydrological problems (Bargaouia and Chebbib, 2009), mapping topsoil fertility (Webster and McBratney, 1987), and As contamination (Goovaerts et al., 2005; Juang and Lee, 1998; Liu et al., 2004). Geostatistical tools are increasingly coupled with the geographic information system (GIS) for applications that characterize spatiotemporal structures, and spatially interpolate scattered measurements are used to construct spatially exhaustive layers of information (Pijanowski et al., 2002; Goovaerts et al., 2005). The estimation of individual-level historical exposure of study participants to arsenic can be obtained from the visualized spatiotemporal information of the spatiotemporal mobility of study participants and their surrounding environment.

The hyper-endemic blackfoot disease in the Yun-Lin County of Taiwan has been verified to be associated with high As concentrations in groundwater (Chen et al., 1995; Chiou et al., 1997). The residents had long-term exposed themselves to As through various paths such as ingestion of aquacultural and agricultural products, and thus dangerously posed carcinogenic risks to their health (Liu et al., 2008). Due to great concern for the potential effects of As on human health, there is a growing need for efficiently modeling the spatial distribution of As contamination in groundwater. One of the popular modeling approaches in use is the multiple linear regression (MLR), but this approach, however, may fail to estimate the spatial distribution of As contamination due to the great variability of As concentration and complex nonlinear processes involved in geology and geomorphology. Lately, using ANNs for the prediction of heavy metal concentration in groundwater has been attempted and gained a reasonably good degree of success (Chang et al., 2010; Cho et al., 2011; Giri et al., 2011; Mondal et al., 2012; Purkait et al., 2008). The modeling results indicated that ANN techniques could produce higher prediction accuracy than the conventional methods such as MLR. These studies were mostly dedicated to exploring the applicability of static ANNs, such as the back propagation neural network (BPNN), for building the relationship between As concentration in groundwater and hydro-geological parameters in arsenic-affected areas. Nevertheless, the natural characteristics of hydrogeological processes are not only complex but also dynamic. The static neural networks might fail to establish reliable models for predicting the dynamical features, such that the delivered relationship might be simply the possible impacts of factors on temporal characteristics of local environments. Consequently, the comprehensive analysis of dynamic hydrogeological features and estimation of the variability in As concentration over arsenic-affected regions remains a great challenge that needs to be overcome.

To construct a reliable estimation model of case competence, it is important to understand the impacts of factors on real competence, the interaction and evolvement of factors within an operation system, and the measurements of factors. In this study, we aim to present a novel model of case competence with good accuracy and predictability, in which certain assumptions are made for the nature of cases and case-bases. Consequently, a novel systematical dynamic-neural modeling (SDM) incorporated with a dynamic ANN and four advanced statistical techniques is developed to build a regional As concentration estimation model for decommissioned wells based on the easily-measured water quality variables of nearby functioning wells. The proposed SDM is expected to offer an applicable and useful reference to decision makers for dealing with groundwater management and preventing residents from drinking or using toxic groundwater.

2. Materials

2.1. Study area

Yun-Lin County is located in the southwestern alluvial fan of the Chou-Shui River in Central Taiwan (Fig. 1). Based on hydrogeological settings, the southern Choushui River alluvial fan is classified mainly into the proximal-fan, the mid-fan and the distal-fan areas (Central Geological Survey, 1999), in which the coastal region of the Yun-Lin County is located in the distal-fan area. The hydrogeological formation of the distal-fan can be divided into six inter-layered sequences: three marine sequences and three non-marine sequences. The non-marine sequences with coarse sediments (from medium sand to highly permeable gravel) are considered as aquifers, whereas the marine sequences with fine sediments are considered as aquitards. The annual average precipitation is 1417 mm and mainly occurs during wet season (i.e. May and September). Aquaculture is the primary revenue source for the inhabitants in the coastal region of Yun-Lin County. Due to high demand but limited water supply, groundwater has become a vital water resource in this area for decades. In 1992 the Water Resources Agency installed 26 groundwater monitoring wells (well depths range from 8 m to 110 m) distributed in this area for recording groundwater quality, particularly As pollution and other potential contamination in groundwater. Approximately 757 million m^3 of groundwater was extracted annually from the aquifers in this area, of which 268 million m^3 was considered to be over-pumped (Liu et al., 2001). High As concentration ($93.2 \pm 161\ \mu g\ l^{-1}$) was detected in monitoring wells in this area (WHO drinking water