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

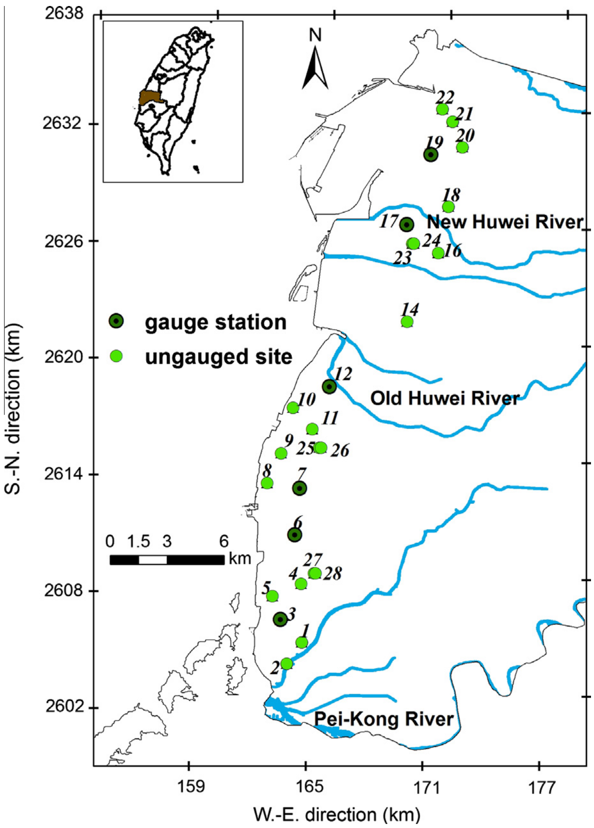

Fig. 1. Locations of 26 groundwater wells at Yun-Lin coastal area, Taiwan.

standard: 10 μg l^{-1}). Liu et al. (2006) indicated over-pumping groundwater induces dissolved oxygen and increases As mobility in water and the relatively high As content has accumulated and been deposited in the marine sequences with fine sediments.

2.2. Data collection and preliminary analysis

In this study, sampling data of groundwater quality variables were collected quarterly at 26 wells between 1992 and 1999, and the field sampling methods of As concentration was determined by hydride generation followed by atomic absorption spectroscopy, APHA Method 3500-arsenic Part B (APHA, 1992). The maintenance of groundwater monitoring wells is laborious and cost intensive, and therefore only six wells (#3, #6, #7, #12, #17 and #19) have continued monitoring groundwater quality after 1999. The proposed method intends to estimate the As fluctuations of 20 un-monitored wells based on other water quality variables that are easier to measure. We assume that 20 un-monitored wells are ungauged sites and 6 monitored wells are gauge stations (Fig. 1).

A total of 270 (=45 ∗ 6 wells) data sets of twelve water quality variables [power of hydrogen (pH), alkalinity (Alk), cadmium ion (Ca^{2+}), chlorine ion (Cl^{-}), total dissolved solid (TDS), electrical conductivity (EC), sodium ion (Na^{+}), sulfate ion (SO_4^{2-}), potassium ion (K^{+}), dissolved oxygen (DO), magnesium ion (Mg^{2+}) and temperature (Temp)] were collected at six gauge stations (wells) between 1992 and 2005, which are used for model construction in this study. Table 1 shows the well depth, mean and standard deviation (SD) of groundwater quality variables at these 6 gauge stations, in which high mean and variation of As concentration occur, especially at wells #6 and #7. The depths interval of the three aquifers were <60, 120–200, and 280–350 m, respectively (Agricultural Engineering Research Center, 2008). This indicates that the 6 monitored wells (well depth: 8.4–22.8 m) are in the same confined aquifer. Table 2 shows that all the correlation coefficients between As and twelve water quality variables are smaller than 0.34 (in an absolute sense), which implies the difficulty in determining non-trivial factors that affect As concentration based solely on such traditional correlation analysis. Therefore, we adopt a more sophisticated method to effectively extract non-trivial factors from water quality variables for building an As concentration estimation model.

3. Methods

The proposed SDM incorporates a dynamical-neural network with four advanced statistical techniques to tackle regional estimation problems, and its implementation procedure is shown in Fig. 2. The SDM first effectively extracts the non-trivial factors that significantly affect the fluctuations of As concentrations through the Gamma test (GT). The NARX network is then utilized to obtain As concentration at ungauged sites with inputs consisting of the extracted non-trivial factors and the estimated As concentrations from recurrent connections, and the Bayesian regularization method is configured to control the network complexity for preventing over-fitting. The cross validation technique is used to produce a low-bias estimator of the generalizability and thus provides a sensible criterion for model selection in the calibration stage. Finally, the indicator kriging is implemented to derive the probability map of As concentrations for detecting unsampled areas with As concentrations exceeding the WHO drinking water standard. The methods for use in this study are introduced as follows:

3.1. Nonlinear Autoregressive with eXogenous input (NARX) network

The NARX network is an important class of dynamic discrete-time nonlinear systems. Fig. 3 shows the architecture of the NARX network used in this study. The NARX network consists of three layers (input, hidden and output layers) and produce recurrent connections from the output which may delay several unit times to form new inputs. τ^{-1} is the unit time delay, and $d_Z \geqslant 1$ is the output-memory order. Therefore, this nonlinear system can be mathematically represented by the following equation:

$$z(t) = f[z(t-1), \ldots, z(t-d_z);\ U(t)] \tag{1}$$

where $U(t)$ and $z(t)$ denote the input vector and output value at the discrete time step t, respectively. And $f(\cdot)$ is the nonlinear mapping function.

There are two ways to train the NARX network. The first mode is the Series-parallel (SP) mode, where the output's regressor in the input layer is formed only by the target (actual) values of the system, $d(t)$:

$$z(t) = f[d(t-1), \ldots, d(t-d_z); U(t)] \tag{2}$$

The other alternative is the Parallel (P) mode, where estimated outputs are fed back into the output's regressor in the input layer, which can also be mathematically represented as Eq. (1). It is quite common for a regional estimation model to perform poorly at ungauged sites, which is mainly because the information of target variables is not always available (either lack of or missing) at ungauged sites. To mitigate this defect, the NARX network can be trained in the SP mode by using a few but available target values. Then the trained network is adopted to P mode for estimating the missing target values at the ungauged site.

Table 1
Well depth, mean and SD (standard deviation) of groundwater quality variables during 1992 and 2005 at six gauge stations (wells).

Item	Unit	#3	#6	#7	#12	#17	#19
Well depth	m	22.8	17.0	19.0	19.6	8.4	14.9
		Mean SD[a]	Mean SD	Mean SD	Mean SD	Mean SD	Mean SD
As	ug/L	75.9 ± 67.6	177.0 ± 109.5	450.4 ± 314.3	43.7 ± 30.7	39.5 ± 47.6	38.1 ± 30.7
Temp.	°C	25.8 ± 1.0	25.7 ± 0.9	25.8 ± 1.0	25.9 ± 1.3	26.1 ± 1.4	26.1 ± 1.3
pH		7.7 ± 0.4	7.9 ± 0.4	7.9 ± 0.2	7.7 ± 0.5	7.6 ± 0.4	7.6 ± 0.3
EC	μmho/cm 25 °C	23,383 ± 18,221	16,509 ± 8317	2209 ± 912.2	21,795 ± 13,886	1408 ± 970.8	17,295 ± 7677
DO	uS/cm	1040 ± 7347	419.0 ± 2923	51.0 ± 349.1	1.3 ± 1.0	1.3 ± 1.0	1.3 ± 1.0
Alk	ug/L	356.2 ± 137.3	560.2 ± 138.5	504.4 ± 84.8	384.1 ± 123.9	315.0 ± 55.7	504.2 ± 143.0
TDS	ug/L	15,822 ± 10,285	10,963 ± 4149	1432 ± 626.2	13,994 ± 7302	885.9 ± 637.7	11,790 ± 5426
Cl-	ug/L	6851 ± 4652	4479 ± 1819	391.9 ± 215.6	5945 ± 3419	233.6v236.3	4937 ± 2155
SO_4^{2-}	ug/L	690.9 ± 784.1	512.0 ± 270.8	102.9 ± 82.8	960.0 ± 588.1	64.5 ± 57.4	515.6 ± 467.0
Na^+	ug/L	3772 ± 2604	2708 ± 998.5	293.5 ± 90.8	3297 ± 1806	179.2v121.7	2756 ± 1133
K^+	ug/L	201.1 ± 105.6	145.2 ± 46.1	38.7 ± 15.5	133.4 ± 57.6	17.0v11.4	142.4 ± 91.9
Mg^{2+}	ug/L	598.8 ± 954.6	254.7 ± 205.5	73.2 ± 30.3	427.5 ± 296.4	32.2v22.6	323.8 ± 156.3
Ca^{2+}	ug/L	216.0 ± 133.1	74.9 ± 50.1	59.0 ± 19.7	281.8 ± 174.3	88.0 ± 46.7	150.4 ± 75.8

[a] Standard deviation.

Table 2
Correlation matrix of As concentration and water quality variables collected at six gauge stations (wells) during 1992 and 2005.

	As	pH	Alk	Ca^{2+}	Cl^-	TDS	EC	Na^+	SO_4^{2-}	K^+	DO	Mg^{2+}	Temp.
As	1.00	0.32	0.13	−0.34	−0.32	−0.31	−0.30	−0.30	−0.27	−0.26	0.19	−0.18	−0.07
pH		1.00	0.46	−0.56	−0.48	−0.43	−0.51	−0.44	−0.47	−0.37	0.26	−0.33	−0.03
Alk			1.00	−0.41	−0.27	−0.20	−0.28	−0.22	−0.33	−0.16	0.14	−0.17	0.04
Ca^{2+}				1.00	0.77	0.71	0.76	0.73	0.74	0.62	−0.22	0.52	0.02
Cl^-					1.00	0.93	0.97	0.97	0.88	0.87	−0.16	0.55	−0.05
TDS						1.00	0.93	0.93	0.81	0.84	−0.18	0.51	−0.07
EC							1.00	0.96	0.88	0.86	−0.19	0.53	−0.06
Na^+								1.00	0.84	0.88	−0.19	0.52	−0.09
SO_4^{2-}									1.00	0.72	−0.08	0.42	−0.06
K^+										1.00	−0.11	0.47	−0.07
DO											1.00	−0.12	−0.02
Mg^{2+}												1.00	0.02
Temp.													1.00

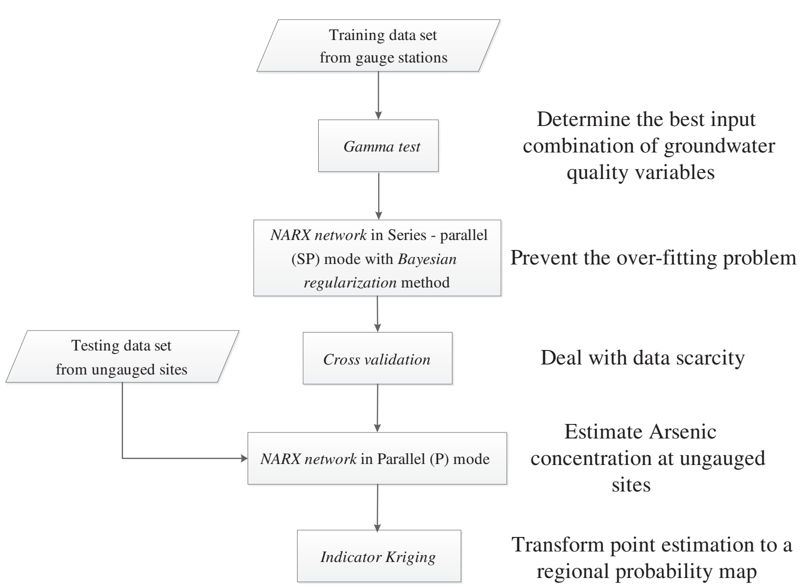

Fig. 2. Implementation procedure of the proposed SDM for regional analysis.

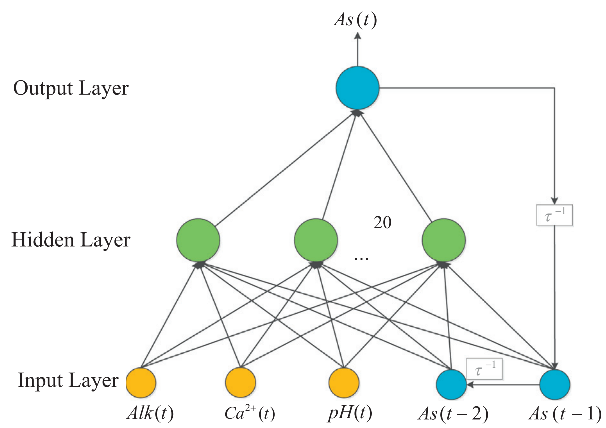

Fig. 3. Architecture of the NARX with recurrent connections from output-delay terms in this study.

3.2. Gamma test (GT)

The Gamma test (GT), presented by Koncar (1997) and Agalbjorn et al. (1997), is a data analysis technique for assessing the extent to which a given set of M data points can be modeled by an unknown smooth nonlinear function.

Suppose a set of input–output observation data is given in the form of:

$$[(x_i, y_i), 1 \leqslant i \leqslant M] \tag{3}$$

where vectors x_i are d dimensional vectors (with a record length of M) and the corresponding outputs y_i are scalars. The underlying relationship of the system is expressed as

$$y = f(x_1 \ldots x_d) + r \tag{4}$$

where f is an unknown smooth function, and r denotes a random variable that represents noise. The Gamma statistic (Γ) is an estimate of the model output's variance that cannot be accounted for through a smooth data model. The GT is assessed based on the kth ($1 \leqslant k \leqslant p$) nearest neighbor $X_{N(i,k)}$ for each vector X_i, and then the GT can be derived from the Delta function of input vectors:

$$\delta_M(k) = \frac{1}{M}\sum_{i=1}^{M}|X_{i,k} - X_i|^2 (1 \leqslant k \leqslant p) \tag{5}$$

where $|\cdots|$ is the Euclidean distance, and the corresponding Gamma function of the output values is given in Eq. (4). The number of p depends on the density of sampling (Koncar, 1997). In this study, the number of p is determined as the value that produces the minimum Γ value through trial and error (p ranges from 10 to 50), and consequently p is determined as 10.

$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}|y_{N(i,k)} - y_i|^2 (1 \leqslant k \leqslant p) \tag{6}$$

where $y_{N(i,k)}$ is the corresponding y-value for the kth nearest neighbor of X_i, in Eq. (3). For computing Γ, a least squares regression line is constructed for p points $(\delta_M(k), \gamma_M(k))$ as Eq. (5):

$$\gamma = A\delta + \Gamma \tag{7}$$

where A is the gradient.

Performing a single Gamma test is a fast procedure, which can provide the noise estimate (Γ value) for each subset of input variables. When the subset for which its associated Γ value is closest to zero, it can be considered as "the best combination" of input variables.

Several studies discussed about the GT theory and its applications in time series forecasting (Durrant, 2001; Tsui et al., 2002). Lately, research findings indicate it is suitable and effective to combine ANNs with the GT for identifying non-trivial input variables and thus reduces the input dimensions as well as produces precise outputs of ANNs (Moghaddamnia et al., 2009; Noori et al., 2010a,b, 2011). Therefore, the NARX network combines the GT to first extract non-trivial factors affecting As concentrations from twelve water quality variables in this study.

3.3. Bayesian regularization method

The regularization method proposed by MacKay (1992) can improve the generalizability of a neural network through minimizing an objective function that constrains the value of network weights. The idea is based on that the true underlying function is assumed to have a degree of smoothness controlled by the network parameters, and the network response will be smooth as the values of parameters are kept small. Thus the network is able to sufficiently represent the true function, rather than capture the noise. The objective function of the network in the regularization method is given by

$$M(W) = \beta E_D + \alpha E_W \tag{8}$$

where α, β are regularization parameters. E_D is the mean square error of network outputs. E_W is a penalty term for network complexity in the regularization method, in which smaller values of network weights imply lower connection complexity for network weights. E_D and E_W are defined as follows:

$$E_D = \frac{1}{2}\sum_{t=1}^{m}(d(t) - z(t))^2 \tag{9}$$

where m is the sample size.

$$E_W = \frac{1}{2}\sum_{i=1}^{n}w_i^2 \tag{10}$$

where w_i is weight value of the network, and n is the number of weights.

Then the Bayesian theory can be utilized to determine the regularization parameters (α and β) at the minimum W_{MP} of $M(W)$ and γ_p presenting the effective number of network parameters can also be calculated. The detailed formulation can be found in MacKay (1992).

3.4. Cross validation

Cross-validation, which involves partitioning data into training and testing sets, is commonly used to obtain a reliable estimation for model performance (Kohavi, 1995; Stone, 1974). The k-fold cross-validation is utilized in this study. The k results from the folds can be further averaged to produce a single estimation error. Therefore, the averaged estimation errors derived from different initial parameter settings are compared to choose the most appropriate model for use in the testing stage. In brief, cross validation can produce a low-bias estimator for the generalization abilities of a statistical model, and therefore provides a sensible criterion for model selection and performance comparison, especially for samples that are hazardous, costly or difficult to collect, such as the As concentration in this study.

3.5. Indicator kriging (IK) method

The IK is a non-parametric geostatistical technique that involves the transformation of one variable to a binary response (0,1) (Cressie, 1992; Journel, 1983). In this study, to reduce the influence of the extreme values of the estimated As concentration on the variogram and mitigate the uncertainty produced by the

NARX model, the IK is utilized to illustrate the variation of As concentration for the whole study area.

The geostatistical method is used to estimate the probability of exceeding a specific cut-off value (threshold) at a given location. In this study, the cut-off value is adequately set as 10 $\mu g\, l^{-1}$, the WHO drinking water standard for As concentration. Therefore, the IK can derive the probability map that discloses the probability of As concentration exceeding the WHO drinking water standard in the study area.

4. Results and discussion

4.1. Extracting effective water quality factors

The six wells (#3, #6, #7, #12, #17 and #19) that have sufficient water quality data are assumed as gauge stations for As concentration and are utilized by the GT. Data sets of twelve water quality factors are first scaled to $[-1, 1]$, and a total of 4095 $(2^{12}-1)$ Γ values corresponding to all possible input combinations are derived through the GT. The derived Γ values are next sorted in an ascending order, in which Γ values smaller than the 10th percentile (Γ_{10} = 0.0089) are classified as the best group ($F_{\Gamma \leqslant \Gamma_{10}}$) whereas Γ values bigger than the 90th percentile (Γ_{90} = 0.136) are classified as the worst group ($F_{\Gamma \geqslant \Gamma_{90}}$). Fig. 4 shows the result of the GT, where blue bars represent the occurrence frequency of variables in the best group ($F_{\Gamma \leqslant \Gamma_{10}}$) and red bars represent the occurrence frequency of variables in the worst group ($F_{\Gamma \geqslant \Gamma_{90}}$). Non-trivial factors that significantly affect fluctuations of As concentration can then be identified as the variables associated with higher blue bars and lower red bars simultaneously, and such ratios are shown by the dotted line in Fig. 4. And therefore we can extract a subset of input variables that ranks top three in the ratio of $F_{\Gamma \leqslant \Gamma_{10}}$ to $F_{\Gamma \geqslant \Gamma_{90}}$. The GT results indicate that Alk, Ca^{2+} and pH value are the non-trivial factors for use in the estimation models (the NARX network and BPNN).

These results are consistent with several studies, which indicated the increase in As leaching efficiency depends on high pH values and Alk concentration (Anawar et al., 2004; Kim et al., 2000; Kuo and Chang, 2010; Liu et al., 2003; Park et al., 2006; Pierce and Moore, 1982). The major C-containing species in the reducing condition n groundwater are HCO_3^- and H_2CO_3, which cause high pH values and Alk concentration (Wang et al., 2007). In addition, salinization and As enrichment are two main hydrogeochemical characteristics in the Yun-Lin coastal area, and they were estimated by the factor analysis (FA) (Wang et al., 2007). Respectable cation, such as calcium ions, and anion contents carried by seawater intrusion initially increased the ion strength in groundwater and induced As desorption (Appelo et al., 2002; Keon et al., 2001). On the one hand, As anions could sorb or bind using carbonates in natural systems (Bauer et al., 2008; Rothwell et al., 2009). Therefore the relationship between As and calcium ions might be caused by the dissolution of calcium arsenates and/or the competitive desorption of calcium (Bothe and Brown, 1999; Mihaljevic et al., 2003; Nishimura and Robins, 1998). In the Yun-Lin coastal area, the shallow aquifer has suffered serious salinization that affects the concentration of the calcium ion due to the over-pumping of groundwater. This exercise gives evidence that the GT can effectively identify non-trivial and meaningful factors that affect the fluctuations of As concentration, compared with the identification difficulty raised by the traditional correlation matrix shown in Table 2.

4.2. Estimating As concentration at ungauged sites by the NARX network

In this study, the NARX network is proposed to estimate the regional As concentration in Yun-Lin County. Variables Alk, Ca^{2+} and pH determined by the GT are used as exogenous inputs to the NARX network. The data sets collected from six gauge stations between 1992 and 2005 are used for model calibration. Therefore, the NARX network in the SP mode trained by the Bayesian regularization method is calibrated by a 30-fold cross validation. The log-sigmoid function and the linear function are the transfer functions used in the hidden and output layers of the NARX network, respectively. The most appropriate NARX network comprises two output-memory orders and 20 neurons in the hidden layer, and the effective number of network parameters (γ_p) is 23.74.

To demonstrate the effectiveness and usefulness of the NARX network established, the backpropagation neural network (BPNN) that represents a classical type of static ANNs is implemented for comparison purpose. The constructed BPNN consists of the same input variables as those of the NARX network and six neurons in the hidden layer. The hyperbolic tangent sigmoid function and the linear function are the transfer functions used in the hidden and output layers of the BPNN, respectively. The BPNN trained with the Levenberg–Marquardt optimization algorithm is also calibrated by a 30-fold cross validation. The results show the average RMSE of the NARX network in the training and validation phases are 95.11 and 106.13 $\mu g\, l^{-1}$, respectively, whereas the average RMSE of the BPNN in the training and validation phases considerably increases to 121.54 and 143.37 $\mu g\, l^{-1}$, respectively. The results demonstrate that the NARX network has much better performance than the BPNN. It is found large errors (average RMSE) produced by both models, this is mainly due to the high uncertainty attached to the sampled values, where the mean (138.26 $\mu g\, l^{-1}$) and standard deviation (205.25 $\mu g\, l^{-1}$) of As concentration contributes to the poor model accuracy.

It is worth noting that the effective number of network parameters (γ_p) has been optimized from 141 to 23.74 after the re-calibration of the NARX network by using the Bayesian

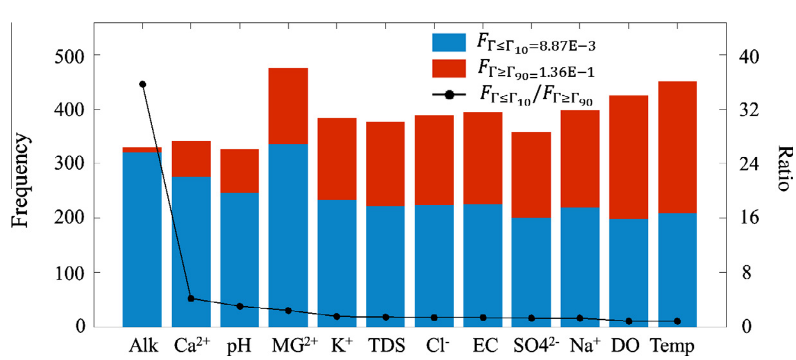

Fig. 4. Determination of non-trivial factors by the GT results.

regularization method. This demonstrates that the Bayesian regularization method can significantly reduce the effective number of network parameters and avoid the over-fitting problem caused in a rather complex network structure. As a result, the NARX network produces suitable results and has similar performance in the training and validation phases (average RMSE: 95.11 μg l^{-1} and 106.13 μg l^{-1} accordingly). In contrast, the BPNN requires fewer neurons in the hidden layer to prevent the over-fitting problem but still performs worse in the validation phase (average RMSE: 143.37 μg l^{-1}).

After model configuration, a total of 100 (=20 $*$ 5 months) As concentration data collected at the assumed 20 ungauged sites in the five monitoring months (January 1995, October 1996, October 1997, September 1998 and January 1999) are utilized to test the two constructed models. In addition to RMSE, the normalized mean squared error (NMSE), *R*-square value (R^2) and *F* test are also used as performance criteria in the testing phase. The NMSE is defined as:

$$NMSE = \frac{\sum_{i=1}^{n}(O_i - \widehat{Z}_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (11)$$

where O_i and $\hat{Z}_i$ are the observed and estimated As concentration from the *i*th assumed ungauged sites in the same year, respectively, $\bar{O}$ represents the average of observed As concentrations in a certain year, and n is the length of data.

The results of model comparison in the testing phase are summarized in Table 3, which indicates the NARX network has much smaller RMSE as well as NMSE values and higher R^2 values than the BPNN. Besides, when assessing the results of the *F* test, the null hypothesis is rejected only at the 5% level of the estimation values in October 1996 for the NARX network, whereas the null hypothesis is rejected in January 1995, October 1996 and January 1999 for the BPNN. Fig. 5 shows the scatter plots of observed and estimated As concentrations in five different months during 1995 and 1999 derived from the NARX network and BPNN. The estimation values obtained from the NARX network are close to the ideal line and only have few underestimations at extremely high As concentrations, whereas the BPNN overestimates As concentrations at values lower than 200 μg l^{-1} and seriously underestimates As concentrations at values higher than 200 μg l^{-1}.

In sum, the NARX network adequately utilizes the information of model outputs through recurrent connections to the network itself for producing reliable estimations of As concentrations at 20 ungauged sites. Owing to the implementation of the Bayesian regularization method into the NARX network, the network shows impressive generalizability and performs well in the testing phase,

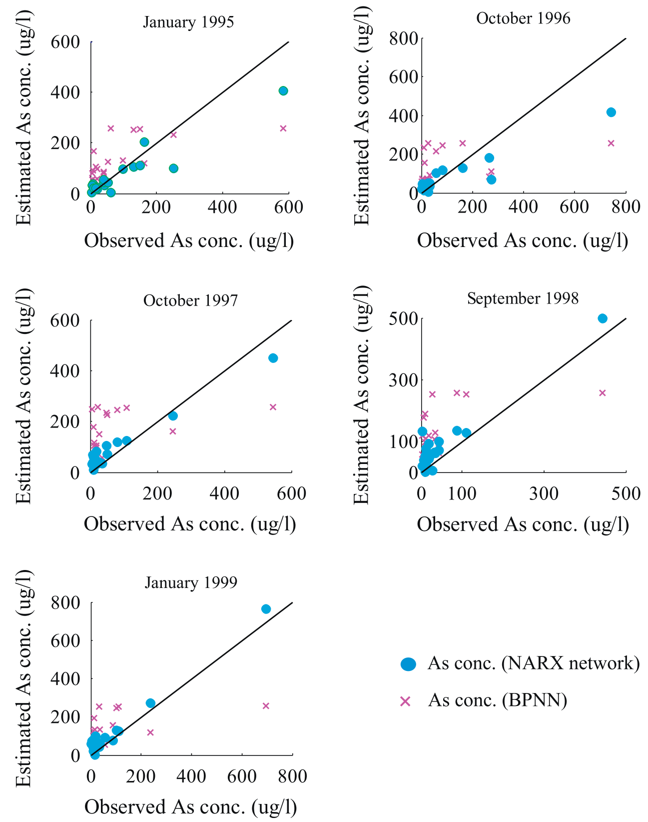

Fig. 5. Scatter plots of observed and estimated As concentration (conc.) derived from the NARX network and BPNN at 20 ungauged sites from 1995 to 1999.

which can be proved through similar NMSE values in five testing years (Table 3).

4.3. Deriving the risk map of As concentration through the IK

From the previous section, the NARX network can provide reliable point estimation of As concentration at 20 ungauged sites. The IK is employed to estimate the regional spatial distribution and to compute the probability of the exposure to high As pollution in unsampled areas. Because the WHO drinking water standard for As concentration is 10 μg l^{-1}, the investigation of this study mainly focuses on the threshold of 10 μg l^{-1}. Therefore, the estimated As concentration from the NARX network at 20 ungauged sites and

Table 3
Estimation performance of the NARX network and BPNN for As concentration at 20 ungauged sites from 1995 to 1999 in the testing phase.

Estimation time	RMSE (μg l^{-1})	NMSE	R^2	*F* test *p*-value	Data mean (μg l^{-1})	Data SD[a] (μg l^{-1})
NARX network						
1995 January	57.31	0.19	0.89	0.105	85.63	134.80
1996 October	91.53	0.29	0.89	<u>0.010</u>[b]	90.71	173.61
1997 October	40.24	0.11	0.96	0.291	64.51	125.99
1998 September	48.34	0.26	0.91	0.731	47.83	97.02
1999 January	41.35	0.07	0.98	0.816	75.57	155.91
BPNN						
1995 January	109.02	0.69	0.41	<u>0.025</u>	85.63	134.80
1996 October	158.70	0.88	0.17	<u>0.002</u>	90.71	173.61
1997 October	142.42	1.35	0.18	0.070	64.51	125.99
1998 September	114.63	1.47	0.29	0.302	47.83	97.02
1999 January	140.21	0.85	0.25	<u>0.004</u>	75.57	155.91

[a] Standard deviation.
[b] The null hypothesis is rejected at the 5% level (p-value < 0.05).

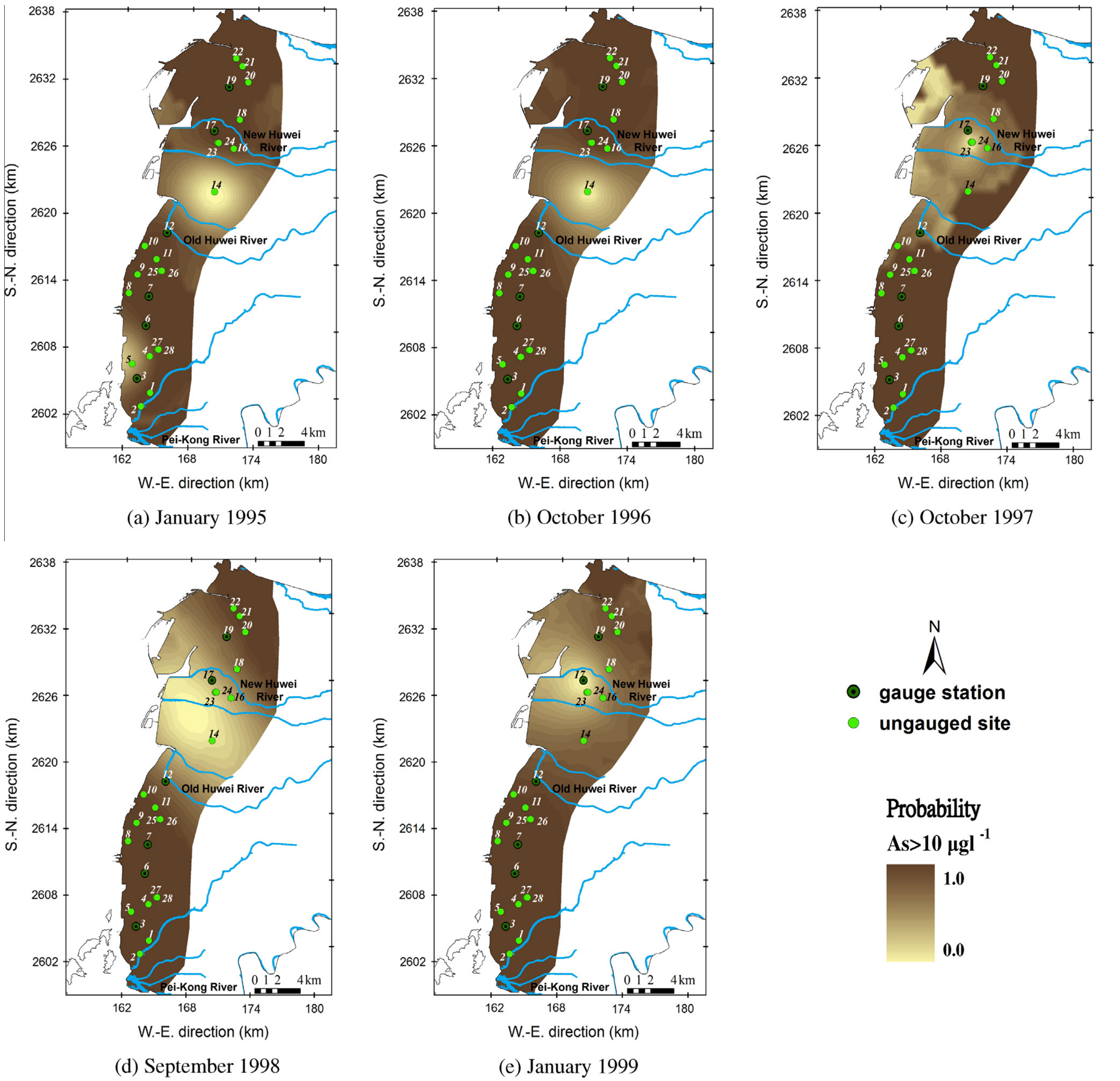

Fig. 6. Exceeding probability maps of As concentration under the threshold of WHO drinking water standard (10 $\mu g\, l^{-1}$) from (a) 1995 to (e) 1999.

the observed As concentration at six gauge stations are utilized to construct the semivariogram models for the IK.

The NARX network coupled with the IK can illustrate the unknown probability of the exposure to high As concentrations at neighboring areas of all 26 wells. If the observed and estimated As concentrations exceed the threshold set in the adjacent region, the IK will assign a high probability of concentration in the region of interest. The probability maps of As concentration under the threshold of WHO drinking water standard (10 $\mu g\, l^{-1}$) in different time spans are shown in Fig. 6. In January 1995, high exceeding probabilities (>10 $\mu g\, l^{-1}$) of As concentration occurred in northern and southern areas, whereas both the surrounding area of well #5 and the central area (located between Old Huwei River and New Huwei River) had low exceeding probabilities of As concentration. In October 1996, the exceeding probability was high in the southern area of the Old Huwei River. In contrast, the exceeding probability of As concentration was gradually and significantly mitigated in the central and northern areas from October 1997 to January 1999, and the Old Huwei River could be deemed as a clear boundary between high and low As concentrations in an exceeding probability sense. These risk maps reveal the high arsenic-prone areas. As a result, the information of the risk maps derived from the IK of the proposed SDM can consequently help decision makers manage groundwater quality and thus prevent residents from drinking or using toxic groundwater.

5. Conclusion

The blackfoot disease in the Yun-Lin Countyof Taiwan has been verified to be associated with high As concentrations in groundwater. Residents had used high-Artesian well water for years and had long-term exposed themselves to As and thus dangerously posed carcinogenic risks to their health. Due to great concern for the potential effects of As on human health, there is a growing need for efficiently modeling the presence and amount of As in groundwater. In this study, we propose a systematical dynamic-neural modeling (SDM) that incorporates a dynamic-neural network with four advanced statistical techniques to adequately estimate As concentration in the area of Yun-Lin County in southern Taiwan. The modeling processes and related results suggest that (1) the GT can effectively extract non-trivial factors that affect the target variable; (2) the Bayesian regularization method that constrains the network's weight values does improve the generalizability of the network; (3) the cross validation can produce a low-bias estimator of the generalization ability of networks; (4) the NARX network can provide reliable estimation of As concentration at both gauged and ungauged sites; and (5) the IK suitably derives the probability maps of As concentration under the threshold of WHO drinking water standard in the study area.

The results demonstrate that the NARX network has much better performance than the BPNN. The average RMSE of the NARX network in the training and validation phases are 95.11 $\mu g\, l^{-1}$ and 106.13 $\mu g\, l^{-1}$, respectively, whereas the average RMSE of the BPNN in the training and validation phases considerably increases to 121.54 $\mu g\, l^{-1}$ and 143.37 $\mu g\, l^{-1}$, respectively. The configured NARX network can suitably and accurately estimate As concentrations at 20 ungauges sites in five testing years (all R^2 are high (0.82–0.95)), whereas the BPNN fails to provide suitable estimations (all R^2 are low (0.17–0.41)). It proves the recurrent connections of model output information (As concentrations) to the NARX network itself makes significant contribution to the accuracy of the regional estimation model.

Finally, the IK can suitably derive the probability maps of As concentration under the threshold of the WHO drinking water standard in the study area, which is meaningful and useful for the authorities to manage water resources so that prevent residents from using and drinking As-contaminated groundwater. In particular, the construction of the proposed SDM requires As concentration data at six gauge stations and data of three easily-measured water quality variables (Alk, Ca^{2+} and pH) at six gauge stations and the other 20 ungauged sites. It merely requires data of three three-easily measured water quality variables for the constructed SDM to effectively and suitably estimate As concentrations at ungauged sites. This approach will significantly reduce the manpower cost of monitoring wells and effectively provide reliable estimation of As concentration at ungauged sites.In summary, the proposed SDM modeling approaches to the estimation of As concentration using on-site measurement data of other water quality variables can be an alternative way to quantify the As contamination and to provide predictive information for better public health management.

Acknowledgments

This study was funded by the National Science Council, Taiwan, ROC (Grant Number: 100-2313-B-002-011-MY3). The gauge data provided by the Water Resources Agency, Taiwan, are very much appreciated. The authors are indebted to Editors and Reviewers for their valuable comments and constructive suggestions.

References

BGS-DPHE, 2001. Arsenic Contamination of Groundwater in Bangladesh; BGS Technical Report WC/00/19. British Geological Survey, Keyworth, UK.
Agalbjorn, S., Koncar, N., Jones, A.J., 1997. A note on the gamma test. Neural Comput. 5, 131–133.
Agricultural Engineering Research Center, 2008. Survey, Analysis and Assessment of Groundwater Quality in Taiwan Area. Water Resource Agency, Taiwan.
Appelo, C.A.J., Van Der Weiden, M.J.J., Tournassat, C., Charlet, L., 2002. Surface Complexation of Ferrous Iron and Carbonate on Ferrihydrite and the Mobilization of Arsenic. Environ. Sci. Technol. 36 (14), 3096–3103.
Ali, A., 2009. Nonlinear multivariate rainfall-stage model for large wetland systems. J. Hydrol. 374 (3–4), 338–350.
Anawar, H.M., Akai, J., Sakugawa, H., 2004. Mobilization of arsenic from subsurface sediments by effect of bicarbonate ions in groundwater. Chemosphere 54 (6), 753–762.
Antar, M.A., Elassiouti, I., Allam, M.N., 2006. Rainfall–runoff modelling using artificial neural networks technique: a Blue Nile catchment case study. Hydrol. Process. 20, 1201–1216.
APHA (American Public Health Association), 1992. Standard Methods for the Examination of Water and Waste Water, 18th ed. APHA, American Water Works Association, and Water Pollution Control Federation, Washington, DC.
Ardalani-Farsa, M., Zolfaghari, S., 2010. Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks. Neurocomputing 73 (13–15), 2540–2553.
Assaad, M., Boné, R., Cardot, H., 2005. Study of the behavior of a new boosting algorithm for recurrent neural networks artificial neural networks: formal models and their applications-ICANN. Lect. Notes Comput. Sci., 169–174.
Ayotte, J.D., Montgomery, D.L., Flanagan, S.M., Robinson, K.W., 2003. Arsenic in groundwater in Eastern New England: occurrence, controls, and human health implications. Environ. Sci. Technol. 37 (10), 2075–2083.
Bargaouia, Z.K., Chebbib, A., 2009. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. J. Hydrol. 365 (1–2), 56–73.
Bauer, M., Fulda, B., Blodau, C., 2008. Groundwater derived arsenic in high carbonate wetland soils: Sources, sinks, and mobility. Sci. Total Environ. 401, 109–120.Keon N.E., Swartz C.H., Brabander D.J., Harvey C., Hemond H.F., 2001. Validation of an arsenic sequential extraction method for evaluating mobility in sediments. Environ. Sci. Technol. 35, 2778–2784.
Bothe, J.V., Brown, P.W., 1999. The stabilities of calcium arsenates at 23 ± 1 °C. J. Hazard. Mater. 69, 197–207.
Central Geological Survey, 1999. Project of Groundwater Monitoring Network in Taiwan during First Stage-research Report of Chou-Shui River Alluvial Fan, Taiwan. Water Resource Agency, Taiwan.
Chakraborti, D., Rahman, M.M., Das, B., Murrill, M., Dey, S., Mukherjee, S.C., Dhar, R.K., Biswas, B.K., Chowdhury, U.K., Roy, S., Sorif, S., Selim, M., Rahman, M., Quamruzzaman, Q., 2010. Status of groundwater arsenic contamination in Bangladesh: a 14-year study report. Water Res. 44, 5789–5802.
Chang, F.J., Chang, L.C., Huang, H.L., 2002. Real time recurrent learning neural network for streamflow forecasting. Hydrol. Process. 16 (13), 2577–2588.
Chang, F.J., Kao, L.S., Kuo, Y.M., Liu, C.W., 2010. Artificial neural networks for estimating regional arsenic concentrations in a blackfoot disease area in Taiwan. J. Hydrol. 388, 65–76.
Chen, C.J., Hsueh, Y.M., Lai, M.S., Shyu, M.P., Chen, S.Y., Wu, M.M., Kuo, T.L., Tai, T.Y., 1995. Increased prevalence of hypertension and long-term arsenic exposure. Hypertension 25, 53–600.
Chiang, Y.M., Chang, L.C., Chang, F.J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling. J. Hydrol. 290, 297–311.
Chiang, Y.M., Hsu, K.L., Chang, F.J., Yang, H., Soroosh, S., 2007. Merging multiple precipitation sources for flash flood forecasting. J. Hydrol. 340, 183–196.
Chiang, Y.M., Chang, L.C., Tsai, M.J., Wang, Y.F., Chang, F.J., 2010. Dynamic neural networks for real-time water level predictions of sewerage systems-covering gauged and unguaged sites. Hydrol. Earth Syst. Sci. 14, 1309–1319.
Chiou, H.Y., Hsueh, Y.M., Hsieh, L.L., Hsu, L.I., Hsu, Y.H., Hsieh, F.I., Wei, M.L., Chen, H.C., Yang, H.T., Leu, L.C., 1997. Arsenic methylation capacity, body retention, and null genotypes of glutathione S-transferase M1 and T1 among current arsenic-exposed residents in Taiwan. Mutat. Res. 386, 197–207.
Cho, K.H., Kao, S.S., Pachepsky, Y.A., Kim, K.W., Kim, J.H., 2011. Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. Water Res. 45 (17), 5535–5544.
Cressie, N., 1992. Statistic for spatial data. Terra Nova 4 (5), 613–617.
Durrant, P.J., 2001. winGamma: a Non-linear Data Analysis and Modeling Tool with Applications to Flood Prediction. PhD Thesis. Department of Computer Science, Cardiff University, Wales, UK.
Ford, R.G., Wilkin, R.T., Fendorf, S., 2006. Introduction: controls on arsenic transport in near-surface aquatic systems. Chem. Geol. 228 (1–3), 1–5.
Giri, A.K., Patel, R.K., Mahapatra, S.S., 2011. Artificial neural network (ANN) approach for modelling of arsenic(III) biosorption from aqueous solution by living cells of *Bacillus cereus* biomass. Chem. Eng. J. 178, 15–25.
Goovaerts, P., AvRuskin, G., Meliker, J., Slotnick, M., Jacquez, G., Nriagu, J., 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. Water Resour., 41.
Hong, Y.S.T., 2012. Dynamic nonlinear state-space model with a neural network via improved sequential learning algorithm for an online real-time hydrological modeling. J. Hydrol. 468–469, 11–21.
Journel, A.G., 1983. Nonparametric estimation of spatial distributions. J. Int. Assoc. Math. Geol. 15 (3), 445–468.
Juang, K.W., Lee, D.Y., 1998. Simple indicator kriging for estimating the probability of incorrectly delineating hazardous areas in a contaminated site. Environ. Sci. Technol. 32, 2487–2493.
Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. J. Hydrol. 405 (3–4), 277–287.

Kim, M.J., Nriagu, J., Haack, S., 2000. Carbonate ions and arsenic dissolution by groundwater. Environ. Sci. Technol. 34 (15), 3094–3100.
Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proc. 14th Int. Jt. Conf. Artif. Intell., vol. 2, Quebec, Canada.
Koncar, N., Optimisation Methodologies for Direct Inverse Neurocontrol. PhD Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, University of London, 1997.
Krishna, B., Rao, Y.R.S., Vijaya, T., 2008. Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. Hydrol. Process. 22, 1180–1188.
Kuo, Y.M., Chang, F.J., 2010. Dynamic factor analysis for estimating groundwater arsenic trends. J. Environ. Qual. 39, 176–184.
Liao, V.H.C., Chu, Y.J., Su, Y.C., Hsiao, S.Y., Wei, C.C., Liu, C.W., Liao, C.M., Shen, W.C., Chang, F.J., 2011. Arsenite-oxidizing and arsenate-reducing bacteria associated with arsenic-rich groundwater in Taiwan. J. Contam. Hydrol. 123, 20–29.
Lin, T., Horne, B.G., Tino, P., Giles, C.L., 1996. Learning long-term dependencies in NARX recurrent neural networks. IEEE Trans. Neural Netw. 7 (6), 1424–1438.
Liu, C.W., Lin, W.S., Shang, C., Liu, S.H., 2001. The effect of claydehydration on land subsidence in Yun-Lin coastal area, Taiwan. Environ. Geol. 40 (4–5), 518–527.
Liu, C.W., Lin, K.H., Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Total Environ. 313 (1–3), 77–89.
Liu, C.W., Jang, C.S., Liao, C.M., 2004. Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). Sci. Total Environ. 321 (1–3), 173–188.
Liu, C.W., Wang, S.W., Jang, C.S., Lin, K.H., 2006. Occurrence of arsenic in ground water in the Choushui River Alluvial Fan, Taiwan. J. Environ. Qual. 35, 68–75.
Liu, C.W., Huang, Y.K., Hsueh, Y.M., Lin, K.H., Jang, C.S., Huang, L.P., 2008. Spatiotemporal distribution of arsineic species of oysters (Crassostreagigas) in the coastal area of southwestern Taiwan. Environ. Monit. Assess. 138, 181–190.
Ma, Q.L., Zheng, Q.L., Peng, H., Zhong, T.W., Qin, J.W., 2008. Multi-step-prediction of chaotic time series based on co-evolutionary recurrent neural network. Chin. Phys. B 17 (2), 536.
MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4, 415–447.
Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58 (8), 1246–1266.
McNamara, J.P., Kane, D.L., Hobbie, J.E., Kling, G.W., 2008. Hydrologic and biogeochemical controls on the spatial and temporal patterns of nitrogen and phosphorus in the Kuparuk River arctic Alaska. Hydrol. Process. 22, 3294–3309.
Menezes Jr, J.M.P., Barreto, G.A., 2008. Long-term time series prediction with the NARX network: an empirical evaluation. Neurocomputing 71 (16–18), 3335–3343.
Mihaljevic, M., Ponavic, M., Ettler, V., Šebek, O., 2003. A comparison of sequential extraction techniques for determining arsenic fractionation in synthetic mineral mixtures. Anal. Bioanal. Chem. 377, 723–729.
Moghaddamnia, A., GhafariGousheh, M., Piri, J., Amin, S., Han, D., 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Adv. Water Resour. 32, 88–97.
Mondal, P., Mohanty, B., Balomajumder, C., Saraswati, S., 2012. Modeling of the removal of arsenic species from simulated groundwater containing As, Fe, and Mn: a neural network based approach. Clean: Soil, Air, Water 40 (3), 285–289.
National Research Council, 1999. Arsenic in Drinking Water. National Academy Press, Washington, DC.
Nikolos, I.K., Stergiadi, M., Papadopoulou, M.P., Karatzas, G.P., 2008. Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design. Hydrol. Process. 22, 3337–3348.
Nishimura, T., Robins, R.G., 1998. Are-evaluation of the solubility and stability regions of calcium arsenites and calcium arsenates in aqueous solution at 25 °C. Mineral Proc. Extr. Met. 18, 283–308.
Noori, R., Hoshyaripour, G., Ashrafi, K., Nadjar-Araabi, B., 2010a. Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. Atmos. Environ. 44, 476–482.
Noori, R., Sabahi, M.S., Karbassi, A.R., 2010b. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste predicting. J. Environ. Manage. 91, 767–771.
Noori, R., Karbassi, A.R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M.H., Farokhnia, A., Gousheh, M.G., 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. J. Hydrol. 401, 177–189.
Park, J.M., Lee, J.S., Lee, J.U., Chon, H.T., Jung, M.C., 2006. Microbial effects on geochemical behavior of arsenic in As-contaminated sediments. J. Geochem. Explor. 88 (1–3), 134–138.
Pierce, M.L., Moore, C.B., 1982. Adsorption of arsenite and arsenate on amorphous iron hydroxide. Water Res. 16 (7), 1247–1253.
Pijanowski, B.C., Brown, D.G., Shellito, B.A., Manik, G.A., 2002. Using neural networks and GIS to forecast land use changes: a land transformation model. Comput., Environ. Urban Syst. 26 (6), 553–575.
Polizzotto, M.L., Harvey, C.F., Li, G., Badruzzman, B., Ali, A., Newville, M., Sutton, S., Fendorf, S., 2006. Solid-phases and desorption processes of arsenic within Bangladesh sediments. Chem. Geol. 228 (1–3), 97–111.
Purkait, B., Kadam, S.S., Das, S.K., 2008. Application of artificial neural network model to study arsenic contamination in groundwater of Malda District, Eastern India. J. Environ. Inf. 12 (2), 140–149.
Rahman, M., Tondel, M., Chowdhury, I.A., Axelson, O., 1999. Relations between exposure to arsenic, skin lesions, and glucosuria. Occup. Environ. Med. 56 (4), 277–281.
Rothwell, J.J., Taylor, K.G., Ander, E.L., Evans, M.G., Daniels, S.M., Allott, T.E., 2009. Arsenic retention and release in ombrotrophic peatlands. Sci. Total Environ. 407, 1405–1417.
Sahoo, G.B., Ray, C., De Carlo, E.H., 2006. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. J. Hydrol. 327 (3–4), 525–538.
Serre, M.L., Kolovos, A., Christakos, G., Modis, K., 2003. An application of the holistochastic human exposure methodology to naturally occurring arsenic in Bangladesh drinking water. Risk Anal. 23 (3), 515–528.
Shin, H.S., Salas, J.D., 2000. Spatial analysis of hydrologic and environmental data based on artificial neural networks. In: Govindaraju, R.S., Rao, A.R. (Eds.), Artificial Neural Networks in Hydrology. Kluwer Academic Publishers, pp. 259–286.
Stollenwerk, K.G., 2003. Geochemical processes controlling transport ofarsenic in groundwater: a review of adsorption. In: Stollenwerk, K.G., Welch, A.H. (Eds.), Arsenic in Groundwater. Springer US, New York, pp. 67–100.
Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B 36, 111–147.
Tsui, A.P.M., Jones, A.J., Guedes de Oliveira, A.G., 2002. The construction of smooth models using irregular embeddings determined by a gamma test analysis. Neural Comput. Appl. 10, 318–329.
Wang, S.W., Liu, C.W., Jang, C.S., 2007. Factors responsible for high arsenic concentrations in two groundwater catchments in Taiwan. Appl. Geochem. 22 (2), 460–476.
Wang, S.W., Kuo, Y.M., Kao, Y.H., Jang, C.S., Maji, S.K., Chang, F.J., Liu, C.W., 2011. Influence of hydrological and hydrogeochemical parameters on arsenic variation in shallow groundwater of southwestern Taiwan. J. Hydrol. 408 (3–4), 286–295.
Webster, R., McBratney, A.B., 1987. Mapping soil fertility at Broom's Barn by simple kriging. J. Sci. Food Agric. 38 (2), 97–115.
Winkel, L., Berg, M., Amini, M., Hug, S.J., Johnson, C.A., 2008. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. Nat. Geosci. 1 (8), 536–542.
Xie, X., Wang, Y., Ellis, A., Su, C., Li, J., Lia, M., Duan, M., 2013. Delineation of groundwater flow paths using hydrochemical and strontium isotope composition: a case study in high arsenic aquifer systems of the Datong basin, northern China. J. Hydrol. 476, 87–96.
Yu, W.H., Harvey, C.M., Harvey, C.F., 2003. Arsenic in groundwater in Bangladesh: a geostatistical and epidemiological framework for evaluating health effects and potential remedies. Water Resour. Res. 39 (6), 1146.