Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2015.1049411

57 (2016) 12227–12236
June

Development of enhanced groundwater arsenic prediction model using machine learning approaches in Southeast Asian countries

Yongeun Park[a], Mayzonee Ligaray[b], Young Mo Kim[c], Joon Ha Kim[c], Kyung Hwa Cho[b,*], Suthipong Sthiannopkao[d,*]

[a] *Environmental Microbial and Food Safety Laboratory, USDA-ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA*
[b] *School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 689-798, Korea, email: khcho@unist.ac.kr (K.H. Cho)*
[c] *School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), 261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea*
[d] *Department of Environmental Engineering, Dong-A University, Busan 604-714, Korea, email: suthisuthi@gmail.com*

Received 5 February 2015; Accepted 4 May 2015

ABSTRACT

Groundwater contamination with arsenic (As) is one of the major issues in the world, especially for Southeast Asian (SEA) countries where groundwater is the major drinking water source, especially in rural areas. Unfortunately, quantification of groundwater As contamination is another burden for those countries because it requires sophisticated equipment, expensive analysis, and well-trained technicians. Here, we collected approximately 350 groundwater samples from three different SEA countries, including Cambodia, Lao PDR, and Thailand, in an attempt to quantify total As concentrations and conventional water quality variables. After that, two machine learning models (i.e. artificial neural network (ANN) and support vector machine (SVM)) were applied to predict groundwater As contamination using conventional water quality parameters. Prior to modeling approaches, the pattern search algorithm in MATLAB software was used to optimize the ANN and SVM model parameters, attempting to find the best parameters set for modeling groundwater As concentrations. Overall, the SVM showed the superior prediction performance, giving higher Nash–Sutcliffe coefficients than ANN in both the training and validation periods. We hope that the model developed by this study could be a suitable quantification tool for groundwater As contamination in SEA countries.

Keywords: Groundwater; Arsenic contamination; Machine learning; Support vector machine; Artificial neural network; Southeast Asian countries

1. Introduction

In Southeast Asian (SEA) countries, groundwater is a major source of drinking water supplies. However, groundwater contamination with arsenic (As) has become a critical issue in SEA. In particular, it is regarded as a public health issue because it is a carcinogenic element, which typically presents as an inorganic species [1–5]. Groundwater contaminated by As through natural aquatic chemical reactions has been monitored in tube wells/hand pump drinking water

*Corresponding authors.

supplies in South and Southeast Asia including Thailand, Vietnam, Lao PDR, Cambodia, Myanmar, Bangladesh, India, Nepal, and Pakistan [6–15]. Sun et al. [9] reported that approximately 200 million people are exposed to the potentially toxic effects of As in these countries.

Long-term and intensive monitoring programs are insufficient to characterize As contamination in SEA. It is very challenging to provide scientific guidelines for public health due to a lack of groundwater As observations. Quantification of As contamination, however, could be another aspect that needs the substantial labor and cost burden in SEA countries. The measurement of groundwater As needs expensive equipment and highly experienced technicians with a high maintenance cost. Thus, indirect quantification of As by modeling approaches could be an alternative way to detect As contamination and provide predictive information for public health management. In particular, machine learning approaches can be useful for predicting As fate that has complex relationship between structural geology, mineral chemistry, and mobilization characteristics [16].

Among various modeling approaches, machine learning models could be useful for predicting As concentration based on the analysis of non-linear relationship between environmental variables and As. The artificial neural network (ANN) is a well-known information-processing modeling approach [17]. The ANN is one of the powerful pattern recognition approaches which have been widely applied in various areas [18–27]. One of the drawbacks is that the ANN is not only difficult to construct the model, but also often causes over-fit problems in predictions. A few researchers have applied ANN to the prediction of As in groundwater [28–30]. The support vector machine (SVM) is another machine learning algorithm which has been widely used in various fields [31–33]. It is based on statistical learning theory using a linear high dimensional hypothesis space. It is regarded as the effective alternative way to overcome the weakness of ANN modeling, having all the positive characteristics of ANN modeling [34]. Very few studies were found in the application of SVM for groundwater As contamination, including the previous study by Purkait et al. [28]. For the practical application, efforts for developing more enhanced and more reliable prediction models than conventional models are needed.

Therefore, the objective of this study was to suggest enhanced statistical modeling approach for predicting As concentrations through the comparison of prediction performance between the ANN and SVM, using sufficient and comprehensive dataset that has been measured from three different SEA countries during the years 2008–2012 with a wide range of As concentration levels.

2. Materials and methods

2.1. Field sampling

We collected groundwater samples to investigate the As concentrations and six different parameters from three countries, Cambodia, Laos, and Thailand, from 2008 to 2012. Table 1 shows mean and standard deviation values of As concentrations and conventional water quality parameters. In Cambodia, 153 groundwater samples were collected from seven villages in Kandal, Prey Veng, Kamphong Cham, and Kratie provinces. In Laos, 182 samples were collected and analyzed. In Thailand, the concentrations of As in groundwater were examined in Tambon Ongphra ($n = 10$) in Suphanburi province in 2008. In the three countries, samples were collected from tube wells by following this sequence: (1) pumping out the standing water in the tube wells for about 10 min, (2) rinsing clean polyethylene bottles which were previously washed with water drawn from the tube well, and (3) taking tube well water without filtering (raw water). All samples were preserved with concentrated HNO_3, kept at 4°C, and delivered to the laboratory. During

Table 1
Mean and standard deviation of groundwater observations

Country	Total As (ppb)	Conductivity (μS/cm)	Temperature (°C)	Redox (mV)	pH	Well depth (m)	TDS (mg/L)
Cambodia	428.83	597.62	29.60	−70.71	7.38	34.82	203.27
	(278.80)	(274.01)	(1.02)	(63.12)	(0.31)	(6.40)	(399.96)
Laos	6.55	389.99	29.25	115.17	7.68	283.58	77.63
	(14.21)	(407.16)	(1.45)	(66.83)	(18.00)	(452.15)	(77.63)
Thailand	1.77	261.84	29.18	123.45	5.43	30.35	138.97
	(1.38)	(95.80)	(2.15)	(45.89)	(0.73)	(20.71)	(50.93)

the sample collection in three countries, a series of in situ measurements were conducted: pH, Eh, water temperature (W_t) (HORIBA d-54 m), electrical conductivity, and total dissolved solids (TDS) (ORION 3 STAR, Thermo Electron Corporation).

2.2. Sample analysis

Total As concentrations in the groundwater samples were measured by an inductively coupled plasma spectrometer (Agilent 7500ce, with a detection limit of 0.05 μg L^{-1}). Accuracy and precision of measurements were checked using a reagent blank, instrument calibration standard, and standard reference material for trace metals in natural water (SRM 1640). After every tenth sample during analysis, the SRM sample and calibration standards were analyzed to check the analysis accuracy. All samples were measured at least twice in order to assess the measurement reliability; samples were reanalyzed if the error either from the SRM or from the calibration standards exceeded 10% or the relative standard deviation of the measurement exceeded 5%. Dilution was made with 2% HNO_3 when the concentration of the sample was over the upper limit of the standard range (100 μg L^{-1}).

2.3. Back-propagation artificial neural network and SVM

2.3.1. Back-propagation artificial neural network

The back-propagation artificial neural network (BPANN) is one of the machine learning methods for modeling complex and complicated relationships between explanatory and dependent variables [30]. Basically, the BPANN consists of multiple layers of nodes which include an input layer, hidden layer, and output layer. These layers are sequentially connected by links which transfer the signal to the next layers. The signal strength is changed by multiplying the weights and then transferred to the next node in the network where linear or nonlinear transfer functions are used to transform the signal. After estimating the errors between As prediction and observation, the learning algorithm will be activated to update weight factors in an attempt to make a better prediction of groundwater As concentrations. This process is continued until the model satisfies the performance goal. The BPANN used for predicting groundwater As concentration consists of three layers (input, hidden, and output layers) with N input nodes, L hidden nodes, and K output nodes. It can be expressed as follows [17]:

$$O_{Pk} = f_1\left(\sum_{i=1}^{L} w^0_{jk} f_2\left(\sum_{j=1}^{N} w^h_{ij} x_{Pi} + b^j_1\right) + b^k_2\right), k \in 1, 2, \ldots, K \tag{1}$$

where O_{Pk} is the output from the kth node of the output layer in the network for the Pth input data vector, x_{Pi} is the ith element of the Pth input vector (accepted by the ith node of the input layer), w^0_{jk} is the connection weight between the jth node of the hidden layer and the kth node of the output layer, w^h_{ij} is the connection weight between the ith node of the hidden layer and the jth node of the input layer, b^j_1 and b^k_2 are bias terms, N, L, and K are the number of nodes in the input layer, hidden layer, and output layer, respectively, and f_1 (·) and f_2 (·) are the activation functions in the input and hidden layers, respectively. Here, we tested log-sigmoid, pure linear, and tan-sigmoid as transfer functions for the hidden and output layers.

2.3.2. Support vector machine

The SVM has drawn much attention as an excellent tool for classification and regression because of its many attractive advantages and its generalization ability [33,35]. The advantage of an SVM over other machine learning methods is that it has superior generalization ability with a relatively small number of observations [36]. The main idea of the SVM is to map the training samples from the input space into a higher dimensional feature space using a nonlinear mapping function Φ. The function is typically unknown and performs linear regression in the feature space [32]. Here, the regression addresses a problem of estimating a function based on a given dataset $G = \{(x_i, d_i)\}^1_{i=1}$ (x_i is input vector, d_i is the desired value). SVM approximates the function from the following equation:

$$y = \sum_{i=1}^{l} w_i \Phi_i(x) + b \tag{2}$$

where $\{\Phi_i\ (x)\}^1_{i=1}$ are the feature of inputs, $\{w_i\}^1_{i=1}$ and b are coefficients. They are calculated by minimizing the regularized risk function ($R(C)$).

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}||w||^2 \tag{3}$$

where

$$L_\varepsilon(d,y) = \begin{cases} |d-y| - \varepsilon|d-y| \geq \varepsilon \\ 0 \qquad \text{others} \end{cases} \tag{4}$$

where ε is a prescribed parameter. The L_ε (d, y) is an ε-insensitive loss function which does not discipline errors less than ε. The term $\frac{1}{2}||w||$ measures function flatness. C is a regularized constant determining the trade-off between training error and the model flatness.

2.4. Modeling parameter optimization

The learning algorithm of the ANN contains learning and momentum rates which significantly influence the training process and the model performance. In addition, the size of the network (i.e. the number of hidden nodes) has a considerable effect on the model performance [37]. If the size of structure is set up with too many hidden nodes, it could cause an over-fitting process which means that the model could not be useful for the prediction problem. Otherwise, the model performance could not be satisfactory with the simplified structure and non-optimized model parameters [38].

Just as with the ANN model, the model parameters of the SVM also have a significant influence on the model performance, implying that the parameters need to be optimized with proper methods [39]. There are three significant parameters in the SVM model, including the number of C, epsilon, and sigma. As well, we are also facing the global optimization problem when we search for the optimal model parameters [40]. Hence, the pattern search algorithm was applied to find the globally optimal parameters for the ANN and SVM models. The ranges of the ANN parameters were chosen by previous studies [30,41,42]. For the SVM, the pattern search algorithm was applied to determine the number of C, epsilon, and sigma within its parameter range [43].

2.5. Model training strategy

The total dataset was divided to separate training and validation datasets for the ANN and SVM models. A total of 10 disjoint subsets of data, which contain 10% of the total data, were used to evaluate the validation errors. The remaining 90% of the data were randomly assigned into two different subsets of data: training (70%) and testing (20%). The training set was used to determine the optimal network weights ($w_{ij}{}^{h}$, $w_{jk}{}^{h}$), and the test set was used to decide the optimal iterations for ANN training. After the determined stopping point, the validation set was then used to determine the validation error of the ANN. This process was performed 10 times to calculate the validation errors for each of the validation subsets; these validation errors were subsequently compared with the errors of SVM models.

3. Results and discussion

3.1. Relationship between As and conventional parameters

Among the three countries, groundwaters in Cambodia can be characterized by a high level of groundwater As concentration with very low redox potential (Table 1). For most of the samples from Cambodia, arsenic concentrations exceeded 10 μg L^{-1} (the WHO drinking water guide value) and 50 μg L^{-1} (Cambodian drinking water legal limit), reaching to 1110.23 μg L^{-1}. In Laos, mean As concentration does not exceed the WHO drinking water guide value, but samples collected from the downstream of the Mekong River showed relatively high concentrations ranging from 0.59 to 71.06 μg L^{-1}. All samples from Thailand do not exceed the WHO dirking water guide value for arsenic. We found significant relationships among As concentrations and conventional parameters (Fig. 1). We see a strong negative relationship (r = −0.72) between redox potential and total As concentrations which indicates that high concentrations of As might be caused by reducing conditions [30,44,45]. As well, groundwater As concentrations are positively correlated with TDS, which is strongly associated with well depth.

Cho et al. [30] demonstrated that the predictive performance of linear models is not satisfactory in terms of Nash–Sutcliffe coefficients (NSE) values, indicating that the vigorous variation of groundwater As concentration could not be reproduced by linear models. This is the reason that we apply machine learning theories that include nonlinear transfer or mapping functions to reproduce the nonlinear relationship between groundwater As concentrations and conventional water quality parameters.

3.2. Optimization of ANN and SVM models

Prior to modeling approaches, the model parameters in the ANN and SVM were optimized in an attempt to construct the best model and to maximize the model performance. The pattern search algorithm from MATLAB, a generic algorithm, was used to determine optimal parameters for both the ANN and SVM models [46]. For the ANN model, learning and

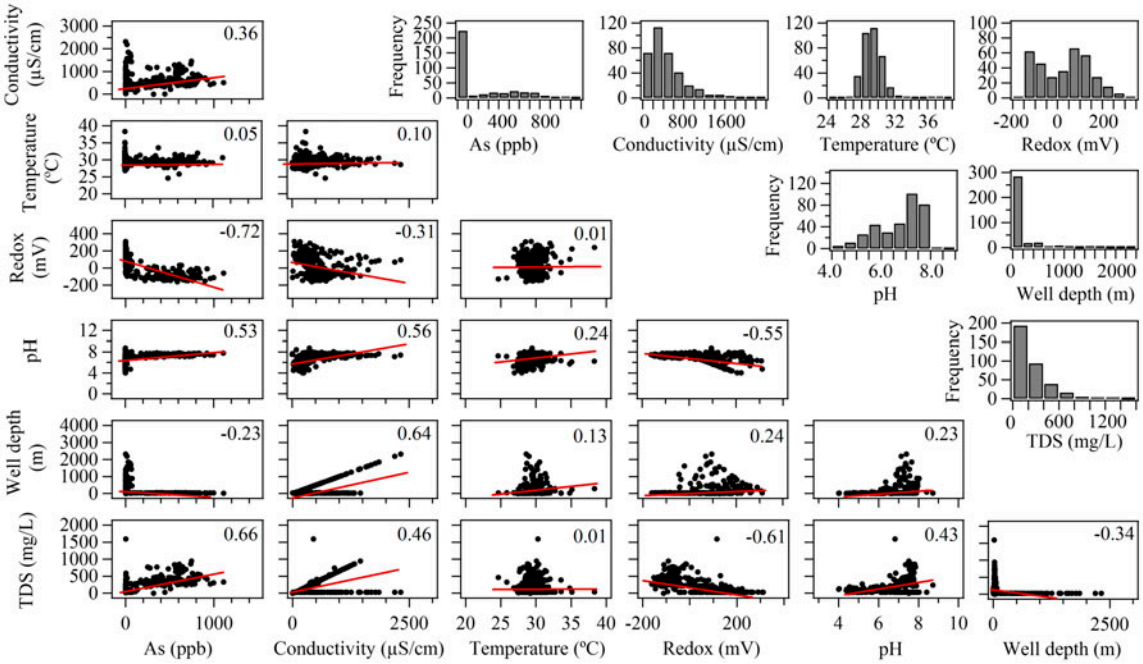

Fig. 1. Correlation matrix among total As concentration (ppb), conductivity (μS/cm), temperature (°C), redox (mV), pH, well depth (m), TDS (mg/L); the red line is the linear regression line; the number in each box is Pearson's correlation value; the gray bar means the histogram of each parameter.

momentum rates and the number of hidden layer nodes were optimized, while the type of Kernel function, *C*, epsilon, and sigma was adjusted to find the optimal set for the SVM model. The lower and upper limit of target parameters was selected from the previous environmental application of the ANN model [30,41]. As well, the reported range of the SVM model parameters from the previous study by Wang et al. [43] was selected to determine the optimal parameters set. Table 2 shows the optimization results of ANN and SVM parameters for modeling groundwater As concentration. The optimal parameter set is the combination of pure linear and log-sigmoid transfer functions with a learning rate of 0.8 and a momentum rate of 0.3. Overall, NSE values in both the training and validating steps did not change significantly with different parameter sets. Similar to the ANN model, we found that the performance of SVM was not significantly influenced by the model parameters and kernel functions. As shown in Table 2, an RPF function with *C* of 100, epsilon of 0.1, and sigma of 26.63 was the optimal parameter set for the SVM model which results in 0.76 and 0.58 for each training and validation step.

Table 2
Optimization of ANN and SVM model parameters for modeling groundwater As concentration

	Model parameter			NSE		Kernel function	Model parameter			NSE	
Transfer function	Lr	Mo	# of hidden neuron	Tr	Vl		C	Epsilon	Sigma	Tr	Vl
Logsig-Logsig	0.70	0.11	11	0.66	0.50	Exponential RBF	51.13	0.069	6.38	0.71	0.45
Logsig-purelin	0.90	0.90	16	0.65	0.52	Gaussian RBF	50	0.069	9.63	0.61	0.48
Logsig-Tansig	0.59	0.25	10	0.62	0.48	RBF	100	0.1	26.63	0.76	0.58
Purelin-Logsig	0.80	0.30	11	0.66	0.52						
Purelin-Purelin	0.10	0.10	2	0.61	0.47						
Purelin-Tansig	0.10	0.10	5	0.60	0.46						
Tansig-Logsig	0.90	0.40	16	0.67	0.51						
Tansig-Purelin	0.10	0.35	3	0.62	0.52						
Tansig-Tansig	1.00	0.50	5	0.66	0.51						

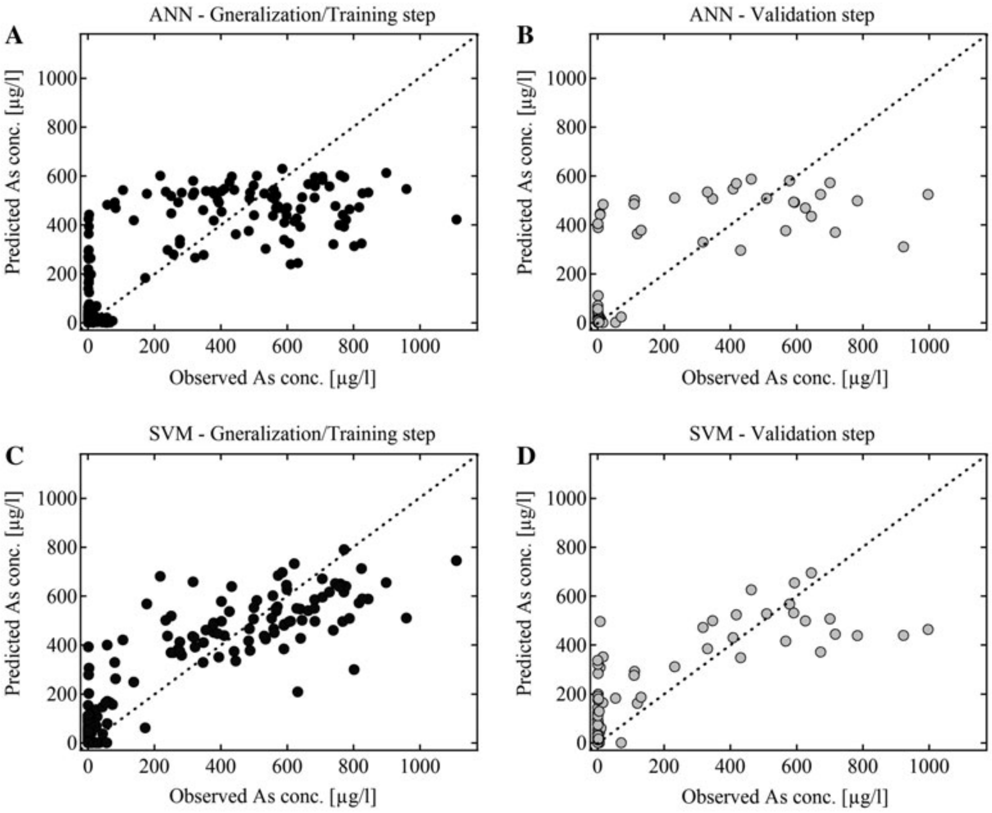

Fig. 2. Comparison between observed and predicted As by ANN and SVM models.

3.3. SVM and ANN modeling results

After obtaining the optimal parameters set, ANN and SVM models were trained to model groundwater As concentrations with conventional parameters as input data. A total of 313 samples were used to train and validate the two models. Fig. 2 illustrates the training and validating results of the ANN and SVM models. As well, the prediction accuracies of the two models were quantified using the NSE [47]. NSEs of SVM (training = 0.76 and validation = 0.58) are slightly higher than ANN models (training = 0.66 and validation = 0.52), showing acceptable prediction accuracies [48]. However, both models have limited ability to reproduce the higher level of As concentrations which were mostly collected from Cambodia. This can be explained by the fact that the nonlinearity of both models cannot fully reproduce the high level of As concentrations. In addition, the range of groundwater As concentrations from the three different countries is too broad to be modeled by ANN and SVM models. The concentrations measured in this study ranged from undetectable to 1110.23 μg L^{-1}, implying that As observation is much too variable to be modeled by two models.

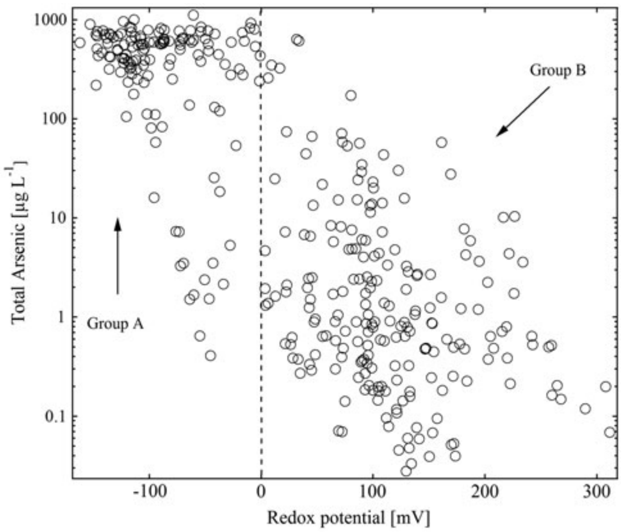

Fig. 3. Comparison between observed and predicted As by ANN and SVM models.

3.4. The relationship between redox potential and As concentration

Eh is significantly related to As concentration as shown in Fig. 3 where the dataset was divided into two different groups, Group A and Group B. Most of the samples in Group A collected from Kandal Province in Cambodia and samples in Group B were measured from Lao PDR and Thailand. It is reported that a reduced state of As(III) is the major species in Kandal Province where the majority of As exist in highly reducing aquifer regions having a low Eh level [10,30,49,50]. Here, Group B can be characterized by As in a reducing region, showing a strong negative correlation with Eh. The strong correlation might result from a combination of low ferrous and high Eh in Lao PDR [30,51]. Consequently, we found a significantly different dependency of As concentration in response to Eh.

Table 3
Optimization of the model parameters for modeling groundwater As concentration in Group A

	Model parameter			NSE		Kernel function	Model parameter			NSE	
Transfer function	Lr	Mo	# of hidden neuron	Tr	Vl		C	Epsilon	Sigma	Tr	Vl
Logsig-Logsig	0.90	0.65	17	0.56	0.49	Linear	80.22	0.10	10.00	0.47	0.48
Logsig-purelin	0.33	0.80	16	0.56	0.52	Exponential RBF	86.41	0.05	5.14	0.87	0.59
Logsig-Tansig	0.80	0.30	10	0.51	0.47	Gaussian RBF	8.77	0.09	0.25	**0.72**	**0.62**
Purelin-Logsig	1.00	0.13	11	0.47	0.48	RBF	14.75	0.10	6.44	0.71	0.61
Purelin-Purelin	0.10	0.73	2	0.47	0.45	MLP	79.61	0.44	6.15	−0.01	0.01
Purelin-Tansig	0.30	0.30	6	0.48	0.46						
Tansig-Logsig	0.87	0.93	17	**0.60**	**0.54**						
Tansig-Purelin	0.10	0.10	15	0.57	0.51						
Tansig-Tansig	0.32	0.30	12	0.58	0.51						

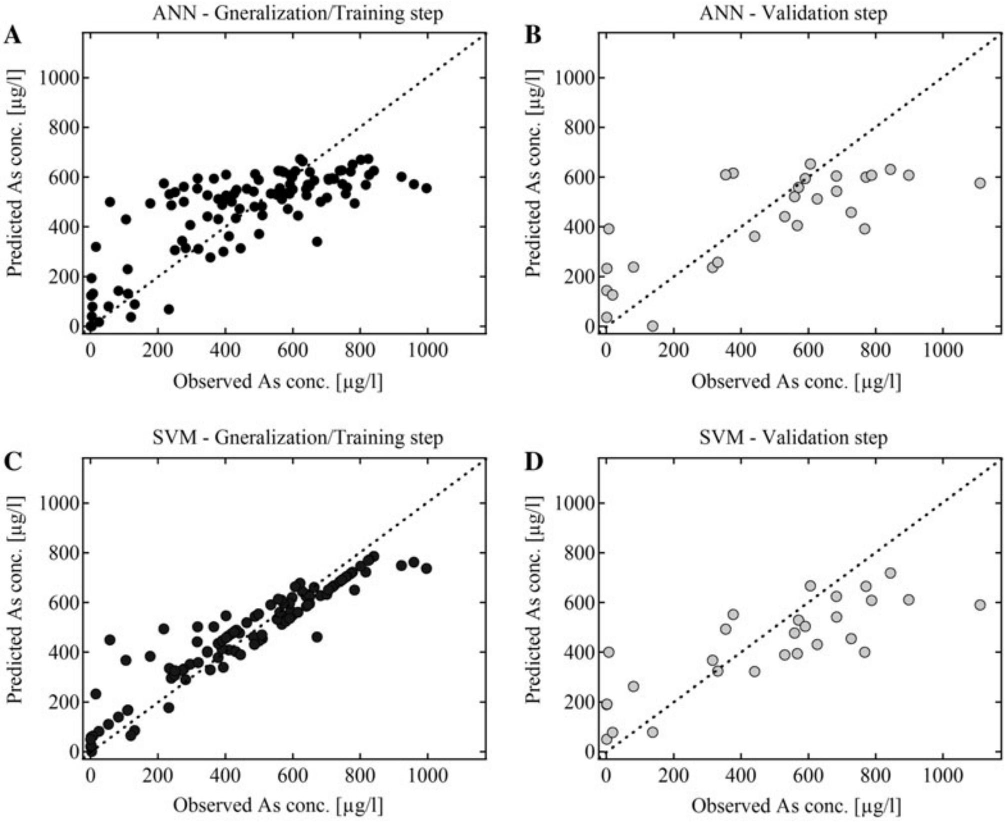

Fig. 4. Comparison between observed and predicted As by ANN and SVM models in Group A.

This is the reason that application of ANN and SVM could not be very satisfactory to predict As concentrations from three different countries. Here, we only applied ANN and SVM models for predicting As concentration in Group A where the models show the limitation of predicting higher As concentration levels.

3.5. Prediction of As concentration in Group A

The ANN and SVM were applied to predict groundwater As concentration in Group A which is characterized by high groundwater As concentrations and low Eh. Table 3 shows the model results (NSE values) in response to different transfer functions and model parameters in Group A. We found that the model performance is not very sensitive to different learning and momentum rates. For the ANN, the combination of Tan-sigmoid function in the hidden layer and Log-sigmoid in the output layer showed the highest NSE values. Relatively higher learning and momentum rates were determined as the optimal parameters. In particular, the momentum rate for Group A is three times greater than that of the ANN for the whole dataset. As well, the number of hidden nodes (17) in Group A is also more than the whole dataset (11). For the SVM model, even though the NSE value of Group A (0.72) is slightly less than the whole dataset (0.76) in the training step, it showed improved predictive accuracy in the validation step from 0.58 to 0.62. Fig. 4 compares the observed As concentrations with predicted As concentrations. As shown in Fig. 4(A) and (B), the ANN model could not make a good fitting on higher As concentrations. In particular, we clearly see the limitation in the training step. Conversely, the SVM model demonstrated the better performance on Group A, showing higher NSE values compared to the ANN model. Fig. 4 also demonstrates that the SVM could be a superior model for predicting As concentrations in Group A. Here, we found better performance of SVM for modeling groundwater As concentrations, compared to the ANN model.

Modeling works on groundwater As concentrations have been conducted with geological information and soil properties [52]. The accurate prediction of As, however, is still a challenging task. This study was focused on high resolution-scale prediction by incorporating 5 years of field study and modeling works. It will be more robust and reliable with a new dataset from another country. The models developed in this study can be utilized by the public via web-based services or local government offices in different SEA countries.

4. Conclusions

Groundwater is a very important drinking water resource in SEA countries. Comprehensive investigations of As in SEA countries have been conducted by many researchers in the world. However, accurate quantification of groundwater As contamination can be a burden for SEA countries because it requires advanced and expensive equipment and well-trained technicians. Here, we proposed an alternative way to quantify groundwater As contamination using conventional parameters. ANN and SVM were used to model groundwater As concentrations. Major findings in this study are:

(1) Among three countries (Cambodia, Laos, and Thailand), the Cambodian samples showed the highest level of groundwater As concentrations with very low redox potential. It exceeded the WHO drinking water guide value of 10 μg L^{-1} and the Cambodian drinking water legal limit of 50 μg L^{-1}, reaching to 1110.23 μg L^{-1}.
(2) Two machine learning models (i.e. ANN and SVM) showed acceptable prediction accuracy for modeling groundwater As concentrations using conventional water quality parameters, but tend to underestimate high levels of As concentration.
(3) The total dataset was separated into two different groups in terms of redox potential. We extracted a higher As concentration group where a reduced state of As(III) is the major species.
(4) Consequently, we found that the performance of the SVM model is slightly better than the ANN model for predicting the higher concentrations of groundwater As.

This study provides two different machine learning models for quantifying groundwater As concentrations for SEA countries, including Cambodia, Laos, and Thailand. We hope that these models could be a useful assessment tool for establishing better strategies for public health concerning As toxicity from consuming contaminated groundwater.

Acknowledgments

This study was funded by the Toyota Foundation, grant number D11-R-81. And this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2059680).

References

[1] P. Bagla, J. Kaiser, India's spreading health crisis draws global arsenic experts, Science 274 (1996) 174–175.

[2] American Water Works Association (AWWA), Arsenic Rule, Mainstream, 45, 2001.

[3] M. Berg, S. Luzi, P.T. Kim, P.H. Viet, W. Giger, D. Stuben, Arsenic removal from groundwater by household sand filters: Comparative field study, model calculations, and health benefits, Environ. Sci. Technol. 40 (2006) 5567–5573.

[4] B.D. Kocar, C. Fendorf, Thermodynamic constraints on reductive reactions influencing the biogeochemical of Arsenic in soils and sediments, Environ. Sci. Technol. 43 (2009) 4871–4877.

[5] S. Sthiannopkao, K.W. Kim, K.H. Cho, K. Wantala, S. Sotham, C. Sokuntheara, J.H. Kim, Arsenic levels in human hair, Kandal Province, Cambodia: The influences of groundwater arsenic, consumption period, age and gender, Appl. Geochem. 25 (2010) 81–90.

[6] M. Berg, H.C. Tran, T.C. Nguyen, H.V. Pham, R. Schertenleib, W. Giger, Arsenic contamination of groundwater and drinking water in Vietnam: A human health threat, Environ. Sci. Technol. 35 (2001) 2621–2626.

[7] M. Berg, H.C. Tran, T.C. Nguyen, H.V. Pham, R. Schertenleib, W. Giger, Magnitude of arsenic pollution in the Mekong and Red River Deltas—Cambodia and Vietnam, Sci. Total Environ. 372 (2001) 413–425.

[8] P.L. Smedley, D.G. Kinniburgh, A review of the source, behaviour and distribution of arsenic in natural waters, Appl. Geochem. 17 (2002) 517–568.

[9] G. Sun, J. Liu, T.V. Luong, D. Sun, L. Wang, Endemic Arsenicosis: A Clinical Diagnostic with Photo Illustrations, UNICEF East Asia and Pacific Regional Office, Bangkok, 2002.

[10] D.A. Polya, A.G. Gault, N.J. Bourne, P.R. Lythgoe, D.A. Cooke, Coupled HPLC-ICP-MS analysis indicates highly hazardous concentrations of dissolved arsenic species in Cambodian groundwaters, in: J. Holland, S.D. Tanner (Eds.), Plasma Source Mass Spectrometry: Applications and Emerging Technologies, Royal Society of Chemistry, Cambridge, 2003, pp. 127–140.

[11] D.A. Polya, A.G. Gault, N. Diebe, P. Feldman, J.W. Rosenboom, E. Gilligan, D. Fredericks, A.H. Milton, M. Sampson, H.A.L. Rowland, P.R. Lythgoe, C. Middleton, D.A. Cooke, Arsenic hazard in shallow Cambodian groundwaters, Mineral Mag. 69 (2005) 807–823.

[12] G. Stanger, T.V. Truong, K.S.L.T. My Ngoc, T.V. Luyen, T.T. Tuyen, Arsenic in groundwaters of the Lower Mekong, Environ. Geochem. Health 27 (2005) 341−357.

[13] A. Kohnhorst, Arsenic in groundwater in selected countries in South and Southeast Asia: A review, J. Tropical Medicine Parasitology 28 (2005) 73–82.

[14] A. Tetsuro, K. Takashi, F. Junko, K. Reiji, B.M. Tu, T.K.T. Pham, I. Hisato, S. Annamalai, H.V. Pham, T. Shinsuke, Contamination by arsenic and other trace elements in tube-well water and its risk assessment to humans in Hanoi, Vietnam, Environ. Pollut. 139 (2006) 95–106.

[15] H. Chiew, M.L. Sampson, S. Huch, S. Ken, B.C. Bostick, Effect of groundwater iron and phosphate on the efficacy of arsenic removal by iron-amended biosand filters, Environ. Sci. Technol. 43 (2009) 6295–6300.

[16] N.K.C. Twarakavi, D. Mishra, S. Bandopadhyay, Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data, Nat. Resour. Res. 15 (2006) 15–26.

[17] M. Norgaard, Neural Network Based System Identification Toolbox, Version 2, Department of Automation, Technical Report 00-E-891, Technical University of Denmark, Lungby, 2000.

[18] B. Widrow, D.E. Rumelhart, M.A. Lehr, Neural networks: Applications in industry, business and science, Commun. ACM 37 (1994) 93–105.

[19] H.R. Maier, G.C. Dandy, The use of artificial neural networks for the prediction of water quality parameters, Water Resour. Res. 32 (1996) 1013–1022.

[20] C.G. Wen, C.S. Lee, A neural network approach to multiobjective optimization for water quality management in a river basin, Water Res. 34 (1998) 427–436.

[21] G.M. Brion, S. Lingireddy, A neural network approach to identifying non-point sources of microbial contamination, Water Res. 33 (1999) 3099–3106.

[22] J.H. Lee, M.J. Yu, K.W. Bang, J.S. Choe, Evaluation of the methods for first flush analysis in urban watersheds, Water Sci. Technol. 48 (2003) 167–176.

[23] S. Riad, J. Mania, L. Bouchaou, Y. Najjar, Rainfall-runoff model using an artificial neural network approach, Math. Comput. Model. 40 (2004) 839–846.

[24] A. Sarangi, A.K. Bhattacharya, Comparison of artificial neural network and regression models for sediment loss prediction from Banha watershed in India, Agr. Water Manage. 78 (2005) 195–208.

[25] G. Tayfur, D. Swiatek, A. Wita, V.P. Singh, Case study: Finite element method and artificial neural network models for flow through Jeziorsko earthfill dam in Poland, J. Hydraul. Eng. 131 (2005) 431–440.

[26] M. Holmberg, M. Forsius, M. Starr, M. Huttunen, An application of artificial neural networks to carbon, nitrogen and phosphorus concentrations in three boreal streams and impacts of climate change, Ecol. Model. 195 (2006) 51–60.

[27] J.T. Kuo, P.H. Hsieh, W.S. Jou, Lake eutrophication management modeling using dynamic programming, J. Environ. Manage. 88 (2008) 677–687.

[28] B. Purkait, S.S. Kadam, S.K. Das, Application of artificial neural network model to study arsenic contamination in groundwater of Malda Disrict, Eastern India, J. Environ. Inf. 12 (2008) 140–149.

[29] F.J. Chang, L.S. Kao, Y.M. Kuo, C.W. Liu, Artificial neural networks for estimating regional arsenic concentrations in a blackfoot disease area in Taiwan, J. Hydrol. 388 (2010) 65–76.

[30] K.H. Cho, S. Sthiannopkao, Y.A. Pachepsky, K.W. Kim, J.H. Kim, Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network, Water Res. 45 (2011) 5535–5544.

[31] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.

[32] V.N. Vapnik, S. Golowich, A.J. Smola, Support vector method for function approximation, regression estimation, and signal processiong, Adv. Neural Inf. Process. Syst. 9 (1997) 281–287.

[33] V. Vapnik, Statistical Learning Theory, Wiley, New York, NY, 1998.

[34] R.S. Govindaraju, Artificial neural networks in hydrology. II: Hydrologic applications, J. Hydrol. Eng. 5 (2000) 124–137.

[35] B. Schölkopf, A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.

[36] S. Osowski, K. Garanty, Forecasting of the daily meteorological pollution using wavelets and support vector machine, Eng. Appl. Artif. Intell. 20 (2007) 745–755.

[37] H.R. Maier, G.C. Dandy, The effect of internal parameters and geometry on the performance of back-propagation neural networks: An empirical study, Environ. Model. Softw. 13 (1998) 193–209.

[38] G. Bebis, M. Georgiopoulos, Feed-forward neural networks, IEEE Potentials 13 (1994) 27–31.

[39] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, Neural Networks 17 (2004) 113–126.

[40] Y. Ren, G. Bai, Determination of optimal SVM parameters by using GA/PSO, J. Comput. 5 (2010) 1160–1168.

[41] H.R. Maier, G.C. Dandy, Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications, Environ. Model. Softw. 15 (2000) 101–124.

[42] R. Patuelli, A. Reggiani, P. Nijkamp, N. Schanne, Neural networks for regional employment forecasts: Are the parameters relevant? J. Geog. Sci. 13 (2011) 67–85.

[43] W. Wang, Z. Xu, W. Lu, X. Zhang, Determination of the spread parameter in the Gaussian kernel for classification and regression, Neurocomputing 55 (2003) 643–663.

[44] M. Berg, C. Stengel, P.T.K. Trang, P. Hung Viet, M.L. Sampson, M. Leng, S. Samreth, D. Fredericks, Magnitude of arsenic pollution in the Mekong and Red River Deltas—Cambodia and Vietnam, Sci. Total Environ. 372 (2007) 413–425.

[45] J. Buschmann, M. Berg, C. Stengel, M.L. Sampson, Arsenic and manganese contamination of drinking water resources in Cambodia: Coincidence of risk areas with low relief topography, Environ. Sci. Technol. 41 (2007) 2146–2152.

[46] R.M. Lewis, T.A. Virginia, Globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds, SIAM J. Optimiz. 12 (2002) 1075–1089.

[47] J.E. Nash, J.V. Sutcliffe, River flow forecasting through conceptual models part I—A discussion of principles, J. Hydrol. 10 (1970) 282–290.

[48] D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, T.L. Veith, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, T. ASABE 50 (2007) 885–900.

[49] H.A.L. Rowland, A.G. Gault, P. Lythgoe, D.A. Polya, Geochemistry of aquifersediments and arsenic-rich groundwaters from Kandal Province, Cambodia, Appl. Geochem. 23 (2008) 3029–3046.

[50] M.L. Polizzotto, B.D. Kocar, S.G. Benner, M. Sampson, S. Fendorf, Near-surface wetland sediments as a source of arsenic release to ground water in Asia, Nature 454 (2008) 505–508.

[51] P. Chanpiwat, S. Sthiannopkao, K.H. Cho, K.W. Kim, V. San, B. Suvanthong, C. Vongthavady, Contamination by arsenic and other trace elements of tube-well water along the Mekong River in Lao RDR, Environ. Pollut. 159 (2011) 567–576.

[52] L. Winkel, M. Berg, M. Amini, S.J. Hug, C.A. Johnson, Predicting groundwater arsenic contamination in Southeast Asia from surface parameters, Nat. Geosci. 1 (2008) 536–542.