Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2015.1049411

(2015) 1–10

Taylor & Francis
Taylor & Francis Group

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.