Geosmin and 2-Methylisoborneol Adsorption using Different Carbon Materials: Isotherm, Kinetic, Multiple Linear Regression, and Deep Neural Network Modeling using a Real Drinking Water Source

Taeho Bong[a], Jin-Kyu Kang[b], Viviane Yargeau[c], Hye-Lim Nam[d], Sang-Hyup Lee[e], Jae-Woo Choi[e,f], Song-Bae Kim[b,g], Jeong-Ann Park[h,*]

[a] *Ecology & Environment Research Division, Gyeonggi Research Institute, Suwon 16207, Republic of Korea*

[b] *Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, 08826, Republic of Korea*

[c] *Department of Chemical Engineering, McGill University, 3610 University St., Montréal, H3A 0C5, Québec, Canada*

[d] *FITI Testing & Research Institute, Republic of Korea, Chungbuk 28115, Republic of Korea*

[e] *Water Cycle Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea*

[f] *Division of Energy & Environment Technology, KIST School, Korea University of Science and Technology, Seoul, 02792, Republic of Korea*

[g] *Environmental Functional Materials and Water Treatment Laboratory, Department of Rural Systems Engineering, Seoul National University, Seoul, 08826, Republic of Korea*

[h] *Department of Environmental Engineering, Kangwon National University, Chuncheon 24341, Republic of Korea*

Abstract

High concentrations of geosmin (GSM) and 2-methylisoborneol (MIB) caused by cyanobacterial blooms are problematic issues in drinking water treatment plants on the Nak-Dong River in South Korea. The goal of this study was to identify the best-performing carbon materials to treat GSM and MIB in river water and to predict the final concentrations of GSM and MIB according to different water quality parameters using computational predictive modeling, multiple linear regression (MLR), and a deep neural network (DNN) for practical applications. Three types of powdered activated carbon (PAC) and three types of mesoporous carbon (MC) were compared in terms of their adsorption capacities using a source of drinking water from the Nak-Dong River. The highest maximum adsorption capacities were achieved when using C-PAC (1485.06 μg/g) in both distilled water and river water. C-PAC possesses a high micropore volume (0.45 cm^3/g), low mesopore volume (0.11 cm^3/g), and small, narrow pore size distribution, yielding optimal adsorption conditions. PACs yielded better results than MCs because dominant mesopore structures can hinder the adsorption of small molecules (0.6 to 0.8 nm) and allow them to pass through pores. GSM was removed more effectively than MIB because GSM is more hydrophobic and has a flatter structure. The high dissolved organic carbon concentration in the river water caused no reduction in GSM and MIB adsorption, but actually enhanced adsorption because a small portion of natural organic matter can compete with adsorption sites and provide additional adsorption sites based on the shrunk pore effect. The MLR and DNN models were used to predict the removal efficiency of C-PAC for GSM and MIB using 72 individual datasets with cross-validation for robust prediction and sensitivity

analysis. MLR predicted the relationships between the input variables and GSM and MIB concentrations with an acceptable mean absolute error (MAE) of 6.43 ng/L for GSM and 5.80 ng/L for MIB. However, the DNN provided better agreement with a higher R^2 value (>0.99) and lower MAE of 1.67 ng/L for GSM and 1.24 ng/L for MIB.

Keywords: carbon materials; deep neural network modeling; dissolved organic carbon; drinking water; multiple linear regression modeling; taste and odor compounds.

1. Introduction

Two of the most common earthy-musty taste and order compounds, namely geosmin (GSM) and 2-methylisoborneol (MIB), found in drinking water sources cause significant problems in drinking water treatment plants (DWTPs) (Butakova, 2013; Qian et al., 2014; He et al., 2016). High concentrations of GSM and MIB are caused by periodic cyanobacterial blooms of several cyanobacteria genera, including *Anabaena, Phormidium*, and *Planktothrix* (Antonopoulou et al., 2014; Hafuka et al., 2019). Non-treated GSM and MIB, even at low ng/L levels, reduce the perceived quality of drinking water and increase operation costs based on unpleasant smells resulting in consumer complaints (Im et al., 2019). Therefore, the Ministry of the Environment in South Korea has regulated the concentrations of GSM and MIB in drinking water to 20 ng/L (MOE, 2015). Additionally, Japan, China, and Australia have set even stricter limits at 10 ng/L (MOH, 2015; MOH & SAC, 2006; NHMRC, 2016).

Because conventional processes such as coagulation, filtration, and chlorination are ineffective in removing GSM and MIB, adsorption using carbon materials (CMs) has been shown to be a simple and effective method for removing these compounds in DWTPs (Pan et al., 2016; Zamyadi et al., 2015). Various types of CMs, including granular activated carbon

(GAC) (Greenwald et al., 2015; Huang et al., 2019), powdered activated carbon (PAC) (Cook et al., 2001; Newcombe et al., 1997; Yu et al., 2007; Berton et al., 2018), superfine PAC (Matsui et al., 2012; Pan et al., 2016), graphene oxide (Hafuka et al., 2019), and activated carbon fiber (Srinivasan & Sorial, 2009), have been used for treating GSM and MIB. However, only GAC and PAC are commercially available and easily applicable in DWTPs. The addition of PAC is the most common and cost-effective method for treating GSM and MIB because these compounds only appear periodically in DWTPs (Cook et al., 2001; Berton et al., 2018). However, the GSM and MIB adsorption capacities of PAC are highly variable and dependent on physicochemical characteristics such as hydrophobicity (Matsui et al., 2015), particle size (Matsui et al., 2015), pore size distribution (Newcombe et al., 1997), raw material (Thiel & Cullum, 2007), and micropore volume (Yu et al., 2007). Although there is knowledge available regarding the adsorption of these compounds by PAC, information related to the pore characteristics affecting both GSM and MIB adsorption is insufficient and the important criteria for selecting specific CMs to remove GSM and MIB are still controversial. Additionally, the effects of mesopores on GSM and MIB adsorption using PAC and GAC have been reported to minimize the natural organic matter (NOM) pore blockage mechanism (Newcombe et al., 2002; Nowack et al., 2004; Yu et al., 2007). However, the role of mesopore structures is still unclear because mesoporous carbons (MCs) have not been fully investigated for GSM and MIB removal in real drinking water sources. Therefore, we investigated six types of CMs, including three PACs and three MCs, for the removal of GSM and MIB in both distilled water (DW) and Nak-Dong River water.

Although adsorption isotherms and kinetics are traditional models for evaluating adsorption behavior using DW, they provide limited information for the practical application of CMs in DWTPs, because the adsorption of GSM and MIB in real drinking water sources heavily

depends on water quality factors (NOM concentration, alkalinity, turbidity, color, pH), initial GSM and MIB concentrations, doses of CMs, and contact time. Therefore, computational predictive modeling is required for the practical application of CMs in DWTPs based on data from experiments using actual drinking water sources and a range of different water quality parameters and initial GSM and MIB concentrations, which have not been fully explored. One of the largest drinking water sources in South Korea, namely Nak-Dong River water, annually suffers from periodic cyanobacterial blooms (Ha et al., 2002). This water source was selected for our study because it is frequently impacted by GSM and MIB. Multiple linear regression (MLR) analysis and a deep neural network (DNN) were considered for comparing the prediction accuracy of required CM doses and contact times according to water quality parameters and initial GSM and MIB concentrations in Nak-Dong River water. MLR analysis is one of the most widely used statistical techniques for considering two or more independent variables to predict a dependent variable. Artificial neural networks (ANN) with the back-propagation algorithm have been widely used to solve various forecasting problems (e.g., Zakeri et al., 2021) based on their advantages of not requiring the prior specification of a suitable fitting function and precise modeling of nonlinear complex systems (Desai et al., 2008). Recently, DNN has been used to overcome the overfitting and vanishing gradient problems observed in ANN. DNN model has recently been applied to predict the removal efficiency of contaminants in water for optimizing processes without performing massive experiments. Özdemir et al. (2011) successfully predicted the adsorption of sodium dodecyl benzene sulfonate onto polyaniline in DW using MLR and ANN models and Ahmad et al. (2020) predicted dye adsorption on waste-based adsorbents in DW using a DNN. Additionally, Sit et al. (2020) conducted a comprehensive review of deep learning approaches in the water industry and Mamandipoor et al. (2020) predicted faults in wastewater treatment plants using a DNN.

However, there has been a lack of predictive modeling studies using natural water. Furthermore, modeling studies on GSM and MIB treatment using real water are very scarce, although the selection of the best-performing PACs for real drinking water has been studied continuously (Bruce et al., 2002; Cook et al., 2001; He et al., 2016).

This study aimed to 1) analyze GSM and MIB adsorption isotherms and kinetics using six types of CMs, including three different origins of PACs and three different pore types of mesoporous carbons, to identify the characteristics of CMs affecting adsorption; 2) select the best-performing CM in both DW and a real source of drinking water (Nak-Dong River Water, South Korea); 3) investigate the effects of water quality factors (dissolved organic carbon concentration, turbidity, color, alkalinity, and initial GSM and MIB concentrations) on GSM and MIB adsorptions using river water; 4) predict the required dose of CMs and contact time to remove GSM and MIB to meet drinking water guidelines in river water by using MLR and DNN modeling for better understanding of removal trends according to water quality parameters; and 5) compare the results of both models to obtain more accurate predictions for practical application in DWTPs.

2. Materials and methods

2.1 Selected CMs

Six CMs were considered in this study. Three CMs were mesoporous carbons, namely MC1 (699632; Sigma-Aldrich, USA), MC2 (702102, Sigma-Aldrich, USA), and MC3, which were synthesized in a laboratory (Park et al., 2017a). The three other CMs were commercially used PACs of different origins, coal (S-PAC; 2.15 $/kg), coconut (C-PAC; 1.88 $/kg), and wood (W-PAC; 1.97 $/kg), which were provided by Shin-Kwang Chemicals (South Korea). The

characteristics of the six CMs were partially reported in our previous studies (Park et al., 2017a; 2018) and were evaluated using various techniques such as a surface area analyzer (ASAP2020, Micrometrities Co., USA), Fourier transform infrared spectroscopy (Nicolet 6700, Thermo Scientific, USA), dynamic light scattering spectrophotometer (DLS-7000, Otsuka Electronics, Japan), and electrophoretic light scattering analyzer (ELS-Z2, Otsuka Electronics, Japan). The characteristics of the six types of CMs are presented in the supporting materials (Tables S1 and S2).

2.2 GSM and MIB analysis using gas chromatography-mass spectrometry

GSM and MIB were purchased from Sigma-Aldrich (USA). The concentrations of GSM and MIB in the water samples were measured using the method described in our previous paper (Park et al., 2017b). Briefly, the GSM and MIB in water samples were extracted using DVB/CAR/PDMS fibers (50/30 μm; Supelco, USA) for 30 min at 70 °C and 400 rpm, and then passed through an HP-5MS UI capillary column (30 cm × 0.25 mm × 0.25 μm, Agilent, USA) for separation. Gas chromatography-mass spectrometry (Agilent 7890A, Agilent 5975C MSD, Germany) with a multi-purpose sampler (Gerstel, Germany) was used to measure the concentrations of GSM and MIB.

2.3 Equilibrium and kinetic studies for GSM and MIB adsorption on CMs

Equilibrium and kinetic experiments were conducted in batches using a 530 mL cap-sealed glass bottle. A solution containing both GSM and MIB was added to the top until no air remained in the glass bottle to except volatiles of GSM and MIB. The initial concentrations of

each GSM and MIB were 10 μg/L. The bottles were shaken at 25 °C and 150 rpm in a shaking incubator (JSSI-100T, JS Research Inc., Korea) with a specific dose of CM. All experiments were conducted in quintuplicate with duplicate analyses. For the experiments conducted to investigate the effects of the initial CM concentration (5.7 to 168.1 mg/L) on the adsorption of GSM (10 μg/L) and MIB (10 μg/L), the bottles were sampled after 24 h. The results were analyzed using the Langmuir, Freundlich, and Temkin isotherm models based on the values of R^2, SAE, and χ^2. Kinetic experiments were conducted with a consistent CM dose of 25.2 mg/L over 480 min. The data were fitted using pseudo-first-order, pseudo-second-order, and Elovich models, and the values of R^2, SAE, and χ^2 were calculated. All of the equations used in this process are explained in the supplementary materials.

2.4 GSM and MIB adsorption on CMs in a real drinking water source of Nak-Dong River water

Nak-Dong River water was collected on five different dates from February to April of 2016 in Daegu, South Korea. The collected water was passed through GF/C filters (Whatman; diameter: 4.7 cm, pore size: 1.2 μm) and stored in a refrigerator prior to characterization. All experiments were conducted within 48 h following collection. Water quality parameters were measured using an ultraviolet-visible spectrophotometer (Optizen POP QX, Mecasys Co. Ltd., Korea), turbidimeter (2100Q, Hach, USA), TOC-L analyzer (Shimadzu, Japan), spectrophotometer (DR/300, Hach, USA), alkalinity testing kit (24443-01. Hach, USA), and pH/ion meter (Orion, Thermo Fisher Scientific, USA). Table 1 lists the water quality parameters of the Nak-Dong River water with indexes of one to five representing the different sampling times. Aliquots of the river water samples designated from No. 1 to No. 3 were spiked with GSM and MIB to obtain initial concentrations of 10 μg/L for each compound. Each type

of CM (25.2 mg/L) was tested to select the best-performing CMs for treating Nak-Dong River water. To determine the effects of dissolved organic carbon (DOC) in real water on GSM and MIB adsorption using C-PAC, Suwannee River NOM (IHSS, Atlanta, GA, USA) was added to the Nak-Dong River water (No. 5, initial DOC concentration = 2.76 mg/L) to obtain different total DOC concentrations, which were measured at 3.42, 5.18, and 6.56 mg/L using a TOC analyzer (Shimadzu, Japan). Additionally, the NOM characteristics were analyzed using liquid chromatography-organic carbon detection (LC-OCD; DOC-Labor Huber, Germany). Finally, various initial concentrations (50, 100, and 200 ng/L) of each GSM and MIB with 32.1 mg/L of C-PAC were analyzed using aliquots of water sample No. 3 to estimate the appropriate adsorbent doses for different initial GSM and MIB concentrations in Nak-Dong River water.

2.5 MLR and DNN modeling based on experiments using a real drinking water source

All results from the GSM and MIB adsorption experiments (72 individual datasets) performed using C-PAC and Nak-Dong River water were considered for MLR and DNN modeling. Table S3 lists the data used in this study.

2.5.1 MLR analysis

MLR is a statistical technique used to predict outcomes based on independent variables (x_i) and is expressed as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon, \quad (1)$$

where ε is the prediction error and β_0 is the intercept of MLR. β_0, β_1, …, and β_n are the regression coefficients representing the changes in y according to changes in x. Here, the relationships between the dependent variable and each independent variable are linear.

Therefore, if modeling is performed using standard MLR, the concentration changes with time exhibit a linear relationship. However, the concentrations over time in the experimental results exhibit distinctly nonlinear decay trends. Therefore, the following modeling approach was adopted for accurate and robust modeling. Linear regression can be used on inherently nonlinear relationships by transforming the raw data. Exponential (Eq. 2) and power models (Eq. 3) are often applied as follows:

$$\text{Exponential model:} \quad y = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+ \cdots +\beta_n x_n}, \tag{2}$$

$$\text{Power model:} \quad y = e^{\beta_0} x_1^{\beta_1} x_2^{\beta_2} \cdots x_n^{\beta_n}. \tag{3}$$

The optimal transformation method depends on the raw data and a mixed model may be appropriate at times. Nonlinear transformation using a mixture of exponential and power models is defined as follows:

$$\text{Mixed model:} \quad y = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+ \cdots +\beta_n x_n} x_1^{\beta_1} x_2^{\beta_2} \cdots x_m^{\beta_m}. \tag{4}$$

These three equations were transformed into a linear sum of functions to calculate the regression coefficients using the least squares method. Various independent variables can be considered for regression modeling and appropriate variables must be selected to establish an optimal regression equation. Determining such equations typically requires three selection procedures: forward selection, stepwise selection, and backward elimination. However, these methods do not always yield optimal variables. Helsel and Hirsch (2002) discussed the disadvantages of existing variable selection methods and stated that there are many benefits to selecting an appropriate model by evaluating all possible combinations of input variables. Therefore, the number of independent variables was set to 3, 4, and 5, and the performances of all possible combinations were evaluated for three transformed nonlinear models through cross-validation error analysis. The multicollinearity between the input variables was evaluated

using a variance inflation factor (VIF) calculated as VIF=1/(1−R^2), where R^2 is the coefficient of determination. To determine the influence of input variables on GSM and MIB concentrations, global sensitivity analysis, which is a useful method for determining the importance of input variables (Bassaganya-Riera, 2015), was conducted using Monte Carlo simulations.

2.5.2 DNN modeling

Deep learning, which is a subfield of machine learning, utilizes an ANN with multiple hidden layers between the input and output layers. Hinton et al. (2006) demonstrated that proper initialization using a restricted Boltzmann machine can mitigate the overfitting problem and resolve several problems related to backpropagation. The vanishing gradient problem in traditional ANNs depends on the type of activation function and is not a fundamental problem related to neural networks. Therefore, in this study, the rectified linear unit function proposed by Nair and Hinton (2010) was selected as an activation function based on its ability to overcome the disadvantages of the sigmoid function and effectively alleviate the overfitting problem. TensorFlow, which is an open-source deep learning library developed by Google, was used to implement our DNN. Raw data contain variables with various scales and data preprocessing often plays an important role in deep learning. Common data preprocessing methods include normalization and standardization. Normalization refers to rescaling values to a range of [0, 1] and standardization refers to rescaling data to have a mean of zero and standard deviation of one (unit variance). Here, min-max normalization was applied for data preprocessing as follows:

$$z = \frac{x - min(x)}{[max(x) - min(x)]}. \quad (5)$$

Although DNN can alleviate the overfitting problem to some extent compared to traditional ANN, it is still an important problem. A common method for verifying the performance of a DNN model is to divide a dataset into training and validation sets and then evaluate model performance with a validation set. However, different splits of data may result in different performance levels and having insufficient data makes it difficult to evaluate the performance of a model accurately. K-fold cross-validation is a method for improving the estimated performance of a DNN model by repeating the cross-validation procedure multiple times and evaluating the performance of the model based on average results. For K-fold cross-validation, the 72 experimental datasets were randomly divided into 24 groups. Training was performed using 23 groups and verification was performed on the one excluded group. This procedure was repeated for all groups and the prediction error was evaluated as the average of the mean absolute error (MAE) for each group.

3. Results and discussion

3.1 Characterization of CMs

The characteristics of the three types of MC (MC1, MC2, and MC3) and the three types of PAC (S-PAC, C-PAC, and W-PAC) are presented in Tables S1 and S2, respectively. All of the CMs consist mainly of oxygen and carbon (Table S1). However, several characteristics of the CMs vary. The Brunauer-Emmett-Teller (BET) surface area values are high for MC3 (1488 m^2/g), W-PAC (1222 m^2/g), and C-PAC (1216 m^2/g), which are nearly double the values for MC2 and S-PAC. The BET surface area of MC1 is only 256 m^2/g. The total pore volume decreases in the following order: MC3 > W-PAC > MC2 > MC1 > C-PAC > S-PAC. However, C-PAC has the highest micropore volume (0.45 cm^3/g) and lowest average pore diameter (2.84

nm) among the selected CMs. The mesoporosities of the MCs all exceed 81%, followed by W-PAC (63.5%), while C-PAC has the lowest value (16.9%). Excluding MC1, the average hydrodynamic diameters of the CMs are similar and in the range of 2.36 – 4.54 μm. The zeta potential of the CMs is negatively charged in DW at pH 5.7 – 6.5.

3.2 GSM and MIB adsorption: Isotherm and kinetic model analysis

The GSM and MIB equilibrium isotherm model analyses are presented in Fig. 1. For GSM adsorption, the MC1, MC2, S-PAC, and W-PAC equilibrium data are relatively well described by the Freundlich model (Table S4), which can be attributed to multilayer adsorption. However, the Langmuir model fits the MC3 and C-PAC equilibrium data better, indicating that single-layer adsorption may occur during GSM adsorption (Table S4). The maximum adsorption capacity (q_m) for GSM (initial conc. = 10 μg/L) was calculated using the Langmuir model and the highest value is 1485.06 μg/g for C-PAC, followed by S-PAC (1216.82 μg/g) and W-PAC (1032.42 μg/g). The q_m values for GSM decrease in the following order: C-PAC > S-PAC > W-PAC > MC3 > MC1 > MC2. For MIB adsorption, excluding MC3, most of the equilibrium data agree well with the Freundlich model (Table S5). The MC3 equilibrium data are better fitted by the Langmuir model (Table S5). Similarly, the highest maximum adsorption capacity (q_m) for MIB (initial concentration of MIB = 10 μg/L) was calculated using the Langmuir model as 1372.76 μg/g for C-PAC. Excluding MC3, the equilibrium data for MIB are well fitted by the Freundlich model. The Temkin model exhibits the worst fit for all of the GSM and MIB equilibrium data. MIB is less effectively removed by CMs compared to GSM, which has been reported in previous studies (Cook et al., 2001; Bertone et al., 2018; Yu et al., 2007; Zoschke et al. 2011). This phenomenon occurs because hydrophobic interactions can be enhanced between hydrophobic CMs and GSM because GSM (log K_{ow} = 3.70) is more

hydrophobic than MIB (log K_{ow} = 3.13) (Hafuka et al., 2019), although both GSM and MIB are hydrophobic compounds. Similarly, Zoschke et al. (2011) and Cook et al. (2001) found that PAC achieved a higher removal efficiency for GSM than MIB based on the flatter and more hydrophobic structure of GSM, which is favorable for adsorption in the slit-shaped pores of PAC.

The kinetic model parameters for GSM and MIB adsorption on the CMs are presented in Tables S6 and S7, respectively. Fig. 2 presents the experimental data. The kinetic experimental data for GSM and MIB exhibit trends similar to those observed in our model analysis. The Elovich model fits the GSM and MIB data for MC1 and MC2 well, whereas the pseudo-second-order model effectively represents those for MC3 and W-PAC. The pseudo-first-order model accurately describes the data obtained using S-PAC and C-PAC. The PACs exhibit faster adsorption of GSM and MIB compared to the MCs. GSM and MIB adsorption on C-PAC are the fastest ($k_{1,\ GSM}$ = 0.607/min, $k_{1,\ MIB}$ = 0.596/min). Most of the GSM and MIB adsorption using C-PAC was completed within 15 and 10 min, respectively. As shown in Fig. S1, among the characteristics of the CMs, the adsorption affinity of GSM and MIB are only correlated with three factors: micropore volume (R^2 = 0.60 (GSM) and 0.67 (MIB)), mesopore volume (R^2 = 0.56 (GSM) and 0.52 (MIB)), and average pore diameter (R^2 = 0.50 (GSM) and 0.31 (MIB)). This indicates that a high micropore volume, low mesopore volume, and small average pore diameter of CM are beneficial for GSM and MIB adsorption. Yu et al. (2007) determined that micropore volume is the main factor affecting MIB and GSM adsorption when using PACs. The adsorption capacities for GSM and MIB were higher for the PACs than for the MCs because a dominant mesopore structure can hinder the adsorption of small molecules (less than 1 nm) (Newcombe, 2006) and even allow them to penetrate through pores.

Overall, the best GSM and MIB adsorption capacities were achieved using C-PAC, which

has the highest micropore volume (0.45 cm^3/g) and lowest mesopore volume (0.11 cm^3/g) among the CMs. C-PAC also has a small and narrow pore size distribution, providing optimal conditions for GSM and MIB adsorption. The second-highest adsorption capacity was observed for S-PAC, which has an average pore diameter (2.89 nm) similar to that of C-PAC (2.84 nm). Previous studies have indicated that C-PAC has a higher adsorption affinity for MIB because it has a more hydrophobic surface (lower oxygen content) than W-PAC. Additionally, Greenwald et al. (2015) reported that pore volumes in the range of 2.3 – 3.2 nm are highly correlated with MIB adsorption on coconut-based GAC. Therefore, GSM and MIB adsorption can be significantly enhanced by using C-PAC with a large micropore volume in DW.

3.3 GSM and MIB adsorption on CMs in a real drinking water source

Various water quality parameters such as the presence of NOM can decrease the adsorption capacity of CMs for micropollutant removal (Newcombe et al., 2002; Park et al., 2018). Our samples of Nak-Dong River water contained 2.56 to 3.04 mg/L of DOC with similar alkalinity (65 to 70 mg/L $CaCO_3$) and a pH range of 7.35 – 7.96 (Table 1). Although the GSM and MIB removal efficiencies of the CMs were lower in the Nak-Dong River water than in DW, C-PAC still achieved the highest removal efficiency for GSM with values of 42.0% and 62.5% after 20 min and 120 min of contact time, respectively (Fig. 3). Although it was removed less effectively than the GSM, the MIB exhibited higher adsorption efficiency (approximately 41.7%) when using C-PAC and W-PAC within 15 min of contact time compared to the other CMs. Similarly, Yu et al. (2007) reported that fruit-shell-based PAC exhibited superior adsorption capacity compared to S-PAC and W-PAC when using raw water in a DWTP. Additionally, C-PAC is the most cost-effective adsorbent for the removal of GSM and MIB because it is the cheapest among the six types of CMs. The MCs still exhibited worse

performance than the PACs because the MCs only provide sufficient adsorption sites for large NOM molecules and are unfavorable for small molecules such as GSM and MIB. Therefore, the MCs removed GSM and MIB less effectively than the PACs in both DW and river water, and C-PAC was the most efficient adsorbent for removing GSM and MIB from the river water. In Table 2, the removal efficiencies of GSM and MIB are compared to those reported previously for various drinking water sources. Most studies have evaluated the removal efficiency of GSM or MIB with a long contact time (3 – 7 days) with a PAC dose of 10 to 30 mg/L. Compared to those results, the C-PAC considered in this study exhibited higher removal efficiencies for both GSM and MIB (>80%) with a very short contact time (20 min).

Fig. 4a and 4b present the effects of NOM concentration on the Nak-Dong River water (No. 4). Generally, the presence of NOM inhibits adsorption on PAC as a result of pore blockage induced by large NOM molecules (humic substances and building blocks) and direct adsorption site competition caused by low molecular weight (LMW) NOM molecules (Zoschke et al., 2011; Bertone et al., 2018; Newcombe et al., 2002). However, the removal efficiencies of GSM and MIB gradually enhanced by increasing the DOC concentration and contact time in this study. There are two possible explanations. First, GSM and MIB could reach adsorption sites before pore blockage by large NOM molecules occurs based on their small size (0.6 to 0.8 nm). Second, GSM and MIB could continually adsorb onto additional shrunken pores because some large pores may be reduced to the micropore range (<2 nm) through the external surface adsorption of NOM.

LC-OCD analysis indicated that the NOM in the Nak-Dong River water (No. 4) consisted of 18.8% LMW NOM (< 350 g/mol), 44.1% humic substances, and 17.4% building blocks. The Suwannee River NOM contains over 80% humic substances and building blocks, and 17% LMW NOM. Therefore, pore blockage by large NOM molecules is the dominant phenomenon

and only a small portion of the NOM competes with the adsorption of GSM and MIB. Furthermore, MIB is more sensitive to an increasing DOC concentration than GSM, which has a flatter structure that can more easily enter shrinking pores. Matsui et al. (2012) found that NOM in natural water mainly adsorbs close to external PAC surfaces and that the competitive adsorption between NOM (MW of 180 Da) and MIB is insignificant (0.2 to 2 % of total NOM), despite the increased NOM uptake in natural surface water.

The GSM and MIB concentrations both met the guidelines for drinking water in South Korea (20 ng/L) after 20 min of contact time (Fig. 4c and 4d) when each of their initial concentration was100 ng/L (No. 4). However, more than 1 h was required to meet the guidelines when GSM and MIB were present at a higher concentration of 200 ng/L. To predict the required dose of C-PAC to meet the guidelines for GSM and MIB in a specific drinking water source, the Nak-Dong River, all 72 datasets on Nak-Dong River water with various water quality parameters and initial GSM and MIB concentrations were applied to the MLR and DNN modeling discussed below.

3.4 Setting up a model equation for predicting the dose of C-PAC and contact time using MLR analysis

Our experiments included eight variables: DOC concentration, turbidity (T_{NTU}), color, pH, alkalinity, dose of C-PAC (C_d), contact time (t), and initial GSM and MIB concentration (C_0). Two variables, namely contact time and initial GSM and MIB concentrations, must be used for modeling. Therefore, the other six variables were considered as additional variables for varying the number of input variables used for modeling. The number of input variables for MLR was varied from three to five to derive an optimal predictive equation. Additionally, input variable

combinations were generated according to different numbers of input variables. Leave-one-out cross-validation was performed for all combinations of input variables to quantify the predictive performance of MLR on independent data. Leave-one-out cross-validation is a form of K-fold cross validation taken to its logical extreme with K being equal to the number of data. This is an exhaustive technique for estimating the prediction error of a model by subtracting samples sequentially from a dataset (Stone 1974).

Tables S8 and S9 reveal that a mixed model always yields the best performance, regardless of the number of input variables. Fig. 5a and 5b compare the observed and predicted GSM and MIB concentrations for the optimal models that exhibited the best performance according to the number of input variables. Although the prediction accuracy shows no significant differences depending on the number of input variables, the lowest MAE can be observed when five input variables are selected for both GSM and MIB. The final regression equations were established using the five selected variable forms and all experimental data were used to calculate the regression coefficient. The functional forms of the final predictive models are expressed as

$$\text{GSM}: \quad C_{t,GSM} = t^{-0.45} e^{\left(9.245+0.16\sqrt{C_{o,MIB}}-0.94\left(\frac{1}{T_{NTU}}\right)-4.3\sqrt{T_{NTU}}-0.00107C_d^2\right)} \tag{6}$$

$$\text{MIB}: \quad C_{t,MIB} = {C_{o,MIB}}^{0.808} {T_{NTU}}^{-3.71} e^{\left(2.72-0.0087t-1.928\left(\frac{1}{T_{NTU}}\right)-0.00096C_d^2\right)} \tag{7}$$

where the units of C_0 and C_d are ng/L, and t and T_{NTU} correspond to minutes and nephelometric turbidity units (NTU), respectively. If the input variables used for MLR have high correlation, then multicollinearity problems that lead to reliability issues during regression coefficient estimation may occur. Therefore, it is necessary to evaluate multicollinearity, which can be quantified by a VIF. The VIF is close to one if the independent variables are not correlated with any other independent variables. In general, a VIF greater than 10 indicates strong correlation

and is a cause for concern. A more conservative level of five is considered to be ideal (Montgomery & Runger, 2010). Table 3 summarizes the statistical properties and VIF values for each parameter. Both MLR models have variables with low VIFs of two or less, except for those derived from NTU. A high VIF is sometimes acceptable when the variables with high VIFs are derived from the same variable in the MLR model. Therefore, the estimates of the fitted coefficients for both MLR models are considered to be appropriate.

Fig. 5c and 5d compare the observed and predicted final GSM and MIB concentrations based on the MLR models (Eqs. 3 and 6). For the final MLR models, the MAE, bias, and coefficient of variation in bias were 6.4 ng/L, 1.00, and 18.6% for GSM and 5.8 ng/L, 1.01, and 15.7 % for MIB, respectively. The predictions of both MLR models agree will with the observed results and the MLR model for MIB provides slightly better accuracy than that for GSM. Sensitivity analysis was performed to identify the effects of the input variables on the final GSM and MIB concentrations. Spearman rank correlations were calculated from global sensitivity simulation results. The tornado diagram in Fig. 6 represents the effects of each input variable on the final GSM and MIB concentrations (C_t), where C_0 is the most influential variable, regardless of GSM or MIB. Both MLR models exhibit strong correlations in the order of $C_d > t > T_{NTU}$. The sensitivity of t in the MLR model for GSM is 1.8 times higher than that for MIB, meaning t has a greater impact on the change in final GSM concentration than on the change in the final MIB concentration. The rank correlation for T_{NTU} exhibits similar values, but different signs for the GSM and MIB final concentrations. However, these values represent relatively weak correlations compared to the other variables and have little effect on the concentration changes.

3.5 Predicting final GSM and MIB concentrations using DNN analysis

Typically, the prediction accuracy of a DNN can be improved by increasing the number of hidden layers or training epochs. However, in this study, this approach did not improve the prediction accuracy for the testing data, which can be attributed to overfitting. Therefore, establishing an optimal structure that considers the accuracy of independent data is crucial for robust modeling. The trends of predictive performance for various learning conditions were similar for GSM and MIB, although the effects of the input variables on the changes in concentration were different. Compared to the use of one hidden layer, the use of two hidden layers yielded a significant error reduction and the use of three hidden layers yielded a slight additional error reduction, which was insignificant compared to the use of two hidden layers.

Regarding the number of nodes, the predictive performance improved as the number of nodes increased, but the result with 30 nodes and 40 nodes did not differ significantly. Although 40 nodes generally yielded a lower MAE, 30 nodes yielded slightly improved predictive performance. Regarding the number of epochs, the error significantly decreased as the epochs increased initially. However, the error often increased above a certain epoch based on the effects of overtraining. Specifically, the error for the training dataset decreased as the epochs increased, but the error for the testing dataset increased. As a result, the trained DNN model with hidden layers = 3, nodes = 40, batch size = 20, and epochs = 10,000 provided the best predictive performance for GSM, while that with hidden layers = 3, nodes = 40, batch size = 30, and epochs = 16,000 provided the best predictive performance for MIB (Table S10). The final DNN models were established using all the data under the optimal learning conditions described above. Fig. 7 compares the observed and predicted final GSM and MIB concentrations for cross-validation and the final models using all data in DNNs with the optimal structures. For the final models using all data, the predicted concentrations agree well with the measured concentrations. The predictive performance of the deep learning models

improved when all of the data were used for the DNNs. Specifically, the MAE decreased by 59% and 68% for GSM and MIB, respectively. Table 4 summarizes the statistical results of the performance of the predictive models using MLR and DNN.

When comparing the results of the two modeling approaches, the DNN model cannot identify the extent to which input variables affect outputs because it is a black-box model, whereas the MLR model is able to understand the relationships between input variables and outputs with acceptable prediction errors. However, the DNN model yields better agreement between the predicted and measured values with higher determination coefficients and a lower MAE. This may be because a DNN model is developed without assuming a specific fitting function in advance and can effectively model nonlinear systems. Additionally, a DNN model becomes more robust if additional data are accumulated.

Conclusion

A series of experiments were conducted to select the best-performing CM among three types of PACs and three types of MCs. The PACs adsorbed GSM and MIB better than the MCs and C-PAC exhibited the highest adsorption capacity, which was attributed to its high micropore volume and small, narrow pore size distribution. The relatively high concentrations of DOC and turbidity in river water did not hinder the adsorption of GSM and MIB, but actually enhanced it, which was attributed to rapid adsorption characteristics and shrinkage of pores. The developed MLR and DNN models predicted the required dose of C-PAC and contact time to bring GSM and MIB concentrations below the drinking water guidelines for various water quality parameters and initial GSM and MIB concentrations in Nak-Dong River water, which is one of the main drinking water sources in South Korea. The developed models can be used

by DWTP operators for the rapid assessment of additional required doses of C-PAC based on sudden events related to GSM and MIB. Therefore, these predictive models could reduce the cost and energy consumption incurred by the overuse of C-PAC and minimizing sludge production. Additionally, the developed models and equations can be used as alternatives to large-scale experiments to derive optimal processes for DWTPs, which can save significant resources and energy. This study broadened the knowledge of using C-PAC to treat GSM and MIB in DWTPs and provided a basis for the development of prediction models to be used by operators of DWTPs to adjust operating conditions and adapt to changing conditions, although the developed prediction models are site specific (Nak-Dong River water). Caution should be taken when these models are applied to new conditions that lie outside the bounds of the database used for modeling in this study. In future studies, the proposed models could be applied to the entire DWTP process at the same site for verification or extended to additional sites.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) Grants funded by the Korea government (MEST) (NRF-2014M3C8A4034282) and (MSIT) (2019R1C1C1010053).

References

Ahmad, M.B., Soomro, U., Muqeet, M., Ahmed, Z., 2021. Adsorption of indigo carmine dye onto the surface-modified adsorbent prepared from municipal waste and simulation using deep neural network. J. Hazard. Mater. 408, 124433. https://doi.org/10.1016/j.jhazmat.2020.124433.

Antonopoulou, M., Evgenidou, E., Lambropoulou, D., Konstoantinou, I., 2014. A review on advanced oxidation processes for the removal of taste and odor compounds from aqueous media. Water Res. 53, 215-234. https://doi.org/10.1016/j.watres.2014.01.028.

Bassaganya-Riera, J., 2015. Computational Immunology: Models and Tools, Elsevier, Academic Press.

Bertone, E., Chang, C., Thiel, P., O'Halloran, K., 2018. Analysis and modeling of powdered activated carbon dosing for taste and odour removal. Water Res. 139, 321-328. https://doi.org/10.1016/j.watres.2018.04.023.

Butakova, E.A., 2013. Specific features of odor-causing compounds (geosmin and 2-methylisoborneol) as secondary metabolites of cyanobacteria. Russ. J. Plant Physiol. 60, 507-510. https://doi.org/10.1134/S102144371304002X.

Bruce, D., Westerhoff, P., Brawley-Chesworth, A., 2002. Removal of 2-methylisoborneol and geosmin in surface water treatment plants in Arizona. J. Water Supply Res. T. 51(4), 183-197. https://doi.org/10.2166/aqua.2002.0016**.**

Cook, D., Newcombe, G., Sztajnbok, P., 2001. The application of powdered activated carbon for MIB and geosmin removal: predicting PAC doses in four raw water. Water Res. 35(5) 1325-1333. https://doi.org/10.1016/S0043-1354(00)00363-8.

Desai, K.M., Survase, S.A., Saudagar, P.S., Lele, S.S., Singhal, R.S., 2008. Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: Case study of fermentative production of scleroglucan. Biochem. Eng. J. 41(3), 266-273. https://doi.org/10.1016/j.bej.2008.05.009.

Greenwald, M.J., Redding, A.M., Cannon, F.S., 2015. A rapid kinetic dye test to predict the adsorption of 2-methylisoborneol onto granular activated carbons and to identify the influence of pore volume distributions. Water Res. 68, 784-792. https://doi.org/10.1016/j.watres.2014.10.022.

Ha, K., Jang, M.H., Joo, G.J., 2002. Spatial and temporal dynamics of phytoplankton communities along a regulated river system, the Nakdong River, Korea. Hydrobiologia. 470, 235-245. https://doi.org/10.1023/A:1015610900467.

Hafuka, A., Nagasato, T., Yamamura, H., 2019. Application of grapheme oxide for adsorption removal of geosmin and 2-methylisoborneol in the presence of natural organic matter. Int. J. Environ. Res. Public Health. 16(11), 1907. https://doi.org/10.3390/ijerph16111907.

He, Q., Zhong, L., Wang, H., Zou, Z., Chen, D., Yang, K., 2016. Odor removal by powdered activated catbon (PAC) in low turbidity drinking water. Water Sci. Tech-W Sup. 16 (4), 1017-1023. https://doi.org/10.2166/ws.2016.017.

He, X., Liu, Y.L., Conklin, A., Westrick, J., Weavers, L.K., Dionysiou, D.D., Lenhart, J.J., Mouser, P.J., Szlag, D., Walker, H.W., 2016. Toxic cyanobacteria and drinking water: impacts, detection, and treatment. Harmful Algae. 54, 174-193. https://doi.org/10.1016/j.hal.2016.01.001.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J., 2002. Statistical Methods in Water Resources. USGS, TWRI-4-A3. https://doi.org/10.3133/tm4A3.

Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18(7), 1527-1554. https://doi.org/10.1162/neco.2006.18.7.1527.

Im, S.J., Shon, H.K., Kim. I.S., Jang. A., 2019. The effects of naturally occurring operation factors on the removal mechanism of major algae metabolized materials in forward osmosis process. J. Clean. Prod. 239, 118009. https://doi.org/10.1016/j.jclepro.2019.118009.

Huang, Y., Nie, Z., Yuan, J., Murray, A., Li, Y., Woods-Chabane, G., Hofmann, R., 2019. Experimental validation of a test to estimate the remaining adsorption capacity of granular activated carbon for taste and odour compounds. Environ. Sci.; Water Res. Technol. 5(3), 609-617. https://doi.org/10.1039/C8EW00600H.

Mamandipoor, B., Majd, M., Sheikhalishahi, S., Modena, C., Osmani,V., 2020. Monitoring and detecting faults in wastewater treatment plants using deep learning, Environ. Monit. Assess. 192(2), 148. https://doi.org/10.1007/s10661-020-8064-1.

Matsui, Y., Yoshida, T., Nakao, S., Knappe, D.R.U., Matsushita, T., 2012. Characteristics of competitive adsorption between 2-methylisoborneol and natural organic matter on superfine and conventionally sized powdered activated carbons. Water Res. 46(15), 4741-4749. https://doi.org/10.1016/j.watres.2012.06.002.

Matsui, Y., Nakao, S., Sakamoto, A., Taniguchi, T., Zamyadi, L., Pan, L., Matsushita, T., Shirasaki, N., 2015. Adsorption capacities of activated carbons for geosmin and 2-methylisoborneol vary with activated carbon particle size: effect of adsorbent and adsorbate characteristics. Water Res. 85, 95-102. https://doi.org/10.1016/j.watres.2015.08.017.

Ministry of Environment (MOE), 2015. Water Treatment Plant Algae Control Guideline. South

Korea.

Ministry of Health (NOH), 2015. Labor and Welfare in Japan.

Ministry of Health of the People's Republic of China (MOH), Standardization Administration of The People's Republic of China (SAC), 2006. Standards for Drinking Water Quality.

Montgomery, D.C., Runger, G.C., 2010. Applied statistics and probability for engineers. 5th Ed., Wiley, New York.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning, Haifa, Israel, 807-814.

National Health and Medical Research Council of Australia, 2016. Australian drinking water guidelines 6 2011-version 3.3.

Newcombe, G., Drikas, M., Hayes, R., 1997. Influence of characterized natural organic material on activated carbon adsorption: II. Effect on pore volume distribution and adsorption of 2-methylisoborneol. Water res. 31(5), 1065-1073. https://doi.org/10.1016/S0043-1354(96)00325-9.

Newcombe, G., Morrison, J., Hepplewhite, C., Knappe, D.R.U., 2002. Simultaneous adsorption of MIB and NOM onto activated carbon: II. Competitive effects. Carbon 40, 2147-2156. https://doi.org/10.1016/S0008-6223(02)00098-2.

Newcombe, G., Dixon, D., 2006. Chapter 8: Removal of natural organic material and algal metabolites using activated carbon. Interface Science in Drinking Water Treatment. Academic Press, Messachusetts, USA. https://doi.org/10.1016/S1573-4285(06)80077-3.

Nowack, K.O., Cannon, F.S., Mazyck, D.W., 2004. Enhancing activated carbon adsorption of

2-methylisoborneol: methane and steam treatments. Environ. Sci. Technol. 38(1), 276-284. https://doi.org/10.1021/es026397j.

Özdemir, U., Özbay, B., Veli, S., Zor, S., 2011. Modeling adsorption of sodium dodecyl benzene sulfonate (SDBS) onto polyaniline (PANI) by using multi linear regression and artificial neural networks, Chem. Eng. J. 178, 183-190. https://doi.org/10.1016/j.cej.2011.10.046.

Pan, L., Matsui, Y., Matsushita, T., Shirasaki, N., 2016. Superiority of we-milled over dry-milled superfine powdered activated carbon for adsorptive 2-methylisoborneol removal. Water Res. 102, 516-523. https://doi.org/10.1016/j.watres.2016.06.062.

Park, J.A. Jung, S.M., Yi, I.G., Choi, J.W., Kim, S.B., Lee, S.H., 2017a. Adsorption of microcystin-LR on mesoporous carbons and its potential use in drinking water source. Chemosphere 177, 15-23. https://doi.org/10.1016/j.chemosphere.2017.02.150.

Park, J.A. Nam, H.L., Chio, J.W., Ha, J.S., Lee, S.H., 2017b. Oxidation of geosmin and 2-methylisoborneol by the photo-Fenton process: Kinetics, degradation intermediates, and the removal of microcystin-LR and trihalomethane from Nak-Dong River water, South Korea. Chem. Eng. J. 313, 345-354. https://doi.org/10.1016/j.cej.2016.12.086.

Park, J.A., Jung, S.M., Choi, J.W., Kim, J.H., Hong, S., Lee, S.H., 2018. Mesoporous carbon for efficient removal of microcystin-LR in drinking water sources, Nak-Dong River, South Korea: Application to a field-scale drinking water treatment plant. Chemosphere 193, 883-891. https://doi.org/10.1016/j.chemosphere.2017.11.092.

Pendleton, P., Wong, S.H., Schumann, R., Levay, G., Denoyel, R., Rouquerol, J., 1997. Properties of activated carbon controlling 2-methylisoborneol adsorption. Carbon 35(8), 1141-1149. https://doi.org/10.1016/S0008-6223(97)00086-9.

Qian, F., Dixon, D.R., Newcombe, G., Ho, L., Dreyfus, J., Scales, P.J., 2014. The effect of pH on the release of metabolites by cyanobacteria in conventional water treatment processes, Harmful Algae 39, 253-258. https://doi.org/10.1016/j.hal.2014.08.006.

Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82(12), 2635-2670. https://doi.org/10.2166/wst.2020.369**.**

Srinivasan, R., Sorial, G.A., 2009. Adsorption of geosmin and MIB on activated carbon fibers-single and binary solute system. Water Air Soil Pollut: focus. 9, 223-235. https://doi.org/10.1007/s11267-009-9217-y.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society, Series B. 36(2), 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x.

Thiel, P., Cullum, P., 2007. Evaluating the performance of different powdered activated carbons for taste and odour reduction, 32nd Annual Old Water Industry operations Workshop, 39-44.

Yu, J., Yang, M., Lin, T.F., Guo, Z., Zhang, Y., Gu, J., Zhang, S., 2007. Effects of surface characteristics of activated carbon on the adsorption of 2-methylisoborneol (MIB) and geosmin from natural water. Sep. Purif. Technol. 56(3), 363-370. https://doi.org/10.1016/j.seppur.2007.01.039.

Zakeri, H.R., Yousefi, M., Mohammadi, A.A., Baziar, M., Mojiri, S.A., Salehnia, S., Hosseinzadeh, A., 2021. Chemical coagulation-electro fenton as a superior combination process for treatment of dairy wastewater: performance and modelling. Int. J. Environ. Sci. Te.1-14. https://doi.org/10.1007/s13762-021-03149-w.

Zamyadi, A., Henderson, R., Stuetz, R., Hofmann, R., Ho, L., Newcombe, G., 2015. Fate of geosmin and 2-methylisoborneol in full-scale water treatment plants. Water Res. 83, 171-183. https://doi.org/10.1016/j.watres.2015.06.038.

Zoschke, K., Engel, C., Börnick, H., Worch, E., 2011. Adsorption of geosmin and 2-methylisoborneol onto powdered activated carbon at non-equilibrium conditions: influence of NOM and process modeling. Water Res. 45 (15), 4544-4550. https://doi.org/10.1016/j.watres.2011.06.006.

Table 1

Water quality parameters of sampled Nak-Dong River water.

River water Sample No	Sampling date	DOC (mg/L)	$SUVA_{254}$ (L/mg/m)	Turbidity (NTU)	Color	EC (μs/cm)	Alkalinity (mg/L $CaCO_3$)	pH
1	February 16, 2016	2.82	1.81	0.20	6	358	65	7.96
2	March 07, 2016	3.04	2.37	0.28	7	388	70	7.84
3	March 27, 2016	2.56	2.46	0.31	5	384	65	7.35
4	April 4, 2016	2.58	2.54	0.32	6	379	65	7.93
5	April 20, 2016	2.76	2.42	0.5	7	355	65	7.96

Table 2

Comparison of the removal efficiencies of GSM and MIB using PACs in drinking water sources.

Drinking water source	Initial conc. of GSM & MIB	Type of PAC	Dose of PAC	Contact time	q_m (or removal efficiency)	Ref.
Reservoir	GSM: 100 ng/L MIB: 100 ng/L	C-PAC	GSM: 10 mg/L MIB: 25 mg/L	3-5 day	93-95% 85-95%	Cook et al. (2001)
Distilled water with NOM (8 mg/L)	MIB: 327 & 880 ng/L	S-PAC	13 mg/L	5 days	95-99, 96-99%	Newcombe et al. (1997)
Natural water	①GSM: 100 ng/L ②MIB: 100 ng/L	S-PAC	2 – 30 mg/L	3-5 days	①24 – 26 ng/mg ②10 - 20 ng/mg	Yu et al. (2007)
		W-PAC			①37 ng/mg ②18 ng/mg	
		Fruit shell-based PAC			①58 ng/mg ②30 ng/mg	
Influent from WTP	GSM: 100 ng/L MIB: 100 ng/L	S-PAC	30 mg/L	Stirring for 40 min and settling for 30 min	95% 77%	Berton et al. (2018)
Lakes & river	MIB = 100 ng/L	W-PAC	8 mg/L	1 week	0.4 – 3 nmol/mg (67.31 – 504.84 ng/mg)	Matsui et al. (2012)
Natural water	①GSM ②MIB (concentration was	S-PAC	15 mg/L	30 min	①86% ②79%	Thiel & Cullum, (2007)
		W-PAC			①80%	

	not provided)				②62%	
		C-PAC			①50%	
					②48%	
River water	GSM = 100 ng/L MIB = 100 ng/L	C-PAC	32.14 mg/L	20 min	Both > 80%	In this study

Table 3

Summary of the multicollinearity analysis results

GSM				MIB			
Variable	Standard error	*t*-value	VIF	Variable	Standard error	*t*-value	VIF
Intercept	1.403	6.59	*NA*	Intercept	0.344	7.92	*NA*
$\ln(t)$	0.031	-14.60	1.0	t	0.001	-8.01	1.00
$\sqrt{C_0}$	0.010	15.28	1.0	$\ln(C_0)$	0.047	17.16	1.13
$1/\mathrm{T}_{NTU}$	0.218	-4.30	53.0	$1/\mathrm{T}_{NTU}$	0.270	-7.13	125.19
$\sqrt{\mathrm{T}_{NTU}}$	1.244	-3.45	49.3	$\ln(\mathrm{T}_{NTU})$	0.555	-6.68	119.51
C_d^2	0.000	-13.46	1.7	C_d^2	0.000	-14.25	1.86

Table 4

Performance of prediction methods concentration.

Prediction method		Data	MAE (ng/L)	R^2	Bias, λ	
					Mean	COV (%)
GSM	MLR	Cross validation	7.12	0.87	1.02	20.3
		All data	6.43	0.89	1.00	18.6
	DNN	Cross validation	4.09	0.96	1.00	16.7
		All data	1.67	0.99	1.01	9.2
MIB	MLR	Cross validation	6.34	0.95	1.02	17.2
		All data	5.80	0.96	1.01	15.7
	DNN	Cross validation	3.86	0.98	1.00	13.6
		All data	1.24	1.00	1.02	3.1

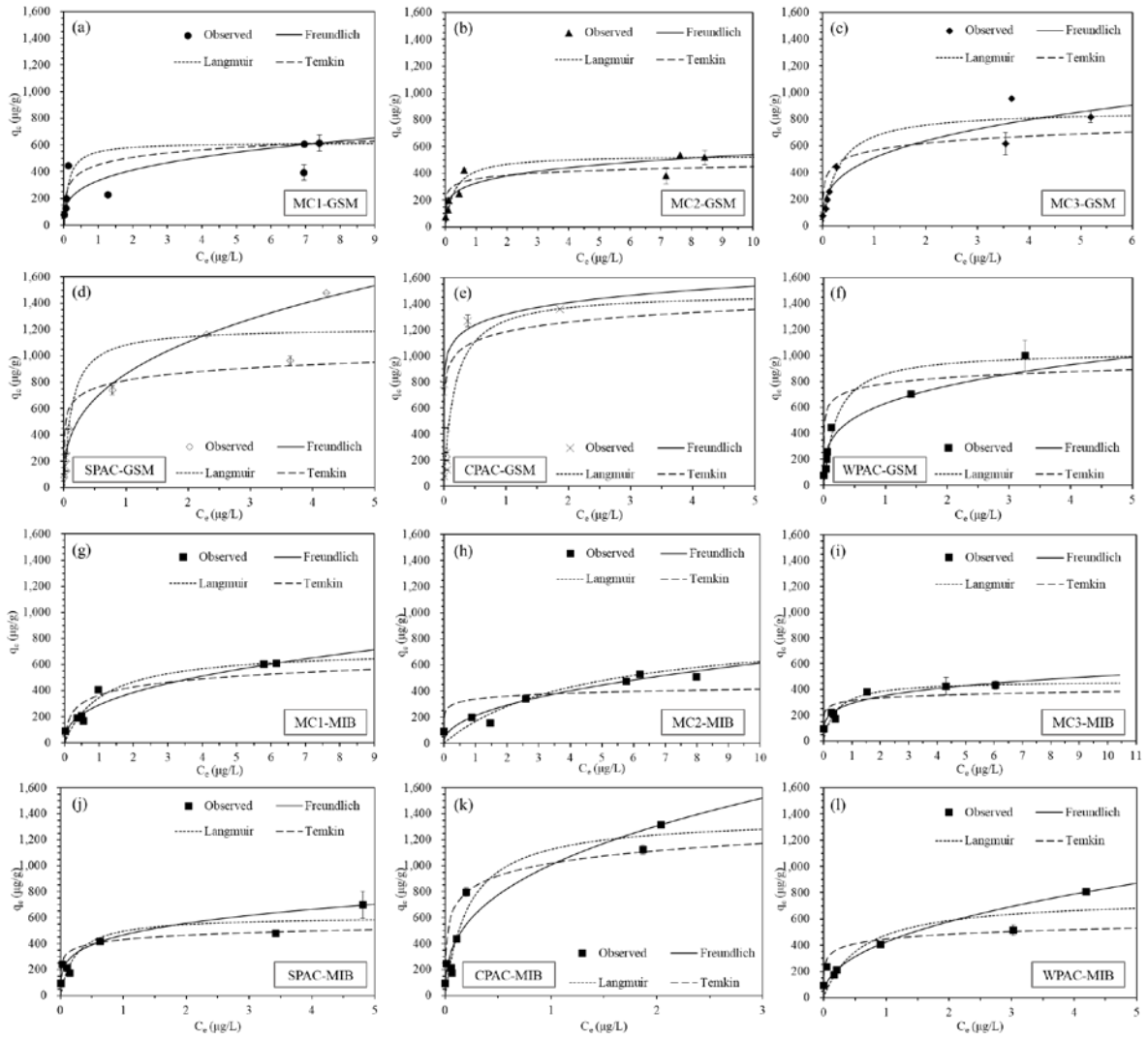

Figure 1. Equilibrium isotherm model analysis for comparison GSM (a-f) and MIB (g-l) adsorption capacity of the six CMs ($[GSM]_0 = [MIB]_0 = 10$ μg/L; dose = 5.7 – 168.1 mg/L; 24 h reaction). Error bars represent the standard deviation of the mean (n=5).

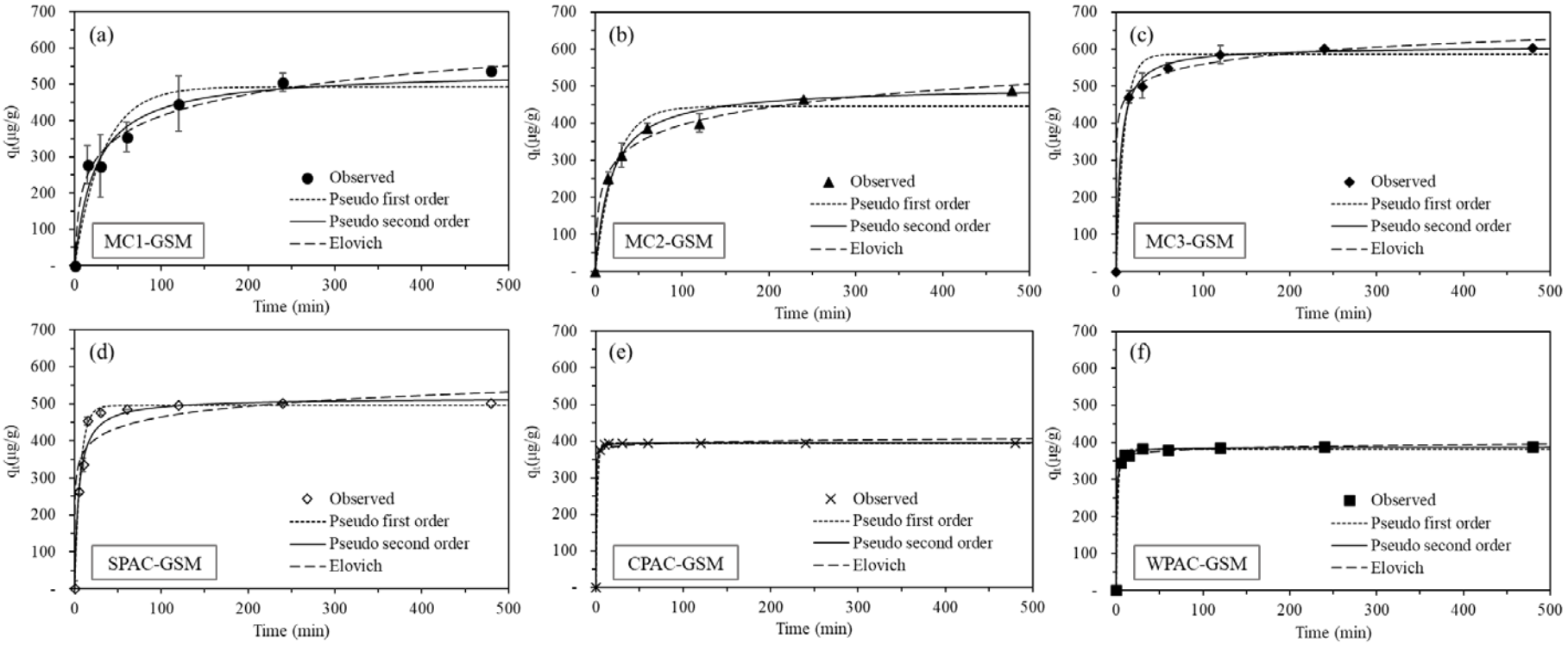

Figure 2. Kinetic model analysis for comparison GSM (a-f) and MIB (g-l) adsorption capacity of the six CMs ($[GSM]_0 = [MIB]_0 = 10$ μg/L; dose = 25.2 mg/L). Error bars represent the standard deviation of the mean (n=5)

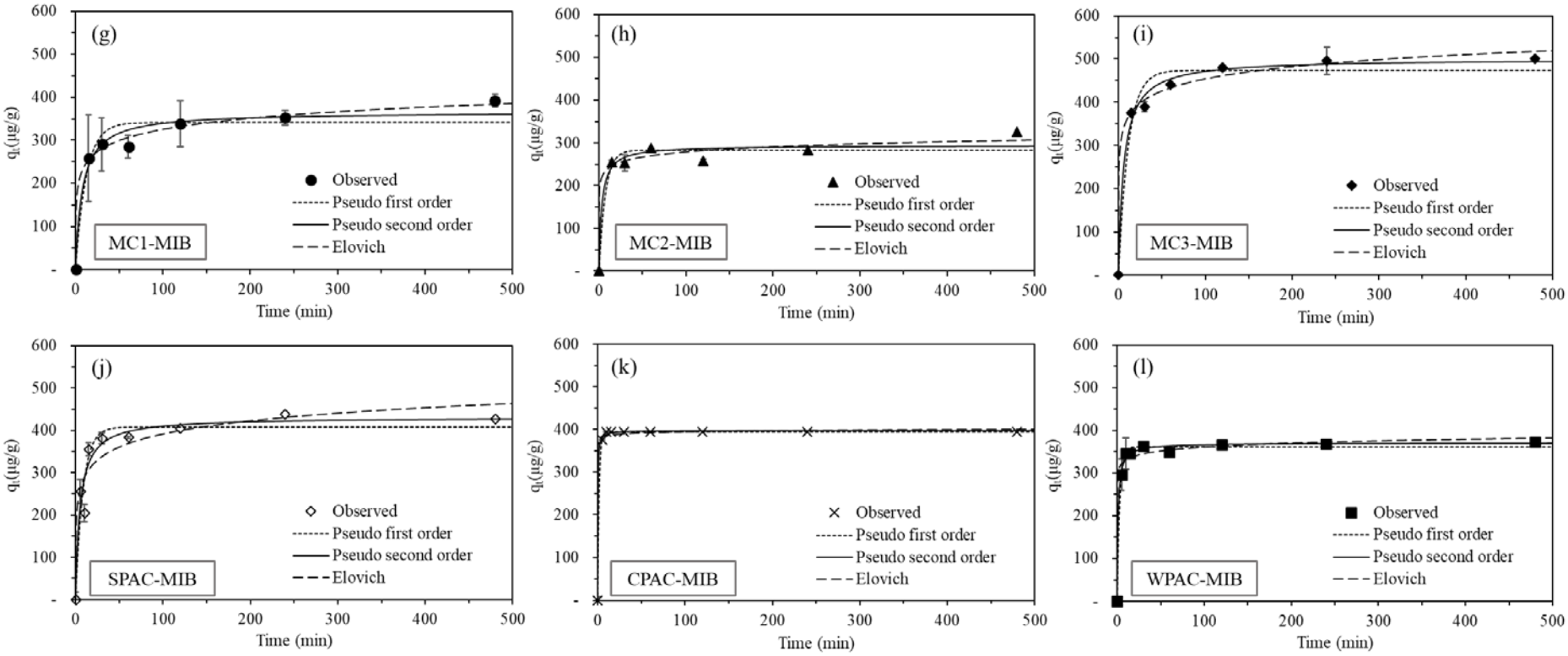

Figure 2. continued

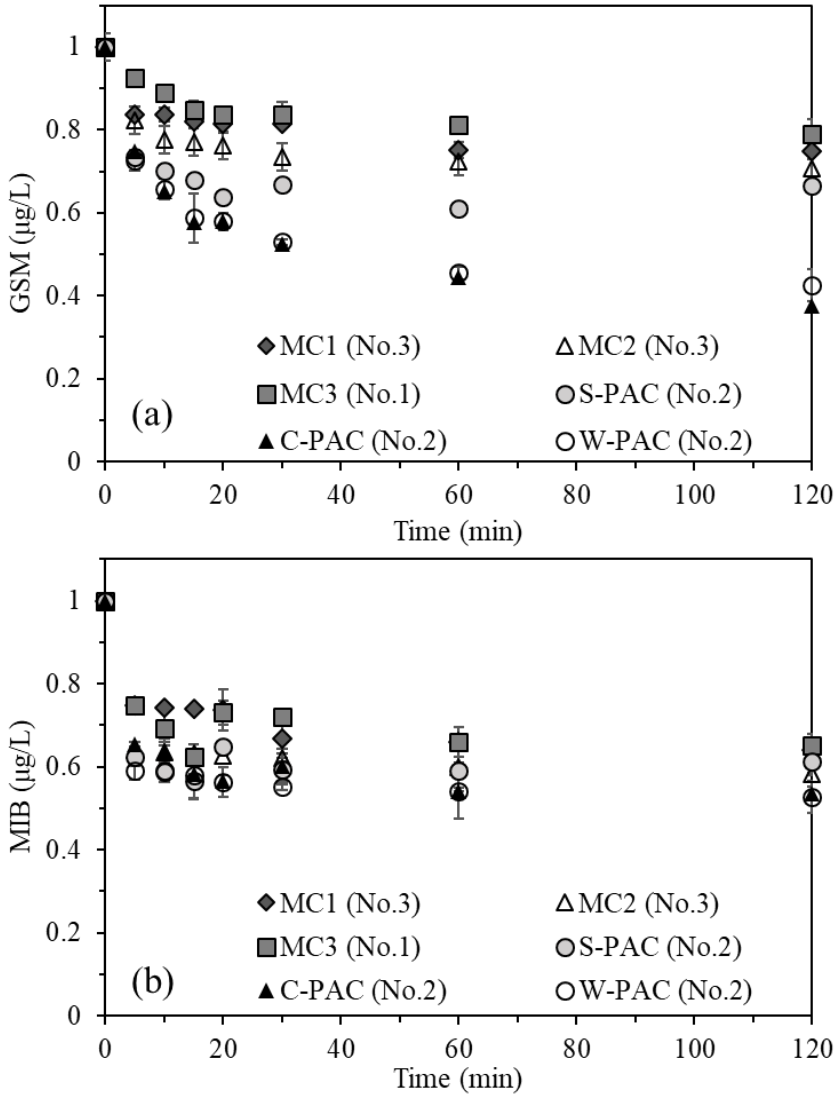

Figure 3. Comparison (a) GSM and (b) MIB adsorption efficiencies of the six CMs using Nak-Dong River water (No. 1 – 3) ($[GSM]_0 = [MIB]_0 = 10$ μg/L; dose = 25.2 mg/L). Error bars represent the standard deviation of the mean (n=5).

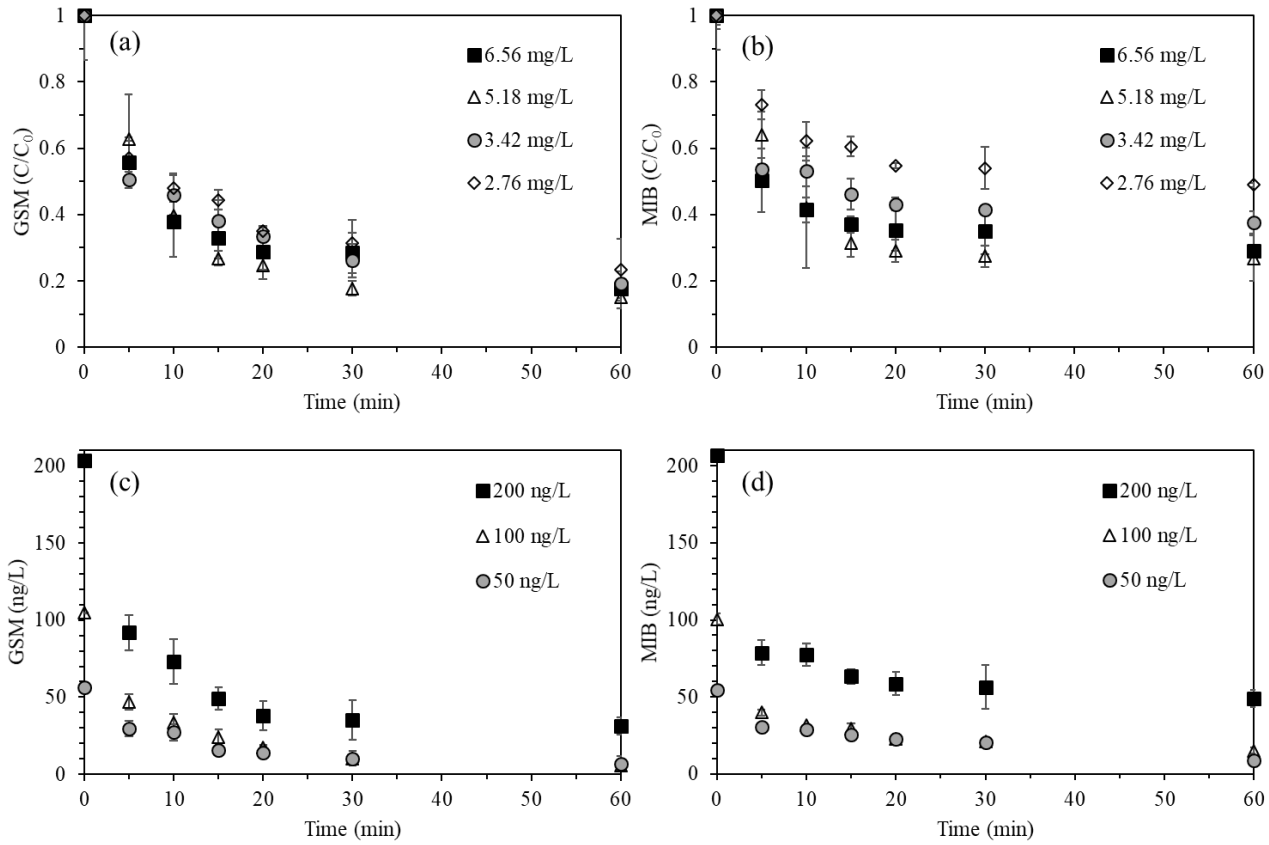

Figure 4. Effect of DOC concentration (2.76 – 6.56 mg/L) on (a) GSM and (b) MIB adsorption; and effect of initial concentration (50 – 200 ng/L) of (c) GSM and (d) MIB using C-PAC (32.14 mg/L). Error bars represent the standard deviation of the mean (n=5).

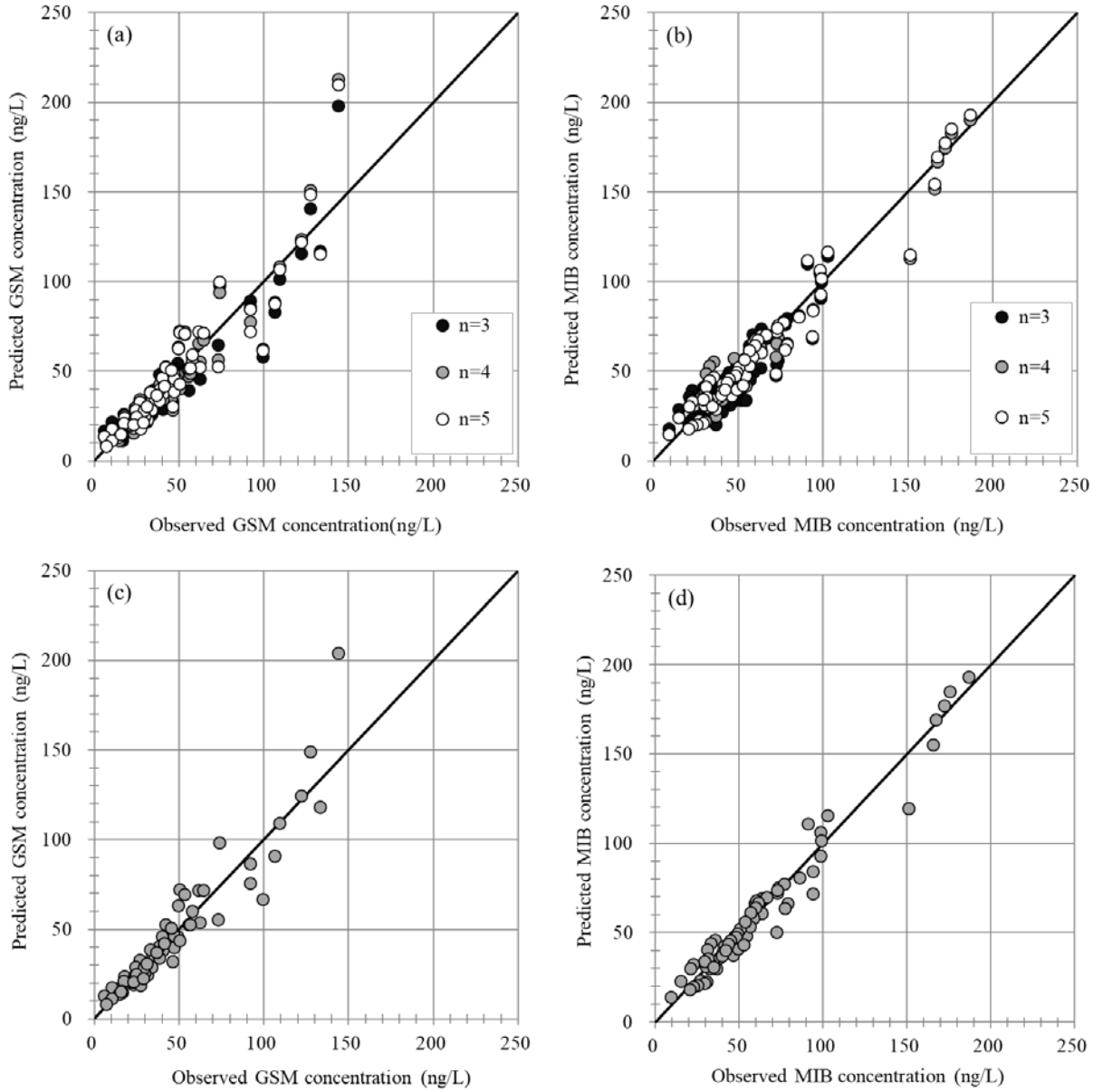

Figure 5. Comparison of predicted and observed (a) GSM and (b) MIB concentration by cross validation and (c) GSM and (d) MIB concentration by the proposed MLR equation (n is the number of variables)

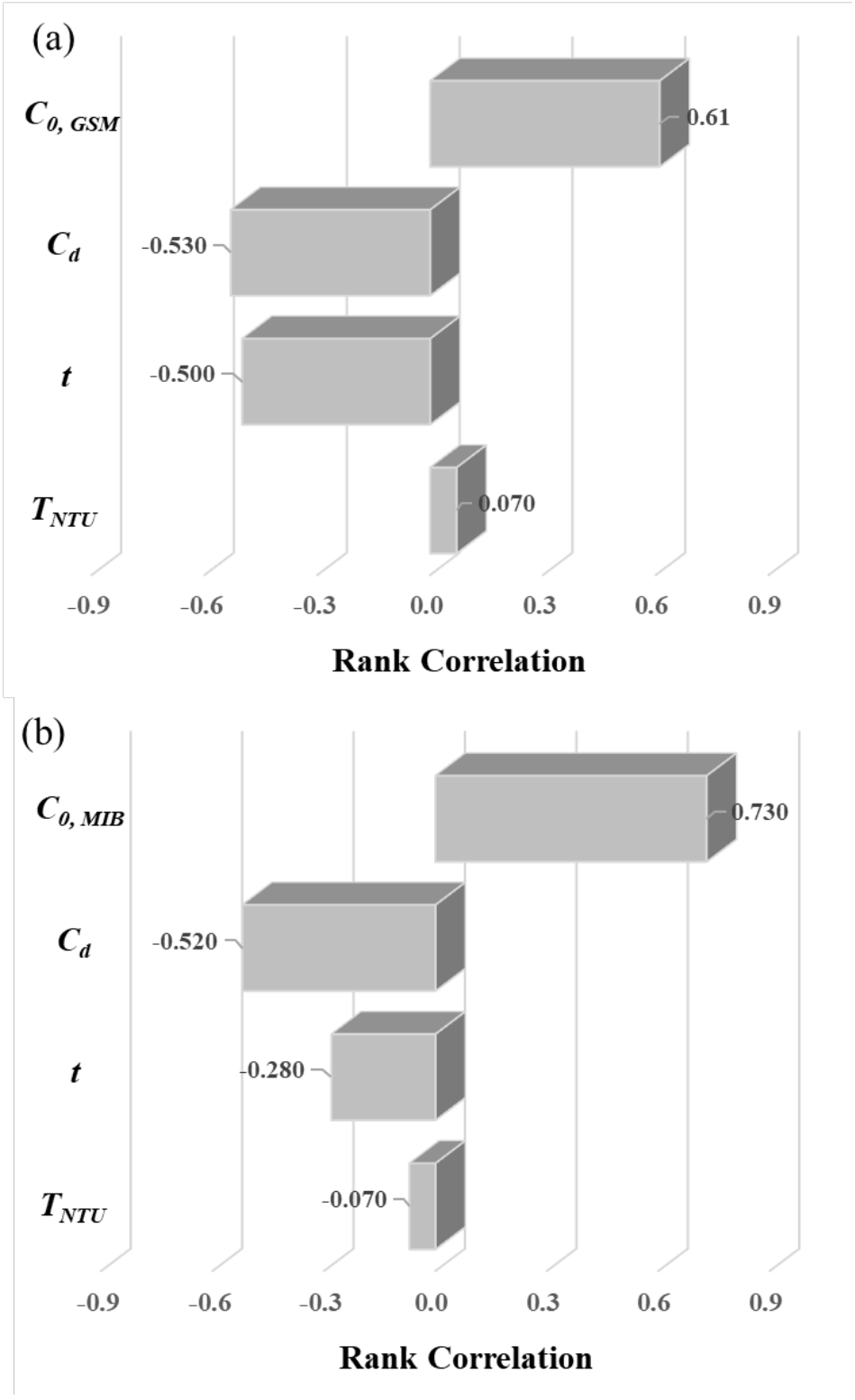

Figure 6. Tornado diagram for the input variables in the MLR models.

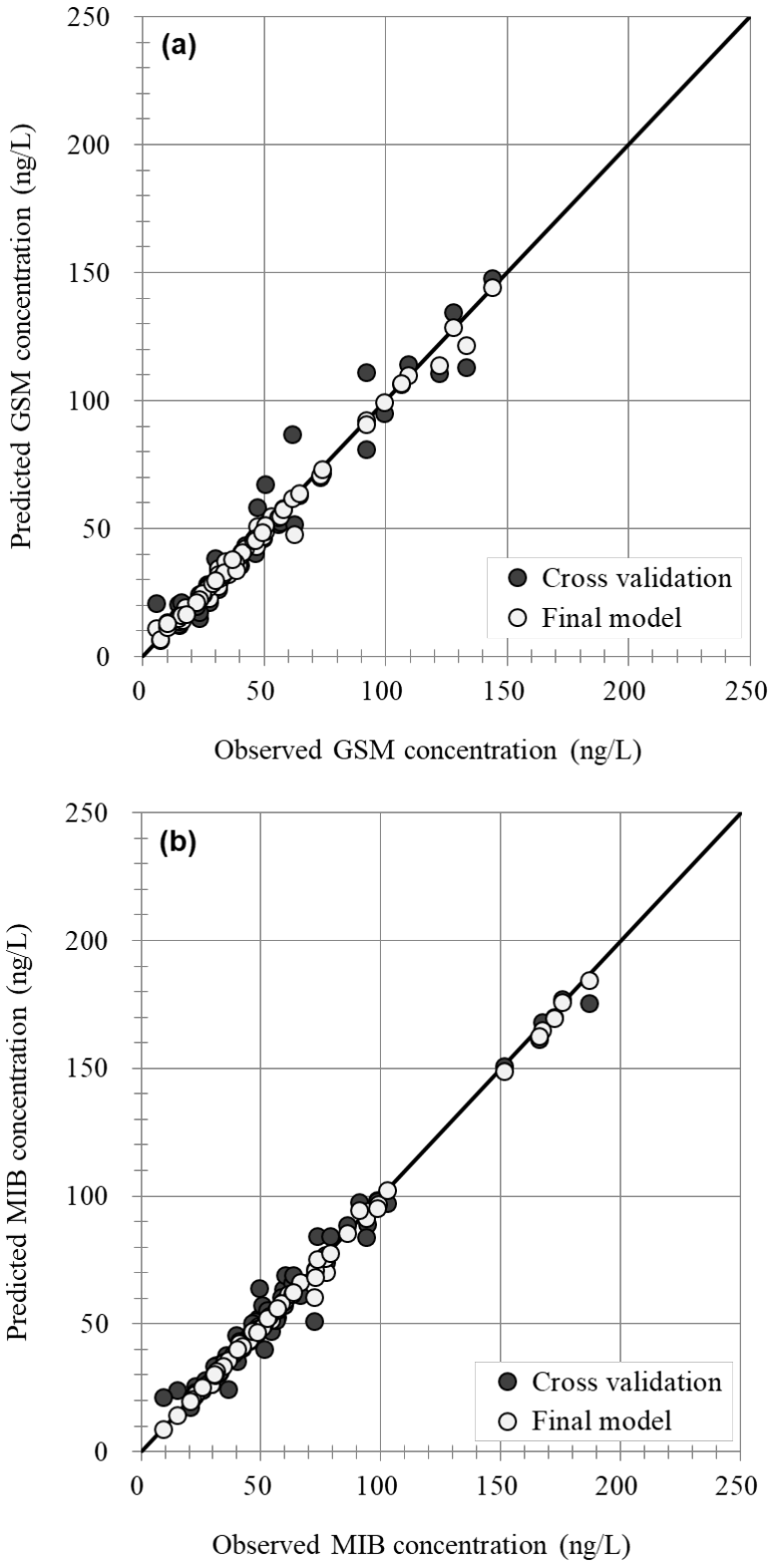

Figure 7. Comparison of predicted and observed (a) GSM and (b) MIB concentration by DNN.