Article

A Stepwise-Cluster Inference Model for Phenanthrene Immobilization at the Aqueous/Modified Palygorskite Interface

Shan Zhao [1,2], Guohe Huang [2,3,*], Guanhui Cheng [2], Wei Sun [2], Qian Su [4], Zeyu Tao [5] and Shuguang Wang [1,*]

[1] Shandong Key Laboratory of Water Pollution Control and Resource Reuse, School of Environmental Science and Engineering, Shandong University, Jinan 250100, China; szhao@sdu.edu.cn

[2] Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, SK S4S 0A2, Canada; guanhuicheng@gmail.com (G.C.); sunweihunu@hotmail.com (W.S.)

[3] Center for Energy, Environment and Ecology Research, UR-BNU, Beijing Normal University, Beijing 100875, China

[4] School of Energy and Power Engineering, Shandong University, Jinan 250100, China; shandasuqian@sina.com

[5] School of Municipal and Environmental Engineering, Shandong Jianzhu University, Jinan 250101, China; zeyutao@sina.cn

* Correspondence: huang@iseis.org (G.H.); wsg@sdu.edu.cn (S.W.)

Abstract: A stepwise-cluster inference (SI) model was established through introducing stepwise-cluster analysis (SCA) into the phenanthrene immobilization process at the aqueous/modified palygorskite interface. SCA has the advantages of tackling the nonlinear relationships among environmental factors and the phenanthrene sorption amount in the immobilization process. The essence of SCA is to form a tree-based classification on a series of cutting or mergence procedures under given statistical criteria. The results indicated that SI could help develop a statistical relationship between environmental variables and the phenanthrene sorption amount, where discrete and nonlinear complexities exist. During the experiment, data were randomly sampled 10 times for model calibration and verification. The R^2 (close to one) and root mean squared error (RMSE) (close to zero) values guaranteed the prediction accuracy of the model. Compared to other statistical methods, the calculation of R^2 and RMSEs showed that SI was more straightforward for describing the nonlinear relationships and precisely fitting and predicting the immobilization of phenanthrene. Through the calculation of the input effects on the output in the SI model, the influence of environmental factors on phenanthrene immobilization were ranged in descending order as: initial phenanthrene concentration, ionic strength, pH, added humic acid dose, and temperature. It is revealed that SCA can be used to map the nonlinear and discrete relationships and elucidate the transport patterns of phenanthrene at the aqueous/modified palygorskite interface.

Keywords: phenanthrene; immobilization; stepwise-cluster analysis; palygorskite; gemini surfactants

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a group of nonpolar hydrophobic contaminants with two or more fused benzene rings from natural as well as anthropogenic sources. PAHs can transport and be accumulated in groundwater and surface water for a long period of time, and are difficult to biodegrade. Immobilization of PAHs to solid sorbents places an important effect on its fate, transport, and bioavailability in natural aquatic environments [1]. Clay minerals, such as palygorskite, have received substantial interest as a potential alternative to conventional sorbents such as activated

carbon. Due to the intrinsic negative charges of palygorskite, it has been widely utilized as a sorbent for heavy metals [2]. However, being intrinsically hydrophilic, palygorskite shows low immobilization for hydrophobic organic contaminants. One effective way for the modification of palygorskite is organomodification, that is, to replace the natural inorganic exchangeable cations with large organic cations of surfactant molecules. Previous studies regarding the immobilization of organic pollutants by organopalygorskite have concentrated on polar organic compounds such as 2,4-D herbicide [3], atrazine pesticide [4], and phenolic compounds [5]. However, the removal of nonpolar compounds from the aqueous phase by organopalygorskite is not well documented.

It has been reported that increasing the surfactant chain length and the number of alkyl chains per surfactant molecule can increase the immobilization of organic contaminants by organoclays [6]. Containing two hydrophilic heads and two long hydrophobic tails in one surfactant molecule, cationic gemini surfactants have received emerging interest recently. Compared to their single-chain counterparts, gemini surfactants have increasing hydrophobicity and can endow solids with increasing organic matter. The sorbed surfactants can act as an effective partitioning media for nonpolar or weakly polar organic pollutants [7]. In our previous studies, the enhanced solubilization capabilities of cationic gemini surfactants towards PAHs has been proved [8]. The enhanced retention for phenanthrene by soils with the addition of gemini surfactants has been well established [9]. However, there is limited data on modifying palygorskite with gemini surfactants for the removal of PAHs from the aqueous phase. Enriched knowledge regarding the solution chemistry effects on PAH immobilization to modified palygorskite still needs to be clarified. Moreover, it is of great importance to elucidate the environmental factors controlling the sorption process, including the PAH concentration, pH, ionic strength, and temperature, etc.

Mathematical modeling is one of the essential approaches for tackling the complicated relationships among environmental factors and the sorption amount in the immobilization process. Among them, factorial design are extensively used to obtain information regarding whether or not several initial conditions had an impact on specific immobilization characteristics [10]. However, factorial analysis is not straightforward when evaluating the relation between multiple factors and the sorption amount. In addition, many model parameters need calibration and verification in order to achieve a satisfactory fit [11]. More recently, a few studies have been conducted through the development of statistical tools, such as the fuzzy factorial method [1], fuzzy logic modeling [12], support vector regression [13], multiple linear regression and tree regression analysis [14], and artificial neural networks [15]. However, many variables in the immobilization system can be either continuous or discrete, and relations among them are inherently nonlinear, which leads to difficulties in performing these methods.

Stepwise-cluster analysis (SCA) is a type of non-parameter regression technology. The essence of SCA is to form a classification tree in the sense of probability, based on a series of cutting or mergence procedures under given statistical criteria. A cluster could be cut into two sub-clusters, while two could be merged into a new cluster during the iterative training process. Step by step, a tree can be established when no clusters can be further cut or merged. SCA can not only deal with nonlinear relationships among continuous and/or discrete variables, but also clearly show the significance levels of different branches. There are a few applications of SCA for the prediction of air quality in the atmospheric environment [16], groundwater remediation [17], hydrological processes [18–21], streamflow prediction [22,23], composting [11,24], and climate change [25,26]. However, it was not reported that SCA can help reveal the complicated relationships between environmental factors and process characteristics in the PAH immobilization process. Moreover, the potential utilization of SCA for tackling discrete and nonlinear complexities in the immobilization system is desired.

Therefore, this study aims to develop a stepwise-cluster inference (SI) model through introducing SCA into the PAH immobilization process to tackle the nonlinear relationships among environmental parameters and the PAH sorption amount. In detail: (1) establishing a stepwise-cluster inference model for predicting variations of PAH immobilization at the aqueous/modified palygorskite

interface; (2) verifying the proposed modeling system based on the data obtained from the PAH immobilization process; and (3) evaluating the effects of multiple factors on the distribution of PAHs in the water/modified palygorskite system.

2. Materials and Methods

2.1. Materials

Phenanthrene was selected as the representative PAH, and was purchased from Sigma-Aldrich Canada Co. (Oakville, ON, Canada), with a purity greater than 98%. Cationic gemini surfactant (N1-dodecyl-N1, N1, N2, N2-tetramethyl-N2-octylethane-1,2-diaminium bromide, 12-2-12) was obtained from Chengdu Organic Chemicals Co., Ltd. (Chengdu, China), with a purity of 98%. The water solubility of phenanthrene is 1.06 mg/L at 298 K and the octanol-water coefficient (log Kow) is 4.57. The stock solution of phenanthrene was made by diluting the desired amount of pure phenanthrene into High Performance Liquid Chromatography (HPLC)-grade methanol solution, and was stored in a dark place at 277 K in an amber borosilicate bottle to minimize photodegradation and volatilization. Palygorskite was obtained from Huaiyuan Mining Co. Ltd. (Xuyi, China). The gemini surfactant modified palygorskite was prepared according to the method reported in our previous studies [27].

2.2. Sorption Studies

The batch sorption experiments were conducted in 20 mL glass vials. 0.1 g of clay sample was first added into the glass vial; after that, the appropriate amount of deionized water was added into the vial. The background solution contained 0.01 M NaCl as an electrolyte and the pH was 7. Then a pre-calculated volume of phenanthrene stock solution was added to each vial, and the initial concentration of phenanthrene was pre-determined for each sample. The vials were sealed with Teflon-lined screw caps and were vortexed for 20 s, and then were placed in a reciprocal shaker at 200 rpm for 24 h to reach the sorption equilibrium. Preliminary experiments showed that 24 h were sufficient for the sorption process to reach equilibrium and the experimental loss of phenanthrene was negligible [28–30]. Before testing, the samples were placed for 10 min to separate the soil from the solution. An appropriate aliquot of supernatant was then carefully withdrawn with a volumetric pipette to further determine the residual amount of phenanthrene. Meanwhile, controlled experiments without phenanthrene were conducted and the supernatant of the controlled samples was analyzed as the background concentration for PHE. The quality assurance/quality control test was conducted and the phenanthrene recovery rate was between 96% and 103%.

To investigate the influence of temperature on the immobilization of phenanthrene, the sorption behaviors of phenanthrene were carried out at 283, 293, and 303 K. The solutions contained 0.01 M NaCl and pH was kept at 7, and then the solutions were placed in a reciprocal shaker at 200 rpm for 24 h. To study the effect of pH on the immobilization of phenanthrene, the immobilization behaviors of phenanthrene were investigated at pH 3, 7, and 11. The pH value was adjusted with standard HCl or NaOH solution, and the ion concentration in the system was kept constant at 0.01 M. The initial concentrations of phenanthrene ranged from 0.3 to 1 mg/L. The temperature was kept constant at 293 K. Batch experiments were performed in the manner as mentioned for the immobilization studies. The effect of humic acid (HA) on the immobilization behaviors of phenanthrene was examined in a modified palygorskite/water system. The experiments were conducted in the presence of HA ranging from 0 to 80 mg organic carbon/L (OC/L). An appropriate volume of phenanthrene stock solution was added and the initial concentrations for phenanthrene varied from 0.3 to 1 mg/L. Temperature was kept constant at 293 K. Batch experiments were performed following the sorption test procedures. To investigate the influence of ionic strength, immobilization tests were conducted. NaCl was added at different concentrations (0.01, 0.1 and 1 M). The initial concentrations of phenanthrene were from 0.3 to 1 mg/L and the pH value was 7. The vials were placed on a reciprocal shaker at 293 K and

200 rpm for 24 h to reach the sorption equilibrium. After preliminary experiments, a 3^5 full factorial design approach was adopted to represent various factorial effects on phenanthrene immobilization, including the initial phenanthrene concentration, added HA dose, ionic strength, temperature, and pH. Three levels of each parameter obtained from the preliminary experiments were indicated in Table 1. A total of $3^5 \times 2 = 486$ experiments with all possible combinations of variables were conducted in duplicate.

Table 1. The 3^5 experimental design of phenanthrene immobilization at the aqueous/modified palygorskite interface.

Symbol	Factor	Level		
		Low (−1)	Medium (0)	High (+1)
X_1	Initial phenanthrene concentration, mg/L	0.4	0.8	1.0
X_2	Added HA dose, mg/L	2	4	42
X_3	Ionic strength, M	0.05	0.10	0.55
X_4	Temperature, K	283	293	303
X_5	pH	3.95	7.90	10.95

2.3. Analytical Methods

Phenanthrene was analyzed using HPLC. The HPLC instrument, an Agilent 1260 Infinity LC System (USA), was equipped with a vacuum degasser, binary pump, autosampler, thermostated column compartment (set to 303 K), diode array detector (DAD), and ZORBAX Eclipse PAH column (3.5 μm particle size, 4.6 mm × 150 mm ID). A mobile phase consisting of acetonitrile/water (75/25, *v/v*) was used at a flow rate of 1.0 mL/min. Phenanthrene was monitored with DAD at 250 nm. The amounts of the phenanthrene/surfactants sorbed to the soil were the difference between the initial amount added and the amount remaining in the solution. The pH measurements were conducted through a SevenEasy S20K pH meter (Mettler-Toledo, Columbus, OH, USA). All tests were conducted in duplicate and the typical error in the measurement was less than ±10%.

2.4. Data Collection and Analysis

The candidate inputs (environmental parameters) were state variables, which had potential effects on phenanthrene immobilization; the inputs included the initial phenanthrene concentration (X_1), added HA dose (X_2), ionic strength (X_3), temperature (X_4), and pH (X_5). The phenanthrene sorption amount (Y) were the outputs (dependent variables). A total of 486 samples were obtained. The data were divided into training and test sets. The training set was used for shaping the SCA trees and the test set for verifying the developed model. The model calibration and verification were conducted 10 times though SCA methods to assure the stability of the model. In each experimental run, 162 and 81 samples were randomly selected as the training and the test set, respectively. The correlation coefficient (R) and the root mean squared error (RMSE) were employed to evaluate the performance of the SCA model. The *RMSE* is represented as:

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{j=1}^{n} (v_j - y_j)^2}{n-1}} \tag{1}$$

where *n* is the number of samples in the training or test set, and v_j and y_j are the predicted and observed values for the phenanthrene immobilization amount in the *j*th sample.

2.5. Stepwise-Cluster Analysis

The classical methods are mainly classification methods associated with a few complexities. For example, the variable observations collected for classical methods may be uncertain due to causes such as instrumental or operational errors. The multiple variables selected to characterize the sorption process may be dependent on each other. A common condition for most statistical classification methods is that the variable samples should come from a normally distributed population. In addition, many classical methods rely on subjective judgements to support the screening of variables, selection of classification thresholds, and setting of sample sizes or numbers prior to classification practices. These issues are challenging the effectiveness of existing classification methods, the reliability of classification results, and the reasonability of impact studies or the other related research. Recently, nonparametric statistical methods have received much attention because of their superior capability in capturing nonlinear and discrete relationships between state and response variables [17]. Stepwise-cluster analysis (SCA) is an emerging non-parameter regression technique. It consists of a series of cutting or merging operations according to given statistical criteria and finally generates a cluster tree in the sense of probability. SCA is capable of reflecting differences both between and within clusters, therefore improving the prediction accuracy [24].

The whole or part of the training set can be treated as one cluster (α), including n_α samples, m independent variables (X), and one dependent variable (Y), indicated as follows:

$$\alpha = [X_1, X_2, X_3, \ldots, X_m, Y] = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \, y_1 \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2m} \, y_2 \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3m} \, y_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n_\alpha 1} & x_{n_\alpha 2} & x_{n_\alpha 3} & \cdots & x_{n_\alpha m} \, y_{n_\alpha} \end{bmatrix} \quad (2)$$

Here please refer to the flowing SCA chart with detailed descriptions in our previous study [11]. Let cluster α be cut into two sub-clusters β and γ (with n_β and n_γ samples, respectively). According to Wilks' likelihood-ratio criterion, the cutting point is optimal only if Wilks' Λ value is a minimum [31]. According to Wilk's likelihood-ratio criterion, the smaller the Λ value, the bigger the difference between the sample means of β and γ. Since the Λ is directly related to the *F* statistic, the sample means of the two sub-clusters can be compared for significant differences through an *F* test. Therefore, the criteria of cutting (or merging or not merging) clusters is based on the *F* tests. All sub-clusters produced from the original dataset will enter a set of iterative cutting (or merging) runs until all hypotheses of further cut (or mergence) are rejected or the minimum number of samples (N_{min}) within every cluster is reached. After all calculations and tests are completed, an SCA tree can be built which indicates that the training is done. When a new sample (x_1, x_2, ..., x_m, y_1; y_1 is unknown) enters the tree at a cutting point, the sample will finally drop into a tip cluster which cannot be either cut or merged further according to the routes decided by new independent variables (x_1, x_2, ..., x_m). The predicted value of y_1 will be the mean of the dependent variables of the training samples in the tip cluster. Therefore, the SCA tree can predict new dependent variables when new samples enter the tree from top to bottom.

3. Results and Discussion

3.1. Sorption Studies

The performance of gemini surfactant modified palygorskite for the removal of phenanthrene under the impacts of different state variables was investigated. From Figure 1a–d, it can be seen that the phenanthrene immobilization was affected by the initial concentration, with a significant sorption increase as the initial concentration increased from 0.3 to 1.0 mg/L. The initial concentration provided an important driving force for overcoming all mass transfer resistance of phenanthrene molecules between the aqueous and solid phases [1]. In Figure 1a, the effects of ionic strength on

the immobilization of phenanthrene on modified palygorskite were evaluated, with phenanthrene from 0.3 to 1.0 mg/L, ionic strength from 0.01 to 1 M, and a constant pH of 7. The phenanthrene sorbed amounts were found to increase significantly with the increasing NaCl concentration from 0.01 to 1 M. The maximum phenanthrene immobilization was 0.476 mg/g at NaCl of 1 M. It could, to a certain extent, be explained by the "salting-out" effect, which refers to the reduced solubility of organic compounds in salt solutions [32]. Phenanthrene is a nonionic compound and the immobilization was partially due to hydrophobic interaction mechanisms. The decrease in solubility caused an increase in the immobilization capacity of the sorbents due to increasing hydrophobic interactions induced by the increasing NaCl concentration [33]. Generally, the amount of phenanthrene sorbed by modified palygorskite increased with the increasing NaCl concentrations at a given phenanthrene aqueous concentration. The addition of more salt ions, therefore, can enhance the retention of phenanthrene on modified palygorskite. This is of special interest for the removal of PAHs through modified palygorskite from an aqueous phase with a high salinity level.
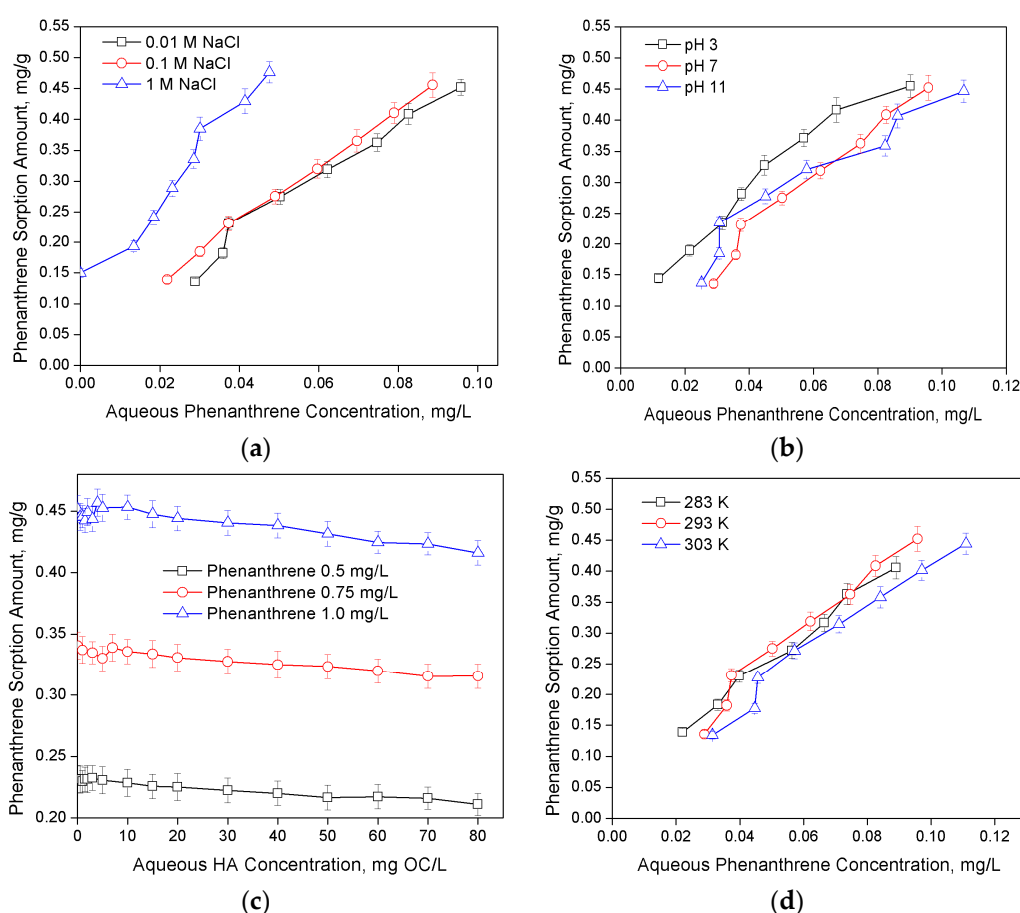
Figure 1. The phenanthrene sorption data at different (**a**) ionic strength; (**b**) pH; (**c**) humic acid (HA); and (**d**) temperature levels.

The effects of pH on phenanthrene immobilization onto modified palygorskite were studied through batch experiments at pH 3, 7, and 11. Figure 1b shows that the pH level can significantly affect phenanthrene immobilization on modified palygorskite. The maximum uptake of phenanthrene was 0.455 mg/L at pH 3 for the gemini surfactant modified palygorskite. Then the sorption reduced markedly from pH 3 to 7 at all given concentrations of phenanthrene. A further increase of pH from 7 to 11 had less pronounced effects on phenanthrene immobilization. Three mechanisms applicable to PAH sorption were proposed: the hydrophobic interactions, the electron donor–acceptor interaction, and the π–π interaction [34]. Because palygorskite clay minerals contained low amounts of humic

substances [35], mechanisms such as enhanced dipole interaction between the charged surface (electron acceptors) and phenanthrene with electron-rich π systems (electron donors) might be the major cause of phenanthrene immobilization. In addition, the solution pH affected the surface properties of the sorbent and the degree of ionization of the sorbate [36]. The pH_{pzc} of modified palygorskite was 7.9. The mineral surface would be covered with negative charges when the pH was higher than pH_{pzc}. The modified palygorskite was subjected to protonation at low pH or deprotonation at elevated pH [37]. This would affect the sorption sites for PAHs on sorbent surface. The above-mentioned mechanisms were to some extent related to solution pH, resulting in a dependence of PAH sorption on aqueous pH. The detailed mechanisms depended on the physical and chemical properties of the interactive sorbate–sorbent system and still needed further investigation.

The effects of HA on phenanthrene immobilization to gemini-surfactant-modified palygorskite as a function of HA dose (0–80 mg OC/L) were investigated and the results are shown in Figure 1c. The initial phenanthrene concentrations were 0.5, 0.75, and 1 mg/L, respectively. The immobilization of phenanthrene increased at the low HA level and then decreased if more HA was added. Specifically, when the added dose of HA was less than 4 mg OC/L, the sorbed phenanthrene amount varied from 0.442 to 0.460 mg/g. However, the binding affinity of phenanthrene to the modified palygorskite decreased with the further addition of HA. By gradually increasing the HA concentration to 80 mg OC/L, the retention of phenanthrene was hindered from 0.460 to 0.416 mg/g. At a low HA dose, the initial phenanthrene sorption increase was presumably due to the binding of phenanthrene to HA along with the sorption of HA onto modified palygorskite, forming a complexation of sorbed HA with phenanthrene [38]. The hydrophobic interaction mechanism was reported as the major factor responsible for the binding of PAHs to HA [39]. This process could also be produced by the $\pi-\pi$ interactions between phenanthrene and humic substance acceptor groups in aqueous solutions [40]. The presence of HA at the low concentration facilitated the uptake of phenanthrene due to the cosorption of HA and the phenanthrene complex on the modified palygorskite surface. At higher HA concentrations, phenanthrene immobilization decreased with increasing HA concentration. Instead of cosorption, the HA molecules would inhibit phenanthrene retention through competition for limited sites on modified palygorskite surfaces [41]. With the increasing dose of HA, HA molecules would occupy the pores of modified palygorskite surfaces that were large enough to accommodate phenanthrene molecules, thus reducing the accessibility of phenanthrene molecules to sorption sites and reducing the phenanthrene sorption on modified palygorskite [42]. In addition, the high hydrophobicity of phenanthrene enhanced its ability to be dissolved into the hydrophobic HA molecules in the solution and impeded the phenanthrene retention on the modified palygorskite. The pH_{PZC} value of modified palygorskite was 7.9. Under the test pH, the HA sorption would decrease and more HA molecules would enter the aqueous phase. This would also result in enhanced phenanthrene solubility in the solution and thus in reduced phenanthrene sorption from sorbents.

The effects of temperature on phenanthrene immobilization were investigated in Figure 1d. There was slight difference in phenanthrene sorption when the temperature varied from 283 to 293 K. However, the sorption decreased from 0.452 to 0.445 mg/g with temperatures varying from 293 K to 303 K. This indicated that temperature played a negative effect on phenanthrene immobilization. The uptake of phenanthrene was found to reduce with increasing temperature, indicating that phenanthrene immobilization on modified palygorskite was favored at lower temperatures. The reduction in phenanthrene immobilization capacity with increasing temperature from 293 to 303 K indicated an exothermic nature of the immobilization process. One explanation was that increasing temperature could enhance phenanthrene solubility, and decrease immobilization [43].

3.2. SCA Trees

An SCA tree was established to reflect the relationships between the state variables and the dependent variables. There were several factors affecting the performance of the SCA model, including the quality of original experimental data, internal configuration parameters (α and N_{min}),

the combination of state variables and the data partition strategy [11]. The internal parameters had significant effects on the shape of the SCA trees, since the SCA relied on these variables for cutting and merging procedures as well as for stopping the iterations. The criteria for the cutting and merging clusters were: cut cluster when $p \leq \alpha$ and merge clusters when $p > \alpha$, where the p values used at cutting and merging knots were significance levels of the F test. In this study, a default significance level of 0.01 was set for both cutting and merging exercises. In general, the higher the α value, the lower the F level (i.e., a decreased strictness in the cutting which would result in more cutting operations). Similarly, the lower the α value, the higher the F level (i.e., more merging operations because of the reduced strictness of mergence). The N_{min} was set as 3 in this study and it also placed significant effects on the scales of the cluster trees since it was used as one of the end criteria in training the SCA tree.

As for experimental run No. 3, the corresponding SCA tree is described in Figure 2; it formed a forecasting system to reflect the phenanthrene immobilization process at the water/modified palygorskite interface. Based on the tree, the phenanthrene immobilization can be predicted. For example, let $X_1 = 0.9$, $X_2 = 20$, $X_3 = 0.3$, $X_4 = 300$, and $X_5 = 9$ be new inputs. To predict the corresponding phenanthrene immobilization process, we have: $X_1 > 0.6$ for the first branch knot so that the sample enters cluster 3; $X_1 \leq 0.9$, so that it enters cluster 6; $X_3 \leq 0.325$, so that it enters cluster 12; $X_5 > 8.75$, so that it enters cluster 19; and $X_2 > 7$, so that it finally enters cluster 29 with a prediction value of 0.337 ± 0.004. Let $X_1 = 0.4$, $X_2 = 6$, $X_3 = 0.2$, $X_4 = 290$, and $X_5 = 6.5$ as new inputs. To predict the corresponding phenanthrene immobilization, we have: $X_1 \leq 0.6$ for the first branch knot so that the sample enters cluster 2; $X_3 \leq 0.325$, so that it enters cluster 4; $X_5 > 5.25$, so that it enters cluster 9; and $X_2 \leq 7$, so that it enters cluster 16 and finally merges into cluster 38 with a prediction value of 0.182 ± 0.011. Let $X_1 = 0.65$, $X_2 = 12$, $X_3 = 0.08$, $X_4 = 310$, and $X_5 = 8.5$ as new inputs. To predict the corresponding phenanthrene immobilization, we have: $X_1 > 0.6$ for the first branch knot so that the sample enters cluster 3; $X_1 \leq 0.9$, so that it enters cluster 6; $X_3 \leq 0.325$, so that it enters cluster 12; $X_5 \leq 8.75$, so that it enters cluster 18; $X_3 > 0.075$, so that it enters cluster 27; and $X_5 > 5.25$, so that it enters cluster 37 and finally merges into cluster 40 with a prediction value of 0.359 ± 0.021. Similarly, we can obtain other prediction data with respect to any combination of X_1–X_5 through the SCA tree.
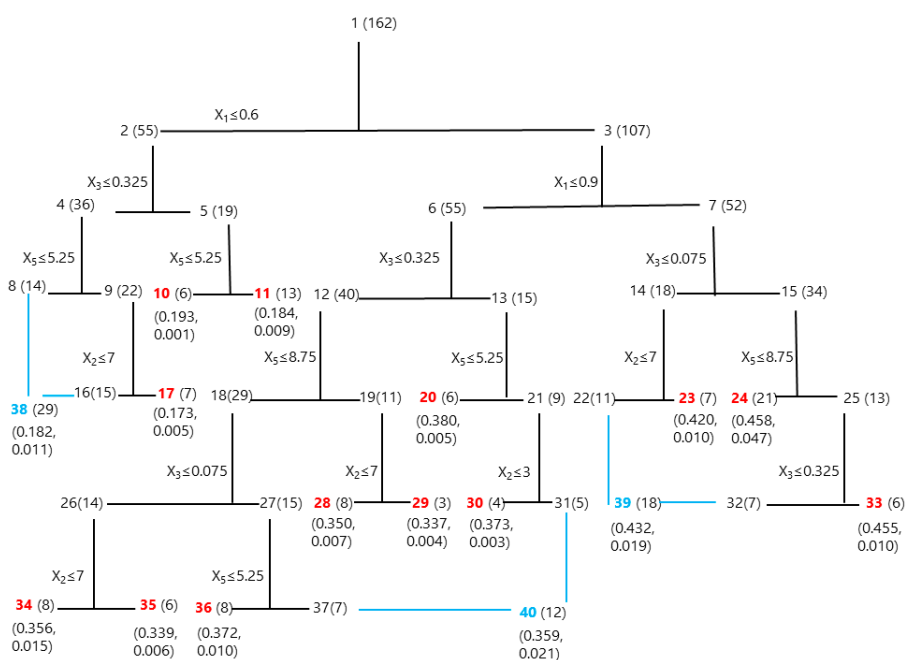
Figure 2. The stepwise-cluster analysis (SCA) tree for phenanthrene immobilization at aqueous/modified palygorskite interface (Run No. 3, the cutting nodes are highlighted in red while the merging nodes are highlighted in blue). X_1: Initial phenanthrene concentration; X_2: Added HA dose; X_3: Ionic strength; X_4: Temperature; X_5: pH.

The tree clearly shows the role of every environmental parameter in mapping the relationship. However, the SCA tree could not directly quantify the effects of environmental factors on phenanthrene immobilization. To this end, the effects of input (X_k) on the output (Y) in SCA are defined as [11]:

$$Effects(X_k, Y) = \sum_{i=1}^{NN} N_{X_k}(i) \tag{3}$$

where, *NN* is the number of total nodes in the SCA tree; and N_{Xk} *(i)* is the number of patterns (samples) at node *i* of the SCA tree where the corresponding X_k variable is used as the cutting criteria at node *i*. Correspondingly, N_{X_k} *(i)* represents the number of samples on which the X_k variable has influence at node *i*. To a certain extent, the layers (nodes) where X_k is located in the SCA tree represent the weight effect which X_k has on Y. This is because the higher the independent variables are in the SCA tree, the earlier the classification criteria would depend on the variables, and the more data patterns on which X_k would have an effect. Therefore, based on the obtained SCA tree, the effects of the environmental factors on phenanthrene immobilization were ranged in a descending order as: initial phenanthrene concentration, ionic strength, pH, added HA dose, and temperature. The results are consistent with our previous studies using the fuzzy factorial method [1]. One of the advantages of SCA is that it has no requirement for the format of training and test data, while factorial method needs preliminary experimental design for the levels of input data.

Figures 3–5 and Table 2 show the results obtained using least squares support vector machines (LSSVM) and random forest (RF) models for experimental run No. 3. The reason for choosing these models for comparison was that the current models used for constructing relationships in sorption studies are mainly based on these methods. The LSSVM-related model indicated a higher predictive capability than the linear method for the sorption of methylene blue [44]. RF proved to be more powerful than multiple linear regression to predict the sorption of bromophenol blue using activated carbon sorbents [45]. As for the immobilization of phenanthrene on modified palygorskite, the coefficient of the determination of R^2 for training and test sets of SCA (0.994 and 0.992) were higher than those of LSSVM (0.843 and 0.710) and those of RF (0.881 and 0.883). The RMSEs for training and test sets of SCA (0.00837 and 0.0101) were lower than those of LSSVM (0.0454 and 0.0642) and those of RF (0.0577 and 0.0603). From the data, the calibration and verification efficiency of RF was higher than 0.8 and the whole dataset distributed along a straight line. However, the dataset was unbalanced in terms of regional distribution and all the data were away from the line. Moreover, the calibration and verification ability of LSSVM was limited. These results confirmed the better fitting and predictive ability of SCA relative to LSSVM and RF when building nonlinear relationships between environmental factors and the sorption amount of phenanthrene in the immobilization process.
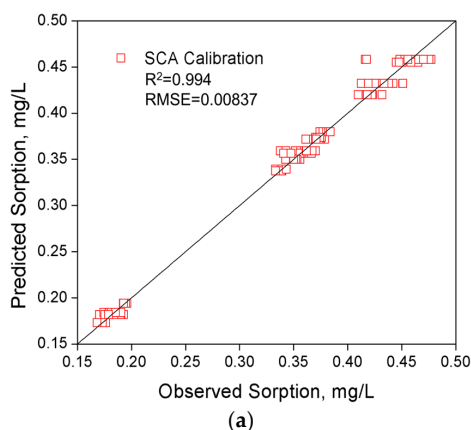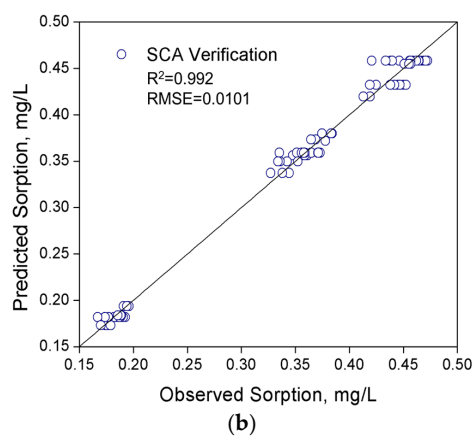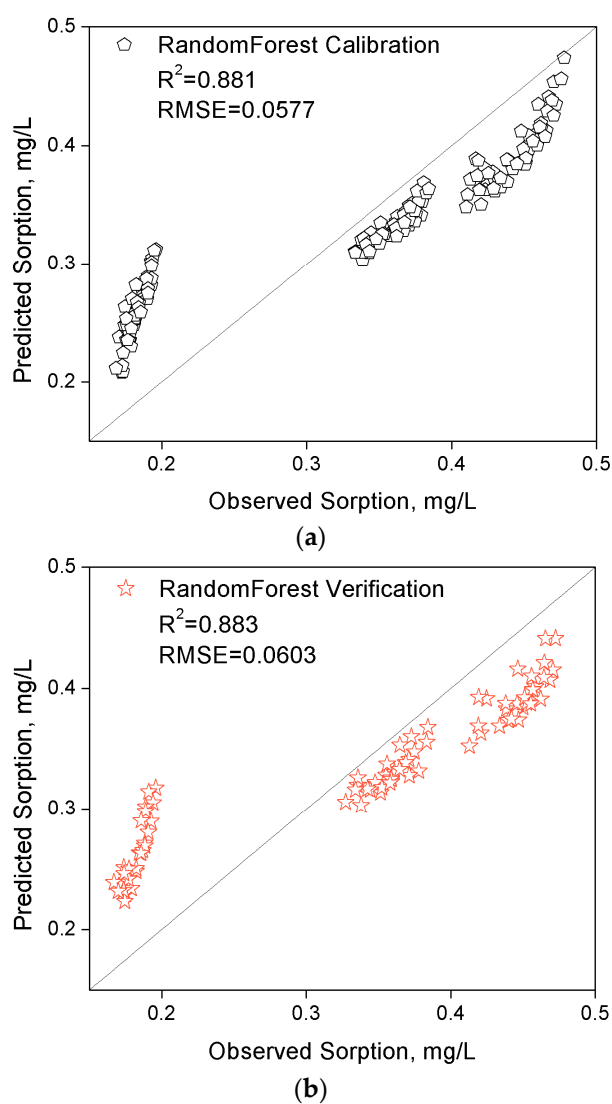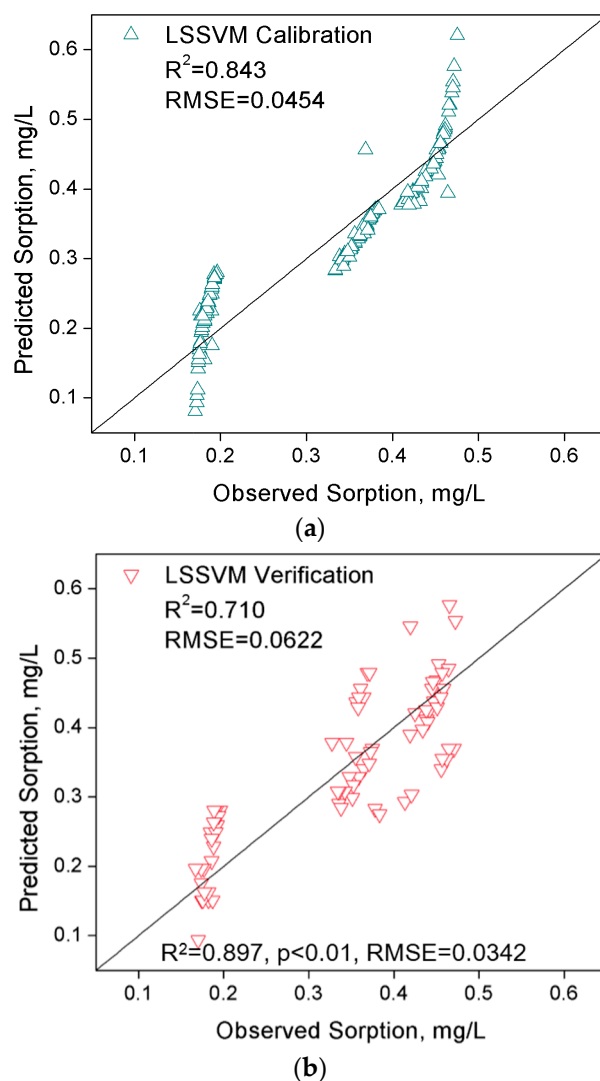

(a)

Figure 3. *Cont.*

(**b**)

Figure 3. The calibration (**a**) and verification (**b**) of phenanthrene sorption data using the stepwise-cluster analysis (SCA) method (Run No. 3).

(**a**)

(**b**)

Figure 4. The calibration (**a**) and verification (**b**) of phenanthrene sorption data using the random forest (RF) method (Run No. 3).

(a)

(b)

Figure 5. The calibration (**a**) and verification (**b**) of phenanthrene sorption data using the least squares support vector machines (LSSVM) method (Run No. 3).

Table 2. Comparisons among SCA, LSSVM and random forest for the fitting of phenanthrene sorption data (Run No. 3).

Run No. 3	SCA		Random Forest		LSSVM	
	Training	Test	Training	Test	Training	Test
R^2	0.994	0.992	0.881	0.883	0.843	0.710
RMSE	0.0084	0.0100	0.0577	0.0603	0.0454	0.0622

In this study, the data were collected 10 times for the model calibration and verification through SCA methods. The calibration and verification of the model were conducted 10 times to assure the stability of the model. Each experimental run consists of 243 samples, 162 samples for the training set and 81 samples for the test set. The data were randomly sampled based on uniform distribution. The correlation coefficients R^2 and RMSE of SCA from ten experimental runs are listed in Table 3. It is indicated in the table that all the R^2 are approaching one and the RMSE approaching zero. This confirmed the excellent prediction performance of the model when building nonlinear relationships through SCA methods.

Table 3. R^2 and RMSE through SCA methods through 10 experimental runs.

SCA	R^2	RMSE	SCA	R^2	RMSE
Calib_run1	0.994	0.00806	Verif_run1	0.990	0.0104
Calib_run2	0.996	0.00701	Verif_run2	0.994	0.00827
Calib_run3	0.994	0.00837	Verif_run3	0.992	0.0101
Calib_run4	0.996	0.00748	Verif_run4	0.990	0.0107
Calib_run5	0.994	0.00796	Verif_run5	0.990	0.0110
Calib_run6	0.996	0.00767	Verif_run6	0.988	0.0118
Calib_run7	0.994	0.00813	Verif_run7	0.992	0.0103
Calib_run8	0.996	0.00757	Verif_run8	0.988	0.0120
Calib_run9	0.992	0.00906	Verif_run9	0.992	0.00941
Calib_run10	0.994	0.00801	Verif_run10	0.992	0.00956

After collecting the experimental data 10 times, 10 SCA trees were built accordingly. The sequence of the significance of the factors is listed in Table 4. It is obvious that X_1 (initial phenanthrene concentration) and X_3 (ionic strength) are the two most significant factors influencing the sorption process for all the experimental runs. X_2 and X_5 are less significant than X_1 and X_3. This is because the sample size becomes smaller after several cutting/merging processes, and the system is not that stable with a small dataset. Therefore, the sequence of X_2 (added HA dose) and X_5 (pH) might be different. The most insignificant factor compared to the other four parameters is X_4 (temperature), though it still has certain effects on the immobilization process (Figure 1). In all, the ordering of the effect of the input variables was basically the same for multiple trees, indicating that SCA can be used to map the nonlinear and discrete relationships and elucidate the effects of factors on phenanthrene immobilization at the aqueous/modified palygorskite interface.

Table 4. The sequence of the significance of the factors in the SCA tree.

Run Number	Sequence of the Environmental Factor			
1	X_1	X_3	X_2	X_5
2	X_1	X_3	X_2	X_5
3	X_1	X_3	X_5	X_2
4	X_1	X_3	X_5	X_2
5	X_1	X_3	X_5	X_2
6	X_1	X_3	X_2	X_5
7	X_1	X_3	X_5	X_2
8	X_1	X_3	X_2	X_5
9	X_1	X_3	X_2	X_5
10	X_1	X_3	X_5	X_2

4. Conclusions

In this paper, a stepwise-cluster inference model was developed to analyze the nonlinear relationships among environmental factors and phenanthrene immobilization at the water/modified palygorskite interface, including the initial phenanthrene concentration, pH, ionic strength, temperature, and added HA dose. The results indicated that SI could help establish a statistical relationship between environmental parameters and phenanthrene immobilization where discrete and nonlinear complexities exist. Compared to LSSVM and RF methods, SI was more straightforward for describing the nonlinear relationships and precisely fitting and predicting the immobilization of phenanthrene. R^2 and RMSE were calculated to demonstrate the accuracies of the developed forecasting trees. Through the calculation of input effects on the output in SI model, the effects of environmental factors on phenanthrene immobilization were ranged in a descending order as: initial phenanthrene concentration, ionic strength, pH, added HA dose, and temperature. The maximum phenanthrene immobilization on modified palygorskite was 0.476 mg/g under the experimental

conditions tested and the corresponding removal efficiency of phenanthrene was higher than 94%. To the best of our knowledge, this study was the first attempt to introduce SCA into mapping the nonlinear relationships involved in PAH immobilization processes. It was expected that SCA would expand the potential applications to other complicated relationships during multiple types of sorbent/sobate systems in the immobilization processes.

Author Contributions: Shan Zhao designed and conducted all the experiments and established the SI model. Guohe Huang contributed to the acquisition of data, and analysis and interpretation of data. Guanhui Cheng proposed the methodology and wrote the code. Wei Sun conducted the calibration and verification of the model through SCA methods. Qian Su and Zeyu Tao collected the data and did the preliminary analysis. Shuguang Wang commented on the experimental design and was involved in drafting the manuscript. Guohe Huang and Shuguang Wang offered many insightful suggestions which substantially improved the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhao, S.; Huang, G.; Wang, S.; Wang, X.; Huang, W. Insight into sorption mechanism of phenanthrene onto gemini modified palygorskite through a multi-level fuzzy-factorial inference approach. *J. Environ. Sci. Health Part A* **2016**, *51*, 759–768. [CrossRef] [PubMed]
2. Thanos, A.G.; Katsou, E.; Malamis, S.; Psarras, K.; Pavlatou, E.A.; Haralambous, K.J. Evaluation of modified mineral performance for chromate sorption from aqueous solutions. *Chem. Eng. J.* **2012**, *211–212*, 77–88. [CrossRef]
3. Xi, Y.; Mallavarapu, M.; Naidu, R. Adsorption of the herbicide 2,4-D on organo-palygorskite. *Appl. Clay Sci.* **2010**, *49*, 255–261. [CrossRef]
4. Rodríguez-Cruz, M.S.; Sánchez-Martín, M.J.; Andrades, M.S.; Sánchez-Camazano, M. Modification of clay barriers with a cationic surfactant to improve the retention of pesticides in soils. *J. Hazard. Mater.* **2007**, *139*, 363–372. [CrossRef] [PubMed]
5. Sarkar, B.; Megharaj, M.; Xi, Y.; Naidu, R. Surface charge characteristics of organo-palygorskites and adsorption of p-nitrophenol in flow-through reactor system. *Chem. Eng. J.* **2012**, *185–186*, 35–43. [CrossRef]
6. Zhu, L.; Yimin Li, A.; Zhang, J. Sorption of organobentonites to some organic pollutants in water. *Environ. Sci. Technol.* **1997**, *31*, 1407–1410. [CrossRef]
7. Zhu, L.; Chen, B.; Tao, S.; Chiou, C.T. Interactions of organic contaminants with mineral-adsorbed surfactants. *Environ. Sci. Technol.* **2003**, *37*, 4001–4006. [CrossRef] [PubMed]
8. Wei, J.; Huang, G.; Yu, H.; An, C. Efficiency of single and mixed Gemini/conventional micelles on solubilization of phenanthrene. *Chem. Eng. J.* **2011**, *168*, 201–207. [CrossRef]
9. Zhao, S.; Huang, G.; An, C.; Wei, J.; Yao, Y. Enhancement of soil retention for phenanthrene in binary cationic gemini and nonionic surfactant mixtures: Characterizing two-step adsorption and partition processes through experimental and modeling approaches. *J. Hazard. Mater.* **2015**, *286*, 144–151. [CrossRef] [PubMed]
10. Gottipati, R.; Mishra, S. Process optimization of adsorption of Cr(VI) on activated carbons prepared from plant precursors by a two-level full factorial design. *Chem. Eng. J.* **2010**, *160*, 99–107. [CrossRef]
11. Sun, W.; Huang, G.H.; Zeng, G.; Qin, X.; Sun, X. A stepwise-cluster microbial biomass inference model in food waste composting. *Waste Manag.* **2009**, *29*, 2956–2968. [CrossRef] [PubMed]
12. Hosseini Asl, S.M.; Masomi, M.; Hosseini, M.; Javadian, H.; Ruiz, M.; Sastre, A.M. Synthesis of hydrous iron oxide/aluminum hydroxide composite loaded on coal fly ash as an effective mesoporous and low-cost sorbent for Cr(VI) sorption: Fuzzy logic modeling. *Process Saf. Environ. Prot.* **2017**, *107*, 153–167. [CrossRef]
13. Parveen, N.; Zaidi, S.; Danish, M. Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI). *Process Saf. Environ. Prot.* **2017**, *107*, 428–437. [CrossRef]

14. Vega, F.A.; Andrade, M.L.; Covelo, E.F. Influence of soil properties on the sorption and retention of cadmium, copper and lead, separately and together, by 20 soil horizons: Comparison of linear regression and tree regression analyses. *J. Hazard. Mater.* **2010**, *174*, 522–533. [CrossRef] [PubMed]

15. Khandanlou, R.; Fard Masoumi, H.R.; Ahmad, M.B.; Shameli, K.; Basri, M.; Kalantari, K. Enhancement of heavy metals sorption via nanocomposites of rice straw and Fe_3O_4 nanoparticles using artificial neural network (ANN). *Ecol. Eng.* **2016**, *91*, 249–256. [CrossRef]

16. Huang, G. A stepwise cluster analysis method for predicting air quality in an urban environment. *Atmos. Environ. Part B Urban Atmos.* **1992**, *26*, 349–357. [CrossRef]

17. Wang, S.; Huang, G.H.; He, L. Development of a clusterwise-linear-regression-based forecasting system for characterizing DNAPL dissolution behaviors in porous media. *Sci. Total Environ.* **2012**, *433*, 141–150. [CrossRef] [PubMed]

18. Cheng, G.; Huang, G.; Dong, C.; Zhou, X.; Zhu, J.; Xu, Y. Climate classification through recursive multivariate statistical inferences: A case study of the Athabasca River Basin, Canada. *Int. J. Climatol.* **2017**. [CrossRef]

19. Cheng, G.; Huang, G.; Dong, C.; Zhu, J.; Zhou, X.; Yao, Y. High-resolution projections of 21st century climate over the Athabasca River Basin through an integrated evaluation-classification-downscaling-based climate projection framework. *J. Geophys. Res. Atmos.* **2017**, *122*, 2595–2615. [CrossRef]

20. Li, Z.; Huang, G.; Wang, X.; Han, J.; Fan, Y. Impacts of future climate change on river discharge based on hydrological inference: A case study of the Grand River Watershed in Ontario, Canada. *Sci. Total Environ.* **2016**, *548–549*, 198–210. [CrossRef] [PubMed]

21. Li, Z.; Huang, G.; Han, J.; Wang, X.; Fan, Y.; Cheng, G.; Zhang, H.; Huang, W. Development of a stepwise-clustered hydrological inference model. *J. Hydrol. Eng.* **2015**, *20*, 4015008. [CrossRef]

22. Fan, Y.R.; Huang, G.H.; Li, Y.P.; Wang, X.Q.; Li, Z. Probabilistic prediction for monthly streamflow through coupling stepwise cluster analysis and quantile regression methods. *Water Resour. Manag.* **2016**, *30*, 5313–5331. [CrossRef]

23. Fan, Y.R.; Huang, W.; Huang, G.H.; Li, Z.; Li, Y.P.; Wang, X.Q.; Cheng, G.H.; Jin, L. A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 1557–1569. [CrossRef]

24. Sun, W.; Huang, G.H.; Zeng, G.; Qin, X.; Yu, H. Quantitative effects of composting state variables on C/N ratio through GA-aided multivariate analysis. *Sci. Total Environ.* **2011**, *409*, 1243–1254. [CrossRef] [PubMed]

25. Wang, X.; Huang, G.; Lin, Q.; Nie, X.; Liu, J. High-resolution temperature and precipitation projections over Ontario, Canada: A coupled dynamical-statistical approach. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 1137–1146. [CrossRef]

26. Wang, X.; Huang, G.; Lin, Q.; Nie, X.; Cheng, G.; Fan, Y.; Li, Z.; Yao, Y.; Suo, M. A stepwise cluster analysis approach for downscaled climate projection—A Canadian case study. *Environ. Model. Softw.* **2013**, *49*, 141–151. [CrossRef]

27. Zhao, S.; Huang, G.; Mu, S.; An, C.; Chen, X. Immobilization of phenanthrene onto gemini surfactant modified sepiolite at solid/aqueous interface: Equilibrium, thermodynamic and kinetic studies. *Sci. Total Environ.* **2017**, *598*, 619–627. [CrossRef] [PubMed]

28. Asztalos, J.; Kim, Y. Lab-scale experiment and model study on enhanced digestion of wastewater sludge using bioelectrochemical systems. *J. Environ. Inform.* **2017**, *29*, 98–109.

29. Ben-Awuah, E.; Elkington, T.; Askari-Nasab, H.; Blanchfield, F. Simultaneous production scheduling and waste management optimization for an oil sands application. *J. Environ. Inform.* **2015**, *26*, 80–90.

30. Forsythe, K.; Irvine, K.N.; Atkinson, D.; Perrelli, M.; Aversa, J.; Swales, S.; Gawedzki, A.; Jakubek, D. Assessing lead contamination in Buffalo River sediments. *J. Environ. Inform.* **2015**, *26*, 106–111.

31. Wilks, S.S. *Mathematical Statistics*; John Wiley and Sons: New York, NY, USA, 1962.

32. Ko, S.O.; Schlautman, M.A.; Carraway, E.R. Effects of solution chemistry on the partitioning of phenanthrene to sorbed surfactants. *Environ. Sci. Technol.* **2014**, *32*, 3542–3548. [CrossRef]

33. Zhang, S.; Shao, T.; Bekaroglu, S.S.K.; Karanfil, T. Adsorption of synthetic organic chemicals by carbon nanotubes: Effects of background solution chemistry. *Water Res.* **2010**, *44*, 2067–2074. [CrossRef] [PubMed]

34. Yuan, M.; Tong, S.; Zhao, S.; Jia, C.Q. Adsorption of polycyclic aromatic hydrocarbons from water using petroleum coke-derived porous carbon. *J. Hazard. Mater.* **2010**, *181*, 1115–1120. [CrossRef] [PubMed]

35. Fang, J.; Shan, X.; Wen, B.; Lin, J.; Lu, X.; Liu, X.; Owens, G. Sorption and desorption of phenanthrene onto iron, copper, and silicon dioxide nanoparticles. *Langmuir* **2008**, *24*, 10929–10935. [CrossRef] [PubMed]

36. Sarkar, B.; Xi, Y.; Megharaj, M.; Krishnamurti, G.S.R.; Naidu, R. Synthesis and characterisation of novel organopalygorskites for removal of p-nitrophenol from aqueous solution: Isothermal studies. *J. Colloid Interface Sci.* **2010**, *350*, 295–304. [CrossRef] [PubMed]

37. Zhang, W.; Zhuang, L.; Yuan, Y.; Tong, L.; Tsang, D.C.W. Enhancement of phenanthrene adsorption on a clayey soil and clay minerals by coexisting lead or cadmium. *Chemosphere* **2011**, *83*, 302–310. [CrossRef] [PubMed]

38. Radian, A.; Mishael, Y. Effect of humic acid on pyrene removal from water by polycation-clay mineral composites and activated carbon. *Environ. Sci. Technol.* **2012**, *46*, 6228–6235. [CrossRef] [PubMed]

39. An, C.; Huang, G. Stepwise Adsorption of phenanthrene at the fly ash–water interface as affected by solution chemistry: Experimental and modeling studies. *Environ. Sci. Technol.* **2012**, *46*, 12742–12750. [CrossRef] [PubMed]

40. Zhu, D.; Hyun, S.; Pignatello, J.J.; Lee, L.S. Evidence for $\pi-\pi$ electron donor-acceptor interactions between π-donor aromatic compounds and π-acceptor sites in soil organic matter through pH effects on sorption. *Environ. Sci. Technol.* **2004**, *38*, 4361–4368. [CrossRef] [PubMed]

41. Wang, X.; Tao, S.; Xing, B. Sorption and competition of aromatic compounds and humic acid on multiwalled carbon nanotubes. *Environ. Sci. Technol.* **2009**, *43*, 6214–6219. [CrossRef] [PubMed]

42. Lou, L.; Liu, F.; Yue, Q.; Chen, F.; Yang, Q.; Hu, B.; Chen, Y. Influence of humic acid on the sorption of pentachlorophenol by aged sediment amended with rice-straw biochar. *Appl. Geochem.* **2013**, *33*, 76–83. [CrossRef]

43. Ping, L.; Luo, Y.; Wu, L.; Qian, W.; Song, J.; Christie, P. Phenanthrene adsorption by soils treated with humic substances under different pH and temperature conditions. *Environ. Geochem. Health* **2006**, *28*, 189–195. [CrossRef] [PubMed]

44. Ghaedi, M.; Ghaedi, A.M.; Hossainpour, M.; Ansari, A.; Habibi, M.H.; Asghari, A.R. Least square-support vector (LS-SVM) method for modeling of methylene blue dye adsorption using copper oxide loaded on activated carbon: Kinetic and isotherm study. *J. Ind. Eng. Chem.* **2014**, *20*, 1641–1649. [CrossRef]

45. Ghaedi, M.; Ghaedi, A.M.; Negintaji, E.; Ansari, A.; Vafaei, A.; Rajabi, M. Random forest model for removal of bromophenol blue using activated carbon obtained from Astragalus bisulcatus tree. *J. Ind. Eng. Chem.* **2014**, *20*, 1793–1803. [CrossRef]