Analytical Chemistry: A Journal (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/analc.2025.040101
ISSN 2616-1958 Vol. 4 Num. 1

Study on adsorption of modified biochar based on multiple regression model

Lingge Shi[1,#], Qifei Yan[1,#], Xinyang Li[2,#], Jianing Zhang[1,#]

[1]School of Pharmacy, Shenyang Pharmaceutical University, Shenyang, 110016, China
[2]School of Pharmaceutical Engineering, Shenyang Pharmaceutical University, Shenyang, 110016, China
[#]These authors contributed equally.

Keywords: Modified Biocarbon, Arsenic Ion (As (V)), Roxarsone (ROX), Ridge Regression, Multinomial Regression Model

Abstract: With the rapid development of industrialization and urbanization, industrial wastewater discharge, excessive use of agricultural fertilizers and pesticides, and improper waste disposal have all led to the intensification of water pollution, making the development of efficient and economical water treatment technologies increasingly important. Aiming at the adsorption effect of modified biochar on arsenic ions (As(V)) and roxarsone (ROX) in water bodies, a model was established and systematically analyzed based on polynomial regression theory. This study explores the effects of reaction temperature, solution pH, and adsorbent dosage, three key factors, on adsorption efficiency. By combining ridge regression with GridSearchCV, mathematical optimization techniques were employed to identify the conditions for maximum adsorption capacity. The optimal adsorbent dosage was found to be 0.2 g/L, the optimal temperature was 17.47 ℃, and the optimal pH value was 5.02. These findings provide a scientific basis and decision support for the application of modified biochar in water treatment, with significant environmental and economic implications.

1. Introduction

Due to natural processes and human emissions, a significant amount of waste is released into water bodies, posing severe threats to human and wildlife survival [1]. Common methods for treating arsenic and ROX in water currently include physical-chemical methods, chemical methods, biological methods, and adsorption methods [2]. Among these, adsorption is a relatively mature and straightforward wastewater treatment technique, with the preparation of efficient and economical arsenic adsorption materials being the technological core of this method.

In a study conducted by Benis K Z et al., it was discovered that under the experimental conditions, a hybrid adsorption mechanism predominantly involving heterogeneous multilayer chemisorption was more likely to occur [3]. This finding highlights the complexity of the adsorption process and the need for a deeper understanding of the underlying mechanisms. Sun Y et al., in their research, identified that the dosage of the adsorbent and the pH level of the solution are the key factors that influence the adsorption capacity of arsenic on modified biochar. They also found that the adsorption

process of arsenic onto biochar is endothermic, meaning it requires heat to proceed [4]. Furthermore, Kirmizakis P et al. assessed the effectiveness of iron-modified biochar in treating arsenic-laden simulated wastewater. They employed both standard chemical monitoring techniques and real-time monitoring of the adsorption process to evaluate its potential [5]. This comprehensive approach provides a robust evaluation of the biochar's performance in removing arsenic from water.

Previous studies, although having achieved certain progress, still have some shortcomings. Some research may rely too heavily on a single adsorption mechanism theory and fail to fully consider the complexity of the adsorption process. Additionally, previous studies may lack systematic approaches for optimizing adsorbents and precisely controlling adsorption conditions, leading to inaccuracies in predicting adsorption efficiency and stability. In contrast, this paper constructs an adsorption model based on polynomial regression theory, incorporating Ridge Regression and GridSearchCV, providing a solid theoretical foundation and data support for practical water treatment.

2. The model in this paper

2.1 Polynomial Regression

Polynomial fitting is a widely used tool in the field of mathematical modeling, which involves fitting a series of data points obtained through experiments or observations into a polynomial function to reveal the underlying relationships between these data points. The core of this method lies in approximating these data points using a polynomial function containing unknown parameters, aiming to gain a deeper understanding of the trends and patterns hidden behind the data.

In the process of polynomial fitting, we first assume that these data points are generated by a specific polynomial function that contains some yet-to-be-determined parameters. Subsequently, we employed the least squares method for fitting the data points, which involves seeking a set of parameter values that minimize the discrepancy between the polynomial function determined by these parameters and the actual data points. This method finds the optimal parameter values by minimizing the sum of squared errors, thus ensuring the accuracy of the fitting results.

Through this fitting process, we can determine the specific form of the polynomial function and obtain the best estimated values for each parameter. Ultimately, the polynomial function derived from this fitting can not only be used to predict the values of new data points but also help us analyze the trends and patterns within the data. This makes polynomial fitting play an important role in scientific research, engineering applications, and other fields requiring data analysis. By unveiling the relationships hidden within the data, polynomial fitting provides decision-makers with a powerful tool, enabling them to better understand complex phenomena and make more informed decisions based on this understanding.

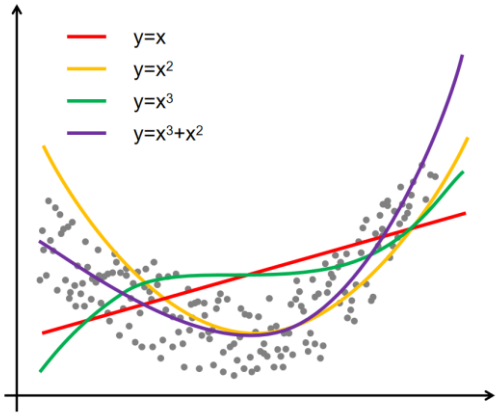

Figure 1 Schematic of Polynomial Fitting Principle

As shown in Figure 1, polynomial fitting that simultaneously considers multiple degrees of independent variables exhibits high accuracy. The general form of polynomial fitting is as shown in Equation (1).

$$g(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (1)$$

2.2 Ridge Regression and GridSearchCV

Ridge regression, also known as Tikhonov regularization, is a particularly effective technique for addressing regression problems, especially when dealing with issues caused by multicollinearity among features. This method introduces an L2 regularization term into the loss function, penalizing the size of the model coefficients, thereby reducing model complexity and effectively preventing overfitting [6]. The addition of the L2 regularization term not only enhances the model's generalization ability but also mitigates the impact of feature correlation, making the model more stable.

GridSearchCV is a widely used cross-validation technique in the field of machine learning. It systematically traverses through given parameter candidate combinations, training and validating each set of parameters to find the optimal hyperparameter settings. The core of this method lies in ensuring that each parameter combination is thoroughly evaluated during the model selection process, thereby enhancing the model's predictive performance and stability. By employing cross-validation, GridSearchCV not only reduces the risk of overfitting but also optimizes across multiple parameter dimensions, which is crucial for improving the model's generalization ability and adaptability. Through this approach, we can identify the hyperparameters that best suit the current data and problem from a large number of parameter combinations, thereby constructing a model with superior performance.

(2) The least squares method commonly used in regression analysis is an unbiased estimation. For a well-posed problem, X is typically column full-rank.

(3) By employing the least squares method, the loss function is defined as the square of the residuals, with the objective of minimizing this loss function.

(4) This optimization problem can be solved using the gradient descent method or directly using the following formula [7].

$$X\theta = y, \quad (2)$$

$$\|X\theta = y\|^2 \ . \quad (3)$$

$$\theta = (X^T X)^{-1} X_y^T. \quad (4)$$

3. Results

3.1 Polynomial Regression Model Results

We will explore the specific effects of reaction temperature, solution pH, and adsorbent dosage on the removal rates of As(V) and ROX. By establishing and validating generalized linear models (GLM), we will examine the main effects of each factor and their interactions to determine the influence of different factors on adsorption. We selected a quintic polynomial model for evaluation. Additionally, given the use of a feature model that includes polynomials, we will also refer to a quadratic polynomial model when interpreting the regression coefficients in practice.

A three-dimensional surface plot depicting the removal rates of As(V) and ROX at fixed pH is shown in Figure 2.

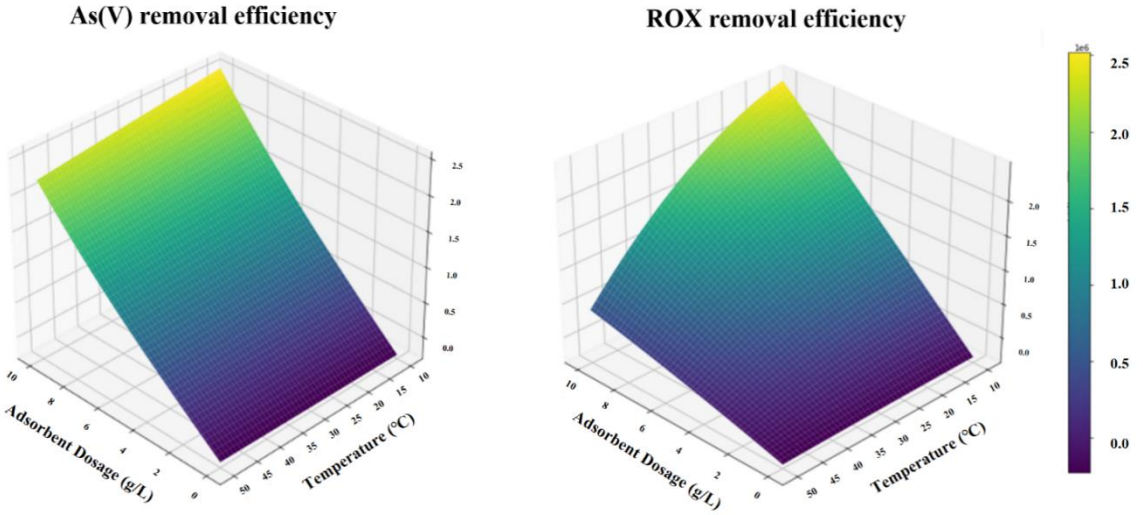

Figure.2. Three-dimensional surface of As (V) and ROX removal efficiency

As can be seen from the figure, the removal efficiency of pentavalent arsenic does not change much throughout the treatment process, showing a relatively stable state, which may mean that the removal efficiency is less affected by the treatment conditions, or the treatment technology has limited effect on pentavalent arsenic. In contrast, the removal efficiency of the ROX showed more significant fluctuations, especially under certain conditions, where the removal efficiency increased, suggesting that the removal effect of the ROX may be more sensitive to the treatment conditions.

Specific model parameters are shown in Table 1.

Table.1. Specific model parameters

Item	As(V) model coefficients	ROX model coefficients
Adsorbent Dosage	255850.618	2892543.52
Temperature	328.598791	-88.1166162
pH	-0.002961164	-0.017484137
Adsorbent dosage ×Temperature	6.22195928	-48.9782416
Adsorbent dosage ×pH	-778.507618	1028.95203
Temperature ×pH	2.9021034	21.1537008
Adsorbent dosage ^{2}	2.73536065	-111.010026
Temperature ^{2}	1.59246448	2.59883645
pH ^{2}	307.493516	-125.441753
Adsorbent dosage ×Temperature ×pH	-0.009174942	-0.277500053

These coefficients indicate:

1) The removal rate of As(V) is significantly affected by temperature, with each unit increase in temperature leading to an approximately 328.60% increase in removal rate. An increase in pH value slightly reduces the removal rate, but the nonlinear effect of pH has a positive impact on the removal rate. Increasing the dosage of adsorbent is very effective in enhancing the removal rate of As(V), and its positive impact strengthens with increasing dosage.

2) The removal rate of ROX decreases significantly with increasing temperature, with each unit increase leading to an approximately 88.11% reduction in removal rate, indicating that high temperatures may be unfavorable for ROX adsorption. The pH value has a minor effect on ROX removal efficiency, but the nonlinear effect has a significant negative impact. An increase in adsorbent dosage has a substantial positive effect on ROX removal efficiency; however, beyond a certain dosage,

the positive effect diminishes, possibly due to the saturation of adsorption sites.

3.2 Solving Multi-Objective Optimization

Based on the adsorption model established in Section 3.1, we have conducted optimizations using various algorithms to seek the optimal adsorption conditions for maximizing benefits.

Adsorption isotherms are curves that represent the relationship between the amount of adsorbate adsorbed by the adsorbent and the concentration of the substance at a constant temperature. Commonly used adsorption isotherm models include the Langmuir and Freundlich isotherm models [8].

Taking the Langmuir equation as an example, the adsorption capacity of each component can be expressed as:

$$Q_{AS(V)} = \frac{Q_{max,\ AS(V)} \times K_{L,\ AS(V)} \times m}{1 + K_{L,\ AS(V)} \times m} \tag{5}$$

$$Q_{ROX} = \frac{Q_{max,\ ROX} \times K_{L,\ ROX} \times m}{1 + K_{L,\ ROX} \times m} \tag{6}$$

where $Q_{max,\ AS(V)}$ and $Q_{max,\ ROX}$ is the maximum adsorption capacity of the Langmuir adsorption isotherm

where $K_{L,\ AS(V)}$ and $K_{L,\ ROX}$ is the adsorption coefficient of the Langmuir adsorption isotherm

Combining Equations 5 and 6, the total adsorption can be expressed as a function of the decision variables T, pH, and m.

$$Q_{total}(T, pH, m) = \frac{Q_{max,\ AS(V)} \times K_{L,\ AS(V)} \times m}{1 + K_{L,\ AS(V)} \times m} + \frac{Q_{max,\ ROX} \times K_{L,\ ROX} \times m}{1 + K_{L,\ ROX} \times m} \tag{7}$$

When designing this optimization adsorption experiment, not only should the maximization of the total adsorption capacity be pursued, but also the adsorption capacities of As(V) and ROX should be balanced. To this end, the coefficient of variation (CV) was introduced to weigh the adsorption capacities of As(V) and ROX.

First, calculate the standard deviation and mean of the adsorption amounts of As(V) and ROX separately, and then compute their coefficient of variation. Next, to balance the adsorption amounts of As(V) and ROX during the optimization process, the coefficient of variation is incorporated into the objective function. The mean, standard deviation, and coefficient of variation of As(V) and ROX are calculated separately to assess the key indicators of central tendency and dispersion in the dataset. The optimization results of different models, along with the mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2), are shown in Table 2.

Table.2. Model parameters

	MSE	MAE	R_2	Condition (Adsorbent dosage, temperature, pH)	Maximum adsorption capacity
RSM	10.287	2.522	0.934	0.20g, 15.00℃, 3.00	47.08
RSM+PSO	10.287	2.522	0.934	0.20g, 15.00℃, 3.00	47.08
Ridge Regression and GridSearchCV	9.609	2.353	0.939	0.20g, 15.00℃, 3.00	45.84
Random Forest and RandomizedSearchCV	43.42	3.778	0.557	0.20g, 15.00℃, 3.00	45.31

The comparison in the table indicates that Ridge Regression and GridSearchCV perform the best in maximizing total adsorption, with the highest R^2 value and lower error. The formula for total

adsorption is shown in Equation 8.

$$Q = 73.6476 - 126.3301m + 0.0363T - 0.287pH - 0.0035T^2 + 2.408pH^2 - 52.8201m^2 + 0.1063m \times T + 4.4693m \times pH + 0.0085T \times pH \quad (8)$$

Through multi-factor regression analysis of the adsorption process, the optimal adsorption conditions were determined to be an adsorbent dosage of 0.20g, a temperature of 15.00℃, and a pH value of 3.00. Under these conditions, the total adsorption capacity was maximized, reaching 47.08.

4. Conclusions and outlooks

We established a removal rate model through polynomial regression fitting. A systematic study on the effects of different conditions on adsorption revealed that higher temperatures promoted the adsorption of As(V) but inhibited that of ROX; an increase in pH adversely affected the removal rate of As(V) but had a lesser impact on that of ROX; an increase in adsorbent dosage significantly improved the removal rates of both As(V) and ROX, but as the dosage of adsorbent increased, the improvement in ROX removal rate began to diminish, indicating saturation. Furthermore, by comparing different optimization algorithm models, we identified the ridge regression and GridSearchCV total adsorption optimization models. Based on this, the optimal conditions were determined: an adsorbent dosage of 0.20 grams, a temperature of 15.00 ℃, and a pH of 3.00. Under these conditions, the total adsorption capacity reached a maximum of 47.08. This model was established based on the systematic collection and analysis of a large amount of experimental data, ensuring the adequacy and representativeness of the data. This not only improved the prediction accuracy of the model but also enhanced its reliability and credibility in practical applications.

When attempting to improve the fit by using higher-order polynomials, there is a risk of overfitting. We can limit the complexity of the model through L1 and L2 regularization, which helps prevent the model parameters from becoming too large during training, thereby reducing the risk of overfitting. Additionally, training and validating the model on different subsets of data can provide a more accurate estimate of its performance on unseen data. In the future, the model could also be extended to include more types of pollutants, studies on multi-pollutant interactions, targeted optimization, and other areas. Provides a solid theoretical foundation and data support for the promotion of modified biochar in practical water treatment applications.

References

[1] Wang C Y, Zhang X W, Wang J W, et al. A new one-dimensional coordination polymer synthesized from zinc and guanazole: Superior capture of organic arsenics [J]. Applied Organometallic Chemistry, 2020, 34(6): e5637.

[2] Lim K T, Shukor M Y, Wasoh H. Physical, chemical, and biological methods for the removal of arsenic compounds[J]. BioMed research international, 2014, 2014(1): 503784.

[3] Benis K Z, Soltan J, McPhedran K N. Electrochemically modified adsorbents for treatment of aqueous arsenic: Pore diffusion in modified biomass vs. biochar [J]. Chemical Engineering Journal, 2021, 423: 130061.

[4] Sun Y, Yu F, Han C, et al. Research progress on adsorption of arsenic from water by modified biochar and its mechanism: A review [J]. Water, 2022, 14(11): 1691.

[5] Kirmizakis P, Tawabini B, Siddiq O M, et al. Adsorption of arsenic on Fe-modified biochar and monitoring using spectral induced polarization[J]. Water, 2022, 14(4): 563.

[6] Hoerl R W. Ridge regression: a historical context [J]. Technometrics, 2020, 62(4): 420-425.

[7] Tsigler A, Bartlett P L. Benign overfitting in ridge regression [J]. Journal of Machine Learning Research, 2023, 24(123): 1-76.

[8] Rajahmundry G K, Garlapati C, Kumar P S, et al. Statistical analysis of adsorption isotherm models and its appropriate selection[J]. Chemosphere, 2021, 276: 130176.