Desalination and Water Treatment
www.deswater.com

doi:10.5004/dwt.2019.23647

145 (2019) 107–119
March

Comparison of regression and design models for biosorption process

Seda Çetintaş[a], Sermin Elevli[b], Deniz Bingöl[b,*]

[a]*Kocaeli University, Department of Chemistry, Kocaeli, Turkey, Tel. +90 262 303 20 27, email: sdctntsd@gmail.com (S. Çetintaş), Tel. +902623032030, Fax +902623032003, email: denizbingol1@gmail.com, deniz.bingol@kocaeli.edu.tr (D. Bingöl)*
[b]*Ondokuz Mayıs University, Department of Industrial Engineering, Samsun, Turkey, Tel. +90 (362) 312-1919, email: sermin.elevli@omu.edu.tr (S. Elevli)*

Received 3 April 2018; Accepted 22 December 2018

ABSTRACT

In this study, the biosorption process were optimized and compared for the first time in terms of their accuracy and predictive ability for the sorption of Cu(II) ions onto date palm (*Phoenix dactylifera* L.) seeds used as an agricultural waste product by using four models, multi-linear regression (MLR), full factorial design with center points (FFD), Box-Behnken design (BBD) and central composite design (CCD). The responses were evaluated based on the regression equations formulated according to the results of the analyses of models. It was found that MLR and FFD models had a lower predictive capability than response surface methodologies (RSM).

Keywords: Biosorption; Box-Behnken design (BBD); Central composite design (CCD); Full factorial design (FFD); Multi linear regression (MLR)

1. Introduction

Heavy metals that accumulate in soil, water, and air environments pose a significant hazard to living things that use or live in the environment by contamination through food and water. This accumulation causes various diseases, as well as biological and genetic disorders [1]. In certain systems, the mechanism of heavy metals varies depending on their concentration. They are only toxic when they exceed a certain limit. On the other hand, the activity of heavy metals is not only affected by their concentration in living bodies, but it is also dependent on the structure of the metal ion and living specimen, such as solubility, chemical structure, redox and complex forming capability, body uptake, frequency in the environment, and local pH [2]. Copper is a heavy metal which has a high toxic effect and spreads into the environment through the wastewater from various industries, such as metal cleaning and plating baths, refineries, paper and pulp, fertilizer, and wood preservatives [3]. Divalent copper (Cu(II)) is carcinogenic and is deposited in the liver, causing several symptoms, such as headache, nausea, vomiting, respiratory difficulty, and abdominal pain, as well as failure of the liver and kidneys. It can even result in gastrointestinal bleeding if excess amount is ingested [4]. To prevent health-related problems, the maximum contaminant level of copper was determined by the Environmental Protection Agency as 1.5 mg/l or 1.5 ppm [5].

In the treatment of wastewater, adsorption/biosorption presents as a feasible, easy-to-operate and economical alternative to traditional methods, such as precipitation, membrane filtration, electrolyte or liquid extraction, electrodialysis, and reversed osmosis since it utilizes adsorbents with a lower cost and causes less sludge disposal problems [6].

In the literature, numerous applications of sorption have been successfully implemented to solve environmental problems related to the removal of Cu(II) ions from wastewater. Many of these solutions are based on experimental designs that aim to provide optimum heavy metal removal both economically and in a shorter time. Algal and microbial biomass and agricultural waste products have been effectively used as adsorbents to remove Cu(II) from aqueous solutions; e.g., *Carica papayaL.* [6], *Penicillium* [7], *Trametes versicolor* [8], *Ulva fasciata*[9], *Enteromorpha prolifera* [10], *Bacillus brevis* [11], *Trichoderma viride*[12], clay [13],

*Corresponding author.

pumice [14], chitosan [15], rice husk [16], rice straw [17], chestnut shell [18], coconut tree sawdust, eggshell, sugarcane bagasse [19], Durian tree sawdust, coconut coir; oil palm empty fruit bunch [20], orange peel [21], almond shell [22], kenaf fiber [23], olive stone [24], coconut shell, coconut husk, sawdust, Moringa oleifera seeds [25], and sunflower shells [26].

In the present study, the sorption process was first optimized using the MLR, FFD with center points, BBD and CCD models, and then the accuracy and predicting capabilities of these models for Cu(II) ion sorption were assessed and compared using the seeds of date palm as an agricultural waste product. For all the methods, regression equations were also formulated to demonstrate the effect of key parameters.

2. Experimental

2.1. Material

The chemical content and physical characteristics of the palm kernel on a dry-weight basis were determined in a previous study as follows; 22.61% carbohydrate, 6.43% lipid, 54.35% fiber, 0.97% ash, 4.94% protein, 10.70% moisture, 0.03% Ca, 0.12% P, 0.08% Mg and 1.68 kcal/g energy [27]. Furthermore, according to the Fourier Transform Infrared (FTIR) spectra, the functional groups were identified as hydroxyl, carboxyl and carboxylate [27].

Copper sulfate ($CuSO_4$; AR grade) was used to prepare a stock solution of Cu(II) (1.0 g/l) in deionized water. The 5–100 mg/l Cu (II) solutions used in the experiments were prepared by diluting from this stock solution. The chemicals that were used were of analytical reagent grade. All glassware was washed with dilute HNO_3 and rinsed with deionized water.

2.2. Batch sorption studies

Experiments on the biosorption of the aqueous solution of Cu(II) ions with palm seeds were performed according to the experimental designs of the models used. The solutions were prepared at an appropriate pH (2–6), initial Cu(II) concentrations (5 to 100 mg/l) and amount of adsorbent ranging from 0.050 to 0.500 g depending on the conditions of each experiment. For batch experiments, 50 ml of Cu(II) solution was magnetically stirred at 20°C at 300 rpm until equilibrium time were achieved, which were determined as 60 min in preliminary studies. To remove the biosorbent from the solution, centrifuging (Elektromag M815P model) was performed at 1000 rpm for 5 min, followed by filtering with a Whatman paper (No. 42). A flame atomic absorption spectrophotometer (Perkin Elmer model AAnalyst 800) was used to analyze the amount of Cu(II) ion that remained in the supernatant. The absorbance was found to be linear for 1–5 mg/l of standard Cu(II) solutions and the correlation coefficient was calculated as 0.999. The amount of sorbed-Cu(II) (q_e, mg/g) and sorption efficiency of Cu(II) (R, %) were calculated by the following equations:

$$q_e(mg/g) = (C_0 - C_e) \times \frac{V}{M} \tag{1}$$

$$R(\%) = \frac{(C_0 - C_e)}{C_0} \times 100 \tag{2}$$

where C_o and C_e are the initial and equilibrium concentrations of Cu(II) ion (mg/l), respectively; V refers to the volume of the Cu(II) solution (l); and W indicates the mass of the seed sample (g).

2.3. Optimization methods

In the present study, MLR, FFD with center points, BBD and CCD were used to determine the optimum process conditions in a sorption process. Tests were performed to investigate the factors, pH, initial Cu(II) ion concentration and biosorbent mass, affecting Cu(II) sorption onto date palm seeds.

The coefficients in a model equation are generally determined by an MLR analysis, which models the relationship between experimental and response variables based on the following equation for a given number of p observations:

$$y_i = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon \tag{3}$$

where y_i indicates the predicted response; β_o is the intercept of the plane, the β_1 and β_2 parameters are the partial regression coefficients; $x_i (i = 1,\ldots,p)$ are experimental variables; and ε is a random or an unexplained error [27,28].

In the conventional one-variable-at-a-time (OVAT) approach, 20 experiments are performed with the MLR analysis to evaluate the statistical significance of the effect of initial Cu(II) ion concentration, biosorbent mass and pH on the sorption process.

Factorial designs are particularly important in experiments that involve multiple factors in terms of providing data about the cooperative effect of each factor on the response. A 2^k design is usually preferred in early stages of an experiment when it is possible to investigate several factors. This design gives the minimum number of runs with k number of factors that can be studied in an FFD. Since there are only two levels for each factor, we expected that the response would be approximately linear in the selected ranges [29]. Generally, the curvilinear effect on the response is so strong that the first-order model (even with the interaction term) is not adequate. In order to evaluate the curvilinear effect, there is a need for a polynomial function with quadratic terms. In this instance, the experimental design assumes that all the variables have at least three factors [30].

In a 2^k FFD, each factor level is investigated to test all the combinations based on the following regression equation:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{1 \le i \le j}^{k} \beta_{ij}\, x_i x_j + \varepsilon \tag{4}$$

where y is the response; x_i and x_j refer to the coded variables; βs indicate the regression coefficients; and ε is a random error. We used a 2-level FFD to evaluate the statistical significance of the effect and interactions of pH (X_1), initial Cu(II) ion concentration (X_2), and biosorbent mass (X_3) and curvature on the batch sorption process. The levels of experimental factors chosen for FFD were coded as −1 and +1, respectively (Table 1). 2^3 experiments were performed for eight runs (in duplicate). The experiments were carried

Table 1
The levels of experimental factors chosen for FFD, BBD and CCD

Factors	*Symbols*	Levels of factors (FFD)				
		–1		+1		
Initial pH	X_1	2.5		5.5		
Initial Cu(II) concentration, C_o (mg/l)	X_2	20		80		
Biosorbent mass, *m* (mg)	X_3	100		400		
Factors	*Symbols*	Levels of factors (BBD)				
		–1	0			+1
Initial pH	X_1	2	4			6
Initial Cu(II) concentration, C_o (mg/l)	X_2	20	60			100
Biosorbent mass, *m* (mg)	X_3	50	275			500
Factors	*Symbols*	Levels of factors (CCD)				
		$-\alpha$(–1.68)	–1	0	+1	$+\alpha$(+1.68)
Initial pH	X_1	2	3	4	5	6
Initial Cu(II) concentration, C_o (mg/l)	X_2	20	36	60	84	100
Biosorbent mass, *m* (mg)	X_3	50	141	275	409	500

out by selecting factor levels according to the data obtained from preliminary experiments. Furthermore, the curvature was assessed using a center points.

If the curvature is statistically significant ($P < 0.05$) at the 5% probability level, special designs, such as BBD and CCD that allow the use of second-order models are required. The actual values of the factors were selected as −1, 0 and +1 for the coded values in BBD and as –α (–1.68), −1, 0, +1 and +α (+1.68) for those in CCD (Table 1). The following second-order polynomial equation was used to identify all the possible interactions of the selected factors obtained from the models:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{1 \le i \le j}^{k} \beta_{ij}\, x_i x_j + \varepsilon \quad (5)$$

where $x_1, x_2, \ldots x_k$ are the input variables with an impact on the response y; β_0, β_i ($i = 1, 2, \ldots k$); β_{ii} and β_{ij} ($i = 1,2, \ldots k$; $j = 1,2, \ldots k$) refer to the intercept, linear, quadratic and interaction constant coefficients, respectively; and ε is a random error. The selected independent variables X_i were coded as x_i based on the following relationship:

$$x_i = (X_i - X_o)/\Delta x \quad (6)$$

where X_o is the uncoded value of X_i at the center and Δx represents the step change [29].

3. Results and discussion

The data were analyzed using the coefficients of determination (R^2), response surface plots, and analysis of variance (ANOVA) in the Minitab 16 Statistical Software. The *F* and probability (*P*) values were estimated using the coded values obtained from the models with the former being the ratio of the mean square of the parameter to that of the error term and the latter being obtained from the *F* distribution. The *P* value is defined as the smallest level of significance that results in the rejection of a null hypothesis [31]. At 5% significance, a regression model is accepted when either the *F* value is higher than the corresponding *F* critical value or the *P* value is lower than 0.05. The *F* critical values based on the degrees of freedom of the model were taken from the literature [29]. Each variable term in the model was evaluated in terms of their degree of significance based on the size of their regression coefficient. A higher regression coefficient of a variable term generally indicates that the term is more significant [32].

The R^2value is defined as the ratio of the variability in the data "explained" by the ANOVA model. Suitable in $0 \le R^2 \le 1$ range, but larger values are more desirable. Other statics similar to R^2 are also displayed in the output. The "adjusted" R^2 is a variation of ordinary R^2 statistics reflecting the number of factors in the model. It may be a useful statistic to assess the effect of increasing or decreasing the number of model terms for more complex experiments with various design factors. Before accepting ANOVA's results, the adequacy of the recommended model should be checked. As before, the residue analysis is the primary diagnostic tool [29].

A very useful procedure is to generate a normal probability plot to control the assumption of normality. In ANOVA, it is more efficient (and simpler) to do so with residues. If the emphasize error distribution is normal, the plot will look like a straight line. When visualizing the straight line, give more importance to the central values of the plot than to the extremes. The general impression from examining this display is that the error distribution is approximately normal. The residuals plots to determine the deviation and express in terms of percent deviation. The residual plots show whether the residues are in normal distribution or not. If residues do not show normal distribution, the reliability of *P* values obtained from ANOVA decreases. To verify the assumptions, it is tested whether the distribution of residues is normal and whether residues have changed [29].

In addition, the statistical analyses of models are also performed by adding significance parameters such as variance inflation factor (VIF) for multicollinearity, and root mean square error (RMSE) is used to measure the differences between values predicted by a model and the values observed.

Multicollinearity, also called collinearity, is usually occurs when two or more predictive variables are associated in applied models. Models with collinearity have predictors with lower precision that cause problems in testing hypotheses and estimating [33]. Variance inflation factors (VIFs) quantify the severity of multicollinearity in an ordinary least squares regression analysis, and are used to detect multicollinearity among predictors. VIF can be calculated through the following formula to check for the multicollinearity between explanatory variables:

$$VIF = 1/\left(1 - R^2\left(adj\right)\right) \tag{7}$$

If VIF = 1, there is no correlation, if VIF is more than 5 but less than 10, there is moderate correlation and if VIF is greater than 10, there is high correlation. The common rule of thumb is that VIF should be less than 3 [34].

RMSE is the square scoring rule that measures the average size of the error and is calculated by taking the square root of the mean squared errors between the estimation and the actual observation [Eq. (8)]. RMSE increases as the variance associated with the frequency distribution of the error magnitudes increases. The lower values of RMSE are better [35].

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j + \hat{y}_j\right)^2} \tag{8}$$

where y_j is observed values and $\acute{y}_j$ is modelled values.

Response surface plots provide a better understanding of how each variable affects and interacts with the response. Each plot visualizes the effect of two variables within the selected range with a third variable being kept at a specific value.

3.1. MLR

According to ANOVA results summarized in Table 2, the models were fitted for q_e and R with adjusted-R^2 of 88.26% and 70.02%, and standard deviations of 0.57 and 14.64, respectively. There was a low correlation between the observed results and predicted values. In addition, the models were not able to account for 11.74% and 29.98% (residuals) of the total variation.

The following regression equations obtained from MLR model show the relationship between the factors and the responses as q_e and R.

$$q_e\,(\text{mg/g}) = -1.28 + 0.78\ \text{pH} + 0.02\ C_o - 6.35\ \text{m} \tag{9}$$

$$\text{R}\ (\%) = -26.00 + 16.17\ \text{pH} - 0.64\ C_o + 5.50\ \text{m} \tag{10}$$

A graphical presentation of the models are given in Fig. 1. The response surface graphs in Fig. 1 show that the MLR model can represent only the linear effects of all factors on the sorption process.

3.2. FFD

FFD is assumed that there is a linear relationship when examining two levels of factors. However, this assumption is incorrect if the relationship between two points is curvilinear. It is very useful to investigate the curvature in experimental designs. In doing so, 3-level instead of 2 full factorial is not used. By adding midpoints in 2^k designs, the

Table 2
ANOVA for (a) q_e (mg/g) and (b) R (%) based on MLR at coded units

(a)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Regression	3	22.859	7.6198	23.56	0.001	
pH	1	6.643	6.6428	20.54	0.004	1.35
C_o (mg/l)	1	3.362	5.5949	10.39	0.018	1.31
m (g)	1	5.595	3.3615	17.30	0.006	1.24
Error	6	1.940	0.3234			
Total	9	24.800				

S = 0.57 R-Sq = 92.18% R-Sq(pred) = 65.37% R-Sq(adj) = 88.26%

(b)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Regression	3	5145.01	1715.00	8.01	0.016	
pH	1	1829.31	2829.31	13.21	0.011	1.35
C_o (mg/l)	1	2984.74	2984.74	13.93	0.010	1.31
m (g)	1	4.20	4.20	0.02	0.893	1.24
Error	6	1285.40	214.23			
Total	9	6430.41				

S = 14.64 R-Sq = 80.01% R-Sq(pred) = 47.26% R-Sq(adj) = 70.02%

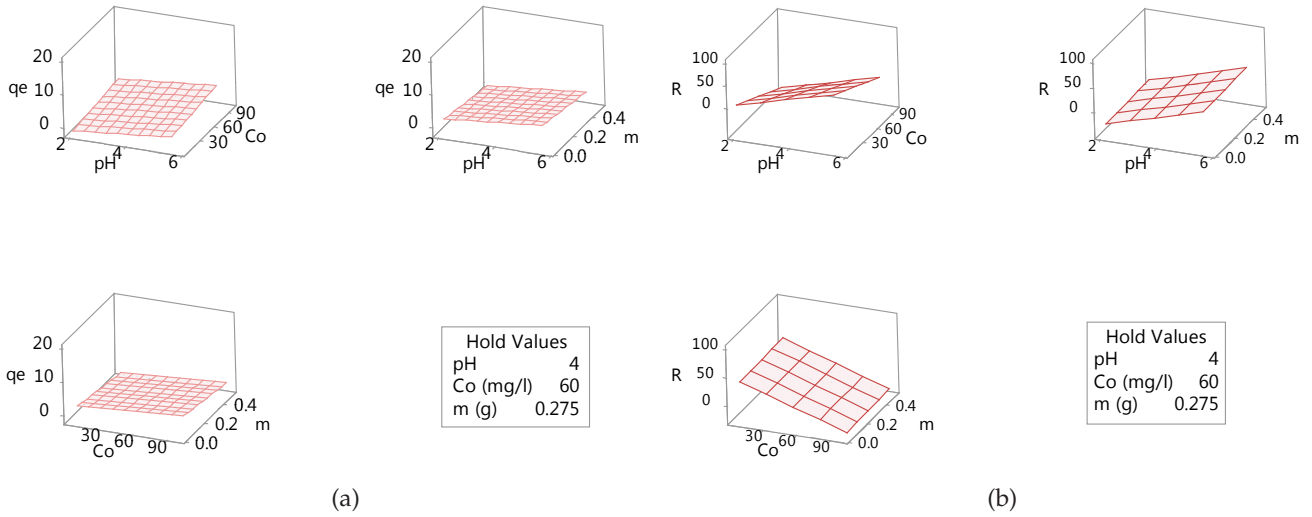

Fig. 1. Response surface plot based on (a) q_e and (b) R for MLR.

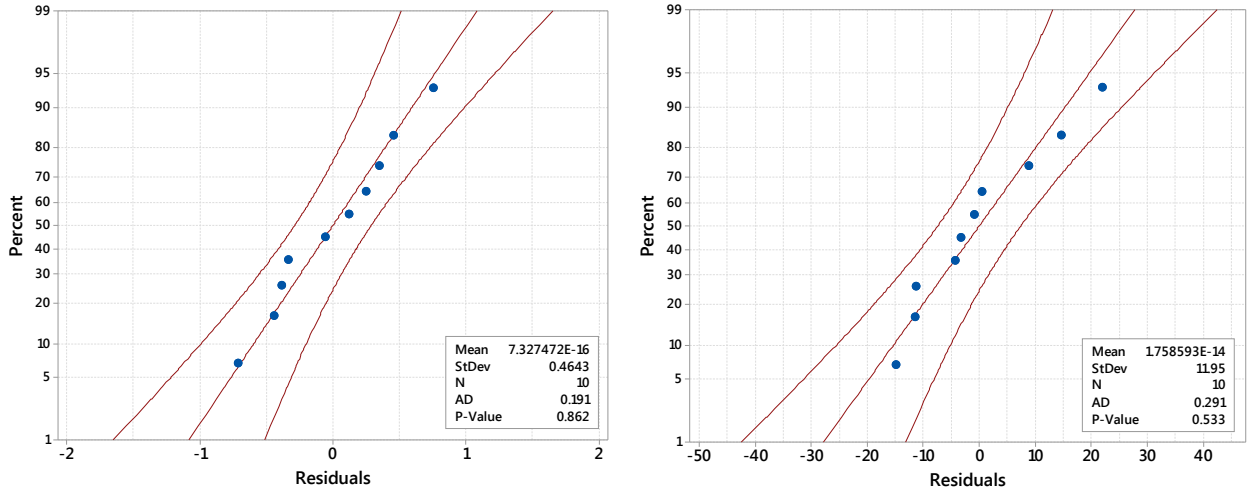

Fig. 2. Probabilitiy plots based on (a) q_e and (b) R for MLR.

number of experiments is designed with fewer experiments than 3 levels. The points used to test the system for curvature are called as center points.

ANOVA results of FFD with center points in Table 3 show that the effects and interactions of all experimental factors were indicated to be highly significant ($P < 0.05$) for the sorbed amounts of Cu(II) (q_e) and sorption efficiency (R), at 5% probability. In Table 3b, P-value indicates that 3-way interactions are significant. Therefore, the lower-level interactions were not excluded from the model. In Table 3, the curvature seemed to be an important factor, which means that the linear model was not sufficient for this region and a higher-order model was required. The presence of curvature indicates that the experiment region may be close to the optimum.

The adjusted-R^2 values with standard deviations as indicators of good fit of the models were calculated as 99.98%±0.04 and 99.22%±1.77 for q_e and R, respectively. Only 0.02% for q_e and 0.78% for R of the total variation could not be explained.

$$q_e\,(\mathrm{mg/g}) = -5.50 + 2.07\ \mathrm{pH} + 1.24C_o - 2.48\ \mathrm{m} + 0.35\ \mathrm{pH}^*C_o - 0.91\ \mathrm{pH}^*\mathrm{m} - 0.55\ C_o{}^*\mathrm{m} \quad (11)$$

$$\mathrm{R}(\%) = 42.51 + 16.32\ \mathrm{pH} - 7.91\ C_o + 1.94\ \mathrm{m} - 3.49\ \mathrm{pH}^*C_o \quad (12)$$

There was a high correlation between the experimental results and the predicted values by Eqs. (11) and (12). In FFD models, the center points were ignored to draw the response surface plots. According to response surface plots Fig. 3, the sorbed amounts of Cu(II) increase linearly as pH and C_o increase, and decreases as m increases. The sorption efficiency increase with increasing pH and m, decreases with increasing Co.

The normal probability plot of residuals is shown in Fig. 4. The residual plots show that the residual is normally distributed.

Table 3
ANOVA for (a) q_e (mg/g) and (b) R (%) based on FFD with center points at coded units

(a)						
Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	8	243.832	30.4791	298.58	0.000	
Linear	3	191.418	63.8061	625.05	0.000	
pH	1	68.683	68.6827	672.82	0.000	1.00
C_o	1	24.676	24.6761	241.73	0.000	1.00
m	1	98.060	98.0595	960.60	0.000	1.00
2-Way Interactions	3	19.927	6.6424	65.07	0.000	1.00
pH*C_o	1	1.925	1.9252	18.86	0.000	1.00
pH*m	1	13.195	13.1951	129.26	0.000	1.00
C_o*m	1	4.807	4.8071	47.09	0.000	1.00
3-Way Interactions	1	0.030	0.0298	0.29	0.596	
pH*C_o*m	1	0.030	0.0298	0.29	0.596	1.00
Curvature	1	32.457	32.4572	317.95	0.000	
Error	19	1.940	0.1021			
Total	27	245.772				
Values for the model with significant coefficients						
S = 0.32 R-Sq = 99.21% R-Sq(pred) = 98.88% R-Sq(adj) = 97.19%						
Values for the reduced model with significant coefficients						
S = 0.04 R-Sq = 99.99% R-Sq(pred) = 99.96% R-Sq(adj) = 99.98%						
(b)						
Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	8	10850.7	1356.34	431.80	0.000	
Linear	3	5320.2	1773.40	564.57	0.000	
pH	1	4259.5	4259.52	1356.05	0.000	1.00
C_o	1	1000.8	1000.77	318.60	0.000	1.00
m	1	59.9	59.91	19.07	0.000	1.00
2-Way Interactions	3	200.3	66.78	21.26	0.000	
pH*C_o	1	194.6	194.60	61.95	0.000	1.00
pH*m	1	4.6	4.64	1.48	0.239	1.00
C_o*m	1	1.1	1.09	0.35	0.562	1.00
3-Way Interactions	1	13.8	13.84	4.41	0.049	
pH*C_o*m	1	13.8	13.84	4.41	0.049	1.00
Curvature	1	5316.3	5316.33	1692.49	0.000	
Error	19	59.7	3.14			
Total	27	10910.4				
Values for the model with significant coefficients						
S = 1.77 R-Sq = 99.45% R-Sq(pred) = 98.48% R-Sq(adj) = 99.22%						

3.3. BBD

Table 4 presents the ANOVA results with the predicted F and P values for q_e and R. In the BBD models, at the 5% probability level, the linear effects of experimental factors and the square interactions indicating curvature were determined to be statistically significant ($P < 0.05$) for the sorbed amounts of Cu(II) (q_e) and sorption efficiency (R). The BBD results were evaluated according to adjusted-R^2 values obtained from ANOVA analyses and response surface plots.

According to the results, the following regression models were developed, giving the relationship between the responses (q_e and R) and all the variable terms in coded units:

$$q_e\,(\text{mg/g}) = 8.13 + 2.48\,\text{pH} + 4.37\,C_o - 6.53\,\text{m} - 6.22\,\text{pH}^2 + 4.20\,\text{m}^2 - 5.61\,C_o{*}m \quad (13)$$

$$R\,(\%) = 64.84 + 23.19\,\text{pH} - 10.52C_o + 10.32\,\text{m} - 30.59\,\text{pH}^2 - 9.83\,\text{m}^2 - 7.97\,\text{pH}{*}C_o + 13.43\,\text{pH}{*}m \quad (14)$$

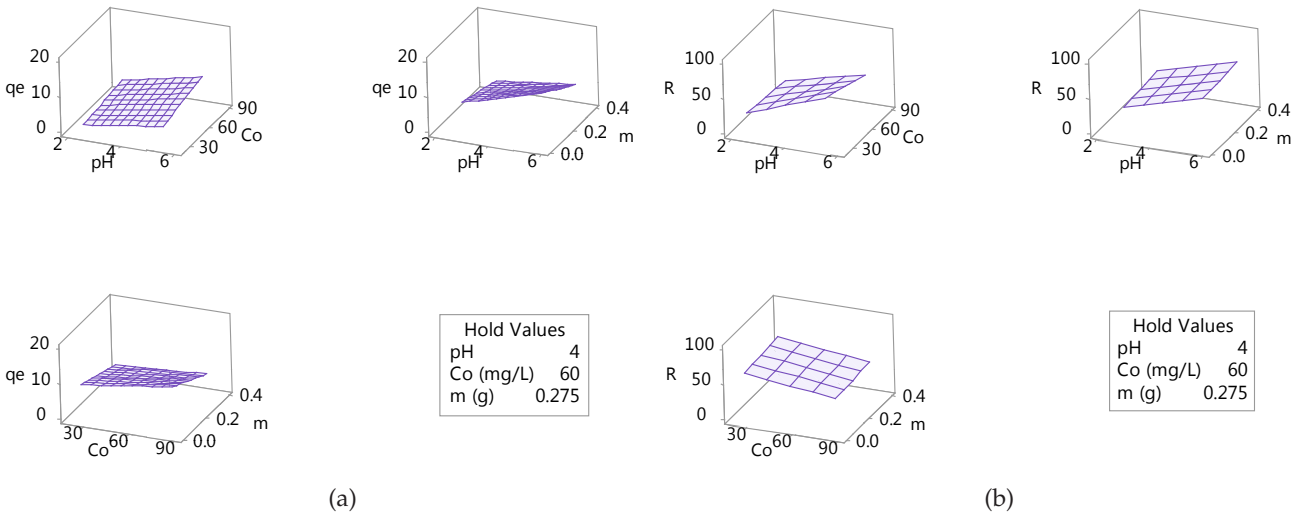

Fig. 3. Response surface plot based on (a) q_e and (b) R for FFD.

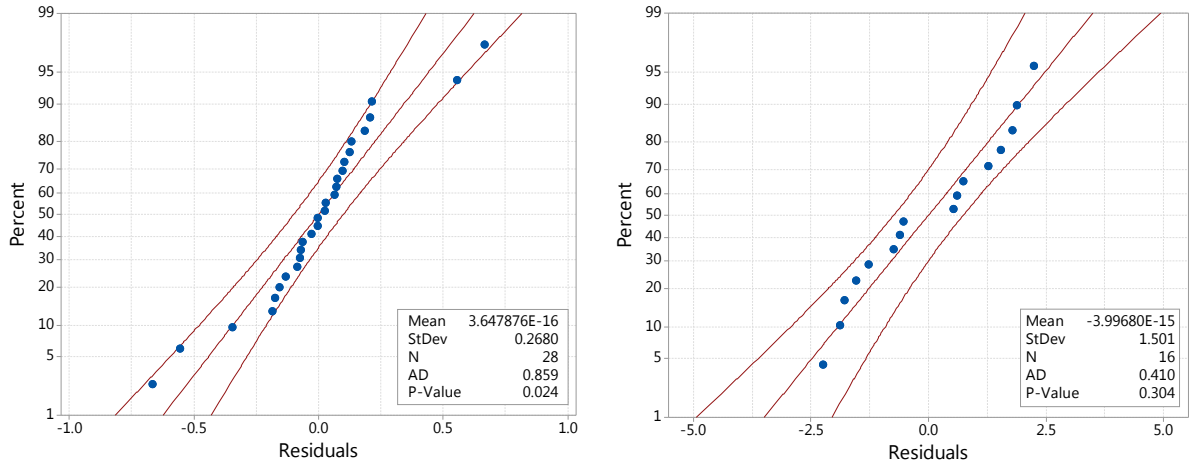

Fig. 4. Probabilitiy plots based on (a) q_e and (b) R for FFD.

In Eqs. (13) and (14), the positive signs of coefficients for the main effects and interactions of factors demonstrate their synergistic effect on q_e and R, while the negative signs denote their antagonistic effects. Based on their effect on q_e and R the order of decreasing significance of each variable term were as follows; $m \sim pH^2 > m^2 \sim C_o > pH$ and $pH^2 > pH > pH{*}m > C_o \sim m > m^2 > pH{*}C_o$, respectively. According to q_e and R, ANOVA results showed that the most important variables for the biosorption process were sorbent amount and pH, respectively. These observations are confirmed by the response surface plots presented in Fig. 5. The high similarity between the experimental results and the predicted values by Eqs. (13) and (14) show a good fit of the BBD model, with the adjusted-R^2 of 95.69% and 99.43% the standard deviation being 1.66 and 2.06. In addition, the probability plots for q_e and R indicated that these models explained 95.69% and 99.43% of the total variation, respectively.

The residual plots in Fig. 6 shows that the residues are normally distributed.

3.4. Central composite design (CCD)

Table 5 presents the results of ANOVA with F and P values estimated by CCD for q_e and R. At the 5% probability level, the linear and square coefficients were found to be highly significant ($P < 0.05$), except for 2-way interaction.

The relationship between the responses and the factors can be presented using the coded variables as follows:

$$q_e\,(\text{mg/g}) = 7.77 + 2.01\,pH + 1.45\,C_o - 4.30\,m - 2.05\,pH^2 - 1.15\,C_o^2 + 2.10\,m^2 \quad (15)$$

$$R\,(\%) = 70.62 + 16.88\,pH - 9.36\,C_o + 3.27\,m - 14.93\,pH^2 - 2.19\,C_o^2 - 7.68\,m^2 - 3.49\,pH{*}C_o \quad (16)$$

The experimental results were highly correlated with the values predicted by Eqs. (15) and (16). Furthermore, the adjusted-R^2 of 97.67% and 99.96% for q_e and R confirmed that the model was a good fit, respectively. The percentage

Table 4
ANOVA for (a) q_e (mg/g) and (b) R (%) based on BBD at coded units

(a)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	9	1849.67	205.519	6.55	0.000	
Linear	3	1086.77	362.257	11.54	0.000	
pH	1	98.03	98.029	3.12	0.092	1.00
C_o (mg/l)	1	305.61	305.608	9.74	0.005	1.00
m (g)	1	683.13	683.134	21.76	0.000	1.00
Square	3	498.10	166.035	5.29	0.008	
pH*pH	1	267.49	267.488	8.52	0.008	1.01
C_o (mg/l)*C_o (mg/l)	1	49.94	49.943	1.59	0.222	1.01
m(g)*m(g)	1	142.83	142.830	4.55	0.045	1.01
2-Way Interaction	3	264.80	88.266	2.81	0.066	
pH*C_o(mg/l)	1	10.65	10.654	0.34	0.567	1.00
pH*m(g)	1	2.65	2.650	0.08	0.774	1.00
C_o(mg/l)*m(g)	1	251.49	251.493	8.01	0.010	1.00
Error	20	627.74	31.387			
Lack-of-Fit	3	360.92	120.308	7.67	0.002	
Pure Error	17	266.82	15.695			
Total	29	2477.41				

Values for the model with significant coefficients
S = 5.60 R-Sq = 74.66%R-Sq(pred) = 35.14%R-Sq(adj) = 63.26%
Values for the reduced model with significant coefficients
S = 1.66R-Sq = 96.58%R-Sq(pred) = 94.24%R-Sq(adj) = 95.69%

(b)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	9	21510.6	2390.06	36.74	0.000	
Linear	3	12079.5	4026.49	61.90	0.000	
pH	1	8604.2	8604.19	132.27	0.000	1.00
C_o (mg/l)	1	1770.9	1770.94	27.22	0.000	1.00
m (g)	1	1704.3	1704.35	26.20	0.000	1.00
Square	3	7388.7	2462.91	37.86	0.000	
pH*pH	1	6930.9	6930.86	106.55	0.000	1.01
C_o (mg/l)*C_o (mg/l)	1	3.1	3.13	0.05	0.829	1.01
m(g)*m(g)	1	720.6	720.60	11.08	0.003	1.01
2-Way Interaction	3	2042.4	680.79	10.47	0.000	
pH*C_o(mg/l)	1	508.7	508.73	7.82	0.011	1.00
pH*m(g)	1	1443.6	1443.64	22.19	0.000	1.00
C_o(mg/l)*m(g)	1	90.0	90.01	1.38	0.253	1.00
Error	20	1301.0	65.05			
Lack-of-Fit	3	477.5	159.17	3.29	0.046	
Pure Error	17	823.5	48.44			
Total	29	22811.6				

Values for the model with significant coefficients
S = 8.07R-Sq = 94.30%R-Sq(pred) = 85.44%R-Sq(adj) = 91.73%
Values for the reduced model with significant coefficients
S = 2.06 R-Sq = 99.57%R-Sq(pred) = 99.28%R-Sq(adj) = 99.43%

of residuals was only 2.33% *and* 0.04% for q_e and R, respectively. The significance of each variable term regarding their effect on q_e and R were in the following decreasing order: m > m^2 ~ pH ~ pH^2 > C_o > Co^2 and pH > pH^2 > C_o > m^2 > pH*C_o > m > Co^2, respectively. The response surfaces were plotted to visualize the effects of three factors on q_e and R (Fig. 7). According to the results, when the sorption improved at the pH range of 2 to 4, decreases after pH 4. In Fig. 8, the residuals show normal distribution.

3.5. Comparison of models

Screening methods, in particular the two-level FFD, are used to determine the factors with a significant effect

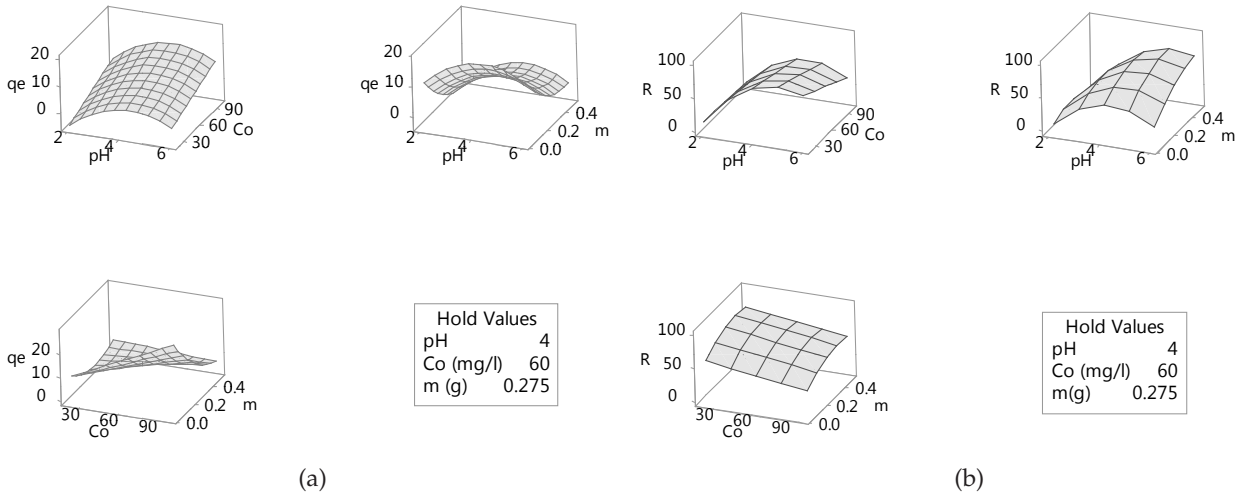

Fig. 5. Response surface plot based on (a) q_e and (b) R for BBD.

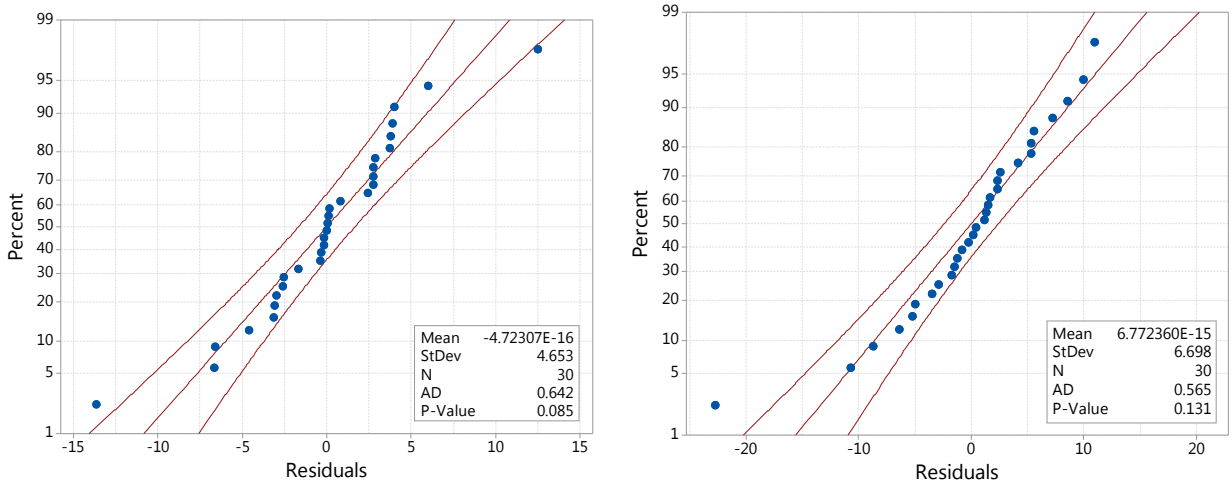

Fig. 6. Probabilitiy plots based on (a) q_e and (b) R for BBD.

on a response and the level of that significance. RSM are multivariable techniques that fit mathematically with a response function on the selected experimental region examined in the theoretical design. Among these designs, CCD and BBD are commonly preferred [36]. The former is based on FFD (N= 2^k), in which center points, axial or start points are represented by a star allowing the estimation of the curvature BBD; however, it is not directly based on an FFD in terms of using center points rather than corner points [36]. This design is limited to cases, in which the experimenter is not interested in predicting the response at extreme points (corners of the cube) [36]. BBD requires fewer experiments (N = $2k(k–1)+C_o$) than CCD (N= 2^k+2k+C_o) but covers a slightly smaller experimental region (N: experimental points, k: number of variables and C_o: number of center points) [37,38].

In this study, the accuracy and the predicting capability of MLR, FFD with center points, BBD and CCD for the sorption of Cu(II) ions were evaluated and compared. The difference in adjusted R^2 and RMSE values between the experimental and predicted responses were used to assess and compare the accuracy of the models (Table 6). Among the models, CCD model had capable of best estimation due to high R^2 and low RMSE values for the biosorption process. This was also confirmed by the results given in Table 7 for comparison of experimental and predicted results from models under random conditions. Except for MLR, the adjusted R^2 values of all the models were greater than 95% and were very close to the R^2 values. It confirms that the RSM models were highly significant, and the relationship between the experimental and predicted responses was quite good. Due to their R^2 values being greater than 0.800; FFD, BBD and CCD models had a better performance than MLR model [39,40]. In addition, the RSM models have the advantage of identifying relationships and developing an efficient regression model with a considerably lower number of data compared to MLR. Furthermore, in MLR model, the effects of parameters on a response may not be linear

Table 5
ANOVA for (a) q_e (mg/g) and (b) R (%) based on CCD at coded units

(a)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	9	1010.07	112.230	17.94	0.000	
Linear	3	675.22	225.074	35.97	0.000	
pH	1	110.58	110.577	17.67	0.000	1.00
C_o (mg/l)	1	57.43	57.427	9.18	0.005	1.00
m (g)	1	507.22	507.220	81.07	0.000	1.00
Square	3	314.89	104.963	16.78	0.000	
pH*pH	1	121.30	121.304	19.39	0.000	1.02
C_o (mg/l)*C_o (mg/l)	1	38.41	38.411	6.14	0.019	1.02
m (g)**m* (g)	1	126.84	126.839	20.27	0.000	1.02
2-Way Interaction	3	19.95	6.651	1.06	0.379	
pH*C_o (mg/l)	1	1.92	1.916	0.31	0.584	1.00
pH**m* (g)	1	13.21	13.209	2.11	0.157	1.00
C_o (mg/l)**m* (g)	1	4.83	4.828	0.77	0.387	1.00
Error	30	187.70	6.257			
Lack-of-Fit	5	185.58	37.116	438.19	0.000	
Pure Error	25	2.12	0.085			
Total	39	1197.77				

Values for the model with significant coefficients
S = 2.50 R-Sq = 84.33%R-Sq(pred) = 66.43% R-Sq(adj) = 79.63%
Values for the reduced model with significant coefficients
S = 0.78R-Sq = 98.02%R-Sq(pred) = 97.28% R-Sq(adj) = 97.67%

(b)

Source	*df*	Sum of squares	Mean squares	*F*-value	*P*-value	*VIF*
Model	9	18215.5	2023.95	88.65	0.000	
Linear	3	10468.6	3489.53	152.85	0.000	
pH	1	7782.6	7782.56	340.89	0.000	1.00
C_o (mg/l)	1	2394.1	2394.09	104.86	0.000	1.00
m (g)	1	291.9	291.93	12.79	0.001	1.00
Square	3	7546.5	2515.51	110.18	0.000	
pH*pH	1	6428.7	6428.73	281.59	0.000	1.02
C_o (mg/l)*C_o (mg/l)	1	137.6	137.62	6.03	0.020	1.02
m (g)**m* (g)	1	1699.1	1699.14	74.42	0.000	1.02
2-Way Interaction	3	200.4	66.81	2.93	0.050	
pH*C_o (mg/l)	1	194.7	194.66	8.53	0.007	1.00
pH**m* (g)	1	4.7	4.66	0.20	0.655	1.00
C_o (mg/l)**m* (g)	1	1.1	1.09	0.05	0.828	1.00
Error	30	684.9	22.83			
Lack-of-Fit	5	616.5	123.30	45.06	0.000	
Pure Error	25	68.4	2.74			
Total	39	18900.4				

Values for the model with significant coefficients
S = 4.78 R-Sq = 96.38%R-Sq(pred) = 92.35%R-Sq(adj) = 95.29%
Values for the reduced model with significant coefficients
S = 0.42 R-Sq = 99.97%R-Sq(pred) = 99.95%R-Sq(adj) = 99.96%

[41]. RSM providing more flexibility and having greater capability is considered to be more effective in modeling non-linearity. For all models, the normal probability plots of the residuals indicate linearity between the experimental results and the predicted results by the model. Also, the tested VIF values to control the multicollinearity show that there is no perfect linear relationship between the explanatory variables and the predictive variables. That is, there is no correlation because VIF values are equal to 1 in applied models, except for MLR.

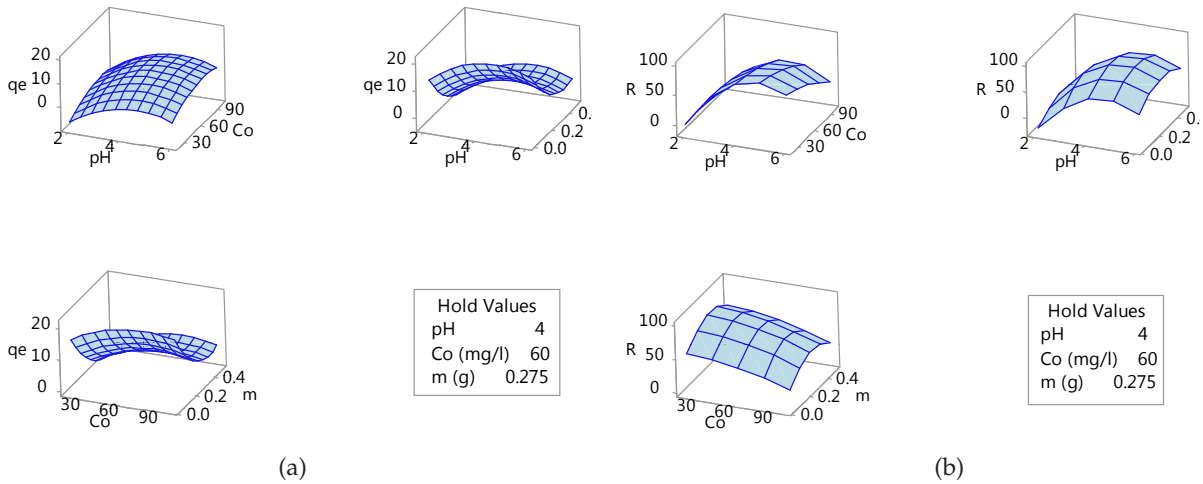

Fig. 7. Response surface plot based on (a) q_e and (b) R for CCD.

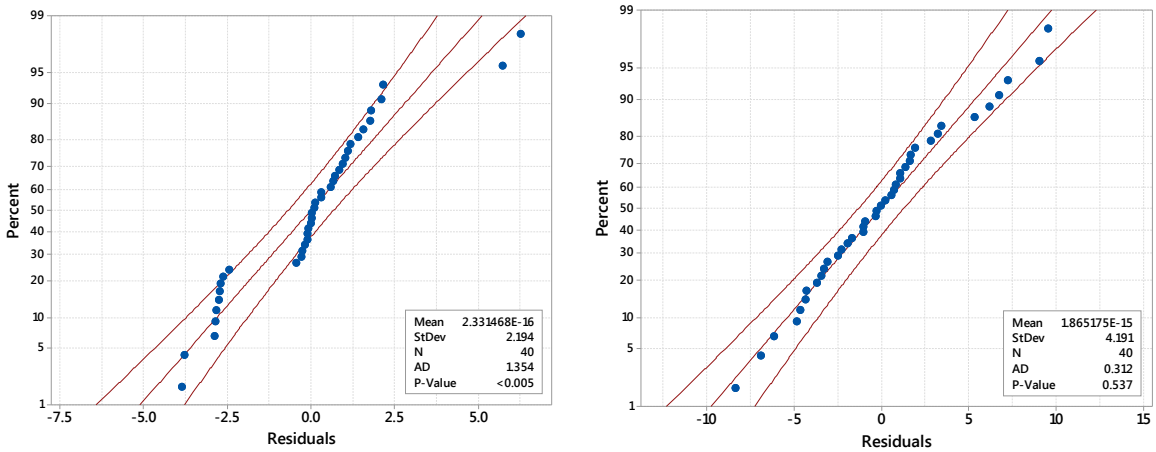

Fig. 8. Probabilitiy plots based on (a) q_e and (b) R for CCD.

Table 6
Comparison of the models performance

Parameters	q_e (mg/g)				R (%)			
	MLR	FFD	BBD	CCD	MLR	FFD	BBD	CCD
RMSE	0.2584	0.0034	0.2673	0.1346	2.0948	0.0800	0.1083	0.0130
R-Sq	0.9218	0.9999	0.9658	0.9902	0.8001	0.9965	0.9957	0.9997
R-Sq(adj)	0.8826	0.9998	0.9569	0.9767	0.7002	0.9952	0.9943	0.9996
S	0.5687	0.0575	1.6583	0.7776	14.637	1.3340	2.0576	0.4240

4. Conclusions

In this study, MLR, FFD with center points, BBD and CCD models were used to develop a function to model and optimize three experimental parameters that affect the biosorption process. The accuracy and predicting capability of these models were compared for q_e and R as response. The results showed that RSM models performed better than the MLR and FFD models in predicting of q_e and R. However, the FFD model with center points was found to be feasible for identifying the relationship between variables that affect curvature. In RSM models, the existence of square and two-way interactions of the parameters and the magnitude of the coefficients in the model equations indicate that the nonlinear effect is more effective. Also, RSM is important in modern research since they require less process time. Contrary to the classical univariate methods, which take time to investigate the response of each factor since all other factor need to be kept at a constant level, RSM presents a relatively simple and effective solution for independent factors.

Table 7
Comparison of the experimental and the predicted results by the models in random conditions

Factors			Experimental results		Predicted results							
pH	C_o (mg/l)	m (mg)	q_e (mg/g) ± S (N = 3)	R (%) ± S (N = 3)	q_e (mg/g)				R (%)			
					MLR	FFD	BBD	CCD	MLR	FFD	BBD	CCD
2.5	80	250	3.11 ± 0.15	19.44 ± 0.23	0.81	4.32	6.04	3.25	–35.40	21.77	27.83	18.97
4.0	80	400	4.90 ± 0.15	49.00 ± 1.07	1.03	3.72	6.41	5.98	–10.32	36.53	62.28	57.55
5.5	50	400	4.21 ± 0.38	67.36 ± 1.21	1.56	4.19	3.83	4.03	33.14	60.76	77.45	69.90

In addition, while experimental design techniques take into account all the factors and possible interactions, univariate methods fail to account for the interaction effects between the factors.

Symbols

P	—	Probability value
R-Sq	—	Coefficient of determination
R-Sq(adj)	—	Adjusted coefficient of determination
R-Sq(pred)	—	Predicted coefficient of determination
S (sigma)	—	Standard deviation
F	—	Calculated test statistic
α (alpha)	—	Significance level in hypothesis test

References

[1] J.O. Duruibe, M.O.C. Ogwuegbu, J.N. Egwurugwu, Heavy metal pollution and human biotoxic effects, Int. J. Phys. Sci., 2 (2007) 112–118.

[2] R. Singh, N. Gautam, A. Mishra, R. Gupta, Heavy metals and living systems: An overview, Indian J. Pharmacol., 43 (2011) 246–253.

[3] A.İ. Şengil, M. Özacar, Biosorption of Cu(II) from aqueous solutions by mimosa tannin gel, J. Hazard. Mater., 157 (2008) 277–285.

[4] M. Bilal, J.A. Shah, T. Ashfaq, S.M.H. Gardazi, A.A. Tahir, A. Pervez, H. Haroon, Q. Mahmood, Waste biomass adsorbents for copper removal from industrial wastewater-A review, J. Hazard. Mater., 263 (2013) 322–333.

[5] US Environmental Protection Agency, Drinking Water Contaminants–Standards and Regulations. http://water.epa.gov/drink/contaminants/basicinformation/copper.cfm/,2016 (accessed 28.11.16).

[6] Z.N. Garba, I. Bello, A. Galadima, A.Y. Lawal, Optimization of adsorption conditions using central composite design for the removal of copper (II) and lead (II) by defatted papaya seed, Karbala Int. J. Mod. Sci., 2 (2016) 20–28.

[7] G. Xiao, X. Zhang, H. Su, T. Tan, Plate column biosorption of Cu(II) on membrane-type biosorbent (MBS) of Penicillium biomass: Optimization using statistical design methods, Bioresour. Technol., 143 (2013) 490–498.

[8] T. Şahan, H. Ceylan, N. Şahiner, N. Aktaş, Optimization of removal conditions of copper ions from aqueous solutions by Trametes versicolor, Bioresour. Technol., 101(12) (2010) 4520–4526.

[9] Y. Prasanna Kumar, P. King, V.S.R.K. Prasad, Removal of copper from aqueous solution using Ulva fasciata sp.–A marine green algae, J. Hazard. Mater. B., 137 (2006) 367–373.

[10] A. Özer, G. Gürbüz, A. Çalimli, B.K. Körbahti, Biosorption of copper(II) ions on Enteromorpha prolifera: Application of response surface methodology (RSM), Chem. Eng. J., 146 (2009) 377–387.

[11] R. Kumar, R. Singh, N. Kumar, K. Bishnoi, N.R. Bishnoi, Response surface methodology approach for optimization of biosorption process for removal of Cr(VI), Ni(II) and Zn(II) ions by immobilized bacterial biomass sp. Bacillus brevis, Chem. Eng. J., 146 (2009) 401–407.

[12] R. Singh, R. Chadetrik, R. Kumar, K. Bishnoi, D. Bhatia, A. Kumar, N.R. Bishnoi, N. Singh, Biosorption optimization of lead(II), cadmium(II) and copper(II) using response surface methodology and applicability in isotherms and thermodynamics modeling, J. Hazard. Mater., 174 (2010) 623–634.

[13] S. Uwadiae, E. Aluyor, C. Okieimen, I. Oboh, Statistical design analysis for adsorption of Cu(II) and Pb(II) onto kaolinitic clay, Pac. J. Sci. Technol., 14 (2013) 310–317.

[14] D. Öztürk, T. Şahan, Design and optimization of Cu(II) adsorption conditions from aqueous solutions by low-cost adsorbent pumice with response surface methodology, Pol. J. Environ. Stud., 24 (2015) 1749–1756.

[15] A.R. Cestari, E.F.S. Vieir, I.A. Oliveira, R.E. Bruns, The removal of Cu(II) and Co(II) from aqueous solutions using cross-linked chitosan–Evaluation by the factorial design methodology, J. Hazard. Mater., 143 (2007) 8–16.

[16] N.K.E.M. Yahaya, M.F.P.M. Latiff, I. Abustan, M.A. Ahmad, Effect of preparation conditions of activated carbon prepared from rice husk by $ZnCl_2$activation for removal of Cu (II) from aqueous solution, IJET, 10 (2010) 28–32.

[17] K. Kalantari, M.B. Ahmad, H.R.F. Masoumi, K. Shameli, M. Basri, R. Khandanlou, Rapid adsorption of heavy metals by Fe_3O_4/Talc nanocomposite and optimization study using response surface methodology, Int. J. Mol. Sci., 15 (2014) 12913–12927.

[18] G. Vázquez, M. Calvo, M.S. Freire, J. González-Alvarez, G. Antorrena, Chestnut shell as heavy metal adsorbent: Optimization study of lead, copper and zinc cations removal, J. Hazard. Mater., 172 (2009) 1402–1414.

[19] W.P. Putra, A. Kamari, S.N.M. Yusoff, C.F. Ishak, A. Mohamed, N. Hashim, I.M. Isa, Biosorption of Cu(II), Pb(II) and Zn(II) ions from aqueous solutions using selected waste materials: Adsorption and characterisation studies, J. Encapsul. Adsorpt. Sci., 4 (2014) 25–35.

[20] S.N.M. Yusoff, A. Kamari, W.P. Putra, C.F. Ishak, A. Mohamed, N. Hashim, I.M. Isa, Removal of Cu(II), Pb(II) and Zn(II) ions from aqueous solutions using selected agricultural wastes: Adsorption and characterisation studies, J. Environ. Prot., 5 (2014) 289–300.

[21] A. Ghosh, K. Sinha, P. Das Saha, Central composite design optimization and artificial neural network modeling of copper removal by chemically modified orange peel, Desal. Water Treat., 51 (2013) 7791–7799.

[22] M.C. Hoces, G.B. García, A.R. Gálvez, A.E. Álvarez, M.Á. Martín-Lara, Biosorption of Cu^{2+} in a packed bed column by almond shell: optimization of process variables, Desal. Water Treat., 51 (2013) 1954–1965.

[23] Z.Z. Chowdhury, S.M. Zain, R.A. Khan, A.A. Ahmad, K. Khalid, Application of response surface methodology (RSM) for optimizing production condition for removal of Pb(II) and Cu(II) onto kenaf fiber based activated carbon, Res. J. Appl. Sci. Eng. Technol., 4 (2012) 458–465.

[24] T.M. Alslaibi, I. Abustan, M.A. Ahmad, A.A. Foul, Application of response surface methodology (RSM) for optimization

of Cu^{2+}, Cd^{2+}, Ni^{2+}, Pb^{2+}, Fe^{2+}, and Zn^{2+} removal from aqueous solution using microwaved olive stone activated carbon, J. Chem. Technol. Biotechnol., 88 (2013) 2141–2151.
[25] M.A. Acheampong, J.P.C. Pereira, R.J.W. Meulepas, P.N.L. Lensa, Biosorption of Cu(II) onto agricultural materials from tropical regions, J. Chem. Technol. Biotechnol., 86 (2011) 1184–1194.
[26] E. Oguz, M. Ersoy, Removal of Cu^{2+} from aqueous solution by adsorption in a fixed bed column and Neural Network Modelling, Chem. Eng. J., 164 (2010) 56–62.
[27] D. Bingöl, M. Inal, S. Çetintaş, Evaluation of copper biosorption onto date palm (Phoenix dactylifera L.) seeds with MLR and ANFIS models, Ind. Eng. Chem. Res., 52 (2013) 4429–4435.
[28] R.G. Brereton, Applied Chemometrics for Scientists, John Wiley & Sons, New York, 2007.
[29] D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, New York, 2008.
[30] C. Stalikas, Y. Fiamegos, V. Sakkas, T. Albanis, Developments on chemometric approaches to optimize and evaluate microextraction, J. Chromatogr. A., 1216 (2009) 175–189.
[31] S.H. Hasan, P. Srivastava, M. Talat, Biosorption of Pb(II) from water using biomass of Aeromonas hydrophila: Central composite design for optimization of process variables, J. Hazard. Mater., 168 (2009) 1155–1162.
[32] S.H. Chang, T.T. Teng, N. Ismail, Optimization of Cu(II) extraction from aqueous solutions by soybean-oil-based organic solvent using response surface methodology, Water Air Soil Pollut., 217 (2011) 567–576.
[33] L. Murray, H. Nguyen, Y.F. Lee, M.D. Remmenga, D.W. Smith. Variance inflation factors in regression models with dummy variables, Conference on Applied Statistics in Agriculture, (2012) https://doi.org/10.4148/2475-7772.1034.
[34] W. Yoo, R. Mayberry, S. Bae, K. Singh, Q.P. He, Jr. J.W. Lillard, A Study of effects of multicollinearity in the multivariable analysis, Int. J. Appl. Sci. Technol., 4(5) (2014) 9–19.
[35] MAE and RMSE—Which Metric is Better? https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d (accessed 23.11.18).
[36] G. Hanrahan, K. Lu, Application of factorial and response surface methodology in modern experimental design and optimization, Crit. Rev. Anal. Chem., 36 (2006) 141–151.
[37] W.J. Gong, Y.P. Zhang, S.H. Choi, Y.J. Zhang, K. Lee, Application of response surface methodologies in capillary electrophoresis, Microchim. Acta, 156 (2007) 327–335.
[38] J. Zolgharneina, A. Shahmoradia, J.B. Ghasemi, Comparative study of Box–Behnken, central composite, and Doehlert matrix for multivariate optimization of Pb(II) adsorption onto Robinia tree leaves, J. Chemom., 27 (2013) 12–20.
[39] M. Sedighi, K. Keyvanloo, J. Towfighi, Modeling of thermal cracking of heavy liquid hydrocarbon: Application of kinetic modeling, artificial neural network, and neuro-fuzzy models, Ind. Eng. Chem. Res., 50 (2011) 1536–1547.
[40] S.S. Chong, A.R. Abdul Aziz, S.W. Harun, H. Arof, S. Shamshirband, Application of multiple linear regression, central composite design, and ANFIS models in dye concentration measurement and prediction using plastic optical fiber sensor, Measurement, 74 (2015) 78–86.
[41] T.M. Patel, N.M. Bhatt, RSM and MLR model for equivalent stress prediction of Eicher 11.10 chassis frame: a comparative study, Syst. Technol., 51 (2016) 387–395.