Journal of Engineering Science and Technology
Vol. 15, No. 4 (2020) 2324 - 2338
© School of Engineering, Taylor's University

TOBIT MODELING AND CENTRAL COMPOSITE DESIGN FOR OPTIMIZATION OF SPECIFIC SURFACE AREA OF CLAY BY ACTIVATION

JAKA NUGRAHA[1,*], IS FATIMAH[2]

[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Kampus Terpadu UII, Jl. Kaliurang Km 14, Yogyakarta, Indonesia
[2]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Kampus Terpadu UII, Jl. Kaliurang Km 14, Yogyakarta, Indonesia
*Corresponding Author: jnugraha@uii.ac.id

Abstract

Clay is one of the interesting materials in the adsorption process for water treatment. According to some reports, physical and chemical treatments of clay are usually performed to increase the specific surface area as well as the adsorptivity for specified water contaminants, and modeling related to optimization is usually used for this purpose. Technically, optimization to obtain a high specific surface area by chemical experiments is inefficient in time and chemical consumption, so a statistical approach may be required. In this study, optimization of bentonite treatment was conducted. Clay activation was conducted by acid treatment, calcination, and their combinations refers to the application for chromate and rhodamine B adsorption. In such circumstances regarding the variables of clay treatments, many studies have utilized logit, probit, and complementary log-log models, but these models are difficult to implement. The aims of the present research were to evaluate the Tobit model as a new method for predicting the specific surface area of clay as a function of the activation method (reflux and calcination) and the variables of acid concentration and the temperature of calcination. The Tobit model produces roughly the same results as the ordinary regression model for clay activation and adsorption application. In order to identify the maximum point, a central composite design provides a more accurate result with a maximum specific area of 167.34 m^2/g at an activation temperature of 415.029 °C and an acid concentration of 0.4876 M.

Keywords: Central composite design, Generalized linear model, Logit model, Tobit model.

1. Introduction

The specific surface area (SSA) of clay minerals is an important parameter that is useful for predicting the capabilities of the mineral in adsorption and catalysis applications. Like many factors in adsorption and catalysis, the SSA parameter represents the adsorption capability, which is important to understand specifically for further technical applications [1, 2]. The SSA quantifies the mineral surface area per unit mass and is expressed in meters squared per gram (m^2 /g). The SSA varies widely depending on mineralogy and treatments, ranging from as much as approximately 800 m^2/g for expansive clay minerals such as smectite, depending on the extent of the exposed interlayer surface, to as little as approximately 10 m^2/g for non-expansive clay minerals such as kaolinite. Because the surface areas of silt, sand, and gravel-sized particles are negligible by comparison, the surface areas of natural soils containing a distributed mixture of such particles are dominated by the SSA of the clay fraction, even if this fraction is very small. Such chemical and physical treatments are reported for determining the optimum surface area of clay minerals. Chemical treatments that can be conducted are acid and base treatments, while physical treatment consists of calcination (heating) [3, 4]. The combination of both methods is also sometimes performed. In order to minimize steps or variations in the optimization method, some research has employed statistical analysis for optimization. One such analytical method is regression analysis. By designating the temperature of calcination (T) and the concentration of acid in the chemical treatment (C) as crucial parameters, the optimization of both parameters was conducted in this study.

Regression analysis assumes a normal distribution of the dependent variables with no limit value. In many cases, the only independent variable has a value with certain restrictions. For example, the specific surface area has a minimum value of zero. In such cases, using multiple linear regression with ordinary least squares (OLS) is less precise because of the limited nature of the dependent variables. As OLS continues to be used, the result will be a biased and inconsistent estimator. Tobit regression can be used to overcome this problem [5].

The Tobit regression model as a regression analysis method to determine the dependent variable has limitations. The estimation of parameters in the Tobit model is done using maximum likelihood estimation (MLE). According to previous research, the MLE method produces consistent and efficient estimators for large samples [6]. The Tobit regression model is used frequently to model censored variables in econometrics research. It has been concluded that the simulation data of Tobit model parameter estimators tend to be biased upward, while OLS results tend to be biased downward. However, it can be guaranteed that the application of Tobit models with the maximum likelihood method will reduce the bias compared with the use of OLS in cases of censored data. Tobit estimates will provide a useful guide to failures of homoskedasticity or normality [7].

OLS is based on the assumption that data have a normal distribution and lie in the $-\infty$ to ∞ range. In many cases, data with limited values are also faced, either at the lower or upper limit. For example, the data on the temperature of a solution cannot be greater than 100 °C. A zero value is calculated as the lower censored value with 100 as the upper censored value, and the data is called censored data. This value can be applied to assumptions of data distribution.

Monte Carlo simulation studies were examined to contrast the performance of the Tobit model for censored data to that of OLS regression. It was demonstrated that in the presence of a ceiling effect, if the conditional distribution of the measure of health status had uniform variance, then the coefficient estimates from the Tobit model had superior performance compared with estimates from OLS regression. However, if the conditional distribution had non-uniform variance, then the Tobit model performed at least as poorly as the OLS model. The Tobit model has also been widely applied in research in the field of health [8-10]. The present research aims to evaluate the effect of the statistical method for optimizing the specific surface area of clay as a function of activation methods (reflux and calcination) and the variables of temperature and acid concentration. With multiple factors influencing the specific surface area parameters, a design strategy for prediction and optimization are important economic-processing in an approach similar to that of previous works [11-13]. In chemometric optimization, central composite design (CCD) is widely studied and compared with other techniques. Asghar et al. [11] reported the comparison of CCD and Taguchi optimization by the Fenton process and concluded that the Taguchi method is a suitable alternative to CCD for several chemical engineering applications. In this research, the ordinary regression model and the Tobit model were evaluated for the prediction, and the optimum point was determined by comparing the models with CCD. Theoretically, the Tobit model can be utilized for this case because the data is censored data.

Furthermore, all models were evaluated based on the coefficient of determination (R^2), the mean absolute error (*MAE*), and the mean absolute percentage error (*MAPE*).

2. Tobit Model

The model was first developed by Tobin [14], who correlated a study based on probit analysis and called it the Tobit model (Tobit Probit). Censored data are normally distributed with the assumption $N(\mu;\sigma^2)$. The standard Tobit model formulation is an independent variable with a left-censored y-independent standard of zero value:

$$y_i^* = x_i'\beta + \epsilon_t \quad \text{with} \quad y_i = \begin{cases} 0 & if\ y_i^* \leq 0 \\ y_i^* & if\ y_i^* > 0 \end{cases} \tag{1}$$

for $i = 1,2,...,n$ stated the observation, y_i^* is an unobserved variable, x_i is a vector of explanatory variables, β is a vector of unknown parameters, and ε_i is a disturbance term.

In a generalization of the standard Tobit model, the dependent variable can be either left-censored, right-censored, or both left-censored and right-censored, where the lower and/or upper limit of the dependent variable can be any number:

$$y_i^* = x_i'\beta + \epsilon_t \quad \text{with} \quad y_i = \begin{cases} a & if\ y_i^* \leq a \\ y_i^* & if\ a < y_i^* < b \\ b & if\ y_i^* \geq b \end{cases} \tag{2}$$

Here, *a* is the lower limit and *b* is the upper limit of the dependent variable [15, 16].

Tobit models are usually estimated using maximum likelihood estimation (MLE). Assuming that the disturbance term ε follows a normal distribution with mean 0 and variance σ^2, the *likelihood function*, denoted *L*, is:

$$L = \left(\prod_{y_i \le a} f(y_i)\right)\left(\prod_{a<y_i<b} f(y_i)\right)\left(\prod_{y_i \ge b} f(y_i)\right) \quad (3)$$

and $f(y_i)$ is the probability density function of normal distribution

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right)$$

So, the *log likelihood function*, denoted *log L*, is:

$$logL = \left(\sum_{y_i \le a} log\,(f(y_i))\right) + \left(\sum_{a<y_i<b} log\,(f(y_i))\right) + \left(\sum_{y_i \ge b} log\,(f(y_i))\right) \quad (4)$$

Eq. (4) can be stated as

$$logL = \sum_{i=1}^{n} I_i^a log\,\Phi\left(\frac{a - x_i'\beta}{\sigma}\right) + \sum_{i=1}^{n} I_i^b log\,\Phi\left(\frac{x_i'\beta - b}{\sigma}\right) + \sum_{i=1}^{n}\left((1 - I_i^a - I_i^b) log\,\Phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) - log(\sigma)\right) \quad (5)$$

$$with\ I_i^a = \begin{cases} 1\ if\ y_i = a \\ 0\ if\ y_i > 0 \end{cases} \quad and\ I_i^b = \begin{cases} 1\ if\ y_i = b \\ 0\ if\ y_i < 0 \end{cases} \ and\ \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sigma} f(y_i) dy_i.$$

The maximum likelihood estimations for β and σ are the optimization of Eq. (5). Those equations can be solved by using the *maxLik* function of the maxLik package in R software [17]. Tobit models can be estimated in R software with the *censReg* function, which is available in the *censReg* package [15, 16].

3. Optimization in Linear Model

The basic concept of optimization in a linear model is that the function of two variables, $z=f(x,y)$ has continuous second partial derivatives in ($x=a$, $y=b$) and

$$\left.\frac{\partial f(x,y)}{\partial x}\right|_{(x=a,y=b)} = \left.\frac{\partial f(x,y)}{\partial y}\right|_{(x=a,y=b)} = 0 \quad (6)$$

Let $\Delta = AD - B^2$ with

$$A = \left.\frac{\partial^2 f(x,y)}{\partial x^2}\right|_{(x=a,y=b)}, B = \left.\frac{\partial^2 f(x,y)}{\partial x \partial y}\right|_{(x=a,y=b)}, D = \left.\frac{\partial^2 f(x,y)}{\partial y^2}\right|_{(x=a,y=b)}.$$

with the properties of $f(x,y)$ as below:

a. $f(a,b)$ is a local minimum value if $A > 0$ and $\Delta > 0$.
b. $f(a,b)$ is a local maximum value if $A < 0$ and $\Delta > 0$.
c. if $\Delta < 0$; then $f(a,b)$ is not an extreme value.
d. If $\Delta = 0$, there is no conclusion.

In order to evaluate whether there are any effects of *x* and *y* for *z*, linear models of two-way analysis of variance (ANOVA) and regression analysis were conducted.

4. Materials and Method

The materials used for the experiments were natural montmorillonite/ bentonite from East Java, Indonesia; sulfuric acid (Merck-Millipore, Germany); and N_2 gas (Samator, Indonesia). Montmorillonite activation was conducted by refluxing montmorillonite powder in varied concentrations of sulfuric acid (*C*) for 4 h, neutralization using water followed by drying and calcination at varied temperatures (*T*). The sulfuric acid concentrations were 0, 0.5, 1.0, and 2.0 M and the temperature of calcination ranged from 25 °C to 700 °C. The specific surface areas (*SSA*) of the activated montmorillonite samples were determined based on the Brunauer-Emmett-Teller (BET) equation from the N_2 adsorption-desorption data obtained using a NOVA 1200e gas sorption analyzer (Quantachrome, New York, US).

The maximum value was determined from the most accurate of three models: linear, Tobit, and composite design.

5. Results and Discussion

Figure 1 presents the results of the SSA at various temperatures and concentrations of acid.

In each treatment, it can be seen that the effect of *T* on the SSA achieves the maximum at *T*=400 °C, except for *C*=0, where the maximum SSA is at *T*=500 °C.

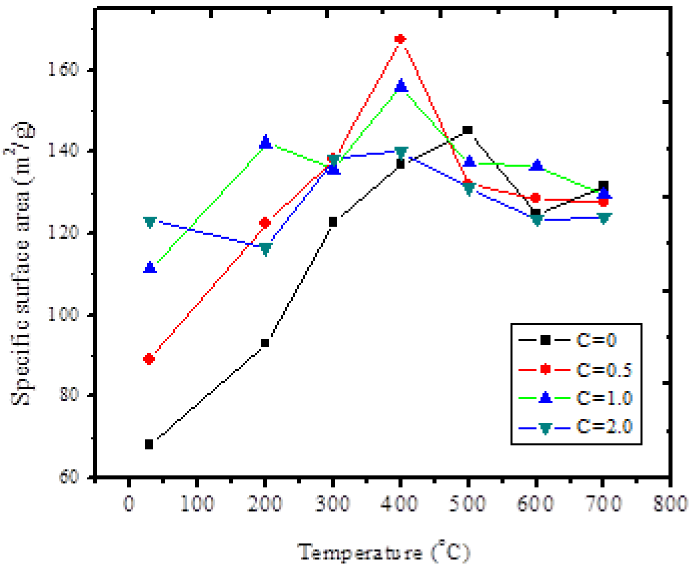

Fig. 1. Surface area at varied temperatures and concentrations of acid.

Furthermore, the specific surface area at each temperature affects the optimum concentration. At room temperature (25°C), each SSA is achieved at *C*=2 and *T*= 200 and moves to *C*=1. At *T*=400, the optimum values were achieved at *C*=0.5 and *T*=500; the optimum was obtained at *C*=0, so the opposite effect identifies the interaction. In general, the optimum points are at *C*=0.5 and *T*=400.

5.1. Linear model

Analysis of the effect of *C* and *T* using 2-way ANOVA provides the output presented in Table 1.

Table 1. Results of 2-way ANOVA.

Source	Df	SS	MS	F	p-value
C	3	3551.1	1183.69	107.63	0.000
T	6	19088.9	3181.48	289.28	0.000
Interaction	18	8588.8	477.16	43.39	0.000
Error	56	615.9	11.00		
Total	83	31844.6			

It is clearly found from the *F* statistic that *C* and *T* have a significant effect on SSA.

Points to be noted are:

a. There is an influence of *C* on SSA (*Z*).

The highest *Z* value occurred at *C*=1, which is 135.55, and the lowest value was at *C*=0, corresponding to 117.4662. At *C*=2, the *Z* value was 128.13, which indicates that the concentration of acid (*C*) exerts influence in a quadratic pattern.

b. There is an influence of *T* on SSA (*Z*).

The description of the effect of *T* on Z can be seen in Fig. 1. The lowest mean of Z occurs at room temperature, with the value of 97.89, while the maximum value, at *T*=400, is 150.13. The Z value is reduced at increasing temperatures of more than 400 °C.

c. There is an interaction between *C* and *T* on SSA (*Z*).

The interaction between *C* and *T* is described by the curve in Fig. 1. The minimum Z is 68.05 through the combination of T at room temperature and the concentration of acid (*C*) at zero (*C*= 0). The maximum value for Z is 167.43, which results from the combination of *T*= 0.5 and *C*= 400.

By using the linear regression model, it was noted that the linear model of the second-order was obtained. The regression equation is:

$$Z = 65.4 + 47.3\, C + 0.257\, T - 0.0427\, CT - 12.6\, C^2 - 0.000243\, T^2 \quad (7)$$

(5.25) (7.93) (13.05) (-6.24) (-5.10) (-9.88)

Values in the parentheses are the student's t-test values of the corresponding parameter, which leads to the conclusion that the model obeys the second-degree polynomial with the value of coefficient determination R^2 = 75.31%.

The optimization of Eq. (7) can be conducted using the properties of Eq. (6). The first order-derivative of the equation is:

$$\frac{\partial Z}{\partial C} = 47.3 - 0.0427T - 25.2C \quad (8)$$

$$\frac{\partial Z}{\partial T} = 0.257 - 0.0427C - 0.000486T \quad (9)$$

By using d*Z*/d*C*= 0 and d*Z*/d*T* = 0, it is found that:

T_0 = 428.581 and C_0 = 1.151

The second order derivative at T_0 = 428.581 and C_0 = 1.151 is

$$\frac{\partial^2 Z}{\partial C^2} = -25.2\ ;\ \frac{\partial^2 Z}{\partial T^2} = -\ 0.000486 \text{ and } \frac{\partial^2 Z}{\partial C \partial T} = -0.0427$$

Based on Eq. (6), it is found that Δ = 1.0760. Because $\frac{\partial^2 Z}{\partial C^2} < 0$ and Δ > 0 it is concluded that $f(T_0, C_0)$ is the maximum at *f*. The maximum value of the specific surface area as determined using the OLS method is

Z_{max} = 147.5967

It is concluded that the maximum value obtained from OLS is lower than the maximum value obtained from observation.

On T=28 and C= 0, the SSA is found to be 68.05, meanwhile the optimum condition is on T_0 = 428.581 and C_0 = 1.151. Z_{max} = 147.597. The data suggest that there is an increasing SSA of 216.89%. The result is refer to previous research on the increasing specific surface area of silica alumina material at higher temperature and acid concentration in activation process [4] .The higher concentration of activation agent has a direct correlation to produce porous structure and active sites and cavities in the silica-alumina material [1]. Particularly in this research, effect of acid concentration on *SSA* is by polynomial second order at the maximum concentration of *C*= 1.151.

5.2. Tobit model

The Tobit model provides the capability to ensure the left or right censor points. From those points, it is found that the left censor type of 90 is the sufficient point. The determination of the censored value of the lower limit is due to aim of the study, which was to increase the SSA with varied temperatures and acid concentrations. The fit equation is:

$$Z = 64.278 + 46.581C + 0.262T - 0.0432CT - 0.000248T^2 - 12.159C^2 \qquad (10)$$

(10.920) (6.928) (11.002) (-5.613) (-9.170) (-4.722)

Values in the parentheses are the student's t-test values of the corresponding parameter. Based on the student's t-test value, it is concluded that all parameters in Eq. (10) are significant. The model has a *log-likelihood* value of -291.2797 at a degree of freedom (*Df*) of 7. It was noted that G^2 = -2 *log-likelihood* has Chi-square distribution, so the obtained value was G^2 = 582.5594, which has a value of *p-value* = $1.386.10^{-121}$. These statistics provide evidence that the model fit with the data. The correlation between observation and prediction on the Tobit model is 75.23%.

The optimization of Eq. (1) can be performed using the properties of Eq. (6). The first order derivative of the equation is:

$$\frac{\partial Z}{\partial C} = 46.581 - 0.04318T - 24.318C \qquad (11)$$

$$\frac{\partial Z}{\partial T} = 0.262 - 0.04318C - 0.000496T \qquad (12)$$

By using d*Z*/d*C*= 0 and d*Z*/d*T* = 0, it is found that

T_0 = 427.524 and C_0 = 1.15675

The second order derivative at T_0 and C_0 is

$$\frac{\partial^2 Z}{\partial C^2} = -24.31 \; ; \frac{\partial^2 L}{\partial T^2} = -0.000496 \text{ and } \frac{\partial^2 Z}{\partial C \partial T} = -0.04318$$

Based on Eq. (6), it is found that Δ = 0.010193. Because $\frac{\partial^2 z}{\partial c^2} < 0$ and Δ > 0 it is concluded that $f(T_0, C_0)$ is the maximum local at *f*. The maximum value of specific surface area as determined using the Tobit model is

$Z_{max} = 147.2096$

As with the result from OLS, the maximum value obtained from the Tobit model is lower than the maximum value obtained from observation.

5.3. Central composite design (CCD)

CCD is one of the more responsive surface methodologies, and it has been widely used for optimization in catalysis and chemical processes, such as seeking the optimum catalytic condition, preparation of catalyst extraction of specific components, etc. [13, 18, 19].

An experiment was conducted using CCD with a center of *T*= 400 and *C*= 0.5. The result is presented in Table 2.

X and *Y* variables can be determined with the following equation:

$$Y = \frac{T-400}{400}, \text{ and } X = \frac{C-0.5}{0.5} \quad (13)$$

Based on Table 2, the regression equation is

$$Z = 167 + 4.47\,Y - 5.21\,XY - 15.8\,X^2 - 15.3\,Y^2 \quad (14)$$

(42.48) (2.62) (-2.64) (-5.22) (-5.39)

Table 2. Results of CCD.

Independent Variables		Transformation Code		Dependent Variables
C	***T***	***X***	***Y***	***Z***
0	300.0	-1	-1	123.30
0	300.0	-1	-1	135.50
0	300.0	-1	-1	109.20
0	500.0	-1	1	140.70
0	500.0	-1	1	150.45
0	500.0	-1	1	144.21
1	300.0	1	-1	135.30
1	300.0	1	-1	136.00
1	300.0	1	-1	136.00
1	500.0	1	1	137.40
1	500.0	1	1	137.00
1	500.0	1	1	137.80
0.5	400.0	0	0	167.89
0.5	400.0	0	0	168.50
0.5	400.0	0	0	165.90
0.5	541.4	0	1.41	139.54
0.5	258.6	0	-1.41	140.06
1.207	400.0	1.41	0	141.93

The values in the parentheses are the student's t-test values of the corresponding parameter. The ANOVA and test model statistics, which are presented in Table 3, show that the model is very good. This can be seen from the *F* value, the *value of Lack of Fit*, and the R^2 value.

a. The *F* value in regression showed that not all of the regression coefficients were equal to zero. The conclusion to be noted is that all variables in Eq. (14) influence the Z values significantly.

b. From the *value of Lack of Fit*, *p-value* > α. *P-value* is larger than α, there is no evidence that the model does not fit the data. We can conclude that the model fit the data well.

c. R-squared (R^2) is a statistical measure of how close the data are to the fitted regression line. It means that the model is capable of describing the variation among the data at 83.5%. Using the central composite design can result in very low *variance inflation factor* (*VIF*) values (close to 1), which means that multicollinearity does not occur. In other words, the model has the predictive capability.

Table 3. Analysis of variance.

Source	Df	SS	MS	F	p-value
Regression	4	3069.42	767.35	16.47	0.000
Residual Error	13	605.77	46.60		
Lack of Fit	3	206.21	68.74	1.72	0.226
Pure Error	10	399.56	39.96		
Total	17	3675.19			
S = 6.82627 *R-Sq* = 83.5% *R-Sq(adj)* = 78.4%					

A scatter plot of the observation values of prediction is shown in Fig. 2. It can be seen that the observation values lie around the line, indicating the correlation of observation and prediction values with a correlation of 0.9183.

The optimization of Eq. (14) can be done using the properties of Eq. (6). The first-order derivative of the equation is

$$\frac{\partial Z}{\partial Y} = 4.47 - 5.21X - 30.6Y \qquad (15)$$

$$\frac{\partial Z}{\partial X} = -5.21Y - 31.6X \qquad (16)$$

By using $\partial Z/\partial X = 0$ and $\partial Z/\partial Y = 0$ the solution of the equation

$Y_0 = 0.150298$ and $X_0 = -0.02478$

The second order derivative at X_0 and Y_0 is

$$\frac{\partial^2 Z}{\partial y^2} = -30.6\ ;\ \frac{\partial^2 L}{\partial x^2} = -31.6 \text{ and } \frac{\partial^2 Z}{\partial x \partial y} = -5.21$$

Because $\frac{\partial^2 Z}{\partial x^2} < 0$ and $\Delta > 0$ it is concluded that $f(X_0, Y_0)$ is the maximum local at *f*. The maximum value of the specific surface area using CCD is

$Z_{max} = 167.3359$

Based on Eq. (13), *X* and *Y* values can be converted into *C* and *T* and it is found that

$C_0 = 0.48761$ and $T_0 = 415.0298$.

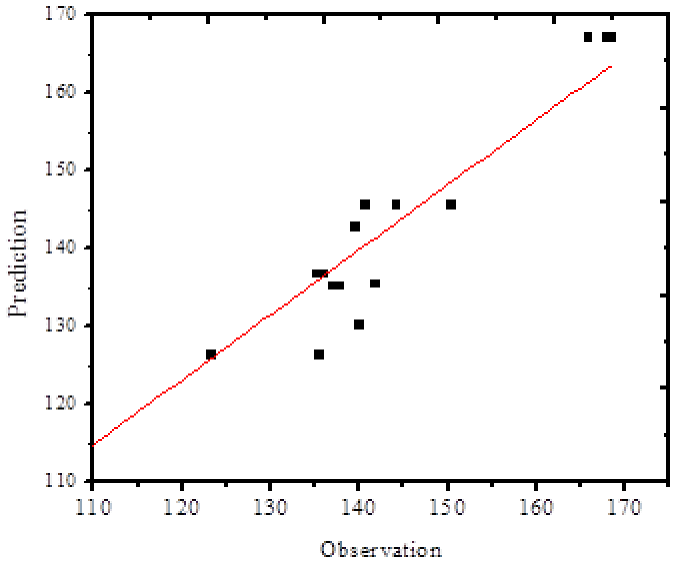

Fig. 2. Scatter plot of observation values of prediction using CCD.

At T=28 and C= 0, the SSA is 68.05, while the optimum conditions are C_0 = 0.48761 and T_0 = 415.0298, Z_{max} = 167.3359. An increase of 245.90% was obtained.

6. Comparison Tobit, OLS in two-way ANOVA, and Central Composite Design

A comparison of the model can be derived from:

1) Correlation of observed and predicted values

The correlation between observed and predicted values is given in Table 4 and the scatter plots in Figs. 1 and 2.

Table 4. Correlation between observation and prediction.

Model	R^2 (%)	MAE	MAPE (%)
OLS	75.308	7.663	6.050
Tobit	75.230	9.408	7.865
Composite Design	83.513	4.048	3.076

Based on the coefficient determination value denoted R^2 in Table 4, the OLS and Tobit models are relatively similar. However, the regression model obtained using CCD resulted in a better R^2 value than the factor design in Table 1. It is noted that the obtained linear model from CCD as Eq. (14) is applied for $258.6 \leq T \leq 541$ and $0.5 \leq C \leq 1.207$. Otherwise, the insignificant fitness will be obtained at ranges of $28 \leq T \leq 700$ and $0 \leq C \leq 2$ with R^2 of 7.17% and a MAPE value of 38.363%. Linear model Eq. (7) and Tobit model Eq. (10) can be utilized for predicting SSA at varied C and T. Meanwhile Eq. (14) from CCD can be employed for determining optimum value, but cannot be used for predicting SSA [11, 12, 13, 19].

From both the MAE and MAPE values, it can be seen that OLS is better than the Tobit model. In advance, the Tobit model has advantages as there is no censored observation, the results will be the same as OLS and the determination of censored points can be obtained from visual graphical analysis.

2) Maximum values from the points

The optimum value uses OLS and the Tobit model based on factor design, as in Eqs. (7) and (10), compared with the optimization value in CCD.

Based on the maximum value in Table 5, the regression model obtained using CCD gives the value closest to the observed value, as in the two-way ANOVA.

Table 5. Comparison of optimization on 2-way design and CCD.

Model	T(°C)	$C(M)$	L_{max}(m²/g)
2-way ANOVA	400	0.50	167.43
OLS	428.58	1.15	147.59
Tobit	427.52	1.16	147.21
Composite Design	415.03	0.49	167.34

From the statistics F, R^2, MAE and MAPE it is concluded that CCD is the best among OLS and the Tobit model. The optimum value obtained using CCD is T=415.029 °C, C=0.4876, for a SSA of 167.34.

The comparison on statistical optimization for specific surface area of clay and other adsorbent materials are listed in Table 6.

Table 6. Comparison on statistical optimization for specific surface area of clay and other adsorbent materials.

Material	Statistical method, variable of optimization	Optimization Result and correlation value	Reference
Bentonite	Response surface methodology (RSM), (time, acid concentration, and microwave heating power in microwave-assisted activation method)	RSM represents good method for optimization with R^2= 0.914. The optimum conditions to obtain the maximum specific surface area for microwave assisted acid-activated bentonite were acid concentration = 5.2 M, time = 7.4 min, and microwave power = 117 W.	[20]
Hydroxyapatite	CCD	CCD is effectively optimizing the activation for increasing specific surface area of Hydroxyapatite prepared from catfish. The optimum conditions for the production n were found to be calcination temperature of 300 °C and calcination time of 1 h.	[21]

Bentonite	Response surface methodology based on central composite rotatable design (CCRD)	The small *p-value* (0.0001) and a suitable coefficient of determination (R^2=0.9270) showed that the quadratic polynomial model based on optimization is good fit with the experimental data. The optimal conditions for the acid activation of the bentonite were obtained at 90 °C, 4.5 N, 450 rpm,1:4.5 and time 3 h.	[22]
Smectite clay	Factorial design	The statistical method showed good optimization for specific surface area based on the variables of temperature, acid (HCl) concentration, stirring speed, solid/liquid ratio, and time. The coefficient of determination of 0.995 indicates a good predictability of the model.	[23]
Bentonite	CCD	A central composite design was used to determine the pertinence of variables: the activation temperature (X_1), activation time (X_2) and massic ratio for adsorptivity for Cd(II). Coefficient of determination = 0.92.	[24]
Bentonite	OLS, Tobit model, and CCD	CCD is the best model for optimizing specific surface area with the condition at temperature (T) of 415.029 °C and a sulfuric acid concentration (C) of 0.488 M, with the optimum value of 167.34 m^2/g. coefficient of determination between prediction and observation = 0.918.	This work

Refer to the data in Table 6, it can be noted that the CCD method applied in this research is comparable with previous works as shown by the similar coefficient of determination between prediction and observation at around 0.92, even though there is the better model with higher correlation value (0.995) obtained by factorial design. The better performance is due to the more variables as predictors for the optimization. The results imply that CCD has good capability for optimization on less data.

7. Conclusions

OLS, the Tobit model, and CCD can be applied to optimize the specific surface area of clay by the activation method. In this study, OLS and the Tobit model showed similar performance to CCD, but CCD showed the best performance. The optimum values for maximizing the SSA by clay activation are a temperature (*T*) of 415.029 °C and a sulfuric acid concentration (*C*) of 0.488 M, with the optimum value of 167.34 m^2/g.

Central composite design can be used for modeling the optimum conditions for clay activation with temperature and acid concentration variables. Central composite design will show the most accurate optimum condition as it can be applied to a range of optimum conditions with better fitness than OLS regression.

Acknowledgement

The authors acknowledge KEMENRISTEK-DIKTI for supporting this research.

Nomenclatures

α	Significance level (alpha)
F	F-test, the result of a test where the null hypothesis is that all of the regression coefficients are equal to zero
p-value	Estimated probability of rejecting the null hypothesis (H0) of a study question when that hypothesis is true
R^2	Coefficient of determination

Greek Symbols

β	Parameter of coefficient regression
ε	Disturbance term
μ	Mean of normal distribution
σ^2	Variance of normal distribution
X	Independent variable
Y	Dependent variable

Abbreviations

ANOVA	Analysis of Variance
C	Concentration of Acid in the Chemical Treatment
Df	Degree of Freedom
L	Likelihood Function
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimation
MS	Mean of Sum Square
OLS	Ordinary Least Squares
SS	Sum of Square
SSA	Specific Surface Area
T	Temperature of Calcination
VIF	Variance Inflation Factor
Z	Surface Area

References

1. Nandiyanto, A.B.D; Putra, Z.A.; Andika, R.; Bilad, M.R.; Kurniawan, T.; Zulhijah, R.; and Hamida, I. (2017). Porous activated carbon particles from rice straw waste and their adsorption properties. *Journal of Engineering Science and Technology (JESTEC)*, Special Issue, 12, 1-11.
2. Ragadhita, R; Nandiyanto, A.B.D.; Nugraha, W.C.; and Mudzakir, A. (2019). Adsorption isotherm of mesopore-free submicron silica particles from rice husk. *Journal of Engineering Science and Technology* (*JESTEC*), 14(4), 2052-2062.
3. Garba, Z.N.; Bello, I.; Galadima, A.; and Lawal, A.Y. (2015). Optimization of adsorption conditions using central composite design for the removal of copper (II) and lead (II) by defatted papaya seed. *Karbala International Journal of Modern Science*, 2, 20-28.
4. Sun, Q.; Hu, X.; Zheng, S.; Sun, Z.; Liu, S.; Li, H. (2015). Influence of calcination temperature on the structural, adsorption and photocatalytic properties of TiO2 nanoparticles supported on natural zeolite. *Powder Technology*, 274, 88-97.
5. Rawlings, J.O.; Pentula, S.G.; David, A. Dickey. (1998). *Applied regression analysis: A research tool* (2nd ed.). Springer-Verlag New York, Inc.
6. Hosmer, D.W.; and Lemeshow, S. (2000). *Applied logistic regression* (2nd ed). Wiley publisher.
7. Deaton, A. (1998). *The analysis of household surveys. John Hopkins* University *Press*. London.
8. Hernández, B.; and Garcés, T. (2002). Application of tobit regression models in modelling censored epidemiological variables. *Gaceta Sanitaria* , 16, 188-195.
9. Mourabet, M; El Rhilassi, A.; El Boujaady, H.; Bennani-Ziatni, M.; and Taitai, A. (2017). Use of response surface methodology for optimization of fluoride adsorption in an aqueous solution by Brushite. *Arabian Journal of Chemistry*, 10(2), S3292-S3302.
10. Sudamalla, P.; Saravanan, P.; and Matheswaran, M. (2012). Optimization of operating parameters using response surface methodology for adsorption of Crystal Violet by activated Carbon prepared from Mango Kernel. *Sustainable Environmental Research*, 22, 1-7.
11. Asghar, A.; Raman, A.A.A.; and Daud, W.M.A.W. (2014). A Comparison of central composite design and Taguchi method for optimizing Fenton process. *Scientific World Journal*. Article ID 86912.
12. Sadhukhan, B.; Mondal, N.K.; and Chattoraj, S. (2016). Optimisation using central composite design (CCD) and the desirability function for sorption of methylene blue from aqueous solution onto Lemna major. *Karbala International Journal of Modern Science,* 2, 145-155.
13. Sarrai, A.E.; Hanini, S.; Merzouk, N.K.; Tassalit, D.; Szabó, T.; Hernádi, K.; and Nagy, L. (2016). Using central composite experimental design to optimize the degradation of Tylosin from aqueous solution by Photo-Fenton reaction. *Materials*, 9*(*6), 428.
14. Austin, P.C.; Escobar, M.; and Kopec, J.A. (2000). The use of the Tobit model for analyzing measures of health status. *Quality of Life Research,* 9, 901-910.

15. Henningsen, A. (2011). CensReg: Censored regression (Tobit) models. R package version 0.5. Retrieved January 16, 2020, from https://cran.r-project.org/web/packages/censReg/censReg.pdf.
16. Henningsen, A. (2012). Estimating censored regression models in R using the censReg package. Retrieved January 16, 2020, from https://cran.r-project.org/web/packages/censReg/vignettes/censReg.pdf.
17. Toomet, O. and Henningsen, A. (2020). Tools for maximum likelihood estimation. R package version 0.7. Retrieved January 16, 2020, from http://cran.r-project.org/package=maxLik.
18. Dhawane, S.H.; Kumar, T.; and Halder, G. (2015). Central composite design approach towards optimization of flamboyant pods derived steam activated carbon for its use as heterogeneous catalyst in transesterification of Hevea brasiliensis oil. *Energy Conversion and Management* , 100, 277-287.
19. Vaez, M.; Moghaddam, A.Z.; and Alijani, S. (2012). Optimization and modeling of photocatalytic degradation of Azo dye using a response surface methodology (RSM) based on the central composite design with immobilized Titania Nanoparticles. *Industrial Engineering and Chemical Research*, 11, 4199-4207.
20. Petrović, S.; Rožić, L.; Vuković, Z., Novaković, T.; and Stanisavljev, D. (2012). Response surface optimization for activation of bentonite using microwave irradiation. *Clays and Clay Minerals*, 60, 32-39.
21. Farombi, A.G.; Amuda, O.S.; Alade, A.O.; Okoya, A.A.; Adebisie, S.A. (2018). Central composite design for optimization of preparation conditions and characterization of hydroxyapatite produced from catfish bones. *Beni-Suef University Journal of Basic and Applied Sciences*, 7, 474-480.
22. Rožić, L.; Novaković, T.; and Petrović, S. (2010). Modeling and optimization process parameters of acid activation of bentonite by response surface methodology. *Applied Clay Science*, 48, 154-158.
23. Novaković, T.; Rožić, L.; Petrović, S.; and Rosić, A. (2008). Synthesis and characterization of acid-activated Serbian smectite clays obtained by statistically designed experiments. *Chemical Engineering Journal*, 137, 436-442.
24. Elhammoudi, N.; Oumam, M.; Abourriche, A.; Mansouri, S.; Chham, A.; and Hannache, H. (2018). Modeling using the response surface methodology the activation process of Moroccan oil shale for removal of Cd (II). *International Journal of Chemical Sciences*, 16(2), 1-10.