MDPI

Article

Statistical Modeling and Optimization of Process Parameters for 2,4-Dichlorophenoxyacetic Acid Removal by Using AC/PDMAEMA Hydrogel Adsorbent: Comparison of Different RSM Designs and ANN Training Methods

Irvan Dahlan [1,*], Emillia Eizleen Md Azhar [1], Siti Roshayu Hassan [2], Hamidi Abdul Aziz [3] and Yung-Tse Hung [4]

1 School of Chemical Engineering, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, Nibong Tebal 14300, Pulau Pinang, Malaysia
2 Faculty of Bioengineering and Technology, Universiti Malaysia Kelantan, Jeli Campus, Jeli 17600, Kelantan, Malaysia
3 School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, Seri Ampangan, Nibong Tebal 14300, Pulau Pinang, Malaysia
4 Department of Civil and Environmental Engineering, Cleveland State University, Cleveland, OH 44115, USA
* Correspondence: chirvan@usm.my; Tel.: +60-4-599-6463

Abstract: In this study, the response surface methodology (RSM) and artificial neural network (ANN) were employed to study the adsorption process of 2,4-dichlorophenoxyacetic acid (2,4-D) by using modified hydrogel, i.e., activated carbon poly(dimethylaminoethyl methacrylate) (AC/PDMAEMA hydrogel). The effect of pH, the initial concentration of 2,4-D and the activated carbon content on the removal of 2,4-D and adsorption capacity were investigated through the face-centered composite design (FCCD), optimal design and two-level factorial design. The response surface plot suggested that higher removal of 2,4-D and adsorption capacity could be achieved at the higher initial concentration of 2,4-D and lower pH and activated carbon content. The modeling and optimization for the adsorption process of 2,4-D were also carried out by different design methods of RSM and different training methods of ANN. It was found that among the three design methods of RSM, the optimal design has the highest accuracy for the prediction of 2,4-D removal and adsorption capacity (R^2 = 0.9958 and R^2 = 0.9998, respectively). The numerical optimization of the optimal design found that the maximum removal of 2,4-D and adsorption capacity of 65.01% and 65.29 mg/g, respectively, were obtained at a pH of 3, initial concentration of 2,4-D of 94.52 mg/L and 2.5 wt% of activated carbon. Apart from the optimization of process parameters, the neural network architecture was also optimized by trial and error with different numbers of hidden neurons in the layers to obtain the best performance of the response. The optimization of the neural network was performed with different training methods. It was found that among the three training methods of the ANN model, the Bayesian Regularization method had the highest R^2 and lowest mean square error (MSE) with the optimum network architecture of 3:9:2. The optimum condition obtained from RSM was also simulated with the optimized neural network architecture to validate the responses and adequacy of the RSM model.

Keywords: adsorption; response surface methodology; artificial neural network; activated carbon; adsorption capacity; modeling; optimization

1. Introduction

Recently, water pollution has become one of the most common issues in the world. Agriculture practices, which contribute about 70% usage of surface water supplies, are one of the major causes of water pollution due to the usage of herbicides or pesticides by the farmer [1,2]. Among different types of herbicides, 2,4-dichlorophenoxyacetic acid (2,4-D)

is the most popular due to its cheaper price [3–5]. Its residue is often found in surface and groundwater due to its high potential of leach ability, low sorption of soil and poor biodegradability [5–7]. The International Agency for Research on Cancer has classified 2,4-D as very toxic due to its possible carcinogen and mutagen impact on humans [8]. Therefore, several methods such as photocatalytic degradation, advanced oxidation, electrochemical oxidation, biological treatment, ion exchange, membrane technology and activated carbon adsorption have been proposed by researchers to remove 2,4-D from water and soils. Among these methods, adsorption is the most effective and widely utilized in the industry to remove hazardous organic and inorganic pollutants in water due to its simple, cost-effective process and high flexibility in design and operation [4,9,10].

Currently, activated carbon (AC) is used as a common adsorbent in the adsorption process. However, AC adsorbents have high manufacturing costs mainly contributed by the raw materials and additional agents used to improve their adsorption capacity [4]. Therefore, many studies have been conducted to find a low-cost adsorbent with high adsorption efficiency. Recently, the use of modified hydrogel adsorbents has been found to greatly improve the adsorption efficiency and overcome the problem of the high cost of activated carbon [4,11]. The hydrophilic properties of hydrogel due to its three-dimensional network and porous structure make them able to adsorb large amounts of water [11,12]. Apart from that, the hydrogel is also able to improve adsorption efficiency by entrapment of different particles inside its network [11].

In the past few decades, many studies have been conducted based on the conventional method, one variable at a time (OVAT), which requires a lot of time. This is because a large number of experiments is needed to screen all variables independently [13]. Consequently, the study has a high cost due to the high number of tests and runs needed. Therefore, multivariate statistics techniques such as response surface methodology (RSM) and artificial neural network (ANN) are preferred nowadays to overcome the limitations of traditional methods since they offer a lower cost and a smaller number of experiments and are able to describe the interaction between the independent variables. RSM works by building an empirical model that fits the best for the quantitative data [14]. It is also useful in the optimization of processes apart from analyzing the interaction between the independent variable and dependent variable [13]. Meanwhile, ANN is an attractive approach for the modeling of multiple nonlinear factors, where it consists of three layers: input, hidden and output layers [15]. The modeling and optimization of the adsorption process with different designs and training methods in RSM and ANN, respectively, could offer a variety of performances in terms of optimum conditions for the process and prediction capability.

Recently, many comparative analyses between RSM and ANN have been conducted to compare the performance of the adsorption process. Kang et al. [16] compared both ANN and RSM for the prediction of diclofenac (anti-inflammatory drug) removal rate and final pH. Both the diclofenac removal rate and the final pH could be predicted using the developed RSM model (i.e., central composite design) using the quartic formula. From the ANOVA test, it was revealed that two input parameters (i.e., initial pH and adsorbent dosage) had a significant impact on both the diclofenac removal rate and the final pH. Additionally, the developed ANN model with the 4:11:6:2 topology accurately predicted the diclofenac removal rate and the final pH. From this comparison study, the ANN model showed better predictability for the diclofenac removal rate and the final pH than the RSM model. Prasad and Yadav [17] studied the effectiveness of water hyacinth (WH) as an adsorbent to remove methylene blue dye from colored wastewater by using a three-level central composite design (CCD) of the RSM model, and Levenberg–Marquardt (LM) backpropagation algorithms were applied for training the ANN model. Both models provided good quality prediction for three independent variables (i.e., initial pH, dye concentration and WH dose). Meanwhile, Sen et al. [18] investigated the optimization and modeling of RSM and ANN in the biosorption of chromium (VI) ions from an aqueous solution using cyanobacterial biomass. Both the central composite design of RSM and the

developed ANN model are able to predict and optimize the removal of chromium (VI) at various operating conditions with reasonably high accuracy.

Selecting an RSM design is important to get better prediction and optimum results as there are many different designs in RSM. On top of that, selection of architecture in the neural network especially the number of hidden layers or nodes could affect the performance of the process [19,20]. Nevertheless, none of the reported studies have compared the performance of the adsorption process with different RSM designs and ANN training methods. This is because most of the researchers only focused on the predictive capability between the commonly used design methods and training methods of both RSM and ANN. In addition, no information exists with respect to the modeling and optimization of 2,4-D adsorption by using activated carbon poly(dimethylaminoethyl methacrylate) (AC/PDMAEMA hydrogel), except for the study conducted by Taktak et al. [4]. Taktak et al. [4], however, only focused on the prediction and optimization of 2,4-D removal with modified hydrogel by using a face-centered composite design (FCCD) of RSM. It is important to ensure the optimum condition that can maximize the performance of the adsorption process by taking into account the selection of the right RSM design and the ANN training method. Therefore, the key objective of this study was to predict and optimize the 2,4-D removal by modified hydrogel (AC/PDMAEMA hydrogel) using different RSM designs and ANN training methods by employing the experimental data obtained by Taktak et al. [4]. The interaction of independent variables (such as pH, initial concentration of 2,4-D and activated carbon content) toward the removal of 2,4-D and adsorption capacity were also investigated.

2. Materials and Methods

The experimental data on the batch adsorption of 2,4-D with modified hydrogel by using a face-centered composite design (FCCD) of RSM were obtained from the study conducted by Taktak et al. [4]. The modified hydrogel was prepared using activated carbon (extracted from pomegranate husk), (dimethylamino) ethyl methacrylate (DMAEMA), N, N'-methylenebisacrylamide and ammonium persulfate. In their study, the percentage removal of 2,4-D (%) and adsorption capacity of 2,4-D (mg/g) were taken as the responses and were calculated using the following equations.

$$\text{Removal of } 2,4-\text{D } (\%) = \frac{C_0 - C_e}{C_0} \times 100\% \tag{1}$$

$$\text{Adsorption capacity}(\text{mg/g}) = \frac{(C_0 - C_e)}{m} \times V \tag{2}$$

where C_0 and C_e are the initial and equilibrium concentrations of 2,4-D (mg/L), respectively, m is the mass of the hydrogel adsorbent (g), and V is the solution volume (mL).

2.1. Response Surface Modeling

In the present study, the adsorption of 2,4-D was studied to determine the optimum condition for the maximum removal of 2,4-D and adsorption capacity. The influence of pH, initial concentration of 2,4-D and activated carbon content on the responses was also investigated through the RSM study. In general, the independent variable was varied at two levels, as shown in Table 1. The range for the independent variables was selected based on the previous study [4].

Table 1. Range of independent variables' coded and actual values.

Factors (Independent Variables)	Ranges of Coded and Actual Values	
	(−1)	(+1)
pH	3	9
Initial concentration of 2,4-D (mg/L)	20	100
Activated carbon content (%)	2.5	20

For the RSM modeling, Design Expert V12.0 (Stat-Ease, Inc., Minneapolis, MN, USA) software was used, and various design techniques were employed, i.e., face-centered composite design (FCCD), optimal design and two-level factorial design. Based on the data inserted in the design matrix, Design Expert V12.0 software generated an empirical model (i.e., quadratic model) as shown below to describe the relationship between the variables and the responses [14].

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{3}$$

where $\hat{y}$ is the response or output, x_i and x_j are the input factors, β_0, β_i, β_{ii}, β_{ij} are coefficients for intercept, linear, quadratic and interaction parameters, respectively, and ε is the residual associated with the experiments [21]. Since different design methods have different steps in designing the experiment, the details on each of these steps are further explained in the following sections.

2.1.1. Face-Centered Composite Design (FCCD)

The modeling was carried out by selecting the numeric factors as shown in Table 1. The generated experimental design matrix in the FCCD (with the coded distance (α) of 1) consisted of 20 runs where 8 of the runs were the factorial point, 6 of them were the axial point, and another 6 were the center point.

2.1.2. Optimal Design

In this design, the constraint was built by considering the problem vertex and the constraint point for each factor. The problem vertex was selected based on the level of factors that gave low responses. Based on the previous experimental data [4], the responses were low when a pH of 9, 20 wt% activated carbon and 20 mg/L initial concentration of 2,4-D were used. Hence, these values were chosen as the problem vertex to be excluded from the design space. Based on the problem vertexes, the constraint points for each factor were chosen by considering the point that would be feasible to be run by the design. Table 2 summarizes the input parameters for each factor that was required to be inserted in the constraint tool.

Table 2. Value of input parameters for the building of constraint equation.

Label	Factors	Low Actual	High Actual	Vertex	< > Skip	Constraint Point
A	pH	3	9	9	A <	3
B	Initial concentration of 2,4-D	20	100	20	B >	100
C	Activated carbon content	2.5	20	20	C <	2.5

Based on the value of the input parameters, the constraint equation was built. Equation (4) below shows the constraint equation generated by the parameters from Table 2.

$$280A - 21B + 96C \leq 2340 \tag{4}$$

Prior to the final step of design, the search method of point and optimality for the design were chosen such that it could give the best performance to the data. There were 20 runs required where 10 of them were the required model point, 5 of the runs were the replicate point, and the rest were the lack-of-fit point. Since the custom design generated an unusual combination of factors in the experimental design, the response for the factors was generated based on an empirical model built by the FCCD [4].

2.1.3. Two-Level Factorial Design

The building of this design started by selecting the number of factors that needed to be studied. Based on the selected number of factors, there were only 8 runs needed in this

design. The design of the experiment continued with the power calculation which included determination of the signal and noise. Signal is the smallest change in response that could be considered an achievement, and its value was defined by the user itself. Meanwhile, the noise was taken from the standard deviation as stated in the previous study [4]. Table 3 summarizes the input parameters inserted for the design power calculation. It was important to ensure that the design power was greater than 80% as it represents the probability of success for the effect that we wanted to detect in the study [22].

Table 3. Summary of input parameters inserted for the design power calculation.

Name	Units	Difference to Detect Delta (Signal)	Estimated Standard Deviation (Noise)	Signal/Noise Ratio
Removal of 2,4-D	%	10	2.21	4.525
Adsorption capacity	mg/g	10	1.82	5.495

The performance of the responses predicted by the empirical model built by the software was analyzed using analysis of variance (ANOVA). The values of the coefficient of variation (C.V) and coefficient of determination (R^2) were used to identify the model with the best fit. The model with the best fit had a lower C.V and a higher R^2. The adequacy of the model was determined from the significance of the model and the lack of fit. The significance of the model or lack of fit was determined based on the *p*-value or 'Prob>F' displayed in the ANOVA section of the software. The model must have a value of 'Prob > f' or a *p* value of less than 0.05 to significantly display the relationship between the response and the factors. Meanwhile, the lack of fit of the model must be insignificant ($p > 0.05$) for the model to fit well with the data [22]. The transformation of the data or model reduction was performed to improve the statistical performance of the selected model, whereby the Box–Cox graph in the diagnostic tab of the Design Expert software was used as a reference. Numerical optimization was performed for each design method to obtain the optimum condition of the adsorption process. All the independent variables were kept in range, while the responses were kept at a maximum. The restriction for the upper and lower limits of the responses was used to ensure a unique optimum condition (only 1 solution) at high desirability (more than 0.9) suggested by the software.

2.2. Neural Network Modeling

In the present study, Matlab R2021a (The MathWorks, Inc., Natick, MA, USA) software was used to build the ANN model, using the feedforward backpropagation network with the learning method 'learngdm' (gradient descent with momentum weight and bias learning function). The neural network created consisted of an input layer (i.e., pH, initial concentration of 2,4-D and activated carbon content), an output layer (i.e., removal of 2,4-D and adsorption capacity) and a hidden layer. The hidden nodes in the hidden layer were adjusted from 1 to 10. Trifonov et. al. [23] suggested that the optimum number of neurons in the hidden layer could be estimated by N/2, where N is the number of input variables or experimental data. The tangent sigmoid transfer function (tansig) and linear transfer function (purelin) were applied to the hidden layer and output layer, respectively. Three different training methods (i.e., Levenberg–Marquardt, Bayesian and Scaled Conjugated Gradient) were used in the ANN models, and the data division in the simulation was set at default to divide randomly.

The experimental data taken from the previous study [4] were used to constitute the optimum architecture of the ANN model, whereby the network was trained until a high overall correlation coefficient (R) value for training, testing, validation and all data sets was obtained. In other words, the R value was the stopping point for the training of the network. According to Mourabet et al. [24], the value of the overall correlation coefficient (R) could be used as a measure of the network's predictive capability. Prior to the training of the network, the data were normalized within a range of 0 (new x_{min}) to 1 (new x_{max})

using the following equation to obtain fast convergee and minimal mean square error (MSE) values [25].

$$x_n = 0.8 \left(\frac{x_i - x_{min}}{x_{max} - x_{min}} \right) + 0.1 \tag{5}$$

where x_n is the normalized value of x_i, and x_{min} and x_{max} are the minimum and maximum values of x_i, respectively. After a desirable result for a network was achieved, the weight, bias and the output data predicted were recorded. Based on the experimental response and predicted response of the network, the values of the MSE and the coefficient of determination (R^2) were calculated using the following equations.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(\left| y_{prd,i} - y_{exp,i} \right| \right)^2 \tag{6}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(\left| y_{prd,i} - y_{exp,i} \right| \right)^2}{\sum_{i=1}^{N} \left(\left| y_{exp,i} - y_m \right| \right)^2} \tag{7}$$

where N is the number of data, $y_{prd,i}$ is the ith predicted property characteristic, $y_{exp,i}$ is the ith measured value, and y_m is the mean value of $y_{exp,i}$. The calculation for the MSE and R^2 was performed manually since the 'nntool' command was used instead of 'nnstart' to generate the neural network toolbox. Note that each network was trained separately, and the best performance of networks (with different training methods) over different numbers of hidden nodes in the architecture was chosen according to the value of MSE, R^2 and overall R value.

The chosen optimized architecture of the ANN for each training method was used in the post analysis of the result to validate both responses predicted by RSM models and its model adequacy. The post analysis of the result was performed by running the optimum condition generated by RSM designs in the selected optimized ANN model. The predicted results under optimum conditions were inserted in the confirmation tab in the Design Expert V12.0 software, and the average result was observed. If the average result was predicted to fall within the 95% prediction interval (PI), then the empirical model generated by different designs of RSM was useable even though they had a significant lack of fit [22].

2.3. Comparative Analysis of RSM and ANN Models

The comparisons were made with respect to different training methods and designs in the ANN and RSM, respectively. The values of calculated R^2 and MSE were used in determining the performance of the ANN model. However, the performance of RSM models was determined by observing the values of R^2 and the coefficient of variation (C.V, %) generated by the Design Expert V12.0 software.

3. Results and Discussion

3.1. Response Surface Methodology (RSM)

Design Expert V12.0 software was used to study the effect of pH, initial concentration of 2,4-D and the activated carbon content on the removal of 2,4-D and adsorption capacity of modified hydrogel. The predictive modeling and optimization were performed using different designs of RSM.

3.1.1. Predictive Modeling

All the empirical equations (suggested by the software) for predicting the removal of 2,4-D (Y1) and adsorption capacity (Y2) were expressed in coded form, where A is the pH, B is the initial concentration of 2,4-D, and C is the activated carbon content. The generated empirical equation for the FCCD is shown in Equation (8) as a reduced quadratic model for the removal of 2,4-D, while Equation (9) represents the two-factor interaction (2FI) model for predicting adsorption capacity.

$$Y1 = 26.7455 - 17.7766\ A + 5.8641\ B - 6.5686\ C - 1.96863\ AB + 1.74688\ AC + 7.19018\ A2 - 9.59932\ B^2 + 5.41618\ C^2 \tag{8}$$

$$Y2 = 22.0559 - 13.4036\ A + 16.225\ B - 4.197\ C - 9.01788\ AB + 1.26788\ AC - 2.34113\ BC \tag{9}$$

The developed empirical equation generated in the optimal design is shown in Equations (10) and (11). The reduced quadratic model was suggested by the software for the removal of 2,4-D, and the quadratic model was recommended for predicting adsorption capacity.

$$Y1 = 5.07314 - 1.68647\ A + 0.561398\ B - 0.607189\ C + 0.0849027\ BC + 0.47414\ A^2 - 0.897096\ B^2 + 0.479117\ C^2 \tag{10}$$

$$Y2 = 4.7828 - 1.5333\ A + 1.92787\ B - 0.552366\ C - 0.386627\ AB - 0.072154\ AC - 0.0298015\ BC - 0.322986\ A^2 \\ - 0.48975\ B^2 - 0.0519347\ C^2 \tag{11}$$

On top of that, the empirical model generated in the two-level factorial design is shown in Equations (12) and (13). The main effect model was suggested by the software for the removal of 2,4-D. Meanwhile, a reduced two-factor interaction (2FI) model was recommended for predicting adsorption capacity.

$$Y1 = 5.1444 - 1.6624\ A + 0.5010\ B - 0.6624\ C \tag{12}$$

$$Y2 = 22.3056 - 13.0759\ A + 15.6536\ B - 9.01787\ AB \tag{13}$$

According to Aklilu et al. [25], the positive and negative signs in the model represent both the synergetic and antagonistic effects of the factors, respectively. Hence, the negative symbol of the coefficient of the pH (A) and activated carbon content (C) indicate that these factors had a negative impact on both the removal of 2,4-D and adsorption capacity. In contrast, the positive coefficient of the initial concentration of 2,4-D (B) indicated the increase in the removal of 2,4-D and adsorption capacity. These interactions of model terms with the responses were in agreement with the study conducted by Taktak et al. [4].

3.1.2. Statistical Analysis

The statistical analysis for each design was performed using ANOVA to determine the adequacy of the models generated by the software as shown in Tables 4–6. In this study, both models (with respect to the removal of 2,4-D and adsorption capacity of modified hydrogel) for all the designs were significant ($p < 0.05$). The lack of fit for the FCCD was found to be significant ($p < 0.05$) for both models. The software cannot generate/calculate the lack-of-fit F-statistic for the other two designs. The lack of fit was found to be significant as the pure error was very low since the original data taken from the previous study [4] have very little variation in the responses for factors at the center points. This may indicate that the center points could not capture all normal process variations in the system. Despite all the effort in the transformation of the model and reduction of the model, the lack of fit for all models was still found to be significant. Therefore, the post analysis of the result predicted by the RSM model was performed to determine the model's usefulness.

Although the lack-of-fit analysis shows that the models do not fit well with the data, the adequacy of the model could also be determined from other fit statistic parameters such as R^2 and C.V (as shown in Tables 4–6 above). The values of R^2 for the removal of 2,4-D were 0.9886, 0.9958 and 0.9960 while the values of R^2 for adsorption capacity were 0.9920, 0.9998 and 0.9516 in the FCCD, optimal design and two-level factorial design, respectively. According to Aklilu et al. [25], the value of R^2 must be at least 0.8 for the model to be a good fit. Therefore, the predicted data in all the models were in good agreement with the experimental data since these three designs have values of R^2 of more than 0.8. In addition, C.V is the ratio of standard error in predicting the mean value of the actual response. In other words, C.V could be used to represent the repeatability of the model. The values of C.V for the removal of 2,4-D were 7.58, 2.39 and 3.22 while the regression

model for adsorption capacity has a C.V of 8.25, 0.9651 and 31.89 in the FCCD, optimal design and two-level factorial design, respectively. Lower C.V indicated a better precision and reliability of the experiment [25].

Table 4. ANOVA for experimental results in FCCD.

Source	df	Removal of 2,4-D (%)			
		Sum of Squares	Mean Square	F-Value	*p*-Value
Model	8	4364.32	545.54	119.11	<0.0001
A	1	3160.08	3160.08	689.96	<0.0001
B	1	343.88	343.88	75.08	<0.0001
C	1	431.47	431.47	94.2	<0.0001
AB	1	31	31	6.77	0.0246
AC	1	24.41	24.41	5.33	0.0414
A^2	1	142.17	142.17	31.04	0.0002
B^2	1	253.4	253.4	55.33	<0.0001
C^2	1	80.67	80.67	17.61	0.0015
Residual	11	50.38	4.58		
Lack of Fit	6	50.26	8.38	334.67	<0.0001
Pure Error		0.1251	0.025		
Cor Total	4414.7				
Standard Deviation	2.14				
C.V (%)	7.58				
Adjusted R^2	0.9803				
Predicted R^2	0.9387				
R^2	0.9886				
Adeq Precision	42.09				

Source	df	Adsorption Capacity (mg/g)			
		Sum of Squares	Mean Square	F-Value	*p*-Value
Model	6	5312.5	885.42	267.73	<0.0001
A	1	1796.56	1796.56	543.25	<0.0001
B	1	2632.51	2632.51	796.02	<0.0001
C	1	176.15	176.15	53.26	<0.0001
AB	1	650.58	650.58	196.72	<0.0001
AC	1	12.86	12.86	3.89	0.0703
BC	1	43.85	43.85	13.26	0.003
Residual	13	42.99	3.31		
Lack of Fit	8	42.96	5.37	744.8	<0.0001
Pure Error	5	0.036	0.0072		
Cor Total	19	5355.49			
Standard Deviation	1.82				
C.V (%)	8.25				
Adjusted R^2	0.9883				
Predicted R^2	0.9747				
R^2	0.9920				
Adeq Precision	62.88				

Table 5. ANOVA for experimental results in optimal design.

Source	df	Removal of 2,4-D (%)			
		Sum of Squares	**Mean Square**	**F-Value**	***p*-Value**
Model	7	49.7	7.1	408.82	<0.0001
A	1	38.06	38.06	2191.21	<0.0001
B	1	4.27	4.27	245.79	<0.0001
C	1	5.06	5.06	291.59	<0.0001
BC	1	0.083	0.083	4.78	0.0493
A^2	1	0.621	0.621	35.75	<0.0001
B^2	1	2.36	2.36	135.65	<0.0001
C^2	1	0.7051	0.7051	40.6	<0.0001
Residual	12	0.2084	0.0174		
Lack of Fit	7	0.2084	0.0298		
Pure Error	5	0	0		
Cor Total	19	49.91			
Standard Deviation	0.1318				
C.V (%)	2.39				
Adjusted R^2	0.9934				
Predicted R^2	0.9855				
R^2	0.9958				
Adeq Precision	58.05				

Source	df	Adsorption Capacity (mg/g)			
		Sum of Squares	**Mean Square**	**F-Value**	***p*-Value**
Model	9	83.94	9.33	4805.91	<0.0001
A	1	21.55	21.55	11102.94	<0.0001
B	1	34.44	34.44	17747.62	<0.0001
C	1	3.02	3.02	1555.35	<0.0001
AB	1	1.07	1.07	553.8	<0.0001
AC	1	0.0463	0.0463	23.84	0.0006
BC	1	0.0076	0.0076	3.91	0.0763
A^2	1	0.2681	0.2681	138.12	<0.0001
B^2	1	0.6381	0.6381	328.81	<0.0001
C^2	1	0.0083	0.0083	4.26	0.0659
Residual	10	0.0194	0.0019		
Lack of Fit	5	0.0194	0.0039		
Pure Error	5	0	0		
Cor Total	19	83.96			
Standard Deviation	0.0441				
C.V (%)	0.9651				
Adjusted R^2	0.9996				
Predicted R^2	0.9989				
R^2	0.9998				
Adeq Precision	219.504				

From the above statistical analysis, the optimal design was found to have the best performance in predicting both responses. This is because the values of standard deviation and C.V were found to be the lowest compared to other design methods. In addition, the value of R^2 for response predicted by optimal design surpassed most of the other design methods. However, it should be noted that the result in the optimal design could be misleading since the experimental data for optimal design were obtained from the empirical model while the data for other designs were taken from the result of the experiment conducted by Taktak et al. [4]. Data generated from the empirical model may have less error compared to the raw data from the experiment itself. In addition, the empirical model

resulted from the modeling and fitting of the data was generated by the software such that its performance to predict the responses was the best that it could be.

Table 6. ANOVA for experimental results in two-level factorial design.

Source	df	Removal of 2,4-D (%)			
		Sum of Squares	Mean Square	F-value	*p*-Value
Model	3	27.63	9.21	336.04	<0.0001
A	1	22.11	22.11	806.75	<0.0001
B	1	2.01	2.01	73.28	0.001
C	1	3.51	3.51	128.09	0.0003
Residual	4	0.1096	0.0274		
Cor Total	7	27.74			
Standard Deviation	0.1655				
C.V (%)	3.22				
Adjusted R^2	0.9931				
Predicted R^2	0.9842				
R^2	0.9960				
Adeq Precision	48.2812				

Source	df	Adsorption Capacity (mg/g)			
		Sum of Squares	Mean Square	F-Value	*p*-Value
Model	3	3978.69	1326.23	26.21	0.0043
A	1	1367.83	1367.83	27.04	0.0065
B	1	1960.29	1960.29	38.75	0.0034
AB	1	650.58	650.58	12.86	0.023
Residual	4	202.37	50.59		
Cor Total	7	4181.06			
Standard Deviation	7.11				
C.V (%)	31.89				
Adjusted R^2	0.9153				
Predicted R^2	0.8064				
R^2	0.9516				
Adeq Precision	11.4243				

3.1.3. Analysis of Response Surface

To facilitate a straightforward examination of the effect of independent variables and their interaction, the developed mathematical model was utilized to construct three-dimensional (3D) response surfaces. Since the effect of process parameters on the responses had the same pattern as they originated from the same adsorption study, it was sufficient to analyze the response surface plot for one of the design methods only. In this study, the 3D response surface plot in a two-level factorial design was chosen to be analyzed. This model was selected to distinguish it from the other model (FCCD) that had been studied by Taktak et al. [4], as well as to cross-check with previous findings.

Figure 1a,b show the interaction between the initial concentration of 2,4-D and pH on the removal of 2,4-D and the adsorption capacity of the adsorbent. It can be seen from the figure that both responses increased when the initial concentration was raised from 20 to 100 mg/L. This is because the initial concentration of 2,4-D provided an important driving force to overcome all mass transfer resistance of the adsorbate between the solid and aqueous phase. According to Fick's second law, by increasing the driving force, the rate of mass transfer along the concentration gradient would consequently increase. These results are in agreement with Taktak et al. [4] and Bazrafshan et al. [9]. In contrast, both responses tended to decrease when the pH of the solution was adjusted from 3 to 9. According to Bazrafshan et al. [9], pH was the key parameter in the removal of the pollutant by the adsorbent as it could control the electrostatic force between the adsorbent and the adsorbate.

Their study on the adsorption of 2,4-D with single-wall carbon nanotubes as the adsorbent found that both the removal of 2,4-D and adsorption capacity fluctuated as the pH was increased from 3 to 13. A similar study was also reported by Safa and Bhatti [26], where they studied the usage of rice husk as an adsorbent for the removal of Everdirect Orange-3GL and Direct Blue-67 textile dyes. It was found that as the pH was increased from 2 to 12, the adsorption capacity decreased from 20.5 to 18.8 mg/g.
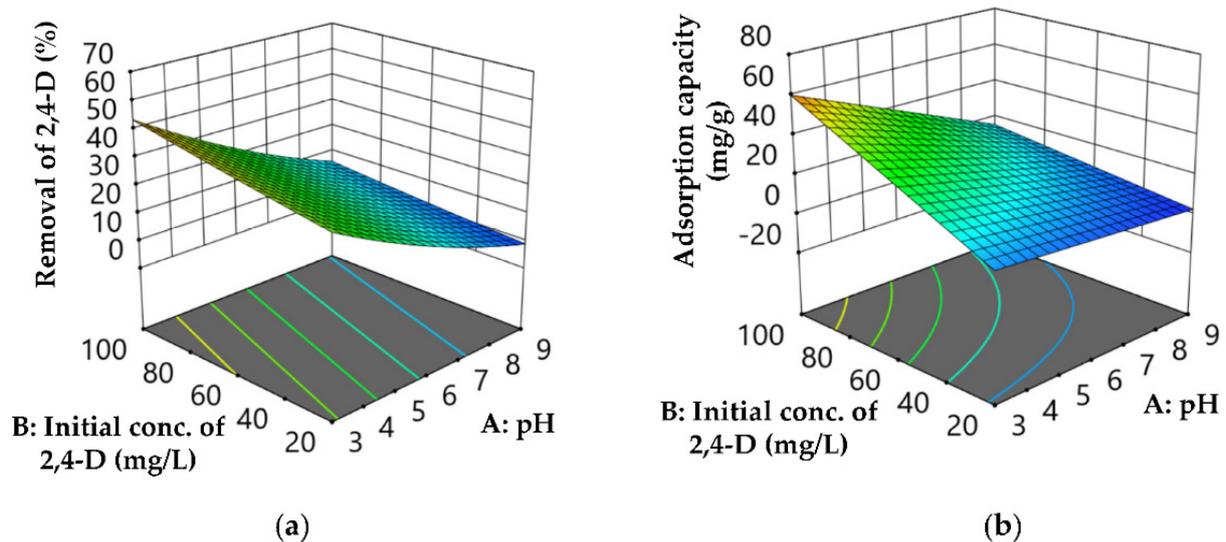

(a)   (b)

Figure 1. Two-Level Factorial Design Response surface plot for removal of 2,4-D (**a**) and adsorption capacity of adsorbent (**b**) as a function of initial concentration of 2,4-D solution (mg/L) and pH (activated carbon = 11. 25%).

The effect of pH of the solution and the activated carbon content on the removal of 2,4-D and adsorption capacity of modified hydrogel is shown in Figure 2a,b. It could be seen from the figure that both responses decreased when activated carbon content was increased from 2.5 to 20 wt%. This is because the adsorbent was made by introducing activated carbon from pomegranate husk into the polymeric network of hydrogel. Therefore, as more activated carbon was introduced, the structure of the adsorbent became less porous. Consequently, the uptake capacity of the adsorbent as well as the removal of 2,4-D decreased. A similar result was also reported by Xu et al. [27] whereby the adsorption capacities of the biosorbent prepared from rice husk toward heavy metals from simulated wastewater decreased with increasing adsorbent content.

Figure 3a,b show the interaction between the initial concentration of 2,4-D and activated carbon content on the responses. As discussed previously, a high initial concentration of 2,4-D and low activated carbon content cause a positive effect on both responses. Hence, the maximum removal of 2,4-D and adsorption capacity were obtained when the lowest level of activated carbon and the highest level of initial concertation of 2,4-D were applied in the experiment. In summary, pH plays the most important role in the adsorption of 2,4-D followed by activated carbon content and the initial concentration of 2,4-D. This could be proven from the F value and *p* value of ANOVA shown previously.
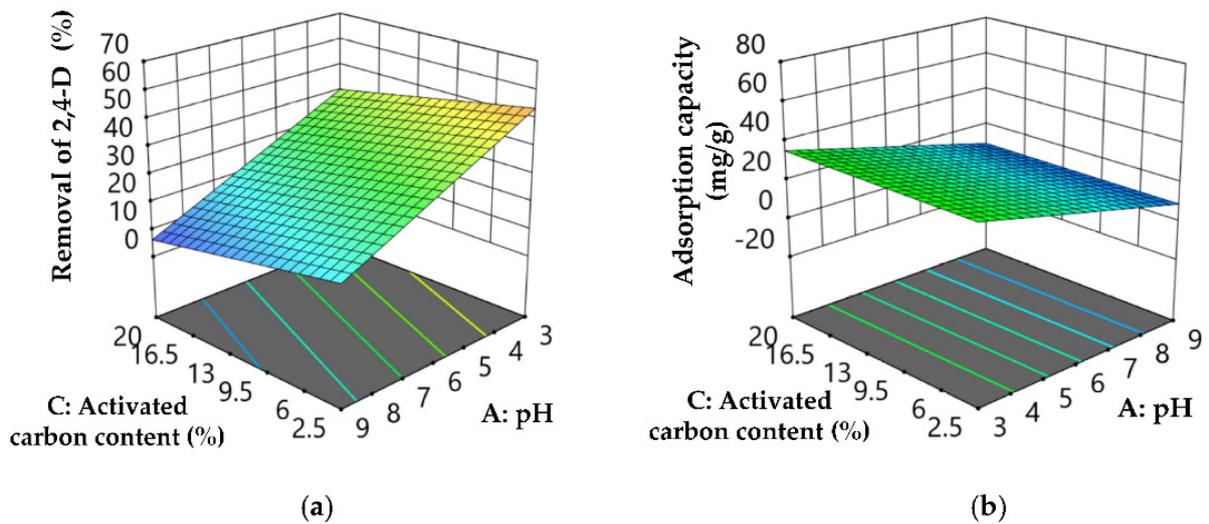
(a)

(b)

Figure 2. Two-Level Factorial Design Response surface plot for removal of 2,4-D (**a**) and adsorption capacity of adsorbent (**b**) as a function of activated carbon content (%) and pH (initial concentration of 2,4-D = 60 mg/L).
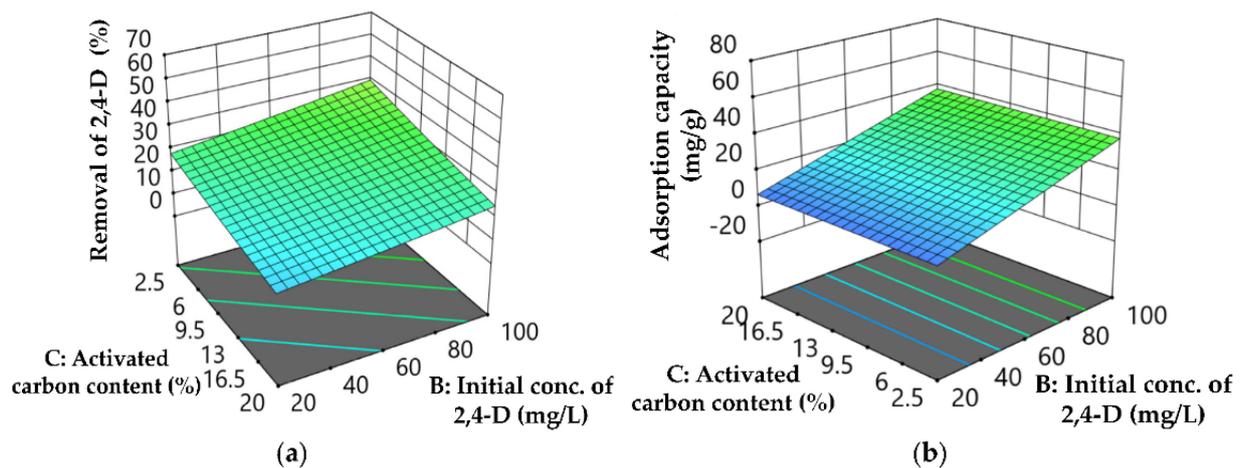

(a)

(b)

Figure 3. Two-Level Factorial Design Response surface plot for removal of 2,4-D (**a**) and adsorption capacity of adsorbent (**b**) as a function of activated carbon (%) and initial concentration of 2,4-D (pH = 6).

3.1.4. Optimization

In this study, it was aimed to achieve the maximum removal of 2,4-D and adsorption capacity of modified hydrogel. Thus, the goal chosen in the numerical optimizations was the maximum removal of 2,4-D and adsorption capacity while the process variable (pH, initial concentration of 2,4-D and activated carbon content) was set to be in a range. In the first optimization, there were too many solutions with the same desirability suggested by the software for each design. Thus, the limit for the goals was restricted. Then, the second optimization was performed by increasing the upper and lower limits of goals. By doing so, a stretch could be put in the maximization of the goal. Otherwise, many potential optimum conditions may come up such as the one encountered in the first optimization. Nevertheless, the range of the stretch had to be carefully chosen so that the desirability of the optimum condition suggested was not too low. In the current study, the range was adjusted such that the maximization of response was obtained at a unique optimum condition with a desirability of more than 0.9. The result of the second optimization is tabulated in Table 7.

Table 7. Unique solution obtained by different design method of RSM.

Design	Optimum Condition			Predicted Response		Desirability
	pH	Initial Concentration of 2,4-D (mg/L)	Activated Carbon Content (wt%)	Removal of 2,4-D (%)	Adsorption Capacity (mg/g)	
FCCD	3.00	99.97	2.52	63.63	68.47	1.000
Optimal	3.00	94.52	2.50	65.01	65.29	0.911
Two-level	3.00	100.00	2.50	63.52	60.05	0.912

3.2. Artificial Neural Network (ANN)

The architecture of ANN for different training methods (i.e., Levenberg–Marquardt, Scaled Conjugate Gradient and Bayesian Regularization) were constructed by varying the number of neurons in the hidden layer. The optimization of the ANN model was performed by training the network until the value of the correlation coefficient (R) for training, testing, validation and all prediction sets of more than 0.9 was obtained. In other words, the R value was the stopping point for the training of the network in the measurement of the network's predictive capability [24]. Meanwhile, the best architecture of the optimized ANN model for each training method was determined by the value of MSE and R^2 calculated for both responses. The following Figures 4 and 5 show the value of MSE and R^2, respectively, for both responses over different numbers of hidden nodes for different training methods.
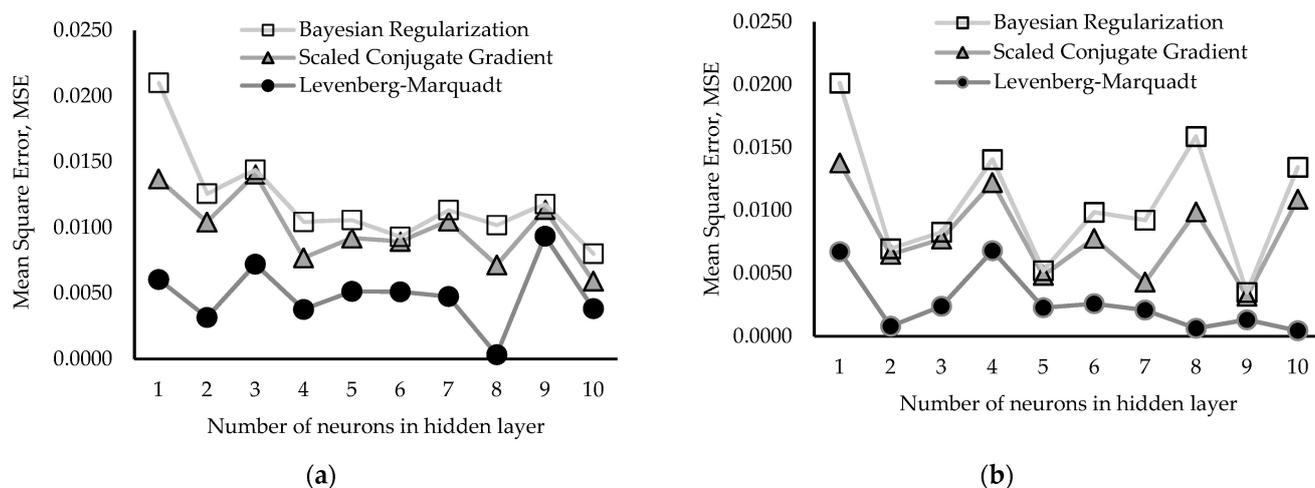

(a)

(b)

Figure 4. MSE plot over the number of hidden nodes for (**a**) removal of 2,4-D and (**b**) adsorption capacity.

The results presented in Figures 4 and 5 show that the optimized network with the Levenberg–Marquardt training method had the best performance (MSE of 0.0003 and 0.0006; R^2 of 0.9928 and 0.9833) for the removal of 2,4-D and adsorption capacity, respectively, when eight hidden nodes were used. Meanwhile, the best performance of the optimized network with the Scaled Conjugate Gradient training method was obtained when nine hidden neurons were used. The value of MSE for this network was 0.002 and 0.0019, while the value of R^2 was 0.9576 and 0.9552 for the removal of 2,4-D and adsorption capacity, respectively. On top of that, the best architecture for the Bayesian Regularization training method was found to be 3:9:2 as this model has a very low MSE (0.0004 and 0.0003) and high R^2 (0.9910 and 0.9924) for the removal of 2,4-D and adsorption capacity, respectively. High R^2 and low MSE values show that the model has high accuracy in predicting both responses. To sum up, the architecture of the optimized network for the Levenberg–Marquardt, Scaled Conjugate Gradient and Bayesian Regularization training methods was 3:8:2, 3:9:2 and 3:9:2, respectively. Hence, it was decided that these networks were used in the post analysis

of results whereby the responses under optimum conditions generated by different designs of RSM models were predicted.
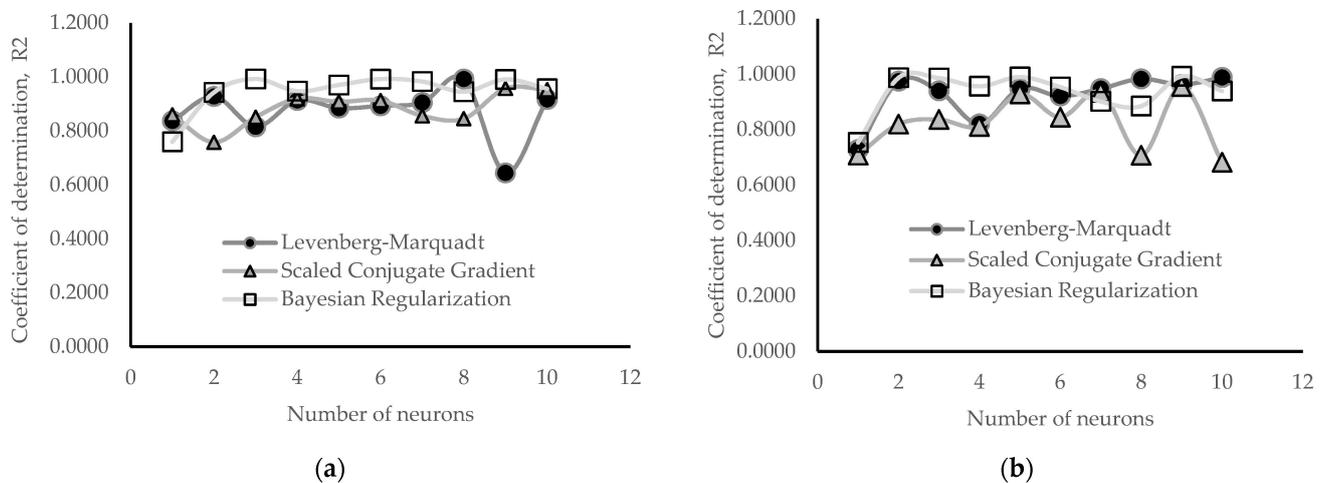

Figure 5. R^2 plot over the number of hidden nodes for (**a**) removal of 2,4-D and (**b**) adsorption capacity.

The comparison between the experimental and the computed ANN data's spread plot in training, testing and validation for the selected networks is shown in the following Figure 6. Figure 6a shows the R value of 0.99998, 0.97368, 0.98974 and 0.99451 for training, testing, validation and all data sets, respectively, for the optimized network with the Levenberg–Marquardt training method. The R value for the optimized network with the Scaled Conjugate Gradient training method is shown in Figure 6b with the value of 0.98161, 0.93959, 0.99975 and 0.97985 for training, testing, validation and all data sets, respectively. The high value of R indicated that these optimized models have good precision as the R value was a measure of the predictive ability of the networks [24]. Additionally, Figure 6c shows the R value of 0.99999, 0.9922 and 0.99632, respectively, for training, testing and all data sets for the optimized network with the Bayesian Regularization training method. There was no R value for validation data sets available in this training method since the validation stop was disabled by default in the training parameter settings. Nevertheless, this training method has its own validation built into its algorithm, where the validation was performed in the form of regularization [22]. This result was analogous to those reported by Jazayeri et al. [28] whereby no data were available on the validation performance when they estimated the output power of a photovoltaic (PV) module.

3.3. Post Analysis of Results

The post analysis was performed to validate the responses predicted by the RSM models under optimum operating conditions recommended by the software. A similar study was reported by Xu et al. [27], where three independent experiments were performed under the predicted optimum condition to confirm the model prediction. In addition, according to StateEase [22], for a case where the lack of fit was found to be significant, an additional experiment could be run under optimal conditions. In the current study, the additional experiment was replaced with the simulation of a selected optimized network of the ANN model. The comparison between the responses predicted by the RSM model and the selected optimized ANN model under optimum conditions recommended by the Design Expert software is shown in Table 8. By using the confirmation tab in the Design Expert software, the average result for all RSM models was found to fall within the 95% prediction interval (PI). Hence, the model was useable, and the result predicted by the models was validated.
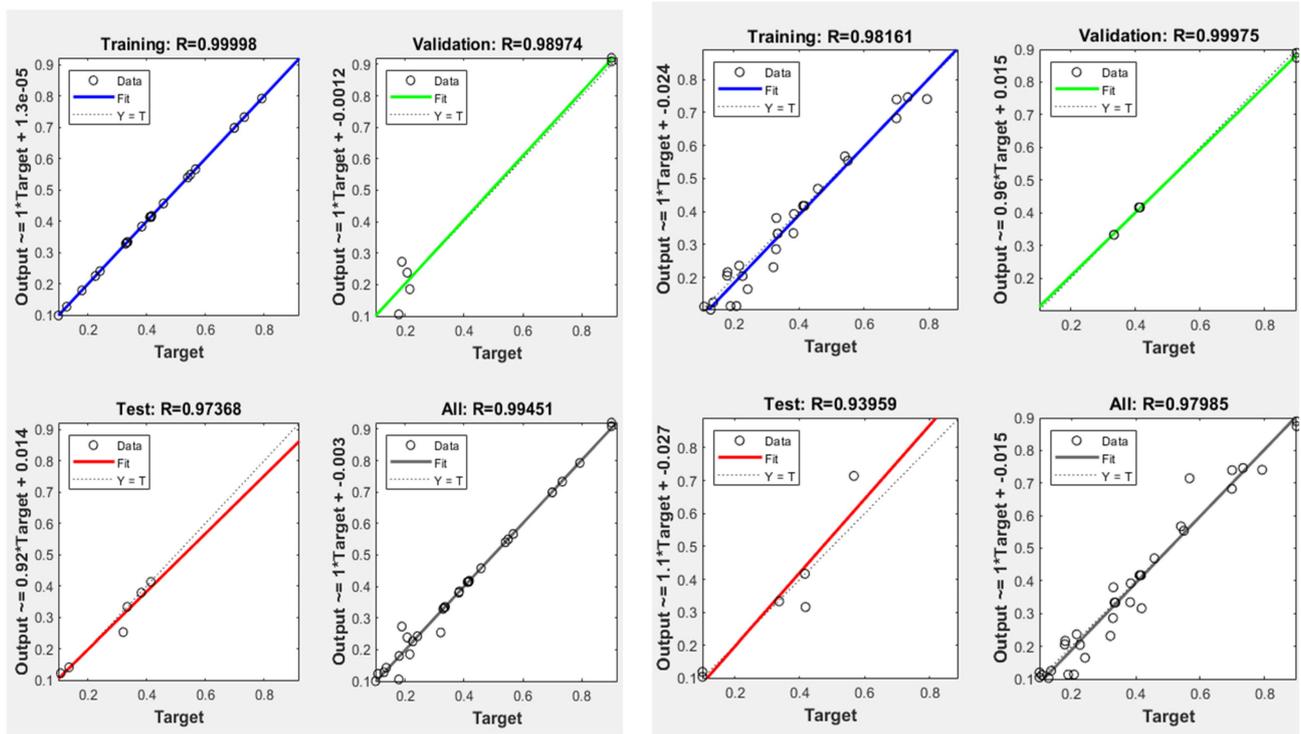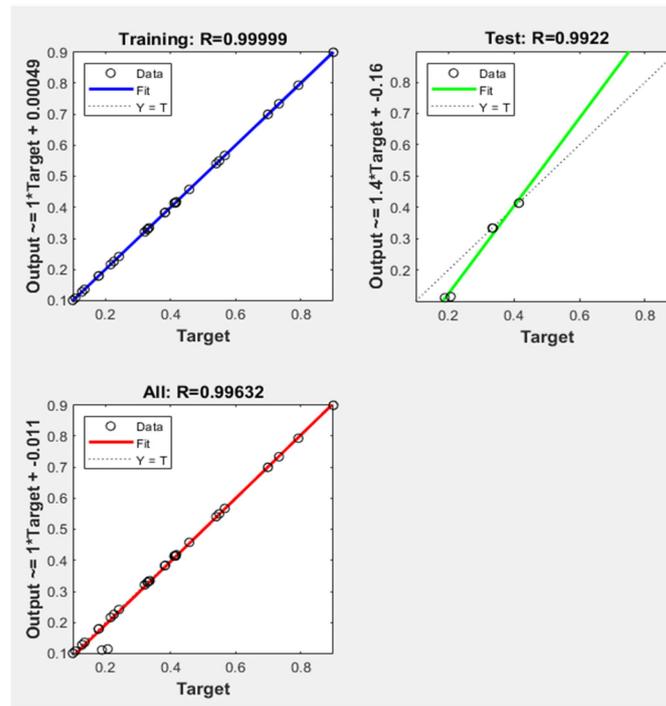
Figure 6. R value for training, testing, validation and all data sets for the best architecture of optimized ANN model with (**a**) Levenberg–Marquardt training method, (**b**) Scaled Conjugate Gradient training method and (**c**) Bayesian Regularization training methods.

Table 8. Comparison between the responses predicted by RSM model and optimized ANN model under optimum conditions.

Design	Optimum Condition			Removal of 2,4-D (%)				Adsorption Capacity (mg/g)			
	pH	Initial Concentration of 2,4-D (mg/L)	Activated Carbon Content (wt%)	RSM Predicted	Levenberg–Marquardt	Conjugated Gradient	Bayesian	RSM Predicted	Levenberg–Marquardt	Conjugated Gradient	Bayesian
FCCD	3	99.97	2.52	63.630	62.555	60.461	61.112	68.470	69.206	66.306	68.308
Optimal	3	94.52	2.5	65.010	64.713	60.323	61.094	65.280	66.555	65.677	68.234
Two-Level	3	100	2.5	63.520	62.657	60.462	61.112	60.050	69.252	66.308	68.309

4. Conclusions

In this study, the modeling and optimization for the adsorption process of 2,4-D using modified hydrogel (AC/PDMAEMA hydrogel) were conducted with different designs of RSM and different training methods of ANN. The 3D response surface plot in RSM shows that the maximum removal of 2,4-D and adsorption capacity was achieved when a high initial concentration of 2,4-D was used. In contrast, the high pH of the solution and the activated carbon content in the polymeric network of the adsorbent had a negative impact on both responses. Among different design methods of RSM, the empirical model generated by optimal design was found to have the highest value of R^2 for both responses with the lowest C.V value. However, further study needs to be conducted in the future to confirm the result since this result could be misleading due to the different sets of experimental data used in this design. In the ANN modeling, the Bayesian Regularization training method (with an architecture of 3:9:2) was found to have the best performance among different training methods with a very low MSE and a high R^2 for both responses. From the numerical optimization, the maximum removal of 2,4-D and adsorption capacity in optimal design were 65.01% and 65.29 mg/g, respectively, which were obtained at a pH of 3, initial concentration of 2,4-D of 94.52 mg/L and 2.5 wt% of activated carbon. The post analysis with the optimized ANN model found that the responses predicted by RSM under optimum conditions were validated and the generated models were useable even though they had a significant lack of fit.

References

1. Nord, N.B.; Berthelsen, N.M.S.; Milter, H.; Bester, K. Removal of herbicides from landfill leachate in biofilters stimulated by ammonium acetate. *Water* **2020**, *12*, 1649. [CrossRef]
2. Mateo-Sagasta, J.; Zadeh, S.M.; Turral, H. *Water Pollution from Agriculture: A Global Review*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
3. US EPA. Herbicides. Available online: https://www.epa.gov/caddis-vol2/caddis-volume-2-sources-stressors-responses-herbicides (accessed on 25 January 2022).

4. Taktak, F.; İlbay, Z.; Şahin, S. Evaluation of 2,4-D removal via activated carbon from pomegranate husk/polymer composite hydrogel: Optimization of process parameters through face centered composite design. *Korean J. Chem. Eng.* **2015**, *32*, 1879–1888. [CrossRef]

5. Liu, Y.; Sun, B. Unusual catalytic effect of Fe^{3+} on 2,4-dichlorophenoxyacetic acid degradation by radio frequency discharge in aqueous solution. *Water* **2022**, *14*, 1719. [CrossRef]

6. Lv, X.; Ma, Y.; Li, Y.; Yang, Q. Heterogeneous Fenton-like catalytic degradation of 2,4-dichlorophenoxyacetic acid by nano-scale zero-valent iron assembled on magnetite nanoparticles. *Water* **2020**, *12*, 2909. [CrossRef]

7. Shankar, M.V.; Anandan, S.; Venkatachalam, N.; Arabindoo, B.; Murugesan, V. Fine route for an efficient removal of 2,4-dichlorophenoxyacetic acid (2,4-D) by zeolite-supported TiO_2. *Chemosphere* **2006**, *63*, 1014–1021. [CrossRef] [PubMed]

8. Mir, M.; Zahedi, M.M.; Pourmortazavi, S.M. Removal of 2,4-dichlorophenoxyacetic acid from aqueous samples using electrospun polyacrylonitrile nanofiber-based supported liquid membrane transport. *J. Iran. Chem. Soc.* **2020**, *18*, 631–639. [CrossRef]

9. Bazrafshan, E.; Kord Mostafapour, F.; Faridi, H.; Farzadkia, M.; Sargazi, S.; Sohrabi, A. Removal of 2, 4-dichlorophenoxyacetic acid (2, 4-d) from aqueous environments using single-walled carbon nanotubes. *Health Scope* **2013**, *2*, 39–46. [CrossRef]

10. Herrera-García, U.; Castillo, J.; Patiño-Ruiz, D.; Solano, R.; Herrera, A. Activated Carbon from Yam Peels Modified with Fe_3O_4 for removal of 2,4-dichlorophenoxyacetic acid in aqueous solution. *Water* **2019**, *11*, 2342. [CrossRef]

11. Tran, V.V.; Park, D.; Lee, Y.-C. Hydrogel applications for adsorption of contaminants in water and wastewater treatment. *Environ. Sci. Pollut. Res.* **2018**, *25*, 24569–24599. [CrossRef]

12. Ahmed, E.M. Hydrogel: Preparation, characterization, and applications: A review. *J. Adv. Res.* **2015**, *6*, 105–121. [CrossRef]

13. Witek-Krowiak, A.; Chojnacka, K.; Podstawczyk, D.; Dawiec, A.; Pokomeda, K. Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process. *Bioresour. Technol.* **2014**, *160*, 150–160. [CrossRef]

14. Montgomery, D.C. *Design and Analysis of Experiments, 10th ed*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2019.

15. Kayri, M. Predictive abilities of Bayesian regularization and Levenberg–Marquardt algorithms in artificial neural networks: A comparative empirical study on social data. *Math. Comput. Appl.* **2016**, *21*, 20. [CrossRef]

16. Kang, J.-K.; Kim, Y.-G.; Lee, S.-C.; Jang, H.-Y.; Yoo, S.-H.; Kim, S.-B. Artificial neural network and response surface methodology modeling for diclofenac removal by quaternized mesoporous silica SBA-15 in aqueous solutions. *Microporous Mesoporous Mater.* **2021**, *328*, 111497. [CrossRef]

17. Prasad, R.; Yadav, K.D. Use of response surface methodology and artificial neural network approach for methylene blue removal by adsorption onto water hyacinth. *Water Conserv. Manag.* **2020**, *4*, 83–89. [CrossRef]

18. Sen, S.; Nandi, S.; Dutta, S. Application of RSM and ANN for optimization and modeling of biosorption of chromium(VI) using cyanobacterial biomass. *Appl. Water Sci.* **2018**, *8*, 148. [CrossRef]

19. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62. [CrossRef]

20. Le, T.-H.; Jang, H.; Shin, S. Determination of the optimal neural network transfer function for response surface methodology and robust design. *Appl. Sci.* **2021**, *11*, 6768. [CrossRef]

21. Nair, A.T.; Makwana, A.R.; Ahammed, M.M. The use of response surface methodology for modelling and analysis of water and wastewater treatment processes: A review. *Water Sci. Technol.* **2013**, *69*, 464–478. [CrossRef]

22. Statease. Tutorials» Two-Level Factorial. Available online: https://www.statease.com/docs/v11/tutorials/two-level-factorial/ (accessed on 29 November 2021).

23. Trifonov, R.; Yoshinov, R.; Pavlova, G.; Tsochev, G. Artificial neural network intelligent method for prediction. *AIP Conf. Proc.* **2017**, *1872*, 020021.

24. Mourabet, M.; El Rhilassi, A.; Bennani- Ziatni, M.; Taitai, A. Comparative study of artificial neural network and response surface methodology for modelling and optimization the adsorption capacity of fluoride onto apatitic tricalcium phosphate. *Univers. J. Appl. Math.* **2014**, *2*, 84–91. [CrossRef]

25. Aklilu, E.G.; Adem, A.; Kasirajan, R.; Ahmed, Y. Artificial neural network and response surface methodology for modeling and optimization of activation of lactoperoxidase system. *S. Afr. J. Chem. Eng.* **2021**, *37*, 12–22. [CrossRef]

26. Safa, Y.; Bhatti, H.N. Adsorptive removal of direct textile dyes by low cost agricultural waste: Application of factorial design analysis. *Chem. Eng. J.* **2011**, *167*, 35–41. [CrossRef]

27. Xu, M.; Yin, P.; Liu, X.; Tang, Q.; Qu, R.; Xu, Q. Utilization of rice husks modified by organomultiphosphonic acids as low-cost biosorbents for enhanced adsorption of heavy metal ions. *Bioresour. Technol.* **2013**, *149*, 420–424. [CrossRef]

28. Jazayeri, K.; Jazayeri, M.; Uysal, S. Comparative analysis of Levenberg-Marquardt and Bayesian Regularization backpropagation algorithms in photovoltaic power estimation using artificial neural network. In *Advances in Data Mining. Applications and Theoretical Aspects, ICDM 2016. Lecture Notes in Computer Science*; Perner, P., Ed.; Springer: Cham, Switzerland, 2016; Volume 9728, pp. 80–95.