CHAPTER 4

MODELLING
CHAPTER 4

MODELLING

4.1 GENERAL

In this research work, activated carbons were prepared from the three locally available carbonaceous materials. The locally available carbonaceous materials used for the preparation of the activated carbon are (1) fruit of Leucanea Glauca Benth, (2) Bark of Casuarina Equisettifolia wood, and (3) Tamarind fruit shell. Various experiments have been performed to test the abilities of these activated carbons in removing Nickel and Lead from aqueous solutions. They are categorised into (1) Batch Experiments and (2) Column Experiments.

In the Batch experiments, various types of experiments were conducted to test the efficiency of these carbons in removing Nickel and Lead from aqueous solutions. The results obtained were recorded and used for the generation of models. Similarly, Column experiments were also conducted using all these three activated carbons to remove Nickel and Lead from aqueous solutions. The results obtained were recorded and used to generate models.

4.2 REGRESSION ANALYSIS

According to Blair, “Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.”

Regression helps to estimate one variable or the dependent variable from the other variables or the independent variables. In other words, we can estimate the value of one variable, provided the values of the other variable are given. The statistical method which helps us to estimate the unknown value of one variable from the known values of the related variables is called regression. The dictionary meaning of the word regression is ‘return’ or ‘giving back’. In 1877, Sir Francis Galton first introduced the word ‘regression’.
4.2.1 SIMPLE LINEAR REGRESSION

According to Taro Yamane, "the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related casually is regression analysis."

According to Walli's and Robert, "It is often more important to find out what the relation actually is in order to estimate or predict one variable (the dependent variable), and the statistical technique appropriate in such a case is called Regression Analysis".

According to Ya-Lun-Chow, "Regression analysis is a statistical device. With the help of the regression analysis, we can estimate or predict the unknown values of one variable from the known values of another variable. In regression analysis, the independent variable is also known as the "Regressor" or "Predictor" or "Explanator" and the dependent variable is known as "regressed" or "explained" variable.

Regression equation is an algebraic method. It is an algebraic expression of the regression line. It can be classified into regression equation, regression coefficient, individual observation and group distribution.

As there are two regression lines, there are two regression equations. For the two variables X and Y, there are two regression equations. They are regression equation of X on Y and regression equation of Y on X. The former states the change in the value of X for a given change in Y and the latter states the change in the value of Y for the change in X.

Regression Equation of X on Y

$$X_{(c)} = a + b \ Y$$

Here ‘a’ and ‘b’ the two unknown constants at the parameters of the line.

By the least square method, we can find out the value of ‘a’ and ‘b’ and determine the regression line, which is known as the line of best fit. The formulae are:

$$SX = Na + bSY$$

$$SXY = aSY + bSY^2$$

28
where $N$ is number of observed pairs of values; $S X$, $S Y$, $S XY$, $S Y^2$ are the totals computed from the value of the two variables $X$ and $Y$. Thus we can fit a least square line.

**Regression Equation of $Y$ on $X$:**

$$Y_c = a + b X$$

To find out the value of $a$ and $b$ we can use the following formulae:

$$S Y = Na + b S X$$

$$S XY = a S X + b S X^2$$

We can call these equations normal equations.

### 4.2.2 MULTIPLE LINEAR REGRESSION

In the multiple linear regression, the regression co-efficients as well as the variance of the errors, are usually unknown and have to be estimated from observations. As in the simple linear regression model, we can use the least-squares criterion to determine the estimates of regression coefficients. Normal equations are arrived. It is quite easy to solve the normal equations and determine the least-squares estimates.

**Table 4.1 ANOVA table for the multiple linear regression model**

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression error</td>
<td>SSR</td>
<td>P</td>
<td>MSR=SSR/p</td>
<td>MSR/MSE</td>
</tr>
<tr>
<td>SSE</td>
<td>n-p-1</td>
<td></td>
<td>MSE=SSE/(n-p-1)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>SSTO</td>
<td>n-1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.3 USE OF REGRESSION ANALYSIS

1. Regression analysis is used in statistics and in all those fields where two or more relative variables have a tendency to go back to the average. It is used more than the correlation analysis in many scientific studies. It is widely used in social sciences like economics, natural and physical sciences. It is used to estimate the relationship between two economic variables like income and expenditure. If we know the income, we can find out the probable expenditure. Thus it is a highly valuable tool in economics and business. Most of the economic issues are based on cause and effect relationship. It is very useful for prediction purposes.

   In business also, it is very helpful to study the business predictions. Cost of production is affected by the sale of the product. Economists have arrived at many predictions and theories on the basis of regression.

2. Regression analysis predicts the value of dependent variables from the values of independent variables.

3. The regression analysis is highly useful and the regression line equation helps to estimate the value of dependent variable, when the values of independent variables are used in the equation.

4. We can calculate the coefficient of correlation ($r$) and the coefficient of determination ($r^2$) with the help of regression coefficient.
5. Using regression analysis, the statistical estimation of demand curves, production function, consumption function, etc. can be predicted.

4.3 ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) are capable of modeling highly nonlinear relationships and can be trained to accurately generalize the relationship, when presented with new, unseen data. For this reason ANN has been chosen for modelling activated carbon treatment.

ANN has been widely used for modelling control parameters, pattern recognition, signal processing, prediction, etc. The ANN is taught to model a relationship during a supervised training procedure by using series of input and associated output data.

4.3.1 BACK PROPAGATION

The very general nature of the back propagation training method is that a back propagation net (a multi-layer, feed forward net trained by back propagation) can be used to solve problems in many areas.

The main aim of back propagation is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to give responsible (good) response to the input that is similar, but not identical to that used in training (generalization).

The training of network by back propagation involves three stages: the feed forward of the input training pattern, the calculation, and back propagation of the weights. After training, application of the net involves only the computations of the feed forward phase. Even if training is slow, a trained net can produce its output very rapidly. Numerous variations of back propagation have been developed to improve the speed of the training process.

4.3.2 TRAINING OF ANN

The artificial neural network paradigm adopted in this study utilizes the back propagation algorithm. Training of neural network is essentially carried out through
the presentation of a series of example patterns of associated input and output. The input and output neurons process these inputs by their weights, summing the product and then passing the sum through a non-linear transfer function to produce a result. The S-shaped sigmoid curve is commonly used as the transfer functions.

The artificial neural network learns by modifying the weights of the neurons in response to the errors between the actual output values and the target output values. This is carried out through the gradient descent on the sum of squares of the errors, for all the training patterns. The changes in weights are in propagation to the negative of the derivative of the error term. Training is carried out by repeatedly presenting the entire set of training patterns until the average sum squared error over all the training patterns are minimized and within the tolerance specified for the problem.

The associated weights of the connections from training are stored in memory and then the network is fed with a separate set of data. In this training-testing phase, the neural network predictions are compared with the target output values. This evaluates the reliability of the neural network to generalize correct responses for the testing patterns that only broadly resemble the data in the training set. No additional learning or weight adjustments occur during this phase. Once the training and testing phase are found to be successful, the neural network can then be put to use in practical applications. The trained network will produce almost instantaneous results for the given input patterns.

The development of a back propagation neural network model essentially involves a number of stages. First, the variables have to be identified. The training set must provide a representative sample of the data containing the various distinct characteristics of the problem the neural networks are likely to encounter in the finished application.

4.3.3 USEFULNESS OF ANN

An Artificial Neural Network model has an advantage over other models, in that experts are not needed to use the model. Hence the artificial neural network models are generally used to arrive at the relationship between variables.
Recent renewed artificial neural networks are attributed to several factors. Training techniques have been developed for the more sophisticated network architectures that are able to overcome the shortcomings of the early, simple neural nets. High-speed digital computers make the simulation of neural process more feasible.

Neural nets are of interest to researchers in many areas for different reasons. Electrical engineers find numerous applications in signal processing and control theory. Computer engineers are intrigued by the potential of hardware to implement neural nets efficiently and by the application of neural nets to robotics. Computer scientists find that neural nets show promise for difficult problems in areas such as artificial intelligence and pattern recognition. For applied mathematicians, neural nets are a powerful tool for modelling probables for which explicit form of the relationship among certain variables is not known.