Identification of Algal Biomass Production with Partial Least Squares & Neural Network

Diamond Das and Madhusree Kundu

***Abstract*—Parameters like pH, nitrate concentration, depth of liquid column and temperature influenced the yield of Algal biomass in a batch culture. For 64 combinations of the aforesaid parameters, the experimentation was done and each run was observed for 10 days. The Taguchi method of design of experiments was used for creating the 64 combinations. The yield of Algal biomass was modeled as a function of pH, nitrate concentration, depth of liquid column and temperature using partial least squares (PLS) and neural networks. The predictions using both the techniques were in excellent agreement with the experimental data generated in this work.**

***Index Terms*—Algae, Bio-mass, Batch culture,Partial Least Square, Artificial Neural Network, Modeling.**

I. Introduction

Microalgae are photo synthetic microorganisms that utilize the solar energy and form potential bio-fuels, food, feeds and high value bioactives. There are a number of different types of biofuels produced by algae which includes methane produced by anaerobic digestion of algal biomass, biodiesel produced by extracting micro algal oil, and photo biologically produced biohydrogen [1]. The concept of using algae as a source is not new; rather its significance was understood with the emergence of global concerns like shortage of fossil fuels, global warming and rise in the price of fossil fuels. Algae are highest yielding field stock for biodiesel producing 250 times the amount of oil produced per acre of soybeans and 7- 31 times more oil than palm oil. Waste water can be a source of algal growth; the algal growth in it helps in bioremediation of water [1].

In the present growing demand of fuels; biofuels from algae has been considered as a blessing. As estimated by earlier studies the cost of production of a kilogram of algal biomass is $2.95 and $3.80 for bioreactors and raceway ponds respectively [2]. Chisti (2007) also proposed that the cost of algal biomass production can be reduced to roughly $0.47 and $0.60 for photobioreactors and raceway ponds respectively by increasing the annual biomass production to around 10,000 tons [2]. It is assumed that algae contains 30% biomass by weight, the cost of biomass for obtaining a liter of oil is $1.41 and $1.81 for photo bioreactors and raceway ponds. The employment of a high-value co-product strategy through the integrated biorefinery approach is expected to significantly enhance the overall cost-effectiveness of microalgal biofuels production. Besides bio-fuels, algae is also known to produce a vast array of high value bioactive compounds that can be used as pharmaceutical compounds, health foods, and natural pigments [3]. Some well-studied examples include acetylic acids, α-carotene [2 & 4], vitamin B, Ketocarotenoid astaxanthin, polyunsaturated fatty acids, and lutein [5]. The economic feasibility of algal bio-fuel production can be significantly enhanced by understanding its growing pattern.

Thus the understanding of growing pattern of algae can be successfully utilized for extracting bioreactive products from harvested algal biomass, thermal processing (pyrolysis, liquefaction, or gasification) and extracting high-value chemicals from the resulting liquid; vapor and/or solid phases, and reforming/ upgrading bio-fuels for different applications. For the production of the microalgae on a larger scale the process identification and the growing pattern recognition is very important. One should also understand the importance of different parameters and their influence on algae growth.

Algae are mostly photo-autotrophs requiring only sun light, water and inorganic nutrients for its growth, but there are some species which can be considered as heterotrophs. The temperature required for the growth of microalgae is around 20 – 30 °C and the growth medium must also provided with inorganic elements that constitute algal cells. The essential elements include nitrogen, phosphorus, iron and in some cases silicon. Among all the elements provided; nitrogen (i.e. mostly in the form of nitrates) is considered as the most significant element for the algal growth [5]. The energy for growth and reproduction in microalgae is provided by normal photosynthesis. These photosynthetic microbes are known to as a potential source for oil and biodiesel production [5]. The utilization of solar energy for the production of bio-fuel is an absolute necessity for large scale production of microalgae, the light intensity and nitrate concentrations are known to affect the growth kinetics of microalgae [7 and 8]. This particular characteristic of algae has created interest among researchers for its commercial cultivation from an industrial perspective. To meet the growing demand of bio-fuels the indoor cultivation of bioreactors may not be suitable rather an open culture system is required like the photobioreactors for the large scale production of algal biomass. The algal biomass is known to have approximately 50% concentration of carbon by dry weight [9]. The carbon is obtained from the atmospheric CO_2 whose affinity in algae is known to be affected by changing pH. At times it is seen that the algae forms a layer on the surface of the media thus preventing the sunlight to further penetrate into the fluid. As a result of it there is more growth of biomass at the surface than the bottom. So the height of liquid i.e. light penetration depth is also considered as a parameter for algal growth.

Thus the different parameters known to have an effect on biomass production in an outdoor condition includes Average Temperature, pH of the medium, nitrate concentration of the

Manuscript received June 10, 2011; revised July 22, 2011.

Diamond Das and Madhusree Kundu are with Department of Chemical Engineering, National Institute of Technology, Rourkela, P.O. Box 769008, orissa, India (corresponding author to provide :e-mail: mkundu@nitrkl.ac.in)

medium and Height of the liquid column [10 & 7].

In view of this, the present work was taken up for the identification of algal growth process. In any process, there may be a large number of input variables (factors) that may be assumed *a priori* to affect the process. Screening experiments are conducted to determine the inputs and interactions of inputs that influence the process significantly. one-factor-at-a-time' method fails whenever the maximizing value of the varying factor is not independent of the levels of the other factors, i.e. interactions may exist among the variables. In addition, factorial design permits more precision in estimation of the effects of the factors than the 'one-factor-at-a-time' design, at the same number of experimental runs. Moreover, the whole set of experimental tests designed permits the estimation of an empirical model relating the yield with the parameters, expressed by polynomial equations, that comes in handy when the physics involved are unknown or too complicated to study and postulate a theoretical model [11]. Taguchi method for design of experiments is such a logical, economical and statistical method for an experimental design. In this study, Taguchi method for design of experiments was used for the first time on algal growth, to create a four level design (Table 1) for the considered parameters. The first three parameters are used to create a L_{16} array with 16 combinations. The 16 combinations are repeated at four different average temperatures making it a total 64 combinations of the aforesaid input parameters for performing the experiments; hence creating the design database. The influence of selected parameters on the algal biomass yield has been modeled using neural network as well as Partial least squares.

TABLE 1 FOUR LEVELS OF THE CONSIDERED INPUT PARAMETERS

Factors	Level 1	Level 2	Level 3	Level 4
pH	6.5	7	7.5	8
Nitrate Concentration (mg/l)	0.2	0.25	0.3	0.35
Depth of Liquid Column (cm)	5	10	15	20
Temperature	18	20	22	24

II. MATERIALS AND METHODS

A. Isolation of Algae

Water samples were collected from stagnant water ponds due to over flow near the water reservoirs, in general, and used as the source for isolating native algal strain. BG11 medium was used for the enrichment experiments [12] and the composition of the medium is presented in Table 2. 100 mL of BG11 medium was dispensed in 250 mL Erlenmeyer flasks and sterilized. 5 mL of the collected water sample was added to the flasks and mixed separately with the medium and agitated for 30 min on a rotary shaker. The flasks were then incubated under the sun light with Light & Day cycles of 12-12 h for 3 weeks. Indigenous algae were isolated from the flasks through serial dilution.

TABLE 2 COMPOSITION OF BG11 MEDIUM

Composition	Stalk (grams/100ml)	Medium (ml)
$NaNO_3$	15	10
$K_2HPO_4.7H_2O$	0.4	10
$MgSO_4.7H_2O$	0.75	10
$CaCl_2.2H_2O$	0.36	10
Citric Acid	0.06	10
Ferric Ammonium Cutrate	0.06	10
EDTA	0.01	10
Na_2CO3	0.2	10
Micronutrient		1
Micronutrient Composition	1000 ml solution(mg)	
H_3BO_3	61.0	
$MnSO_4.H_2O$	169.0	
$ZnSO_4.7H_2O$	287.0	
$CuSO_4.5H_2O$	2.5	
$(NH_4)_6 Mo_7 O_{24}.7H_2O$	12.5	

B. Experimental procedure with analysis of product

Algal samples were grown in glass beakers and for each sample 20 mg of algal biomass isolated from enrichment culture was added to 100 mL of media (i.e. BG11). As the normal pH of BG11 medium was 7, the pH was altered using 0.1 % NaOH and 0.1% HCl. The depth of the liquid column was altered by increasing the volume of the solution keeping the concentration of biomass constant. All the observations were taken at different weeks and the average temperature during the days for the period is taken into account. Each run was observed for 10 days and the final biomass was calculated on the last day. The above experiment was performed to observe the biomass production of algae at different levels of pH, nitrate concentration of medium, depth of liquid column and temperature. The biomass in the solution was observed by filtering the solution with filter paper and drying it for 24 hours at 40ºC. The weight of filter paper was compared with its initial weight (i.e. before filtration) to give the net dry cell weight of the biomass.

III. MODELING

The functional relationship between the Algal biomass yield and influencing parameters like temperature, nitrate concentration, depth of liquid column and pH was derived using two kinds of supervised data based modeling; neural network and partial least squares (PLS).

A. Development of partial least squares (PLS) based model

Use of linear PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regressions. PLS attempts to find latent variables that capture the maximum variance in the data at the same time achieve maximum correlation between predictor (X) matrix and predicted/response (Y) matrix. PLS model consists of outer relations (X & Y data individually to their respective scores T & U) and inner relations that links X data to Y data (through their scores T & U). Input - output data matrices were generated considering the experimental data. The predictor X matrix (50×4) consists of features as temperature, nitrate concentration, depth of liquid column and pH and the response Y matrix (50×1) consists of Algal biomass yield are scaled in the following way before they are processed by iterative PLS algorithm developed in the MATLAB environment.

$$X = XS_X^{-1} \text{ and } Y = YS_Y^{-1} \tag{1}$$

where $S_X = \begin{bmatrix} s_{x1} & 0 \\ 0 & s_{x2} \end{bmatrix}$ and $S_Y = \begin{bmatrix} s_{y1} & 0 \\ 0 & s_{y2} \end{bmatrix}$

The outer relationship for the input matrix and output matrix with predictor variables can be written as

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{t_1 p_1^T + t_2 p_2^T + \dots\dots\dots} \\ &\boldsymbol{\dots\dots + t_n p_n^T + E = TP^T + E} \end{aligned} \tag{2}$$

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{u_1 q_1^T + u_2 q_2^T + \dots\dots} \\ &\boldsymbol{\dots + u_n q_n^T + F = UQ^T + F} \end{aligned} \tag{3}$$

where, T and U represents the matrices of scores of X and Y while P and Q represent the loading matrices for X & Y. If all the components of X and Y are described, the errors E & F become zero. The inner model that relates X to Y is the relation between the scores T & U.

$$U = TB \tag{4}$$

where B is the regression matrix? The response **Y** can now be expressed as:

$$Y = TBQT + F \tag{5}$$

To determine the dominant direction of projection of X and Y data, the maximization of covariance within X and Y is used as a criterion. The first set of loading vectors p_1 and q_1 represent the dominant direction obtained by maximization of covariance within X & Y. Projection of X data on p_1 and Y data on q1 resulted in the first set of score vectors t1 and u1, hence, the establishment of outer relation. The matrices X and Y can now be related through their respective scores, which is called the inner model, representing a linear regression between t_1 and u_1: $\hat{u}_1 = t_1 b_1$. The calculation of first two dimensions is shown in Fig 1.

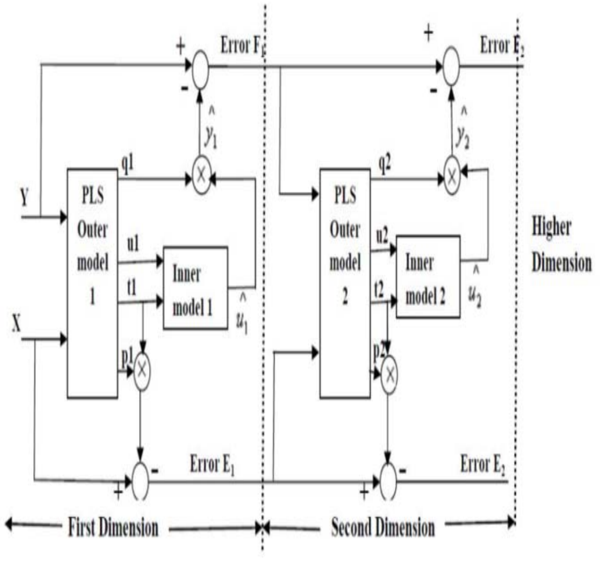

Fig. 1 Standard linear PLS algorithm.

The residuals are calculated at this stage is given by the following equations.

$$E_1 = X - t_1 p_1^{'} \tag{6}$$

$$F_1 = Y - u_1 q_1^{'} = Y - t_1 b_1 q_1^{'} \tag{7}$$

The procedure for determining the scores and loading vectors is continued by using the newly computed residuals till they are small enough or the number of PLS dimensions required are exceeded. In practice, the number of PLS dimensions is calculated by the developed iterative PLS algorithm for capturing percentage of variance explained and cross validation. The irrelevant directions originating from noise and redundancy are left as $\boldsymbol{E}$ and $\boldsymbol{F}$. The multivariate regression problems decomposed into several univariate regression problems with the application of PLS.

B. Development of Artificial Neural Network (ANN) Based Models

Neural networks, either supervised or unsupervised have emerged as an important tool in various engineering applications; especially for modeling of non-linear systems [13]. On the basis of supplied training data, the network learns the hidden relationship between the process input and output. The trained network then undergoes simulation to predict the output for unknown inputs. More than 50 types of neural network exist among them the most popular networks today are the Hopfield network, Artificial adoptive resonance (ART) theory network, Radial basis (RBF) network, Probabilistic neural network (PNN) and the most simple back propagation feed forward network [14]. In this work networks used were simple as well as robust Multiple Layer Perceptrons (MLP) for the model building. MLP is a feed forward back propagation neural network containing three layers namely input layer, hidden layer and output layer. MLP has an adaptive learning ability MLP does not make any assumptions regarding the underlying probability density functions or other probabilistic information regarding pattern classes under consideration.

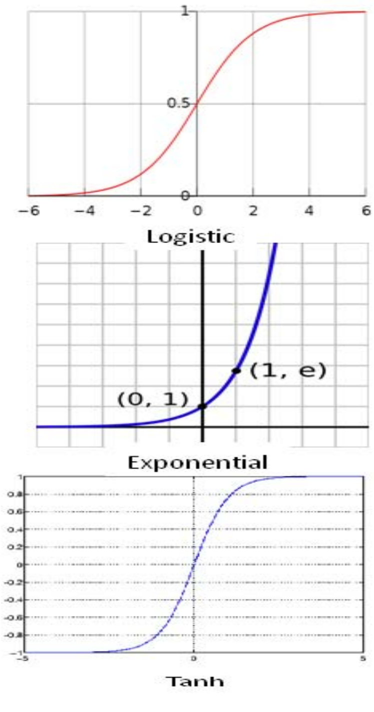

Fig. 2 Input and output activation functions in neural network.

For the model development numbers of neurons in the hidden layers were varied and best of five network configurations were considered for the model building. The Broyden– Fletcher– Goldfarb– Shanno (BFGS) method is used as the training algorithm. BFGS is a powerful second order training algorithm with very fast convergence but high memory requirements due to storing Hessian matrix and the error function included for model development was sum of squares (SOS). The transfer functions used for the hidden and output activation included Logistic sigmoid, Tanh and Exponential. The activation functions are presented in Fig 2.

The present work is an attempt to develop a neural model of the biomass yield affected by the various influencing factors considered over here.

IV. Results and Discussion

Sixty four numbers of datasets were generated using Minitab 14.0 for the Algal biomass yield using sixteen combinations of the variables like pH, nitrate concentration, and depth of liquid column at 18, 20, 22, and 24^0C temperature as presented in the Table 3.

TABLE 3 Coded Values of Input Parameters in Taguchi Design for Algal biomass Production

Runs	pH	Nitrate conc	Depth	Biomass Yield (uncoded)			
				18^0C	20^0C	22^0C	24^0C
1	1	1	1	8.949	8.937	8.925	8.913
2	1	2	2	8.949	8.821	8.809	8.797
3	1	3	3	8.717	8.705	8.693	8.681
4	1	4	4	8.601	8.589	8.577	8.565
5	2	1	1	8.946	8.935	8.923	8.911
6	2	2	2	9.064	9.053	9.041	9.029
7	2	3	3	8.714	8.703	8.691	8.679
8	2	4	4	8.832	8.821	8.809	8.797
9	3	1	1	8.944	8.932	8.920	8.908
10	3	2	2	8.828	8.816	8.804	8.792
11	3	3	3	9.180	9.168	9.156	9.144
12	3	4	4	9.064	9.052	9.040	9.028
13	4	1	1	8.941	8.930	8.918	8.906
14	4	2	2	9.059	9.048	9.036	9.024
15	4	3	3	9.177	9.166	9.154	9.142
16	4	4	4	9.295	9.284	9.272	9.260

Fig 3 shows that 100% variance of the data was captured by 3 principal components in the PLS regression. The mean square error (MSE) calculated for the PLS regression method was found to be 10-6. The experimental Algal biomass yields were in excellent agreement with the PLS predictions, which is verified well by the residual plot (Fig 3). The PLS regression coefficient matrix (β) for input variables is as follows:

$$\beta = [7.6800 \quad 0.2290 \quad 0.0201 \quad -0.0234]\ T. \qquad (8)$$

where, $\beta 11$= coefficient for variable pH

$\beta 12$= coefficient for variable nitrate concentration

$\beta 13$= coefficient for variable depth of liquid column

$\beta 14$= coefficient for variable temperature

The best five network architectures along with the training algorithm used, training, testing and validation performances and activation functions used are presented in the Table 4(a, b). The neural model was developed using Statistica 9.0; the learning rate was taken a low value of 0.1. Learning rates are used to adjust weights; a higher learning rate may be used for the quicker convergence but it may also lead to exhibit greater instability. The residual errors in PLS prediction are depicted in Fig 4. Fig 5 presents the experimental versus the neural network predictions of Algal biomass yield. The proposed networks predicted exceptionally well as evidenced in Fig 5. The residual errors in neural prediction are depicted in Fig 6. The residuals in the neural prediction are mostly within an error range of -0.01 to 0.01; establishing the efficiency of the developed neural model.

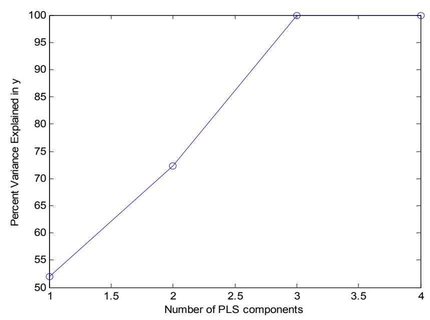

Fig. 3 Variation of percentage variance explained in response matrix with number of principal component chosen in predictor matrix

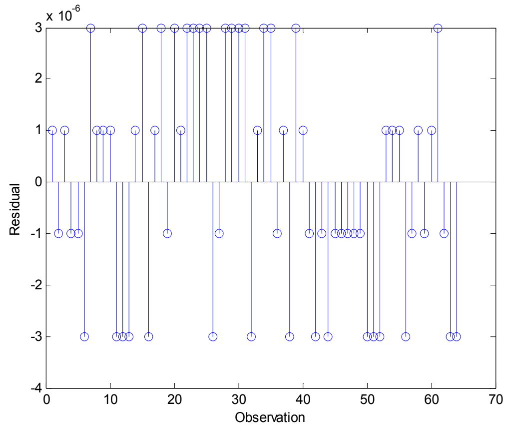

Fig. 4 Residual error in PLS prediction

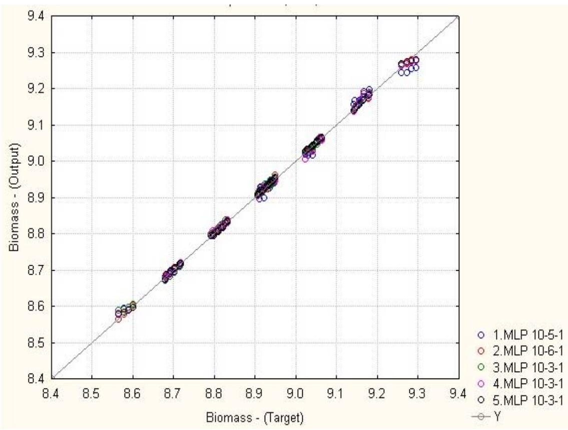

Fig. 5 Performance of Neural Network

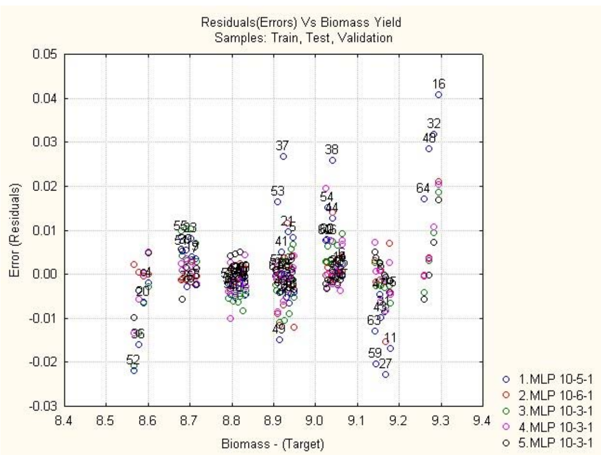

Fig. 6 Residual error for Neural prediction

TABLE 4(A) NETWORK PERFORMANCE OF THE DEVELOPED NEURAL MODEL PREDICTING ALGAL BIOMASS YIELD

Net. name	Training performance.	Test performance.	Validation performance.	Mean Squared Error
MLP 10-5-1	0.999	0.997	0.990	0.00019
MLP 10-6-1	0.999	0.999	0.994	0.00000
MLP 10-3-1	0.999	0.999	0.998	0.00020
MLP 10-3-1	0.999	0.999	0.996	0.00000
MLP 10-3-1	0.999	0.999	0.999	0.00025

TABLE 4(B) NETWORK ARCHITECTURE OF THE DEVELOPED NEURAL MODEL PREDICTING ALGAL BIOMASS YIELD.

Net. name	Training algorithm	Error Function	Hidden layer activation function	Output activation function
MLP 10-5-1	BFGS 56	SOS	Logistic	Logistic
MLP 10-6-1	BFGS 94	SOS	Exponential	Tanh
MLP 10-3-1	BFGS 42	SOS	Exponential	Logistic
MLP 10-3-1	BFGS 168	SOS	Exponential	Logistic
MLP 10-3-1	BFGS 113	SOS	Logistic	Logistic

V. CONCLUSIONS

Algal biomass was produced for 64 different combinations of parameters like pH, nitrate concentration, depth of liquid column and temperature. The Taguchi method of design of experiments was used to create 64 combinations of the aforesaid input parameters for performing the experiments; hence creating the design database.The yield of Algal biomass as a function of pH, nitrate concentration, depth of liquid column and temperature has been modeled by PLS and neural networks. PLS and feed forward back propagation Artificial Neural Network could predict the biomass yield with excellent agreement. But PLS technique was found to have lower residual errors in prediction thus considered as better technique for algal biomass modeling.

REFERENCES

[1] A.B.M Sharif Hossain and A. Salleh, "Biodiesel Fuel Production from Algae as Renewable Energy". America Journal of Biochemistry and Biotechnology. 2008, 4(3): pp 250-254.

[2] Y. Chisti. "Biodiesel from microalgae". Biotechnology advances. 2007, 25(3): pp 294-306.

[3] P. Spolaore, C. Joannis-Cassan, E. Duran and A. Isambert. "Commercial applications of microalgae". Journal of bioscience and bioengineering. 2006, 101(2): pp 87-96.

[4] P. Metzger and C. Largeau. Botryococcus braunii: a rich source for hydrocarbons and related ether lipids. Applied microbiology and biotechnology. 2005, 66(5): pp 486-96.

[5] L. Xu, P.J. Weathers, X-R. Xiong and C-Z. Liu. "Microalgal bioreactors: Challenges and opportunities". Engineering in Life Sciences. 2009, 9(3): pp 178-189.

[6] E. Molina Grima. Microalgae, mass culture methods. In: M.C. Flickinger, S.W. Drew, editors. Encyclopedia of bioprocess technology: fermentation, biocatalysis and bioseparation, vol. 3.Wiley; 1999; p. 1753–69.

[7] H. De La Hoz Siegler, A. Ben-Zvi and R.E. Burrell, McCaffery WC. "The Dynamics of Heterotrophic Algal Culture". Bioresour Technol. 2011; (in press): doi:10.1016/j.biortech.2011.01.081.

[8] S. Baquerisse, S.S. Nouals, A. Isambert, P.F. Santos and G. Durand. "Modelling of a continuous pilot photobioreactor for microalgae production". Journal of. Biotechnology. 1999, 70: pp 335-342.

[9] S.A. Mirón. "Shear stress tolerance and biochemical characterization of Phaeodactylum tricornutum in quasi steady-state continuous culture in outdoor photobioreactors". Biochemical Engineering Journal. 2003, 16(3): pp 287-297.

[10] R. Harun, M. Singh, G.M. Forde and M.K. Danquah. "Bioprocess engineering of microalgae to produce a variety of consumer products". Renewable and Sustainable Energy Reviews. 2010, 14: pp 1037-1047

[11] G. E. P. Box, W. G. Hunter, J. S. Hunter. Statistics for experimenters: an introduction to design, data analysis and model building. John Wiley and Sons, New York, 1978.

[12] S. Chinnasamy, A. Bhatnagar, R.W. Hunt and K.C. Das. "Microalgae cultivation in a wastewater dominated by carpet mill effluents for

biofuel applications". Bioresource technology. 2010, 101(9): pp 3097-105.
[13] D. Das and S.J. Dandapat. "Artificial Neural Network Modelling for the Study of pH on the Fungal Treatment of Red Mud". Research Journal of Chemical Sciences. 2011, 1(2): pp 107-111.
[14] D. Shathi, G. Sahoo and N. Saravanan. "Designing an Artificial Neural Network Model for the Prediction of Thrombo- Embolic Stroke". International Journals of Biometric and Bioinformatics. 2009, pp 3: 10-18

Madhusree Kundu was born in Kolkata, 21 January 1965. She did her Bachelors in Chemistry and Chemical Engineering from university of Calcutta, India. Dr. Kundu got her M.Tech. in Chemical Engineering from the same university in the year 1992. She completed her Doctoral study from Indian Institute of Technology, Kaharagpur, India and was awarded Ph.D. in the year 2005.

She served as a Process Engineer in the Simon carves India Ltd. during 1993-1998. She served as a lecturer and then Assistant Professor in The Birla Institute of Technology & Science, Pilani, India during 2004-2006. Presently, she is the Associate Professor, Department of Chemical Engineering, NIT, Rourkela, Orissa, India. She has published a couple of research articles and book chapters to her credit. Fluid Phase Equilibria, Advanced Process Control, Fault Detection and Diagnosis, & Process Monitoring are the current research interests of her.

Dr. Kundu is the life member of Indian Institute of Chemical Engineers (IIChE).

Diamond Das, M.Tech Student, Department of Chemical Engineering, NIT, Rourkela, Orissa, India. e-mail: diamond.das@gmail.com. He got his B.Tech Degree from IASE University, Sardarshar, India.