A Residual Analysis for the Removal of Biological Oxygen Demand through Rotating Biological Contactor

Muhmammad Rafique Daudpoto[1a], Mir Ghulam Haider Talpur[1b], Feroz Shah[2], Aijaz Khooharo[3]

ABSTRACT

Regression is a statistical method that is generally used for forecasting and prediction. It helps us to estimate the relationship between a dependent variable and one or more independent variables. This is the most widely used technique that best approximates the individual data points. It has found numerous successful applications in Engineering, Science, business and other fields. Getting average removal % of Biological Oxygen Demand (BOD$_5$) from greywater through Rotating Biological Contactor (RBC), following experiment was conducted in Sindh University hostels using different parameters such as Hydraulic Retention Time (HRT) *i.e.* 2 hours (0.42 liter per min), 2.5 hours (0.33 l/min) and 3 hours (0.28 l/min) and multiple number of discs *i.e.* 40, 42, 44, 46, 48, 50 and 52. Consequences reveal that linear estimate of HRTs and numbers of disc are considerable whereas linear and quadratic estimates of number of discs are highly significant, which evidence the significance of time and discs. However, as p-value is greater than 0.05, hence quadratic estimate of HRT is not significant. By using coefficients of the table the regression equation is Removal = - 79.995 + 6.88 time + 2.90 disc, where the sample standard deviation is 7.151, coefficient of correlation is 0.86 and coefficient of determination is 0.742. Distributions of errors are approximately normal as probability plot of the residuals is approximately linear. Residual analysis shows that against each predicted variable, residuals plot falls approximately in a horizontal band symmetric and centered about the horizontal axis and against predicted y-values. Moreover, Residual plot shows the constant standard deviations and linearity assumptions appear to be met.

Keywords: Regression, Response Surface Modeling, Residual Analysis, Biological Oxygen Demand, Hydraulic Retention Time

1. INTRODUCTION

Resources of fresh water are under immense stress and the need of fresh water can never be denied at all especially in the populous countries. Falkenmark and Lindh [1] state that a state or country is said to be water stressed when water provisions drop below 1000 cubic meters per person per year. Gleick [2] mentions that throughout last 20 years, in all regions of the world, shortages are observed due to decline of available water in per capita. The swift urbanization, industrialization, climatic change and ever increasing population spread have significantly threatened the assets of fresh water which are depleting day by day as stated by Nghiem *et al.* [3]. In urbanized countries, the most important consideration for treated wastewater reuse is implementation and planning of water assets. According to Ali *et al.* [4] reuse of treated effluents is being considered as reliable source of water in recent

[1] Department of Statistics, University of Sindh, Jamshoro, Sindh, Pakistan.
  Email: [a]daupoto.rafique@usindh.edu.pk (Corresponding Author), [b]ghtalpur@yahoo.com
[2] Department of Basic Sciences and Related Studies, Mehran University of Engineering and Technology, Jamshoro, Pakistan. Email: feroz.shah@faculty.muet.edu.pk
[3] Sindh Agriculture University, Tando Jam, Sindh, Pakistan. Email: khooharo@gmail.com

years. In this regard a number of countries like Saudi Arabia, Singapore and Jordan have made significant progress in reclaiming the treated effluents and this has been made a part of their national policies.

RBC is an attached growth aerobic treatment method which requires the occurrence of molecular oxygen for the metabolic movement of microorganisms. In this process by the standard of physiochemical adsorption removal of colloidal particles were achieved first and secondly by the embarrassing situation of balanced particulate matters on the biological flocs respectively [5]. On the other hand, the removal of soluble organic fractions like BOD_5 was done by microbial bio combination. The RBC works on the standard of bio amalgamation. These pollutants are transformed by microbes in to simpler end products *i.e.* water and carbon dioxide and create them possess cells.

Cohen *et al.* [6] state that Regression analysis is a statistical procedure for estimating the association among various variables. It comprises of techniques for analysis and modeling of many variables when the center of attention lies in cooperation between dependent variable and one or more independent variables. Doudpoto *et al.* [7] states that Regression analysis, helps us to determine how a variable changes its value by altering one or more independent variables. The change may be linear or nonlinear. Hayes [8] concludes that regression analysis estimates the conditional expectation of the dependent variable given the independent variables – that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. The goal is to estimate, in all cases, the function of the independent variables, called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution. The regression analysis predicts a value a variable y, given known values if the variable x. Jamal [9] categorize the regression analysis as single regression analysis and multiple regression analysis. Single regression analysis of independent variable is called simple regression and when there are two or more independent variables involved in the analysis, it is called a multiple regression analysis. Li-Kun [10] affirm that the coefficient of correlation is generally used to compute the strength of the relationship between the two variables. The value of the correlation coefficient range between -1 and 1.

2. NORMALITY ASSUMPTIONS

To decide whether it is reasonable to assume that normality assumptions are met or not, the so called residual analysis is carried-out, which is based on the differences between the observed and predicted values. For multiple regressions, inferences are met by three conditions. Weiss [11] suggests that against each predicted variable, a plot of residuals should fall approximately in a horizontal band symmetric and centered about the horizontal axis, against predicted y-values, and the normal probability plot of the residuals should be approximately linear. Montgomery [12] notes down that those violations of model adequacy and basic assumptions can be investigated without any difficulty by the examinations of Residuals. Doubt casts on the validity of one or more of the assumptions, if failure of any of these three conditions for multiple regression inferences [13]. Histogram of residuals can also be made to check normality assumptions. The plot looks like a sample from a normal distribution centered at zero, if NID $(0, \sigma^2)$ assumptions on errors are satisfied. Fluctuations often occur in the shape of histogram with small samples, unfortunately, but the moderate departure from normality does not imply necessarily a serious violation of the assumptions.

3. MATERIALS AND METHODS

Primary information regarding quantities of BOD_5 and related information of grey water were collected from Sindh University Hostel after processed into RBC so that we can collect different quantities of pollutant levels. Regression Analysis is applied on RBC which is actually used in Simulation for treatment of wastewater, when removal is subjective by one or more than one factors. Two factors were chosen for the present study, to fit the quadratic response i.e. Number of Discs and HRT to analyze the effect. Different procedures were applied in order to optimize the removal response. The system operated under three different positions of HRTs *i.e.* 2 hours (0.42 liter per

Mehran University Research Journal of Engineering and Technology, Vol. 40, No. 2, April 2021 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

460

min), 2.5 hours (0.33 l/min) and 3 hours (0.28 l/min) and multiple numbers of discs i.e. 40, 42, 44, 46, 48, 50 and 52 for greywater treatment through RBC. Statistical software for Social Sciences (SPSS: 20), Minitab: 17, Ms Excel and Origin pro version 7 is used for analysis and to draw residual plots and 3D plots. The developed statistical model was validated with the previous experimental work by the researchers.

A simple linear regression model may be written as

$$y = b_o + b_1 x$$

The number b_o is called the y-intercept and the number b_1 is called the slope of the line. Weiss [11] noted that the values of b_o and b_1 (in straight line equation) can be symbolize by a and b.

Chaudhry and Kamal [14] propose that for multiple regressions with two predictor variables x_1 and x_2, the regression equation is in the form:

$$y = b_o + b_1 x_1 + b_2 x_2$$

And in general; for multiple regression with k predictor variables $x_1, x_2, ..., x_k$ the regression equation is in the form:

$$y = b_o + b_1 x_1 + b_2 x_2 + ... + b_k x_k$$

where b_o is the unbiased estimate of the regression intercept and b_1, b_2, and b_k are the parameters of regression of equation [15].
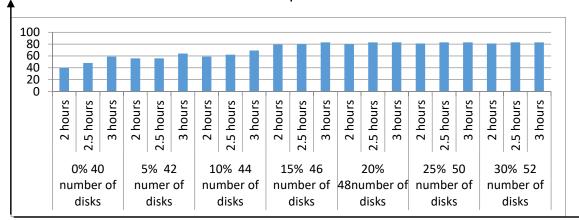

Fig. 1: Average removal of BOD$_5$ on different HRTs levels i-e; 2 hours, 2.5 hours and 3 hours with multiple numbers of discs
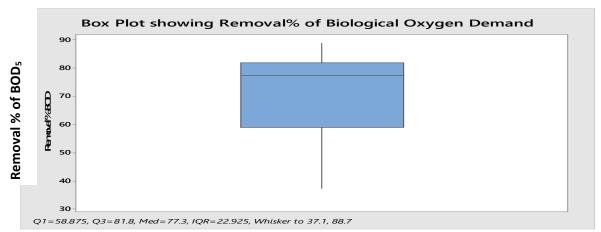

Fig. 2: Whisker Box Plot showing Removal% of BOD$_5$, on 2, 2.5 and 3 Hours HRTs with different descriptive measures
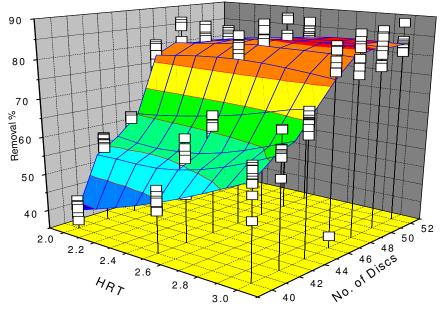
Mehran University Research Journal of Engineering and Technology, Vol. 40, No. 2, April 2021 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

461

Fig. 3: Response Surface effect of Retention Time and Number of Discs on Removal of BOD_5.

Average removal % of BOD_5 in Fig.1 shows different HRTs levels *i.e.* 2 hours, 2.5 hours and 3 hours, multiple numbers of discs and with corresponding area increment. It is observed that as the HRT and number of discs increase, the removal % is increased simultaneously. Dot plot in Fig.2 represents the descriptive statistics including quartiles Q_1= 58.875 and Q_3= 81.8, Median= 77.3, Inter Quartile Range= 22.925 and whisker observations i.e. 37.1 to 88.7 of the average removal% of BOD_5. Fig.3 shows the effect of 3D Response Surface regarding HRT and Number of Discs on Removal of BOD_5.

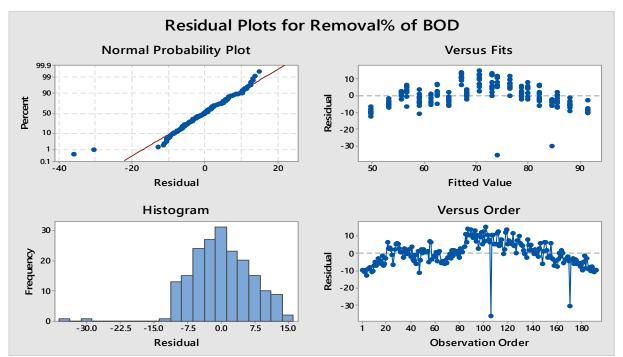
Fig. 4: Normal Probability, Fits, Histogram residual Predicted (Fitted) values of removal ofBOD_5**.**

Mehran University Research Journal of Engineering and Technology, Vol. 40, No. 2, April 2021 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

462

4. RESULTS AND DISCUSSION

Key information about the removal % amount of BOD_5 through RBC was obtaied from Sindh University Hostels by different procedures concerning the multiple numbers of discs, and the HRT.

Table 1 represents Predictor variables, coefficients as well as their consequent t and p values for Response Surface Model. Table 2 contains t and p-values showing the effect of intercept, retention time and number of discs on the removal of BOD_5.

Table 1: t and p value for response surface model				
Predictor	Coefficient	Standard Deviations	t-values	p-values (sig. level)
Constant	-807.00	70.65	-11.42	0.000
Time	53.09	20.66	2.57	0.011
Disc	32.237	2.748	11.73	0.000
$time^2$	0.731	3.451	0.21	0.832
$Disc^2$	-0.28971	0.02913	-9.95	0.0000
Timedisc	-1.0861	0.2467	-4.40	0.0000

Table 2: t and p values showing the coefficient, standard deviation, and p & t values.				
Predictor	Coefficient	Standard Deviation	t-values	P- values
Constant	-79.995	6.684	-11.97	0.000
Time	6.881	1.264	5.44	0.000
Disc	2.9008	0.1278	22.70	0.000

The regression equation is
Removal = - 79.995 + 6.88 time + 2.90 disc
S = 7.151   R-Sq = 74.2%   R-Sq(adj) = 74.0%

Table 3: F and p-values showing the effect of the removal of BOD_5					
Source	DF	SS	MS	F	P
Regression	2	27857	13928	272.37	0.000
Error	189	9665	51		
Total	191	37522			

By calculating t-values and consequent p-values the factors be validated. Literature revealed that a factor possessing p-value less than 0.05 is considered as significant, if not, factor is considered as not significant. The coefficients of the above equation in Table 1 with their consequent "t" and "p" values specified reveals that linear estimate of HRT and number of discs are significant as their p-values are less than 0.05, however p-value less than 0.01 indicates that linear and quadratic estimates of number

of discs are highly significant. Quadratic estimate of HRT is, however not significant as its p-value is greater than 0.05. Interaction effect of HRT and number of discs is also highly significant as the p-value of interactive effect is less than 0.01.

Table 2 reveals that the estimate of HRT and number of discs have highly significant effect on the removal of BOD_5, even as level of significance was 0.05, as their resulting p-values are less than 0.01. By using coefficients of the Table the regression equation of Removal = - 79.995 + 6.88 time + 2.90 disc, where the sample standard deviation is 7.151, coefficient of correlation is 0.86 and coefficient of determination is 0.742. Moreover, by applying Regression Analysis the ANOVA in Table 3 was constructed to validate the consequences obtained. It indicates that even when level of significance was 0.05, HRT and number of discs has highly significant effect on removal % of BOD_5, since the consequent p-values are less than even 0.01.

Normal Probability Plot shows that that the distributions of errors are approximately normal. Tendency of normal probability plot upward slightly on right side and to bend down slightly left side implies that the tails of error distribution are to some extent thinner than would be anticipated in a normal distribution. The normal probability plot of the residuals is approximately linear.

Residual analysis shows the differences between the observed and predicted values. Fig.4 shows that against each predicted variable, residuals plot falls approximately in a horizontal band symmetric and centered about the horizontal axis and against predicted y-values, residuals plot fall approximately in a horizontal band symmetric or centered about horizontal axis. Residual plot shows the constant standard deviations and linearity assumptions appear to be met. Histogram of residuals shows normality assumptions as the plot looks like a sample from a normal distribution centered at zero and NID $(0, \sigma^2)$ assumptions on errors are satisfied as standardized residuals are approximately normal with mean zero and unit variance.

Mehran University Research Journal of Engineering and Technology, Vol. 40, No. 2, April 2021 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

463

5. CONCLUSION

The output indicates that, retention time and number of discs has linear positive effect on the removal % of (BOD_5). The coefficients for retention time, number of discs are 53.1 and 32.2 respectively. The quadratic effect of retention time is 0.73. It indicates that as the time increases the removal of Biological Oxygen Demand is also increased. The model further reveals that number of discs has positive effect on the removal of BOD_5; however, after some level of increment of discs, it tends towards decline, as the quadratic response of number of discs is -0.290.

The distribution of errors (Histogram) is approximately normal. The normal probability plot of the residuals is approximately linear. In Normal Probability Plots and Residual Plots, it was observed that by removing the outliers models can fit best.

ACKNOWLEDGEMENTS

REFERENCES

1. Falkenmark and Lindh, (1976) quoted in UNEP/WMO. "*Climate Change 2001*: *Working Group* II: *Impacts, Adaptation and Vulnerability*," UNEP. Retrieved 3 February 2009.
2. Gleick P. H., "Water Use", *Annual Review of Environment and Resources"*, Vol. 28, No.1, pp. 275-314, 2003.
3. Nghiem D., Oschmann N., Schafer, I. "Fouling in Greywater Recycling by Direct Ultrafiltration*," Desalination,* Vol. 187, pp. 283-290, 2006.
4. Ali, H.Q., Aamir F., Ahmed M., "Monitoring the Wastewater Treatment Efficiency of Oxidation Ponds at Chokera, Faisalabad", *Mehran University Research Journal of Engineering and Technology*, Vol. 36, No. 4, pp. 987-994, October 2017.
5. Pathan A. A.; Mahar R.B.; Memon F.A., "The effect of flow rate and disc area increment on the efficiency of rotating biological contactors for treating greywater", *Mehran University Research Journal of Engineering and Technology*, V. 34, No.2, 2015, pp 99-106, 2015.
6. Cohen J., Cohen., P., West S.G., Aiken L.N., "*Multiple Regression/Correlation Analysis for the Behavioral Sciences*", Mahwah New Jersey, Lawrence Erlbaum Associates, London, 2003.
7. Daudpoto M. R., Talpur M.G.H, Shah F, Khooharo A. A., "Regression and correlation analysis on rotating biological contactor, *Sindh University Research Journal (Science Series)*, Vol. 49, No. 4, pp. 849-852, 2017.
8. Hayes A.F., "*Introduction to Mediation, Moderation and Conditional Process Analysis*: *A Regression Based Approach"*, The Guilford Press, New York, September, 2013.
9. Jamal S., "*Statistics Problems and Practice*", Ahmed Academy, Urdu Bazar Karachi, 2016.
10. Li-Kun Y., "Application of Correlation Coefficient and Biased Correlation Coefficient in Related Analysis", *Journal of Yunnan university of Finance and Economics*, Vol. 19, No. 3, pp. 78-80, 2003.
11. Weiss N.A., "*Introductory Statistics*", Pearson Education, ISBN 9780321989178, 2017.
12. Montgomery D.C.,"*Design and Analysis of Experiments*", John Wiley & Sons, ISBN: 978-1-118-14692-7, 2013.
13. Myers R.H., Montgomery D.C., Anderson-Cook C.M., "*Response Surface Methodology: Process and Product Optimization Using Designed Experiments*", Wiley , ISBN: 978-1-118-91601-8, 2016.
14. Chaudhry S. M., Kamal S., "Introduction to Statistical Theory", Part-I, Ilmi Kitab Khana, Urdu Bazaar Lahore Pakistan, 2012.
15. Groebner D. F., Shannon P. W., Fry P.C., Smith K.D., "Business Statistics, A Decision Making Approach", Pearson, 2011.

Mehran University Research Journal of Engineering and Technology, Vol. 40, No. 2, April 2021 [p-ISSN: 0254-7821, e-ISSN: 2413-7219]

464