

# Predicting COD and BOD Parameters of Greywater Using Multivariate Linear Regression

Samir Sadik Shaikh<sup>a, 1</sup>, Dr. Rekha Shahapurkar<sup>a</sup>

<sup>a</sup> Computer Science and Engineering Department, Oriental University, Indore, India

**Abstract.** Greywater reuse furthermore, reusing can be an incredible method to get non-consumable water. Since it contains broke down pollutions, greywater can't be utilized straightforwardly. As an outcome, it is critical to decide the nature of water prior to utilizing it. Body estimations require five days to finish, while COD estimations require only a couple of hours. Not exclusively improve models for evaluating water quality are required; however, a more coordinated methodology is additionally getting more normal. Most of these models require a wide scope of information that isn't in every case promptly available, making it a costly and tedious activity. Because of different issues in the enlistment with estimation included in water quality boundaries like BOD as well as COD, the principal objective of this investigation is to track down the best multivariate direct relapse models for foreseeing complex water quality outcomes. The code was written in Python for multi-variable information sources, and a Linear Regression Model was created. The projected COD versus estimated COD chart shows that the noticed and expected qualities are practically the same. The R-squared worth was 0.9973. A plot of extended BOD as an element of COD is likewise remembered for the outcome.

**Keywords.** ANN, BOD, COD, Greywater, Multivariate Linear Regression.

## 1. Introduction

Alternative water management strategies have been set up in dry territories because of the absence of fresh water. Almost 97% of the world's absolute water supply is found in the seas, yet only 3% of it appropriate for direct use [1]. Greywater is squandered water that is generally made by kitchen sinks, showers, clothing or clothes washers, cooling outlets, and other comparable gadgets. As indicated by information, greywater age fluctuates somewhere in the range of 39 to 85 percent in various nations [2] Greywater treatment and reuse will incorporate non-consumable water for latrine flushing, cultivating, vehicle cleaning, and floor washing, in addition to other things. Table 1 shows characteristics of greywater [3, 4].

---

<sup>1</sup> Samir Sadik Shaikh, Computer Science and Engineering Department, Oriental University, Indore  
Email: samirshaikh81@gmail.com.

**Table 1.** Characteristics of Greywater

Parameters	Units	Values
pH	---	7.3 - 8.1
EC	$\mu\text{S cm}^{-1}$	489 - 550
Turbidity	NTU	20.6 - 38.7
Total Suspended Solids (TSS)	$\text{mg L}^{-1}$	12 - 17.6
Nitrate ( $\text{NO}_3^-$ )	$\text{mg L}^{-1}$	0.5 - 0.63
Total Nitrogen (TN)	$\text{mg L}^{-1}$	42.8 - 57.7
Phosphate ( $\text{PO}_4^{3-}$ )	$\text{mg L}^{-1}$	1.52 - 3.36
BOD	$\text{mg L}^{-1}$	56 - 100
COD	$\text{mg L}^{-1}$	244 - 284
Total Caliform (TC)	CFU/100 mL	$3.74 \times 10^4$ to $3.8 \times 10^4$
Na	$\text{mg L}^{-1}$	43.8 - 48.1
K	$\text{mg L}^{-1}$	8.3 - 15.2
B	$\text{mg L}^{-1}$	1.3 - 1.5
$\text{Cl}^-$	$\text{mg L}^{-1}$	7.4 - 12.9

The Clean Water Act was sanctioned in the mid-1970s, trailed by the making of the USEPA, which finished in the characterization of wastewater quality for natural benefit dependent on four principle rules [5]:

- **Physical Properties:** e.g. pH, turbidity, temperature, colour, and odour.
- **Solids:** e.g., Total Solids (TS), Total Suspended Solids (TSS), Total Dissolved Solids (TDS), Total Volatile Solids (TVS), and Total Fixed Solids (TFS).
- **Organics:** e.g., Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Total Organic Carbon (TOC), and Oil and Grease (O&G).
- **Nutrients:** e.g., TN (Total nitrogen) and TP (Total phosphorous).

Various water quality lists have been utilized in numerous ordinary investigations comparable to different water sources like lakes, waterways, and dam supplies [6-8]. The Trophic State Index (CTSI), set up via Carlson in 1977 [9], is ordinarily utilized by water the board offices and associations throughout the planet. The CTSI is a typical and valuable water quality record that has been utilized as the essential measurement in numerous examinations [10-11]. CTSI is determined utilizing three separate water quality variables: chlorophyll-a (Chl-a) fixation, total phosphorus (TP) focus, and Secchi depth (SD). Substance tests, estimations, and careful estimations of water tests are utilized to decide the centralizations of chlorophyll-an and complete phosphorus. Secchi depth, then again, can be physically estimated in repositories without the utilization of compound examinations or present-day innovation, however, it is likely the most unpredictable boundary because of its reliance on temperature (counting season, turbidity, and different variables) [12].

Because of the trouble of ascertaining chlorophyll-an and complete phosphorus fixations, various water quality records are used, together with turbidity, electrical conductivity, natural oxygen request focus or biochemical oxygen demand focus (BOD), chemical oxygen demand fixation (COD), and all-out total dissolved solids (TDS). A considerable lot of these are the most normally utilized boundaries for surveying water quality in Artificial Intelligence (AI) methods [13-17]. Chang and Liu (2015) suggested a fluffy back spread neural organization model to order the level of eutrophication because of the shakiness of trophic status dictated by TDS, BOD, and COD due to temperamental turbidity. Not exclusively improve models for evaluating water quality are required; however, a more methodical methodology is additionally getting more normal. Lately, assortments of AI-based approaches have been used towards address water quality issues; also AI holds a ton of guarantees around here (Chau, 2006). The utilization of AI to acquire useful connections among information dependent on chronicled info and yield information is at the core of AI. Fake neural organizations, choice trees (DTs), straight relapse, and the assistance vector machine are the most regularly utilized information-digging calculations for this reason in related works. In ANN-based applications [18-21] and SVM-based applications [22-25], some of them are utilized independently for the forecast.

In a few settings, the practical connection between covariates (otherwise called input factors) and reaction factors (otherwise called yield factors) is of extraordinary interest. When demonstrating complex sicknesses, for instance, potential danger factors and their impacts on the infection are explored to decide hazard factors that can be utilized to improve preventive or mediation techniques. Fake neural organizations can estimate any complex useful relationship. Rather than summed up straight models [26], it isn't critical to characterize the type of connection among covariates and reaction factors ahead of time. Thus, fake neural organizations are a compelling factual instrument. They are GLMs' immediate augmentations, and they can be utilized similarly. The neural organization is prepared utilizing noticed information, and it iteratively adjusts its boundaries to gain proficiency with a guess of the relationship [27].

In any interaction industry, execution lists like biochemical oxygen interest (BOD) and compound oxygen interest (COD) are used to decide the nature of wastewater created. Body plus COD are characteristic boundaries in place of sewer water quality. The body stays an expected pointer for the measure included in biochemically degradable natural matter found in a water test aimed at homegrown wastewater. COD estimations should be possible surprisingly fast versus five days for BOD estimations, regardless of the way that COD qualities are consistently higher than BOD esteems. The at present accessible technique for figuring BOD and COD is tedious and defenseless against estimation blunders. To deal with the accepted procedures for water quality protection, a few water quality models, like ordinary unthinking methodologies, have been made. Most of these models require a wide scope of information that isn't in every case promptly available, making it a costly and tedious activity [28]. Lately, the Artificial Neural Network (ANN) procedure has acquired in prevalence. Dogan et al. [29] investigated the capacity about the ANN model on the way to increase the exactness of natural oxygen request assessment (BOD). By contrasting the discoveries

with noticed BOD levels, the limit of an ANN technique in BOD assessment in the Melen River was investigated in this report. Utilizing the ANN strategy with COD, water release, suspended strong, complete nitrogen, and all-out phosphorus, MSE of 708.01, normal supreme relative mistakes of 10.03 percent, and a coefficient of assurance of 0.919 were gotten. Rene and Saidutta [30] utilized ANNs to assess BOD and COD fixations dependent on quantifiable water quality lists. The ANN's capacity to anticipate BOD was better than COD, as per their outcomes. Akratos et al. [31] utilized an ANN model and plan conditions to foresee BOD and COD evacuation in even subsurface stream planned wetlands. The discoveries of the ANNs and the model plan condition were fundamentally the same as test proof from the writing. The outcomes showed that utilizing the ANN cycle, a reasonable connection could be gotten. COD evacuation was found to be unequivocally connected with BOD expulsion. What's more, a COD evacuation expectation condition was created.

Due to various issues in the registration and measurement of water quality such as BOD and COD, the main goal of this study is to find the best multivariate linear regression models for predicting complex water quality results.

## 2. Material and Method

### 2.1. Case Study

Throughout the span of 11 months, the informational index for this examination was gathered through the kitchen sink (May 2020-and in the long run picked for model development dependent on estimated estimations of various factors and their correlative investigation. The body is estimated by hatching a fixed water test for five days and ascertaining the oxygen misfortune March 2021). The examples were assembled and shipped off Vashi's Water Quality Testing Lab, and a Cumulative Report of Water Quality was acquired (see Figure 1). Components (factors) like pH, complete suspended strong (TSS), absolute suspended (TS), and water temperature (T) that influence water quality (BOD and COD) were distinguished from start to finish. In the event that examples are not weakened until hatching, microscopic organisms will drain the entirety of the oxygen in the jug before the test is finished. The test outcomes were determined utilizing the Standard Procedures of the American Public Health Association [32].

From the cumulative water quality report, values of std. deviation and deviation coefficients were calculated, as represented in Table 2. The value of SDx and CV are calculated as:

$$SDx = \sqrt{\frac{\sum_{i=1}^n (Xmean - Xi)^2}{n}} \quad (1)$$

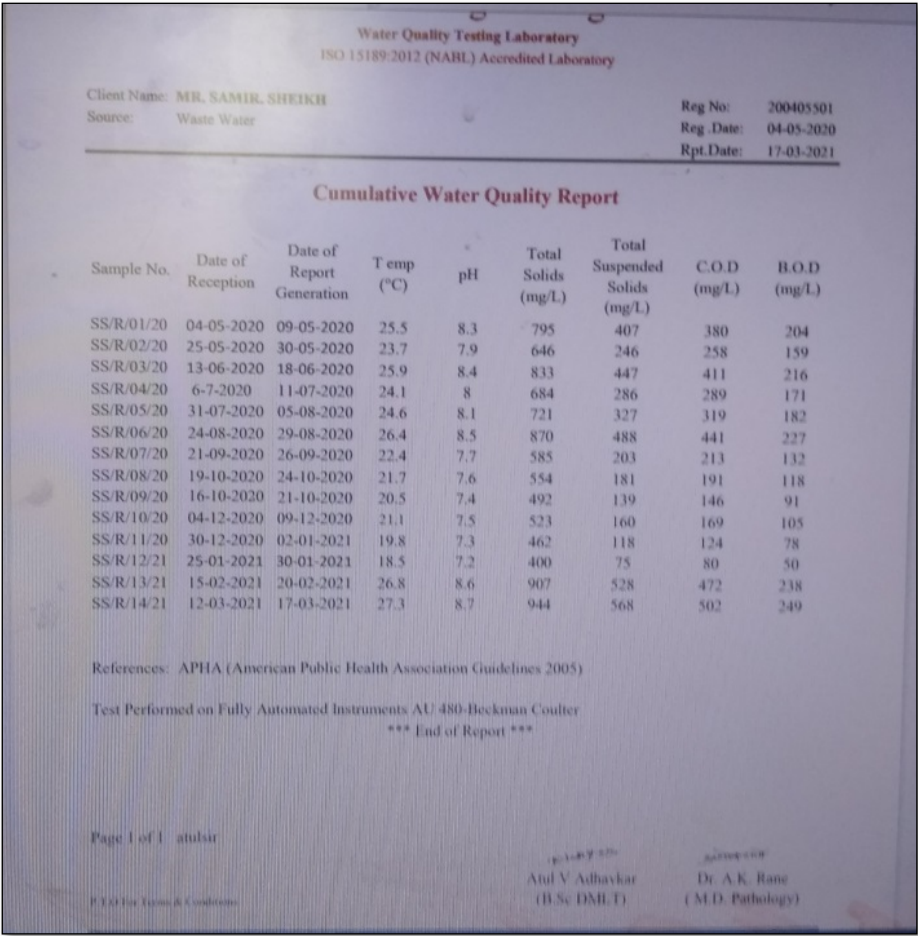


Figure 1. Cumulative Water Quality Report.

$$CV = \frac{SDx}{xmean}$$

(2)

Table 3 shows the model domain boundary of the water quality parameter. Xmean, Xmax, Xmin, SDx, and CV denote the data set's mean, maximum, minimum, standard deviation, and deviation coefficient, respectively (derived from cumulative report and Table 1). Table 2 shows that the CV value for pH (0.06) is the lowest and it is highest for TSS (0.53).

**Table 2.** Calculation of Standard deviation (SDx) and Deviation coefficient (CV)

Sample No.	Temp (°C)	pH	Total Solids (mg/L)	Total Suspended Solids (mg/L)	C.O.D (mg/L)	B.O.D (mg/L)
SS/R/01/20	4	0.16	14981.76	11859.21	8949.16	2061.16
SS/R/02/20	0.04	0	707.56	2714.41	750.76	0.16
SS/R/03/20	5.76	0.25	25728.16	22171.21	15775.36	3294.76
SS/R/04/20	0.36	0.01	129.96	146.41	12.96	153.76
SS/R/05/20	1.21	0.04	2342.56	835.21	1128.96	547.56
SS/R/06/20	8.41	0.36	38966.76	36062.01	24211.36	4678.56
SS/R/07/20	1.21	0.04	7673.76	9044.01	5241.76	707.56
SS/R/08/20	3.24	0.09	14065.96	13712.41	8911.36	1648.36
SS/R/09/20	9	0.25	32616.36	25312.81	19432.36	4569.76
SS/R/10/20	5.76	0.16	22380.16	19071.61	13548.96	2872.96
SS/R/11/20	13.69	0.36	44352.36	32436.01	26049.96	6496.36
SS/R/12/21	25	0.49	74310.76	49773.61	42189.16	11793.96
SS/R/13/21	10.89	0.49	54943.36	52854.01	34819.56	6304.36
SS/R/14/21	14.44	0.64	73657.96	72846.01	46915.56	8172.16
<b>SDx</b>	<b>2.712537</b>	<b>0.488438</b>	<b>170.4736</b>	<b>157.8514</b>	<b>133.0782</b>	<b>61.70288</b>
X <sub>mean</sub>	23.5	7.9	672.6	298.1	285.4	158.6
<b>CV</b>	<b>0.115427</b>	<b>0.061828</b>	<b>0.253455</b>	<b>0.529525</b>	<b>0.466287</b>	<b>0.389047</b>

**Table 3.** Water Quality Properties

Parameters	Unit	Xmin	Xmax	Xmean	SDx	CV
T	°C	18.5	27.3	23.5	2.71	0.1
pH	---	7.2	8.7	7.9	0.48	0.06
TS	mg/L	400	944	672.6	170.47	0.25
TSS	mg/L	75	568	298.1	157.85	0.53
COD	mg/L	80	502	285.4	133.1	0.47
BOD	mg/L	50	249	158.6	61.7	0.39

**3. Multivariate Linear Regression**

In a forecast issue, straight relapse (LR) is a relapse model that was intended to decide the connection among autonomous and subordinate factors [33]. This investigation utilizes multivariate straight relapse, which is one of the numerous types of direct relapse. The most essential strategy for setting up a connection between autonomous factors (noticed or estimated), otherwise called indicators or regressors, and a reliant variable, otherwise called the reaction variable, is various direct relapse (MLR). A summed up articulation for the model can be composed as follows:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + C \tag{3}$$

Where Y is the dependent variable,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ , and  $\beta_4$  are the coefficients of  $X_1$ ,  $X_2$ ,  $X_3$ , and  $X_4$  respectively, and C is the block. The direct relapse strategy is like the condition of a straight line, given by  $Y = ax + b$ .

Statistical methods, for example, regression models, stand as the most effective tools aimed at examining every relationship amongst dependent and independent variables in lesser samples [21]. The MLR is a process for modeling the linear relationship among one or more independent variables and a dependent variable. MLR is based on least squares. In the best model, the sum of square error between observed and predicted parameters should be a minimum value. BOD and COD estimation also can be performed using linear models which explain the linear relationship between parameters. MLR is based on the principle of least squares. The sum of square errors between observed and predicted parameters should be as low as possible in the best model. Linear models that describe linear relationships between parameters can also be used to estimate BOD and COD. In addition, as shown in equation 4, the same input variables used in MLR models can also be used in linear models.

$$Y = \beta_1 T + \beta_2 pH + \beta_3 TS + \beta_4 TSS + e \quad (4)$$

Where,  $Y$  represents COD values,  $\beta_1, \beta_2, \beta_3, \beta_4$  as well as  $e$  are constant coefficients coming from the linear regression model,  $T, pH, TS$  also  $TSS$  are input factors which will determine the predicted value of COD for our model. Also, we will estimate the values of BOD using COD.

#### 4. Results and Discussions

Google gives an online Google Colaboratory that can be utilized to compose and execute AI calculations in Python utilizing the online code supervisor. Along these lines, there is no compelling reason to introduce the libraries of python on a work area or PC. For the execution of the examination, we have utilized Google colab where the code was written in python and the Linear Regression Model was made for multivariable sources of info ( $T, pH, TS, TSS$ ) by bringing in linear\_model utilizing sklearn. OLS (Ordinary Least Square) technique was utilized to create the aftereffects of direct relapse as demonstrated in Figure 2.

Thus from the report, it can be seen that the values of intercept (marked as constant) and coefficients of  $T, pH, TS$ , and  $TSS$  (marked as  $x1, x2, x3$ , and  $x4$ ) can be used to predict the values of COD for the described model. Substituting these values in equation (4) yields the following;

$$Y = -3.1004*T + 7.3189*pH + 0.6417*TS + 0.1808*TSS - 185.5697 \quad (5)$$

Along these lines, if a model is portrayed by equation (5) it will give us a best-fit model. The condition was executed in dominant utilizing  $T, pH, TS, TSS$  as info factors to anticipate COD as yield, characterized by equation (5). A portion of the outcomes that appeared in figure 3 are acquired by plotting single free factor ( $T, pH, TS$ , and  $TSS$ ) against the anticipated estimations of COD. It is done so in light of the fact that a straight fit can be best seen in situations where we have single autonomous and ward factors. In this manner, the element of perception will be a 2D plane (which is administered by the connection  $p+1$  where  $p$  is the quantity of autonomous factors). As the quantity of autonomous variable expands, the element of noticed plane expansions in the same extent ( $p+1$ ), consequently fitting model on a straight line gets

unpredictable. Such models are acknowledged utilizing the dissipate plots as demonstrated in figure 4.

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	1.000			
Model:	OLS	Adj. R-squared:	1.000			
Method:	Least Squares	F-statistic:	6.432e+05			
Date:	Sun, 21 Mar 2021	Prob (F-statistic):	1.54e-24			
Time:	06:16:43	Log-Likelihood:	-0.39479			
No. Observations:	14	AIC:	10.79			
Df Residuals:	9	BIC:	13.98			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	-185.5697	28.918	-6.417	0.000	-250.986	-120.153
x1	-3.1004	2.323	-1.334	0.215	-8.356	2.155
x2	7.3189	4.296	1.704	0.123	-2.399	17.037
x3	0.6417	0.088	7.299	0.000	0.443	0.841
x4	0.1808	0.054	3.376	0.008	0.060	0.302
-----						
Omnibus:	0.394	Durbin-Watson:	1.619			
Prob(Omnibus):	0.821	Jarque-Bera (JB):	0.034			
Skew:	0.106	Prob(JB):	0.983			
Kurtosis:	2.889	Cond. No.	2.70e+05			
=====						

Figure 2. OLS Regression Report.

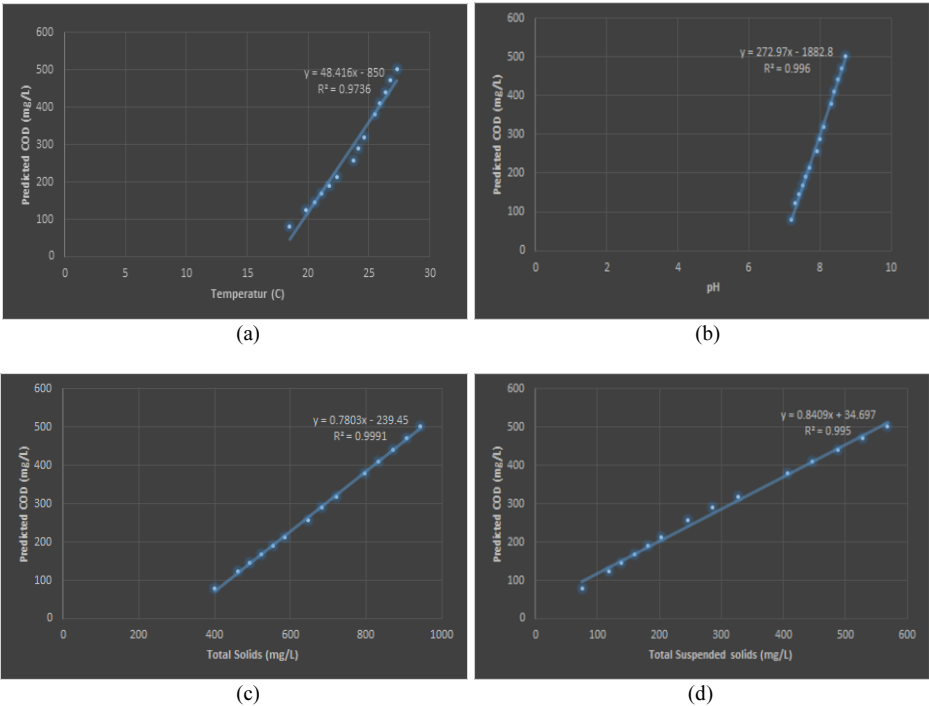
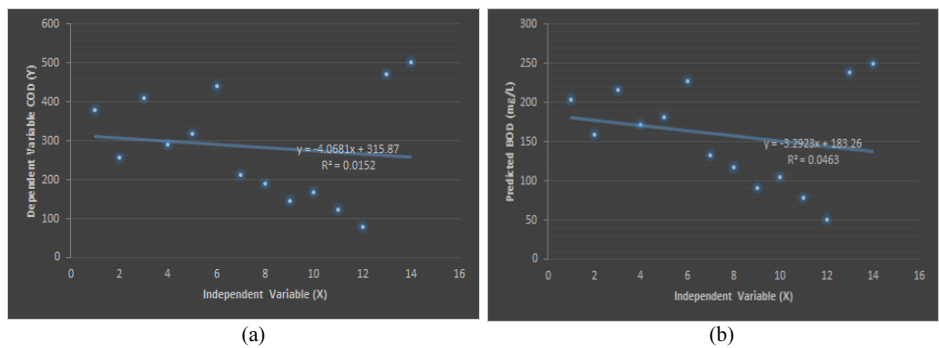


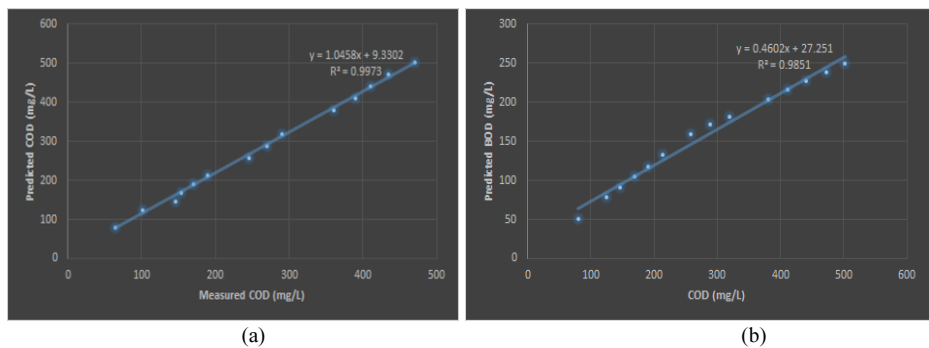
Figure 3 (a) Predicted COD v/s Temperature (b) Predicted COD v/s pH (c) Predicted COD v/s TS (d) Predicted COD v/s TSS





**Figure 4 (a)** Predicted COD v/s T, pH, TS, TSS **(b)** Predicted BOD v/s T, pH, TS, TSS

The main objective of a linear regression model is to estimate the difference between the predicted and observed (measured) value of the variable with the intention of validates the usefulness of the model. A response of predicted COD v/s measured COD is represented in figure 5 (a). Chemical investigation for COD measurement takes few hours, while BOD measurement takes 05 days; therefore it is also possible to predict BOD using values of COD Figure 5(b) represents Predicted BOD as a function of COD.



**Figure 5 (a)** Predicted COD v/s Measured COD **(b)** Predicted BOD v/s COD

## 5. Conclusion

As per information, greywater age fluctuates somewhere in the range of 39 and 85 percent in various nations. Greywater treatment and reuse can be utilized to give non-consumable water to latrine flushing, cultivating, vehicle and floor cleaning, and different employments. In any interaction industry, execution records like biochemical oxygen interest (BOD) with synthetic oxygen interest (COD) are utilized to decide the nature of wastewater produced (COD). The utilization of modern techniques like compound tests, conditions, and complex water test investigations is expected to gauge these amounts. The COD test requires a couple of hours, while the BOD test requires five days. Examinations in the lab are both tedious and costly. Not exclusively improve AI models for surveying water quality should be made, yet there is likewise an expanding interest for a more incorporated methodology. As of late, the Artificial Neural Network (ANN) strategy has acquired prominence. Instead of ANN, measurable

procedures like relapse models are the best techniques for exploring any connection among reliant and free factors with a restricted example size. The code was written in Python with the guide of Google Colaboratory, and a Linear Regression Model for multi-variable sources of info was created. For the model, the accompanying perceptions were made:

- The R-squared coefficient for the model was equal to 1.00
- As the number of independent variables grows, the dimension of the observed plane grows in lockstep ( $p+1$ ), making model fitting on a straight line more difficult. Scatter plots are used to build such models.
- The graph of predicted COD v/s measured COD shows a close approximation between observed and predicted value. The R-squared value was 0.9973.
- It is also possible to predict BOD using observed values of COD.

In our future work, we aim to analyse the data set by using other machine learning techniques such as ANN, SVM, etc., and comparing the attainment of the models based upon MAE also values based on RMSE.

## References

- [1] Oteng-Pepurah M, Acheampong M, DeVries N, (2018) “Greywater Characteristics, Treatment Systems, Reuse Strategies and User Perception—a Review”. Water Air, & Soil Pollution; volume 229, Article number: 255.
- [2] Samayamanthula D, Sabarathinam C, Bhandary H, (2019), “Treatment and effective utilization of greywater”, Applied Water Science; volume 9, Article number: 90.
- [3] Manna S, (2018), “Treatment of Gray Water for Reusing in Non-potable Purpose to Conserve Water in India”. International Journal of Applied Environmental Sciences, Volume 13, Number 8, pp.703-716.
- [4] Sushmitha M, Chanakya H, Khuntia H, (2019) “Efficient Grey Water Treatment and Reuse Options for India—A Review”. Springer.2019; 225-232.
- [5] Barzegar G, Wu J, Ghanbari F, (2019) “Enhanced treatment of greywater using electrocoagulation/ozonation: investigation of process parameters”. Process Safety Environment Prot.; 12:125–132.
- [6] Parparov A, Hambright D, Berman T, (2014). “Water Quality Assessment”, Ecology and Management. Springer Netherlands, Dordrecht, pp. 607-615.
- [7] Sofwan N.M, Fong S.S, (2014). “Evaluation of Various Water Quality Indices for Water Quality Assessment of Sg. Sarawak, In: Aris, From Sources to Solution”, Proceedings of the International Conference on Environmental Forensics 2013. Springer Singapore, Singapore, pp. 227-231.
- [8] Walker D, Jakovljević D, Savić D, Radovanović M, (2015). “Multi-criterion water quality analysis of the Danube River in Serbia: A visualisation approach”. Water Research 79, 158- 172.
- [9] Carlson R.E., (1977). “A trophic state index for lakes”. Limnology and Oceanography 22, 361-369.
- [10] Prasad A.G.D, Siddaraju, (2012). “Carlson's Trophic State Index for the assessment of trophic status of two lakes in Mandya district”. Advances in Applied Science Research 3, 2992-2996.
- [11] Taheri Tizro A, Ghashghaie M, Nozari H, (2016). “Assessment of Carson Trophic Index in Dam Lake: a Case Study of Ekbatan Dam”. Iranian Journal Of Health Sciences 4, 25-33.
- [12] Chang C.L, Liu H.C, (2015). “Applying the Back-Propagation Neural Network model and fuzzy classification to evaluate the trophic status of a reservoir system”. Environmental Monitoring and Assessment 187, 560-567.
- [13] Chen W.C, Chang N.B, Chen J.C, (2003). “Rough set-based hybrid fuzzy-neural controller design for industrial wastewater treatment”. Water Research 37, 95-107.
- [14] Najah A, El-Shafie A, Karim O.A, El-Shafie A.H, (2013). “Application of artificial neural networks for water quality prediction”. Neural Computing and Applications 22, 187-201.
- [15] S.K, T J, A.H.M, M. A.S, (2014). “Prediction of Water Quality of Euphrates River by using Artificial Neural Network Model (spatial and temporal study)”. International Research Journal of Natural Sciences 2, 25-38.

- [16] Seshan H, Goyal M.K, Falk M.W, Wuertz, S, (2014). "Support vector regression model of wastewater bioreactor performance using microbial community diversity indices: Effect of stress and bioaugmentation". *Water Research* 53, 282-296.
- [17] Verma A.K, Singh T.N, (2013). "*Prediction of water quality from simple field parameters*". *Environmental Earth Sciences* 69, 821-829.
- [18] Chau K.W, (2006). "A review on integration of artificial intelligence into water quality modelling". *Marine Pollution Bulletin* 52, 726-733.
- [19] Hamed M.M, Khalafallah M.G, Hassanien E.A, (2004). "*Prediction of wastewater treatment plant performance using artificial neural networks*". *Environmental Modelling & Software* 19, 919- 928
- [20] Maier H.R, Dandy G.C, (2000). "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications". *Environmental Modelling & Software* 15, 101-124.
- [21] Wu W, Dandy G.C, Maier H.R, (2014). "Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling". *Environmental Modelling & Software* 54, 108-127.
- [22] Guo Y, Wang G, Zhang X, Deng W, (2014). "*An improved hybrid ARIMA and support vector machine model for water quality prediction*", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 411-422.
- [23] Tarmizi A, Ahmed A.N, El-Shafie A, (2014). "Dissolved Oxygen Prediction Using Support Vector Machine in Terengganu River. Middle-East" *Journal of Scientific Research* 21, 2182-2188.
- [24] Zhao W.G, Wang L.Y, (2010). "Water quality evaluation based on Multiclass Fuzzy Support Vector Machine", *Advanced Materials Research*, pp. 708-711.
- [25] Jui-Sheng Chou, Chia-Chun Ho, Ha-Son Hoang, (2018), "*Determining quality of water in reservoir using machine learning*", *Ecological Informatics*.
- [26] P. McCullagh. J. Nelder. (1983) "*Generalized Linear Models*". Chapman and Hall, London.
- [27] Frauke Günther, Stefan Fritsch, (2010), "*neuralnet: Training of Neural Networks*", *The R Journal* Vol. 2/1.
- [28] Hamid Zare Abyaneh, (2014), "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters", *Journal of Environmental Health Science & Engineering*, 12:40.
- [29] Dogan E, Ates A, Ceren Yilmaz E, Eren B (2008), "Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand". *Environ Prog*, 27(4):439–445
- [30] Rene ER, Saidutta MB, (2008), "Prediction of Water Quality Indices by Regression Analysis and Artificial Neural Networks". *Int J Environ Res*, 2(2):183–188
- [31] Akratos CS, Papaspyros JNE, Tsihrintzis VA, (2008), "An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands". *Chem Eng J*, 143:96–110.
- [32] APHA (American Public Health Association): (1995), "*Standard Methods for the Examination of Water and Wastewater*". 19th edition. Washington, DC.
- [33] Chou J.S, Yang K.H, Lin J.Y, (2016c), "Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods". *Journal of Computing in Civil Engineering* 30, 1-17.