International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Research Article)

Multiple Linear Regression Analysis to Identify Influential Water Quality Parameters of Dairy ETP

Rinku Chawla *

Department of Chemistry, SSGC Government PG College, Suratgarh, Rajasthan, India.

International Journal of Science and Research Archive, 2022, 07(01), 608-615

Publication history: Received on 27 June 2022; revised on 19 September 2022; accepted on 28 September 2022

Article DOI: https://doi.org/10.30574/ijsra.2022.7.1.0203

Abstract

The focus of this study was the statistical analysis of the secondary treated water quality of SARAS dairy effluent treatment plant, Hanumangarh, Rajasthan and find the most influential quality parameters for rapid water quality monitoring. For this, wastewater quality index (WWQI) was computed using sixteen measured physicochemical parameters like temperature, pH, EC, COD, BOD, TA, TH, TDS, Na^+, K^+, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, NO_3^- and F^- as per standard methods. The WWQI values for recreational and irrigation purpose range from 43.77 to 68.97, indicate excellent to good quality of treated water in terms of physicochemical parameters. Karl pearson correlation results indicated pairs COD-BOD, COD-WWQI, BOD-WWQI, pH– WWQI, TH – Ca^{2+}, TH – Mg^{2+}, TA - WWQI and NO_3^- - SO_4^{2-} are observed with highly significant levels ($0.8 < r < 1.0$). The statistical method, multiple linear regression (MLR) analysis was performed using MS excel software to proportionate contribution of each independent variable to the overall variance. This analysis evaluated BOD as statistically the most significant with p-value < 0.05, so is sufficient to predict treated water quality.

Keywords: Effluent treatment plant; Secondary treated water; Physico-chemical parameters; WWQI; Correlation matrix; Multiple linear regression analysis

1. Introduction

Water treatment process produces clean reusable water. Properly treated wastewater is reliable and cheaper source of irrigation water due to its high nutrient content [1]. Actually, wastewater is an undervalued resource with a potential of water, energy and nutrient source. Proper treatment and management of wastewater is essential for sustainable development and environmental benefits [2]. The water quality is described by a number of physical, chemical and biological parameters depending on the specific use. Treated water quality should be monitored regularly to maintain public and environment health. It requires laboratory setup, costly chemicals and is time consuming. Multiple Linear Regression (MLR) is used to analyze the relative influence of a number of independent variables on a single dependent variable and understand the complex relationship between variables. The systematic calculation of correlation coefficient between water quality variables and multivariate analysis provide indirect means for rapid monitoring of water quality. The linear correlation is very useful to get fairly accurate idea of water quality by determining a couple of parameters experimentally [3]. [4] determined the wastewater quality index and studied statistical interrelationships amongst different parameters. The findings led to development of an equation to predict BOD and WWQI. Statistical regression analysis of twelve samples of underground drinking water from the IM2 hand pump in Moradabad, India, was conducted to study correlations among various physicochemical parameters [5]. This analysis indicated that conductivity is a key parameter, significantly correlated with ten out of twelve water quality indices, suggesting that monitoring conductivity can effectively indicate drinking water quality. Evaluation of ground water quality using multiple linear regression and structural equation modelling was done [6]. The present work applies statistical analysis to secondary treated water from SARAS dairy ETP, Hanumangarh, Rajasthan, employing Karl Pearson correlation matrix and multiple linear regression using MS Excel to identify influential quality parameters for rapid water quality

* Corresponding author: Rinku Chawla

monitoring. This approach will help to eliminate the need for exhaustive laboratory testing by focusing on significantly correlated parameters present in the regression equation.

1.1. Study area

The present study area was confined to SARAS dairy Effluent Treatment Plant, Hanumangarh, Rajasthan. The ETP of SARAS dairy, Sri Ganganagar Zila Dugdh Utpadak Sehkari Sangh Ltd. Hanumangarh is having 500 m3per day of effluent treatment capacity. Hanumangarh is one of the northern most district of Rajasthan, located between 29° 5′ to 30° 6′ North Latitude and 74° 3′ to 75º 3′ east Longitude [7]. The district headquarter Hanumangarh is situated on the bank of Ghaggar river present form of the last mythological Saraswati river (called as 'Nali' in local dialect) that divides the district headquarter into two parts as Hanumangarh town and Hanumangarh junction.

2. Materials and methods

Physicochemical Analysis- Sixteen different physicochemical parameters of water quality were measured following standard methods and procedures of sampling and testing [8] for a period of one year from February,2020 to January,2021. Temperature and pH were measured at the sampling site. DO fixation was performed at the location itself by adding manganese sulphate and alkaline KI solution and samples were stored at 4 ∘ C for determining other parameters. All the reagents used were AR grade and double distilled water was used for solution preparation [9]. To provide overall assessment of water quality the measured parameters were used to compute wastewater quality index (WWQI) [9,10] as following steps-

$$WWQI = \Sigma\, qiWi / \Sigma\, Wi \ldots\ldots\ldots\ldots (1)$$

Where, qi represents the quality rating for each of the water quality parameters used in the index and is given by-

$$qi = 100\ (Vi - V10) / (Si - Vi0) \ldots\ldots\ldots\ldots (2)$$

Vi is measured value of the ith parameter in water sample, Vi0 is the ideal value of this parameter in pure water and Si refers to acceptable limit as given in Indian standard [11].

Wi is unit weight for the ith parameter

$$Wi = k/Si \ldots\ldots\ldots\ldots (3)$$

k is constant of proportionality and is assumed unity for the sake of simplicity.

Statistical Analysis - According to [12], the Statistical Regression Analysis is highly useful tool for correlating different parameters with two-fold advantages. First, using correlation analysis it is easy to find out interrelationship between different parameters. Second, the main parameters affecting the WWQI can be easily figured out. Karl pearson correlation matrix and multiple linear regression analysis (MLR) were performed to find the relationship among WWQI and measured parameters using MS excel software. Karl pearson correlation analysis was performed to examine the relationship among all pairs of the evaluated parameters. It identifies pairs of parameters with significant relationship and helps to recognise pollution sources. MLR was then used to model the relationship among WWQI and the significant parameters. Regression analysis involves two types of variables-dependent variable and regressor or independent variables. The variable whose value is influenced is called dependent variable and the variables influencing the dependent variable are called explanatory (regressors) variables. The model equation used is:

$$Y = \beta 0 + \beta 1X1 + \beta 2\, X2 + \beta 3X3 + \beta 4X4 + \beta 5X5 + e \ldots\ldots\ldots\ldots (4)$$

The developed equation is used to predict the value of dependent variable (WWQI) based on key parameters [4,9].

3. Results and discussion

Physico-chemical analysis: The measured values of selected physico-chemical parameters of secondary treated water samples of the study site are compared with IS- 2296 [11] for agriculture water uses in table-1.

Table 1 Waste Water Quality Index (WWQI) of SARAS ETP, Hanumangarh

Parameter	**IS-2296 (Si)**	**Unit Wt (Wi)**	**EH 1 (qi Wi)**	**EH 2 (qi Wi)**	**EH 3 (qi Wi)**	**EH 4 (qi Wi)**	**EH 5 (qi Wi)**	**EH 6 (qi Wi)**	**EH 7 (qi Wi)**	**EH 8 (qi Wi)**	**EH 9 (qi Wi)**	**EH 10 (qi Wi)**	**EH 11 (qi Wi)**	**EH 12 (qi Wi)**	**EH 13 (qi Wi)**	**EH 14 (qi Wi)**	**EH 15 (qi Wi)**	**EH 16 (qi Wi)**	**EH 17 (qi Wi)**	**EH 18 (qi Wi)**	**EH 19 (qi Wi)**	**EH 20 (qi Wi)**	**EH 21 (qi Wi)**
Temp.	25	0.0400	2.83	2.912	3.328	3.520	4.096	4.448	4.544	4.400	4.224	4.000	3.968	3.872	3.712	3.600	3.520	3.248	2.992	2.720	2.624	2.304	2.480
pH	8.5	0.1176	10.20	10.980	11.765	12.549	13.333	14.118	11.765	10.196	7.059	7.843	8.627	9.412	10.196	10.196	10.980	9.412	8.627	10.196	10.980	10.196	9.412
EC	2250	0.0004	0.01	0.014	0.015	0.015	0.017	0.016	0.015	0.015	0.014	0.014	0.015	0.014	0.014	0.013	0.014	0.014	0.015	0.015	0.015	0.015	0.015
COD	100	0.0100	0.53	0.570	0.720	0.770	0.840	0.704	0.650	0.558	0.486	0.566	0.602	0.488	0.500	0.467	0.520	0.480	0.450	0.470	0.490	0.520	0.480
BOD	10	0.100	28.00	32.00	37.00	40.00	44.00	36.00	32.00	27.00	24.00	28.00	32.00	26.00	28.00	25.00	28.00	24.00	22.00	25.00	23.00	27.00	25.00
TH	300	0.0033	0.10	0.100	0.097	0.113	0.129	0.122	0.120	0.107	0.100	0.098	0.102	0.104	0.108	0.116	0.111	0.107	0.102	0.098	0.108	0.100	0.096
TA	200	0.0050	0.32	0.320	0.330	0.360	0.370	0.350	0.340	0.330	0.320	0.305	0.320	0.308	0.325	0.315	0.338	0.303	0.320	0.328	0.335	0.315	0.305
TDS	500	0.0020	0.10	0.087	0.102	0.114	0.137	0.126	0.142	0.131	0.113	0.107	0.118	0.109	0.100	0.094	0.110	0.118	0.107	0.113	0.094	0.087	0.084
Na+	200	0.0050	0.17	0.180	0.175	0.185	0.205	0.195	0.183	0.163	0.155	0.145	0.160	0.155	0.178	0.168	0.183	0.188	0.200	0.213	0.193	0.185	0.173
K+	15	0.0667	3.33	3.156	3.644	2.667	2.844	3.333	2.756	2.756	2.667	2.844	2.844	2.978	3.022	3.067	3.067	3.200	3.289	3.333	3.422	3.378	3.200
Ca2+	75	0.0133	0.36	0.341	0.327	0.384	0.427	0.412	0.412	0.384	0.327	0.299	0.341	0.348	0.391	0.398	0.384	0.370	0.363	0.327	0.363	0.341	0.299
Mg2+	30	0.0333	1.19	1.133	1.111	1.300	1.511	1.400	1.344	1.133	1.189	1.244	1.189	1.211	1.133	1.300	1.244	1.189	1.111	1.133	1.244	1.133	1.189
NO3-	50	0.0200	0.04	0.048	0.052	0.056	0.068	0.056	0.052	0.064	0.072	0.068	0.056	0.048	0.044	0.052	0.056	0.064	0.052	0.044	0.040	0.040	0.032

SO4-2	400	0.0025	0.00	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.003
Cl-	250	0.0040	0.06	0.067	0.069	0.075	0.077	0.075	0.067	0.069	0.072	0.066	0.064	0.059	0.070	0.078	0.066	0.070	0.074	0.066	0.080	0.072	0.061
F-	1.5	0.6667	5.78	4.889	4.444	6.222	7.111	4.444	5.333	6.222	7.556	7.111	6.667	7.556	7.111	7.556	6.222	7.111	8.000	7.111	7.556	6.667	7.111
∑Wi		1.0899																					
∑qiWi			53.02	56.80	63.18	68.33	75.17	65.80	59.73	53.53	48.36	52.71	57.08	52.67	54.91	52.42	54.82	49.88	47.71	51.17	50.55	52.34	49.92
WWQI = ∑qiWi / ∑Wi			48.65	52.12	57.97	62.70	68.97	60.38	54.80	49.11	44.37	48.37	52.37	48.32	50.38	48.10	50.30	45.76	43.77	46.95	46.38	48.02	45.80

Table 2 Correlation Matrix of Water Quality Parameters of ETP SARAS, Hanumangarh

	Temp.	**pH**	**EC**	**COD**	**BOD**	**TH**	**TA**	**TDS**	**Na^+**	**–K^+**	**Ca^{+2}**	**Mg^{+2}**	**NO_3^-**	**SO_4^{-2}**	**Cl^-**	**F^-**	**WWQI**
Temp.	1.0000																
pH	0.1373	1.0000															
EC	0.2109	0.5877	1.0000														
COD	0.4272	0.7233	0.6417	1.0000													
BOD	0.3727	0.7318	0.5576	0.9762	1.0000												
TH	0.5342	0.7008	0.4809	0.5619	0.5387	1.0000											
TA	0.3560	0.7940	0.7213	0.7853	0.7563	0.7341	1.0000										
TDS	0.7823	0.3162	0.6277	0.5181	0.4268	0.6217	0.5578	1.0000									
Na^+	-0.3253	0.5732	0.6365	0.2031	0.2044	0.3624	0.5318	0.1626	1.0000								
K^+	-0.6730	0.1832	-0.0185	-0.2148	-0.1955	-0.3079	-0.2106	-0.5208	0.4062	1.0000							
Ca^{+2}	0.4738	0.6640	0.3536	0.4206	0.4116	0.9308	0.6746	0.5765	0.4083	-0.2306	1.0000						
Mg^{+2}	0.4925	0.5964	0.5403	0.6288	0.5914	0.8743	0.6567	0.5401	0.2240	-0.3414	0.6366	1.0000					

NO_3^-	0.7161	-0.0962	0.2175	0.3516	0.2792	0.3055	0.2395	0.6436	-0.2320	-0.5836	0.2292	0.3445	1.0000				
SO_4^{-2}	0.6124	-0.0817	0.3090	0.2056	0.1378	0.3621	0.2842	0.6616	0.0247	-0.4924	0.3208	0.3437	0.8886	1.0000			
Cl^-	0.0501	0.3858	0.2785	0.2525	0.1836	0.5197	0.5114	0.0937	0.4391	0.0563	0.5363	0.3968	0.2129	0.2942	1.0000		
F^-	-0.0402	-0.4247	-0.0724	-0.4733	-0.5035	0.0353	-0.2099	-0.0524	0.0443	-0.1457	-0.0188	0.1026	0.0642	0.3782	0.2431	1.0000	

Table 3 Multiple Linear Regression Equations for SARAS ETP, Hanumangarh

Regression Statistics								
Multiple R	0.9975							
R Square	0.9949							
Adjusted R Square	0.9915							
Standard Error	0.5973							
Observations	21							
ANOVA								
	df	**SS**	**MS**	**F**	**Significance F**			
Regression	8	838.8903	104.8613	293.8732	1.42E-12			
Residual	12	4.2819	0.3568					
Total	20	843.1722						
	Coefficients	**Standard Error**	**t Stat**	**P-value**	**Lower 95%**	**Upper 95%**	**Lower 95.0%**	**Upper 95.0%**
Intercept	-14.6197	7.9019	-1.8501	0.0891	-31.8366	2.5971	-31.8366	2.5971
pH	2.3765	1.1907	1.9960	0.0691	-0.2177	4.9707	-0.2177	4.9707
EC	0.0073	0.0072	1.0200	0.3279	-0.0083	0.0229	-0.0083	0.0229
COD	-0.0280	0.0713	-0.3927	0.7014	-0.1832	0.1273	-0.1832	0.1273
BOD	0.9344	0.1279	7.3082	0.0000	0.6559	1.2130	0.6559	1.2130
TH	-1.2650	1.7266	-0.7326	0.4779	-5.0270	2.4970	-5.0270	2.4970
TA	0.0235	0.0509	0.4615	0.6527	-0.0874	0.1343	-0.0874	0.1343
Ca^{+2}	3.4154	4.3622	0.7830	0.4488	-6.0890	12.9197	-6.0890	12.9197
Mg^{+2}	5.8079	7.1284	0.8148	0.4311	-9.7234	21.3393	-9.7234	21.3393

Table 4 Multiple Linear Regression Equations for SARAS ETP, Hanumangarh

Regression Statistics								
Multiple R	0.9816							
R Square	0.9635							
Adjusted R Square	0.9616							
Standard Error	1.2731							
Observations	21							
ANOVA								
	Df	SS	MS	F	Significance F			

Regression	1	812.3784	812.3784	501.2435	4.05831E-15			
Residual	19	30.7938	1.6207					
Total	20	843.1722						
	Coefficients	**Standard Error**	**t Stat**	**P-value**	**Lower 95%**	**Upper 95%**	**Lower 95.0%**	**Upper 95.0%**
Intercept	19.4312	1.4425	13.4704	0.0000	16.4120	22.4504	16.4120	22.4504
BOD	1.0857	0.0485	22.3885	0.0000	0.9842	1.1872	0.9842	1.1872

The results revealed that treated effluent is of alkaline nature (pH range 7.9-8.8) and is within permissible limits of 6.5-8.5 [13] for most of the samples and within [14] public sewer limits of 6.0-9.0 for wastewater discharge into the environment. The soaps and detergents used for cleaning of dairy processing equipment and vessels raise the pH of the effluent. BOD values of 67% samples are higher than permissible limits of FAO and inland surface waters standards (30mg/l) [15] The other physical as well as chemical parameters are within permissible values of Indian standards. The WWQI values ranging from 43.77 to 68.97 are indicative of excellent to good quality of treated water in terms of physicochemical parameters.

Statistical Analysis : Statistical analysis helps to find the main pollution source in a water body and create effective water management strategies. The results of the present statistical analysis study, performed by Karl Pearson Correlation Matrix and Multiple Linear Regression statistics are presented in tables-2 to 4.

Correlation Matrix Analysis - In the present study, Karl Pearson Correlation Matrix was used to identify statistically significantly correlated variables. Correlation coefficient (r) between any two parameters, x and y is calculated for all parameters including WWQI, water temperature, pH, electrical conductivity (EC), COD (Chemical Oxygen Demand), BOD(Biochemical Oxygen Demand), total hardness (TH), total alkalinity (TA), total dissolved solids (TDS), sodium, potassium, nitrate, sulphate, chloride, fluoride, calcium and magnesium of the undertaken samples [16] . The numerical values of correlation coefficients (r) are tabulated in Table-2. The degree of line association between any two of the water quality parameters as measured by the simple correlation coefficient (r) is presented as 17 x 17 correlation matrix. Positive correlation is obtained between 127 combinations (83.0 % of the total number) and the rest 26 combinations (17.0 % of the total number) exhibit negative correlation. WWQI has been found to show positive correlation with all the parameters except K^+ and F^- ions. Out of the 153 correlation coefficients, correlation coefficient (r) between COD-BOD (0.9762), COD-WWQI (0.9698), BOD-WWQI (0.9864), pH– WWQI (0.809), TH – Ca^{2+}(0.93), TH – Mg^{2+} (0.87), TA - WWQI (0.83) and NO_3^- - SO_4^{2-} (0.8886) are observed with highly significant levels ($0.8 < r < 1.0$). 25 correlation coefficients give the moderate significant ($0.6 < r < 0.8$) level of r values and there are 21 values of r which belong to the significant levels ($0.5 < r < 0.6$).

Multiple Linear Regression Analysis - The focus of this study was to identify the main parameters affecting the treated water quality. Multiple linear regression (MLR) analysis was carried out using MS-Excel software to serve the purpose. The parameters exhibiting correlation coefficients (r)>0.5 with WWQI were chosen for regression analysis. Correlation matrix shows parameters pH, EC, COD, BOD, TH, TA, Ca^{2+} and Mg^{2+} exhibit correlation coefficients (r) > 0.5 with WWQI. It shows that out of 17 parameters taken only 8 parameters are significantly correlated with WWQI (Table-2). Multiple linear regression MLR was executed taking WWQI as dependent variable and selected eight parameters as regressors.

Unstandardized coefficients from Table 3, were replaced with the coefficients of equation-4 and Y predicted is presented in equation 5. Table 3 shows that the model fit with 99.49% of accuracy.

$$Y = -14.6197 + 2.3765X1 + 0.0073X2 - 0.0280X3 + 0.9344X4 - 1.2650X5 + 0.0235X6 + 3.4154X7 + 5.8079X8 \quad \text{(Eq. 5)}$$

Where, Y= WWQI, X1 = pH, X2 = EC, X3= COD, X4 = BOD, X5 = TH, X6 = TA, X7 = Ca^{2+}, X8 = Mg^{2+}

This model is fitted significantly as F=293.8732 (Sig F: 1.42E-12) (Table-3). Findings of regression analysis depicts p-value for BOD less than 0.05. Taking BOD, the regression analysis was again executed in order to develop a new equation. So, EC, COD, TH, TA, Ca^{2+} and Mg^{2+} are dropped, recalculated the WWQI, again fitted the regression model and got the regression equation as follows (Equation 6) with 96.35% of accuracy. (Table-3)

Y(E) = 19.4312+1.0857 X1 ……………………………………… (Eq. 6)

Where, Y(E) = WWQI after dropping EC, COD, TH, TA, Ca2+ and Mg2+, X1= BOD

BOD has significant effect on WWQI, with p-value less than 0.05 (Table 4). This model is fitted significantly as F=501.2435 (Sig F: 4.05831E-15) (Table 4).

4. Conclusions

Based on above results and discussions, it is concluded that BOD is an important physicochemical water quality parameter. MLR analysis suggests that water quality of study site can be effectively checked by controlling BOD alone. This may also be applied to other dairy ETP also as dairy industry mainly carries organic pollution load. This methodology will greatly facilitate rapid monitoring of the status of water pollution and thereby ETP functioning. This type of pollution study suggests some quick, effective and economic way of water quality management.

References

[1] Farhadkhani M., Nikaeen M. et.al. (2018). Effects of irrigation with secondary treated wastewater on physicochemical and microbial properties of soil and produce safety in a semiarid area. Water Research144: 356-364.

[2] UN Water (2017) Wastewater: the untapped resource. Facts and figures. The United Nations World Water Development Report

[3] Jothivenkatachalam K., Nithya A. And Chandra Mohan S., (2010). Correlation Analysis of Drinking Water Quality in and around Perur Block of Coimbatore District, Tamil Nadu, India. Rasayan. J. Chem Vol.3 (4), 649-654.

[4] Khambete A.K., Christian R.A. (2014). Statistical Analysis To Identify The Main Parameters Effecting WWQI of Sewage Treatment Plant and Predicting BOD. International Journal of Research In Engineering and Technology, Vol.03, 186-195.

[5] Kumar, N., Sinha, D. K. (2010).” Drinking water quality management through correlation studies among various physico-chemical parameters: A case study.” International journal of environmental sciences, 1(2), pp.253-259.

[6] Chenini I. and Khemiri S. (2009). Evaluation of ground water quality using multiple linear regression and structural equation modelling, Int. J. Environ. Science. Tech., 6(3), 509-519.

[7] Districts of Rajasthan. https://www.rajras.in/rajasthan/districts/hanumangarh/ june,16, 2021

[8] American public health association (APHA) Standard Methods for the Examination of Water and Wastewater, 23rd Edition Washington, DC, USA.

[9] Rinku Chawla and Atul Arora. (2022). Statistical analysis to identify main parameters affecting WWQI – A study on STP Hanumangarh, Rajasthan. IJBPAS, January, Special Issue, 2022, 11(1): 318-328. https://doi.org/10.31032/IJBPAS/2022/11.1.1038.

[10] Ramakrishna Ch., Rao D. M., Rao K. S. and Srinivas N. (2009). Studies on Groundwater Quality in Slums of Visakhapatnam, AP, Asian Journal of Chemistry Vol. 21, No. 6, 4246-4250.

[11] Tolerance limits of inland surface waters, Class-E. IS 2296 (1982). https://www.indiawaterportal.org/sites/default/files/iwp2

[12] Rastogi, G.K. and Sinha, D. K. 2011. A novel approach to water quality management through correlation study. Journal of Environmental Research & Development. 5(4) : 1029-1035.

[13] BIS (2003). Drinking water specification: Bureau of Ind. Standards, New Delhi, IS : 10500, 11.

[14] cpcb. 2007b. advance methods for treatment of textile industry effluents, resource recycling series:reres/&/2007. central pollution control board, India.

[15] Tolerance Limits for Industrial Effluents. IS:2490 (Part I) – 1981. http://www.worldenviro.com/ effstd.htm

[16] Rinku Chawla and Atul Arora (2021) Identification of Influential Parameters Affecting WWQI Using Statistical Analysis with Reference to STP, Sriganganagar, Rajasthan. Poll Res. 40(4): 1433-1439.