Journal of Civil Engineering and Environmental Technology
p-ISSN: 2349-8404; e-ISSN: 2349-879X; Volume 6, Issue 1; January-March, 2019, pp. 1-5

http://www.krishisanskriti.org/Publication.html

Assessment of Various Sewerage Treatment Plants using Multivariate Cluster and Factor Analysis in Delhi, India

Prerna Sharma[a*], Smita Sood[b] and Sudipta K Mishra [c]

[a*]*Department of Basic & Applied Sciences, G D Goenka University, Gurgaon*
[b]*Department of Basic & Applied Sciences, G D Goenka University, Gurgaon*
[c]*Department of Civil Engineering, G D Goenka University, Gurgaon*

Abstract—*Wastewater/Sewage effluent quality and its reutilisation can also be assessed with the help of various multivariate tools. Nowadays multivariate tools have proved to be efficient method for the quality assessment and management of various wastewater treatment technologies. The present study focus on the Assessment of various waste water treatment technologies using multivariate techniques for different Sewerage Treatment Plants in Delhi. Twenty three different STP's from Delhi were focussed by implementing the Cluster and Principal Component /Factor Analysis. Cluster Analysis divided the twenty three STP's in five groups exhibiting similar characteristics with respect to the removal efficiencies of the selected physiochemical parameters. Then the hierarchical clustering analysis was applied for 16 out 23 Sewerage Treatment Plants based on Activated Sludge Process (ASP) technologies. Results of the Factor/Principal Component analysis indicated that TSS and BOD are the two main parameters among the physicochemical parameters which are contributing maximum towards the performance of the STP's.*

Keywords: *Multivariate Techniques, Cluster Analysis, Principal Component Analysis (PCA), Factor Analysis.*

1. INTRODUCTION

Municipal Corporation usually takes care of the various sewerage treatment plants (STP's). In Delhi the same has been taken up by Delhi Pollution Control Board (DPCC) as well as Delhi Jal Board (DJB). Multivariate techniques have been utilised for the assessment of various wastewater/sewage treatment (Boyacioglu H.2006). Multivariate techniques are used worldwide as they are efficient in assessing the potential parameters affecting the Wastewater treatment technologies and further helping deciding the performance and management related to wastewater/sewage or water quality (Vega et al. 1998, Yerel et.al 2012, Wang ZM et.al Al.2014).

Many researchers have also worked on evaluating the efficiency of various STP's in Delhi (Priyanka Jamwal et .al. 2009, Colmenarejo et al. 2006), which primarily focussed on the calculating the integrated efficiency and comparing the same with the standard integrated efficiency to assess the performance of the selected STP's under their course of investigation. The application of Multivariate Techniques not only makes easy to assess the quality of Wastewater/sewage of water quality but along with that it also helps to how one variable can influence the other under defined circumstances/situations and what are the prime most variables affecting the performance or giving the optimum output as the function of input variables (Simeonova et al. 2003, Li X et. al 2014).

In the present study multivariate techniques used for the assessment of various waste water treatment technologies used in different Sewerage Treatment Plants in Delhi are Multiple Regression Analysis, Correlation Analysis, Sensitive Analysis and Principal Component analysis (PCA).Multiple Regression Analysis predicted the relation between the dependent and the independent variables, correlation analysis showed how the variables are associated with one another, sensitive analysis were performed using heat map to determine how different values of an independent variable impact a particular dependent variable under given set of assumptions.

Principal Component Analysis (PCA) in the final portion of the paper emphasis on the reduction of large set of variables into smaller one supporting the fact that the smaller set includes the maximum valuable information of the larger set of variables taken into account for the study (Helena et. Al. 2000).

2. MATERIALS AND METHODS

2.1 Study Area

The present study was carried out on 23 different sewerage treatment plants based upon different sewage treatment technologies in Delhi. The basic study was carried on the three sewage treatment technologies mainly Activated Sludge Process (ASP), Extended Aeration and Densadeck. Hence the STP's based upon these technologies are being focussed in this study.

2.2 **Sampling Points and Frequency**

The sampling points for the above mentioned STP's in the study area was Outlet channel i.e. it focused on the effluents of each selected STP's. Sampling was done every month from the year 2012-2017 (APHA 1998).

2.3 Parameters Analysed

The parameters considered for present study are Total Suspended Solids (TSS), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Oil and grease, Ammonical Nitrogen and phosphates. All the parameters were tested as per (APHA 1998) standards.

2.4 Multivariate Analysis

The multivariate techniques used in the present study are Cluster Analysis (CA) Analysis which was performed on the removal efficiencies for all the physiochemical parameters regarding each selected STP form the year 2012-2017.After that cluster analysis will be performed on 16 STP's out of the selected 23 STP's based on Activated Sludge Process (ASP) technology and finally on the average influent of these 16 STP's. After the cluster analysis, Principal Component Analysis (PCA)/Factor will be implemented on the average influent of the 16 STP's based upon ASP technology. All the Multivariate analysis was carried out on SPSS.

3. RESULTS AND DISCUSSIONS

3.1 Cluster Analysis Results

In the first stage of the paper Cluster analysis is performed on the selected 23 STP's based upon different sewage treatment technologies to foresee that how the cluster are being formed between these STP's i.e. observation which are having similar characteristics/values will be grouped in one category therefore each member of the a group will be different from the member of the another group. Here the Cluster analysis utilised is the Hierarchical Cluster. Hierarchical Clustering orders the rows and/or columns based upon the similarity. It helps to understand correlation very easily in the given set of data. Homogeneity in the cluster is being explained by the agglomeration schedule.

From the fig.1 it is clear that as we have multiple levels of clustering built up in the dendogram hence the name is hierarchical cluster. The dendogram obtained indicated that broadly there are 5 clusters i.e. STP's with the number 6 14 16 23 is the first cluster, 19 20 23 15 is the second cluster, 11 13 9 17 is the third cluster, 5 7 is the fourth cluster and 1 4 is the fifth cluster. Dendogram also indicates the relationship between these clusters. From the figure it can be seen that cluster first and second are more related to one another than with the rest of the clusters. Similarly cluster third and fourth are more related to one another than with the other bunch of clusters. This clearly states that the removal efficiencies of the STP' forming the clusters are almost of the same patter indicating the homogeneity in the group. But the STP's when will be compared to the STP's of another group of cluster then they will be entirely different from each other showing heterogeneous nature with one another.

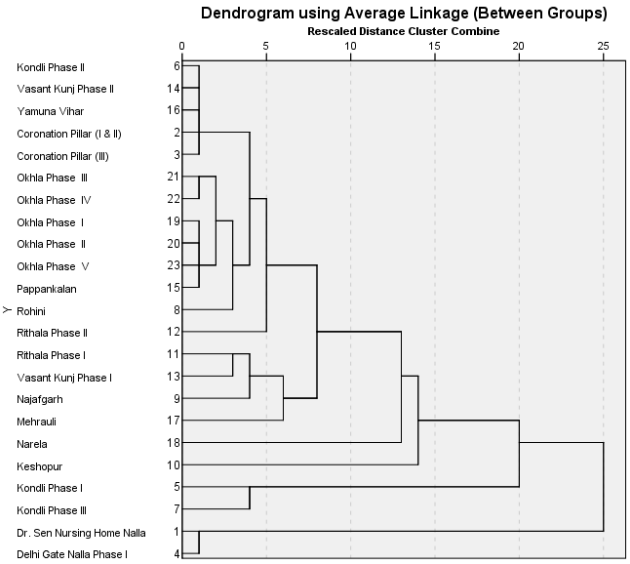

Fig. 1: Dendogram depicting various clusters formed between the different STP's

It is clear from rescaled distance obtained in the fig. 1 that STP 5, 7, 1 and 4 are not highly correlated with the other STP's in each cluster. As out of 23 STP's 16 STP's are based on the same technology i.e. ASP hence hierarchical clustering analysis is conducted for the same in with respect to the removal efficiencies for various physiochemical parameters. The dendogram obtained in fig.2 indicated that broadly there are 7 clusters i.e. STP's with the number 4 10 1 2 is the first cluster, 1 8 is the second cluster, 12 13 16 is the third cluster, 14 15 13 is the fourth cluster, 3 5 is the fifth cluster, 3 7 form the sixth cluster and 3 5 7 form the final i.e seventh cluster. It is clear from fig. 2 that none of the clustering pattern is similar in both of dendograms. In first clustering pattern i.e. 4 10 1 2 Kondli phase-II & Yamuna Vihar STP's have clustered together as they belong to the same Sewerage Zone i.e. Shahdara. Smilarliry all the STP's in Cluster third & fourth belong to the same sewerage zone i.e. Okhla Sewerage Zone, hence they formulate one cluster. Cluster fifth having Kondli phase –I and Kondli phase –II belong to the Shahdara Sewerage Zone (Priyanka et al. 2009).

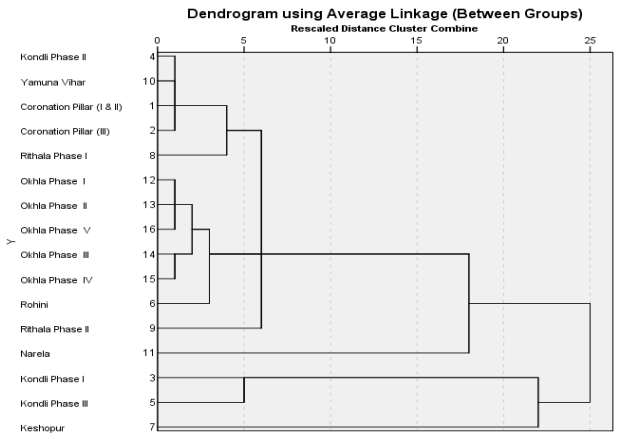

Fig. 2: Dendogram obtained for the removal efficiencies of various STP's based on Activated Sludge Process (ASP) technology.

Hydraulic Retention Time (HRT) is an important criteria taken into design consideration of the plant and also it is very important to assess the performance of the plant. Dendogarm were prepared for the total HRT of the STP's based upon ASP technology using hierarchical cluster analysis.

From fig.3. it is clear that total five clusters are being formed i.e. cluster one as

19 20 18 9 12 8 11 12 7, Cluster second as 5 6 13 14, cluster third as 10 15 16, cluster fourth as 10 16 15 17 and the cluster fifth as the last one i.e. 10 15 16 13. Th first cluster have the combination of the STP's form Shahdara, Rohini- Rithala, Okhla, Kesopur and Coronation Pillar Sewerage Zone of Delhi Watershed. Whereas if we see the third cluster i.e. 10 15 16, STP 10 i.e. Nilothi and STP 16 i.e. Nazafgarh STP belong to the same Sewerage Zone i.e. Kesopur Sewerage Zone. Similarly in the last cluster i.e. cluster fifth, STP 15 & 13 belong to the same Sewerage Zone i.e. Okhla Sewerage Zone and STP 10 & 16 in Kesopur Sewerage Zone.

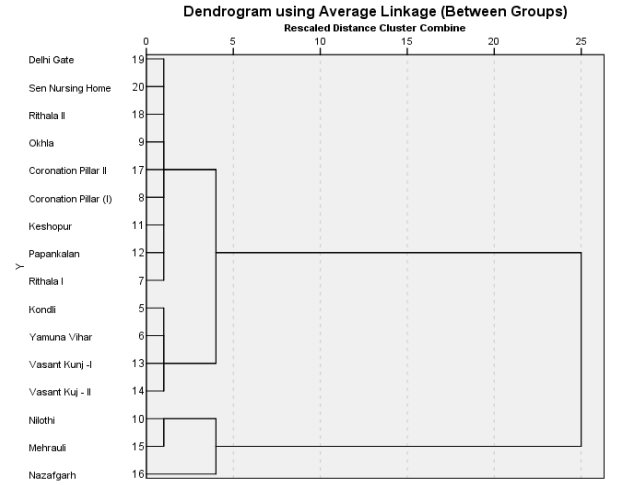

Fig. 3: Dendogram obtained for the Total HRT of various STP's based on Activated Sludge Process (ASP) technology

To foresee the clustering patterns (in terms of average influent) again hierarchical patterns was conducted for these 16 STP's based upon the ASP technology. Fig 3 shows the dendogram for the same. Six clusters are formed, cluster one as 2 5 4 1 8 9 3, cluster two 14 15, cluster three 7 12 13, cluster four 11 6, cluster five 16 13 7 and cluster six 10 8 14. Here too none of the clusters match the clustering pattern with the dendogram shown in fig.2.

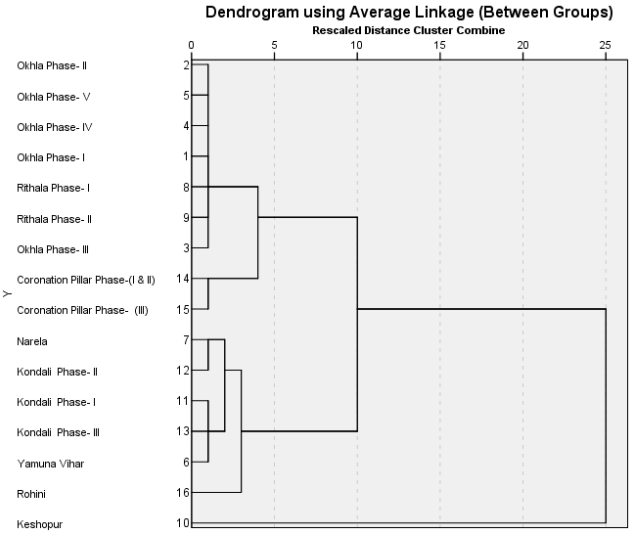

Fig. 4: Dendogram obtained for average influent of various STP's based upon ASP technology.

3.2 Principal Component Analysis (PCA)/Factor Analysis Results

After the cluster analysis done on the 16 STP's based on ASP technology, PCA/Factor Analysis was performed on the same dataset of these 16 STP's. It is the dimension reduction tool which will give the selected parameters whose removal efficiencies are importantly defining the performance of the STP's. And the reduced parameters will be called as "factor" or the "principal component factors".

There are two factor extraction methods which are obtained in results that is Total Variance Explained and the Scree Plot. These two methods tells us that how many factors can be retained that is how many factors do we want to keep as the solutions. Total 3.2.1 describe the significant parameters among the seven selected parameters. For the parameters to be significant the total initial eigenvalues should be greater than 1.so it is clear from the table that component 1and 2 (pH and TSS) are the most important among all as there value is 3.362 and 1.269 respectively. TSS contributes the total percentage variance as 56.040% in the data set and BOD as 21.143%. Out of TSS and BOD 77.183 % of the cumulative % of variance is being explained by BODindicating it very important component. Hence it is clear that out of six parameters only TSS and BOD should be retained.

Table 3.2.1 Total Variance Explained

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3.362	56.040	56.040	3.362	56.040	56.040	3.361
2	1.269	21.143	77.183	1.269	21.143	77.183	1.284
3	.818	13.630	90.813				
4	.448	7.469	98.282				
5	.091	1.513	99.795				
6	.012	.205	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Scree Plot is the visual representation of the variance of the selected component in the form of the Eigen value is being depicted in fig 3.2.1.showing the maximum Eigen value for first component i.e TSS followed by BOD. The rate of change or the slope is quiet minimum as we move for the eigen value of the component from 2-7. Between the component one and two i.e. TSS and BOD big drop is being observed. Hence again indicating that TSS and BOD are the components that has to be retained.

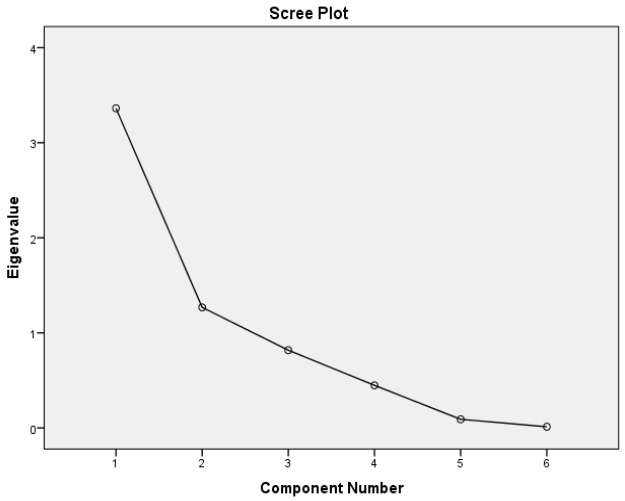

Fig. 3.2.1: Scree Plot

Component Matrix obtained in table 3.2.2 indicates that how each of the individual parameters do in terms of getting at that component. More clearly indicates the Pearson Correlation between the parameters and the components. Components 1 and 2 are also known as the factor loadings. These loadings tell you that how strong is the relationship between the component and the parameter in the solution. From the matrix it is clear that COD, BOD and TSS correlates or loads more on component 1 than component 2.

Highest loading parameter is COD on component 1 with .967 value. If the parameter loading on the component is less than. 3 then it is meant that there is no meaningful loading of the parameter on the component. Least loading is shown by Oil & Grease on component 2 with -.376 value.

Table 3.2.2 Component Matrix

Component Matrix[a]

	Component	
	1	2
TSS	.929	.030
BOD	.901	.015
COD	.967	.172
OilGrease	.788	-.376
AmmonicalNitrogen	-.126	.872
Phosphates	.342	.580

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

The identification of nature of the component is depicted by the Pattern Matrix (table 3.2.3) when the rotation method used is Oblimin. From the table it is clear that BOD, COD, TSS, Oil & Grease are loading nicely on component 1than component 2. Here too if the value of the component is less than .3 than it is not contributing in a meaning full manner towards the parameters. Hence component 2 does not have meaning full role toward Oil & Grease, as its value is -.353. Similarly Component 1 doesn't contributes towards ammonical nitrogen in a meaningful way.

Table 3.2.3 Pattern Matrix

Pattern Matrix[a]

	Component	
	1	2
TSS	.924	.058
BOD	.897	.042
COD	.950	.201
OilGrease	.819	-.353
AmmonicalNitrogen	-.200	.869
Phosphates	.292	.590

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.[a]

a. Rotation converged in 4 iterations.

4. CONCLUSION

In the present study 23 different STP's of Delhi are taken into account and multivariate analysis is being implemented on them. The major two multivariate techniques used were Cluster Analysis and Factor/Principal Component Analysis (PCA). From the Cluster analysis applied to the 23 STP's it is clear that the dataset was segregated into five clusters having similar pattern of the removal efficiencies of the STP's. Out of these seven STP's some of the clusters were closely related to

another and some of them were not related to each other. Thereafter the hierarchical clustering pattern was obtained for 16 STP's based on the ASP technology. From the results obtained it can be concluded that none of the clustering pattern match the pattern obtained for 23 STP's. Dendogarms obtained for the total HRT of the STP's based upon ASP technology revealed that only few STP's form the cluster based upon the Sewerage Zonal distribution hence falls under the same category, else clustering pattern shows the combination of STP's from more than one Sewerage Zone. Again when hierarchical clustering was implemented on the average influent of 16 based on ASP technologies none of the pattern matches the clustering pattern obtained in the earlier two cases. Results obtained for Factor/PCA Analysis it is concluded that TSS and BOD are the major to parameters out of the six selected parameters. 96% and 81% of the variance in COD & BOD is explained by the component 1 and 2 i.e TSS and BOD. Out of the two component obtained the results also concluded that 77.18% of the cumulative variance is being given by BOD. However COD, BOD, TSS and Oil & Grease have more loadings on component 1 than component 2.

REFERENCES

[1] American public health association (APHA) (1998). Standard methods for the examination of waters and wastewaters (20th edn). Washington, DC, USA.

[2] Boyacioglu H. Surface water quality assessment using factor analysis. Water S.A. (2006); 32(3):389-393.

[3] Colmenarejo, M. F., Rubio, A., Sanchez, E., Vicente, J., Gracia, M. G., & Bojra, R. (2006). Evaluation of

[4] municipal wastewater treatment plants with different technologies at Las-Rozas, Madrid (Spain). Journal of Environmental Management, 81, 399–404.

[5] Helena B, Pardo R, Vega M, Barrado E, Fernández JM, Fernández L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Research. 2000;34:807-816.

[6] Li X, Li P, Wang D, Wang Y. (2014) Assessment of temporal and spatial variations in water quality using multivariate statistical methods: a case study of the Xin'anjiang River, China. Frontiers of Environmental Science and Engineering.;8(6):895-904.11.

[7] Priyanka Jamwal, Atul K.Mittal, Jean-Marie Mouchel (2011), Efficiency evaluation of sewage treatment Plants with different technologies in Delhi (India), Environ Monit Assess (2009), 153, 293-305

[8] Simeonova V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimitzis T. (2003) Assessment of the surface water quality in Northern Greece. Water Research.; 37:4119-4124.

[9] Vega M, Pardo R, Barrato E, Deban L. (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research.; 32:35813592

[10] Wang ZM, Chen LD, Zhang HP, Sun RH. (2014); Multivariate statistical analysis and risk assessment of heavy metals monitored in surface sediment of the Luan River and its tributaries, China. Human and Ecological Risk Assessment. 20:1521-1537.

[11] Yerel S, Ankara H. (2012) Application of multivariate statistical techniques in the assessment of water quality in Sakarya River, Turkey. Journal Geological Society of India;79:89-93.