Revista Română de Inginerie Civilă, Volumul 6 (2015), Numărul 1

Application of multivariate statistical analysis to evaluate parameters in a wastewater treatment

Aplicarea analizei statistice multivariate pentru evaluarea parametrilor la o stație de epurare a apei

Marius-Daniel Roman

Technical University of Building Services Engineering Department
Boulevard December 21, no. 128-130, 400604, Cluj-Napoca, ROMÂNIA
E-mail: Marius.ROMAN@insta.utcluj.ro

Rezumat. *Autorul lucrării de față prezintă rezultatele în urma aplicării analizei componentelor principale (PCA). Datele colectate de la influentul și efluentul stației de epurare a apelor uzate Gherla în perioada (mai 2013- mai 2014), au fost analizate pentru a găsi o relație între parametrii fizico-chimici. PCA este conceput pentru a transforma variabilele originale în noi variabile necorelate (axe), numite componente principale, care sunt combinații liniare ale variabilelor originale. Noile axe se află de-a lungul direcției de variație maximă. PCA oferă o modalitate obiectivă de a găsi indicii de acest tip, astfel încât variația în date să poată fi pe cât posibil concisă. Aplicarea PCA este utilizată pentru a obține, pe lângă datele statistice, imaginile care pot fi studiate și interpretate vizual.*

Cuvinte cheie: PCA, analiza cluster, parametrii, apă uzată.

Abstract. *The author of the present paper presents the results following the application of Principal Components Analysis (PCA). The collected data of influent and effluent from Gherla's wastewater treatment plant during (May 2013- May 2014),were analyzed to find a relationship between physical-chemical parameters. PCA is designed to transform the original variables into new, uncorrelated variables (axes) called the principal components, that are linear combinations of the original variables. The new axes lie along the directions of maximum variance. PCA provides an objective way of finding indices of this type so that the variation in the data can be accounted for as concisely as possible. The application of PCA used to obtain, besides statistical data, the images which can be studied and interpreted visual.*

Key words: PCA, cluster analysis, parameters, wastewater.

1. Introduction

The characteristics of municipal wastewaters are varied, depending upon the sources of discharge, the effluents from industries, land uses and groundwater levels [4].

Municipal wastewater characteristics properties its essential as well as concentrations of suspended and dissolved inorganic and organic pollutants. Among the organic pollutants present in sewage are phenols, hydrocarbons, biological oxygen demand (BOD), carbohydrates, lignin, fats, soaps, synthetic detergents, proteins and their decomposition products, as well as various natural and synthetic organic chemicals from the process industries [1]. Inorganic pollutants comprises of compounds of trace minerals, nitrogen and phosphorous.

A major role of wastewater treatment is to restore and maintain the chemical, physical and biological integrity of waters. There is a rapid growth in the re-use of municipal wastewater for irrigation and ground water recharge, which necessitates enhanced treatment to remove nutrients (nitrogen and phosphorus), suspended solids, and other contaminants. The changing of the environmental parameters may change the life cycle of populations from an entire species.

In the present paper, a series of physical-chemical parameters of wastewater, before and after water treatment at Gherla's WWTP, recorded over one year were included into analysis, in order to find relevant relationships between them.

2. Problem formulation

Principal component analysis (PCA) is a vector space transformation often used to transform multivariable space into a subspace which preserves maximum variance of the original space in minimum number of dimensions. The measured process variables are usually correlated to each other. PCA finds linear combinations of variables that describe major trends in data set. Mathematically, PCA is based on an orthogonal decomposition of the covariance matrix of the process variables along the directions that explain the maximum variation of the data.

The collected data at every month at Gherla's wastewater treatment plant (WWTP) include pH, total suspended matter (MTS), biochemical oxygen demand in five days – BOD_5, chemical oxygen demand – COD, and ammonium (NH_4-N) are presented in table 1. Values exceeding maximum standard limits (according to NTPA 001/2005 and NTPA 002/2005, respectively) in table 1.

Table 1

Evolution of parameters in wastewater treatment plant of Gherla, during 2013-2014

Year	Mo	pH		MTS		BOD_5		COD		NH_4-N	
		infl.	effl.	infl.	effl.	infl.	effl.	infl.	effl.	infl.	effl.
2013	May	7,53	7,51	233	21,26	184,42	6,26	267,35	53,87	15,0	1,15
	Jun	7,58	7,54	100,90	20,03	134,97	4,57	204,16	50,83	11,4	0,90
	Jul	7,54	7,53	106,19	20,48	148,39	5,06	229,54	71,33	10,6	0,77
	Aug	7,26	7,10	156,35	22,77	187,63	7,74	264,48	58,42	10	0,62
	Sept	7,34	7,18	139,50	16,70	227,43	9,10	325,12	59	10,9	0,82
	Oct	7,38	7,29	155,42	14,74	237,48	9,66	421,35	54,41	11,0	0,93
	Nov	7,41	7,27	166,97	16,47	213,50	8,73	433,92	56,64	10,2	0,70
	Dec	7,35	6,95	246,77	19,13	246,58	7,32	458,19	64,98	14,9	1,12
2014	Jan	7,35	7,01	221,87	17,42	245,90	8,61	416,43	63,41	12,01	0,72
	Feb	7,37	7,06	165,45	21,76	200,50	7,68	364,63	61,64	11,78	0,97
	Mar	7,01	7,01	241,58	17,19	204,61	7,35	397,59	51,74	10,5	0,57
	Apr	6,79	7,04	220,45	16,87	259,60	10,87	451,30	42,87	10,3	0,48
	May	6,86	7,12	178,98	17,98	254,07	12,97	400,83	37,25	10,5	0,30

4. Problem solution

There are several multivariate statistical methods for the analysis of process. Some of this methods have recently been used successfully for monitoring parameters [3]. These methods are useful because reduces the dimensionality of the original historical data by projecting it onto a lower dimensionality space.

Cluster analysis is a powerful exploratory technique for discovering groups of similar observations within a data set. The idea of cluster analysis is to use values of variables to devise a scheme for grouping objects in such a way that similar objects will belong to the same group (in some sense or another) to each other than to those in other groups.

The association analysis was conducted with 1-r as classification measure, where r represents the Pearson's correlation coefficient [2]. The results are depicted in figure 1.

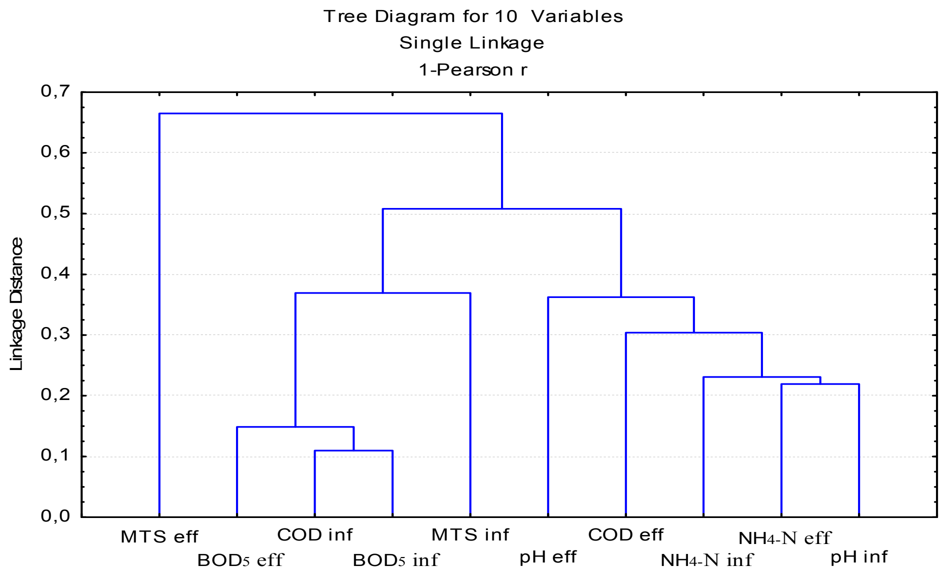

Fig. 1. Single linkage influent and effluent values of the observed variables

The results from figure 1 shows that variables were also included in order to reveal if periodicity is the main factor in the given observable. The strongest association is between BOD_5 (biochemical oxygen demand at 20°C in 5 days) and COD (chemical oxygen demand) values at the effluent of WWTP (being below 0.2 degree of dissimilarity). The next link is at about 0.3 degree of dissimilarity (about 0.5 correlations) and is established between the same observable, but at the influent of the treatment plant. An interesting association is established between the values of CBO_5 and COD at the effluent of the plant and the value of NH_4-N at the influent of the plant, suggesting that somehow the efficiency of the treatment in the WWTP is not affected by the ammonium in treated wastewater.

Total suspended matter at influent (MTS) of WWTP also connects with the cluster formed by BOD_5 influent and COD influent (almost 0.4 degree of dissimilarity), suggesting the influence of season on the quality of inflow. That means that quantity of total suspended matter, which increases along with increase of rain amount, affects oxygen exchange within surface water.

The next figure (fig. 2) shows the explanatory degree of the principal components in the values of the variables depending on the number of the components.

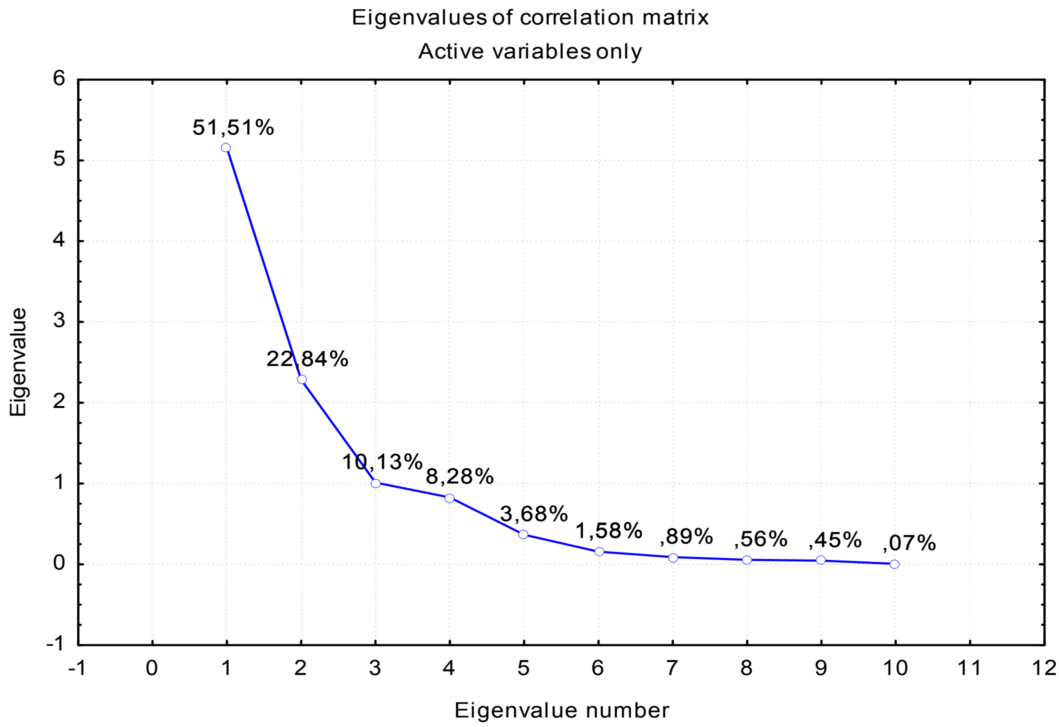

Fig. 2. Explained variance by each consecutive component in principal component analysis

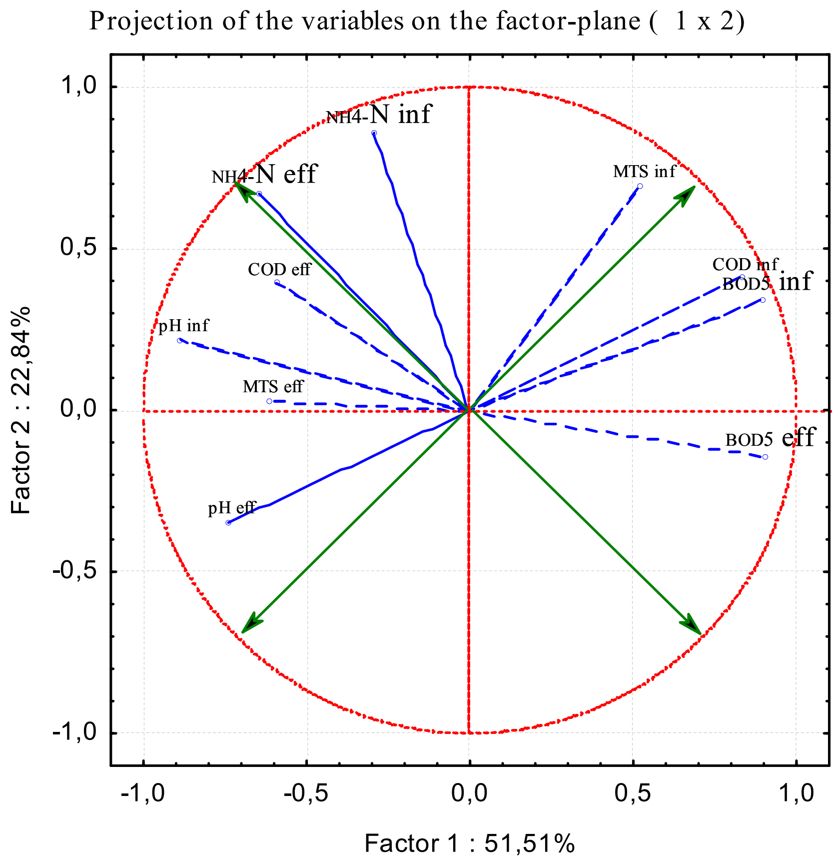

Fig. 3. Projection of the variables in the plane of the first two principal components

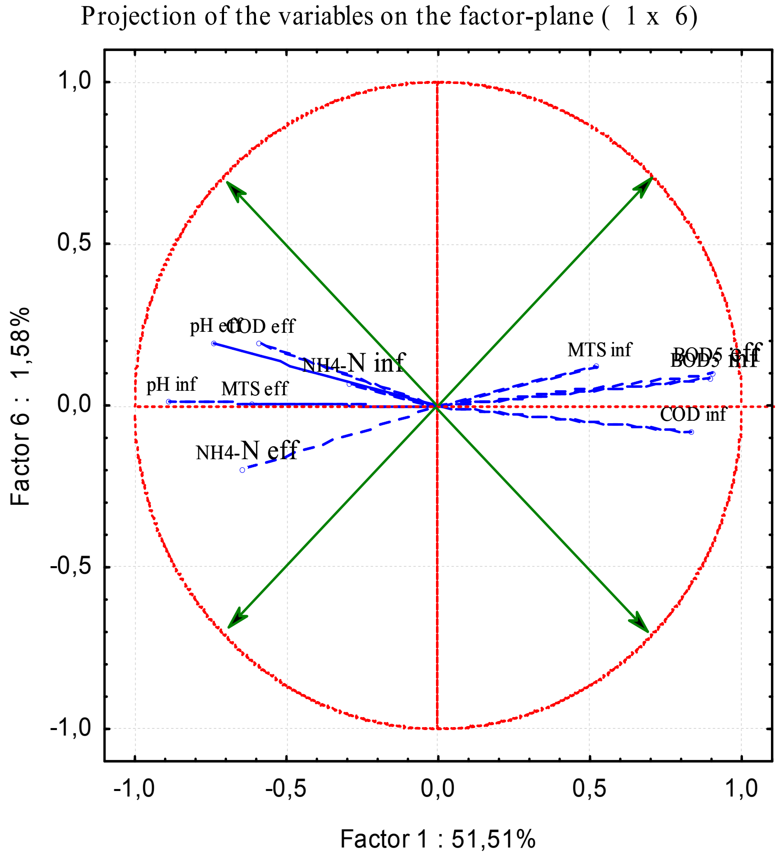

Fig. 4. Projection of the variables in the plane of the first and six principal components

The values from figure 2 reveals that no more than three principal components should be assigned to the deterministic factors in monitored parameters. It is easy to see how the factors starting from factor 3 (with explained variance of 10,13 %) untill factor 10 (with explained variance of 0.07 %) fit to a straight line and their explanatory variance is only an expression of overfitting.

The next figure (fig. 3) depicts the projection of the variables on the plane of the first two principal components (explaining over 60 % of the variance present in the observed values of the active variables).

The analysis of the projections given in figure 3 reveals a strong association of the influent value of the pH at the influent of the WWTP with the BOD_5 factor as well as that both are along the first component of the total variability. This result suggests that the main factor giving variation in the observed variables is the value of pH, and this factor is a slow changing one, varying in approximately good agreement with the BOD_5 of the observation. The analysis of the projections given in figure 4 reveals a strong association of the influent value of the COD at the influent of the WWTP with the ammonium NH_4-N from influent.

6. Conclusions

Statistical methods revealed a strong relationship between chemical and biochemical demand, especially for the effluent of WWTP, which explains why the efficiency of BOD_5 and COD reduction represents a criterion for assessing wastewater treatment degree. These results correlated with the exceeding values reported during the monitored period (May 2013- May 2014) show the influence of ammonium on the efficiency of wastewater treatment process.

Season represents one of the main variation factor, especially concerning the change in suspended matter (MTS) content.

Projection of the variables on the factor plane is able to effluent in detail around which observables gather the identified variation factors.

The obtained statistically significant model allows an analytical expression of intrinsic link between the physical-chemical parameters.

Comparing the values observed in different observation points (influent and effluent) in different times (month, year) are able to highlight on the one hand the extent that the treatment process caused differences in values of environmental parameters controlled and monitored in WWTP and how much systematic changes appeared year to year to influence the water quality if untreated and treated, respectively.

The result of analyzed collected data shows that there is some interrelationship among wastewater parameters. The parameters interrelationship can be assembled into two groups by PCA results. The first group is BOD_5 and COD and the second group is NH_4-N and COD.

Acknowledgements:

This paper is supported by the Sectoral Operational Programme Human Resources Development POSDRU/159/1.5/S/137516 financed from the European Social Fund and by the Romanian Government.

References

[1] Jäntschi, L., S.D. Bolboacă, „The Jungle of Linear Regression Revisited”, Leonardo Electronic Journal of Practices and Technologies, 2007, 6(10):169-187.

[2] Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine 5th Ser 50:157-175.

[3] POP, H.F, EINAX, J.W., Sârbu, C., Classical and Fuzzy Principal Component Analysis of Some Environmental Samples Concerning the Pollution with Heavy Metals, Chemometrics and Intelligent Laboratory Systems, 2009, 97(1), 25-32.

[4] R. R. Bansode, „ Treatment of organic and inorganic pollutants in municipal wastewater by agricultural by-product based granular activated carbons” [M.S. thesis], The Department of Food Science, Louisiana State University and Agricultural and Mechanical College, Baton Rouge, La, USA, 2002.