ELSEVIER

Chemical Engineering Journal 90 (2002) 283–290

Chemical Engineering Journal

www.elsevier.com/locate/cej

Analysis of activated sludge process using multivariate statistical tools—a PCA approach

Rosana K. Tomita*, Song W. Park, Oscar A.Z. Sotomayor

Laboratory of Process Simulation and Control (LSCP), Department of Chemical Engineering, University of São Paulo,
Av. Prof. Luciano Gualberto, 380 trv. 3, 05508-900 São Paulo, SP, Brazil

Abstract

Twelve original physical variables of an activated sludge wastewater treatment system are considered. These cross-correlated variables are transformed in new ones that are not correlated by the use of PCA (principal component analysis), a powerful tool for analysis, monitoring and diagnostics of wastewater treatment processes. Just three principal components explain most of the system total variability (78% of total variance). Thus, the ability to describe the overall characteristics of the process using only three principal components will make the analysis, monitoring and diagnostic of the system easier.

Three groups of variables characterizing the system are detected. The first group identifies variables that represent micro-organisms and inert particulate matter arising from cellular decay, while the second group refers to substrates and flow rate. The third group is related to the pH. Based on these results, the present paper shows how to enlarge the ways of interpreting the characteristics of activated sludge wastewater treatment system.

Keywords: IWA-ASM1; PCA; Multivariate statistical analysis; Wastewater treatment monitoring

1. Introduction

Environmental protection agencies have imposed limits to discharge effluents in body receiver that are more and more strict. Therefore, different approaches of study are challenging objects of research to improve the domestic and industrial effluent treatment process. Our paper presents methodologies of multivariate statistic to improve analysis of operations and involved variables in the wastewater treatment.

Nowadays, increasing degree of industrial automation makes the process monitoring use large quantities of on line data. So, it is important to have the use of efficient techniques to get high quality information from this great number of data. In engineering applications, variables are significantly cross-correlated, made worse by the fact that the data are collected in a noisy environment. In these cases, multivariate analysis has proved to be a powerful statistical technique. Piovoso et al. [1] and Piovoso and Kosanovich [2] present a brief tutorial about applications of multivariate statistics to process analysis, monitoring and control. A matrix approach allied to statistical analysis for multivariate systems is given by Johnson and Wichern [3]. Jackson [4] is another interesting reference about this issue.

Principal component analysis (PCA) utilizes directly the information from the data, compacted in the form of a covariance matrix, to extract more relevant information and to generate new variables named as principal components. This multivariate statistical data has several applications: multilinear regression using principal components (PCR); reduction of number of variables; identification of structures that explain the most relevant variance of the data and for clustering analysis.

The most common application of PCA is statistical process control. Mason et al. [5] present a statistical control chart of PCA including confidence interval based on Hotelling's distribution. Rosén and Olsson [6] show disturbance detection in wastewater treatment system using principal components and partial least-squares (PLS). Raich and Çinar [7] propose to utilize multivariate statistical techniques (PCA and discriminant analysis) to detect states out of control and to diagnose abnormal operation conditions and disturbance that cause poor process behaviour. This methodology was successfully applied in the monitoring of the Tennessee Eastman plant simulation benchmark subjected to different disturbances. Teppola et al. [8] use multivariate statistics to analyse an activated sludge wastewater treatment plant.

---

* Corresponding author. Tel.: +55-11-3091-2237;
fax: +55-11-3813-2380.
E-mail address: rktomita@usp.br (R.K. Tomita).

1385-8947/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S1385-8947(02)00133-X

Nomenclature

b_A	autotrophic decay rate
b_H	heterotrophic decay rate
D	Mahalanobis distance
$\mathbf{D}_x^{1/2}$	standard deviation matrix of **X** data
$\mathbf{e}_k$	eigenvector, *k*th vector of matrix **E**
$\mathbf{E}$	eigenvector matrix
k_a	ammonification rate
k_h	maximum specific hydrolysis rate
K_{NH}	ammonium half-saturation coefficient (hsc) for autotrophs
K_{NO}	nitrate hsc for denitrifying heterotrophs
$K_{O,A}$	oxygen hsc for autotrophs
$K_{O,H}$	oxygen hsc for heterotrophs
K_S	hsc for heterotrophs
K_X	hsc for hydrolysis of slowly biodegradable substrate
n	number of samplings
p	number of variables
$\mathbf{p}_{c_k}$	*k*th vector of principal components
$\mathbf{P}_c$	matrix of principal components
$s_{x_j x_k}$	covariance between variables x_j and x_k
S_{ALK}	alkalinity
S_I	soluble inert organic matter
S_{ND}	Soluble organic nitrogen
S_{NH}	soluble ammonium nitrogen
S_{NO}	soluble nitrate–nitrite nitrogen
S_O	dissolved oxygen concentration
S_S	soluble biodegradable substrate
$\mathbf{S}_x$	covariance matrix of **X** data
$\mathbf{x}$	row vector of variables
x_{jk}	*j*th observation of *k*th variable
$x_{\mathrm{orig}_{jk}}$	original *j*th observation of *k*th variable (without normalization)
$\bar{\mathbf{x}}$	row vector of sample mean
$\bar{x}_k$	mean of *k*th variable
$\mathbf{X}$	data matrix
$X_{B,A}$	active autotrophic biomass
$X_{B,H}$	active heterotrophic biomass
X_I	particulate inert organic matter
X_{ND}	particulate organic nitrogen
$\mathbf{X}_{\mathrm{orig}}$	original data matrix (without normalization)
X_P	particulate products from biomass decay
X_S	particulate substrate biodegradable
T_α^2	Hotelling’s distribution
Greek symbols	
α	significance level
η_g	correction factor for anoxic growth of heterotrophs
η_h	correction factor for anoxic hydrolysis
λ_k	eigenvalue
Λ	eigenvalue matrix
$\hat{\mu}_A$	autotrophic maximum specific growth rate
$\hat{\mu}_H$	heterotrophic maximum specific growth rate
μ_k	percentual variance of *k*th principal component

The main contribution of this paper is to present the method to extract information for easy interpretation, disturbance or deviation detection of normal condition in measurements of wastewater treatment process. Further, this method should suggest the variables responsible for the deviations.

2. Activated sludge wastewater treatment

The present paper utilizes a dynamic model of activated sludge process which is known as IWA activated sludge model no. 1 or ASM1 [9]. This model includes the carbon oxidation both by heterotrophic and autotrophic biomass (organic matter consumption); nitrification (ammonia is oxidized to nitrite and subsequently to nitrate) and denitrification (nitrate is transformed to nitrogen gas).

This model is formed by 13 state variables and by 19 parameters that describe eight biological processes. The involved variables are concentrations of: alkalinity (S_{ALK}), soluble biodegradable organic nitrogen (S_{ND}), ammonia nitrogen (S_{NH}), nitrate (S_{NO}), dissolved oxygen (S_O), readily biodegradable substrate (S_S), active autotrophic biomass ($X_{B,A}$), active heterotrophic biomass ($X_{B,H}$), particulate biodegradable organic nitrogen (X_{ND}), particulate products from biomass decay (X_P), slowly biodegradable substrate (X_S), particulate and soluble inert organic matter (X_I and S_I). These two last ones are not relevant to the present paper, so 11 concentrations and effluent flow rate form the 12 variables analysed here. We are going to review the reaction rates that describe the biological process incorporated in the ASM1 model, because they are important to get a good interpretation of the process (Table 1).

Each state equation has the following form:

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = \frac{q_{in}}{V} c_{i,in} - \frac{q_{out}}{V} c_i + \sum_{j=1}^{8} a_{i,j} r_j \tag{9}$$

where c_i is concentration of state variable i, $c_{i,in}$ the concentration of variable i in the inflow, q_{in} the inflow, q_{out} the outflow, V the bioreactor volume, $a_{i,j}$ the biological process constant coefficient and r_j the reaction rate of biological process.

The relation among the variables of biological processes is schematically represented in Fig. 1. This figure shows, e.g., that concentration of active heterotrophic biomass ($X_{B,H}$) is affected by the processes r_1, r_2 and r_4. In the aerobic growth of $X_{B,H}$ (process r_1), we could see that it is influenced by variables S_S, S_{NH} and S_O. In the same way, $X_{B,H}$ is affected by S_S, S_{NH} and S_{NO} in the anoxic growth (process r_2). When $X_{B,H}$ dies it is converted to X_S and X_{ND} by process r_4.

Table 1
Reaction rates of biological process incorporated in the ASM1 model

Biological process	Kinetic equation	
Aerobic growth of heterotrophs ($X_{\mathrm{B,H}}$)	$r_1=\hat{\mu}_{\mathrm{H}}\left(\frac{S_{\mathrm{S}}}{K_{\mathrm{S}}+S_{\mathrm{S}}}\right)\left(\frac{S_{\mathrm{O}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)X_{\mathrm{B,H}}$	(1)
Anoxic growth of heterotrophs ($X_{\mathrm{B,H}}$)	$r_2=\hat{\mu}_{\mathrm{H}}\left(\frac{S_{\mathrm{S}}}{K_{\mathrm{S}}+S_{\mathrm{S}}}\right)\left(\frac{K_{\mathrm{O,H}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)\left(\frac{S_{\mathrm{NO}}}{K_{\mathrm{NO}}+S_{\mathrm{NO}}}\right)\eta_{\mathrm{g}}X_{\mathrm{B,H}}$	(2)
Aerobic growth of autotrophs ($X_{\mathrm{B,A}}$)	$r_3=\hat{\mu}_{\mathrm{A}}\left(\frac{S_{\mathrm{NH}}}{K_{\mathrm{NH}}+S_{\mathrm{NH}}}\right)\left(\frac{S_{\mathrm{O}}}{K_{\mathrm{O,A}}+S_{\mathrm{O}}}\right)X_{\mathrm{B,A}}$	(3)
Decay of heterotrophs ($X_{\mathrm{B,H}}$)	$r_4=b_{\mathrm{H}}X_{\mathrm{B,H}}$	(4)
Decay of autotrophs ($X_{\mathrm{B,A}}$)	$r_5=b_{\mathrm{A}}X_{\mathrm{B,A}}$	(5)
Ammonification of soluble organic nitrogen (S_{ND})	$r_6=k_{\mathrm{a}}S_{\mathrm{ND}}X_{\mathrm{B,H}}$	(6)
Hydrolysis of entrapped organics (X_{S})	$r_7=k_{\mathrm{h}}\frac{X_{\mathrm{S}}/X_{\mathrm{B,H}}}{K_{\mathrm{X}}+(X_{\mathrm{S}}/X_{\mathrm{B,H}})}\left[\left(\frac{S_{\mathrm{O}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)+\eta_{\mathrm{h}}\left(\frac{K_{\mathrm{O,H}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)\left(\frac{S_{\mathrm{NO}}}{K_{\mathrm{NO}}+S_{\mathrm{NO}}}\right)\right]X_{\mathrm{B,H}}$	(7)
Hydrolysis of entrapped organic nitrogen (X_{ND})	$r_8=k_{\mathrm{h}}\frac{X_{\mathrm{ND}}/X_{\mathrm{B,H}}}{K_{\mathrm{X}}+(X_{\mathrm{S}}/X_{\mathrm{B,H}})}\left[\left(\frac{S_{\mathrm{O}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)+\eta_{\mathrm{h}}\left(\frac{K_{\mathrm{O,H}}}{K_{\mathrm{O,H}}+S_{\mathrm{O}}}\right)\left(\frac{S_{\mathrm{NO}}}{K_{\mathrm{NO}}+S_{\mathrm{NO}}}\right)\right]X_{\mathrm{B,H}}$	(8)

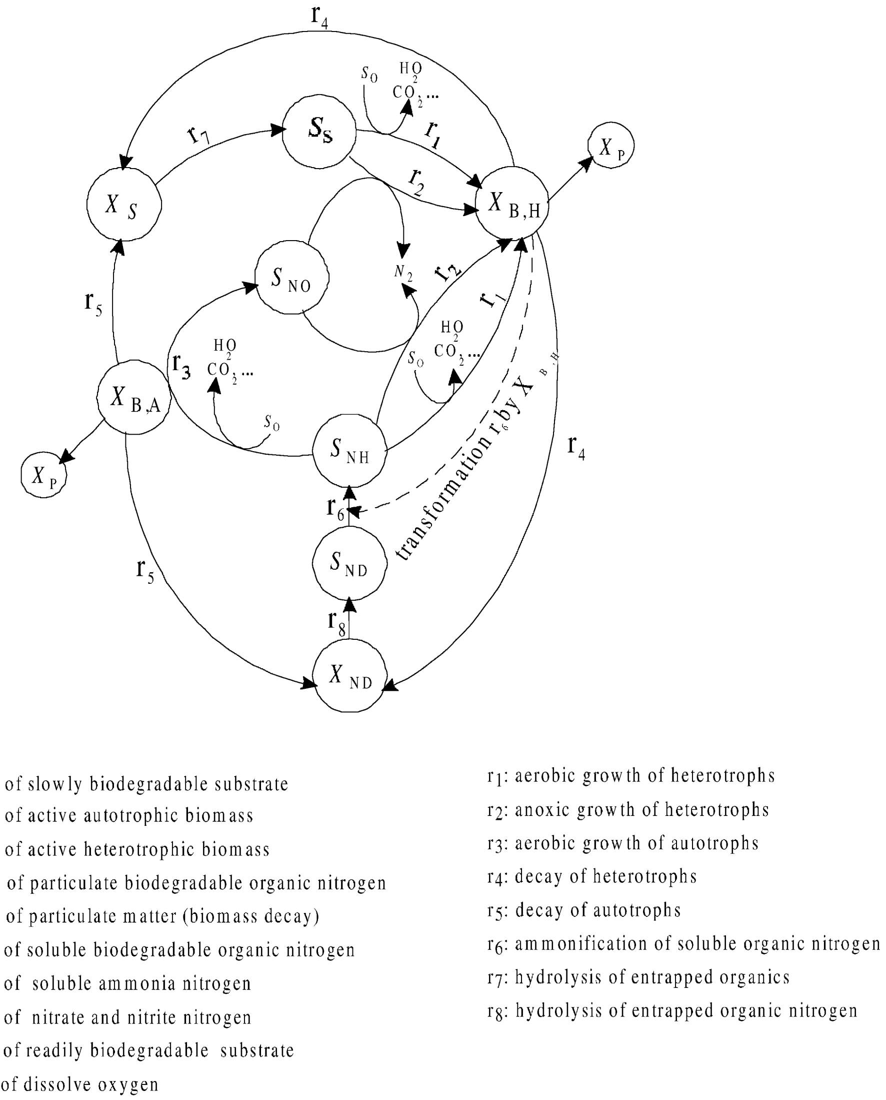

Fig. 1. Representation of biological processes of an activated sludge process.

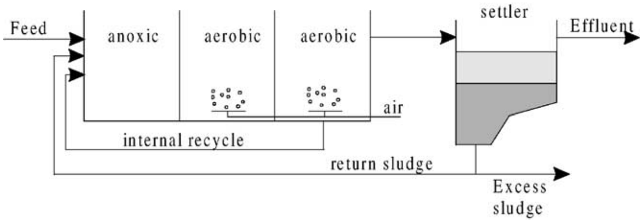

Fig. 2. Activated sludge process with pre-denitrification.

The model ASM1 has an extended literature, so the benchmark of Sotomayor et al. [10] will be utilized here. The simulated system is shown in Fig. 2. This system is composed by three biological treatment zones followed by a settler. The first zone of biological treatment is anoxic, where the pre-denitrification occurs. The remaining two zones are completely mixed aerobic reactors with dissolved oxygen control. The process includes a sludge recycling from the settler to the anoxic reactor to increase the biological mass as well as an internal recirculation from the last aerobic zone to the anoxic zone. This recirculation is due to pre-denitrification.

The operational variables and influent wastewater characteristics for simulation are given in Table 2. The model parameters are not presented here but can be found in [10].

Using this dynamic model the results were obtained in "steady state". In other words, the process was run for 6 h before any data of reactor effluent was taken for each new operating condition. The 12 variables of the influent current are defined by random choice between a pre-defined range for each variable. The influent flow rate and concentrations suffered the disturbance or new operating conditions. To illustrate, Fig. 3 shows a disturbance of the influent flow rate every 6 h, summing 601 sampling points. The pseudo-steady state condition results are equivalent to noisy data collection.

Table 2
Operational variables and influent wastewater characteristics

	Value	Unit
Operational variable		
Influent flow rate, Q_{in}	1080 to 1620	m^3/h
Recycle flow rate	0.8 (of 1080)	–
Internal recycle flow rate	2 (of 1080)	–
Anoxic tank volume	1000	m^3
First aerobic tank volume	3700	m^3
Second aerobic tank volume	3700	m^3
Settler volume	5000	m^3
Air flow rate to 1° aerobic reactor	10.10525	m^3/h
Air flow rate to 2° aerobic reactor	6.6998	m^3/h
Set-point of S_O	2	$g\,O_2/m^3$
Influent wastewater characteristics		
$X_{B,H}$	90	g/m^3
$X_{B,A}$	1	g/m^3
X_S	210, 315	g/m^3
X_{ND}	10	g/m^3
X_P	110, 165	g/m^3
S_O	0	g/m^3
S_{NH}	12.5, 18.75	g/m^3
S_{ND}	10.1	g/m^3
S_{NO}	0.6	g/m^3
S_S	130, 195	g/m^3
S_{ALK}	7	g/m^3

3. Multivariate analysis of activated sludge process

3.1. Data pre-processing

Often the wastewater treatment process variables are measured in different scales or units. Then, for the PCA analysis, first of all 600 data of the 12 variables had to be autoscaled, in other words, the variables were mean centred, by subtracting the mean ($\bar{x}_k$) and scaled to unit variance, by dividing the standard deviation:

$$x_{jk} = \frac{x_{\mathrm{orig}_{jk}} - \bar{x}_k}{\sqrt{s_{x_j x_k}}}, \qquad j = 1, 2, \dots, n, \ k = 1, 2, \dots, p \tag{10}$$

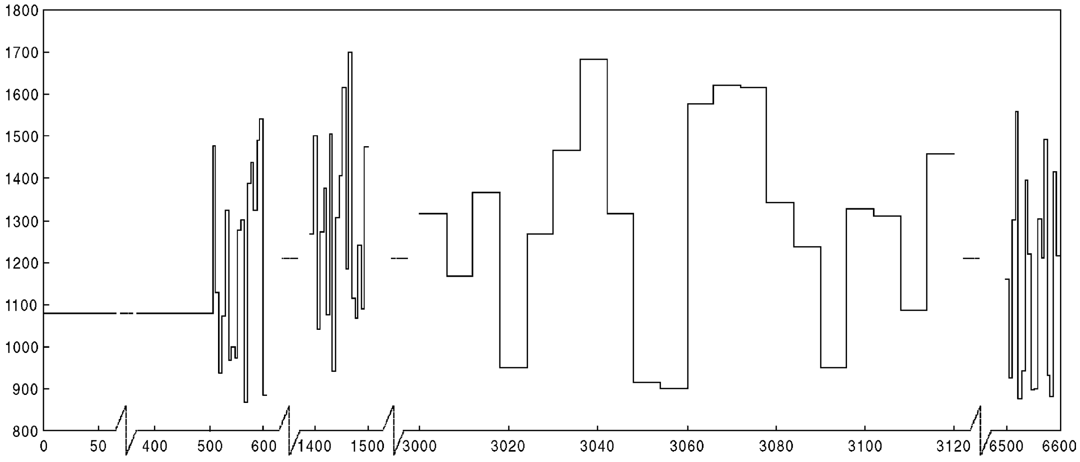

Fig. 3. Example of process disturbance of the influent flow rate.

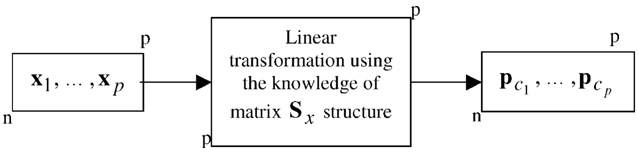

Fig. 4. Transformation of original variables to principal component variables.

In matrix notation, it is

$$\mathbf{X} = (\mathbf{X}_{\mathrm{orig}} - \mathbf{1}_{(n\times 1)}\bar{\mathbf{x}})\mathbf{D}_x^{-1/2} \tag{11}$$

where $\mathbf{D}_x^{-1/2}$ is the matrix of standard deviation

$$\mathbf{D}_x^{-1/2} = \begin{bmatrix} \sqrt{s_{x_1x_1}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{x_2x_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{x_px_p}} \end{bmatrix} \tag{12}$$

The normalized data matrix $\mathbf{X}$ has mean zero and its covariance matrix $(\mathbf{X}'\mathbf{X})$ indicates the directions in the X space where most of the variability occurs [2]. The principal components are obtained from this covariance matrix.

3.2. Principal component analysis

Summing up, the objective of PCA is to make easy interpretation of the measured data by dimensionality reduction. PCA tries to explain the covariance data through linear combinations of variables. In other words, it could be seen as a linear transformation (Fig. 4). In this transformation the data variability is preserved.

The symmetric covariance matrix $\mathbf{S}_x$ could be decomposed into a diagonal matrix Λ and orthogonal matrix $\mathbf{E}$

$$\mathbf{S}_x = \mathbf{E}\Lambda\mathbf{E}' \tag{13}$$

The diagonal elements of Λ, $\lambda_1, \lambda_2, \dots, \lambda_p$, are the eigenvalues of $\mathbf{S}_x$ and columns of $\mathbf{E}$, $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p$, are their correspondent eigenvectors. In PCA, $\mathbf{e}$ is named as loading vector.

These loadings can be visualized in graphic form, as is shown in Fig. 5. This figure is useful for grouping analysis in terms of principal components to distinguish a group of variables. Analysing it we conclude that alkalinity and ammonia nitrogen do not contribute to the total variation in the first principal component, because these variables are placed near the origin of this principal component. However, the influent flow rate (Q_{in}) and the concentration of soluble organic nitrogen (S_{ND}) are important for this component.

In group I are found variables that represent autotrophic ($X_{\mathrm{B,A}}$) and heterotrophic ($X_{\mathrm{B,H}}$) biomass and particulate matter (X_{P}). The concentration of rapidly and slowly biodegradable nitrogen and carbon substrate (S_{ND}, X_{ND}, S_{S} and X_{S}), as well as influent flow rate (Q_{in}) form group II. In the $p_{c_2} \times p_{c_1}$ plane, the concentration of ammonia nitrogen (S_{NH}) is not in the same group of other substrates, despite being the direct source of nitrogen for the micro-organisms. Probably, ammonia nitrogen and alkalinity (S_{ALK}) belong to the same group III, as shown in Fig. 5, because both are variables that characterize the pH of the medium. Both oxygen (S_{O}) and nitrate (S_{NO}) are the terminal electron acceptors of cellular metabolism and these variables are not considered in the groups. A possible explanation for the oxygen to be near the substrates is that the consumption of substrates is greater in aerobic medium than in an anoxic one.

The first principal component is a linear combination that defines the direction of greatest variability

$$\mathbf{p}_{c_1} = \mathbf{X}\mathbf{e}_1 \tag{14}$$

$\mathbf{p}_{c_1}$ is denoted as score vector.

The second principal component $\mathbf{p}_{c_2}$ is orthogonal to the first one and it explain the remaining most variability (in the

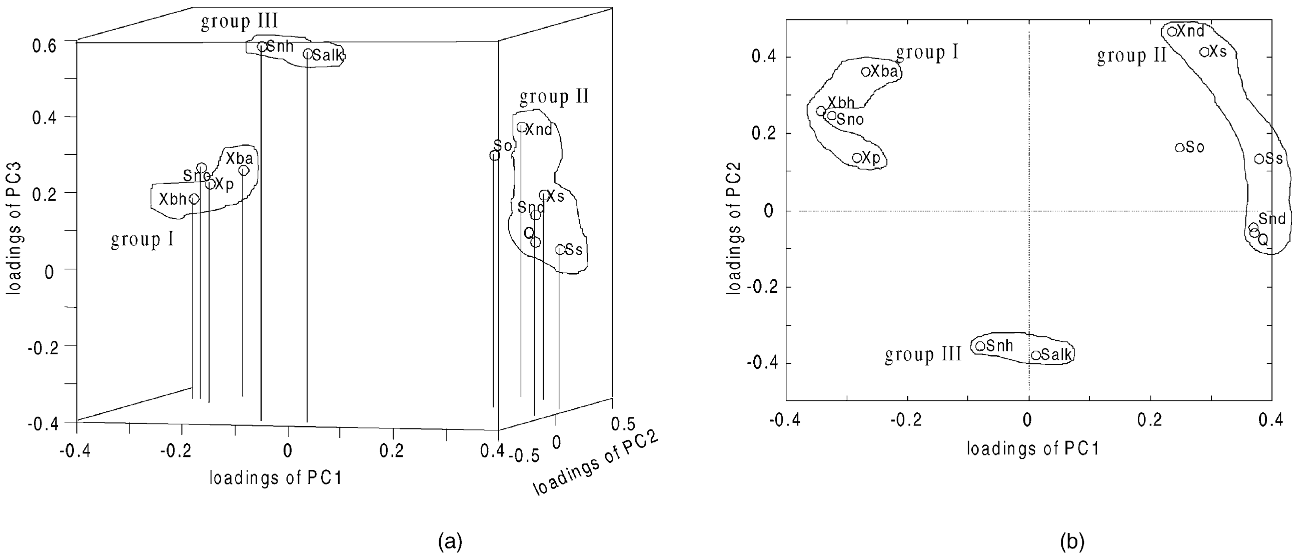

Fig. 5. Grouping analysis in terms of principal components.

direction of $\mathbf{X}-\mathbf{p}_{c_1}\mathbf{e}_1'$). The next one, $\mathbf{p}_{c_3}$, is orthogonal to $\mathbf{p}_{c_1}$ and $\mathbf{p}_{c_2}$, it explains the bit little remain variability, and so on until $\mathbf{p}_{c_p}$. Then,

$$\mathbf{P}_c=\mathbf{XE}=(\mathbf{X}_{\mathrm{orig}}-\mathbf{1}_{(n\times 1)}\bar{\mathbf{x}})\mathbf{D}_x^{-1/2}\mathbf{E} \qquad (15)$$

The scores $\mathbf{p}_{c_1}, \mathbf{p}_{c_2}, \dots, \mathbf{p}_{c_p}$ have variance $\lambda_1, \lambda_2, \dots, \lambda_p$, respectively, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

3.3. Cumulative percentage

Often it is more convenient to express each variance in relation to total variance

$$\mu_k=\frac{\lambda_k}{\lambda_1+\lambda_2+\dots+\lambda_p} \qquad (16)$$

Thus, if the sum of first (i) values of μ, $\mu_1+\dots+\mu_i$, represents, e.g., 80% of total variance, then first (i) principal components could represent all p original variables without loss of information.

The cumulative percentage variance is illustrated in Fig. 6. The first principal component represents 52% of total variance whereas the two first components explain 67% and the three first ones are responsible for 78%. In this way, 12 variables could be represented by three principal components.

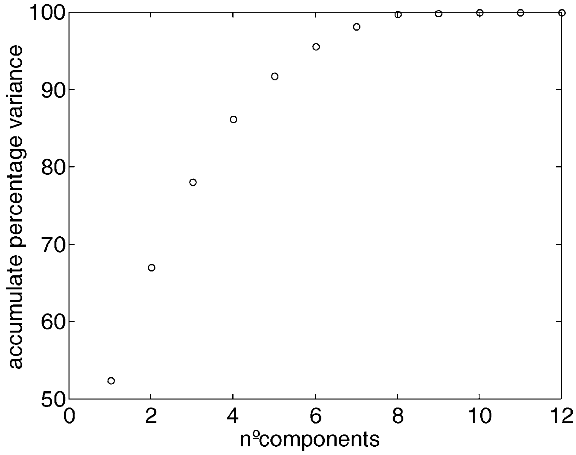

Fig. 6. Percentage variance of principal components.

3.4. Geometrical interpretation of principal components

Principal components represent a new coordinate system that is obtained by rotation of original one. The rotation angles in relation to original coordinate are given by eigenvectors of covariance matrix. The new axes represent the direction of maximum variability.

The PCA decomposition is geometrically interpreted as density constant contours—ellipsoids. These ellipsoids are

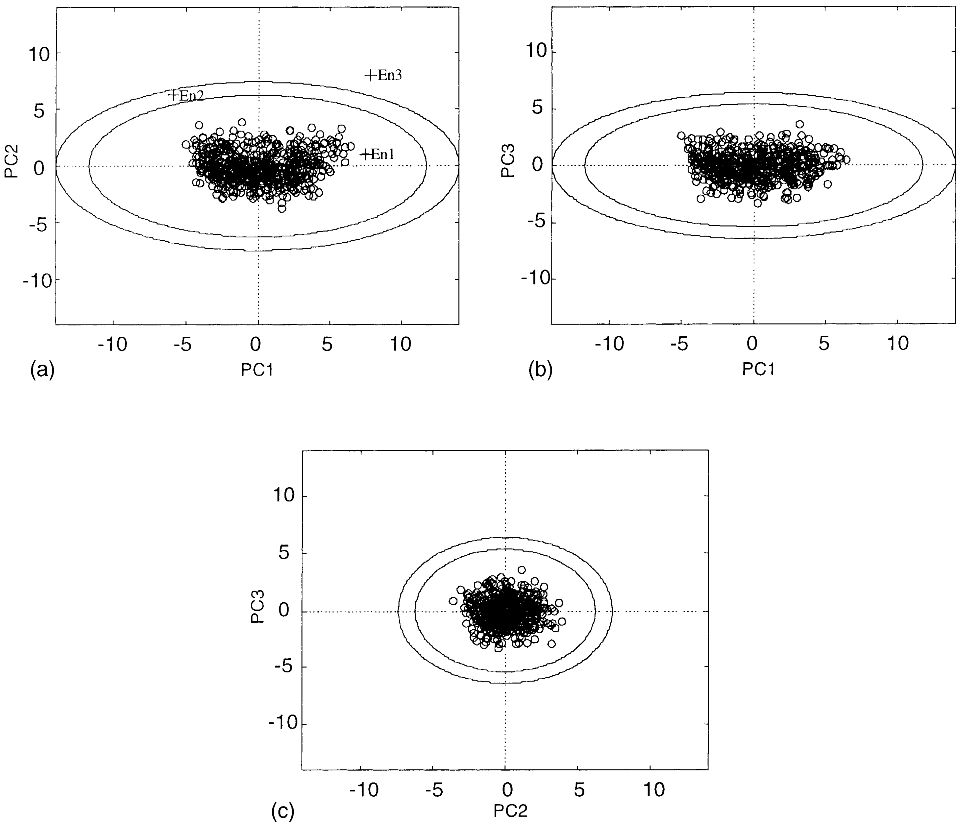

Fig. 7. Control ellipsoids: (a) $\mathbf{P}_{c_1}$ and $\mathbf{P}_{c_2}$; (b) $\mathbf{P}_{c_1}$ and $\mathbf{P}_{c_3}$; (c) $\mathbf{P}_{c_2}$ and $\mathbf{P}_{c_3}$.

built from the concept of Mahalanobis distance between the $\mathbf{x}$ vectors ($\mathbf{x} = [\, x_1 \quad x_2 \quad \dots \quad x_p \,]$) and the mean values $\bar{\mathbf{x}}$ ($\bar{\mathbf{x}} = [\, \bar{x}_1 \quad \bar{x}_2 \quad \dots \quad \bar{x}_p \,]$)

$$(\mathbf{x} - \bar{\mathbf{x}})\boldsymbol{\Lambda}^{-1}(\mathbf{x} - \bar{\mathbf{x}})' = D \tag{17}$$

The probability that a measurement is less than the Mahalanobis distance is given by Hotelling's T^2 distribution

$$(\mathbf{x} - \bar{\mathbf{x}})\boldsymbol{\Lambda}^{-1}(\mathbf{x} - \bar{\mathbf{x}})' \le T^2 \tag{18}$$

The $\mathbf{x}$ variable could be expressed in terms of principal component coordinates. The control ellipse for components p_{c_1} and p_{c_2}, e.g., is obtained from Eq. (17), which can be rewritten as

$$\frac{p_{c_1}^2}{\lambda_1} + \frac{p_{c_2}^2}{\lambda_2} = T_\alpha^2 \tag{19}$$

Fig. 7 shows the points corresponding to principal components calculated by Eq. (15). This figure confirms that the points' dispersion is greater in the direction of first principal components, that is, the component of greatest variance.

One can consider that the data utilized was obtained from normal operating conditions, since the disturbances induced in the system are within the upper and lower limits of operation. This means that the data can be used as calibration data to build a control chart.

The ellipse's major and minor half axes, hereby called radii, are described by the equation: radius $= \sqrt{T_\alpha^2 \lambda_i}$. The internal radius $\sqrt{T_{0,9545}^2 \lambda_i}$ represents the limit for a 95.45% significance level and the external radius $\sqrt{T_{0.9973}^2 \lambda_i}$, a 99.73% confidence interval. In Fig. 7 the ellipses are built with these significance levels. Even so the ellipses shapes are not the same, because they are proportional to the square root of the eigenvalue corresponding to the principal component.

Once the control chart is built, it is used as a reference to check whether or not the process is in control. Suppose, e.g., in Fig. 7(a) three tests called En1, En2 and En3. The test En1 is within the internal limits, so it can be considered a process in normal operating conditions. Contrary to En1, the tests En2 and En3 that are, respectively, within the internal and external limits and beyond the external limits, are processes that present some out of control condition. Control charts using the Mahalanobis distance concept were shown in [11].

The coordinates of the points presented in Fig. 5, that is eigenvalues or loadings, can be shown as bar charts in Fig. 8.

The principal components, as opposed to the physical variables, have no physical meaning, but can be interpreted as a linear combination of these physical variables. The first component illustrates the contribution of all variables, except S_{NH} and S_{ALK}, as an overall level of pollutants in the

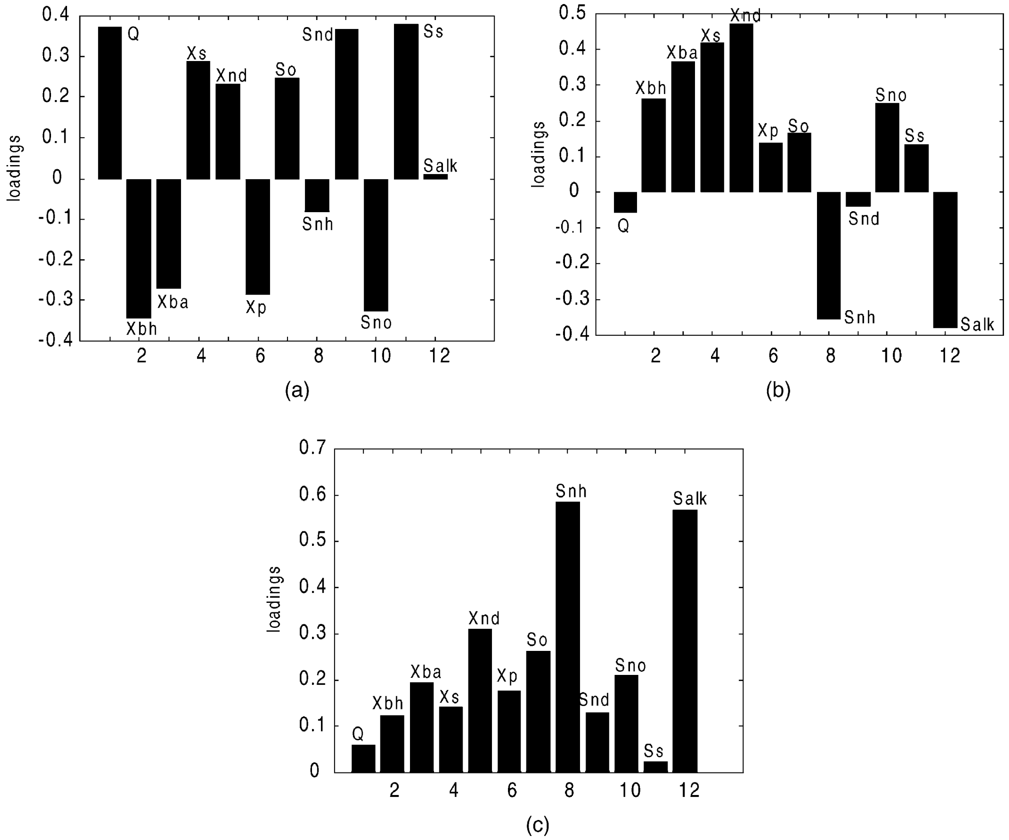

Fig. 8. Loadings for the: (a) first; (b) second; (c) third principal components.

effluent. There is a group mainly made up of substrates which positively influences this component, and another of micro-organisms, particulate matter and nitrate/nitrite that influences it negatively. So, e.g., if a sample point is outside the control window, but is inside the chart for the second and third components, there is strong evidence that the problem is in the flow variable Q_{in}.

From Fig. 8(b), one can conclude that flow rate and soluble organic nitrogen are less important variables for the second principal component. Ammonia nitrogen and alkalinity have a significant negative influence in this component. On the other hand, the contribution of these variables in the third component is positive, as shown in Fig. 8(c). From this figure, one can conclude that the third component is related to the pH of the medium.

4. Conclusions

This work has shown that multivariate analysis of data is an adequate tool for obtaining important process information.

It has shown the possibility of reducing the analysis from 12 process variables down to three principal components which are more relevant to the system. From these principal components the qualitative influence of the variables has been analysed by grouping in the loadings' graphs. The loadings were used to show how the influence of each variable on the principal components can be examined.

All of these aspects enlarge our process know-how and will be used as an important step in a future work involving statistical process control in a wastewater treatment system.

References

[1] M.J. Piovoso, K.A. Kosanovich, J.P. Yuk, Process data chemometrics, IEEE Trans. Instrum. Measur. 41 (2) (1992) 262–268.

[2] M.J. Piovoso, K.A. Kosanovich, The use of multivariate statistics in process control, in: W.S. Levine (Ed.), The Control Handbook, CRC Press, Boca Raton, FL, 1996, pp. 561–573.

[3] R.A. Johnson, D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1998.

[4] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.

[5] R.L. Mason, N.D. Tracy, J.C. Young, A practical approach for interpreting multivariate T^2 control chart signal, J. Qual. Technol. 29 (4) (1997) 396–406.

[6] C. Rosén, G. Olsson, Disturbance detection in wastewater treatment plants, Water Sci. Technol. 37 (12) (1998) 197–205.

[7] A. Raich, A. Çinar, Statistical process monitoring and disturbance diagnosis in multivariable continuous process, AIChE J. 42 (4) (1996) 995–1009.

[8] P. Teppola, S.P. Mujunen, P. Minkkinen, Partial least squares modeling of an activated sludge plant: a case study, Chemom. Intell. Lab. Syst. 38 (2) (1997) 197–208.

[9] M. Henze, C.P.L. Grady Jr., W. Gujer, G.V.R. Marais, T. Matsuo, Activated sludge model No. 1, Report STR1, IAWPRC, London, 1987.

[10] O.A.Z. Sotomayor, S.W. Park, C. Garcia, A simulation benchmark to evaluate the performance of advanced control techniques in biological wastewater treatment plants, Braz. J. Chem. Eng. 18 (1) (2001) 81–101.

[11] R.K. Tomita, Statistical monitoring of wastewater treatment systems, M.Sc. Dissertation, Department of Chemical Engineering, Polytechnic School of University of São Paulo, São Paulo, Brazil, 2001 (in Portuguese).