MULTIVARIATE STATISTICAL PROCESS CONTROL FOR SITUATION ASSESSMENT OF A SEQUENCING BATCH REACTOR

M. Ruiz*, J. Colomer*, J. Colprim† and J. Meléndez*

*Control Engineering and Intelligent Systems (eXiT), University of Girona, Spain, e-mail : (mlruizo,colomer,quimmel)@eia.udg.es
† Laboratory of Chemical and Environmental Engineering (LEQUIA), University of Girona, Spain, e-mail: jesus@lequia.udg.es

Keywords: SBR process, MPCA, situation assessment, fuzzy tool, classification.

Abstract

In this work, a combination between Multivariate Statistical Process Control (MSPC) and an automatic classification algorithm is developed to application in Waste Water Treatment Plant. Multiway Principal Component Analysis is used as MSPC method. The goal is to create a model that describes the batch direction and helps to fix the limits used to determine abnormal situations. Then, an automatic classification algorithm is used to situation assessment of the process.

1 Introduction

The treatment of wastewater has become one of the most important environmental topics. Wastewater treatment is an important part of maintaining the highest possible quality of natural water resources. With the new regulations for quality monitoring of Wastewater Treatment Plants (WWTP), directive 91/271/CEE [8], it is necessary to introduce new technology for control and supervision. In recent years, major advances in wastewater treatment using activated sludge have been developed. This work has been developed using a Sequencing Batch Reactor (SBR) pilot plant shown in Figure 1. The SBR has a cyclic nature, each cycle consist of several stages depending on the objective of operations. This process is highly nonlinear, time-varying and subjected to significant disturbances such equipment defects, atmospheric changes (rain), variation in the composition, among others.

In this work, a Multivariable Statistical Process Control MSPC) is used to determine abnormal situations, because the statistical model considers the correlation structure between variables [4, 5]. MSPC compresses the multidimensional information in few latent variables that explain the variability of the measured variables, including their relationships. In this paper, a model is developed using an extension of the Principal Component Analysis method to detect abnormal batch behaviours. This extension is based on Multiway Principal Component Analysis (MPCA) for batch data [9,14]. An automatic fuzzy classification algorithm is used after MSPC, allowing a good situation assessment.

This paper is organized as follows: Operation of the SBR process is presented in section 2. The section 3 describes those MPCA methods for process monitoring. In section 4, the fuzzy classification method is presented. In section 5 a numerical example is presented by using data recorded from pilot plant SBR during a period of 4 months. Finally, future work and conclusions are summarised in section 6 and 7.

2 SBR pilot plant: influent composition and operational conditions

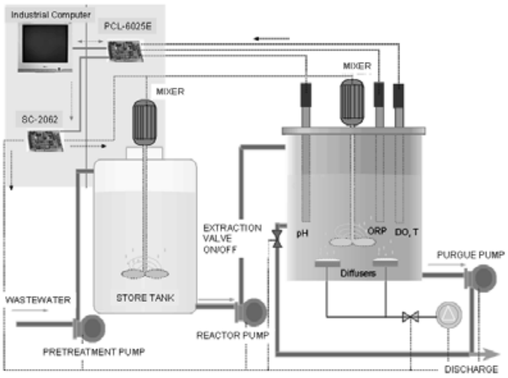


Figure 1: Schematic overview of the SBR Pilot Plant. Data acquisition and control software was responsible for the operation of peristaltic pumps, reactor mixing and air supply control; as well as for the on-line monitoring of reactor pH, ORP, DO and temperature.

SBR Pilot plant is composed of a metal square reactor of 1m3. The minimum and maximum volumes of the reactor are 483 litres and 683 litres, respectively and it defines 200 litres of water to process. Wastewater is taken directly from the Cassà-WWTP (Spain) by means of a peristaltic pump and it is stored in a storage tank. Next, the wastewater is pumped to the reactor by means of another peristaltic pump (Figure 1). The SBR pilot plant carries out advanced treatment, in which the nitrogen is removed. Nitrogen removal in the SBR pilot plant is a two steps process: Nitrification: Ammonia is converted to nitrate by aerobic micro organisms. Denitrification: Nitrate is converted to N_2O or nitrogen gas under anoxic (without oxygen) conditions by anoxic micro organisms. Until now, the duration of operation stages are fixed. Each batch take 8 hours and has 392 samples (obtained every 60 seconds) per variable: pH, Oxidation Reduction Potential (ORP), Dissolved Oxygen (DO) and Temperature. Each cycle is based on alternating anoxic and aerobic reaction. The anoxic period is longer than the aerobic because of increasing denitrification

This work has been developed on the basis of the research project "*Development of a system of control and supervision applied to a Sequencing Batch Reactor (SBR) for the elimination of organic matter, nitrogen and phosphorus DPI2002-04579-C02-01*".

3 MSPC in batch processes

Consider a typical batch run in which j = 1, 2, ..., J variables are measured at k = 1, 2, ...,K time intervals throughout the batch. Similar data will exist on a number of such batch runs i = 1, 2, ..., I [9]. All the data can be summarized in the **X̲** (I x J x K) see figure 2 and 3.

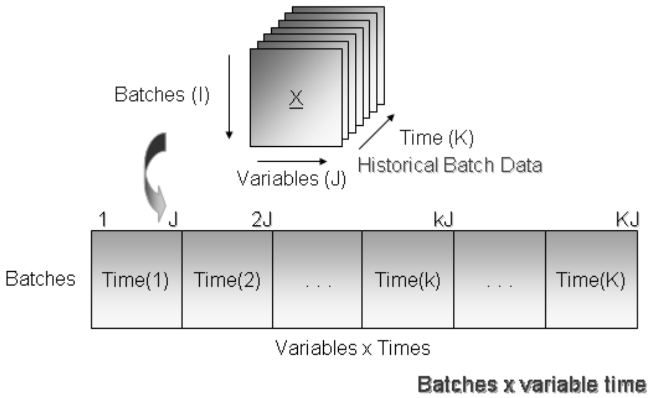


Figure 2: Decomposition of X̲ in the batch direction (*I* x *KJ*).

MPCA is equivalent to performing ordinary PCA on a large two-dimensional (2 − D) matrix constructed by unfolding the three-way. The three-way array **X̲** can be unfolded between the rearranged data matrix in the batch direction (I x KJ) (figure 2) and in direction of variable (IK x J) (figure 3)[5].

The objective of MPCA is to decompose the three-way **X̲**, into a large two-dimensional matrix **X**. It accomplishes this decomposition in accordance with the principle of PCA and separates the data in an optimal way into two parts. The noise or residual part (E), which is as small as possible in a least squares sense, and the systematic part (PR r=1 tr N Pr), which expresses it as one faction (t) related only to batches and a second fraction (P) related to variables and their time variation [9]. The MPCA algorithm derives directly from the NIPALS algorithm and as a result the matrix **X̲**. It is the product of score vector t_r and loading matrices P_r, plus a residual matrix E, that is minimized in a least-squares sense as

$$\underline{X} = \sum_{r=1}^{R} t_r \otimes P_r \qquad (1)$$

$$X = \sum_{r=1}^{R} t_r P_r^T + E = TP + E = \hat{X} + E \qquad (2)$$

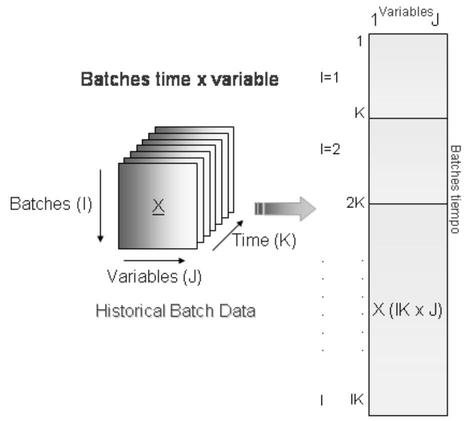


Figure 3: Decomposition of X̲ in the variable direction (*IK* x *J*).

MPCA decomposes the three-way array where ⊗ denotes the Kronecker product (**X̲** = t⊗P is **X̲**(i, j, k) = t(i)P(j, k)) and R denotes the number of principal components retained. The equation (1) is the 3-D decomposition while the equation (2) displays the more common 2-D decomposition [13].

The loading vectors (p_r) define the reduced dimension space (R) and are the directions of maximum variability. Each element of the score vectors (t_r) corresponds to a single batch and depicts the overall variability of this batch with respect to the other batches in the database throughout the entire batch duration [9]. A few principal component can express most of the variability in the data when there is a high degree of correlation among the data (R << min(I, JK))). R is chosen such that most of the systematic variability of the process data is described by these principal components and that the residual matrix, E, is a small as possible in a least-squares sense. When process measurements array X̲ is unfolded to preserving the batch direction and variable direction there are combination of slices of matrices of different sizes (Table 1):

In batch direction	In variable direction
X̲ (I x J x K)	X̲ (I x J x K)
X (I x KJ)	X (IK x J)
T (I x R)	T (IK x R)
P (KJ x R)	P (J x R)
E (I x KJ)	E (IK x J)

Table 1: Dimensions of matrices

Abnormal behaviour of batch can be identified by projecting the batch onto the model. Control charts that are used in monitoring batch processes are generally based on the the *Q*-statistic and *D*-statistic, in which control limits are used to determine whether the process is in control or not. The assumption behind these approximate confidence limits is that underlying process exhibits a multivariate normal distribution with a population mean of zero. This is to be expected, because any linear combination of random variables, according to the central limit theorem, should toward a normal distribution.

The *Q*-statistic is a measure of the lack of fit with the established model. For batch number i, Q_i is calculated as:

$$Q_i = \sum_{j=1}^{J}\sum_{k=1}^{K}(e_{jk})^2 \approx gx_{(h)}^2 \qquad (3)$$

where e_{jk} are the elements of E. Q_i indicates the distance between the actual values of the batch and the projected values onto the reduced space.

The *D*-statistic or Hotelling T^2 statistic, measures the degree to which data fit the calibration model:

$$T_i^2 = t_i^T S^{-1} t_i \approx \frac{I(I-R)}{R(I^2-1)} F_{R,I-R} \qquad (4)$$

where S is the estimated covariance matrix of the scores. The T^2 gives a measure of the Mahalanobis distance in the reduced space between of batch and the origin that designates the point with average batch process behaviour.

4 Fuzzy classification method

In classification purpose the Learning Algorithm for Multivariate Data Analysis (LAMDA) has been used. This method combines both, numeric and symbolic classification algorithms, taking profit of fuzzy logic and hybrid connectives. LAMDA is proposed as a classification technique to determine the current situation of the process [2,6]. The following paragraphs resumes the classification principle used in LAMDA.

One $\hat{X}$ row (object) is a batch. This object has a number of principal components called "descriptors". These descriptors are obtained by MPCA and which are used to describe the batch. Every object is assigned to a "class" in the classification process [10]. Class (k_i) is defined as the universe of descriptors, which characterize one set objects (batch) as pictured in figure 4.

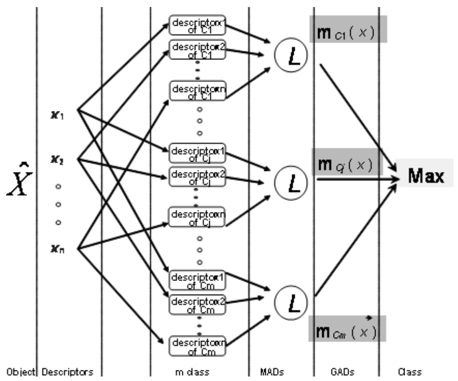


Figure 4: Basic LAMDA recognition methodology

MAD (Marginal Adequacy Degree) concept is a term related to how similar is one object descriptor to the same descriptor of a given class, and GAD (Global Adequacy Degree) is defined as the pertinence degree of one object to a given class as in fuzzy membership functions ($m_{ci}(x)$) [1].

Classification, in LAMDA, is performed according to similarity criteria computed in two stages. FirstMADto each existing class is computed for each descriptor of an object. Second, these partial results are aggregated to get a GAD of an individual to a class [2]. The former implementation of LAMDA included a possibility function to estimate the descriptors distribution based on a "fuzzification" of the binomial probability function computed as equation 5. This approach was used in this work.

$$MAD\left(\frac{d_i x_j}{\rho_{i/k}}\right) = \rho_{i/k}^{d_i x_j}\left(1-\rho_{i/k}\right)^{1-d_i x_j} \qquad (5)$$

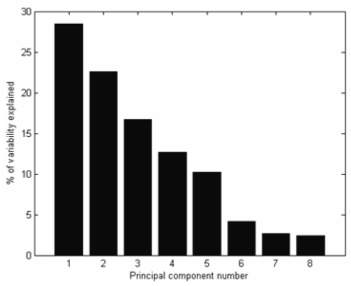
where $d_i x_j$ = Descriptor i of the object j $\rho_{i/k} = \rho$ of descriptor i and class k

On the other hand, GAD computation was performed as an interpolation between a t-norm and a t-conorm by means of the β parameter such that $\beta = 1$ represents the intersection and $\beta = 0$ means the union [7].

$$GAD = \beta T(MAD) + (1-\beta)S(MAD) \qquad (6)$$

5 Results and discussion

The MPCA algorithm were applied to the three-way data array, $\underline{\mathbf{X}}$, with dimensions 179 x 4 x 392, where $k = 392$ is time intervals throughout the batch (samples), $J = 4$ are measured variables and $I = 179$ are the historical data batches. The three-way array $\underline{\mathbf{X}}$ has been unfolded in the batch direction (I x KJ) so, the new dimensionality are (179 x 1568), with dimensionality reduction technique are obtained only (179 x 8). The model is created with eight components which explained 92.79% of the total variability see Table 2.



Principal component number	Eigenvalue of Cov (X)	% Variance Captured This PC	% Variance Captured Total
1	1.06e+000	26.44	26.44
2	8.39e-001	20.98	47.42
3	6.21e-001	15.53	62.95
4	4.69e-001	11.73	74.68
5	3.80e-001	9.49	84.17
6	1.56e-001	3.90	88.08
7	9.85e-002	2.46	90.54
8	9.03e-002	2.26	92.79

Table 2: This is an example of a table caption.

To examine the process data in the reduced projection spaces, is plotted the first two principal component, dashed line is the model (see figure 5). In the score plot a group the batch are out the model. This batch corresponds of disturbance.

The SBR process is very complex so chemistry engineers and control engineers provided the types batch classification depicted in Table 3. It has been used to locate and establish the relationship between classes and types batch. In this analysis were used two methods. First, the analytical methods proposed in [11] where the use of organic matter for denitrification is purposed and Second, the results preliminary of MPCA. This study creates five types of batch process. These are: Electrical fault, Variation in the composition, Equipment defects, Atmospheric changes and Normal behaviour. The normal behaviour was the most common type with higher nitrogen efficiency than legally-required effluent standards and there are 60 batches with abnormal behaviour (Table 4).

Abnormal behaviour of batch is identified by projecting the batch onto the model. Figure 6 shows the Q and D-statistic charts. In the Q chart, only a third of the total the abnormal behaviour is detected, furthermore there are 8 false alarms. The T^2 chart has 20 batches with abnormal behaviour (without false alarm), besides two batches are outside. These batches had electrical fault (EF). Table 5 summarise all information. In conclusion, 31 about 60 of the abnormal behaviour can be detected, 9 batches are in both charts.

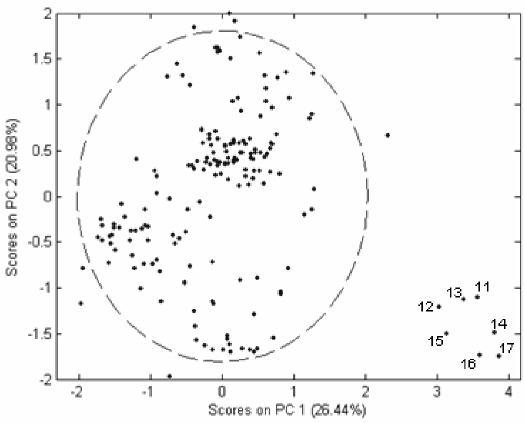


Figure 5: Score plot in batch direction.

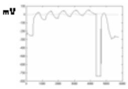
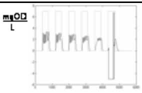
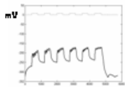
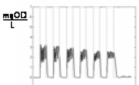
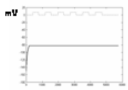
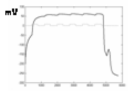
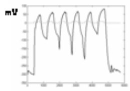
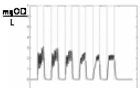
Type	ORP	OD	Effluent Quality
Electrical fault			Regular quality c* nitrification p*denitrification
Variation in the composition			Regular quality p* nitrification p*denitrification
Equipment defect			Bad quality n nitrification denitrification
Atmospheric changes			Normal quality c* nitrification p*denitrification
Normal behaviour			Excellent,good,normal quality c* nitri/denitri

Table 3: Types batch classification (c* complete, p* partial).

Type of batch process	Quantity	%
Atmospheric changes	17	9.50
Equipment defects	8	4.47
Variation in the composition	33	18.44
Electrical Fault	2	1.12
Normal behaviour	119	66.48

Table 4: For each type quantity and percentage.

SBR process is highly nonlinearly and MPCA is a dimensionality reduction lineal technique for this reason there are batches with abnormal behaviour that are not detected. LAMDA is used under a no supervised strategy. The descriptors matrix is based on eight principal components of MPCA. LAMDA classify automatically the data in eleven. The Table 6 compares the classes and type batch presented in process, there are some batches are wrong classified, in this table is showing the number and percentage of data in each class. According to these results, it is possible to identify classes as equipment defects, electrical faults, atmospheric changes and variation in the composition. The classes 1,9 and 10 correspond to normal behaviour. The group 6 is atmospheric changes. The classes 3 and 11 are variation in the composition. The classes 7 and 8 are electrical fault. Finally, the classes 2, 4 and 5 are composed by different types of batch process. The class 5 is abnormal behaviour because there are atmospheric changes and equipment defects. The relation between the class and principal components is studied. The 8th component is less predominant because it does not change. It is the same into the whole classes. If only seven components are selected, the total variability will be 90.54%. It indicates that the control model can be created with seven components.

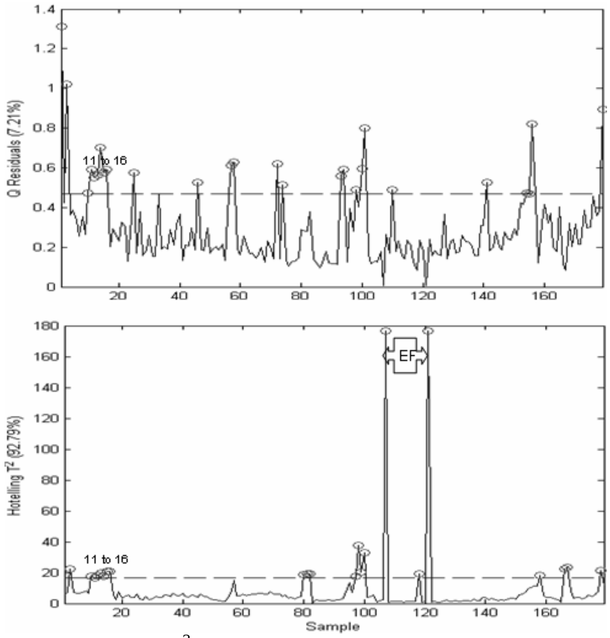


Figure 6: Q and T^2 statistic charts with 92.79% confidence limits

Type of batch process	Q		T^2	
	Quantity	%	Quantity	%
Atmospheric changes	9	5.03	4	2.23
Equipment defects	0	0.00	6	3.35
Variation in the composition	11	6.15	8	4.47
Electrical Fault	0	0.00	2	1.12
Normal behaviour	8	4.47	0	0.00

Table 5: Alarms in Q and T^2.

	Classes										
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
Normal	79	17	0	8	0	0	0	0	17	4	0
A C	2	1	0	10	3	1	0	0	0	0	0
E D	1	3	0	0	2	0	0	0	0	0	2
V C	4	11	7	1	0	0	0	0	3	1	0
E F	0	0	0	0	0	1	1	0	0	0	0
total	86	32	7	19	5	1	1	1	20	5	2
%	48	18	4	10	3	0.6	0.6	0.6	11	3	1

Table 6: Composition by each class.

The three-way array $\underline{\mathbf{X}}$ has been unfolded in the variable direction too (IK x J). The model was developed with dimensions (70168 x 4), where MPCA squeezes in 2 principal components explaining the 87.87% of the total variability. In figure 7 a projection on the first and second component plane of the statistical model. The batches are sequentially ordered and there are 3 sections into the model. Each section corresponds to one period determinate of SBR process. Test of SBR process match the first month with monitoring; Spring is the batches developed in spring season finally summer correspond to cycles in summer season. Temperature is less contribution than others variables (-0.25 by firs component) in consequence, Temperature was omitted and it was developed a new model only with 3 variables, figure 8 shown that process is highly related with environment, here it is explained with 91.88% of the total variability.

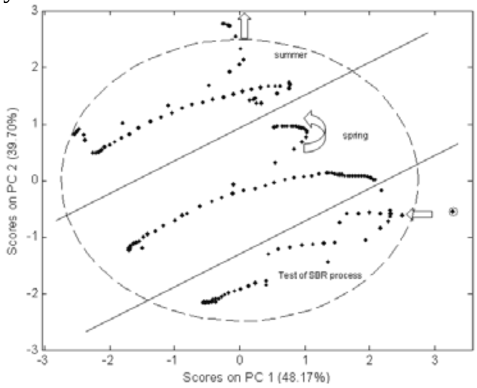


Figure 7: Score plot in variable direction

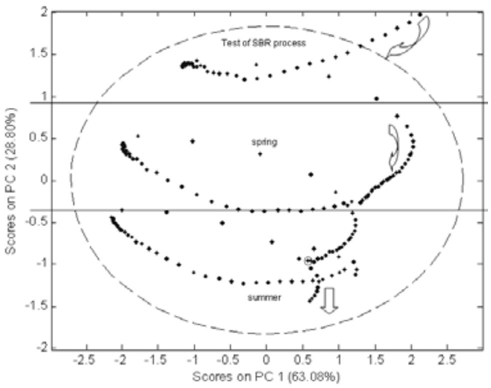


Figure 8: Score plot in variable direction without temperature variable

6 Conclusions

The MSPC has been used to detect abnormal behavior in SBR process, it was projecting the data into a lower dimensional space that accurately characterizes the state of the process. In general, MSPC is a good tool for situation assessment, but it was necessary to complement it with the classification tool. This classification tool can not be used directly because the process dynamics are not taken into account. However MSPC summarises one process (variables and sampling) in 7 variables in the batch direction and 3 in the variable direction with high percentage of total information.

References

[1] J.C. Aguado. A Mixed Qualitative-Quantitave Self-Learning Classification Technique Applied to Situation Assessment in Industrial Process Control. PhD thesis, Universitat Politècnica de Catalunya, 1998.

[2] J. Aguliar-Martin and R. López. The process of classification and learning the meaning of linguistic descriptors of concepts. Approximate Reasoning in Decision Analysis, pages 165–175, 1982.

[3] Ll. Corominas, M. Rubio, S. Puig, M. Vives, M. Balaguer, J. Colomer, and J. Colprim, editors. On-line Optimisation of Step-Fedd Operation of an Urban Wastewater Nitrogen Removal SBR by On-Line OUR Determination and ORP Analysis. 6th Specialist Conference on Small Water and Wastewater Systems (Australian), in Evaluation, feb 2004.

[4] Alberto Ferrer, editor. Control Estadístico MegaVariante para los Procesos del Siglo XXI. 27 Congreso Nacional de Estadística e Investigación Operativa (España), 2003.

[5] Dae Sung Lee and Peter A. Vanrolleghem. Monitoring of a sequencing batch reactor using adaptive multiblock principal component analysis. Biotechnology and Bioengineering, 82(4):489–497, may 2003.

[6] K. Moore. Using neural nets to analyse qualitative data. A Marketing Research, 7(1):35–39, 1995.

[7] J. Mora, D. Llanos, J. Melendez, J. Colomer, J. Sanchez, and X. Corbella, Classification of Sags Measured in a Distribution Substation Based on Qualitative and Temporal Descriptors. CIRED 17th International Conference on Electricity Distribuiton, may 2003.

[8] Diario Oficial n_ L 135 de 30/05/1991 P. 0040 0052. Directiva 91/271/CEE del consejo, de 21 de mayo de 1991, sobre el tratamiento de las aguas residuales urbanas, may 1991.

[9] Paul Nomikos and John F. MacGregor. Monitoring batch processes using multiway principal component analysis. AIChE, 40(8):1361–1375, aug 1994.

[10] N. Piera. Connectius de logiques no estandard com a operadors d'agregacio en classificacio multivariable i reconeixement de formes. Doctorate dissertation, Universitat Politecnica de Catalunya, 1987.

[11] S. Puig, M.T. Vives, Ll. Corominas, M.D. Balaguer, and J. Colprim, editors. Wastewater nitrogen removal in SBRs, applying a step-feed strategy: From Lab-Scale to Pilot Plant Operation. 3rd IWA Specialised Conference on Sequencing Batch Reactor Technology (SBR3), Australian, feb 2004.

[12] Age K. Smilde, Johan A. Westerhuis, and Ricard Boque. Multiway multiblock component and covariates regression models. Journal of Chemometrics, 14:301–331, 2000.

[13] Cenk Undey and Ali Cinar. Statistical monitoring of multistage, multiphase batch processes. IEEE Control Systems Magazine, 22(5):40–52, oct 2002.

[14] Johan A.Westerhuis, Theodora Kourti, and John F. MacGregor. Analysis of multiblock and hierarchical pca and pls models. Journal of Chemometrics, 12:301–321, 1998.