Combination of multivariate statistical process control and classification tool for situation assessment applied to a sequencing batch reactor wastewater treatment

MAGDA L. RUIZ, JOAN COLOMER, MONTSERRAT RUBIO AND JOAQUIM MELÉNDEZ

Control Engineering and Intelligent Systems (eXiT)

University of Girona, Spain

e mail: (mlruizo, colomer, montser ,quimmel)@eia.udg.es

Abstract: A combination of Multivariate Statistical Process Control (MSPC) and an automatic classification algorithm has been developed to be applied in a Waste Water Treatment Plant. Two extensions of the Principal Component Analysis have been used as MSPC method to diagnose the process and Fuzzy Technique used to classify situation assessment of the process. The goal is to perform situation assessment and classify the process with simple groups that describe the batch contributions and helps to fix the limits used to determine abnormal situations..

Introduction

The new regulations for quality monitoring of WWTP(law 91/271/CEE) [8] has been a very important reason for situation assessment. In recent years, major advances in wastewater treatment using activated sludge have been developed. In this work, the WWTP is a Sequencing Batch Reactor (SBR) where nutrient removal reactions and sedimentation are stages of the operation cycle of a batch process that occurs in the same reactor (see figure 1). The SBR has a cyclic nature, each cycle consist of several phases depending on the objective of operations (figure 2). This process is highly nonlinear, timevarying and subjected to significant disturbances such hydraulic changes, atmosphere changes (rain), composition variations, among others. To diagnose situations in a process is essential to take knowledge about the process and the operational behaviour of process variables along the time. Using this knowledge about the process is necessary to blend control approach and artificial intelligence [3]. Now, the problems into modern process are highly complex and operate with a large number of samples and variables. Therefore, the control model must consider the correlation structure between variables [5], characterized by the covariances matrix, due to the existing relation between variables and process. When statistical process control is used to batch process, often false alarms are generated [6]. Fortunately, this problem can be solved using Multivariate Statistical Process Control (MSPC). MSPC compresses the multidimensional information in few latent variables which explain the variability of the measured variables, including their relationships. In this paper, some models are developed using a extension of the Principal Component Analysis method to detect abnormal batch behaviours. This extension is Multiway Principal Component Analysis (MPCA) for batch data [9][17]. When the data collected during the operations have been previously diagnosed, it can be categorized into separate classes where each class belong to a particular condition. Therefore, in order to discover types of batch, an automatic fuzzy classification algorithm working under a non-supervised strategy is proposed.

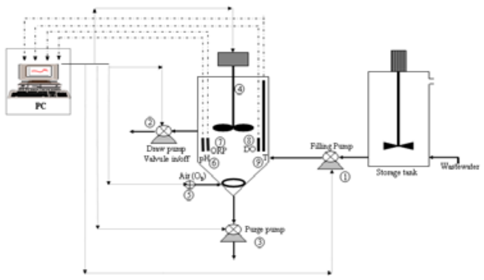


Figure 1. Schematic overview of SBR Pilot Plant. The data acquisition and control software was responsible for operation of (1,2,3) peristaltic pumps, (4) reactor mixing and (5) air supply control; as well as on-line monitoring of reactor (6) pH, (7) ORP, (8) DO and (9) Temperature.

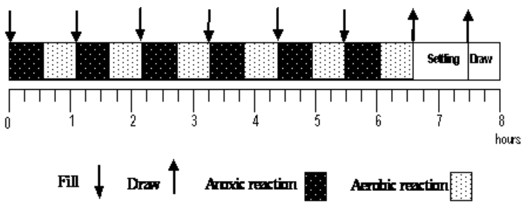


Figure 2. Fixed operational cycle applied in the pilot plant

Operation of the SBR process is presented in next section, PCA extensions for process monitoring is presented consecutively the fuzzy classification method is expounded. Finally a numerical example, future work and conclusions are summarised.

Influent composition and operational conditions of SBR plant

The SBR goal is mainly nitrogen removal. Nitrogen removal has been in two steps: Nitrification : the ammonia is converted to nitrate by aerobic microorganisms and Denitrification : nitrate is converted to nitrogen gas under anoxic conditions by anoxic microorganisms.

Pilot plant SBR is composed by a metal square reactor of 1m3. Minimum 483 liters and maximum volumes of the reactor 683 liters, it defines 200 liters of water to process. Waste water was taken directly from Cass`a-WWTP (Girona-Spain) by means of a peristaltic pump and it was stored in a storage tank, which has not temperature control.

Until now, the operation cycles of the process are fixed. Each batch spend 8 hours of the time, it has 392 samples (obtained every 60 seconds) by each variable: pH, Oxidation Reduction Potential (ORP), Dissolved Oxygen (DO) and Temperature [12]. Each cycle of the pilot plant SBR was based on alternating anoxic and aerobic reaction, where the filling only occurred during anoxic stages. The anoxic period was longer than aerobic period for increasing denitrification. Total filling volume was 200 liters, divided in six feeding parts during the cycle of 8 hours. The settling and draw spend of 1 hour and 0.46 hours respectively, in figure 2 is showing the scheme of operation cycle applied [4].

In this work, the chemistry engineers provided the types of batch classification depicted in table 1. It has been used to locate and establish the relationship between classes and types of batches. In this analysis has been used the analytical methods proposed in [11] where is the specific use of organic matter for denitrification purpose. The normal behavior was the most common type with a higher nitrogen efficiency than legally-required effluent standards. Composition variation is due the ORP disturbance. In table 1 is possible to observe difference between ORP for normal and abnormal behavior. Others type of batch are equipment defects and atmospheric changes.

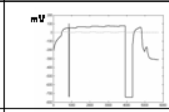
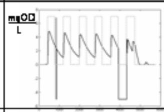
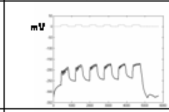
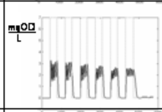
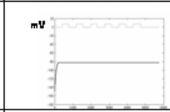
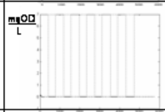
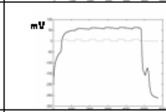
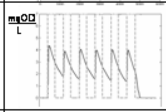
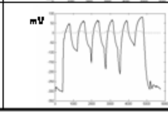
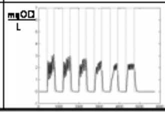
Type	OPR	OD	Effluent Quality
Electrical Fault			Regular Quality Complete nitrification Partial denitrification
Composition variation			Regular Quality Partial nitrification Partial denitrification
Equipment defects			Bad Quality None nitrification and denitrification
Atmosphere changes			Normal Quality Complete nitrification Partial denitrification
Normal behavior			Excellent/good Quality Complete nitrification/denitrification

Table 1. Classification of Batch types

Extension of the Principal Component Analysis

Multiway Principal Component Analysis (MPCA) model 1

Consider a typical batch run in which j = 1, 2, ..., J variables are measured at k = 1, 2, ...,K time instants throughout the batch. Similar data will exist on a number of such batch runs i = 1, 2, ..., I. All the data has been summarized in the X (I x J x K) array illustrated in figure 3, where different batches are organized along the vertical side, the measurement variables along the horizontal side, and their time evolution occupate the third dimension. Each horizontal slice through this array is a (J x K) data matrix representing the time histories or trajectories for all variables of a single batch (i). Each vertical slice is an (I x J) matrix representing the values of all the variables for all batches at a common time interval (k). [9] [17]

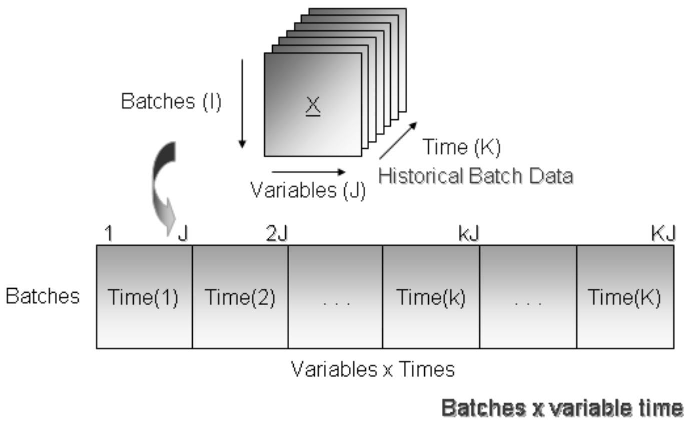


Figure 3. Arrangement of a three-way array X and Decomposition of X to 2-D (IxKJ)

The objective of MPCA is to decompose the three-way X, into a large two-dimensional matrix X. The MPCA algorithm derives directly from the NIPALS algorithm and as a result the matrix X. It is the product of score vector tr and loading matrices Pr, plus a residual matrix E, that is minimized in a least-squares sense as

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

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

where R denotes the number of principal components retained.

Multiway Principal Component Analysis (MPCA) model 2

Another suggestion is unfolding this three-way array into a two-way matrix of size (KIxJ) by preserving the variable direction see figure 4.[15]

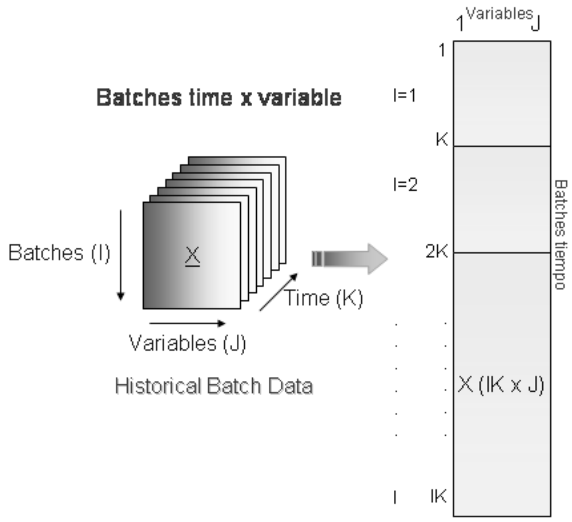


Figure 4: Other decomposition of a three-way data matrix,X, by MPCA

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 follow a normal distribution. The Q-statistic is a measure of the lack of fit with the established model. The D-statistic or Hotelling T2 statistic, measures the degree to which data fit the calibration model (see equation 3).

$$Q_i = \sum_{j=1}^{J}\sum_{k=1}^{K}(e_{jk})^2 \sim gx^2_{(h)} \qquad D_i = t_i^T S^{-1} t_i \sim \frac{I(I-R)}{R(I^2-1)} F_{R,I-R} \tag{3}$$

Fuzzy Classification Method

For classification purpose a 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 and provide more information about the process and a solution for diagnosis when the influent composition affect to the reactions of the process [2][7]. 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 which are used to describe the batch (see table 2).

$\hat{X}$

Batch	X1	X2	X3	X4	X5	X6	X7	X8
1	-1,19040	0,40769	-0,16347	0,68946	-0,37174	0,23784	-0,53304	-0,00425
2	-0,60437	-0,46155	-0,06614	0,76171	-0,24992	0,01876	-0,54621	-0,45446
3	-0,37465	-1,41540	-0,03215	0,70086	-0,18338	0,03463	-0,30230	-0,39654
4	-0,16948	-0,22701	-0,00589	-0,07695	0,00882	-0,15824	-0,02901	-0,08701
5	0,43591	0,19642	-0,02061	-0,19565	0,07733	-0,30996	0,07951	0,35192
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
177	-0,52598	-0,95484	0,73638	-0,70182	1,06240	0,76620	0,09412	0,88625
178	-0,52763	-0,94670	0,73593	-0,70479	1,06320	0,77200	0,08568	0,90762
179	-0,52763	-0,94670	0,73593	-0,70479	1,06320	0,77200	0,08568	0,90762

Table 2. Descriptors used by batch

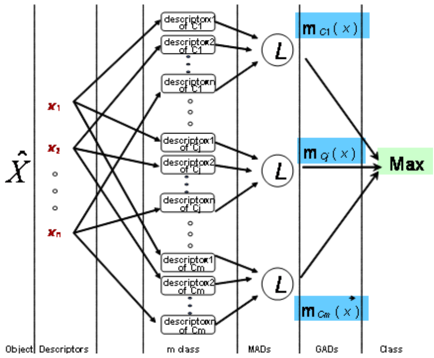


Figure 5. Basic LAMDA recognition methodology

Every object is assigned to a ”class” in the classification process [10]. Type of batch is the result of class assignment (see table 1). Class (ki) is defined as the universe of descriptors, which characterize one set objects (batch) as pictured in figure 5.

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 (mci(x)) [1]. Classification, in LAMDA, is performed according to similarity criteria computed in two stages. First MAD to 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].

Results and Discussion

The MPCA algorithm were applied to the three-way data array, 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 X has been unfolded in the batch direction (I x KJ) so, the new dimensionality are (179 x 1568) (model 1), 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.

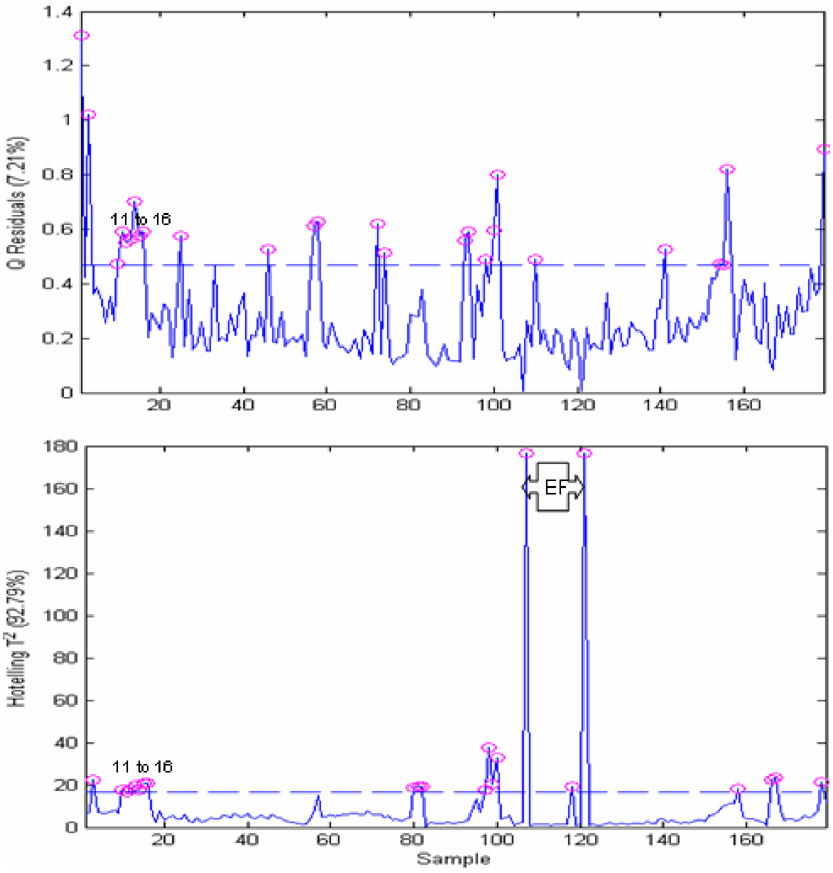


Figure 6. Multiway PCA. Q and D-statistic charts with 92.79% confidence limits

Abnormal behavior of batch can be identified by projecting the batch onto the model. Figure 6 shows the Q and D-statistic charts for all process batch. In the Q-statistic chart, it can be seen that some batches exceed its limits. These batches have several behaviors. In T2, two batches is far the limit. These batch presented electrical fault. For model 2 is possible to determine the connection of variables at one batch to other. The SBR process is exposed to various disturbances such as influent composition variation, atmosphere changes, equipment defects, etc. But all the batches is not classified so Fuzzy classification is used for apply expert knowledge to the different types of batch (table 1). The training method has been not supervised but exigency level was maximum. The classification is performed using eight principal components.

LAMDA classify the all data in eleven classes. Table 3 is showing the number and percentage of batches in each class. Class 1 have a 48.04% of the total the historical data.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11
mesuared	86	32	7	19	5	1	1	1	20	5	2
Porcentage%	48,04	17,88	3,91	10,61	2,79	0,56	0,56	0,56	11,17	2,79	1,12

Table 3. Resulting classes in the analysis for SBR process

Finally, the figure 14 compares the classes and the type of batch presented in the process. According to this results, it is possible to identify classes that only contain batch with equipment defects, electrical faults. The classes 1,9 and 10 correspond to normal behavior batch. The group conformed 4,5 and 6 are into SBR process of water rain. The batch with waste water in season rain is considerate normal batch which the only different is that it has few organic matter. The other classes are abnormal batches. The class 3 is composition variations in ORP. The classes 7 and 8 are electrical fault and class 11 is due equipment defects. Only the class 2 has normal and abnormal batches. This result is useful to identify the possible type of batch.

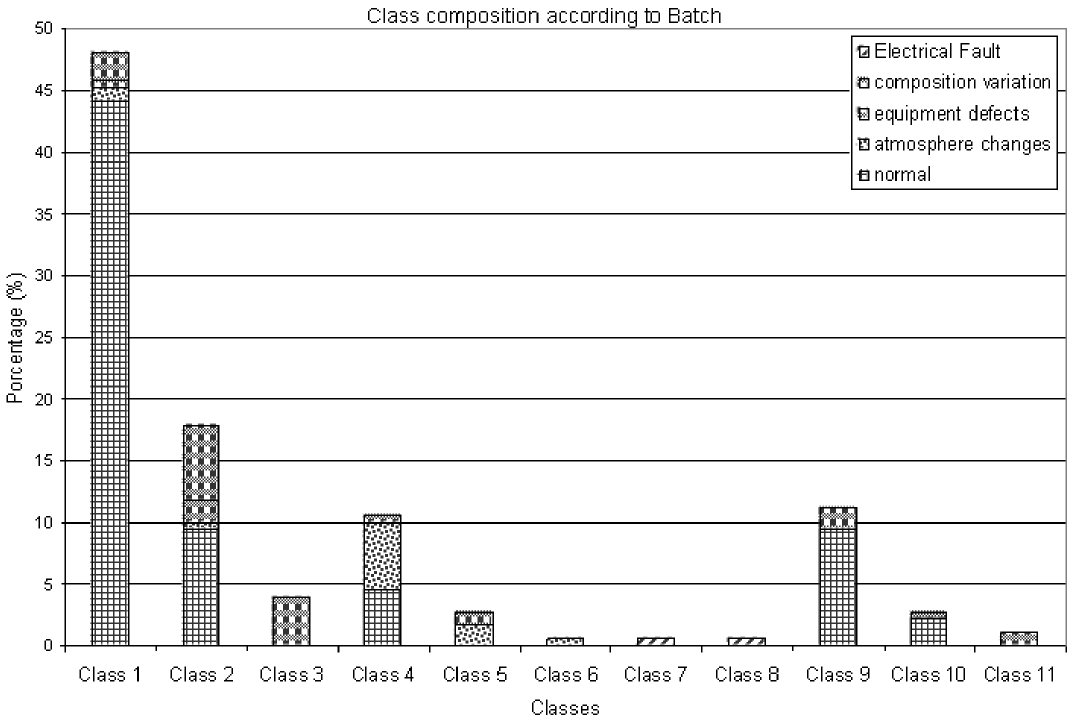


Figure 14. Batch class composition according to type of batch

Conclusions

Multivariate Statistical Process Control has been used to detect abnormal behavior in SBR process, also it was projecting the data into a lower dimensional space that accurately characterizes the state of the process. Therefore, the new variable matrix is smaller. In addition, a fuzzy classification has been used to relate classes with type batches. This information helps to relate the batch behavior with the defined by the chemistry engineers. In general, the results has been satisfactory, but there are type of batch where, if expert applies the knowledge in this moment, the type of batch will be normal so it is necesary to apply Multiblock MPCA.

ACKNOWLEDGEMENTS

This work is supported by Spanish government into research project ”Development of a system of control and supervision applied to a Sequencing Batch Reactor by loads (SBR) for the elimination of organic matter, nitrogen and phosphorus DPI2002-04579-C02-01”. Also, the authors appreciate the valuable contributions made by Sebasti Puig, Lluis Corominas and Maria Dolors Balager.

References

[1] J.C. Aguado, A Mixed Qualitative-Quantative Self-Learning Classification Technique Applied to Situation Assessment in Industrial Process Control, Ph.D. dissertation, Universitat Politcnica de Catalunya, 1998.

[2] J. Aguliar-Martin and R. Lpez, 'The process of classification and learning the meaning of leinguistic descriptors of concepts', Approximate Reasoning in Decision Analysis, 165–175, (1982).

[3] Joan Colomer, Joaquim Melndez, and Jordi Ayza, Sistemas de Supervisin, Romany Valls, S.A., 1ł edn., nov 2000.

[4] Ll. Corominas, M. Rubio, S. Puig, M. Vives, M. Balaguer, J. Colomer, and J. Colprim, eds. 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.

[5] Alberto Ferrer, ed. Control Estadstico MegaVariante para los Procesos del Siglo XXI. 27 Congreso Nacional de Estadstica e Investigacin Operativa (Espaa), 2003.

[6] 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, (mai 2003).

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

[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, mai 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 lgiques no estandard com a operadors d'agregaci en classificaci multivariable i reconeixement de formes, Doctorate dissertation, Universitat Politcnica de Catalunya, 1987.

[11] S. Puig, M.T. Vives, Ll. Corominas, M.D. Balaguer, and J. Colprim, eds. 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] Montse Rubio, Joan Colomer, Magda Ruiz, Jess Colprim, and Joaquim Melndez, 'Qualitative trends for situation assessment in sbr wastewater treatment process', Technical report, Workshop Besai'04 (in Evaluation) Valencia Spain, (aug 2004).

[13] Magda Ruiz, Montse Rubio, Joan Colomer, and Jess Colprim, eds. Multivariate Statistical Process Control to Situation Assessment of a Sequencing Batch Reactor. University of Bath (UK) and IEE CONTROL 2004, (in Evaluation), sep 2004.

[14] Age K. Smilde, Johan A. Westerhuis, and Ricard Boqu, 'Multiway multiblock component and covariates regression models', Journal of Chemometrics, 14, 301–331, (2000).

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

[16] Cenk Undey, Bruce A. Williams, and Ali Cinar, eds. Monitoring of Batch Pharmaceutical Fermentations: Data Synchronization, Landmark Alignment, and Real-Time Monitoring. 15th Triennial World Congress, Barcelona Spain 2002 IFAC, 2002.

[17] 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).