

Multiway Principal Component Analysis and Case Base Reasoning methodology for abnormal situation detection in a Nutrient Removing SBR

M. Garcia, M. Ruiz, J. Colomer and J.Melendez

Abstract—Multiway Principal Component Analysis (MPCA) and Case-Based Reasoning (CBR) approaches are applied in a biological nutrient removal process. The goal is monitoring of normal and abnormal operation conditions in this process. MPCA is used as a compression tool where with few variables the process can be described, as well as, to detect batches with abnormal conditions. However, some abnormal conditions (alarms) are omitted or, otherwise some false alarms are produced. Then, CBR is proposed. It assumes that similar problems should have similar solutions. In biological process, these problems generally are operational situations under normal or abnormal situations: low ORP, high pH, sensors fault, among others. These symptoms can be stored in a Case Base (CB) in order to diagnose future situations. Several proof are made in order to find the better methodology.

I. INTRODUCTION

The environment care has gained an important consideration through the last decades. The European Union with the new rule for control of Wastewater Treatment Plans (WWTP)(91/271) referred to the urban wastewater has started to regulate the final quality. In this work, the goal is to detect the normal and abnormal operation conditions from a WWTP, specifically, of the type Sequencing Batch Reactor (SBR). SBR are mainly characterized by sequential process phases of fill, reaction and settle. The advantages of SBR process can be attributed that: a) the clarification occurs in the same reactor b) biological process take place in a cyclic c) a portion of the treated water is replaced by untreated wastewater for each cycle distinguishing the SBR process from other continuous flow type activated sludge systems d) influent and effluent flows are uncoupled by time sequencing [3]. This process has a high correlation structure between variables, characterized by the covariance matrix. Besides, it is highly nonlinear, time-varying and subject to disturbances with a big amount of data collected. In this way, Multiway Principal Component Analysis (MPCA) has demonstrated to be a powerful data tool for compression and information extraction allowing to find linear combinations of variables that describes major trend

This work was supported by the Spanish Government

M. Garcia is from Modal Interval Automatic Engineering Group - MICE, Department of Electronics, Computer Science and Automatic Control, University of Girona, Campus Montilivi CP 17071 Building PIV, Girona - Spain magarcia@eia.udg.es

M. Ruiz, J. Colomer and J.Melendez are from Control Engineering and Intelligent Systems Group -eXiT Department of Electronics, Computer Science and Automatic Control, University of Girona, Campus Montilivi CP 17071 Building PIV, Girona - Spain mlruizo,colomer,quimmell@eia.udg.es

in a data set [21]. Several studies has been developed using MPCA for analysis and monitoring of WWTP and SBR process, for instance Rosen [12] [13], Ruiz [14] and Villez [17] among others. However, the studies are centered in fault detection or process monitoring.

In this work, the main objective is to start with the diagnosis of the process. Villez [17] obtains good results combining PCA with LAMDA clustering. So, in this sense is proposed Case-Based Reasoning (CBR) as an Artificial Intelligence approach which can be applied to improve expert supervision by exploiting data acquired from the MPCA results. The CBR advantages is that the Case Base is built just once and the maintain and update is made through of the learn capacity of this tool. In Sanchez-Marre [16], Nunez [11] and Wisnes [20] can be appreciate some application of CBR directly from the data sensor of WWTP's. Here, it is proposed an implementation of MPCA approach with CBR. For this, according with Ruiz [14], MPCA has been used as a dimensionality reduction tool which obtains good representation of the data process in few variables. Next, the results from MPCA are used as descriptors or attributes by CBR.

This paper is organized as follows: In sections II to IV provide detailed information about the materials; these are: section II: the definition of the SBR process is presented; section III: the MPCA approach is described in detail; section IV: the Case-Based Reasoning (CBR) definition is presented. The work methodology is explained in section V. The results are shown in section VI. Finally the paper finishes with conclusions and future work in section VII.

II. PILOT-SCALE SEQUENCING BATCH REACTOR (SBR)

In this work, data comes from a nutrient removing Sequencing Batch Reactor (SBR) Wastewater Treatment pilot Plant. 1959 complete batch run from December of 2003 until July of 2005 and stored in a data base. This process has a volume of 80 liters. The input or influent is synthetic and similar to the domestic wastewater characteristics [3].

The process operation consists of 7, that's means 4 cycles per day. The first 60 minutes comprise the fill phase. In the reactor phase, 4 repetitive times of 32.5 aerobic minutes and 20 anoxic minutes sub-phase occur. At the end of each aerobic sub-phase the excess of sludge is wasted. Hereafter,

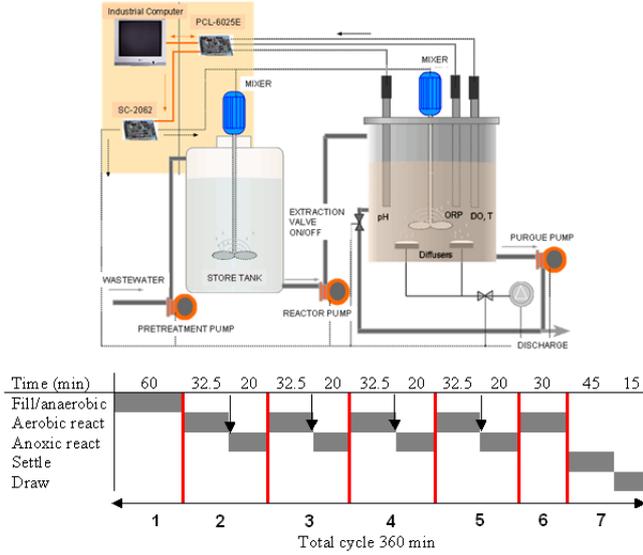


Fig. 1. The operations scheme of the SBR pilot plant

a final aerobic sub-phase of 30 minutes occurs. Finally, the settling and draw phase occurs with a duration of 60 minutes .

Measurements of pH, Oxidation Reduction Potential (ORP), Dissolve Oxygen (DO), temperature, weight and conductivity are recorded every minute, resulting 360 measurements per variable per cycle. However, the last 60 samples of each batch are not take in account due to the next reasons [6] [17]:

- 1) Properties and dynamics of the settling do not provide much information about the process
- 2) As consequence of the last item, these measures are not representative
- 3) Sludge dynamic from the settling phase might lead to batch to batch which is not easy to explain

III. MULTIWAY PRINCIPAL COMPONENT ANALYSIS (MPCA)

The main objective in Principal Component Analysis is the dimensionality reduction. It means to produce a lower dimensionality in which the correlation structure between the process variables is preserved [15]. In the past, processes had a small number of measured variables. Nevertheless, modern processes continually collect massive amounts of data [5].

Before using PCA, data scaling have been applied. Several studies about scaling are presented in literature, in this work, group scaling was selected. The mean trajectories are removed and each variable has equal variance [19]. Once the variables have been standardized, the X covariance matrix is calculated (equation 1).

$$S = \frac{1}{n-1} X^T X \quad (1)$$

The matrix P which it is columns are the eigenvectors of S and the diagonal matrix λ with eigenvalues of S on the main diagonal are found in equation 2.

$$S\hat{P} = \hat{P}\lambda \quad (2)$$

Each eigenvalue is associated to an eigenvector. The eigenvector with the highest eigenvalue represents the most important pattern in the data, i.e. contains the largest quantity of information, therefore: this vector is called the principal component of the data set. Ordering the eigenvectors by its eigenvalue, highest to lowest, gives the components in order of significance. In order to reduce the dimensionality, the less important components can be eliminated (information is lost, but if the eigenvalues are small, this information is not much), then only the n first eigenvectors are chosen (loading vectors and denoted by P) and the final data set will be n -dimensional. The projected matrix T (or score vectors) in the new space is defined in equation 3.

$$T = XP \quad (3)$$

Projecting T over the K-dimensional space the result is:

$$\hat{X} = TP^T \quad (4)$$

where the difference between X and \hat{X} is the residual matrix E:

$$X = \hat{X} + EX = TP^T + E \quad (5)$$

However, the data collected from batch process are stored in a three dimensional matrix as at the top of Figure 2. In this manner, $1 \dots J$ variables and $1 \dots K$ samples per variable are stored together with $1 \dots I$ processes or cycles in a three dimensional data matrix [10]. MPCA is equivalent to performing ordinary PCA on a large two-dimensional matrix constructed by unfolding the three-way data matrix. Several methodologies to unfolding have been suggested in [19]. This work unfolding in direction of the batch is selected as is illustrated in Figure 2.

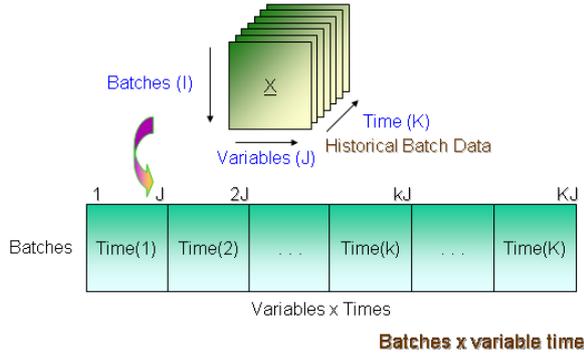


Fig. 2. Dispossession and unfolding of three-way data matrix

The previous process allows decomposing the three-way data matrix \underline{X} into a series of principal components consisting of score vectors T and loading matrices P , plus a residual three dimensional matrix \underline{E} , in this manner, MPCA is defined in equation 6:

$$\underline{X} = T \otimes P + \underline{E} \quad (6)$$

where \otimes denotes the Kronecker product or Tensor product.

Two control charts are needed for process monitoring using MPCA. They are Q -statistic and Hotelling T^2 statistic. The new batch is compared with the control limits determining whether the process is in control or not. The control limits are calculated selecting only batches with Normal Operation Condition (NOC), in others words, any abnormal situation has to be removed at this stage of this depend the successful control chart application [5] [4].

The Q -statistic indicates the distance between the actual values of the batch and the projected values onto the reduced space. This control chart will detect any trouble that can cause the moving of the process away from the model. For batch number i , Q_i is defined as:

$$Q_i = \sum_{j=1}^J \sum_{k=1}^K (e_{jk})^2 \sim g_{x^2(h)} \quad (7)$$

where e_{jk} are the elements of \underline{E} .

The Hotelling T^2 gives a measure of the Mahalanobis distance in the reduced space between one batch and the origin of the model. This control chart mainly verifies if a new batch is inside of the model defines by the principal components.

$$T^2 = t_i^T S^{-1} t_i \sim \frac{I(I-R)}{R(I^2-1)} F_{R,I-R} \quad (8)$$

where S is the estimated covariance matrix of the scores.

IV. CASE BASE REASONING (CBR)

Case Base Reasoning (CBR) adapts and combines old situations to interpret new problems. It is based on the idea that new problems are similar to past situations and therefore past solutions can be used in the new situation. Basically, this procedure is similar than the reasoning human model in which the people remember past information that subsequently it is used in order to solve future situations [7]. A real example of CBR applications is when the doctors diagnose and treatment any new patient with similar symptoms than former patients whit effective cure. The daily routine frequently presents opportunities to apply this approach.

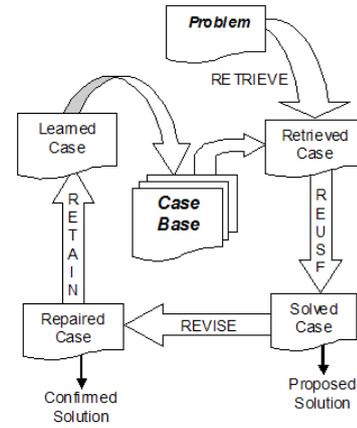


Fig. 3. CBR cycle

A. CBR cycle

CBR approach can be described by the following four steps in accord with [1]:

- Retrieve: The most neighbor case to the new problem. Provides the description of a situation, or problem, and a set of objectives, or tasks, that have to be performed, it is a question of finding a similar case, or a small set of similar cases, that may be useful. Where old situations can be used as inspiration for solving new problems.
- Reuse: The goal in this step is to use the case selected in the previous step in order to solve the new problem.
- Revise: Two phases are needed in this step in order to check up the proposed solution: Evaluation and Learning. The first one evaluates the solution generated explaining the differences, justifying the differences, projecting the outcomes and comparing and raking alternative possibilities. The second one "learning" comes from the success and retaining the case when the solution is correct. Otherwise, old solutions can be update to fit new solutions.

- Retain: Immediately of the Revise step is the moment for retaining new cases into the case base once it has been confirmed or validated. The system must decide which information will be retained and how it will be incorporated into the memory structure where indexing mechanisms are used for an efficient retrieval.

B. Case Representation

In CBR approach is necessary a set of experiences which are called **Case Base** (CB). One case is a piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner. Each case is generally a register comprising a description (minimal representation) of a problem and the functional solution [2].

C. Distance

Several methodologies in order to find the closest cases are reported in the bibliography. Sanchez-Marre [16] and Watson [18] describe the sum of the similarity of all attributes multiplied by a weighting factor. In Nunez seven different distances were checked for two environmental systems: HVDM, IVDM, Euclidean, Manhattan, Clark, Canberra and L'Eixample [11]. Jurgen Wiese made another Euclidean distance implementation for a WWTP [20]. Mujica applies neural network in order to find the better neighbor [9].

D. CBR and Supervision

It is possible to apply CBR approach in the three stages of supervision (fault detection, diagnosis and reconfiguration). In this point, the supervision can be treated as a comparison among symptoms because the data from process with Abnormal Operation Condition (AOC) are stored with specific behavior. In this manner, these cases contain previous experiences representing symptoms, faults, diagnosis and actions, acquired in specific situations [8].

V. WORK METHODOLOGY

The data come from the pilot-plant SBR. 1959 were run from 16-12-2003 to 18-07-2005. Collecting six 6 variables per batch where each variable has 300 samples. In this manner 1800 samples per batch are used to apply MPCA and CBR methodology. To detect batches with abnormal situation several steps are necessary to develop as follow:

A. Step 1: Building Data set

Applying MPCA approach, the data are divided in three sets in accord with Villez [17]. The first one is used to build the model. Second and third sets are used to validate the model.

- Set 1: Building the model: 1119 batches with normal operation condition are selected.
- Set 2: Validation data set of AOC: In this set, there are 560 batches with abnormal operation condition

- Set 3: Validation data set of NOC: It is conformed for 280 batches with normal operation condition

In some implementations, it is necessary to simulate situation of normal and abnormal operation in order to build the Case Base (CB). However, in this work, that is not necessary because the data from the process contains enough amount of batches. However, it is necessary to separate some batches in order to build and check the better option. In this manner, two 2 CB are performed. Both CB are conformed by NOC and AOC data sets. The difference between them is the amount of batches in each CB as follows:

- CB1: It is composed by the whole data set of batches used to build the model in MPCA. It is 1119 batch with NOC besides 20% of batches from the validation data set of AOC, They are 112 batches. In consequence, the CB has a total of 1231 batches as it is shown in figure 4. The whole validation data set of NOC (280) is used in order to test the CBR. As well as, the rest 80% of batches from the validation set of AOC are tested too, in others words 448 batches with AOC are used to check the CBR approach.

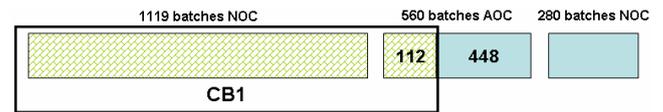


Fig. 4. Case Base 1

- CB2: It is composed for just 448 batches from the data set used to build the model. Same amount of batches are selected from the validation data set of AOC, it means 80%. In consequence, the CB has a total of 896 batches, see figure 5. In the same way as was done before, the whole data set of NOC is used to test the CBR, as well as, the rest 20% of batches from the data set of AOC. Due to CB2 is only using a little portion from the model data set, the other part is used to validate the CBR approach, they are 671 batches and they are called the rest of model data set.



Fig. 5. Case Base 2

B. Step 2: Reduction of dimensionality

Using MPCA, the model was built using only batches with Normal Operation Condition (NOC) in accord with [14]. In total 1119 batches are selected and stored in a three dimensional data matrix. The data matrix was unfolded as suggest Nomikos and McGregor [10], afterwards, the data

were scaled calculating one standard deviation per variable and one mean for each sample through the batch. Commonly, this data scaling is called "group scaling". The model with T^2 -statistic and Q -statistic was developed using 95% of confidence limit. Five principal components are selected in order to perform the model with a 85.27% of the total percentage of variance captured. The validation data sets (NOC and AOC) were stored, unfolding and scaling using the same methodology applied to develop the model. Using T^2 -statistic and Q -statistic was possible to determinate the amount of batches outside of the limit for each validation data set.

C. Step 3: Descriptors

Each new problem and each case stored in CB are described by means of the results from MPCA approach. These results are called "descriptors". In consequence, seven 7 descriptors have been used as following:

- Descriptor 1 corresponding at Q values
- Descriptor 2 corresponding at T^2 values
- Descriptors 3 to 7 are the 5 Principal Components (PC's)

D. Step 4: Retrieval

Calculating the distances, four different options were considered as follow:

- $(Q + T^2 + PC's)$ distance: The simplest distance is calculated employing all the descriptors including T^2 which has not been taken in account by the others distances.
- Q -distance: Only using the Q descriptor for the problem and the CB. The distance is calculated comparing the new problem and batch-by-batch i from the CB.

$$Q - distance = \sqrt{(Q_{new} - Q_{iCB})^2} \quad (9)$$

- $PC's$ -distance: As in the previous distance, This distance is calculated using only the $PC's$ as unique descriptor. It is done comparing each PC $p=1,..,5$ from the new problem with each batch i from the CB.

$$PC's - distance = \sqrt{(PC's_{pnew} - PC's_{ipCB})^2} \quad (10)$$

- Combining distances: Q and $PC's$ distances are implemented in this item. For this, two levels are take in account. These are: First one, Q descriptor is used looking for the 30 nearest neighbor in accord with equation 9. Second one, the $PC's$ distance (equation 10) is used with the 30 cases preselected. Here only the first five 5 nearest neighbor than the problem were inspected. In according at the biggest number of repetitions of the batches selected like neighbors, the problem will be

labeled as them. For instance, if the nearest neighbors are 3 NOC and 2 AOC the problem is NOC. In this work, this technique is called VOTING.

$$\begin{aligned} Q_d &= (Q_{new} - Q_{iCB})^2 \\ T_d^2 &= (T_{new}^2 - T_{iCB}^2)^2 \\ PC's_d &= \left(\sum_{p=1}^P (PC's_{pnew} - PC's_{ipCB}) \right)^2 \\ distance &= \sqrt{Q_d + T_d^2 + PC's_d} \quad (11) \end{aligned}$$

E. Step 5: Testing

To find the better performance of MPCA and CBR methodology, CB1 and CB2 are evaluated using the 4 distances as was explained in section V. In others words, for both CB create, four different distances are checked. Table I summarizes how the respective retrievals were labeled.

	CB1	CB2
Q -distance	CBR 1	CBR 5
$PC's$ -distance	CBR 2	CBR 6
Combining distances	CBR 3	CBR 7
$(Q + T^2 + PC's)$ distance	CBR 4	CBR 8

TABLE I
NAMES FOR EACH DEVELOPED CBR

VI. RESULTS

In this section, first the MPCA results is given and secondly the answer of the CBR methodology is presented.

A. MPCA

The data are stored in a three dimensional data model matrix has a size of (1119 x 6 x 300). Where 1119 corresponds to the number of batch run; 6 are the process variables and 300 are the sample number. In figure 6 the first three PC's is pictured where the bubble is the limit of the model. As was explained in section III, the control charts are used in order to detect NOC or AOC.

1) *Results of validation set of AOC:* In this set, 500 batches are detected in abnormal operation. This is a sensitivity of 89%. Sensitivity means the true alarms. To obtain this results is necessary join the results from Q -statistic and T^2 -statistic charts. Table II shows the percentage of batches detected by each control chart.

2) *Results of validation set of NOC:* In this set, 43 batches are detected in abnormal operation. This is a error or false alarms of 15.36%. To obtain this result is necessary join the results from Q -statistic and T^2 -statistic charts. Table II shows the percentage of batches detected by each control chart.

In accord with the results obtained from the previous study of this process, a robust discrimination between different

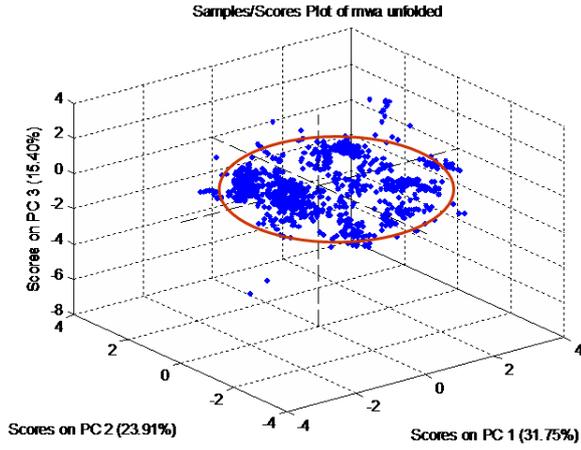


Fig. 6. Disposition and unfolding of three-way data matrix

Parameter	% True NOC	% True AOC
Q-statistic	91.79	82.14
T-statistic	91.43	61.25

TABLE II
METHODOLOGY MPCA

anomalies and operation changes will be used to diagnoses the SBR process [17]. In order to find the better methodology and as previous results, in this work, CBR is only used for detection.

B. CBR

To find the better performance of MPCA and CBR methodology, CB1 and CB2 are evaluated with 4 several distances as was explained in section V. As such, two 2 groups of simulations were done in order to improve the best combination between the CB and the distances calculation. The analysis and discussion of the results are divided in two: First analysis, the simulations from each CB are evaluated separately. For proper evaluation of the performance, the percentage of sensitivity are calculated. Second analysis, all simulations are analyzed together in order to compare CB.

1) *First analysis:* As can be seen, table III exhibits the first group of simulations applied to CB1. There are not performance of the methodology when the validation set of AOC are checked. The reason of these results can be due to the amount of NOC batches that are inside of CB1 is higher than the number of AOC. However, the results improve when the validation set of NOC are inspected. A perfect sensitivity is obtained when the CBR 2 and 3 are used. When CBR 1 and 4 are tested the sensitivity is acceptably good.

In respect to the CB2, table IV shows the results of these proofs. The same validation sets are used like CB1. Nevertheless, the batches that have not taken in account for the CB2 are inspected too. At this set of batches is given the name of NOC rest. There are poor performance when the CBR 5 and 8 are checked in all of these data set. In

	% True NOC	% True AOC
CBR 1	85.49	0
CBR 2	100	0
CBR 3	100	0
CBR 4	89.41	0

TABLE III
SENSITIVITY FOR CASE BASE 1 (CB1)

CBR 6, there are not performance of the methodology when the validation set of AOC are checked, in return, the results improve when the validation set of NOC are inspected, obtaining a perfect sensitivity. Finally, with the CBR 7 a complete sensitivity in the NOC like in the AOC is acquired.

	% True NOC	% True AOC	% True NOC Rest
CBR 5	57.37	56.25	63.62
CBR 6	100	0	100
CBR 7	100	100	100
CBR 8	62.05	54.91	58.71

TABLE IV
SENSITIVITY FOR CASE BASE 2 (CB2)

2) *Second analysis:* In relation with the selection between which Case Base (CB) is better, the results are clear. Using CB2, the performance of the methodology grows respect at validation set of AOC. Nevertheless, there are less sensitivity of the validation set of NOC for CBR 5 and 8 decreasing 30.6% and 32.89% respectively. In spite of, CB2 is the best option for this work.

VII. CONCLUSIONS AND FUTURE WORK

In this paper a study to find the better methodology for supervision of a nutrient removing SBR process using MPCA and CBR is done. From the results shown in section VI, some conclusions can be drawn. First of all, the best Case Base is CB2 which there are the same number of cases with NOC and AOC. Second, about the distances, when it is used distances Q, PC's and $(Q + T^2 + PC's)$ in order to find the most similar case, the results are lower than combining distances (Q and PC's). It is because, the distance Q not distinguish any feature of the descriptors and distances PC's and $(Q + T^2 + PC's)$ make the searching with incomplete information. Finally, combining distances (Q and PC's), which firstly is used the Q descriptor and immediately the PC's, is the best form to apply the distance calculation. As a consequence, the best application of CBR is CBR 7.

As future work, applying the methodology selected in this work, it will be possible to diagnose and reconfigure the SBR process. For this, learning techniques will be added at the methodology. Besides, it will be necessary to check the CB in order to delete repetitive situation maybe using Instance-Based learning algorithm (IB3).

VIII. ACKNOWLEDGMENTS

This work is part of the research project Development of a intelligent control system apply to a Sequencing Batch Reactor by loads (SBR) for the elimination of organic matter, nitrogen and phosphorus DPI2005-08922-C02-02 supported by the Spanish Government and the research project Automatization of the detection, diagnosis and tolerance in faults in uncertainty and distribute systems DPI2006-15476-CO2-02. The authors also appreciate the contributions made by BIOMATH group who have allowed to work with their data; in special attention: Kris Villez, Gurkan Sin and Peter Vanrolleghem.

REFERENCES

- [1] Agnar Aamodt and Enric Plaza. Case-based reasoning: foundational issues, methodological variations, and systems approaches. *AI communications*, 7(1):39–59, mar 1994.
- [2] Leake D.B. *Case-Based Reasoning: experiences, lessons and future directions*. Press, 1996.
- [3] Guclu Insel, Gurkan Sin, Dae Sun Lee, Ingmar Nopens, and Peter Vanrolleghem. A calibration methodology and model-based systems analysis for sbr's removing nutrients under limited aeration conditions. *Journal of Chemical Technology and Biotechnology*, 81:679–687, 2006.
- [4] Theodora Kourti. Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing*, 19:213–246, 2005.
- [5] Theodora Kourti and John MacGregor. Multivariate spc methods for process and product monitoring. *Journal of Quality Technology*, 28(4):409–428, oct 1996.
- [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] Cynthia Marling, Mohammed Sqalli, Edwina Rissland, Hector Muoz, and David Aha. Case-based reasoning integrations. *American Association for Artificial Intelligence*, pages 69–86, 2002.
- [8] Joaquim Melendez, Joan Colomer, and Josep Lluís de la Rosa, editors. *Expert supervision based on cases*. IEEE Conference on Emerging Technologies and Factory Automation (ETFA'01), 2001.
- [9] L.E. Mujica, J. Vehi, J. Rodellar, and P. Kolakowski. A hybrid approach of knowledge-based reasoning for structural assessment. *Smart Materials and Structures*, 14:1554–1562, 2005.
- [10] Paul Nomikos and John F. MacGregor. Monitoring batch processes using multiway principal component analysis. *AIChE*, 40(8):1361–1375, aug 1994.
- [11] H. Nuez, M. Sanchez-Marre, U. Cortes, Q. Comas, M. Martinez, and M. Poch, editors. *Classify environmental system situations by means of case-based reasoning: a comparative study*. iEMSs The International Environmental Modelling and Software Society, in collaboration with: Environmental Modelling and Software, 2002.
- [12] C. Rosen and G. Olsson. Disturbance detection in wastewater treatment plants. *Water Science and Technology*, 37(12):197–205, 1998.
- [13] C. Rosen and Z. Yuan. Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering. *Water Science and Technology*, 43(7):147–156, 2001.
- [14] M. Ruiz, K. Villez, G. Sin, J. Colomer, C. Rosen, and P.A. Vanrolleghem. Effects of scaling and unfolding on the pca-based monitoring of a batch process for biological nutrient removal from wastewater. *submitted to AIChE Journal*, 2006.
- [15] Evan L. Russell, Leo H. Chiang, and Richard D. Braatz. *Data-driven techniques for fault detection and diagnosis in chemical processes "Advances in Industrial Control"*. ISBN 1-85233-258-1, London, 2000.
- [16] Miquel Sanchez-Marre, Ulises Cortes, Ignasi R-Roda, and Manel Poch. Learning and adaptation in wastewater treatment plants through case-based reasoning. *Microcomputers in Civil Engineering*, 12:251–266, 1997.
- [17] K. Villez, M. Ruiz, G. Sin, C. Rose, J. Colomer, and P.A. Vanrolleghem. Combining multiway principal component analysis (mpca) and clustering for analysis of historical data from a nutrient removing sbr. *Water Research*, page accepted, 2006.
- [18] I. Watson. Cbr is a methodology not a technology. *Research and Development in Expert systems*, 15:213–223, 1998.
- [19] Johan A. Westerhuis, Theodora Kourti, and John F. MacGregor. Comparing alternative approaches for multivariate statistical analysis of batch process data. *Journal of Chemometrics*, 13:397–413, 1999.
- [20] Jurgen Wiese, ArminStahl, and Joachim Hansen, editors. *Possible Application for Case-Based Reasoning in the field of Wastewater Treatment*. 16 th European Conference on Artificial Intelligence, BESAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 2004.
- [21] Barry M. Wise, Neal B. Gallagher, Stephanie Watts, Daniel D. White JR, and Gabriel G. Barna. A comparison of pca, multiway pca, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics*, 13:379–396, 1999.