On-line Supervision and Control of an Aerobic SBR process

F. Ciappelloni, D. Mazouni, L. Lardon, J. Harmand*

Laboratoire de Biotechnologie de l'Environnement, INRA-LBE, Avenue des étangs, 11100 Narbonne, France. Telephone +33 468 425 161, Fax +33 468 425 160

Abstract: This paper presents a new software developed in MATLAB for analyzing on-line data of an aerobic SBR, detecting faults and, in this case, proposing the most probable causes of fault. Process diagnosis is achieved using a statistical method divided in two main phases: off-line model building and on-line data diagnosis. The off-line model identifies the correct working conditions of the system (standard operative conditions). It includes the characterization of the deviation of the system from these standard conditions in the case of changing in the biomass properties or carbon and nitrogen load characteristics. The on-line diagnosis aims at collecting and analyzing all the available data available through industrial sensors, and at classifying the behavior of each treatment cycle. The diagnosis performance of the proposed method is tested using a data set of an aerobic SBR pilot plant.

Keywords: monitoring, Sequencing Batch Reactor (SBR), Fault Detection and Isolation (FDI), Principal Component Analysis (PCA), Decision Support System (DSS).

INTRODUCTION

Safety in industrial processes is becoming more and more an important requirement and the ability to early detect and isolate possible causes of process faults is a primary need. In wastewater biological treatment processes, and in particular when dealing with batch processes, the on-line diagnosis is a very important requirement because of the high sensitivity of the whole system to a wide variety of external inputs and internal changing conditions. The operating efficiency can significantly vary over the time and is not easy, even for operators or experts to understand the observed changes. It is proved that the application of an efficient diagnosis system may help operators and engineers to manage in the best way the treatment process to maintain it as close as possible to the optimal conditions before the integrity of the process is affected.

A fault is any unexpected change in a system, due to malfunctions or variation of operating conditions. The aim of an advanced diagnosis system is not only to detect and isolate a fault, but also to propose the operator, as soon as the problem is detected, the most probable causes of fault to intervene quickly and to solve the problem in the fastest way and restore the optimal conditions. The need for an efficient supervisory system able to control bioprocesses in real time is important. Maintaining the efficiency and the quality of the sludge in the reactor is the first requirement. However, it is highly influenced by external inputs. Furthermore, the settleability of the sludge can be durably altered by changes in environment conditions. Too often, the only way to restore the process is to re-start up the process with fresh biomass leading to an important waste of time.

In the present study, an aerobic SBR pilot plant has been chosen as an illustrative example of biological WWTP. In aerobic conditions a series of biological processes take place in presence of oxygen provided by diffusers to the mix of organic matter and biomass. The first step in organic degradation is the consumption of COD (Chemical Oxygen Demand), the second is the transformation of NH4 (ammonium produced by the ammonification of organic nitrogen present in the input) into NO2, while the third and last step is the transformation of NO2 into NO3 that will be discharged at the end of the settling phase. In this treatment process, the most critical phase is clearly the aeration period during which the most influent parameters are Kla (oxygen transfer coefficient in liquid

* Corresponding author

phase), COD and TKN loading rates and the concentrations of biomasses (heterotrophic and autotrophic) in the reactor.

The supervision system analyzes on-line data provided by the acquisition system of the plant comparing each treatment cycle to the "ideal cycle" and to the others obtained by varying some operating conditions; if the real cycle is near to the best cycle, it will be classified as a good cycle, else it will be considered similar to the nearest simulated cycles and so its operating conditions.

The paper is organized as follows. First, the description of the aerobic SBR pilot plant and of its instrumentation is given. Then, the methodology used to analyze and classify on-line data as well as the model implemented to generate the optimal condition and the possible causes of fault are presented. Finally, the application of PCA to compare simulated and real cycles and so to suggest the most probable cause of fault is reported. Finally examples of experiments and results of the application are presented and discussed.

MATERIAL AND METHODS

The SBR pilot Plant

The experiments were performed using an aerobic SBR pilot plant at the "Laboratoire de Biotechnologie de l'Environnement" of INRA in Narbonne, France.

Activated sludge

The sludge used in the SBR comes from the urban wastewater treatment plant (WWTP) of Coursan, France. This plant was chosen because both carbon and nitrogen are removed biologically. The 200 liter tank was filled with about 100 liters of this sludge and operated under aerobic conditions for the first two months. A low organic loading rate (<0.7 kg COD/m3/day) was applied. Then, the anoxic phase was added and after a total start up period of three months, a good purification rate of 98% was obtained while high sludge settleability was observed. The actual average sludge concentration is 12 g/l.

Influent wastewater

Semi-synthetic dairy wastewater is used as influent. This feed is prepared by dilution of concentrated whey collected in a cheese dairy. The characteristics of the raw whey is as follows : 70 g/l COD and 2 g/L TKN. To simulate possible changes in the C/N ratio of the whey, organic nitrogen as Urea (CH4N2O) is used. Regarding these characteristics and given that the process is operated under low loading rate, in "standard conditions", the volume treated is 50 l/day. Its average properties are : 5 g/l COD, 250 mg/l N-TKN, TSS>10 g/l and pH = 6.23.

Pilot plant

The reactor tank is a 47 cm (diameter) times 130 cm (height) cylinder of a total volume of 255 liters. It is equipped with a variable input flow rate pump to fill in the reactor and a controlled valve to withdraw the effluent by gravitation and, when necessary, the sludge in excess. Air is used for aeration and mixing. Its flow is controlled by a gas flow meter. The plant is operated as follows. Filling period 1 (25 liters) : 5 mn, Anoxic period 1 : 2 h, Aerobic period 1 : 8 h 50, Filling period 2 (25 liters) : 5 mn, Anoxic period 2 : 2 h, Aerobic period 2 : 8 h 50, Settling and draw period : 2 h 05.

Measurements

The following sensors are installed : pH, Temperature, Dissolved Oxygen, Redox, CO2 and O2 content of the output gas and sludge level. All sensors and actuators are connected to a PC and the functioning of the process is completely automatized. The on line measurements are acquired through a home made software called MSPC and automatically stored in a data base using a data format defined by *PlantML,* a language of management of information fluxes via XML for biological processes developed within the framework of the European Project TELEMAC (Cf. Neveu *et al*., 2003).

The diagnosis system : classification of reaction phases

The developed system aims at computing a "distance" of a given cycle (one anoxic + one aerobic period) from a reference called "standard". In particular, the objective is to detect and isolate a COD or TKN overload, a problem in the oxygen transfer rate (Kla) or related to the heterotrophic or autotrophic bacteria. The fact that the detection is realized a posteriori (once the cycle is finished) is not a problem because the consequences of an overloading or of a problem on the oxygen transfer or on the biomasses are rather long term consequences rather than short terms ones. For instance, a "durable" COD overload will lead in problems on sludge settleability several cycles after this overload has began. The most probable cause of the identified problem is then displayed to the operator who can adjust the time of the subsequent reaction phases or take any other appropriate decision (calibration, withdrawingt of the sludge, etc).

Regarding the objectives of the study, the aerobic phase is the most important period to investigate the process performance. In fact, during this phase, significative information can be extracted from the dissolved oxygen, pH and Redox curves. In particular, using either the pH or the curve (as shown in figure 1), it is possible to detect when each substrate is completely degraded. The first point corresponds to the end of the COD consumption, the second corresponds to the end of the nitritation while the last one corresponds to the end of the nitratation. Based on this information, notice that an interesting optimization strategy is to switch to the next phase (settle in this case) as soon as the COD and NH4 are degraded.

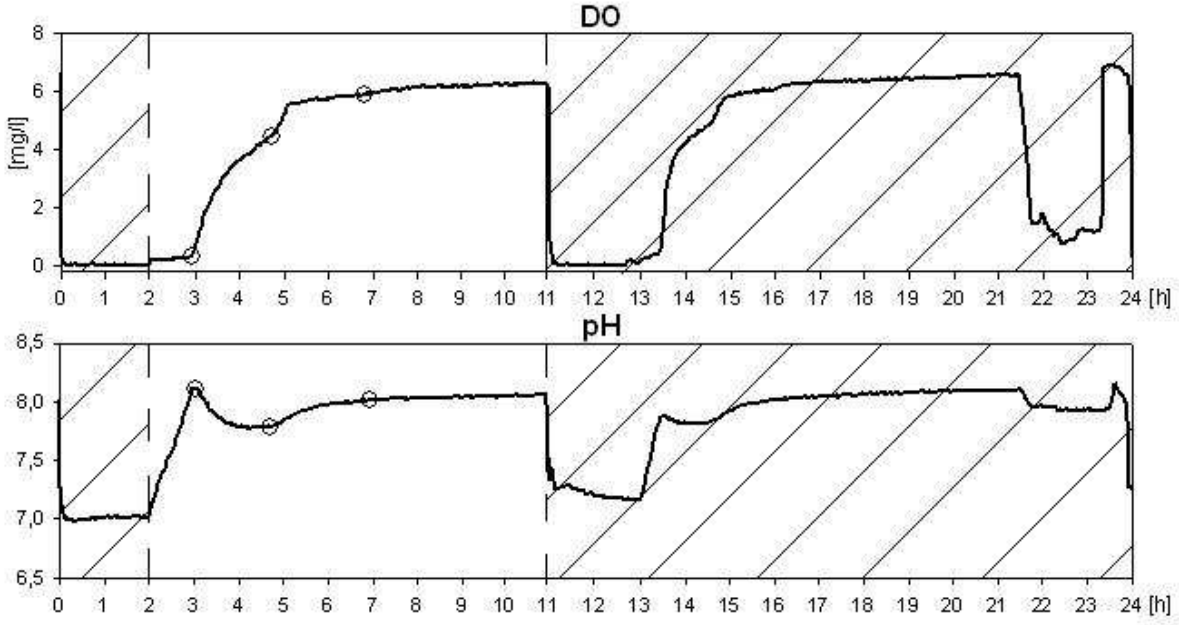

Figure 1 : *pH and Dissolved oxygen curves (25/01/04)*

The detection of the critical points defined hereabove in figure 1 is very useful for classifying a given cycle. These points define three distinct periods : the first one comprised between the beginning of the aerobic period and T1, the ending time of the COD removal (between 2 and 3 hours in figure 1), the second comprised between T1 and T2 defining the end of the nitritation (between 3 and 5 hours in figure 1) and the third one comprised between T2 and T3 defined as the end of the nitration (between 5 and 7 hours in figure 1). In particular, using the temperature to compute the oxygen saturation, notice that it is possible to determine the quantity of oxygen respectively needed in each of these periods (noted H1, H2 and H3). While the Ti parameters give us immediate information about the degradation rates, the Hi parameters are proportional to pollutant quantity. Of course, these values are not exact values since biological reactions evolve simultaneously (for instance both the COD removal and the nitrification begin when the dissolved oxygen differs from zero), however, the oxygen

consumption rates significantly differ between the different phases in order this uncertainty to be neglected.

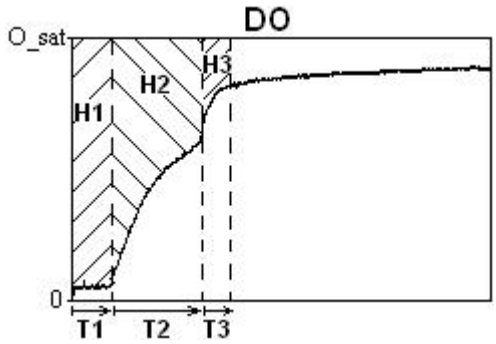

Figure 2 : *Definition of parameters Ti and Hi*

The Ti and Hi parameters are then used as descriptors in the framework of a Principal Component Analysis (PCA) approach. In the proposed software, two different approaches are proposed. The first one is a model-based approach where the projection basis for the PCA procedure is calculated using data generated by a mathematical model of the SBR plant. This approach is particularly interesting because it allows to simulate - in an acceptable time - a large number of simulations corresponding to a large spectrum of environmental conditions (different C/N ratios, different oxygen transfer rates, heterotrophic and autotrophic biomass concentrations, etc). However, it also means that a validated model of the plant is available : such a model is usually difficult to obtain and its validation is time consuming. The other available approach makes use of a limited number of real cycles for which the functioning conditions are well known provided that the available data have been obtained in a large range of functioning conditions. This approach has the advantage of being rather more robust than the model based approach because it does not use any model of the process but only the available on line measurements.

MODEL BASED APPROACH

The model used for the simulations is a modified ASM1 in which the nitritation step has been added. The data to be used in the PCA procedure are generated using this model. A large number of simulations are then run around a standard cycle defined by the following operating parameters : Kla in aerobic and in anoxic conditions, concentrations of COD and NH4 in the input, concentrations of biomass (XH and XA) in the tank, volume of the input and initial volume of the tank, duration of anoxic and aerobic periods, duration and frequency of air pulses during the anoxic phase, parameters Ti imposed as "optimal", parameters successively used to vary the conditions of the simulations (Kla, CODin, NH4in, XH, XA), ranges of parameters variation.

In order to be as open and as generic as possible, the software leaves the possibility to automatically adapt the values of the maximum growth rates model parameters. To do so, it is assumed that the operator is able to enter into the software the theoretical values (defined as "optimal") of the duration of each cycle for the standard conditions. In the previous list, the first seven parameters are used to identify the μ_{maxH}, μ_{maxNS}, μ_{maxNB} (maximum growth rates for heterotrophic, nitrifiant and nitratant biomasses). The last two parameters are needed to compute the operating parameters for each simulation. Usually the most important process faults to detect are related to the Kla, the CODin and/or the NH4in. The operator can define the range of variations of these parameters from their standard value. The possible variations of a parameter are:

- << : very low (e.g. -50% of nominal value);
- < : low (e.g. -25% of nominal value);
- > : high (e.g. +25% of nominal value);
- >> : very high (e.g. +50% of nominal value).

After the compilation and the execution of all the simulations, a database is created and will be used for the next step : the application of a PCA to reduce the space of analysis from six parameters to less.

PCA STUDY

From now, the procedure is the same for the model-based approach and for the data-based approach. The Principal Component Analysis (PCA) is one of the most popular methods for extracting information from data, and has found application in a wide range of disciplines : data rectification in chemical process operation and control (Kramer and Mah, 1994), disturbance detection and isolation (Ku et al., 1995), statistical process monitoring (Kresta et al., 1991; Wise et al., 1990), and fault diagnosis (MacGregor et al., 1994; Dunia et al., 1996). PCA transforms the data matrix in a statistically optimal manner by diagonalizing the covariance matrix and by extracting the cross correlation or relationship between the variables in the data matrix. If the measured variables are linearly related and are contaminated by errors, the first few components capture the relationship between the variables, and the remaining components are comprised only of the errors. Thus, eliminating the less important components reduces the contribution of errors in the measured data and represents it in a compact manner. The popularity of PCA relies on its high ability to reduce the dimension of any data matrix while capturing the underlying variations and relationships between the related variables.

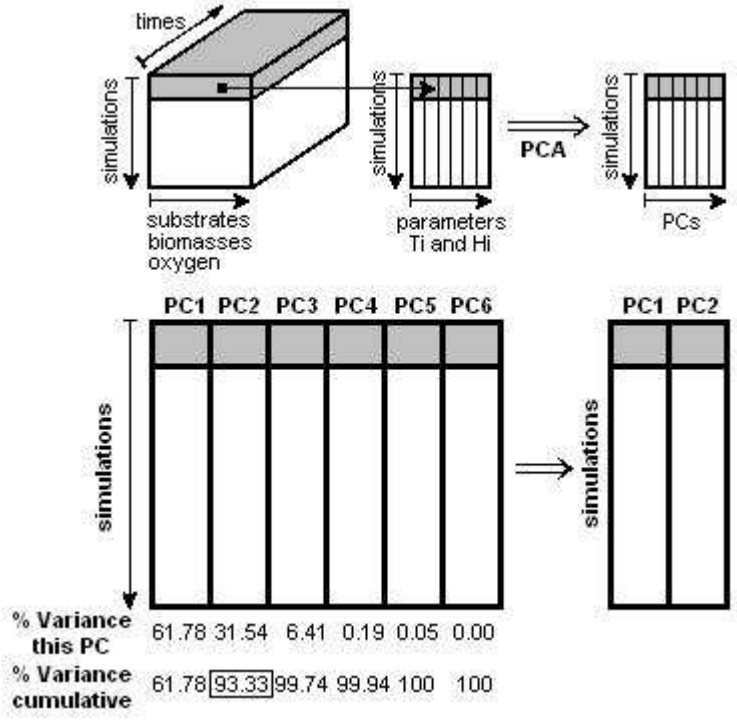

Figure 3 : *Reduction of the data matrix by a PCA*

In our particular case, the data matrix includes the Ti and Hi of all the available experiments (for the data based approach) or simulations (for the model based approach). Once the projection basis has been computed, the original matrix (six parameters for each considered cycle) is transformed into another matrix in which the columns are the principal components. In our case (in both the model based approach and in the data based approach), it was noticed that more than 90 % of the data can be explained in only looking at what happens in the first two principal components. In other words, most of the time, it will be possible to discriminate two different cycles in projecting its six associated descriptors Ti and Hi into on a plan.

THE DECISION TABLE FOR THE DIAGNOSIS

Once the characteristics of a cycle have been projected onto the discriminating plan P via the PCA, the next step is to identify a possible fault affecting the process. It is clear that if two cycles are very

similar (similar Ti, Hi), then their parameters projected onto P will be very close one from each other. Then, once projected, the underlying idea is to search for the closest points. Since this plan has precisely been designed in order to discriminate different data, it is expected that two points that are close reflect the same process behavior. Assuming experiments or simulations used for computing the Principal Components (coordinates that are used to project any new parameters extracted from a given cycle) have covered a range sufficiently large of functioning conditions, it becomes possible to associate the state of a specific point to the state of the closest point.

At this step, a number of decision algorithms can be used. One of the most popular procedure is the k-closest neighbors that consists in associating the actual point to the majority class of its k-closest neighbors. If this algorithm is particularly attracting in a supervised context (all points have been associated to a fixed number of predefined known number of classes), its use is less obvious in the present case. Indeed, only the point corresponding to standard conditions can be classified in a "normal class". The other points correspond to specific conditions that differ from the standard conditions by several parameters. The way we have proceeded is as follows. A parameter x is first defined. If there is a point defined as "normal" (corresponding to the standard conditions) in the vicinity x of the point p, then the point p is assigned to be itself in this set. If it is not the case, a decision table (such as the one pictured in figure 4) is built. In each column, the sum of the inverse of the squared distances to all classified points of the plan are specified. The explanation corresponding to the point p corresponds to the case of the table with the greatest value. This procedure is a kind of weighted explanation table.

It should be noticed here that the developed tool is very generic and not only valid for decisions in a plan. It seems that when the process faults the user is interested in is sufficiently small – as in our case where we are only interested in identifying overloads and problems in the oxygen transfer rate - two dimensions in the discriminating space are sufficient. However, the number of PCs to be finally retained remains under the responsibility of the user : providing that a large number of experiments covering a large range of functioning conditions (including possible simultaneous combination of explanation variables) are available (or a large number of simulations can be run in the model based approach), it would be possible to identify any problem as long as its "signature" in the discriminating plan is known.

Table 1 : *Table for fault causes definition*

	Kla	COD_{in}	$NH4_{in}$	X_H	X_A
<<	0	0	0	0	0
<	0	0	0	0	0
>	0	0	0	0	0
>>	0	0	0	0	0

RESULTS AND DISCUSSIONS

To validate the approach, it was decided to proceed to a number of experiments. The different conditions tested are reported in the table 2. Both approaches (model based and data based) were investigated. For the model based approach, in order to be as close as possible from experiments, the simulation corresponding to the standard conditions were realized using the average values of the parameters obtained experimentally.

Table 2 : *Experiments time table*

DATE	*TEST*	*T1 [min]*	*T2 [min]*	*T3 [min]*	*H1 [mg/l*h]*	*H2 [mg/l*h]*	*H3 [mg/l*h]*	*PC1 sim.*	*PC2 sim.*	*problems found*	*PC1 real*	*PC2 real*
simul.	standard	75	121	130	11.6	11.9	7.4	0.38	-0.16			
15/01/04	standard	82	120	162	12.8	12.2	6.7	0.05	-0.40	OK or COD<	1.03	-0.10
16/01/04	CODin +30%	130	56	96	19.5	7.8	6.4	1.21	2.16	NH4< or COD>	0.66	1.76
17/01/04	standard	78	96	148	11.9	8.3	4.6	1.05	-0.08	OK or COD>	0.92	-0.43
20/01/04	standard	58	158	130	9.3	12.9	6.3	0.20	-0.88	OK or COD<	0.52	-1.02
21/01/04	standard	76	144	126	12.0	14.2	6.9	0.01	-0.29	OK or COD>	0.48	-0.26

22/01/04	CODin +50%	154	92	132	23.0	11.8	9.5	-0.48	1.87	COD> or Kla<	1.00	2.82
23/01/04	standard	96	90	126	14.5	10.4	7.9	0.59	0.64	OK or COD>	1.06	0.55
24/01/04	standard	82	102	130	12.5	11.0	7.8	0.55	0.16	OK or COD>	1.06	0.02
25/01/04	standard	64	110	138	9.9	11.1	8.1	0.57	-0.42	OK or COD<	1.25	-0.63
26/01/04	Kla +50%	38	110	172	5.2	6.7	5.4	1.24	-1.49	COD< or Kla>	1.58	-2.00
27/01/04	NH4 +50%	110	382	16	16.8	33.1	0.7	-2.87	-0.41	NH4> or Kla<	-4.30	0.30
28/01/04	COD -50%	34	318	38	5.5	21.0	1.4	-0.54	-1.77	NH4> or COD<	-2.43	-2.43
29/01/04	Kla -30%	134	274	46	20.3	26.7	2.4	-1.9	0.78	Kla< or COD>	-2.86	1.43
30/01/04	standard	76	136	138	12.2	13.5	7.0	0.03	-0.36	OK or COD<	0.69	-0.25
31/01/04	standard	81	154	126	13.7	17.9	7.9	-0.59	-0.23	OK or NH4>	0.31	0.12
01/02/04	standard	89	110	114	14.7	14.7	8.2	0.15	0.51	OK or COD>	0.63	0.49

Up to this date, only a few cycles are available for validation. Thus, the presented results may rather be considered as a demonstration instead of a the validation of an operational tool. Nevertheless, both for the model based and for the data based approaches, the results are given in Figure 4. In both of the figures, the x axis is PC1 and the y axis is PC2. The red triangles indicate the direction of each of the six parameters (the descriptors) in the space : for example if a cycle has a large T1, its projection will be near the corresponding triangle and so on. It is clear that highly correlated parameters, as T1 and H1, give the same direction. In the left figure, the stars represent all the simulations realized with different conditions automatically generated by the software. The green zone defines the class of "standard cycles". The blue diamonds correspond to the experiments reported in Table 2 used to demonstrate the capabilities of the system to detect and isolate process faults. All 10 standard cycles were correctly classified in the green zone. For the six remaining ones, three were correctly identified while the problems corresponding to the three others were given as the second possible explanation instead of the first one :

- cycle 16/01 : the most probable problem detected is NH4< instead of COD> because T1 is large but T2 is very small ;
- cycle 26/01 : the most probable problem detected is COD< instead of Kla> because T1 is small but T2 and T3 are similar to values of standard cycles ;
- cycle 28/01 : the most probable problem detected is NH4> instead of COD< because T1 is small but T2 is very large.

The most probable explanation of this phenomenon is the particular composition used to artificially increase the nitrogen content of the input : while the input COD is only due to the raw whey, in standard conditions half of the nitrogen content comes from the whey and half of it comes from added Urea. The difference between these two forms of nitrogen is that nitrogen in the whey is probably easily degradable while urea need to be ammonified before the nitrification can take place.

In the right graph, only the real cycles reported in the previous table are represented. To illustrate the data based approach approach, only seven of the standard cycles were used in the learning phase while three were kept for the validation. The colored circles near the origin are the first seven standard cycles while the colored stars are the tests corresponding to functioning conditions differing from the standard ones. The black stars are the last three standard cycles. Their projections onto the plan are all located very close to the origin. In other words, they were correctly identified as standard cycles.

CONCLUSIONS

In this paper a new software for analyzing the state of aerobic SBR has been presented. The classification of on-line data acquired from the plant is realized using a standard PCA procedure together with a decision algorithm. Depending on the available knowledge and data, two distinct approaches (a model based and a data based) were proposed. Using a set of real experiments, both approaches have been validated. Both were able to detect and isolate overload problems. It is also shown that the application of this method of analysis can support in an determinant way the decisions of the operator in the case a process fault occurs in the treatment cycle. In this case, the most probable

causes are provided to the user in order for him to rapidly take adequate preventing or correcting actions to restore standard conditions.

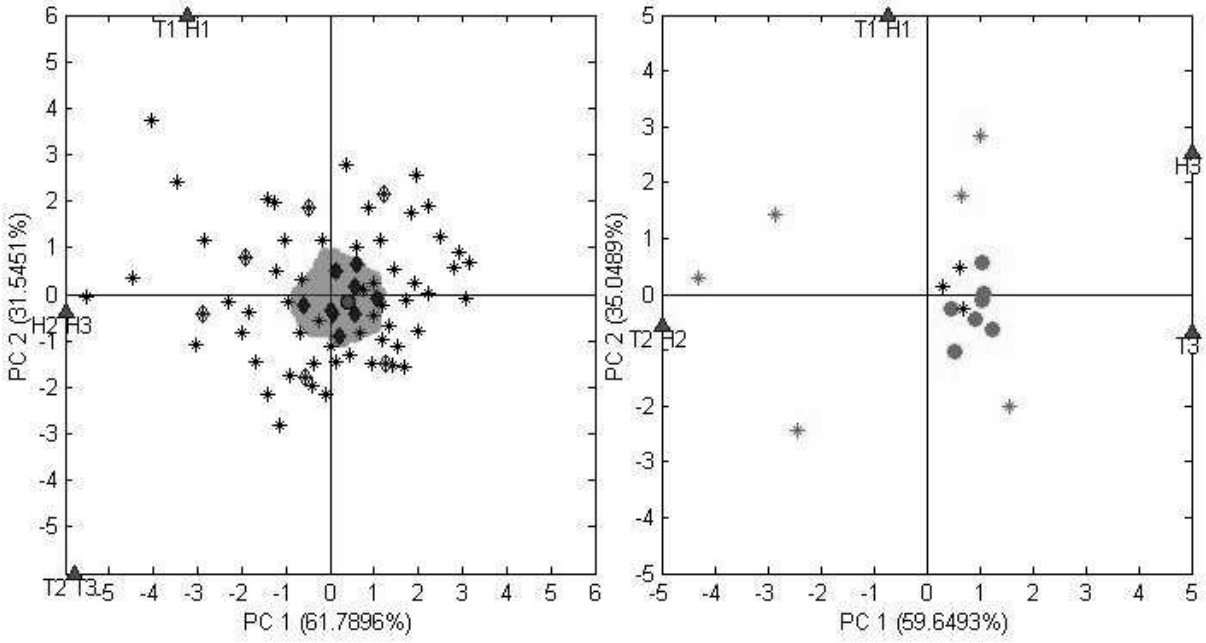

Figure 4 : *Model based and data based approaches*

ACKOWLEDGEMENTS

This paper includes results of the EOLI project that is supported by the INCO program of the European Community (Contract number ICA4-CT-2002-10012). Thank you to Jean Philippe Steyer for fruithful comments about this work.

REFERENCES

Bakshi, B.R. (1998). Multiscale PCA with Application to Multivariate Statistical Process Monitoring. *AIChE Journal, 1998*.

Cho, H.W. and Kim, K.J. (2002). Fault Diagnosis of Batch Processes: a new statistical approach using discriminant model. *2002 QSR Best Student Paper Award*.

Dochain, D. and Vanrolleghem, P.A. (2001), Dynamical Modelling and Estimation in Wastewater Treatment Processes.

Jeppsson, U. (1996). Modelling Aspects Wastewater Treatment Processes, PhD Thesis, Lund University of Technology, Sueden.

Lee, D.S. and Vanrolleghem, P.A. (2002). Monitoring of a Sequencing Batch Reactor Using Adaptive Multiblock Principal Component Analysis. *2003 Wiley Periodicals, Inc.*, 489-497.

Lennox, B., Montague, G. and Marjanovic, O. (2000) Detection of faults in Batch Processes: Application to an industrial fermentation and a steel making process, Water Science and Technology.

Montague, G. and Lennox, B. (2000), Multivariate statistical process control hones a batch. *InTech April 2000*, 42-44.

Neveu, P., Lardon, L., Hacquart, C., Menut, L., Simon, B., Steyer, J. P. 2003. PlantML : un langage pour une gestion répartie des bioprocédés de dépollution. *3ème colloque S&E'2003*, Saint Etienne du Rouvray, France. p. 197-200.

Paulou, L. , Harmand, J. and Cadenat, V. (2003), Analyse et classification de spectres SSCP, Internal report (in french, INRA-LBE, France.

Wilderer, P.A., Irvine, R.L. and Goronszy, M.C. (2000), Sequencing Batch Reactor Technology, IWA Technical Report.

Wise, B.M. and Gallagher, N. B. (2000), PLS_Toolbox 2.0 for MATLAB, Users Guide.