DETECTION OF ABNORMAL CONDITIONS OF AN AEROBIC SBR PROCESS USING FEATURE EXTRACTION

D. Wimberger and C. Verde Rodarte, IINGEN, UNAM, México
Corresponding Author: D. Wimberger
Instituto de Ingeniería, Universidad Nacional Autonoma de México
Edificio 12, Circuito Exterior, Coyoacán 04510 México DF
Phone: +52-(55)-56233683, Fax: +52-(55)-56233681
email: dwimbergerm@iingen.unam.mx

Abstract. Process fault diagnosis allows a control system to maintain the operation of a process under presence of faults. This is undoubtedly a critical feature for an aerobic SBR activated sludge waste water treatment (WWT) process treating waste water contaminated with organic toxic compounds, as in the case of the presented study. A commonly employed methodology for diagnosis based on feature extraction from process signals and their classification is demonstrated on the basis of the respiration rate signal, a known indicator for biological activity in aerobic activated sludge WWT processes. Given the absence of sufficient historical data and diagnostic information, an uncertain mathematical model was used for generation, which represents a transformation of the model that has been selected, identified from laboratory experiments and reported in the EOLI project. The usefulness of the signal for the detection and classification of a set of defined abnormal conditions was verified through sensibility analysis. Results from the analysis not only show the effects of parameter deviations, but also indicated the features to be extracted from the respiration signal for a successful classification. Consequently a feature extraction and classification procedure was designed, to diagnose the formerly defined process faults. Results obtained with the procedure indicate that the methodology can successfully cope with uncertainties. These are very common in this type of bioprocesses and prevent the straightforward application of well studied quantitative PFD approaches.

1. Introduction

Sequencing Batch Reactors (SBR) for waste water treatment (WWT) based on the fill-and-draw principle [7] are commonly employed to deal with variations in flow and composition of residual waters [11]. For autonomous operation, these batch processes require a control system that maintains the operation even under presence of faults. To achieve this goal, the process has to be permanently monitored to detect the presence of faults, diagnose their type and size [9] and take corresponding actions.

Different authors [17, 18, 19, 5, 12] have presented a large variety of methods for process fault diagnosis (PFD) with examples from different areas. Nonetheless, given the uncertainties of the quantitative models used to describe WWT processes, recent publications [14, 13, 2, 15] indicate a tendency towards application of methods from the area of Artificial Intelligence (AI) and statistics for the detection and diagnosis of process faults. Many of these data based PFD methods make use of feature extraction and classification, where the diagnosis is obtained from a classification model that is created from knowledge extracted from historical process data and diagnostic information. The difference between the various methods is marked by the difference in type and obtainment of features, classification models and algorithms.

Generally the procedures involve two parts. The first one is carried out offline and aims at producing a classification model, that can either be constructed by experts, based on statistics, or obtained using machine learning techniques. The second part is the online application of the classifier for the diagnosis of the process behaviour and possible faults. A visualization of this methodology is presented in Figure 1.

The main requirement of data based PFD is that data is available and accompanied by diagnostic information. While in many cases historical data from operations is the main source, it is also possible to generate data and diagnostic information using an uncertain mathematical model. An example of this approach is presented in this paper, that is concerned with data based PFD of an aerobic SBR WWT process treating 4-Chlorophenol (4-CP, an inhibitory organic toxic compound) contaminated waste water. Given the absence of sufficient historical data, a process model that has been selected and identified from laboratory experiments and reported in the EOLI project [3] is introduced and transformed so that a continuous online estimate of the respiration rate during the SBR reaction phase with feeding can be obtained. This process signal has been widely recognized as an indicator for biological activity and employed for monitoring and control purposes [16].

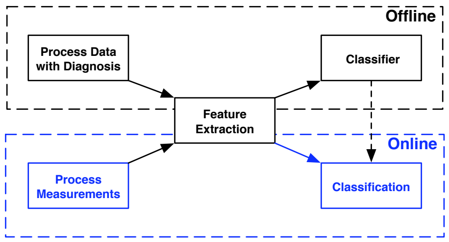

Figure 1: Data Based PFD Methodology

The usefulness of the signal for the detection and classification of a set of defined abnormal conditions (eg. process faults) is verified through sensibility analysis [8]. Results from this analysis not only show the effects of parameter deviations, but also indicate the features that should be extracted from the respiration signal for a successful classification. Consequently a feature extraction and classification procedure is proposed, that is designed to diagnose the formerly defined process faults. The paper concludes with results for diagnosis under conditions affected by uncertainties that are very common in bioprocesses and usually do prevent the straightforward application of well studied quantitative PFD approaches.

2. Methodology

Model and Transformation The EOLI project provided a validated model named EM1 [4], that is based on concentrations and describes the dynamics of a 4-CP contaminated water treatment process through the following set of equations:

$$\begin{aligned} \frac{dX}{dt} &= \mu X - \frac{q_{in}}{V}X \\ \frac{dS_C}{dt} &= -k_1\mu X + \frac{q_{in}}{V}(S_{C,in} - S_C) \\ \frac{dS_O}{dt} &= -k_2\mu X - bX + \frac{q_{in}}{V}(S_{O,in} - S_O) + K_La\,(S_{O,sat} - S_O) \\ \frac{dV}{dt} &= q_{in} \end{aligned} \tag{1}$$

where $X, S_C, S_O, \mu, q_{in}, V, k_1, k_2, b, S_{C,in}, S_{O,in}, K_La$ and $S_{O,sat}$ represent the concentrations of biomass, organic matter and dissolved oxygen, the specific growth rate, the inlet flow rate, the volume, the yield coefficients, the endogenous respiration kinetic constant, the inlet organic matter and dissolved oxygen concentrations, the transfer coefficient and the oxygen saturation concentration, respectively.

The specific growth rate for the modelled single aerobic growth reaction is represented by the following Andrews model [1]:

$$\mu = \frac{\mu_0 S_C}{K_S + S_C + \frac{S_C^2}{K_I}} \tag{2}$$

However, when the volume is not assumed constant, the nonlinear dynamics of the dilution term $\frac{q_{in}}{V}$ presents difficulties which prevent a straightforward application of a linear solution. Nonetheless, the complete model can be transformed into a mass balance model using the facts that $Concentration = \frac{Mass}{Volume}$ and that a measurement is available that allows to derive the actual volume in the reactor tank V(t):

$$\begin{aligned}
\frac{dm_x}{dt} &= \mu(V, m_s)\, m_x \\
\frac{dm_s}{dt} &= -k_1\mu(V, m_s)\, m_x + q_{in}c_{s,in} \\
\frac{dm_o}{dt} &= -k_2\mu(V, m_s)\, m_x - bm_x + q_{in}c_{o,in} + K_{la}(m_{o,sat}(V) - m_o) \\
\frac{dV}{dt} &= q_{in}
\end{aligned} \tag{3}$$

where $m_x, m_s, m_o, V, \mu(V, m_s), k_1, q_{in}, c_{s,in}, k_2, b, c_{o,in}, K_{la}, m_{o,sat}(v)$ are the biomass, the substrate mass, the dissolved oxygen mass, the liquid phase volume, the growth rate, the substrate to biomass conversion coefficient, the inflow rate, the concentration of substrate in the inflow , the oxygen to biomass conversion coefficient, the endogenous respiration coefficient, the dissolved oxygen concentration in the inflow, the oxygen mass transfer parameter and the dissolved oxygen mass at saturation, respectively. The growth rate is described by a volume dependent form of Eqn. 2.

$$: \mu(V) = \frac{\mu_0(V)\, m_s}{K_s(V) + m_s + \frac{m_s^2}{K_i(V)}} \tag{4}$$

where μ_0, K_s, K_i are the specific growth rate, the half saturation coefficient and the inhibition coefficient respectively. The respiration rate is function of μ and m_x and consists of a metabolic and an endogenous part:

$$r_m(t) = -k_2\mu(V.m_s)\, m_x - bm_x \tag{5}$$

The parameters and coefficients from original model can be used without change, assuming that $V(t)$ is available at any time through a measurement. Based on the mass balance model, an estimate of the respiration rate can be obtained with a simple extension of the observer presented in [20].

Definition of Faults An essential part in the formerly presented process model is the metabolic respiration, that is governed by the growth rate model for inhibitory substrates (Eq 2). Deviations in the half saturation parameter K_s and the inhibition parameter K_i have a strong and direct impact on the growth rate and as a consequence on the process dynamics. For this reason, deviations from their nominal values $\pm\epsilon\%$ (to take into account uncertainties in the identification) are defined as process faults.

Fault detection and/or diagnosis will be represented using the following sets of classes:

1. $\{$*normal, abnormal*$\}$, representing detection only
2. $\{K_i$, K_s, *normal*$\}$ representing detection and diagnosis of the parameter that changed.
3. $\{K_{i,low}$, $K_{i,high}$, $K_{s,low}$, $K_{s,high}$, *normal*$\}$ representing detection and diagnosis of parameter that changed as well as the direction of the change.

where *normal* stands for normal behavior, *abnormal* for a change in one of the parameters, $K_{s|i}$ for a change in the corresponding parameter and $K_{s|i,low|high}$ for a change in the respective parameter and the direction of change.

Respiration Rate Signal Features and Extraction In the presented model (Eq. 3), changes in most of the parameters have a direct influence on the respiration rate signal. For the half saturation parameter K_s and the inhibition parameter K_i this is visualized in Figure 2.

The uncertainties in the model considered as most important are the initial biomass $m_x(0)$ and the substrate concentration in the inflow $c_{s,in}$, both of which cannot be measured very exact or online. In this context, sensibility analysis as presented by [8] can be used to study the qualitative effect of small deviations in parameters and initial conditions. The time dependent sensibility function of a variable y for a parameter deviation is defined as follows

$$\sigma(t, \alpha_0) \triangleq \frac{\partial y(t, \alpha)}{\partial \alpha}\ |_{\alpha=\alpha_0} \tag{6}$$

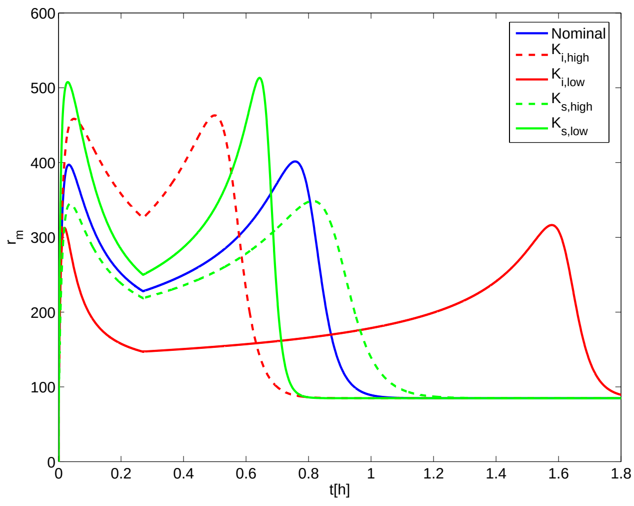

Figure 2: Respiration Rate Signals

where α is the parameter and α_0 is the nominal value of the parameter.

Analogically one can obtain the sensibility function of a variable for an initial condition. Figure 3 presents results of a numerical sensibility analysis performed with [10, 6], that represent the normalized sensibility functions $\overline{\sigma}$ of deviations of $m_x(0), c_{s,in}, K_s$ and K_i with respect to the respiration rate r_m:

$$\overline{\sigma} = \frac{\sigma(t, \alpha_0)}{y(t, \alpha)} \tag{7}$$

These functions indicate not only that the effects of variations in these parameters are be distinguishable, but also the features that should be extracted from the respiration signal.

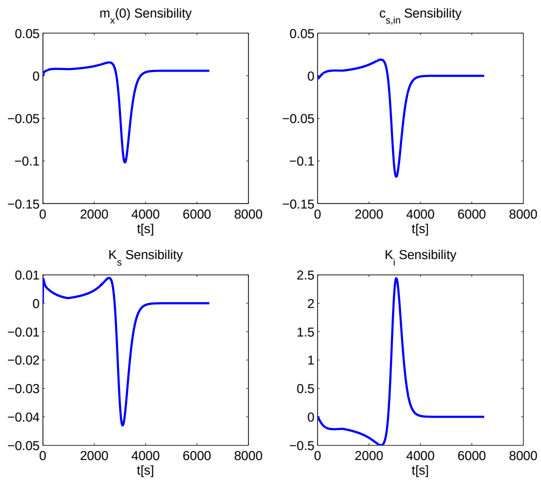

Figure 3: Normalized Sensibility Functions of r_m

Based on the indications from Figures 2 and 3 the features proposed to be extracted from the respiration signal are

- $r_{max,1}$ and $r_{max,2}$ the two maxima

- r_{min} the minimum between the two maxima
- The properties of the triangle $(a, b, c, \alpha, \beta, \gamma)$ defined by $r_{max,1}, r_{max,2}$ and r_{min}
- r_e the endogenous respiration estimate $(= bm_x)$
- $r_{tot}$the total metabolic respiration
- $t_{max,1}$the time until $r_{max,1}$

These are partially indicated in Figure 4.

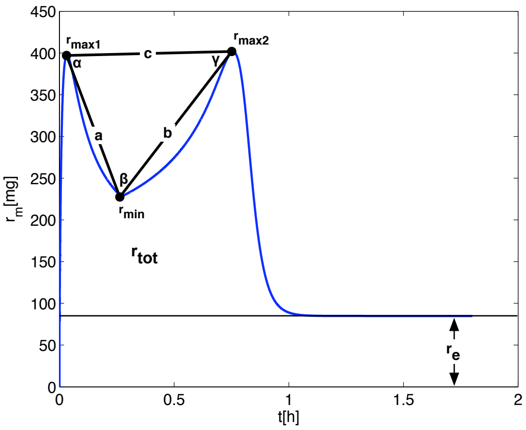

Figure 4: Features of the Nominal Respiration Rate Signal

A corresponding relatively simple feature extraction algorithm has been implemented in MATLAB.

Fault Diagnosis based on Uncertain Mathematical Model The presented mathematical mass balance model of the WWT SBR process can be used for the generation of respiration signal data with diagnostic information. To take into account the uncertainties in initial conditions and inflow concentrations, a Monte Carlo method is employed, that consists of taking uniform random samples from intervals [value $\pm\epsilon\%$] for the uncertain parameters or initial conditions $(c_{s,in}, m_x)$.

For each condition (constant and sampled parameters and initial values), the following sets of data are produced:

1. Changes of $\pm 50\%$ in steps of 10% in the parameters K_s and K_i (separately and not concurrently), which are tagged with the corresponding diagnostic class that indicates the type of fault.
2. To take into account uncertainties in the identification of the parameters, changes of $\pm 5\%$ in steps of 1% (separately and not concurrently), which are tagged with the diagnostic class that indicates normal behavior.

Data sets with diagnostic information are generated and accumulated for a given number of randomly sampled conditions and then used further in the offline part of the procedure, as input for two well known machine learning algorithms for pattern recognition, provided by the WEKA Machine Learning Toolkit [21]:

1. *J48* an implementation of the C4.5 algorithm for machine learning of of decision trees. This algorithm requires few time for learning and has the advantage that it's knowledge is explicit.
2. A Multilayer perceptron (MLP) algorithm, which constructs a simple neuronal network and trains it by means of backpropagation. The number of input nodes, output nodes and of nodes in the single hidden layer is equal to the number of features, the number of diagnostic classes and a fixed ratio between the latter two respectively.

The resulting procedure is a specific example of the generalized methodology described formerly and is visualized in Figure 5.

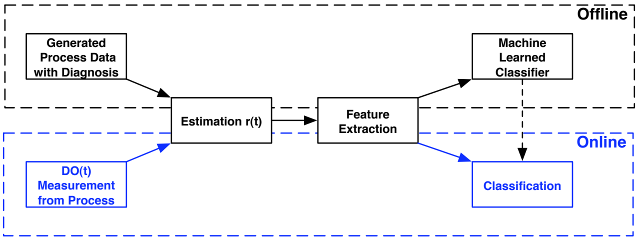

Figure 5: Diagnostic Procedure Overview

3. Results

The method proposed for fault diagnosis based on the uncertain mathematical model, has been verified with a set of simulations. These take into account uncertainties of $\pm 25\%$ in $m_x(0)$, the initial biomass and $c_{s,in}$ the substrate concentration in the inflow, as well as an error of 5-10% in the identification of the parameters K_s and K_i. The inflow rate q_{in} as well as all other parameters have been assumed to be known and constant.

To obtain classification results for feature sets that have not been used for training, the complete generated data set has been randomly permuted and split into training ($\frac{2}{3}$) and test ($\frac{1}{3}$) sets. The former has been used as input for the machine learning algorithms that yield the classification models for the detection (2-class) and diagnostic cases (3-,5-class). The latter has been used to obtain the classification results for unknown feature sets that have not been used for training. This ensures that the results are representative for an online use of the generated classification models. A comparison of the classification results for the two types of classifiers and the three sets of diagnostic classes is presented in Figure 6:

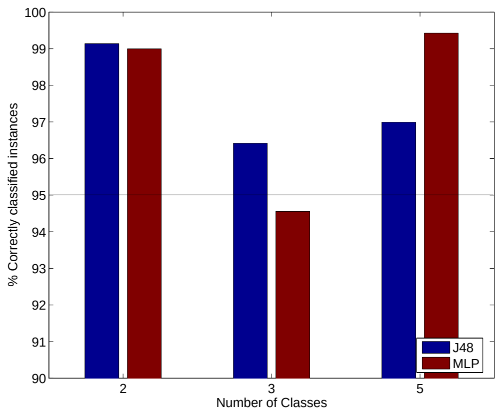

Figure 6: Comparison of Classification Results

4. Conclusion

A PFD procedure based on a feature extraction and classification methodology is presented, that uses an uncertain mathematical model for the generation of the absent historical process data and associated diagnostic information.

The employed model describes the corresponding aerobic SBR process that treats waste water contaminated with 4-Chlorophenol, an inhibitory organic toxic compound. In a first step the original concentration based model is transformed to a mass balance model that allows to estimate the respiration rate during the feed and reaction phase of the process. Variations in the respiration rate signal due to variations in parameters (which are defined as process faults) and the initial condition $m_x(0)$ are evaluated through the application of sensibility analysis.

The results from the analysis not only yield that the respiration rate is a good indicator, but also indicate the features that should be extracted from the signal. Consequently, the extracted features are employed to construct classification models through well known machine learning algorithms for pattern recognition. This part of the procedure takes into account parameter and initial condition uncertainties (biomass and substrate concentration in the inflow) through the use of a Monte Carlo type approach during the generation of data sets through simulation.

Results show that despite of significant uncertainties the methodology is quite successful, given that the percentage of correct detection and diagnosis is above 95% in almost all cases. This indicates that the machine learned classification models can be used online for the detection and diagnosis of certain process faults, which cannot be detected by a straightforward application of well studied quantitative PFD approaches.

5. Acknowledgements

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).

6. References

[1] J.F. Andrews. A mathematical model for the continous culture of microorganisms utilizing inhibiting substrates. *Biotechnol. Bioeng.*, 10:707–723, 1968.

[2] J.A. Baeza, E.C. Ferreira, and J. Lafuente. Knowledge-based supervision and control of wastewater treatment plant: a real-time implementation. *Water Science and Technology*, 41(12):129–137, 2000.

[3] M. Betancur, D. Dochain, and H. Fibrianto. Eoli workpackage 2: Model selection and parameter identification. Technical Report Deliverable D2.3: Validated Model, European Community Research, 2004.

[4] M. Betancur, D. Dochain, and H. Fibrianto. Validated model. EOLI Deliverable 2.3, European Community Research, 2004.

[5] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer Verlag, Berlin Heidelberg, 2003.

[6] John W. Eaton. *GNU Octave Manual*. Network Theory Limited, 2002.

[7] U.S. EPA. *Wastewater Technology Fact Sheet - Sequencing Batch Reactors*. EPA 832-F-99-073, 1999.

[8] Paul Frank. *Empfindlichkeitsanalyse dynamischer Systeme: eine einfuehrende Darstellung*. Oldenburg Verlag GmbH, Muenchen, 1976.

[9] R. Isermann and P. Ballé. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Eng. Practice*, 5(5):709–719, 1997.

[10] Jorge R. Leis and Mark A. Kramer. The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Trans. Math. Softw.*, 14(1):45–60, 1988.

[11] S. Mace and J. Mata-Alvarez. Utilization of sbr technology for wastewater treatment: An overview. *Ind. Eng. Chem. Res.*, 41:5539–5553, 2002.

[12] R. J. Patton, P. M. Frank, and R. N. Clark, editors. *Issues of Fault Diagnosis for Dynamic Systems*. Springer Verlag, London, 2000.

[13] I. Rodriguez-Roda, M. Sànchez-Marrè, J. Comas, J. Baeza, J. Colprim, J. Lafuente, U. Cortés, and Poch M. A hybrid supervisory system to support wwtp operation: implementation and validation. *Water Science and Technology*, 45:289–297, 2002.

[14] Montserrat Rubio, Joan Colomer, Magda L. Ruiz, Jesús Colprim, and Joaquim Meléndez. Qualitative trends for situation assessment in sbr wastewater treatment process. In *BESAI 4th ECAI Workshop on Binding Environmental Sciences and Artifical Intelligence*, Valencia (Spain), August 2004.

[15] R. Sanguesa and P. Burrell. Application of bayesian network learning methods to waste water treatment plants. *Applied Intelligence*, 13:19–40, 2000.

[16] H. Spanjers, P.A. Vanrolleghem, G. Olsson, and P.L. Dold (eds.). Respirometry in control of the activated sludge process:principles. Technical Report No.7, International Water Association, 1998.

[17] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers and Chemical Engineering*, 27:293–311, 2003.

[18] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies. *Computers and Chemical Engineering*, 27:313–326, 2003.

[19] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers and Chemical Engineering*, pages 327–346, 2003.

[20] D. Wimberger and C. Verde Rodarte. Online monitoring of an aerobic sbr process based on dissolved oxygen measurement. In *Proceedings of the IFAC World Congress (IFAC 2005)*, 2005.

[21] I. H. Witten and E. Frank. *Practical Machine LearningTools and Techniques with Java Implementations*. Morgan Kaufmann Publishers, 2000.