Reducing the Anaerobic Digestion Model N°1 for its application to an industrial wastewater treatment plant treating winery effluent wastewater

Carlos García-Diéguez [1], **Olivier Bernard** [2], **Enrique Roca** [1, *]

[1] USC – PRODES. University of Santiago de Compostela, Department of Chemical Engineering/School of Engineering, Sustainable Processes and Products Engineering and Management Research Group. Rua Lope Gomez de Marzoa s/n, 15705 Santiago de Compostela, Spain.

[2] INRIA. Sophia-Antipolis Cedex, 06902 France (e-mail: olivier.bernard@inria.fr).

[*] Corresponding author. Tel +34 881816774. Fax +34 881816702. E-mail: enrique.roca@usc.es

Abstract

The Anaerobic Digestion Model N°1 (ADM1) is a complex model which is widely accepted as a common platform for anaerobic process modeling and simulation. However, it has a large number of parameters and states that hinders its calibration and its use in control applications. A principal component analysis technique was extended and applied to simplify the ADM1. The method shows that the main model features could be obtained with a minimum of two reactions for winery effluent wastewater. A new reduced stoichiometric matrix was identified and the kinetic parameters were estimated on the basis of representative known biochemical kinetics (Monod and Haldane). The obtained reduced model takes into account the measured states in the anaerobic wastewater treatment (AWT) plant and reproduces the dynamics of the process fairly accurately. The reduction methodology presented here can be used to develop on-line control, optimization and supervision strategies for AWT plants.

Keywords

ADM1; anaerobic wastewater treatment; modeling; model reduction; PCA.

1. Introduction

Towards a sustainable society, the production of chemicals from renewable resources and particularly from waste streams is emerging as an important biotechnological application. Efforts are now being made to remove organic contaminants as far as possible without external sources of energy. For these reasons and as consequence of the application of Directive 96/61/EC, the BREF in Common Waste Water and Waste Gas Treatment recommend the use of Anaerobic Wastewater Treatment (AWT) as a best available technique (BAT) for the pre-treatment of wastewater such as those from agro-food industries with high biodegradable soluble substances (European Commission, 2003). Consequently, in the last decades, AWT has evolved into a consolidated technology for the treatment of medium to high load wastewater from agro-food industries (Gomez *et al.*, 2010, van Lier *et al.*, 2001). However, the suitable operation of an AWT plant requires constant surveillance and the targeted adjustment of different process parameters (Garcia *et al.*, 2007, Olsson *et al.*, 2005).

Anaerobic digestion process involves many interactions between species that may not all have been accurately identified. AWT has classically been regarded as a difficult process to be controlled and managed. For these reasons, modeling of AWT process has been a very active area through the last decades (Angelidaki *et al.*, 1993, Bernard *et al.*, 2001, Costello *et al.*, 1991, Dalmau *et al.*, 2010, Gavala *et al.*, 2003, Lee *et al.*, 2009, Mairet *et al.*, 2011, Mosey, 1983, Mu *et al.*, 2008, Muha *et al.*, 2012, Siegrist *et al.*, 2002, Tartakovsky *et al.*, 2008). All this knowledge about the process and its modeling has tried to be condensed into a model that is known as IWA Anaerobic Digestion Model n°1 (ADM1) (Batstone *et al.*, 2002). It is a standard benchmark for developing operational strategies and evaluating the performance of controllers (Batstone and Steyer, 2007).

However, its large number of parameters (86 parameters) and not measurable states (especially a large number of types of biomass) hinders its use in the development of control applications. The ADM1 tries to be as widely applicable as possible and therefore the price to be paid is that the mathematical analysis is very difficult, this influencing further developments in state and parameter estimation, dynamic optimization and control. Thus, a reduction of the model is considered as a reasonable alternative to simplify its complexity and facilitate the development of controllers to specific applications.

There are different methodologies that can be applied for the reduction of models. These are mainly divided in projection (such as the methods based on singular value decomposition or proper orthogonal decomposition) and non-projection based methods (such as singular perturbation). Some research works on model reduction have been focused on extending or applying these methodologies to the reduction of wastewater treatment models (Barrou et al., 2008, Chachuat et al., 2003, Steffens et al., 1997). In general, the majority of these methodologies give as result reduced models that do not take into account the variables that are measured in AWT facilities. Besides, they lead to a reduced model that reproduces properly the behavior but that does not maintain the biochemical sense of variables and parameters.

Bernard *et al.* (Bernard and Bastin, 2005a, Bernard and Bastin, 2005b) have developed a methodology, based on the projection of data through a principal component analysis (PCA). This method allows to reducing the complexity of biochemical models taking into account the measurable variables and maintaining the highest possible biochemical sense. Furthermore, this technique has proven to be effective for reducing the Benchmark Simulation Model Nº 1 (BSM1) (Barrou *et al.*, 2008), and the identification of the minimal reaction network for several biochemical systems (Bernard et al., 2006, Bernard and Bastin, 2005a, Bernard and Bastin, 2005b, Bernard and Queinnec, 2008, Helias et al., 2011).

The present paper aims to extend, to structure and to evaluate the ability of a model reduction technique as basis for the development of a simplified model with control and monitoring purposes. In this sense, a new reduced model was developed from ADM1 using a PCA based technique (Bernard and Bastin, 2005a, Bernard and Bastin, 2005b). The first step consisted of evaluating the number of reactions to be taken into account for preserving the greatest possible variability. In a second step, the new reduced *pseudo*-stoichiometric matrix was computed from the principal axis. Then, the process kinetic was determined and adjusted using already known biochemical kinetic expressions. Finally, the resulting model was validated, where the results of the different states were compared with the results of ADM1. Additionally, an alternative procedure was used to compute and to validate the *pseudo*-stoichiometric matrix derived from PCA.

2. Materials and methods

2.1 Experimental data

The input data set used in this study corresponds to experiments carried out in an anaerobic USBF reactor treating winery effluents. The main characteristics of the input, organic loading rate (OLR) and the influent concentration of chemical oxygen demand (COD_{in}), are shown in Fig. 1. The used data set corresponds to an experiment where a series of increasing influent OLRs up to destabilizing the process had been applied to an anaerobic digester (Ruiz-Filippi, 2005). These data were used as input of an ADM1 simulator and the results of the simulation were used to derive a reduced model.

Fig. 1.

2.2 IWA Anaerobic Digestion Model no. 1 (ADM1)

ADM1 (Batstone *et al.*, 2002) is a complex model of multistep anaerobic process transformations. ADM1 incorporates processes such as hydrolysis of particulates, acidogenesis, acetogenesis and methanogenesis, and it includes 26 dynamic state concentration variables, 19 biochemical kinetic processes, 3 liquid-gas mass transfer kinetic processes, 8 implicit algebraic variables per liquid vessel, and 86 parameters. A detailed description on ADM1 can be seen in the IWA Scientific and Technical Report Nº 13 (Batstone *et al.*, 2002).

A virtual plant of ADM1 in a DAE (differential and algebraic equation) implementation and checked for consistency with the benchmark version from Lund University (Rosen et al., 2006, Rosen and Jeppsson, 2006) was used to the simulations. Additionally, the virtual plant incorporates ethanol degradation pathways (Ruiz-Filippi *et al.*, 2004) and it was calibrated for the anaerobic USBF (Upflow Sludge Bed Filter) reactor which was used to obtain the experimental data.

The dynamic behavior of the anaerobic digestion process in the USBF reactor was described by the following macroscopic mass balance applied to the liquid phase:

$$\frac{d\xi}{dt} = D\xi_{in} - \alpha D\xi - Q_G(\xi) + K \cdot r(\xi) \tag{1}$$

where, ξ_{in} and ξ represent the inlet concentration vector and the state vector respectively, D is the dilution rate, $Q_G(\xi)$ quantifies the loss of mass of each specie by transfer to the gas phase. The parameter α ($0 \leq \alpha_i \leq 1$) describes the solid behavior inside of reactor measuring the deviation from the ideal mixing behavior. The term $K \cdot r(\xi)$ represents the transformation phenomena in the bioreactor where $r(\xi)$ is the (p x 1) vector of reaction rates and K is the *pseudo*-stoichiometric matrix with dimension n x p being n the number of species and p the number of reactions or processes considered in the model. Thus, each k_{ij} element of the matrix K corresponds to the yield coefficient that relates the j^{th} reaction with the concentration of the i^{th} specie. In consequence, a specie i^{th} with a positive value of k_{ij} is related to the products of the reaction j^{th} while a negative k_{ij} value refers to substrate consumption; and if $k_{ij} = 0$ the i^{th} specie is not involved in the j^{th} reaction.

2.3 Model reduction method

A complete structured procedure to reduce ADM1consisting of 4 steps was extended and applied:

1. The first step in the model reduction procedure consists of determining the number of reactions p (i.e., number of columns of the reduced stoichiometric matrix K) applying the methodology based on PCA developed by Bernard *et al.* (Bernard and Bastin, 2005a, Bernard and Bastin, 2005b). PCA was applied in combination with the appropriate mass balances, where the number of measured species n is assumed to be greater than the number of reactions p. It is described in the section 2.3.1.
2. The second step was the determination of a new *pseudo*-stoichiometric matrix (K) from the principal components (i.e., eigenvectors from PCA). It is described in detail in the section 2.3.2.
3. In the third step a state transformation was applied in order to obtain a set of new

kinetic functions and their parameters. These were modeled using well-known kinetic functions used for reproducing the reactions rates of biochemical reactions (i.e., Monod, Haldane, Contois, etc.). It is described in the section 2.3.4.

4. Finally, the accuracy of the reduced model was assessed along with its ability for fitting quantitatively the data.

A scheme that summarizes the model reduction methodology is shown in the Fig. 2.

Fig. 2.

The method was applied to obtain a reduced model for an AWT plant treating winery effluent wastewater. Nonetheless, the methodology described below could be applied to any other type of wastewater (e.g., effluents of agro-food industries) and to other biochemical models based on mass balances.

The following section explains how a new reaction network with a lower dimension than ADM1 can be obtained by means of the PCA method.

2.3.1 Dimension of the reaction network

In the reduction of biochemical models the number of reactions to be considered, is crucial for avoiding an unmanageable boost in complexity. The aim is to define the smallest number of reactions or processes (p) that can represent the process dynamic. For this task, the general mass balance (Eq. (1)) was rewritten such that the transport terms were grouped on the left side or first member in the Eq. (2) while the biochemical transformations $K \cdot r(\xi)$ were located on the right side.

$$\frac{d\xi}{dt} - v(\xi(t)) = K \cdot r(\xi(t)) \tag{2}$$

where $v(\xi)$ represents all fluxes of material in the bioreactor. Then, from Eq. (2) the following equation can be obtained, where $u(t)$ can be computed because of all transport terms are known:

$$u(t) = K \cdot r(t) \tag{3}$$

The determination of the number of *pseudo*-reactions (p) implies the identification of the dimension of the space where $u(t)$ lives. It corresponds to the principal component analysis (PCA). PCA is a statistical technique for the reduction of dimensionality in a multivariable process. PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlay it. To address this question, $u(t)$ is considered at N time instants, with $N > n$, and these vectors are gathered in a matrix $U = [u(t_1)\ u(t_2)\ \ldots\ u(t_N)]$. The number of *pseudo*-reactions (p) is then determined by counting the number of non-zero eigenvalues of UU^T (Bernard and Bastin, 2005a, Bernard and Bastin, 2005b). Nonetheless, in practice there are not zero eigenvalues for the matrix UU^T. Let us remark that the eigenvalues of UU^T correspond to the variance associated with the corresponding eigenvector (i.e., principal component) (Johnson and Wichern, 1992). Thus, the method consists of selecting the p first eigenvectors which represents a total variance larger than a fixed confidence threshold (e.g., 90%).

In order to give the same weighting to all the variables in the PCA, the data vectors $u(t_i)$

were normalized as follows:

$$\tilde{u}_i(t_k) = \frac{u_i(t_k) - \bar{u}_i}{\sqrt{N} \cdot stdv(u_i)} \tag{4}$$

where $\bar{u}_i$ is the mean value of $u_i(t_k)$ and $stdv(u_i)$ is the corresponding standard deviation.

2.3.2 Pseudo-stoichiometric matrix identification

The second step in the model reduction technique consists of determining a new *pseudo*-stoichiometric matrix K. In this sense, the family of possible *pseudo*-stoichiometric matrices K is parameterized by a matrix G according to Eq. (5). Therefore, a number of p^2 constraints must be introduced to identify K.

$$K = \rho \cdot G \tag{5}$$

From a mathematical point of view, the identification of each *pseudo*-reaction (i.e., k_j – column of K) is independent of the identification of the other *pseudo*-reactions. Hence, let us assume that p constraints are available for each j^{th} *pseudo*-reaction and then these can be identified using Eq. (6).

$$k_j = \rho \cdot G_j = g_{1,j}\rho_1 + g_{2,j}\rho_2 + ... + g_{p,j}\rho_p \tag{6}$$

Constraints can be chosen, as follows*: i)* normalizing the j^{th} reaction with respect to one specie, imposing a +1 or a -1 in some element of k_j; *ii)* assuming that a specific component is not involved in the reaction and therefore imposing a zero in the corresponding stoichiometric coefficient; *iii)* imposing the conservation of elementary mass balances; and *iv)* many other valuable assumptions based on *a priori* knowledge of anaerobic digestion process can be set up. However, let us remark that, constraints without biochemical meaning could be also applied to identify K but it should not be the preferred option.

2.3.3 Pseudo-stoichiometric matrix validation

To validate the *pseudo*-stoichiometric matrix an alternative calculation procedure was applied (Bernard and Bastin, 2005b). The idea consists in identifying a group of n-p vectors gathered in β, such that:

$$\beta^T K = 0 \quad \text{with}\, \beta \in \Re^n \tag{7}$$

Once the n-p vectors were identified, it was necessary to solve the multiple linear regressions associated to each vector β_i (refers to Eq. (8)) from where the stoichiometric coefficients were reidentified. Then, the sign of the stoichiometric coefficients and the significance of regressions were checked.

$$\sum_{i=1}^{n} \beta_i u_i(t) = 0 \tag{8}$$

2.3.4 Kinetics and parameter estimation

The third step in the model reduction methodology consists in estimating the reaction rates with respect to the biochemical species in the system. To simplify the parameter identification procedure and to limit the correlation between the kinetic parameters, each reaction was treated separately.

In order to isolate each j^{th} reaction a state transformation was applied. It was assumed that there exists a λ_j vector (with dimension n x 1) according to the following approach (Chen *et al.*, 2000):

$$\lambda_j^T \text{ s.t. } \lambda_j^T K = (0...0\,1\,0...0) \tag{9}$$

$$r_j(\xi(t)) = \lambda_j^T K\, r(\xi(t)) \tag{10}$$

Then combining Eq. (3) and Eq. (10), each reaction rate can be expressed as:

$$r_j(\xi(t)) = \lambda_j^T u(t) \tag{11}$$

Now the objective is to identify the reaction rates (r_j's) as a function of some components of the state ξ. In this sense, the reactions rates are determined by many factors but they often mainly depend on substrate and biomass concentration. However, this step is far from being trivial and different kinetics can be proposed. Here we suggest to investigate the possibility that r_j/X_j *(reaction rate /* biomass for the j^{th} reaction) would be a function of some of the substrate concentrations identified in the previous step. The substrates correspond to negative elements in k_j and the biomass species correspond to some positive elements in k_j. Thus, the procedure consists of determining whether the S_k leads to a functional relationship in the space (r_j/X_j ,S_k). Then, known kinetic models can be chosen in order to assess the adjustment of the reduced model with regard to the data. In general, for the appropriate pair X_j and S_k, the function can be chosen from the set defined by Eq. (12).

$$\frac{r_j(X_j,S_k)}{X_j}=\mu_j(X_j,S_k)=\begin{cases} Monod \\ Haldane \\ Contois \\ Other \end{cases} \tag{12}$$

Once the mappings $r_j(\xi)$ have been selected, a procedure of parameter identification is necessary. As a first step, the parameters of each reaction j^{th} can be estimated by minimizing the following objective function:

$$J_j(\theta_j)=\min\sum_{k=1}^{N}\left(\mu_{\text{model}}(\theta_j,X,S)_k-\mu_{data}(X,S)_k\right)^2 \tag{13}$$

where, θ_j represents the set of parameters associated with the kinetic model selected for the j^{th} reaction. This provides a first estimate of kinetic parameters (θ_j) which will then be used as an initial estimate for a further multi-objective optimization. The aim is to reduce the error between the data generated by the original model and the data reproduced by the reduced model. Thus, the latter optimization consists in minimizing the following objective function along the considered time instants t_k:

$$J(\theta)=\min\sum_{i=1}^{n}\sum_{k=1}^{N}\omega_i\left(\xi_{\text{model}_i}(\theta,t_k)-\xi_{data\,i}(t_k)\right)^2 \tag{14}$$

Here, ω_i is a selected weight for each state variable; ξ_{data} and ξ_{model} are the states of the original and reduced model respectively.

3. Results and discussion

3.1 Number of *pseudo*-reactions

PCA was applied to experimental data from an USBF reactor treating winery effluent wastewater and simulated ADM1 data with the aim of determining the number of *pseudo*-reactions that must be taken into account to simplify the ADM1. Depending on data availability (i.e., measured process variables), different numbers of mass balances were considered for computing the number of *pseudo*-reactions. Table 1 shows the species considered for the application of mass balances (9, 8 or 7 balances) and summarizes the number of reactions obtained in each case for a preservation of data variability of 90%. As it can be observed in the table, when 7 mass balances were considered the resulting number of *pseudo*- reactions was lesser than the other two cases. This apparent discrepancy occurred as result of the methodology used. As a general rule, a lesser number of mass balances implied a lower number of *pseudo*-reactions to be considered in the reduced model structure. In fact, according to the adopted methodology, several combinations are often possible among reaction networks of different complexity (Bernard and Bastin, 2005b) and therefore the choice of a reaction network and its associated stoichiometry matrix is usually the result of modeling assumptions, data availability and modeling objectives. Hence, modeler expertise is crucial for determining the interest of increasing the accuracy of the model by increasing its complexity.

Table 1

In the case of anaerobic digesters, only a small set of measured variables are available on-line (Spanjers and van Lier, 2006). Thus, the reduction method was applied considering only those variables that can be measured or easily estimated in AWT plants. These variables were: chemical oxygen demand (S_1), concentration of volatile fatty acids (S_2) (note that the volatile fatty acids-*VFA* are also a source of chemical oxygen demand-*COD*, however we decided to represent the *VFA* in a separate way from main COD source), methane (*CH4*), inorganic carbon concentration (*IC*), inorganic nitrogen concentration (*IN*), acidogenic biomass concentration (X_1) and methanogenic biomass concentration (X_2). Considering this set of variables, the results obtained by application of PCA suggest that a model with *p* equal to 2 reactions (i.e., a new 7 x 2 pseudo-stoichiometric matrix *K*) would be adequate to describe the behavior of an AWT plant treating winery effluent wastewater. This can be observed in Fig. 3, where the accumulated explained variance is higher than a 90% by considering only 2 reactions.. This result agrees with other scientific works, where two reactions were used to represent the experimental data issued from an anaerobic plant processing ethanol containing wastewater (Bernard et al., 2006, Bernard and Queinnec, 2008).

Fig. 3.

3.2 *Pseudo*-stoichiometric matrix

Once the number of compounds to be modeled was established (i.e., *n* equal to 7) and the number of reactions was determined (i.e., *p* equal to 2), the next step consisted in the identification of the *pseudo*-stoichiometric matrix *K* for the reduced model. As it was explained previously in the section 2.3.2, a number of p^2 constraints ($2^2 = 4$) were required to identify a new *pseudo*-stoichiometric matrix *K*. The structure of *K* is shown in the Eq. (15):

$$K = \begin{array}{c} \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \\ k_{31} & k_{32} \\ k_{41} & k_{42} \\ k_{51} & k_{52} \\ k_{61} & k_{62} \\ k_{71} & k_{72} \end{pmatrix} \end{array} \begin{array}{c} S_1 \\ S_2 \\ CH_4 \\ IC \\ X_1 \\ X_2 \\ IN \end{array} \tag{15}$$

Two constraints by reaction were proposed based on *a priori* knowledge of anaerobic digestion process. Constraints that preserve as much as possible the biochemical meaning were preferred over any other possibilities. Thus, the first column or first reaction was identified assuming that it corresponds to the acidogenic step and therefore the following constraints were considered:

- The reaction was normalized with regard to the acidogenic biomass (X_1) and then a value of +1 was assigned to the stoichiometric coefficient k_{51} (refers to Eq. (15)).
- It was assumed that the solubility of methane is negligible and also that it is only produced in the methanogenic step (i.e., $k_{31} = 0$).

In a similar way, two constraints were used to identify the second column (see Eq. (15)) or 2th *pseudo*-reaction, considering it equivalent to the methanogenic step of the anaerobic digestion process:

- The second reaction was normalized with regard to the methanogenic biomass (X_2); this implies that a value of +1 was imposed for the stoichiometric coefficient

k_{62}.

- It was considered that original source of COD cannot be directly consumed by the methanogenic biomass (i.e., $k_{12} = 0$).

Each pair of constraints associated to a column of K permitted to solve the linear system that corresponds with each *pseudo*-reaction (refers to Eq. (6)). Then, the unknown stoichiometric coefficients k_{ij} were calculated by means of Eq. (5). The obtained *pseudo*-stoichiometric matrix that results of this procedure is the following:

$$K = \left(\begin{array}{cc} -25.50 & 0 \\ 24.10 & -19.03 \\ 0 & 0.278 \\ -0.259 & 0.277 \\ 1 & 0.281 \\ -0.533 & 1 \\ -0.006 & -0.007 \end{array} \right) \begin{array}{c} S_1 \\ S_2 \\ CH_4 \\ IC \\ X_1 \\ X_2 \\ IN \end{array} \tag{16}$$

This new *pseudo*-stoichiometric matrix reduces considerably the 28 x 21 ADM1 stoichiometric matrix.

3.3 Kinetics of the Reduced Model

The functions that describe the kinetics were chosen in order to assess whether the model fits well the data. Fig. 4 shows the projected data and the kinetic fitted data according with Eq. (12).

Fig. 4.

As it is shown in Fig. 4, the reaction rate was expressed with regard to the biomasses (i.e., acidogenic biomass X_1 and methanogenic biomass X_2 concentrations), and plotted against the corresponding substrate concentration (i.e., S_1 and S_2) for each *pseudo*-reaction. A Monod kinetic function (Eq. (17)) was chosen for acidogenesis – 1st *pseudo*-reaction (see Fig. 4a) while a Haldane kinetic function (Eq. (18)) was chosen for methanogenesis – 2nd *pseudo*-reaction (see Fig. 4b).

$$r_1(X_1,S_1)/X_1=\frac{\mu_{\max,1}\cdot S_1}{K_{s,1}+S_1} \tag{17}$$

$$r_2(X_2,S_2)/X_2=\frac{\mu_{\max,2}\cdot S_2}{K_{s,2}+S_2+S_2^{\ 2}/K_{I,2}} \tag{18}$$

Initially, the parameters were identified using Eq. (13), and then these were optimized to improve the response of the reduced model (refers to Eq. (14)). The results of the parameter identification procedure are shown in Table 2.

Table 2

3.4 *pH* - electroneutrality

Electroneutrality is a principle that must be respected in biochemical processes; so the concentration of anions weighted by the number of electrical charges must be equal to the concentration of cations with the same weighting. In ADM1, electroneutrality principle is used to evaluate the *pH* in the medium (an important variable to be monitored). The *pH* is calculated using a charge balance which implies the use of an algebraic equation. This algebraic equation cannot be reduced by the PCA method and therefore it was adapted considering the state variables of the reduced model. *pH* was computed using the charge balance in Eq. (19) which is an adaptation of ADM1 charge balance.

$$HO^{-} + HCO_3^{\ -} + An^{-} + s_2^{\ -} - Cat^{+} - NH_4^{\ +} - H^{+} = 0 \tag{19}$$

Affinity constants were used for computing the dissociated chemical species involved in (19). The affinity constant for the dissociated volatile fatty acid (s_2^-) was assumed equal to acetic acid affinity constant considering that it is the major volatile fatty acid compound in anaerobic digestion process under not overloaded conditions. Cat^+ and An^- represent additional positive and negative ions respectively (e.g. Na^+, Cl^-). Fig. 5 compares the values of *pH* obtained by ADM1 and the results obtained with the charge balance given by Eq. (19). According with the results shown in this figure, the approximation used for the affinity constant seems to be reasonable.

Fig. 5.

3.5 Gas flow

Methane and carbon dioxide flows are other two important operating parameters used to monitor anaerobic processes. As the solubility of methane is very low (with a Henry constant of 0.00116 M/bar in water at 310 K - (Batstone et al., 2002)), it is possible to simplify the differential equation corresponding to the methane in the reduced model structure applying a singular perturbing simplification (Kokotovic et al., 1986; Bastin and Dochain, 1990; Dochain and Vanrolleghem, 2001). Hence, the methane concentration in the liquid phase and its time derivate were set to zero and it was assumed that there not exists mass transfer resistance between the liquid and the gas phase. Thus, the methane gas flow (Q_M) was computed using the Eq. (20).

$$Q_M = k_{32} \cdot r_2 \qquad (20)$$

where r_2 is the kinetic rate associated to methanogenesis and k_{32} is the yield coefficient of methane.

Besides, the carbon dioxide has relatively high solubility and therefore its flow rate (Q_C) was estimated using Henry's law and the following mass transfer equation:

$$Q_C = k_L a \cdot (S_{CO_2} - KH_{CO2} \cdot P_c) \qquad (21)$$

where, $k_L a$ is the overall mass transfer coefficient for carbon dioxide, S_{CO2} is the CO_2 concentration in the liquid phase, KH_{CO2} is the Henry constant for CO_2 in water at 310 K (0.0246 kg mol/m^3·bar - (Batstone et al., 2002)) and P_c is the CO_2 partial pressure in the biogas.

3.6 Model validation

3.6.1 *Validation of the macroscopic reaction network*

The macroscopic reaction network estimated by means of PCA was validated finding the *pseudo*-stoichiometric matrix (K_{val}) by the alternative method described in the section 2.3.3. In consequence, a multiple linear regression was derived for each vector β_i, according to Eq. (8). The numerical results obtained from each regression are shown in Table 3. It resulted that all the regressions were significant. Furthermore, the identified *pseudo*-stoichiometric matrix after validation (Eq. (22)) was compared with that obtained previously by PCA (see Eq. (16)). These matrices and its elements (comparing the order of magnitude and signs) are closer, which validates the estimated reaction scheme.

Table 3

$$K_{val} = \begin{array}{c} \begin{pmatrix} -23.01 & 0 \\ +21.07 & -16.63 \\ 0 & 0.253 \\ -0.214 & 0.239 \\ 1 & 0.169 \\ -0.801 & 1 \\ -0.002 & -0.008 \end{pmatrix} \end{array} \begin{array}{c} S_1 \\ S_2 \\ CH_4 \\ IC \\ X_1 \\ X_2 \\ IN \end{array} \tag{22}$$

3.6.2 Validation of the reduced model

For nonlinear models (e.g., ADM1), it is not possible to check analytically the accuracy of a model, and therefore the quality of a model is assessed by means of simulations. Thus, the main objective of model validation was to verify that the reduced model fits the data properly. Consequently, simulations with the reduced model were carried out, and the different states were compared then with the results obtained by the original model (ADM1). It is important to note that the validation phase was performed with a data set which was not used to establish or to identify the reduced model.

Fig. 6 shows the results of the simulations provided by the reduced model against ADM1 simulation data (i.e., *COD* concentration in the effluent, *VFA* concentration in the effluent and methane flow rate). The dynamic of the system is well explained by the reduced model which is based on a 7 x 2 *pseudo*-stoichiometric matrix in comparison with 28 x 21 ADM1 matrix. Thus, it is demonstrated that this very simple model is able to properly describe the behavior of the principal monitoring and control variables for an AWT plant treating winery effluents. A limitation of the reduced model is that will not be able to differentiate among the different volatile fatty acids and different microorganism populations. For instance, it cannot forecast a propionate accumulation, as this feature has not been taken into account in its development. However, the model is able to reproduce the *pH* properly (Fig. 5). Furthermore, the used reduction methodology is flexible enough for allowing the inclusion of new variables depending on modeling objectives.

Fig. 6.

These results show how ADM1 can be reduced to a much simpler one describing well enough the behavior of the modeled system for a specific type of wastewater.

Table 4 shows the number of parameters, state variables, types of biomass, *pseudo*-reactions and outputs in ADM1 against those used in the reduced model for winery effluents. From these data, it can be observed the degree of reduction or simplification with regard to ADM1. In this sense, the reduced model has a higher complexity ratio (i.e., ratio between the number of available measurements and the number of parameters) than ADM1. As a consequence, the development of controllers and estimators from the reduced model would be easier. On another hand, the reduced model could be recalibrated automatically due to the most of its state variables are measured on-line.

Table 4

4. Conclusions

In this paper, a reduction methodology was extended and applied to evaluate its ability to obtain a reduced model from ADM1 for control and supervision purposes. This methodology appeared to be a powerful tool to simplify ADM1 and models with similar structure for AWT. A reduced model with only two biochemical reactions was derived using this methodology. It was able to represent the dynamics of anaerobic digestion considering the instrumentation available at an AWT plant treating winery effluent wastewater. Future work will focus on the proposal and development of observers, software sensors and controllers based on the methodology presented here.

Nomenclature

Acronyms	**Variable**	**Units**
α	solid behavior model parameter ($0 \leq \alpha_i \leq 1$)	[-]
β	sound left kernel vectors matrix ($n-p$ x p)	

θ_j	set of kinetic parameters	
λ_j	transformation vectors (n x 1) (see Eq. (9))	
μ_j	specific growth rate	1/d
μ_{data}	specific growth rate vector related to the data (1 x N)	1/d
$\mu_{max,1}$	maximum acidogenic bacteria growth rate	1/d
$\mu_{max,2}$	maximum methanogenic bacteria growth rate	1/d
μ_{model}	specific growth rate vector related to the reduced model (1 x N)	1/d
v	fluxes of material in the bioreactor (n x 1) (see Eq. (2))	
ξ	state vector (n x 1)	
ξ_{data}	state vector related to the data (n x 1)	
ξ_{model}	state vector related to the reduced model (n x 1)	
ξ_{in}	inlet concentration vector (n x 1)	
ρ	matrix of eigenvectors (n x p)	
ρ_j	eigenvector (column j of ρ)	
ω_i	weighs during parameter optimization	
An^-	anions concentration	kg mol/m^3
Cat^+	cations concentration	kg mol/m^3
COD	chemical oxygen demand	
COD_{in}	influent concentration of chemical oxygen demand	kg COD/m^3
D	dilution rate	1/d
G	G matrix (p x p) (see Eq. (5))	
H^+	hydron concentration	kg mol/m^3
HCO_3^-	bicarbonate concentrations	kg mol/m^3
HO^-	hydroxide concentration	kg mol/m^3
J	objective function	
K	*pseudo*-stoichiometric matrix (n x p)	
k_{32}	yield coefficient for methane	kg mol-CH_4/kg COD
k_{71}	yield for inorganic nitrogen (acidogenesis)	kg mol-IN/kg COD
k_{72}	yield for inorganic nitrogen (methanogenesis)	kg mol-IN/kg COD
KH_{CO2}	*Henry's constant for CO_2 in water at 310K (0.0246 kg mol/m^3·bar)	kg mol/m^3·bar
K_I	Inhibition constant	kg COD/m^3
k_i	rows of matrix K	
k_{ij}	element i,j of the *pseudo*-stoichiometric matrix K	
k_La	liquid-to-gas mass transfer coefficient	1/d
K_s	half-saturation constant	kg COD/m^3
K_{val}	*pseudo*-stoichiometric matrix for validation (n x p)	
N	number of data set records	
n	number of states	
NH_4^+	ammonium concentration	kg mol/m^3
OLR	organic loading rate	kg COD/m^3·d
p	number of processes or reactions	
P_c	CO_2 partial pressure in biogas	bar
Q_G	mass transferred to the gas phase	kg mol/m^3·d
Q_C	carbon dioxide transferred to the gas phase	kg mol/m^3·d
Q_M	methane transferred to the gas phase	kg mol/m^3·d
q_m	methane gas flow	m^3/d

r	reaction rates vector (p x 1)	
r_1	acidogenesis reaction rate	kg COD/m^3·d
r_2	methanogenesis reaction rate	kg COD/m^3·d
S_1	organic substrate concentration (*COD* source)	kg COD/m^3
S_2	volatile fatty acids concentration	kg COD/m^3
s_2^-	ionized volatile fatty acids concentration	kg mol/m^3
S_{CO2}	CO$_2$ concentration in the liquid phase	kg mol/m^3
t_k	considered time instants	
U	set of data records $U = [u(t_1)\ u(t_2)\ \ldots\ u(t_N)]$ at N time instants (n x N)	
u	transport vector for each time instant (n x 1)	
$\tilde{u}$	transport vector normalized (n x 1)	
VFA	volatile fatty acids	
X_1	acidogenic biomass concentration	kg COD/m^3
X_2	methanogenic biomass concentration	kg COD/m^3

* Taken from (Perry and Green, 1999).

References

Angelidaki, I., Ellegaard, L., Ahring, B.K., 1993. A Mathematical-Model for Dynamic Simulation of Anaerobic-Digestion of Complex Substrates - Focusing on Ammonia Inhibition. Biotechnol. Bioeng. 42 (2), 159-166.

Barrou, O., Karama, A., Lakhal, E.K., Bernard, O., Pons, M.N., Corriou, J.P., 2008. Estimation of a reduced model of the BSM1 activated sludge wastewater treatment plant. Int. J. Chem. React. Eng. 6 (A63).

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhny, S., Pavlostathis, S.G., Rozzi, A., Sanders, W., Siegrist, H., Vavilin, V.A., 2002. Anaerobic Digestion Model No. 1 (ADM1), IWA Publishing, London.

Batstone, D.J., Steyer, J.P., 2007. Use of modelling to evaluate best control practice for winery-type wastewaters. Water Sci. Technol. 56 (2), 147-152.

Bernard, O., Bastin, G., 2005a. Identification of reaction networks for bioprocesses: determination of a partially unknown pseudo-stoichiometric matrix. Bioprocess. Biosyst. Eng. 27 (5), 293-294.

Bernard, O., Bastin, G., 2005b. On the estimation of the pseudo-stoichiometric matrix for macroscopic mass balance modelling of biotechnological processes. Math. Biosci. 193 (1), 51-77.

Bernard, O., Chachuat, B., Helias, A., Rodriguez, J., 2006. Can we assess the model complexity for a bioprocess: theory and example of the anaerobic digestion process? Water Sci. Technol. 53 (1), 85-92.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A., Steyer, J.P., 2001. Dynamical model development and parameter identification for an anaerobic wastewater treatment process. Biotechnol. Bioeng. 75 (4), 424-438.

Bernard, O., Queinnec, I., 2008. Dynamic models of biochemical processes: properties of the models, Hermes Science, Paris.

Chachuat, B., Roche, N., Latifi, M.A., 2003. Reduction of the ASM 1 model for optimal control of small-size activated sludge treatment plants - Réduction du modèle ASM 1 pour la commande optimale des petites stations d'épuration à boues activées. Revue des Sciences de l'Eau. 16 (1), 5-26.

Chen, L., Bernard, O., Bastin, G., Angelov, P., 2000. Hybrid modelling of biotechnological processes using neural networks. Control Eng. Practice. 8 (7), 821-827.

Costello, D.J., Greenfield, P.F., Lee, P.L., 1991. Dynamic Modeling of A Single-Stage High-Rate Anaerobic Reactor .1. Model Derivation. Water Res. 25 (7), 847-858.

Dalmau, J., Comas, J., Rodriguez-Roda, I., Pagilla, K., Steyer, J.P., 2010. Model development and simulation for predicting risk of foaming in anaerobic digestion systems. Bioresour. Technol. 101 (12), 4306-4314.

European Commission, 2003. Reference document on Best Available Techniques in Common Waste Water and Waste Gas Treatment / Management System in the Chemical Sector, European IPPC Bureau (EIPPCB), Seville.

Garcia, C., Molina, F., Roca, E., Lema, J.M., 2007. Fuzzy-based control of an anaerobic reactor treating wastewaters containing ethanol and carbohydrates. Ind. Eng. Chem. Res. 46, 6707-6715.

Gavala, H.N., Angelidaki, I., Ahring, B.K., 2003. Kinetics and modeling of anaerobic digestion process. Adv. Biochem. Eng. Biotechnol. 81, 57-93.

Gomez, A., Zubizarreta, J., Rodrigues, M., Dopazo, C., Fueyo, N., 2010. An estimation of the energy potential of agro-industrial residues in Spain. Resour. Conserv. Recycl. [Gomez, Antonio; Zubizarreta, Javier; Rodrigues, Marcos; Dopazo, Cesar; Fueyo, Norberto] Univ Zaragoza, Fluid Mech Grp, Zaragoza 50018, Spain. 54 (11), 972-984.

Helias, A., Leclercq-Perlat, M.N., Bernard, O., 2011. Identification of the biological reaction network in solid fermentations: aplications to cheese ripening. Biotechnol. Bioeng. To be published.

Johnson, R., Wichern, D., 1992. Applied multivariable statistical analysis, Prentice - Hall, New Jersey.

Lee, M.Y., Suh, C.W., Ahn, Y.T., Shin, H.S., 2009. Variation of ADM1 by using temperature-phased anaerobic digestion (TPAD) operation. Bioresour. Technol. 100 (11), 2816-2822.

Mairet, F., Bernard, O., Ras, M., Lardon, L., Steyer, J.P., 2011. Modeling anaerobic digestion of microalgae using ADM1. Bioresour. Technol. 102 (13), 6823-6829.

Mosey, F.E., 1983. Mathematical modeling of the anaerobic digestion process: regulatory mechanisms for the formation of short-chain volatile acids from glucose. Water Sci. Technol. 15 (8/9), 209-232.

Mu, S.J., Zeng, Y., Wu, P., Lou, S.J., Tartakovsky, B., 2008. Anaerobic digestion model no. 1-based distributed parameter model of an anaerobic reactor: I. Model development. Bioresour. Technol. 99 (9), 3665-3675.

Muha, I., Grillo, A., Heisig, M., Schonberg, M., Linke, B., Wittum, G., 2012. Mathematical modeling of process liquid flow and acetoclastic methanogenesis under mesophilic conditions in a two-phase biogas reactor. Bioresour. Technol. 106, 1-9.

Olsson, G., Martinus, K., Zhiguo, Y., Jensen, A.L., Steyer, J.P., 2005. Instrumentation, Control and Automation in Wastewater Systems, IWA Publishing, London.

Perry, R., Green, D., 1999. Perry's Chemical Engineering Handbook, 7th. Edition ed., McGraw - Hill, USA.

Rosen, C., Jeppsson, U., 2006. Aspects on ADM1 implementation within the BSM2 framework IWABSM TG Technical Report, LUTEDX/(TEIE-7224), Lund University, Lund, Sweden.

Rosen, C., Vrecko, D., Gernaey, K.V., Pons, M.N., Jeppsson, U., 2006. Implementing ADM1 for plant-wide benchmark simulations in Matiab/Simulink. Water Sci. Technol. 54 (4), 11-19.

Ruiz-Filippi,G. (2005) Monitorización y control avanzado de reactores anaerobios. PhD. Thesis , University of Santiago de Compostela.

Ruiz-Filippi, G., Rodriguez, J., Roca, E., Lema, J.M. Modification of the IWA-ADM1 for application to anaerobic treatment of ethanolic wastewater from wine factories. 1341-1344. 2004. Anaerobic Digestion 10th World Congress Anaerobic Bioconversion for Sustainability. 29-8-2004.

Siegrist, H., Vogt, D., Garcia-Heras, J.L., Gujer, W., 2002. Mathematical model for meso- and thermophilic anaerobic sewage sludge digestion. Environ. Sci. Technol. 36 (5), 1113-1123.

Spanjers, H., van Lier, J., 2006. Instrumentation in anaerobic treatment-research and practice. Water Sci. Technol. 53 (4-5), 63-76.

Steffens, M.A., Lant, P.A., Newell, R.B., 1997. A systematic approach for reducing complex biological wastewater treatment models. Water Res. 31 (3), 590-606.

Tartakovsky, B., Mu, S.J., Zeng, Y., Lou, S.T., Guiot, S.R., Wu, P., 2008. Anaerobic digestion model No. 1-based distributed parameter model of an anaerobic reactor: II. Model validation. Bioresour. Technol. 99 (9), 3676-3684.

van Lier, J., Tilche, A., Ahring, B., Macarie, H., Moletta, R., Dohanyos, M., Hulshoff Pol, L., Lens, P., Verstraete, W., 2001. New perspectives in anaerobic digestion. Water Sci. Technol. 43 (1), 1-18.

Tables

Table 1 Mass balances considered depending on the key chemical species involved in the anaerobic process and number of *pseudo*-reactions determined by PCA with a data variability threshold of 90%. (COD: Chemical Oxygen Demand, VFA: Volatile Fatty Acids expressed as acetic).

Table 2 Kinetic parameters obtained for the reduced model.

Table 3 Significance of the regressions (threshold 5%) derived from Eq. (8) and parameter values.

Table 4 Comparison of the complexity between the ADM1 and the reduced model.

Table 1

Mass balances considered depending on the key chemical species involved in the anaerobic process and number of *pseudo*-reactions determined by PCA with a data variability threshold of 90%. (COD: Chemical Oxygen Demand, VFA: Volatile Fatty Acids expressed as acetic).

	9 mass balances	8 mass balances	7 mass balances
Soluble compounds	COD VFA Inorganic carbon Inorganic nitrogen Methane Hydrogen	COD VFA Inorganic carbon Inorganic nitrogen Methane Hydrogen	COD VFA Inorganic carbon Inorganic nitrogen Methane
Biomass	Acidogenic biomass Acetoclastic biomass Hydrogenotrophic biomass	Acidogenic biomass Methanogenic biomass	Acidogenic biomass Methanogenic biomass
Number of pseudo-reactions	3 (3)	3(3)	2(2)

* Inside of parenthesis the number of reactions computed from experimental data.

Table 2

Kinetic parameters obtained for the reduced model.

Parameter	Units	Acidogenesis 1st *pseudo*-reaction	Methanogenesis 2nd *pseudo*-reaction
μ_{max}	1/d	0.065 (0.085)	2.016 (2.029)
K_s	kg COD/m^3	0.279 (0.193)	1.668 (1.450)
K_I	kg COD/m^3	- (-)	0.024 (0.034)

() In brackets the initial values for the optimization by equation (12).

Table 3

Significance of the regressions (threshold 5%) derived from Eq. (8) and parameter values.

Vector	Unknown	Significance	Value	Interval
$\beta_1 = \begin{pmatrix} \frac{k_{21}+k_{22}\cdot k_{61}}{k_{11}} \\ 1 \\ 0 \\ 0 \\ 0 \\ k_{22} \\ 0 \end{pmatrix}$	$\frac{k_{21}+k_{22}\cdot k_{61}}{k_{11}}$	Yes	0.337	[0.327 0.346]
	k_{22}		16.63	[16.24 17.03]
$\beta_2 = \begin{pmatrix} \frac{k_{21}}{k_{11}} \\ 1 \\ \frac{k_{22}}{k_{32}} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$	$\frac{k_{21}}{k_{11}}$	Yes	0.916	[0.915 0.917]
	$\frac{k_{22}}{k_{32}}$		65.72	[65.66 65.78]
$\beta_3 = \begin{pmatrix} \frac{k_{41}-k_{42}\cdot k_{61}}{k_{11}} \\ 0 \\ 0 \\ 1 \\ 0 \\ -k_{42} \\ 0 \end{pmatrix}$	$\frac{k_{41}-k_{42}\cdot k_{61}}{k_{11}}$	Yes	-0.001	[-0.001 -0.0008]
	$-k_{42}$		0.239	[0.232 0.245]
$\beta_4 = \begin{pmatrix} \frac{k_{42}-k_{52}\cdot k_{41}}{k_{11}\cdot k_{42}} \\ 0 \\ 0 \\ -\frac{k_{52}}{k_{42}} \\ 1 \\ 0 \\ 0 \end{pmatrix}$	$\frac{k_{42}-k_{52}\cdot k_{41}}{k_{11}\cdot k_{42}}$	Yes	0.050	[0.0499 0.0502]
	$-\frac{k_{52}}{k_{42}}$		-0.708	[-0.675 -0.741]

$\beta_5 = \begin{pmatrix} \frac{k_{71}-k_{61}\cdot k_{72}}{k_{11}} \\ 0 \\ 0 \\ 0 \\ 0 \\ -k_{72} \\ 1 \end{pmatrix}$	$\frac{k_{71}-k_{61}\cdot k_{72}}{k_{11}}$ $-k_{72}$	Yes	-0.0004 -0.0076	[-0.0004 -0.0004] [-0.0076 -0.0077]

Table 4

Comparison of the complexity between the ADM1 and the reduced model.

	ADM1	Reduced Model
Parameters	89	13
State variables	28	7
Types of biomass	8	2
Number of reactions	21	2
Outputs	32	8
Complexity ratio*	7/89 = 0.08	7/13 = 0.54

* Complexity ratio is defined as the ratio between the number of available measurements and the number of parameters. A higher ratio complexity indicates greater model simplicity.

Figures

Fig. 1. Input conditions for the anaerobic USBF reactor treating winery effluents. The organic loading rate *OLR* (-) (kg COD/m^3·d) and the influent concentration of chemical oxygen demand COD_{in} (-.-) (kg COD/m^3).

Fig. 2. Scheme of the procedure used for model reduction.

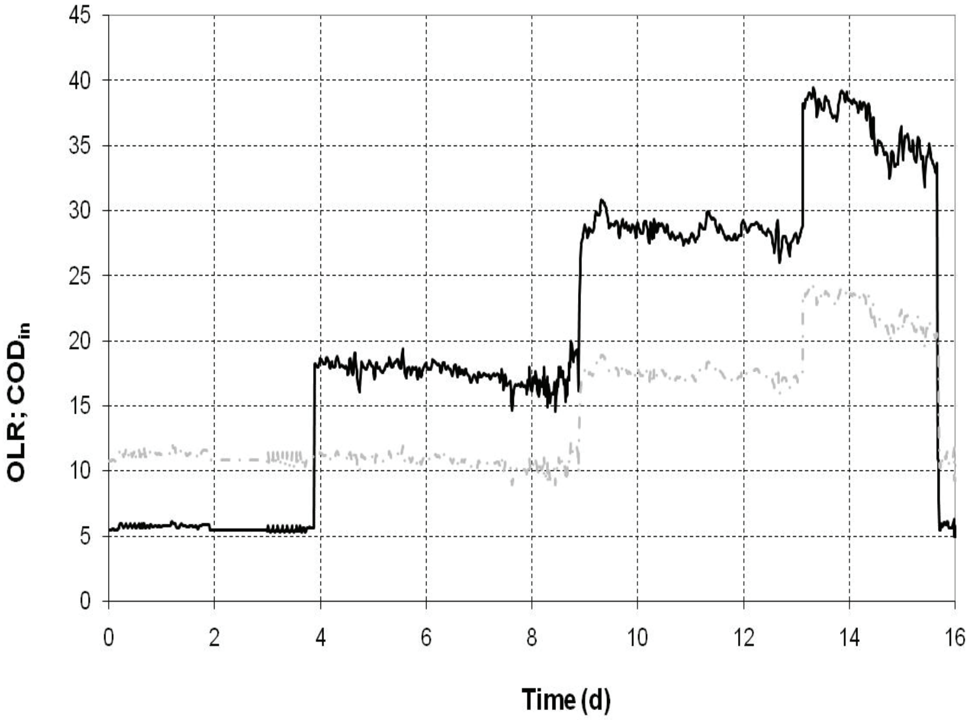
Fig. 3. Explained variance (bar) and cumulated explained variance for 7 mass balances (▲) with respect to the number of reactions for winery effluents treated in a USBF.

Fig. 4. Simulated reaction rate and kinetics of the reduced model. a) Acidogenesis – Monod function. r_1: acidogenic reaction rate; X_1: acidogenic biomass; S_1: primary organic substrate expressed as chemical oxygen demand (COD). b) Methanogenesis – Haldane function. r_2: methanogenic reaction rate; X_2: methanogenic biomass; S_2: volatile fatty acids.

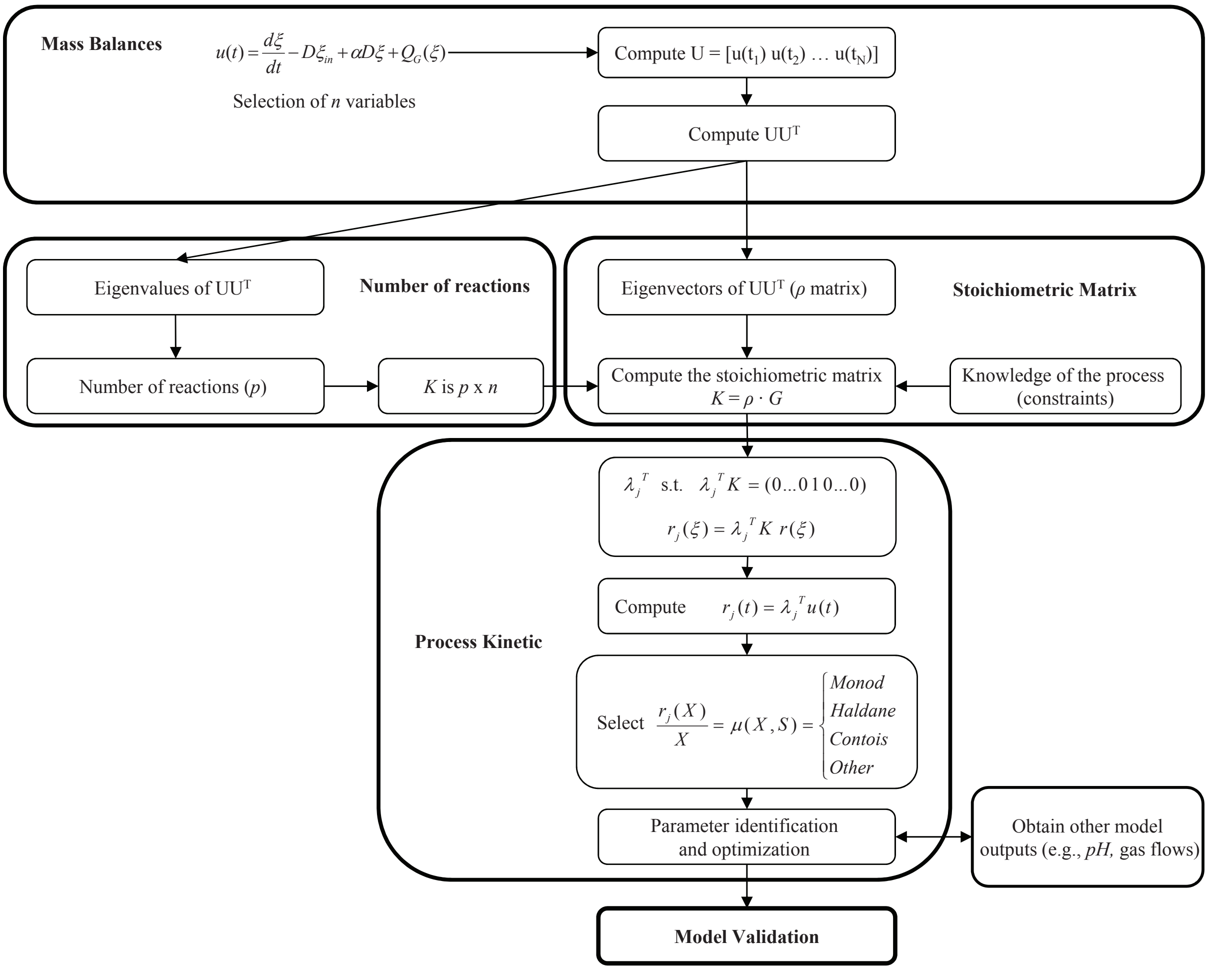
Fig. 5. *pH* simulation comparing the behavior of the reduced model (-.-) against the ADM1 (-).

Fig. 6. Time evolution of methane flow rate q_m (m^3/d), chemical oxygen demand S_1 (kg COD/m^3) and volatile fatty acids concentration S_2 (kg COD/m^3) at the outlet of the USBF reactor for the reduced model (-.-) and ADM1 (-).

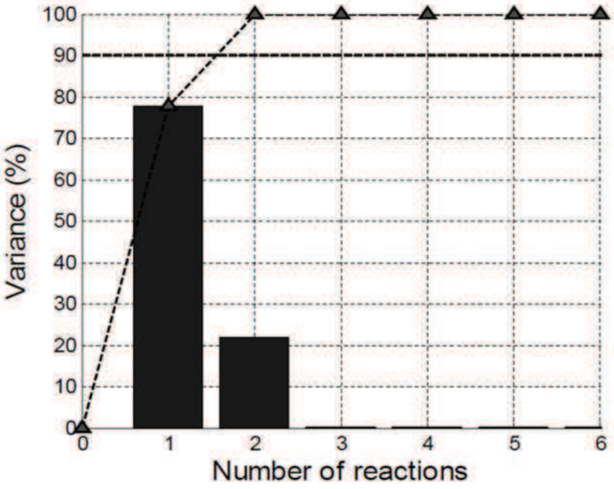
Figure(1)



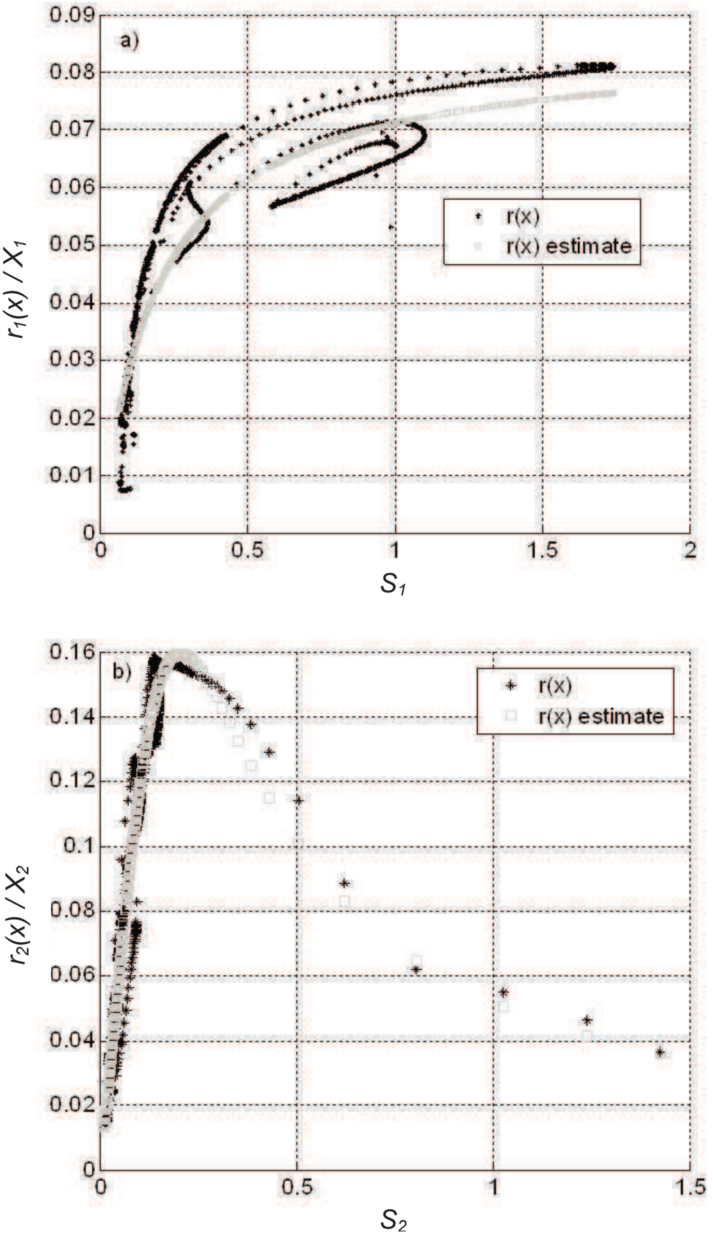
Figure(2)


Mass Balances
$u(t)=\frac{d\xi}{dt}-D\xi_{in}+\alpha D\xi+Q_G(\xi)$
Selection of n variables
Compute U = [u(t$_1$) u(t$_2$) … u(t$_N$)]
Compute UU^T
Eigenvalues of UU^T
Number of reactions
Number of reactions (p)
K is p x n
Eigenvectors of UU^T (ρ matrix)
Stoichiometric Matrix
Compute the stoichiometric matrix $K=\rho\cdot G$
Knowledge of the process (constraints)
λ_j^T s.t. $\lambda_j^T K=(0...0\,1\,0...0)$
$r_j(\xi)=\lambda_j^T K\, r(\xi)$
Compute $r_j(t)=\lambda_j^T u(t)$
Process Kinetic
Select $\frac{r_j(X)}{X}=\mu(X,S)=\begin{cases}Monod\\Haldane\\Contois\\Other\end{cases}$
Parameter identification and optimization
Obtain other model outputs (e.g., pH, gas flows)
Model Validation

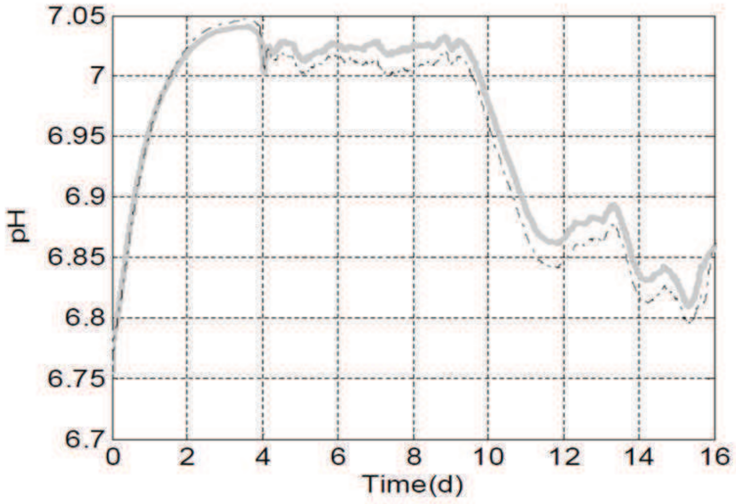
Figure(3)


100
90
80
70
60
50
40
30
20
10
0
Variance (%)
0
1
2
3
4
5
6
Number of reactions

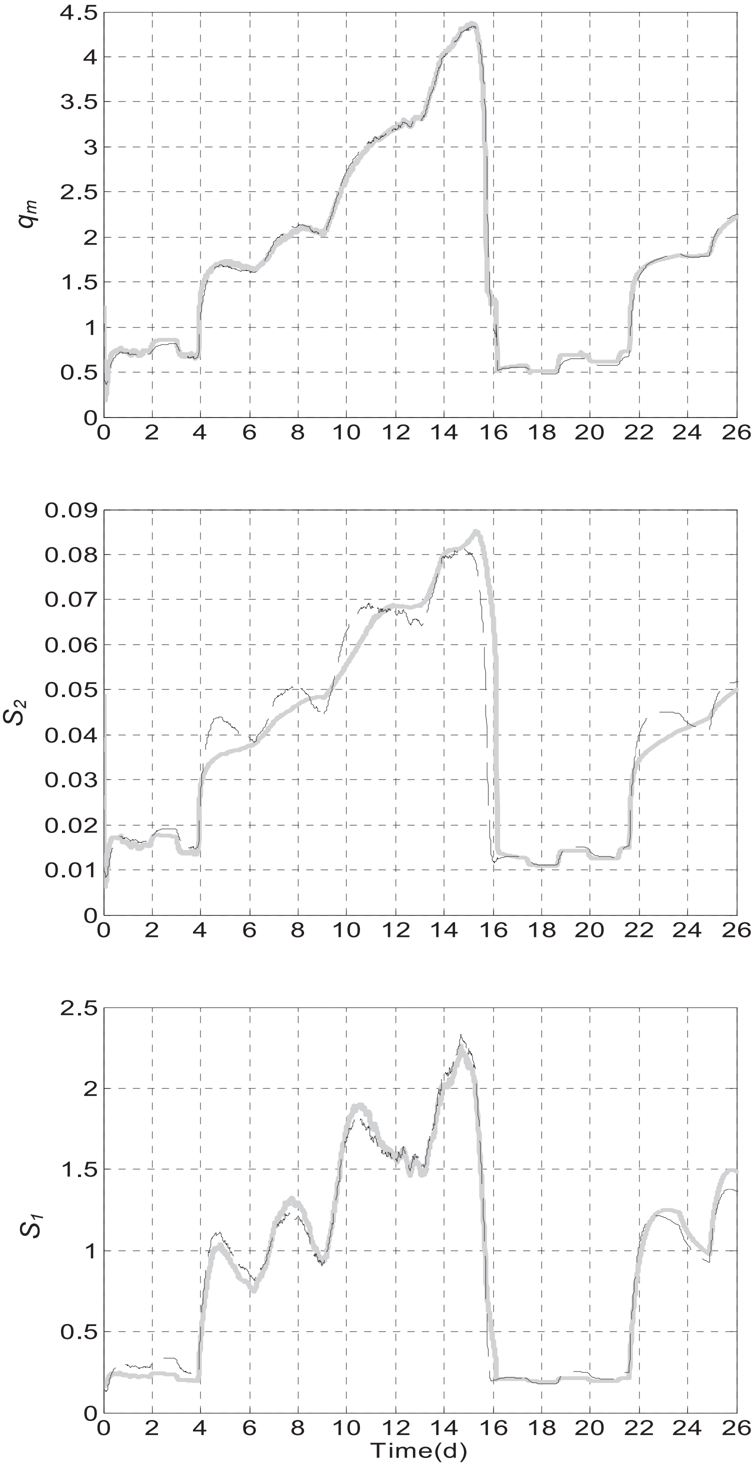
Figure(4)


a)
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
0.5
1
1.5
2
$r_1(x) / X_1$
S_1
r(x)
r(x) estimate
b)
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.5
1
1.5
$r_2(x) / X_2$
S_2
r(x)
r(x) estimate

Figure(5)



Figure(6)



4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
q_m
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
S_2
2.5
2
1.5
1
0.5
0
S_1
0 2 4 6 8 10 12 14 16 18 20 22 24 26
Time(d)