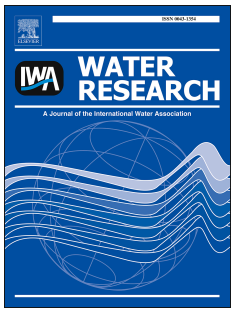
On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen

Giannina Giovannini, Mihaela Sbarciog, Jean-Philippe Steyer, Rolando Chamy, Alain Vande Wouwer

Accepted Manuscript



On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen

Giannina Giovannini, Mihaela Sbarciog, Jean-Philippe Steyer, Rolando Chamy, Alain Vande Wouwer

PII: S0043-1354(18)30050-2

DOI: 10.1016/j.watres.2018.01.036

Reference: WR 13515

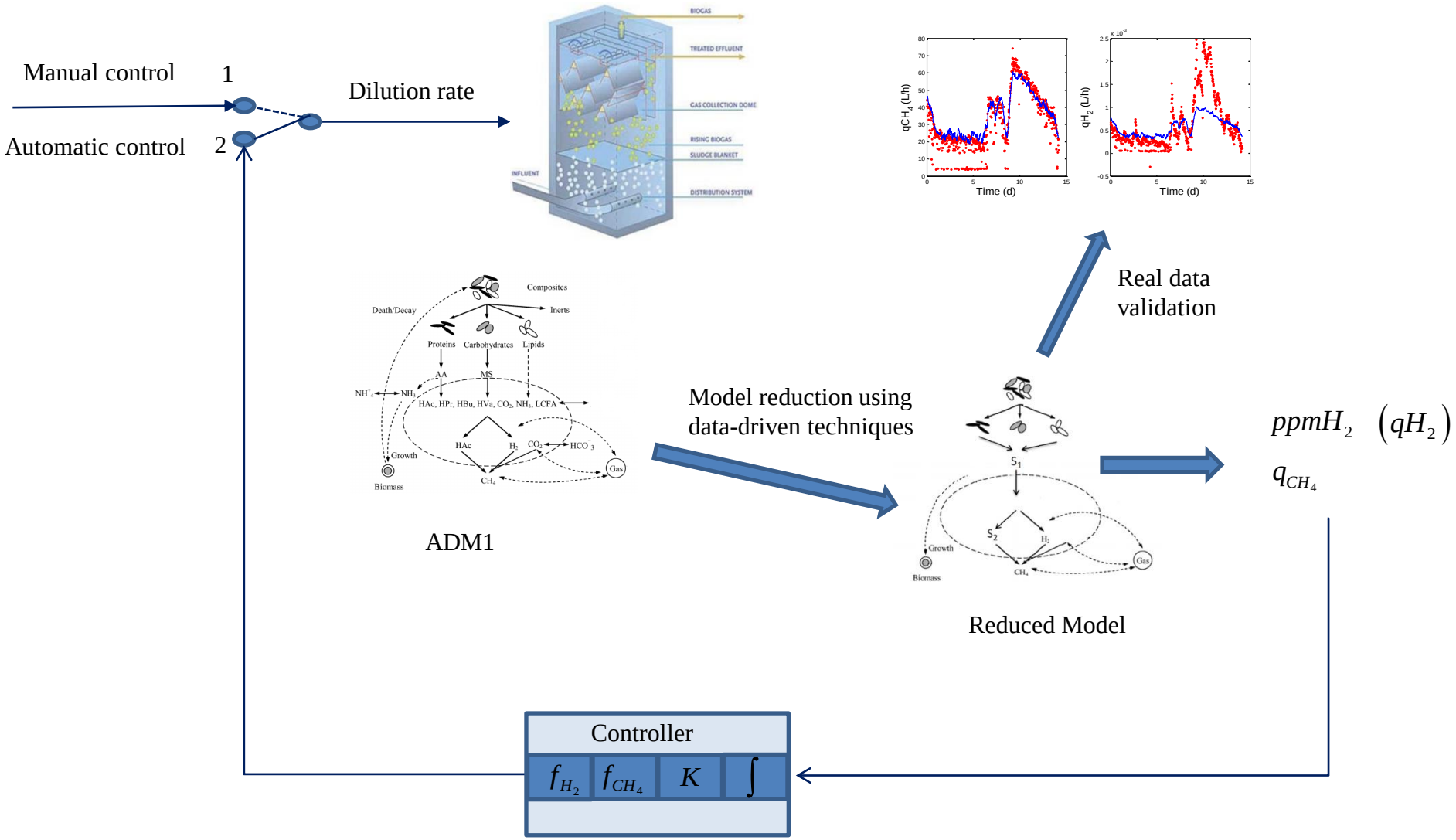
To appear in: *Water Research*

Received Date: 14 June 2017

Revised Date: 5 December 2017

Accepted Date: 15 January 2018

Please cite this article as: Giovannini, G., Sbarciog, M., Steyer, J.-P., Chamy, R., Wouwer, A.V., On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen, *Water Research* (2018), doi: 10.1016/j.watres.2018.01.036.


Manual control
1
Automatic control
2
Dilution rate
BIOGAS
TREATED EFFLUENT
GAS COLLECTION DOME
RISING BIOGAS
SLUDGE BLANKET
INFLUENT
DISTRIBUTION SYSTEM
Time (d)
Time (d)
Real data validation
Composites
Death/Decay
Inerts
Proteins
Carbohydrates
Lipids
AA
MS
HAc, HPr, HBu, HVa, CO2, NH3, LCFA
HAc
H2
CO2
HCO3
Growth
Biomass
Gas
CH4
ADM1
Model reduction using data-driven techniques
S1
S2
H2
Growth
Biomass
Gas
CH4
Reduced Model
ppmH2 (qH2)
qCH4
Controller
fH2
fCH4
K
∫

On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen

In memory of Prof. Gonzalo Ruiz-Filippi

Giannina Giovannini[a,b,*], Mihaela Sbarciog[a], Jean-Philippe Steyer[c], Rolando Chamy[b], Alain Vande Wouwer[a]

[a]*Service d'Automatique, Université de Mons, 31 Boulevard Dolez, 7000 Mons, Belgium.*
[b]*Escuela de Ingeniería Bioquímica, Pontificia Universidad Católica de Valparaíso, Av. Brasil 2085, Valparaíso Chile.*
[c]*Laboratoire de Biotechnologie de l'Environnement, 11100 Narbonne, France*

Abstract

Hydrogen has been found to be an important intermediate during anaerobic digestion (AD) and a key variable for process monitoring as it gives valuable information about the stability of the reactor. However, simple dynamic models describing the evolution of hydrogen are not commonplace. In this work, such a dynamic model is derived using a systematic data driven-approach, which consists of a principal component analysis to deduce the dimension of the minimal reaction subspace explaining the data, followed by an identification of the kinetic parameters in the least-squares sense. The procedure requires the availability of informative data sets. When the available data does not fulfill this condition, the model can still be built from simulated data, obtained using a detailed model such as

*Corresponding author
Email address: giannina.giovannini@umons.ac.be (Giannina Giovannini)

ADM1. This dynamic model could be exploited in monitoring and control applications after a re-identification of the parameters using actual process data. As an example, the model is used in the framework of a control strategy, and is also fitted to experimental data from raw industrial wine processing wastewater.

Keywords: Mathematical modeling, parameter estimation, principal component analysis, sensitivity analysis

1. Introduction

Despite its numerous advantages, anaerobic digestion (AD) is still not used at its full potential, due to the high complexity of the process and its dependency on many operational variables. Under some circumstances the stability of the AD process can be endangered, which may deteriorate the efficiency of the waste treatment and the associated biogas production (Chen et al., 2008). Therefore, an important step towards an optimal operation and control is a better understanding of the interplay between the process dynamics and the operational conditions, which may be achieved by means of a reliable model.

One of the most detailed and well-accepted description of the process is provided by the Anaerobic Digestion Model 1 (ADM1) (Batstone et al., 2002), which can be customized for a wide variety of wastes and plant configurations. However, from a control and optimization viewpoint, ADM1, with its 32 state variables, is too complex.

The underlying idea behind the present work is to use informative data and to investigate the derivation of low-order dynamic models to predict

the time evolution of the key variables of interest. The methodology is data-driven in essence as it infers a low-dimensional subspace spanned by the columns of the stoichiometric matrix using a principal component analysis. The first step is therefore to select the measurement signals, which are potential candidates for describing the process behaviour. Once a candidate reaction scheme has been identified, the kinetic parameters can be estimated using a weighted least squares method. In this procedure, the use of virtual data (i.e., synthetic data generated by a detailed process model) allows to explore a large range of operating conditions, possibly wider than feasible in actual practice with a real plant, to ascertain the derivation of the low-order model structure. Of course, real-life experimental data has to be used at a later stage to estimate the parameters of the low-order model in the framework of a specific application.

Several mathematical models describing anaerobic digestion processes have been developed in past years in order to achieve specific goals (Yu et al., 2013). The majority of them are identified from experimental data starting from an imposed model structure. A few others are developed using data-driven techniques, which first analyze the available data and then propose a structure of the model to accomodate the features of the data. Most of the dynamic models are developed for monitoring and control purposes, and to a less extent, for simulation purposes. Among the models which were identified based on experimental data, one can note the work of Bernard et al. (2001), who developed the model of an AD process treating winery wastewater (the so-called AM2 model), the work of Simeonov and Karakashev (2012), who built a model of a pilot plant treating acti-

vated sludge from municipal wastewater treatment plants, and the work of Owhondah et al. (2016), who, aside model identification, investigated the most appropriate structure in terms of number of reactions and kinetics for an AD process treating green and food waste. Models were also developed based on data generated by simulation models such as ADM1. Among the latter, one can highlight the work of Beltramo et al. (2016), who developed an artificial neural network model to predict the biogas flow rate for a two-substrate co-digestion system simulated by ADM1, or the model presented by García-Diéguez et al. (2013), who used principal component analysis to determine the number of reactions and built a model based on data generated by an adjusted ADM1 to simulate winery effluents. Another research direction is the extension of models identified from experimental data to reproduce the dynamics of ADM1. Such works are reported by Ficara et al. (2012), who emphasized that nitogen has to be included in the model structure introduced by Bernard et al. (2001), and by Hassam et al. (2015), who included the hydrolysis step, along with acidogenesis and methanogenesis, in the model development. Finally, the work of Della Bona et al. (2015) uses a linear fractional transformation to first identify the model parameters from data generated by ADM1 and then from data collected from a lab-scale AD process treating ultra-filtered cheese-whey.

Recently, considerable attention has been paid to the role of hydrogen in anaerobic digestion processes. Indeed, it is an intermediary metabolite present in many reactions, and it is considered a fast indicator of destabilization of the process (Giraldo-Gomez et al., 1992; Ryhiner et al., 1993;

Bhattacharya et al., 1995; Björnsson et al., 2001; Boe et al., 2010; Hou et al., 2014; Giovannini et al., 2016). However, some authors have nuanced its potential as a monitoring indicator (Voolapalli and Stuckey, 1999, 2001). In fact, the type of substrates and operational conditions during anaerobic digestion can trigger certain metabolic pathways which are more sensitive to hydrogen production and/or consumption. For example, it has been reported that high-strength wastewater composed mainly of well-defined sugars, such as glucose, sucrose or lactate will produce more hydrogen (Pauss and Guiot, 1993; Hawkes et al., 2002; Yu et al., 2002) compared to particulate substrates (Boe, 2006). In the latter case, VFA measurements are suggested as better indicators of stability. In this study, having in mind this restriction, the focus is on developing a dynamic model of the AD process including hydrogen as a state variable, in order to be able to subsequently develop monitoring and control systems based on this variable (provided it is pertinent). On the other hand, online hydrogen sensors are available at competitive cost (Pauss and Guiot, 1993; Cord-Ruwisch et al., 1997), which makes this kind of strategy practically feasible.

The resulting mathematical model, built from informative data generated by simulating ADM1, contains two trophic groups, total organic matter (characterized by COD), organic acids, methane, and hydrogen flow rates, which represent the key variables of the process and can be either measured or estimated. This model is envisioned as a good basis for an advanced (model-based) monitoring and control approach of the process. To date most of the control strategies are based on COD, VFA and/or methane. Only a few hydrogen-based control strategies have been proposed so far,

some of them based on empirical principles (Cord-Ruwisch et al., 1997; Rodriguez et al., 2006), others based on dynamic models (Dochain et al., 1991; Ryhiner et al., 1993). In this study, the control strategy proposed by Rodriguez et al. (2006) is exploited to test our dynamic model and its consistency with respect to ADM1. Further, experimental data is used to identify the parameters of the proposed model and to prove its predictive capability.

This paper is organized as follows. The next section explains the development of the dynamic model using informative data sets generated by ADM1. Principal component analysis is used to deduce a macroscopic reaction scheme. Nonlinear least squares is then applied to estimate the kinetic parameters, and parametric sensitivity analysis is carried out to further assess practical parameter identifiability and possibly propose further simplifications. Measurement noise is also discussed, from the point of view of applications to actual process data, with maximum likelihood principal component analysis and estimation. The numerical results show the predictive capability of the low-dimensional model as compared to the original ADM1. Section 3 exploits the reduced model in the context of a simple hydrogen-based controller initially introduced by Rodriguez et al. (2006). In section 4, the model parameters are estimated based on experimental data collected from a $1m^3$ fixed-bed reactor treating industrial wine wastewater, thus demonstrating its predictive capability. Finally, the last section draws some conclusions and perspectives.

2. Model derivation

In this section, synthetic data is generated by simulating ADM1 (Batstone et al., 2002). This data is used to infer a low-dimensional dynamic model involving the main variables of interest, i.e. those with interesting monitoring and control prospects. The next subsection therefore discusses:

- the selection of the model variables and data collection,
- the determination of the minimum number of reactions and pseudo-stoichiometric matrix using principal component analysis,
- kinetic parameter estimation using least-squares identification, and possibly re-estimation of stoichiometric parameters,
- parametric sensitivity analysis for further model simplification, and assessment of the parameter confidence intervals based on the Fisher Information Matrix.

2.1. Model variable selection and data collection

Data has to be informative enough to capture the most important process dynamics. In an actual process study, this step is crucial but can be particularly delicate to conduct as operating a process at optimal conditions is in contradiction with getting information on the several operating ranges where the process operates under abnormal conditions or is even at risk. This is why the use of a reliable, well-accepted, detailed model is appealing. One expects the model to be sufficiently detailed to represent the process in a proper way in several operating ranges, and one is able to generate excitation signals that drive the process in these various ranges.

The synthetic database can therefore be made rich enough to draw conclusion on the significance of a reduced-order model. In contrast, a poor database could lead to oversimplification and models whose predictive capabilities would be drastically limited.

2.1.1. *Liquid phase variables*

Microbial populations and substrates described in ADM1 are clustered in a reasonable and congruent manner in order to include tractable information in the reduced model. Similar clustering has been used by Ficara et al. (2012) and Hassam et al. (2015), who respectively compared the dynamics of AM2 and ADM1 and built AM2HN, an extention of AM2 which includes also the hydrolysis step, to approximate the ADM1 dynamics:

- The total organic matter, denoted by S_1, is the sum of monosaccharides, amino acids, long chain fatty acids, composite, carbohydrates, proteins, and lipids:
  $S_1 = S_{su} + S_{aa} + S_{fa} + X_c + X_{ch} + X_{pr} + X_{li}$

- Organic acids, called S_2, are the sum of valerate, acetate, butyrate, propionate:
  $S_2 = S_{va} + S_{ac} + S_{bu} + S_{pro}$

- The seven microbial populations considered in ADM1 are lumped into two main groups, acidogenic microorganisms called X_1 and acetogens and methanogens called X_2:
  $X_1 = X_{su} + X_{aa} + X_{fa}$
  $X_2 = X_{ac} + X_{H_2} + X_{c4} + X_{pro}$

For the sake of simplicity endogenous decay of microorganisms is not considered.

2.1.2. *Gaseous compounds*

Unlike biomass and substrates which are only present in the liquid phase, gaseous compounds, such as methane, hydrogen and carbon dioxide, are produced in the liquid phase and in large part transferred to the gas compartment. It is interesting to have a closer look at the mass balance equations in order to carefully define the transfer terms. As the reduced model is designed for monitoring and control purposes we are interested in following up the variations in hydrogen and methane, while carbon dioxide is not considered.

The mass balance equations of a biochemical system involving M reactions and N components in the liquid phase can be written as:

$$V_L \frac{d\xi_L}{dt} = F_{in}(\xi_{L,in} - \xi_L) + V_L\, K\, \varphi(\xi_L) - V_L\, Q_G(\xi_G) \tag{1}$$

with $\xi_{L,in}$ and $\xi_L \in \mathbb{R}^N$ the inlet concentration and state vector respectively, $K \in \mathbb{R}^{N\times M}$ the stoichiometric matrix, $\varphi(\xi_L) \in \mathbb{R}^{\mathbb{M}}$ the reaction rate vector, F_{in} the inlet flow rate and V_L the liquid volume, $Q_G(\xi_G) \in \mathbb{R}^N$ the gaseous transfer rates.

Mass balance in the gas phase gives:

$$V_G \frac{d\xi_G}{dt} = -q_G\, \xi_G + V_L\, Q_G(\xi_G) \tag{2}$$

where q_G is the gas outflow, and V_G the headspace volume.

The transfer term between the two phases disappears in the addition of equations (1) and (2) and the following expression is obtained:

$$V_L \frac{d\xi_L}{dt} + V_G \frac{d\xi_G}{dt} = F_{in}(\xi_{L,in} - \xi_L) + V_L K \varphi(\xi_L) - q_G \xi_G \quad (3)$$

The left-hand side term corresponds to the variation of total mass of the compounds under consideration:

$$\frac{dm_G}{dt} = V_L \frac{d\xi_L}{dt} + V_G \frac{d\xi_G}{dt} \quad (4)$$

Considering that the compounds are only produced in the liquid phase and then transported to the gas phase, equation (3) is divided by the liquid volume to obtain a new auxiliary state variable, which corresponds to the total mass concentration in the liquid phase:

$$\frac{1}{V_L} \frac{dm_G}{dt} = D\,(\xi_{L,in} - \xi_L - \frac{q_G}{F_{in}}\xi_G) + K\,\varphi(\xi_L) \quad (5)$$

In this expression the term $D\,(\xi_L + \frac{q_G}{F_{in}}\xi_G)$ corresponds to the total compound outflow either in liquid or gaseous forms. D denotes the dilution rate, defined as the ratio between the inlet flow rate (F_{in}) and the liquid volume (V_L).

2.1.3. *Synthetic data*

In this study, informative data is generated *in silico* by simulation of the ADM1 model as a plant emulator. Simulations are performed considering a 1 m^3 reactor operating in continuous mode and treating mainly soluble waste matter with 50% biomass retention. The influent concentrations used in the ADM1 model are listed in Table **??**.

Two data sets are built, one for parameter identification, and another one for cross-validation, both run over a period of 120 days and sampling time of 3.6 hours. These time scales are used to capture the slow and fast dynamics. Indeed, analysis of ADM1 shows that the dominant time constant is of the order of a month, whereas fast time constants are in the range of a few hours.

The six candidate state variables of our reduced model, namely X_1, X_2, S_1, S_2, CH_4, H_2, are sampled from ADM1, and for the sake of realism are corrupted with independent, normally-distributed, additive white noises, with standard deviation of 1% of the error-free values for biogases and 5% for the rest of the variables:

$$\xi_i^m = \xi_i + \eta_i \qquad (i = 1, ..., n_s) \tag{6}$$

where, ξ_i^m is the noisy measurement, ξ_i is the noise-free variable and η_i is the additive noise.

Figure **??** shows the two data sets. Cross-validation will be performed to challenge the identified model with fresh data (data that has not been used for parameter identification purposes). The data for cross-validation differs from the data for identification in the initial condition values and the dilution rate evolution.

2.2. Minimum number of reactions and pseudo-stoichiometric matrix estimation

A challenge in deriving a macroscopic model of a biological process is to determine the structure of the mathematical model, which consists of defining the adequate number of biological reactions while keeping the complexity of the model as low as possible (Rodriguez et al., 2008).

In this study the minimum number of reactions and pseudo-stoichiometric matrix are obtained using Maximum Likelihood Principal Component Analysis (MLPCA) as described by Mailier et al. (2013).

The reaction number is obtained by successive construction of affine linear subspaces of increasing dimensions p (with p ranging from 1 to $M = 6$) that represent at best the noisy data. To formulate the MLPCA problem, the transport terms of the general mass balance equation are lumped into $v(t)$ as follows:

$$\frac{d\xi}{dt} = K\varphi(\xi) + v(t) \tag{7}$$

The solution of this equation is given by:

$$\xi(t) - \xi_0 = K\int_0^t \varphi(\xi) + \int_0^t v(t) \tag{8}$$

If we define the transport-free concentration $\xi_f(t)$ by:

$$\xi_f(t) = \xi(t) - \int_0^t v(t) \tag{9}$$

the solution can then be expressed as:

$$\xi_f(t) = \xi_0 + K\int_0^t \varphi(\xi) \tag{10}$$

The transport-free state ξ_f evolves within a linear affine subspace κ defined by the column-subspace (or range) of the yield matrix K, and translated from the origin by the initial state vector ξ_0. This affine subspace

is independent of the reaction rates and its dimensionality is equal to the reaction number M.

MLPCA can be used to determine this linear affine subspace by solving a maximum likelihood problem of the form

$$\mathcal{J}_p = \sum_{i=1}^{n_s} (\xi_{f,m_i} - \hat{\xi}^p_{f_i})^T \, \Sigma_i^{-1} \, (\xi_{f,m_i} - \hat{\xi}^p_{f_i}) \tag{11}$$

where n_s denotes the number of measured samples, ξ_{f,m_i} is the noisy measurement, $\hat{\xi}^p_{f_i}$ a p-dimensional maximum-likelihood estimate, Σ_i is the error covariance matrix of the noisy measurements. In general terms, the problem can be formulated to estimate the stoichiometic matrix only, or the stoichiometric matrix together with the initial conditions, and is actually solved using singular value decomposition (see (Mailier et al., 2013) for more details on the method and its solution).

The minimum value of the criterion has a chi-square distribution χ^2 with $N \times (n_s - 1)$ degrees of freedom. The reaction number can be selected by progressively increasing the dimensionality p of the candidate stoichiometric subspace, starting from $p = 1$. A value of J_p larger than $\chi^{2,(1-\alpha)}_{N\times(n_s-1)}$ leads to the rejection of the hypothesis $M = p$ at the significance level α and the next value of p should then be tested (Mailier et al., 2013).

Besides the selection of the minimum number of reactions, MLPCA also allows to obtain information about the pseudo-stoichiometric matrix K, in the form of an orthonormal basis $\rho \in \mathbb{R}^{N\times M}$ of the affine subspace. To find the columns of matrix K, it is then necessary to determine a transformation matrix $G \in \mathbb{R}^{M\times M}$, such that:

$$\hat{K} = \hat{\rho} G \tag{12}$$

The regular matrix $G \in \mathbb{R}^{M \times M}$ can be estimated independently of the reaction kinetics only if sufficient prior knowledge is available. In general, M constraints should be formulated to guarantee this property, which is equivalent to the concept of C-identifiability introduced in (Chen and Bastin, 1996).

Based on the synthetic data collected from ADM1, MLPCA yields the results of Figure **??** for subspace dimensionalities ranging from 1 to 4. The chi-square test shows that a two-reaction scheme ($p = 2$) would be sufficient to represent the given data set, i.e., leads to a log-likelihood cost smaller than any plausible value of a random variable with distribution $\chi^{2,(99.9\%)}_{4806}$. This result is in agreement with the observations of Bernard et al. (2001), who also proposed a two-reaction model (the AM2 model that has been extensively used in the meantime for analysis, estimation and control of AD processes).

MLPCA also delivers the orthonormal basis:

$$\hat{\rho} = \begin{bmatrix} -0.18 & -0.06 \\ -0.07 & -0.13 \\ 0.92 & 0.28 \\ -0.30 & 0.94 \\ -0.06 & -0.04 \\ -2.1 \cdot 10^{-6} & 6.3 \cdot 10^{-6} \end{bmatrix} \tag{13}$$

In order to find the transformation matrix G, one can invoke that only

one microbial population should be involved in each reaction. Normalizing the reactions with respect to these populations, the first two rows of K can be imposed, and G can be deduced as

$$G = \begin{bmatrix} -6.38 & 2.77 \\ 3.33 & -8.61 \end{bmatrix} \tag{14}$$

The matrix K then results:

$$\hat{K} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -4.95 & 0.09 \\ 5.12 & -8.99 \\ 0.26 & 0.21 \\ 3.4 \cdot 10^{-5} & -6.0 \cdot 10^{-5} \end{bmatrix} \tag{15}$$

Each k_{ij} element in the $\hat{K}$ matrix corresponds to a stoichiometric coefficient that relates the j^{th} reaction with the i^{th} compound. Here, the rows correspond to X_1, X_2, S_1, S_2, CH_4, H_2, respectively. A positive value in the $\hat{K}$ matrix corresponds to a product of the reaction, while negative values refer to substrates.

To avoid the use of the indexes (i, j) and simplify the notation, the correspondence given in Table **??** is proposed. Hence, the reaction scheme becomes:

$$k_1 S_1 \xrightarrow{r_1} X_1 + k_3 S_2 + k_7 H_2 + k_5 CH_4 \tag{16}$$

$$k_4 S_2 + k_8 H_2 \xrightarrow{r_2} X_2 + k_6 CH_4 + k_2 S_1 \tag{17}$$

where the first microbial population X_1 transforms the organic matter S_1 into organic acids S_2, methane CH_4, and hydrogen H_2 at a reaction rate r_1. Products of the first reactions would then be used by the second population X_2 to produce methane at reaction rate r_2.

The values of the stoichiometric coefficients in matrix (15) will be used as a starting point for the development of a reduced model, the next identification step being dedicated to the kinetic parameters. Of course, one can question the precision of this first estimation of the stoichiometric matrix as it leads to somewhat unexpected results, such as the production of organic matter in the second reaction. These results have to be considered as an abstract mathematical representation and as a starting point, which is then subject to refinement in the following sections.

It is also interesting to consider the influence of the data set on the MLPCA results. The results of several tests are summarized in Table **??**, where the duration (43 and 120 days) and the level of noise corrupting the gaseous compounds (1% and 5% relative errors), as well as the level of excitation are varied into three types of signals (step function, ramp, and ramp-sinusoidal). The ramp-sinusoidal is chosen as the identification set as it gives the most informative response of the system.

It can be seen that the most affected stoichiometric coefficient is the one associated to S_1 in the second reaction. Longer experiments with relatively rich excitation signals are necessary to extract the information about this coefficient. These observations are in agreement with results obtained after re-estimation of stoichiometric parameters and the relatively large uncertainty on this parameter (see Table **??**), which will lead to its

elimination from the model.

2.3. *Kinetic parameter identification and re-estimation of stoichiometric parameters*

Finding the stoichiometric coefficients independently of the kinetics allows to divide the parameter estimation problem into subproblems which are easier to solve. The kinetic parameters can now be estimated by solving a weighted least squares problem where the deviation between the experimental data $y_{i,meas}(t)$ and the model prediction $y(\theta, t)$ can be quantified by:

$$J(\theta) = \sum_{i=1}^{n_s} \left[(y_i(\theta) - y_{i,meas})^T W_i^{-1} (y_i(\theta) - y_{i,meas}) \right] \tag{18}$$

where n_s denotes the number of samples, θ is the kinetic parameter vector, and W is a weighting matrix, which in the absence of more information on the measurement noise, is usually chosen so as to compensate for different scales and units between the variables (if more information is available, W could be taken as the covariance matrix of the measurement error, thus expressing the criterion (18) in the ML sense).

In a last step, the nonlinear identification problem can also be extended to the stoichiometric parameters. The idea is to subdivide the identification in as many steps as necessary, initializing one problem with the results of the previous subproblems. Here the stoichiometric parameters found by MLPCA could possibly be re-estimated together with the kinetic parameters in a final nonlinear identification problem.

The Nelder-Mead algorithm implemented in the Matlab routine *fminsearch*

is used to find the optimal parameter set. Moreover, positivity constraints for the parameters can be imposed using a logarithmic transformation.

Initialization of the first nonlinear optimization problem is also important since parameter identification can be subject to local minima. Reasonable initial guesses are taken from literature. In addition, a multistart strategy (30 runs) using the Latin Hypercube Sampling (LHS) is used to explore the parameter space in a neighbourhood of this starting point (McKay et al., 1979).

In the application under consideration, Monod laws are considered for both reactions:

$$\mu_{1,2}(S_{1,2}) = \mu_{max1,2} \frac{S_{1,2}}{S_{1,2} + K_{S1,2}} \tag{19}$$

$\mu_{1,2}$ is the specific growth rate of the microorganisms, $\mu_{max1,2}$ is the specific maximum bacterial growth rate, $K_{S1,2}$ is the half-saturation constant associated to the substrate $S_{1,2}$.

The estimated kinetic parameters together with their respective 95% confidence intervals are listed in Table **??** while results for stoichiometric parameters are shown in Table **??**.

Decoupling the identification problem can help the identification process to proceed faster, reducing the number of iterations. An example of such statistics is given in Table **??**.

It is worth noting that kinetic parameters are of the same order of magnitude as values reported in other relevant studies (Bernard et al., 2001; García-Diéguez et al., 2013) and re-estimated stoichiometric values are comparable to those obtained with MLPCA.

Figures **??** and **??** show the comparison between the model prediction, with the best set of parameters, and the noisy data simulated with ADM1 in direct and cross-validation, respectively. The noisy data is represented with red dots, while the model prediction is depicted by solid blue lines. The model fit is satisfactory with sum of squared errors of 7.24 and 12.24, for direct and cross-validation, respectively).

2.4. Sensitivity analysis

As a result of the previous analysis, the following dynamic model equations can be written:

$$\frac{dX_1}{dt} = \mu_1 X_1 - \alpha\, D\, X_1 \tag{20}$$

$$\frac{dX_2}{dt} = \mu_2 X_2 - \alpha\, D\, X_2 \tag{21}$$

$$\frac{dS_1}{dt} = D\,(S_{1in} - S_1) - k_1 \mu_1 X_1 + k_2 \mu_2 X_2 \tag{22}$$

$$\frac{dS_2}{dt} = D\,(S_{2in} - S_2) + k_3 \mu_1 X_1 - k_4 \mu_2 X_2 \tag{23}$$

$$q_{CH_4} = \beta_1\, \mu_1\, X_1 + \beta_2\, \mu_2\, X_2 \tag{24}$$

$$q_{H_2} = \beta_3\, \mu_1\, X_1 - \beta_4\, \mu_2\, X_2 \tag{25}$$

where the factor α accounts for biomass retention ($\alpha = 0$ for an ideal fixed bed reactor and $\alpha = 1$ for an ideal CSTR - here $\alpha = 0.5$ is considered). Knowing that methane and hydrogen have a relatively low solubility they are assumed to go directly to the gas compartment and leave the reactor at a volumetric flow rate proportional to the reaction rates given by equations (24) and (25), respectively.

Sensitivity analysis can be used to assess the influence of each parameter on the measured outputs, as well as the possible interaction (correlation) between these effects. In this study, local sensitivity analysis is carried out using a modification of the Matlab toolbox IDEAS (acronym for IDEntification and Analysis of Sensitivity) (Muñoz-Tamayo et al., 2009), which allows the construction of the dynamic sensitivity equations through symbolic computation, and the numerical simulation of these latter equations.

The first-order sensitivity $s_j = \frac{\partial x}{\partial \theta_j}$ is the solution of:

$$\dot{s}_j = \left[\frac{\partial f}{\partial x}\right]_{(x,\theta,t)} s_j + \left[\frac{\partial f}{\partial \theta_j}\right]_{(x,\theta,t)} \tag{26}$$

where $\dot{x}(t) = f(x(t), \theta, t)$ is the state equation. The parametric sensitivity of the measured outputs is then evaluated as:

$$\left[\frac{\partial y_m}{\partial \theta_j}\right] = \left[\frac{\partial h}{\partial x}\right]_{(x,\theta,t)} s_j + \left[\frac{\partial h}{\partial \theta_j}\right]_{(x,\theta,t)} \tag{27}$$

where $y_m(t) = h(x(t), \theta, t)$ is the measurement equation.

The sensitivity functions are evaluated at the identified parameter values $\hat{\theta}$ and used to compute the Fisher Information Matrix (FIM), whose inverse is a (Cramer-Rao) lower bound of the variance-covariance matrix P of the parameter errors:

$$\hat{P} \succ F^{-1}(\hat{\theta}) \tag{28}$$

The *FIM* is computed as:

$$F(\hat{\theta}) = \sum_{i=1}^{N} \left[\frac{\partial y_m}{\partial \theta}\right]^T_{(t_i,\widehat{\theta})} \Sigma^{-1} \left[\frac{\partial y_m}{\partial \theta}\right]_{(t_i,\hat{\theta})} \tag{29}$$

where Σ is the covariance matrix of the measurement noise. When the latter is unknown it can be replaced by its ML estimate $\hat{\Sigma}$.

The variance of $\tilde{\theta}_j = \theta^* - \hat{\theta}$ can be approximated by $\sigma_j^2 = \hat{P}_{jj}$ and used to evaluate a 95% confidence interval for θ_j, i.e., $\left[\hat{\theta}_j \pm 2\sigma_j\right]$.

Figure **??** gives a graphical representation of the local normalized sensitivities of each variable with respect to each parameter. Highly sensitive variables with respect to the parameters correspond to colors close to red while less sensitive variables are indicated in blue. The element $(k,\ j)$ is computed as:

$$\sum_{i=1}^{n_t} \left| \frac{\hat{\theta}_j}{y_{m_k}(t_i, \hat{\theta})} \left[\frac{\partial y_{m_k}}{\partial \theta_j} \right]_{(t_i, \hat{\theta})} \right| \tag{30}$$

The corresponding correlation matrix of the estimated parameters is shown in Figure **??**. Trajectories of the output sensitivities at each point in time are shown in Figures **??** and **??**.

Figure **??** reveals low sensitivity with respect to parameter k_2, while Table **??** shows its relatively high uncertainty. Hence, the cancellation of k_2 is investigated. Values of the newly identified parameters do not present significant differences, but the cost function decreases from 8.31 to 7.24. This fully justifies the model simplification.

3. Hydrogen-based control

In this section the control developed by Rodriguez et al. (2006) is implemented and tested on respectively the reduced model and ADM1. Intrinsically, the controller is based on experimental observation, and does

not require a model for its design. However a comparative test allows to assess the consistency of the reduced model.

It has been demonstrated in (Pauss and Guiot, 1993) that a functional relation exists in steady state between hydrogen in the biogas and the organic loading rate (OLR). This relation can be computed using ADM1 and the reduced-order model. Both results are represented in Figure **??**.

The controller, described by equation (31) and represented in Figure **??**, includes an integral action with variable gain. In its original form, it uses the hydrogen concentration in the gas phase ($ppmH_2$) and the methane flow rate (q_{CH_4}) to determine the appropriate change in the dilution rate D in order to drive the process to the desired set-point.

$$\frac{dD}{dt} = D\,K\,f_{CH_4} f_{H_2} \tag{31}$$

In this expression, f_{H_2} and f_{CH_4} are a hydrogen factor and methane factor described by equations (32) and (33), respectively,

$$\begin{aligned} f_{H_2} &= \left(1 - \frac{ppmH_2}{ppmH_2^*}\right)^{1/m} \quad if\ ppmH_2 \leq ppmH_2^* \\ f_{H_2} &= \left(\frac{ppmH_2^*}{ppmH_2}\right)^{n} - 1 \quad if\ ppmH_2 > ppmH_2^* \end{aligned} \tag{32}$$

$$f_{CH_4} = \frac{\alpha q_{CH_4}^*}{q_{CH_4} + \alpha q_{CH_4}^*} \tag{33}$$

If the level of hydrogen is below the set-point ($ppmH_2^*$), f_{H_2} is positive, otherwise it is negative. This way, the hydrogen concentration in the gas phase determines the direction of change of the dilution rate, i.e. a positive

or negative time derivative (increase or decrease of the dilution rate). As shown in Figure **??**, f_{H_2} is a smooth function taking values between 1 and -1.

On the other hand, f_{CH_4} is a positive factor decreasing for increasing value of the methane flow rate (see Figure **??**), which will modulate the effect of f_{H_2}.

As a linear relationship exists between the hydrogen concentration (gas and liquid) and the hydrogen gas flow rate (see Figure **??**), equations (32) can be adapted accordingly

$$\begin{aligned} f_{H_2} &= \left(1 - \frac{q_{H_2}}{q^*_{H_2}}\right)^{1/m} \quad if\ q_{H_2} \leq q^*_{H_2} \\ f_{H_2} &= \left(\frac{q_{H_2}}{q^*_{H_2}}\right)^{n} - 1 \quad if\ q_{H_2} > q^*_{H_2} \end{aligned} \tag{34}$$

The performance of the controller is studied in the framework of the process start-up and shown in Figure **??**. In this figure the set-point is indicated in ppm H_2 using the conversion given in Figure **??**. However, it is not necessary to know this relation to apply the controller. Parameter values used for the controller are shown in Table **??**. The controller performance is satisfactory as it allows to start the process in a safe way. The application to both models is comparable, reenforcing the validation of the reduced-order model.

4. Experimental validation

In this section the proposed model structure is identified with experimental data from raw industrial wine processing wastewater. The wastew-

ater is treated in a $1m^3$ fully instrumented up-flow fixed bed digester. More details of the AD plant and its online instrumentation can be found in (Steyer et al., 2002).

All measurements are recorded with online sensors every 2 minutes, while the H_2 sensor performs a series of measurements every half hour. In order to improve simulation time, sampling of experimental data is decreased using the downsampling function from Matlab. Similarly, data is smoothed using a moving average filter with a span of 35 data points. The full non-smoothed set of data is used at the end of the identification when all parameters are identified together. The results of direct and cross-validation are presented in Figure **??** and Figure **??**, respectively.

The identification is performed as explained in Section 2.3. Table **??** shows the estimated parameter values with their respective uncertainties.

While the direct validation results are quite satisfactory, the cross-validation results are perfectible. Several causes can be observed:

- the number of available data sets is limited. More data sets would help consolidate the identification and validation;
- the experiments were not performed with monitoring of hydrogen in mind. Hence hydrogen is measured at a lower frequency, and the sensor was not calibrated accurately. The measurements of hydrogen are therefore more qualitative than quantitative.

5. Conclusions

Using a dynamic simulator based on ADM1, an informative data base is built, which is used to infer a low-order model mimicking the time evolution of the state variables of interest. In the framework of this study, attention is focused on the evolution of hydrogen, which has been identified as an important indicator of the process stability. The resulting model, with 4 states only, captures well the dynamics of ADM1, and is a good basis to develop control strategies. To fit a particular application, the model parameters can be estimated from experimental data, using conventional tools, as exemplified with real experimental data collected from a pilot plant at the Laboratoire de Biotechnologie de l'Environnement, Narbonne, France.

6. Acknowledgements

The authors acknowledge the support of WBI in the framework of the research project REC 11: Optimisation de la production de biogaz d'un réacteur de digestion anaérobie. The present work has been jointly supported by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian State, Science Policy Office (BELSPO). The authors are grateful to the reviewers for their insightful comments. The scientific responsibility rests with the authors.

7. References

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S. G., Rozzi, A., Vavilin, V. A., 2002. The IWA Anaerobic Digestion Model No 1 (ADM1). Water Science and Technology 45 (10), 65–73.

Beltramo, T., Ranzan, C., Hinrichs, J., Hitzmann, B., 2016. Artificial neural network prediction of the biogas flow rate optimised with an ant colony algorithm. Biosystems Engineering 143, 68–78.

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. Biotechnology and Bioengineering 75 (4), 424–438.

Bhattacharya, S., Sluder, J., Uberoi, V., 1995. Effects of 4-nitrophenol on H2 and CO levels in anaerobic propionate systems. Water Research 29 (5), 1249–1258.

Björnsson, L., Murto, M., Jantsch, T. G., Mattiasson, B., 2001. Evaluation of new methods for the monitoring of alkalinity, dissolved hydrogen and the microbial community in anaerobic digestion. Water Research 35 (12), 2833–2840.

Boe, K., 2006. Online monitoring and control of the biogas process. Ph.D. thesis, Institute of Environment and Resources, Technical University of Denmark.

Boe, K., Batstone, D. J., Steyer, J. P., Angelidaki, I., 2010. State indicators

for monitoring the anaerobic digestion process. Water Research 44 (20), 5973–5980.

Chen, L., Bastin, G., 1996. Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. Mathematical Biosciences 132 (1), 35–67.

Chen, Y., Cheng, J. J., Creamer, K. S., 2008. Inhibition of anaerobic digestion process: a review. Bioresource Technology 99 (10), 4044–4064.

Cord-Ruwisch, R., Mercz, T. I., Hoh, C.-Y., Strong, G. E., 1997. Dissolved hydrogen concentration as an on-line control parameter for the automated operation and optimization of anaerobic digesters. Biotechnology and Bioengineering 56 (6), 626–634.

Della Bona, A., Ferretti, G., Ficara, E., Malpei, F., 2015. Lft modelling andnidentification of anaerobic digestion. Control Engineering Practice 36, 1–11.

Dochain, D., Perrier, M., Pauss, A., 1991. Adaptive control of the hydrogen concentration in anaerobic digestion. Industrial & Engineering Chemistry Research 30 (1), 129–136.

Ficara, E., Hassam, S., Allegrini, A., Leva, A., Malpei, F., Ferretti, G., 2012. Anaerobic digestion models: a comparative study. Proceedings of the MATHMOD 2012 45 (2), 1052–1057.

García-Diéguez, C., Bernard, O., Roca, E., 2013. Reducing the Anaerobic Digestion Model No. 1 for its application to an industrial wastewater

treatment plant treating winery effluent wastewater. Bioresource Technology 132, 244–253.

Giovannini, G., Donoso-Bravo, A., Jeison, D., Chamy, R., Ruíz-Filippi, G., Wouwer, A. V., 2016. A review of the role of hydrogen in past and current modelling approaches to anaerobic digestion processes. International Journal of Hydrogen Energy 41 (39), 17713–17722.

Giraldo-Gomez, E., Goodwin, S., Switzenbaum, M. S., 1992. Influence of mass transfer limitations on determination of the half saturation constant for hydrogen uptake in a mixed-culture CH4-producing enrichment. Biotechnology and Bioengineering 40 (7), 768–776.

Hassam, S., Ficara, E., Leva, A., Harmand, J., 2015. A generic and systematic procedure to derive a simplified model from the Anaerobic Digestion Model No. 1 (ADM1). Biochemical Engineering Journal 99, 193–203.

Hawkes, F., Dinsdale, R., Hawkes, D., Hussy, I., 2002. Sustainable fermentative hydrogen production: Challenges for process optimisation. International Journal of Hydrogen Energy 27 (11), 1339–1347.

Hou, Y.-p., Peng, D.-c., Xue, X.-d., Wang, H.-y., Pei, L.-y., 2014. Hydrogen utilization rate: A crucial indicator for anaerobic digestion process evaluation and monitoring. Journal of Bioscience and Bioengineering 117 (4), 519–523.

Mailier, J., Remy, M., Wouwer, A. V., 2013. Stoichiometric identification with maximum likelihood principal component analysis. Journal of Mathematical Biology 67 (4), 739–765.

McKay, M. D., Beckman, R. J., Conover, W. J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Muñoz-Tamayo, R., Laroche, B., Leclerc, M., Walter, E., 2009. Ideas: a parameter identification toolbox with symbolic analysis of uncertainty and its application to biological modelling. 15th IFAC Symposium on System Identification, July 6-8,Saint-Malo, France 42 (10), 1271–1276.

Owhondah, R. O., Walker, M., Ma, L., Bill, N., B., I. D., Poggio, D., Pourkashanian, M., 2016. Assessment and parameter identification of simplified models to describe the kinetics of semi-continuous biomethane production from anaerobic digestion of green and food waste. Bioprocess and Biosystems Engineering 39, 977992.

Pauss, A., Guiot, S. R., 1993. Hydrogen monitoring in anaerobic sludge bed reactors at various hydraulic regimes and loading rates. Water Environment Research 65 (3), 276–280.

Rodriguez, J., Roca, E., Lema, J. M., Bernard, O., 2008. Determination of the adequate minimum model complexity required in anaerobic bioprocesses using experimental data. Journal of Chemical Technology and Biotechnology 83 (12), 1694–1702.

Rodriguez, J., Ruiz, G., Molina, F., Roca, E., Lema, J., 2006. A hydrogen-based variable-gain controller for anaerobic digestion processes. Water Science and Technology 54 (2), 57–62.

Ryhiner, G. B., Heinzle, E., Dunn, I. J., 1993. Modeling and simulation of anaerobic wastewater treatment and its application to control design: case whey. Biotechnology Progress 9 (3), 332–343.

Simeonov, I., Karakashev, D., 2012. Mathematical modelling of the anaerobic digestion including the syntrophic acetate oxidation. IFAC Proceedings Volumes 45 (2), 309–314.

Steyer, J.-P., Bouvier, J., Conte, T., Gras, P., Sousbie, P., 2002. Evaluation of a four year experience with a fully instrumented anaerobic digestion process. Water Science and Technology 45 (4-5), 495–502.

Voolapalli, R. K., Stuckey, D. C., 1999. Relative importance of trophic group concentrations during anaerobic degradation of volatile fatty acids. Applied and Environmental Microbiology 65 (11), 5009–5016.

Voolapalli, R. K., Stuckey, D. C., 2001. Hydrogen production in anaerobic reactors during shock loads-influence of formate production and H2 kinetics. Water Research 35 (7), 1831–1841.

Yu, H., Zhu, Z., Hu, W., Zhang, H., 2002. Hydrogen production from rice winery wastewater in an upflow anaerobic reactor by using mixed anaerobic cultures. International journal of hydrogen energy 27 (11), 1359–1365.

Yu, L., Wensel, P. C., Ma, J., Chen, S., 2013. Mathematical modeling in anaerobic digestion (ad). Journal of Bioremediation & Biodegradation (S4: 003).

Figures caption:

Fig 1: Noise-free data generated with ADM1 model. Upper left graph: evolution of the dilution rate $D(t)$. Continuous lines: identification set - dashed lines: cross-validation set. Variables concentration units are $kgCOD/m^3$. Flow rates units are m^3/d.

Fig 2: Log-likelihood costs related to the MLPCA-estimated subspaces. A plausible range is delimited by two chi-square quantiles (dashed lines). N is the number of state variables (component concentrations), n_s is the number of samples per variable.

Fig 3: Direct validation. Red dots: ADM1 noisy data - Blue lines model prediction. Variables concentration units are $kgCOD/m^3$. Flow rates units are m^3/d.

Fig 4: Cross validation. Red dots: ADM1 noisy data - Blue lines: model prediction. Variables concentration units are $kgCOD/m^3$. Flow rates units are m^3/d.

Fig 5: L1 norm of the normalized sensitivities.

Fig 6: Parameter correlation matrix.

Fig 7 and 8: Parametric sensitivity trajectories.

Fig 9: Steady-state simulation of H_2 gas at different OLR for ADM1 (red dots) and reduced model (blue stars).

Fig 10: Schematic view of the control.

Fig 11: Control functions for a) Hydrogen and b) Methane flow rate.

Fig 12: Linear relationship between hydrogen concentration and hydrogen gas flow rate.

627 **Fig 13:** Controller and system responses for ADM1 (red dashed lines)
628 and reduced model (blue continuous lines).

629 **Fig 14:** Linear regression relation between hydrogen flow rate and
630 dissolved hydrogen.

631 **Fig 15:** Direct validation. Red dots: experimental data - Blue lines
632 model prediction.

633 **Fig 16:** Cross validation. Red dots: experimental data - Blue lines
634 model prediction.

635 **Fig 1:**

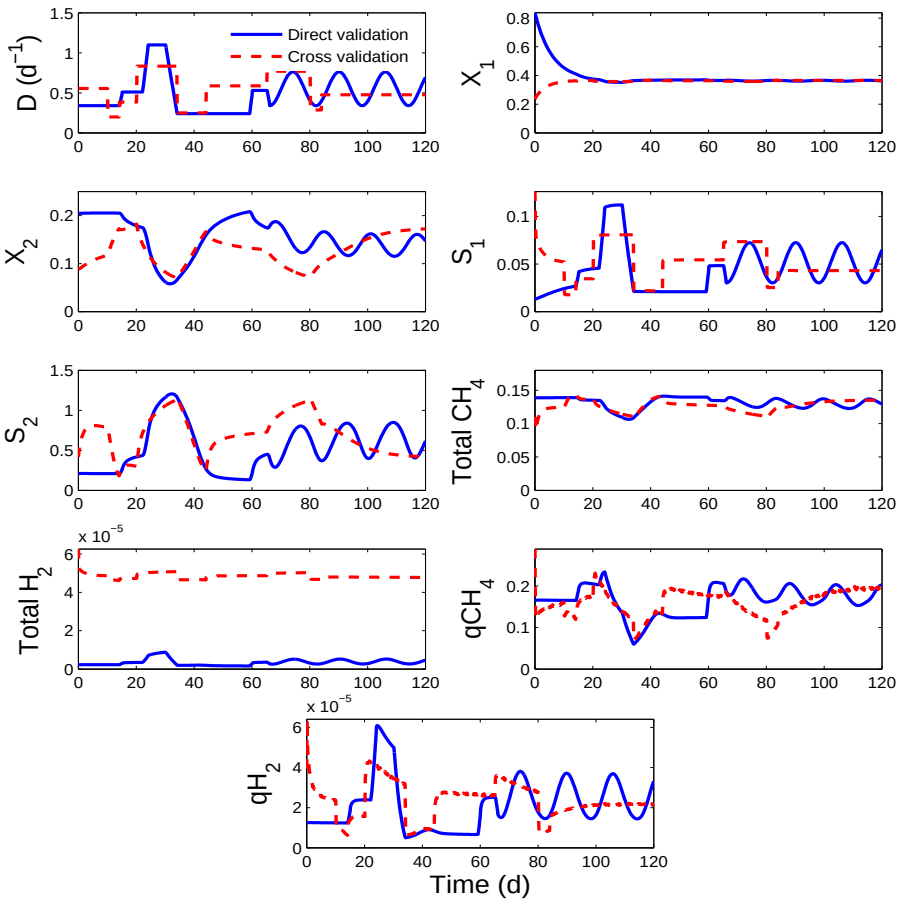


Figure 1:

636 **Fig 2:**

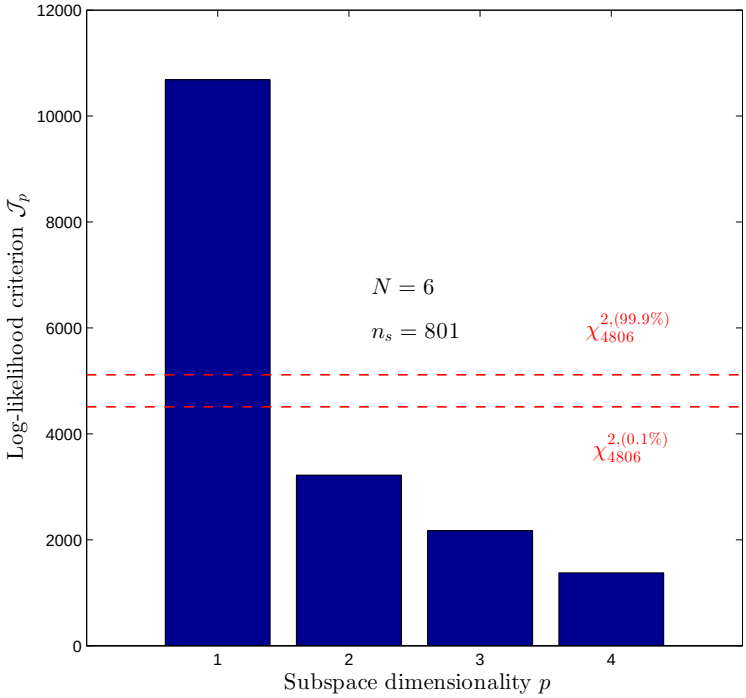


Figure 2:

637 **Fig 3:**

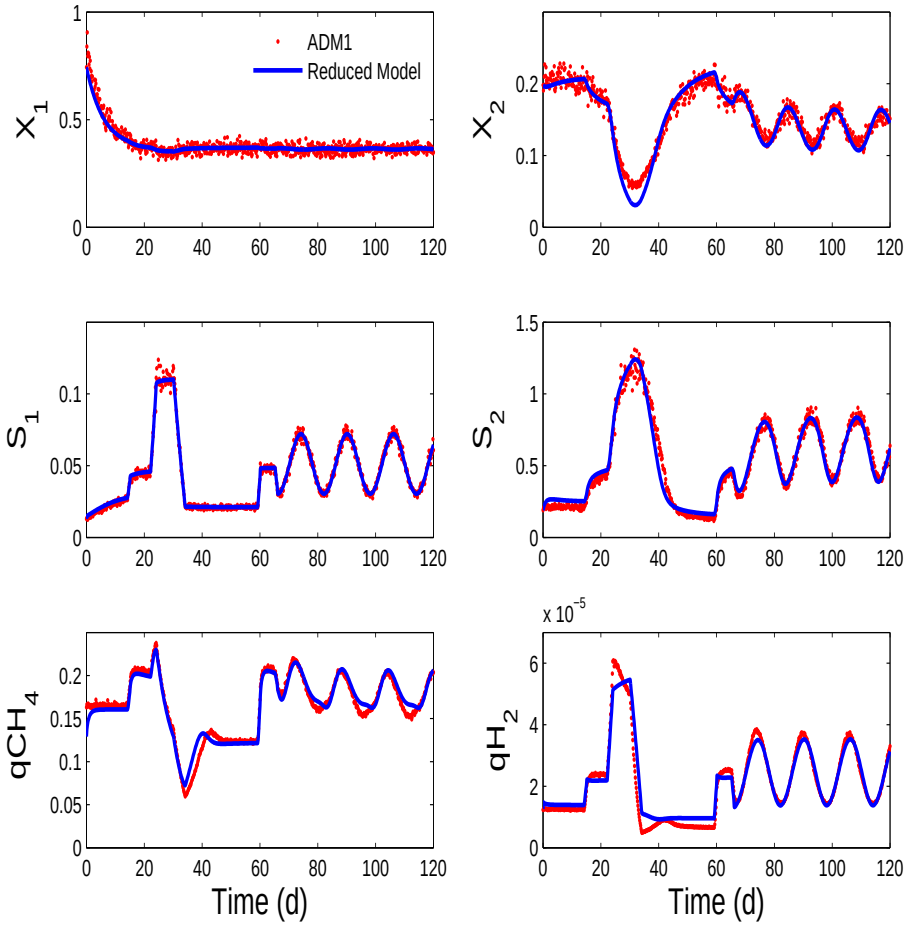


Figure 3:

638 **Fig 4:**

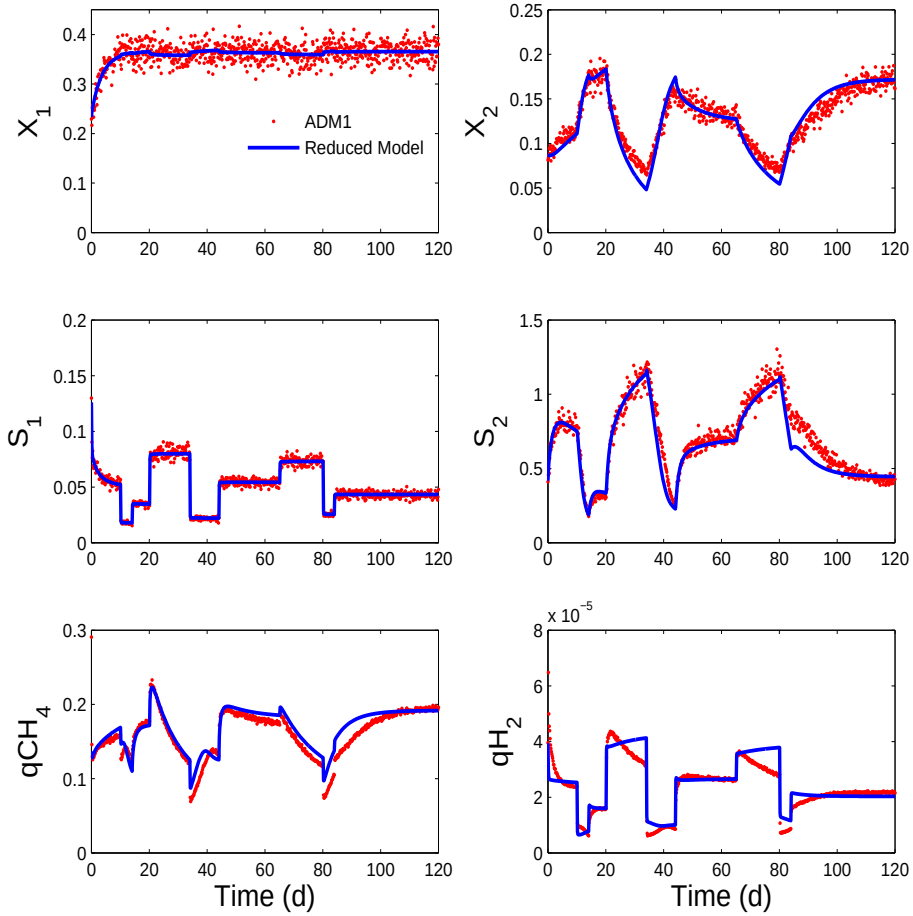


Figure 4:

639 **Fig 5:**

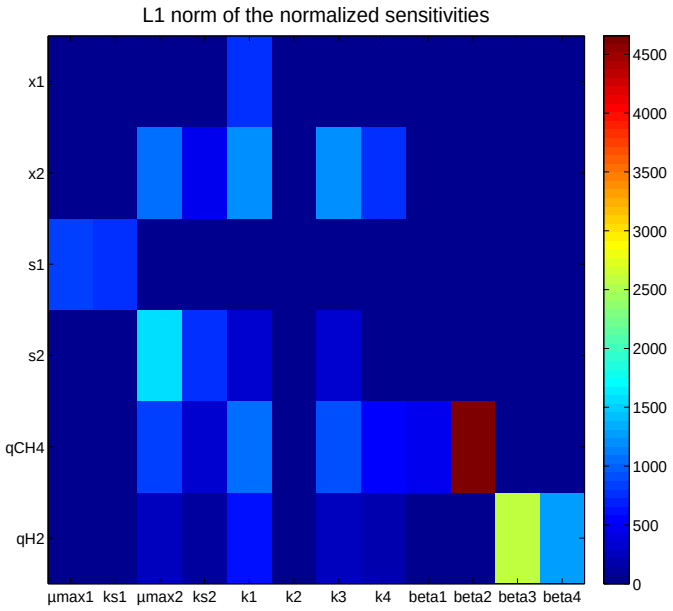


Figure 5:

640 **Fig 6:**

	μ^1_{max}	k_{s1}	μ^2_{max}	k_{s2}	k_1	k_2	k_3	k_4	β_1	β_2	β_3	β_4
μ^1_{max}	1	0.99	−0.12	−0.11	$7 \cdot 10^{-3}$	$2 \cdot 10^{-2}$	$-7 \cdot 10^{-2}$	$-7 \cdot 10^{-2}$	$-3 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$-1 \cdot 10^{-5}$	$-8 \cdot 10^{-6}$
k_{s1}		1	−0.11	−0.10	$4 \cdot 10^{-3}$	$2 \cdot 10^{-2}$	$-7 \cdot 10^{-2}$	$-7 \cdot 10^{-2}$	$-2 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$-1 \cdot 10^{-5}$	$-8 \cdot 10^{-6}$
μ^2_{max}			1	0.98	$-3 \cdot 10^{-2}$	$-2 \cdot 10^{-2}$	$-3 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$6 \cdot 10^{-2}$	$-5 \cdot 10^{-2}$	$-8 \cdot 10^{-7}$	$-4 \cdot 10^{-7}$
k_{s2}				1	$-2 \cdot 10^{-2}$	$-2 \cdot 10^{-2}$	$-5 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$6 \cdot 10^{-2}$	$-4 \cdot 10^{-2}$	$-7 \cdot 10^{-7}$	$-2 \cdot 10^{-7}$
k_1					1	0.95	0.94	0.85	$4 \cdot 10^{-2}$	$-1 \cdot 10^{-2}$	$3 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
k_2						1	0.90	0.89	$3 \cdot 10^{-2}$	$-1 \cdot 10^{-2}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
k_3							1	0.89	$5 \cdot 10^{-2}$	$-3 \cdot 10^{-2}$	$3 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
k_4								1	$5 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
β_1									1	−0.92	$2 \cdot 10^{-6}$	$9 \cdot 10^{-7}$
β_2										1	$-7 \cdot 10^{-7}$	$1 \cdot 10^{-6}$
β_3											1	0.94
β_4												1

Figure 6:

641 **Fig 7:**

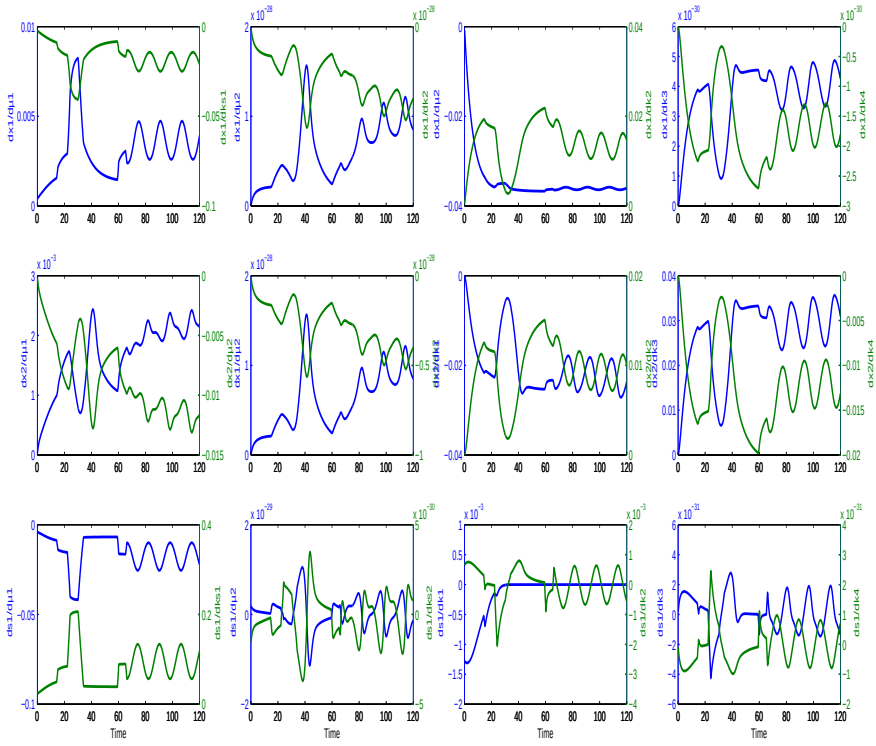


Figure 7:

642 **Fig 8:**

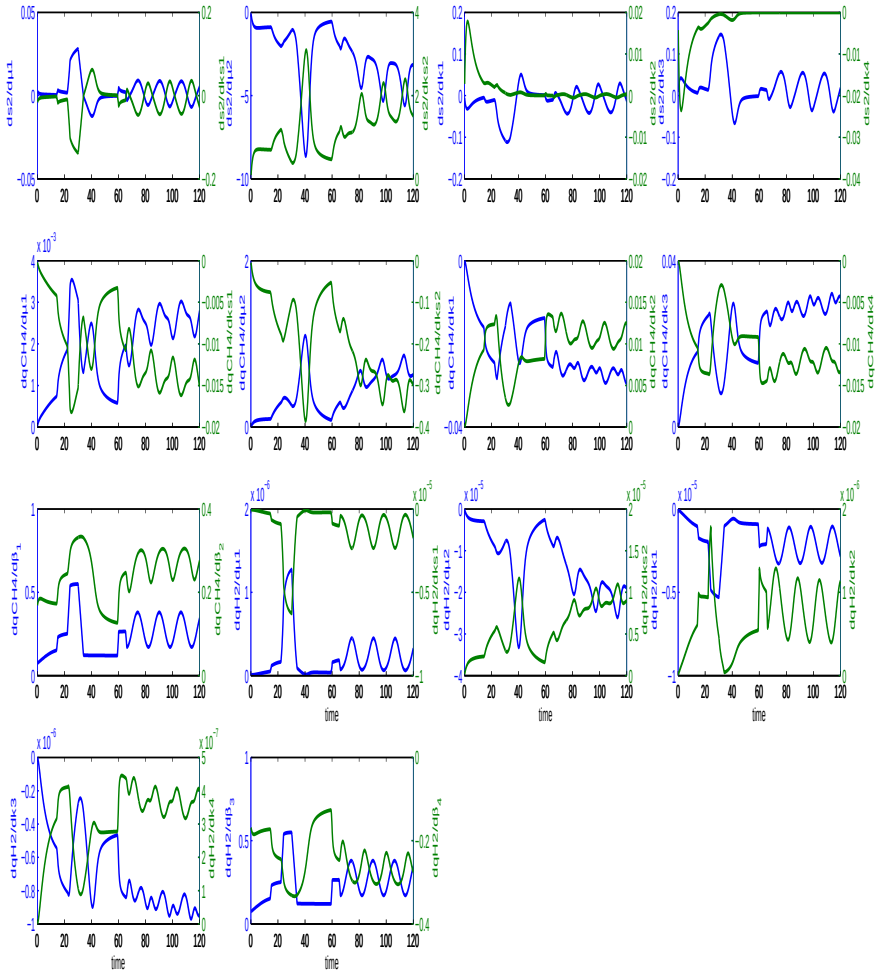


Figure 8:

643 **Fig 9:**

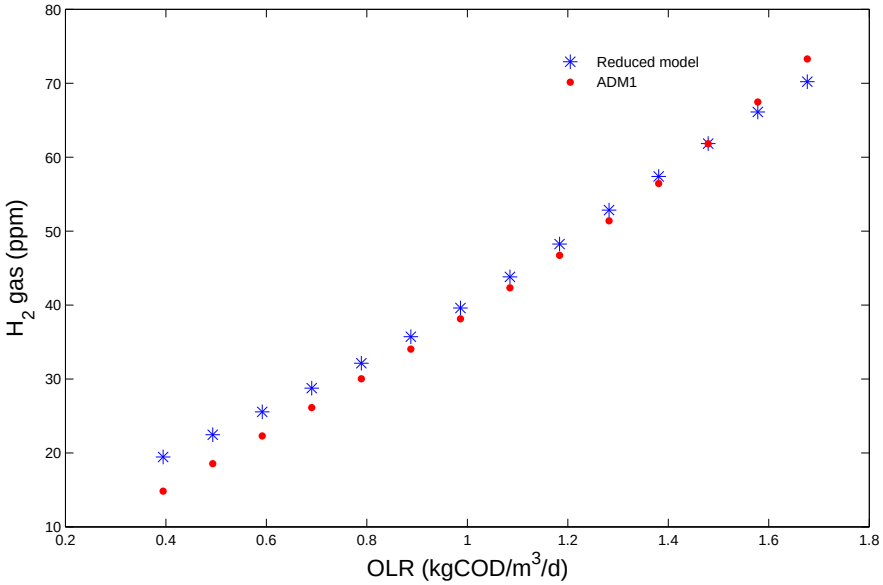


Figure 9:

644 **Fig 10:**

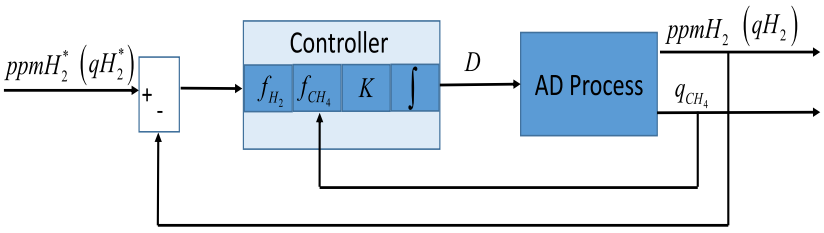


Figure 10:

645 **Fig 11:**

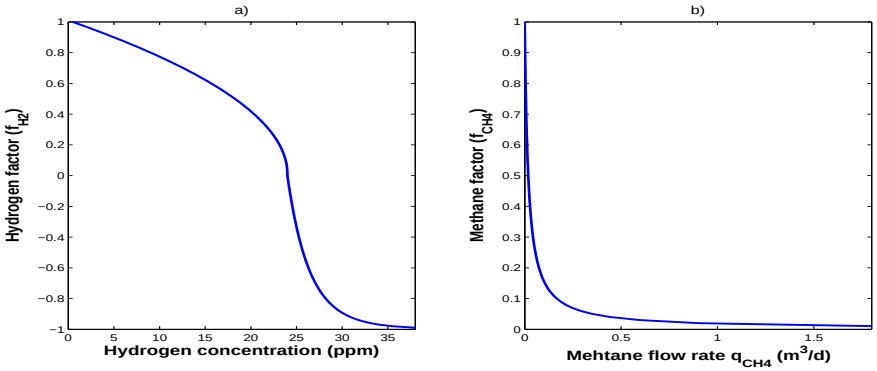


Figure 11:

646 **Fig 12:**

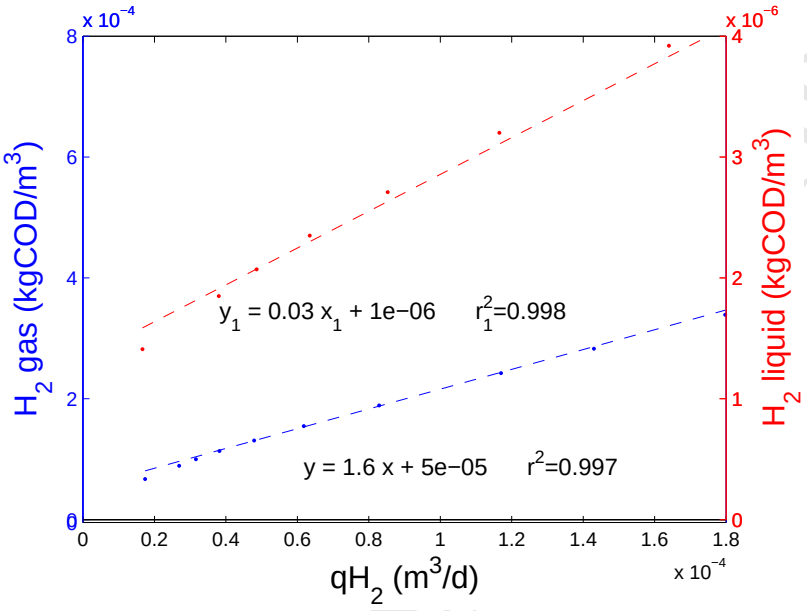


Figure 12:

647 **Fig 13:**

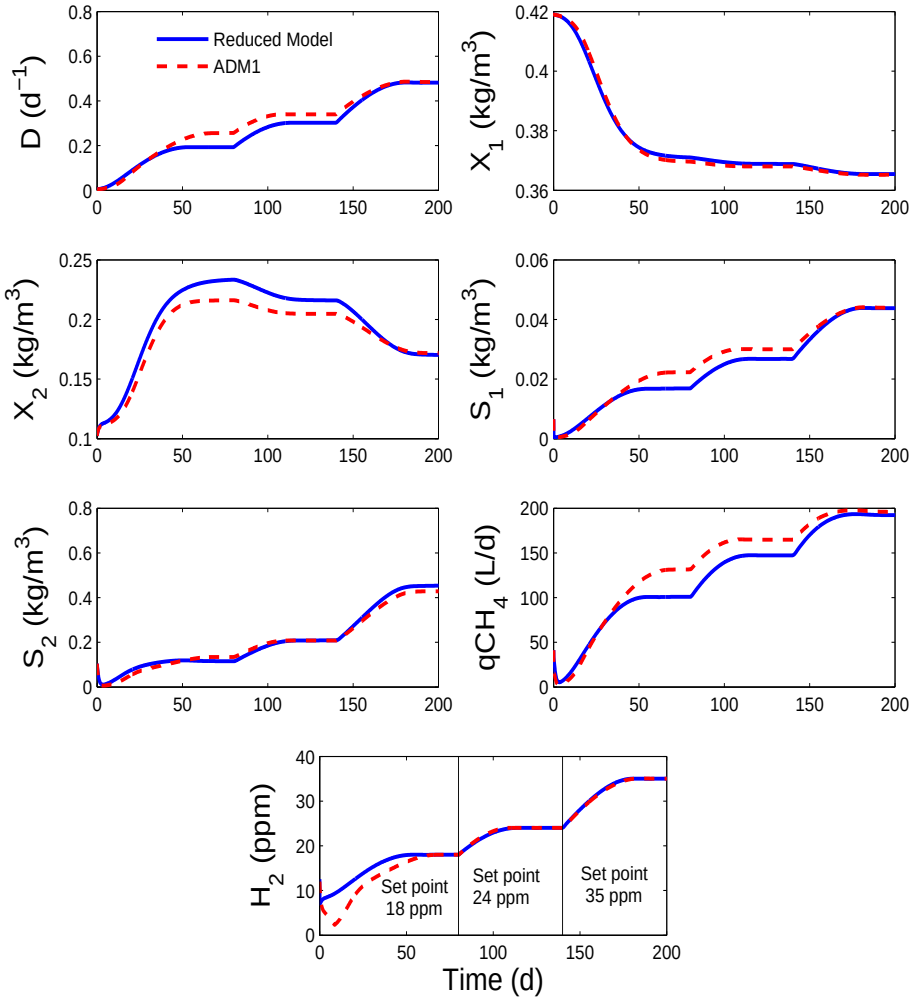


Figure 13:

648 **Fig 14:**

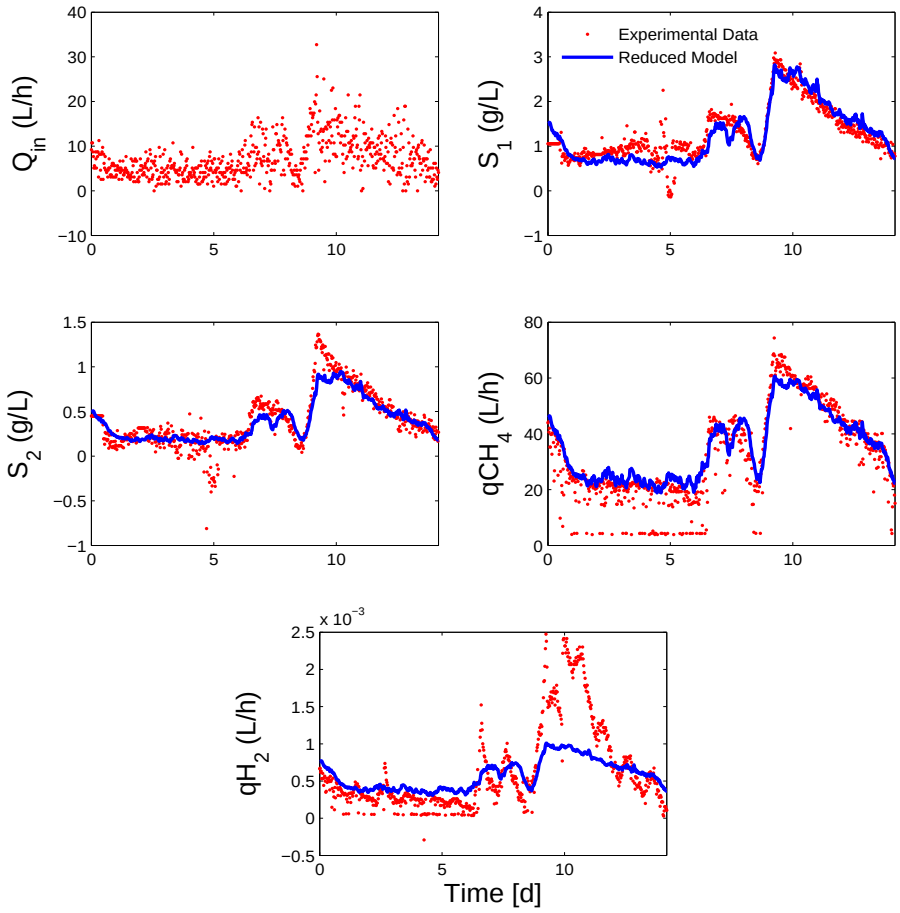


Figure 14:

649 **Fig 15:**

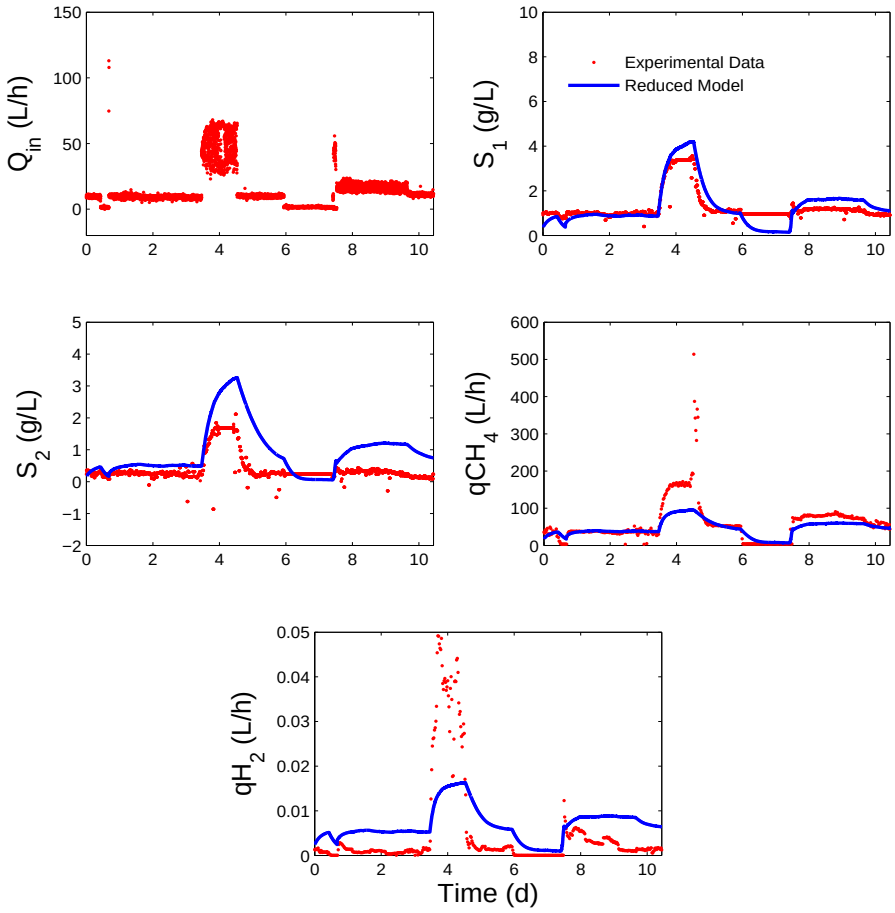


Figure 15:

	ADM1 influent concentrations		ADM1 influent concentrations
$S_{su,in}$	1.87 $kgCOD/m^3$	$S_{IC,in}$	0.065 $kmol/m^3$
$S_{aa,in}$	0.01 $kgCOD/m^3$	$S_{cat,in}$	0.089 $kmol/m^3$
$S_{fu,in}$	0.01 $kgCOD/m^3$	$S_{an,in}$	0.018 $kmol/m^3$
$S_{ac,in}$	0.01 $kgCOD/m^3$	$S_{va,in}$	0.01 $kgCOD/m^3$
$S_{bu,in}$	0.01 $kgCOD/m^3$	$S_{pro,in}$	0.01 $kgCOD/m^3$

Table 1: ADM1 input values used in the model reduction procedure

Parameter	Description	Obtain with MLPCA	Estimated	Units
k_1	S_1consmption yield	4.95	9.99 ±0.42	$kg\,S_1/kg\,X_1$
k_2	S_1 production yield	0.09	1e-12 ±0.99	$kg\,S_1/kg\,X_2$
k_3	S_2production yield	5.12	7.72 ±0.34	$kg\,S_2/kg\,X_1$
k_4	S_2consmption yield	8.99	11.25 ±0.86	$kg\,S_2/kg\,X_2$
β_1	CH_4proportional constant	-	0.41±0.05	$m^3\,m^3/kg\,X_1$
β_2	CH_4proportional constant	-	3.81±0.06	$m^3\,m^3/kg\,X_2$
β_3	H_2proportional constant	-	3e-4±0.05	$m^3\,m^3/kg\,X_1$
β_4	H_2proportional constant	-	1e-4±0.06	$m^3\,m^3/kg\,X_2$

Table 2: Stoichiometric parameters initially estimated using MLPCA, and identified values together with their 95% confidence intervals

Signals	Step		Step		Step		Ramp		Identification set	
Duration	43 days		120 days		120 days		120 days		120 days	
Noise	1%		5%		1%		1%		1%	
X_1	1	0	1	0	1	0	1	0	1	0
X_2	0	1	0	1	0	1	0	1	0	1
S_1	-6.9	4.8	-3.8	-2.0	-3.6	-2.5	-4.6	-0.6	-4.6	0.09
S_2	8.5	-16.6	3.6	-5.6	3.3	-2.5	4.5	-7.5	5.1	-8.9
CH_4	0.39	-9e-2	0.24	0.24	0.22	0.27	0.26	0.21	0.27	0.21
$H2$	5e-5	-1e-4	2e-5	-3e-5	2e-5	-2e-5	3e-5	-4e-5	3e-5	-6e-5

Table 3: Influence of data on the stoichiometric coefficients deduced by MLPCA

Parameter	Description	Estimated	Units
μ_{max1}	Maximum growth rate first reaction	3.62±9e-5	$1/d$
μ_{max2}	Maximum growth rate second reaction	0.48±4e-5	$1/d$
K_{S1}	Half-saturation constant first reaction	0.61±1e-2	$Kg/m3$
K_{S2}	Half-saturation constant second reaction	0.46±4e-3	$Kg/m3$

Table 4: Estimated kinetic parameters together with their 95% confidence intervals

Identification	Cost Function	Set of parameters	Number of iterations
0	509.66		
1	132.45	Kinetics	116
2	90.84	Stoichiometry	1422
3	14.64	Kinetics + stoichiometry	2063

Table 5: Cost function values obtained after each identification step with their respective number of iterations

Parameter	Value	Units
K	0.4	kg/m^3
$q^*_{CH_4}$	0.18	m^3/d
m	2	-
n	10	-
α	0.1	-

Table 6: Controller parameters used during start-up experiments

Parameter	Description	Estimated	Units
μ_{max1}	Maximum growth rate first reaction	0.60±0.02	$1/d$
μ_{max2}	Maximum growth rate second reaction	0.27±0.15	$1/d$
K_{S1}	Half-saturation constant first reaction	5.10±1.32	g/L
K_{S2}	Half-saturation constant second reaction	0.70±0.13	g/L
K_1	S_1consumption yield	21.3±5.23	$g\,S_1/g\,X_1$
K_4	S_2consumption yield	10.0±2.39	$g\,S_2/g\,X_2$
β_1	Proportional constant first reaction	5.0±1.82	L^2/g
β_2	Proportional constant second reaction	3.9±1.23	L^2/g
β_3	Proportional constant first reaction	1e-4±4e-3	L^2/g
β_4	Proportional constant second reaction	1e-5±2e-5	L^2/g

Table 7: Estimated parameters together with their 95% confidence intervals

Highlights:

- A 2-step dynamic model of anaerobic digestion is derived from ADM1.
- The model involves the concentration of hydrogen, which might be used to monitor the process.
- Maximum likelihood principal component analysis is used to infer the reaction number and stoichiometry.
- A hydrogen control strategy is tested based on the 2-step model and ADM1.