Water Research 134 (2018) 209–225

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres

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

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

ARTICLE INFO

Article history:
Received 14 June 2017
Received in revised form
5 December 2017
Accepted 15 January 2018
Available online 3 February 2018

Keywords:
Mathematical modeling
Parameter estimation
Principal component analysis
Sensitivity analysis

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 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.

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

☆ In memory of Prof. Gonzalo Ruiz-Filippi

* Corresponding author. Service d'Automatique, Université de Mons, 31 Boulevard Dolez, 7000 Mons, Belgium.
E-mail address: giannina.giovannini@umons.ac.be (G. Giovannini).

https://doi.org/10.1016/j.watres.2018.01.036
0043-1354/

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 activated 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 nitrogen 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