

Selection of variables using factorial discriminant analysis for the state identification of an anaerobic UASB–UAF hybrid pilot plant, fed with winery effluents

M. Castellano*, G. Ruiz-Filippi**, W. González*, E. Roca*** and J.M. Lema***

*Department of Statistics and O.R., University of Santiago de Compostela, Spain (E-mail: mcaste@usc.es)

**School of Biochemical Engineering, Pontificia Universidad Católica de Valparaíso, Chile

***Department of Chemical Engineering, School of Engineering, University of Santiago de Compostela, Spain

Abstract Anaerobic wastewater treatment has become a widely used method for wastewater depuration, and has been applied in a wide range of situations, from urban wastewater to highly toxic industrial wastewater. Particularly it has been successfully applied to the treatment of the beverage industries effluents. To avoid the destabilization of the system a monitoring diagnosis and control system of the depuration processes is necessary. The cost of this system is an important issue, that depends on the number of parameters that must be controlled for an adequate performance of a wastewater plant control system. This work shows how the classic statistical classification techniques can be applied to determine the number variables that must be monitored to achieve an adequate performance of anaerobic UASB–UAF hybrid Pilot Plant monitoring and control system. The obtained results had not been unique, so different combinations of variables can be selected for a good wastewater treatment process control. Economic or technical criteria may be considered to determine the final variables set in each particular situation.

Keywords Anaerobic digestion; FDA; process state identification; UASB–UAF reactor; winery effluents

Introduction

During recent decades, anaerobic wastewater treatment process has been successfully applied in a wide range of wastewater situations (Huang *et al.*, 2000). This process requires low energy and has low sludge production so it is an interesting treatment for industries from many different sectors, in particular for beverage industries. One of the main difficulties for the operation of such a process is the achievement of stable operational conditions, without the accumulation of intermediate products, such as volatile fatty acids (VFA), hydrogen (H₂), carbon monoxide (CO), among others (Pullammanappallil *et al.*, 2001). Non-stable operation is usually due to changes in the influent characteristics, of both quality and quantity, and sporadic presence of toxic compounds. These changes are frequent in industrial wastewater treatment processes; since influent characteristics depend on the production schedule, stream up of the treatment plant, which is usually not constant or stable, and it can produce variations in the characteristics of the produced wastewater. Perturbations may produce a high destabilization or even the complete failure of the process, so it is highly important to consider an adequate monitoring system, and even better, a monitoring–diagnosis and control (MD&C) system.

A MD&C system for an anaerobic digestion process should be capable an early and automatic detection of overload periods and other types of perturbation such as the presence of toxic or inhibitory compounds or sudden changes in pH. Early detection is needed in order to allow the MD&C system to take the corrective actions to drive

the process to normal operation before process performance is irreversibly affected (Marsili-Libelli and Beni, 1996). To develop a MD&C system, the first step is to select a group of process variables, which can give information about the metabolic state of the process. These variables should have three main characteristics: low response delay, high sensibility and low cost of both, sensor itself and its operation-maintenance requirements.

It has been reported that some variables fulfil these requirements (see Table 1), but there is not a general conclusion about the best variable or group of variables which describe the process and allow the identification of non-stable operation or diagnosis of the process state. However, in the works mentioned in Table 1, only a few variables were compared at the same time and objective tools for variable selection were not applied.

The aim of this work is to determine the minimum number of monitored variables for process state identification, using factorial discriminant analysis, FDA (Peña, 2002). The factorial functions selected define a diagnosis chart, called territorial map, which helps to classify and diagnose the process performance. Process response variables were analysed during several perturbations in a fully instrumented anaerobic wastewater treatment pilot plant operated for the treatment of diluted wine.

Material and methods

Experimental setup and conditions

A UASB–UAF pilot plant fed with diluted wine was used for the experiments. The measurement devices were: feed and recycling flow meters, pH meter; inflow and reactor Pt100, gas flow meter, infrared gas analyser (CH_4 and CO), gas hydrogen analyser and TOC/TIC combustion analyser. The sensors produce a signal every 5 seconds and every 15 minutes a moving average window was saved in the database. Other parameters were calculated using the measured variables: methane flow rate (QCH_4), hydrogen flow rate (QH_2) and organic loading rate (OLR). 26 variables were used to follow the process, Ruiz *et al.* (2001).

The reactor was operated at stable conditions for more than a month at an OLR of 5 kg COD/ $\text{m}^3\cdot\text{d}$ before the experimentation. Three consecutive increases, of the OLR were applied in order to obtain three different steady states (plus the initial steady state). Steady-state data were then labelled according to the state they belong to, (states 1 to 4). Table 2 presents the characteristics of each state. The duration of each state was around 5 days, time considered enough to achieve steady state because the HRT was in the range of 0.6 to 1.5 d. Reactor was fed with wine, diluted to the desired COD concentration.

Table 1 Summary of reported variables used for process state identification on operation of anaerobic digesters

Variable(s)	Reference
Gas flow rate and H_2/CH_4 in the gas phase	Huang <i>et al.</i> (2000)
H_2/CO in the gas phase	Hickey and Switzenbaum (1991)
	Hickey <i>et al.</i> (1989)
	Hickey and Switzenbaum (1990)
	Hickey <i>et al.</i> (1987)
	Switzenbaum <i>et al.</i> (1990)
H_2 in the gas phase	Kidby and Nedwell (1991)
Gas flow rate and CH_4 in the gas phase	Hickey <i>et al.</i> (1991)
Alkalinities (total and partial) in the liquid phase	Ripley <i>et al.</i> (1986)
pH in the liquid phase and gas flow rate	Buffière <i>et al.</i> (1995)

Table 2 Information about the four step organic loading rate increase of the hybrid UASB-UAF anaerobic reactor up to a complete destabilization of the process, being NO: normal operation; HO: hydraulic overload; OO: organic overload

State	Time (d)	OLR (kg COD/m ³ ·d)	Feed flowrate (Qa) (L/h)	TOC influent (mgC/L)
1: NO	0–4	5	22	3,000
2: HO	4–9	15	66	3,000
3: HO + OO	9–14	28	66	4,500
4: HO + OO	14–15.5	32	66	6,000

Multivariate analysis

FDA is a supervised discrimination or classification technique. This kind of supervised discrimination method is used when data sets of well-classified data are available. When no well classified data are available cluster methods are an adequate option. The discrimination rules are based on linear combinations of the observed variables, called discriminant factors. There are many discriminant methods using different transformation of the random variables.

The FDA requires the knowledge of a well classified set of data (matrix X), divided in g groups. Groups are stored in a vector Y with values from 1 to g and data classified in group i is $X(Y=i)$. FDA seeks for factors where the projections of data are, as well as possible, well classified according to the g *a priori* known groups. These factors divide the multivariable space of data using hyperplanes (Peña, 2002). The selection of the factors (or orthogonal hyperplanes) is made in order to minimize the probability of misclassification. There is no assumption about the underlying relationship model between the variables, it only tries to make projections of the variables into subspaces where the g known classes can be separated. The obtained projections can be used to classify new observations, with known or unknown belonging group, in a matrix Z . The selection of factors in FDA can be solved analytically or estimated. The analytical study of the equations involved shows that the discriminant factors, minimizes the Mahalanobis distance into each group and maximizes the distance between groups, providing compact groups that are spread as much as possible in the space.

Instead of selecting the best linear combination using the whole set of variables, as the FDA usually does, the aim in this work is to minimize the number of variables (so of measurements) needed to distinguish between the different states; therefore the FDA was applied to each combination of different number of variables, from 1 to 26, and the discriminant capacity of each combination was evaluated using the rate of good classified data. When one or more groups of variables can separate the different states it will not be necessary go on with greater groups of variables.

Results

The FDA procedure was applied considering all the groups of one and two variables (351 combination). Combination groups of three variables were not needed to be analysed since complete classification was achieved with two variables. There are several combinations that give 100% good classification of steady-state data, so to select only a few number of variables, technical and economical reasons were considered.

Process state classification capability of each variable independently was evaluated by FDA (see Figure 1). Only four variables give 100% good classification of steady-state data. All these variables are related to gas phase (QH_2 , H_2 , Q_{gas} and P). P and Q_{gas} are mutually exclusive variables, since both are related by the head loss in the gas pipe. Furthermore, QH_2 , Q_{gas} and P , present non-bijective behaviour. Therefore, for a given value of these

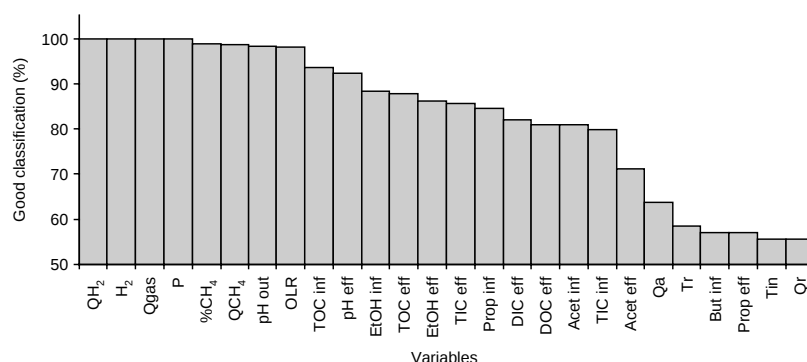


Figure 1 Percentage of good classification applying FDA to one variable. Four variables give 100% classification capability

variables more than one state is associated to it, meaning miss classification. For example in the case of Qgas, an under-loaded state will produce low levels of gas but this also happens in an inhibited state. For this reason, even if these variables present good classification of the studied data, they are not suitable for classification of states others than those presented here. A second variable will be held to circumvent the non-bijectivity behaviour. If only one variable is to be used, H₂ seems to be the best one among the studied ones, since it presents 100% good classification and has a bijective (one to one) behaviour.

FDA was applied to 325 combinations of pairs of variables. A table showing the good-classification rates of all the combinations is too long to present in this paper. In this case, there are 137 pairs of variables which give 100% good classification. Some variables present bad classification capability when they are used alone, but in combination with other variables, classification increased even to 100%. For example, Qa achieves 63.6% correct good classification (Figure 1), but in combination with Qgas, 100% of good classification is achieved.

Any of the 137 couples can be used for classification and this fact can explain the multiplicity of recommendations reported so far (see Table 1). According to the results presented in this work, there is not a unique couple of variables that can be used for correct classification, but there are several combinations that give great classification capability. To reduce the number of couples of variables, other criteria should be applied. For example, normal operation of industrial plants are conducted at constant temperature, influent pH and recirculation flow rate, so these variables can be excluded from the analysis. Moreover, specific substance determinations in the liquid phase are rare in industrial application, so they can also be excluded. Finally, Qgas and P are correlated, as was explained before, so the information of these two variables is redundant, making it possible to exclude one without loss of information. TIC/TOC on-line measurement is not common at industrial scale because of the high cost of the on line equipment and should be excluded from the analysis. But DOC eff can be proportional to the intermediate alkalinity, since most of the dissolved organic compounds are VFA. Also DIC eff is related to partial alkalinity (bicarbonate). These two alkalinity related variables can be determined on line since new analyzers are available on the market at a reasonable cost (De Neve *et al.*, 2004; Bernard *et al.*, 2005). Then these two variables will be kept in the analysis. Using these new criteria the number of couples of variables that yield a 100% of good classification is reduced to 28.

Going further, variables in the liquid phase are supposed to present higher response times than the gas phase variables (Ruiz *et al.*, 2002). From a dynamical point of view,

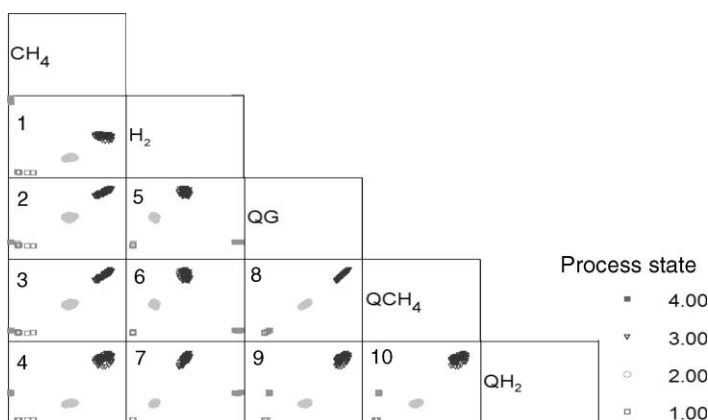


Figure 2 Classification obtained with different couples of variables selected considering technical-economical and dynamic issues

these variables will be of interest because of the need to detect early changes in process state. With this criterion pH eff, DOC eff, DIC eff and TIC eff were excluded from the analysis. The selected variables are related to gas phase (QH₂, H₂, Qgas and P). **Figure 2** shows the FDA results, i.e. the Discriminant functions for the combinations of these variables. Combinations 2, 3 and 8 are not considered since they give bad classification, and the rest obtain a 100% of good classified data.

Usually the basic instrumentation of an industrial wastewater treatment plant includes feed flowrate, effluent pH, influent and effluent temperature. Considering this basic instrumentation and only one couple of the key variables determined in this work, for example methane flow rate and hydrogen concentration in the gas phase, it is possible to build a so-called Territorial Map. Data are plotted in the plane of the first two discriminant functions (DF1 and DF2). FDA was developed for the steady-state data and, according to the Mahalanobis distance between groups, the limits or bounds of each state can be defined.

For future observations or for the non-steady-state data, DF1 and DF2 can be computed and each point can be plotted in the territorial map, giving a simple tool for

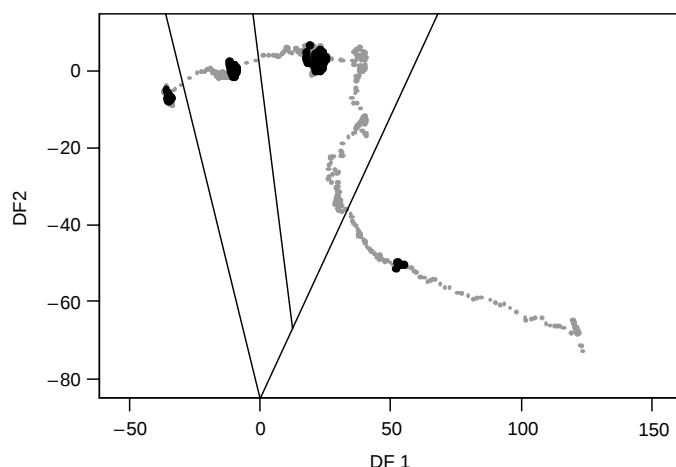


Figure 3 Territorial map for a typical instrumentation scheme. Black points represent the steady-state data used for FDA. Grey points indicate the non steady state data. Four different states are identified (number 1 to 4)

state estimation and diagnosis. Figure 3 presents the territorial map for the data used in this work. Black points correspond to steady-state data used to build the FDA model while grey points represent the DF1 and DF2 projection of the non-steady-state data. It can be seen that there is a clear path from one group to an other, allowing analysis at the same time of the trend and the state of the process.

Conclusions

The results obtained in this study show the utility of statistical methodologies for detecting state conditions of an anaerobic digestion process for the treatment of winery wastewaters. A methodology that, using steady state identification, is capable of selecting the minimum number of variables that allows a complete identification of the anaerobic digestion process state among the four states studied here, including normal operation, hydraulic overload, organic overload and complete destabilization of the process, was presented.

When only one variable for process state identification is used, H_2 concentration in the gas phase seems to be the best, since it has a high discriminatory ability among the process states studied. When two variables for process state identifications are considered, there are several combinations of variables that accomplish complete classification (137 in this case). Using technical, economical and dynamical aspects it was possible to reduce this number of combinations to 7. All of these combinations provide 100% of good classification and are suitable for industrial application. Applying FDA to a set of data of a common instrumentation scheme, it is possible to build a territorial map, which gives information about the state and the trend of the process. This is a useful tool for process monitoring and state diagnosis, which is based on linear relationships of simple application, both for an automatic diagnosis system and by a local operator performing a manual supervision of the process.

Acknowledgements

The Spanish National R&D Program and European Regional Development Fund (ERDF) for the CICYT project Anacom CTQ2004-07811-C02-01 and MTM2005-00820.

References

- Bernard, O., Chachuat, B., Hélias, A., Le Dantec, B., Sialve, B., Steyer, J.-P., Lardon, L., Neveu, P., Lambert, S., Ratini, P., Frattesi, S., Lema, J., Roca, E., Ruiz, G., Rodríguez, J., Franco, A., Vanrolleghem, P., Zaher, U., De Pauw, D.J.W., De Neve, K., Lievens, K., Dochain, D., Schoefs, O., Fibrianto, H., Farina, R., Alcaraz Gonzalez, V., Gonzalez Alvarez, V., Lemaire, P., Martinez, J.A., Duclaud, O. and Lavigne, J.F. (2005). TELEMAC: an integrated system to remote monitor and control anaerobic wastewater treatment plants through the internet. *Water Sci. Technol.*, **52**(1/2), 457–464.
- Buffière, P., Steyer, J.P. and Moletta, R. (1995). A new control strategy for highly loaded anaerobic digestion processes. In: *Proceedings of International Workshop on Monitoring and Control of Anaerobic Digestion Process*, December 6–7th, Narbonne, France, pp. 85–89.
- De Neve, K., Lievens, K., Steyer, J.P. and Vanrolleghem, P. (2004). Development of an on-line titrimetric analyser for the determination of volatile fatty acids, bicarbonate, and alkalinity [Telemac contribution #7]. *10th Congress on Anaerobic Digestion (AD10-2004)*, Montreal, Canada, August 29–September 23, pp. 1341–1344.
- Hickey, R. and Switzenbaum, M. (1990). Behavior of carbon monoxide as a trace component of anaerobic digester gases and methanogenesis from acetate. *Environ. Sci. Technol.*, **24**(11), 1642–1648.
- Hickey, R. and Switzenbaum, M. (1991). The response and utility of hydrogen and carbon monoxide as process indicators of anaerobic digesters subject to organic and hydraulic overloads. *J. Water Pollut. Control Federation*, **63**(2), 129–140.

- Hickey, R., Vanderwielen, J. and Switzenbaum, M. (1987). The effect of organic toxicants on methane production and hydrogen gas levels during the anaerobic digestion of waste activated sludge. *Water Res.*, **21**(11), 1417–1427.
- Hickey, R., Vanderwielen, J. and Switzenbaum, M. (1989). The effect of heavy metals on methane production and hydrogen and carbon monoxide levels during batch anaerobic sludge digestion. *Water Res.*, **23**(2), 207–218.
- Hickey, R., Wu, W., Veiga, M. and Jones, R. (1991). Start-up, operation, monitoring and control of high-rate anaerobic systems. *Water Sci. Technol.*, **24**(8), 207–255.
- Huang, Y., Huang, G., Chou, S. and Cheng, S. (2000). Hydrogen as a quick indicator of organic shock loading in UASB. *Water Sci. Technol.*, **42**(3/4), 43–50.
- Kidby, D. and Nedwell, D. (1991). An investigation into the suitability of biogas hydrogen concentration as a performance monitor for anaerobic sewage sludge digester. *Water Res.*, **25**(8), 1007–1012.
- Marsili-Libelli, S. and Beni, S. (1996). Shock load modelling in the anaerobic digestion process. *Ecol. Model.*, **84**, 215–232.
- Peña, D. (2002). Multivariate data análisis (in Spanish), McGraw Hill-Interamericana de España S.A.U.
- Pullammanappallil, P., Chynoweth, D.P., Lyberatos, G. and Svoronos, S.A. (2001). Stable performance of anaerobic digestion in the presence of a high concentration of propionic acid. *Bioresour. Technol.*, **78**, 165–169.
- Ripley, L., Boyle, W. and Converse, J. (1986). Improved alkalimetric monitoring for anaerobic digestion of high-strength wastes. *J. Water Pollut. Control Federation*, **58**(5), 406–411.
- Ruiz, G., Rodríguez, J., Baeza, J., Roca, E. and Lema, J.M. (2001). Advanced monitoring and supervision of an anaerobic pilot plant. In *Proceedings of the Anaerobic Digestion World Congress (AD9)*, (Vol. 2), van Velsen, A.F. and Verstraete, W. (eds.), Technologisch Instituut vzw, Antwerpen, pp. 207–210.
- Ruiz, G., Roca, E. and Lema, J.M. (2002). Variables selection for process state identification of non steady state data of an anaerobic digester (In Spanish), *Proceedings of the VII Latinamerican workshop and seminar on anaerobic digestion*. Mérida–México, (Vol. 2), pp. 166–172.
- Switzenbaum, M., Giraldo, E. and Hickey, R. (1990). Monitoring of the anaerobic methane fermentation process. *Enzyme Microbial Technol.*, **12**, 722–730.