

## **Anaerobic digestion process parameter identification and marginal confidence intervals by multivariate steady state analysis and bootstrap.**

G. Ruiz<sup>a\*</sup>, M. Castellano<sup>b</sup>, W. González<sup>b</sup>, E. Roca<sup>a</sup> and J.M. Lema<sup>a</sup>

<sup>a</sup> Department of Chemical Engineering, Institute of Technology.

<sup>b</sup> Department of Statistics and Operation Research.

University of Santiago de Compostela. E-15782, Spain

### **Abstract**

There are a few works related to on line steady state detection algorithms and less with parameter estimation. This work used Principal Component Analysis (PCA) for reduction of the dimension of the data space and producing independent variables, allowing the application of the multivariate Cao and Rhinehart algorithm in steady state detection. Once steady states were detected, model parameters can be calculated by steady state mass balance equations and non linear fit of data. Bootstrap method was used in order to approximate the parameters estimators distribution and confidence boundaries for the kinetic model. These methodologies were applied to a case of anaerobic wastewater digestion process where four different organic loading rates (OLR) were applied.

**Keywords:** PCA, steady state, parameter estimation, bootstrap.

### **1. Introduction**

Parameter identification for wastewater treatment process is usually determined by model fit to dynamical data both, in batch and continuous operation (Batstone, 1999; Markel et al, 1996; Lokshina, 1999). These methods have the disadvantage of not consider time delay in dynamic data due to unknown time dependence of data, especially due to non-ideal liquid flow pattern.

Usually steady state identification is analysed by an expert or an operator. In many cases, it is not possible to count with the expert advice. Furthermore, experts are always subjected to human error in recognition of steady state, especially when measurements are noisy and process changes are slow (Szela and Rhinehart, 2002). Consequently, it is important to define an algorithm for steady state detection, specially an algorithm for on line steady state detection. Ruiz et al (2004) have developed a multivariate extension of the Cao and Rhinehart (1995) methodology giving good results.

Because of the reduced number of steady states available from process data and because the use of mean data of steady state data sets, usually high estimation errors can arise

---

\* Author to whom correspondence should be addressed: [gruiz@usc.es](mailto:gruiz@usc.es)

from parameter identification, so it is necessary to increase as much as possible the accuracy of the identification procedure. Re-sampling of data by means of bootstrap, a technique invented by Bradley Efron (Efron, 1979; Efron and Tibshirani, 1993) with inferential purposes, will give more accuracy to parameter identification. Furthermore, it would be possible to obtain marginal intervals at a desired confidence.

Statistical Inference studies how to use the information of a parameter estimator for obtaining probabilistic results about the real and unknown parameter. The bootstrap method rises from the analogy between a population and a sample from it. In the bootstrap word the sample is the population.

In the present work, a steady state detection algorithm for multivariate process was used to generate a data base of steady states data set. Bootstrap was used for re-sampling the data set and to generate a parameter distribution in order to obtain a marginal interval for each variable. This approach will improve the common methodology of model (Haldane Kinetic Model) fit to average data and will give marginal intervals for identified parameters. All these methodologies are applied to an anaerobic digestion process for wastewater treatment, as an example, but it is extensible to any other kind of chemical and biochemical process.

## **2. Materials and Methods**

### **2.1 Experimental setup**

The anaerobic wastewater treatment pilot plant is composed of a hybrid UASB-UAF reactor of 1 m<sup>3</sup>. The measurement devices were: feed and recycling flow meters, pH meter; inflow and reactor Pt100, gas flow meter, infrared gas analyser (CH<sub>4</sub> and CO), gas hydrogen analyser and TOC/TIC combustion analyser. The sensors gave a signal every 5 seconds and every 15 minutes a moving average window was saved in the data base. Other parameters were calculated using the measured variables: methane flow rate (Q\_CH<sub>4</sub>), hydrogen flow rate (Q\_H<sub>2</sub>) and organic loading rate (OLR).

### **2.2 Experimental conditions**

The reactor was operated at stable conditions for more than a month at an OLR of 5 kg COD/m<sup>3</sup>·d. Three consecutive increases of the OLR were performed in order to obtain three different steady states (plus the initial steady state). Table 1 presents the different conditions for each organic load. The duration of each state was around 5 days, considered enough to achieve steady state because the HRT was around of 0.6 to 1.5 d. For all the calculations, the first period (day 0 to 4) was considered as a normal operation state.

### **2.3 Multivariate steady state detection algorithm**

The steady state detection algorithm used was previously presented by Ruiz et al (2004). Principal component analysis (PCA) was applied for reducing the N-dimension space of the multivariate data to two dimensions. By using PCA it is possible to retain the maximum variability (information) of the process with just a few new variables (principal components). Two PCs are usually enough and represent a high level of the total variability of the process. Steady state then is identified using Cao and Rhinehart

(1997) methodology. When both the first and second principal components (PC1 and PC2) are in steady state, the process will be considered at steady state.

Table 1. Experimental conditions applied to the reactor for the four states

| State      | Duration<br>(d) | OLR<br>(kg COD/m <sup>3</sup> ·d) | Feed flowrate<br>(Qinf) (L/h) | TOC influent<br>(mgC/L) |
|------------|-----------------|-----------------------------------|-------------------------------|-------------------------|
| 1: (N.O.)  | 0-4             | 5                                 | 22                            | 3000                    |
| 2: H.O.    | 4-9             | 15                                | 66                            | 3000                    |
| 3: H.O+O.O | 9-14            | 28                                | 66                            | 4500                    |
| 4: H.O+O.O | 14-15.5         | 32                                | 66                            | 6000                    |
| 5: (N.O)   | 15.5-17         | 5                                 | 22                            | 3000                    |

N.O.: Normal operation; H.O. Hydraulic overload; O.O. organic overload

## 2.4 Bootstrap

A parametric bootstrap method is used to estimate the distribution of the kinetic parameters estimators. Lets be  $y$  the methane flow rate,  $x$  the substrate concentration, and  $m_\theta$  the Haldane kinetic model for methane production. The model can be written as:

$$y = m_\theta(x) + \varepsilon \quad (1)$$

Where  $\varepsilon$  is the noise of the process. The N data are separated by groups and the mean of each group is calculated. Using the mean of each group and the minimum mean squared error criterium, the first estimation of the kinetic parameters is done using simplex algorithm. Each data  $y_i$  can be expressed as follows.

$$y_i = m_\theta(x_i) + \varepsilon'_i, 1 \leq i \leq N \quad (2)$$

The residual errors,  $\varepsilon'_i$ , will be used to generate the bootstrap samples. The bootstrap sample is obtained from the Haldane model with estimated parameters plus a residue with the same mean and variance of the original error. The expresion of the B bootstrap samples is

$$y_i^{*j} = m_\theta(x_i) + \varepsilon'_i \cdot Z_i^j, 1 \leq i \leq N, 1 \leq j \leq B \quad (3)$$

Being  $Z_i^j$  a random variable with zero mean and unity standard deviation. Each bootstrap sample, using the agrupation criterium described above, will generate a parameters estimation. By this way B parameters estimation will be obtained, and used to estimate a confidence region for the Haldane curve and for the marginal parameters values.

## 3. Results

### 3.1 Anaerobic digestion process

Figure 1 presents, as an example, some of the on line data obtained from the hybrid UASB-UF anaerobic reactor monitoring. The presented data are a four step increment in the organic loading rate (OLR), both by inflow and inlet concentration increase. The

reactor supported the OLR increases but in the last increase the process arose to a high destabilisation state. The process was followed by 26 variables.

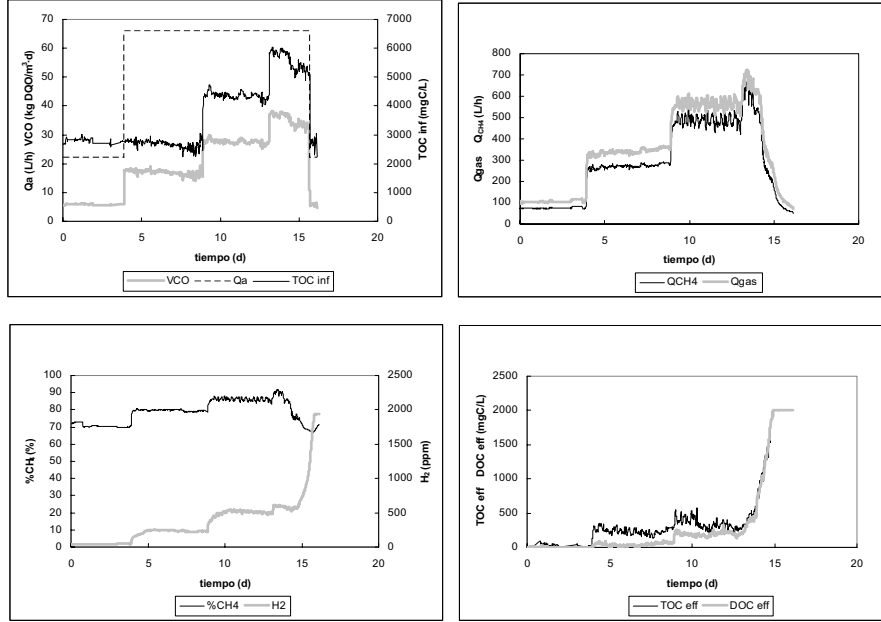


Figure 1: Multivariate on-line data for the four step organic loading rate increase of the hybrid UASB-UF anaerobic reactor up to a complete destabilization of the process. Nine variables are presented as an example.

### 3.2 Dimension reduction

The normal operation state was used for PCA model building. Previously data were normalized. As it can be seen in Figure 2, the use of two PCs is enough to explain the process since it retained almost 80% of the total variability of the process. Subsequently data were normalized and decomposed to PC1 and PC2. Figure 3-A presents the value of the PC1 and PC2 over time. It can be clearly seen the 4 states of the experiments. These data were used for steady state detection, with the Cao and Rhinehart proposed methodology.

### 3.3 Steady state detection using modified Cao and Rhinehart methodology

Once reduction of the process dimension was performed through PCA decomposition, Cao and Rhinehart (1997) multivariate methodology for steady state detection can be applied, because PC1 and PC2 are independent from each other. The used filter parameters were  $\lambda_1 = 0.2$  and  $\lambda_2 = \lambda_3 = 0.1$ . These values give a good balance between type I and type II error (Cao and Rhinehart, 1997).

Data were manually labelled by an expert as steady state and non steady state. Resampling the data, the R-critic was computed in order to obtain a confidence level of  $\alpha_{\text{process}}$  of 0.08 (92%), the R-critic was defined as 2.62 and 2.16, for the R-statistic for PC1 and PC2, respectively. Figure 3-B presents the classification results.

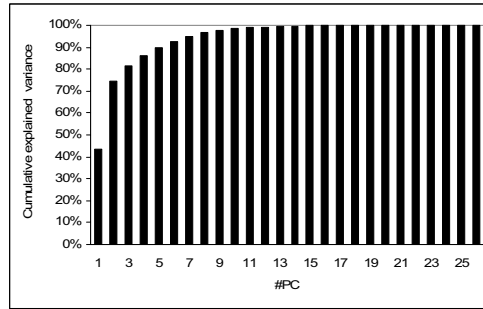


Figure 2: Cumulative explained variance by using more principal components (PC)

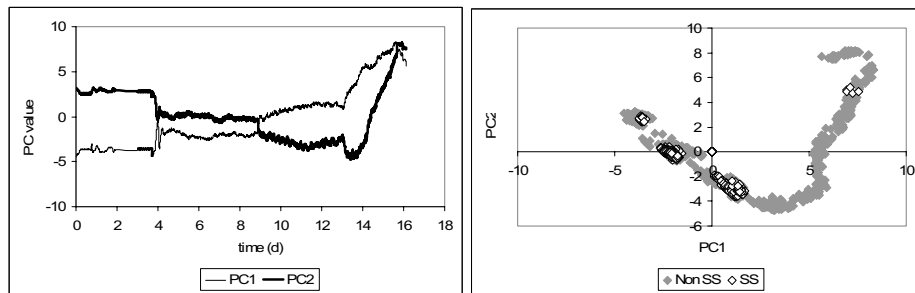


Figure 3: Steady states detected on process data. PC1 and PC2 over time (A) and PC1 against PC2 (B), steady state data marked on the graph.

### 3.4 Steady state data obtained

Figure 4-A presents, as an example, the steady state methane flow rate and the effluent DOC. These two variables can be fitted by a Haldane model, usually by obtain the mean of a group of data, as presented in figure 4-B. In this figure, the variance of the data is lost, loosing information and subtracting accuracy to the final identified parameters, even if the fit is quite good as can be seen in figure 4-B.

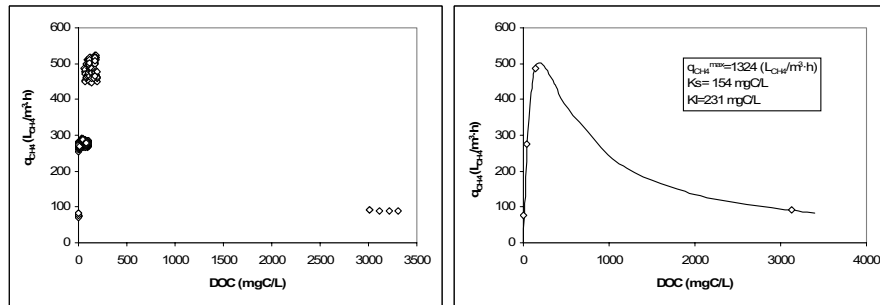


Figure 4 Steady state data detected on process data for methane flow rate and dissolved organic carbon. All data sets (A) and mean data with model fit.

### 3.5 Bootstrap confidence levels.

Figure 5 shows, in summary, the model fit with maximum boundaries computed by 5000 bootstrap parameters estimation (A) and the relationship between the three estimated parameters (B). As it can be seen there is a high correlation between them. The relation between  $Q_{CH_4max}$  and  $K_S$  seems to be linear ( $R^2=0.9998$ ), and a power

relationship between  $K_I$  and both,  $QCH_4\text{max}$  and  $K_S$ . Figures 6C,D and E present the histograms of the parameters distribution obtained by bootstrap.

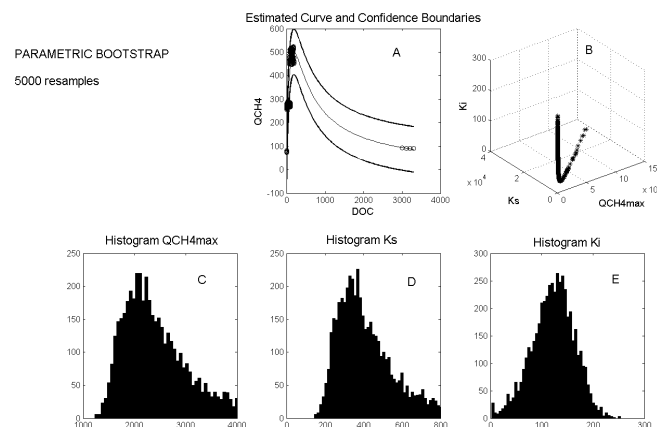


Figure 5 Haldane Estimated Curve, Confidence Boundaries and Parameter Estimation Histograms using Bootstrap

#### 4. Conclusions

A methodology for parameters estimation based on steady state data has been developed using Bootstrap in order to consider the variability of the original data instead of the data mean and an algorithm for multivariate steady state detection was applied. This methodology was applied to an anaerobic digestion waste water treatment process of pilot plant but their fundamentals are extensible to any other kind of processes. As an example methane production fitted to a Haldane model was considered. This methodology allows to obtain parameter estimations as well as marginal confidence boundaries for both, the Haldane curve and the parameters estimation.

#### References

- Batstone, D., 1999, High Rate Anaerobic Treatment of Complex Wastewater. PhD dissertation in Chemical Engineering at The University of Queensland
- Cao, S.L. and R.R. Rhinehart, 1995, Journal of Process Control 5 (6), 363-374.
- Cao, S.L. and R.R. Rhinehart, 1997, Journal of Process Control 7 (2), 149-152.
- Efron, B., 1979, The Annals of Statistics, 7, 1-26.
- Efron, B., and R.J. Tibshirani, 1993, An introduction to the bootstrap. New York: Chapman-Hall.
- Lokshina, L.Ya. and V.A. Vavilin, 1999, Ecological modeling. 117, 285-303.
- Merkel, W., A. Schwarz, S. Fritz, M. Reuss and K. Krauth, 1996, Wat. Sci. and Tech. 34 (5-6), 393-401.
- Ruiz, G., M. Castellano, W. González, E. Roca and J.M. Lema, 2004, Algorithm for steady state detection of multivariate process: Application to anaerobic wastewater digestion process. In: Proceedings of the 2nd International IWA Conference AutMoNet. Pp 181-188
- Szela, J. and R.R. Rhinehart, 2002, Journal of Process Analytical Chemistry 8 (2), 1-4.

#### Acknowledgements

This work has been carried out with the support provided by the MCyT (European FEDER support included) BFM 2002-032123 and ANACON Project (CICYT CTQ2004-07811-C02-01/PPQ).