Validation of a knowledge-based risk model for biological foaming in anaerobic digestion simulation

Dalmau[a] J., Comas[a] J., Rodríguez-Roda[a] I., Latrille[b] E. and Steyer[b] J.P.

[a] *Laboratory of Chemical and Environmental Engineering (LEQUIA), University of Girona, Faculty of Sciences, Campus Montilivi s/n, 17071 Girona, Catalonia, Spain. (E-mail: jordi@lequia.udg.cat; quim@lequia.udg.cat; ignasi@lequia.udg.cat)*
[b] *INRA, UR050, Laboratoire de Biotechnologie de l'Environnement, Avenue des Étangs, 11100 Narbonne, France. (E-mail: latrille@supagro.inra.fr; steyer@supagro.inra.fr)*

Abstract Anaerobic digestion (AD) is a complex biological system which can be affected by several operational problems. Among them, biological foaming is one of the most difficult to deal with. It has many effects, such as causing gas pipe clogging and probe failures, and it can affect mixing devices, etc. Since the mechanisms involved in biological foaming development are not fully understood, it is not included in standard anaerobic digestion models. For this reason, a knowledge-based risk model to determine the suitable conditions for the development of biological foaming during AD simulation was developed. The resulting knowledge-based system, based on organic loading rate and its daily variation, was experimentally validated using real data from a fully instrumented pilot plant (1 m^3 upflow fixed bed digester). Results show a good correlation between the knowledge-based risk model and the estimated biological foaming risk from real data.

Keywords: Anaerobic digestion, foaming, fuzzy logic, knowledge-based systems, validation**.**

1. INTRODUCTION

Activated sludge processes are complex biological systems in which organic matter and nutrients (nitrogen and phosphorous) are removed from wastewater. The system consists of an aeration tank where oxygen is selectively supplied and it is used by the microbial consortia (i.e. biomass and/or sludge) to grow and reproduce by consuming the substrate (i.e. pollutants) present in the wastewater.

The system also includes a secondary settler in which the treated water is separated from the biomass. From the bottom of the clarifier a fraction of the activated sludge is returned to the reactor in order to maintain the biomass constant in the reactor. To prevent overgrowth of the biomass in the system, a small fraction of the sludge is wasted from the system. This fraction represents a significant cost for the activated sludge process, since further treatment is required.

The most common alternative for sludge treatment is AD, as well as for wastewater with high contents of organic matter. In this process (Figure 2), the organic matter (e.g. sludge coming from activated sludge treatment) is biologically degraded in a digester in absence of oxygen. AD advantages are numerous since they provide a treatment for highly loaded wastewater, low sludge production and production of energy in form of methane. AD processes are, like in the activated sludge, very complex biological systems since a huge amount of microbial species are involved in the process.

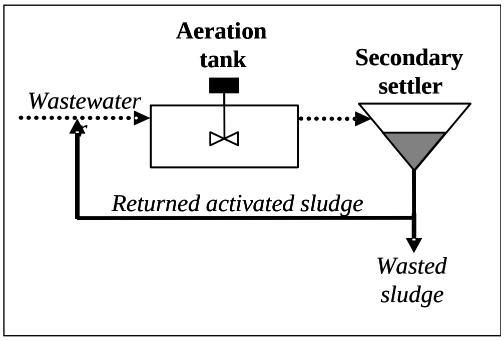

Figure 1. Activated Sludge System.

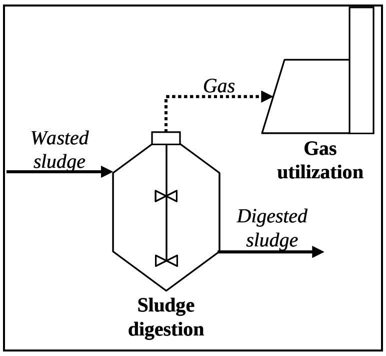

Figure 2. Anaerobic digestion System

Within this complexity some bacteria can have its own growth promoted by certain conditions which can cause imbalances in the digester in the form of a thick foam blanket. According to Pagilla *et al.* (1997), consequences of biological foaming are numerous:

- Blockage of gas mixing devices.
- Inversion of digester solids profiles.
- Foam binding of recirculation pumps.
- Fouling of gas collection pipes (due to entrapped foam solids).
- Foam penetration between floating covers and digester walls.
- Decrease of the digestion efficiency.

There is not yet a complete agreement on the parameters that favours conditions for foaming forming bacteria. Some authors state that a proper control of the feeding will prevent excessive foaming to appear (Massart *et al.*, 2006; Schaffer *et al.*, 2006). Others state that pre-treatment of the feeding is necessary to avoid foaming appearing (Barjenbruch and Kopplow 2003; Elliott and Mahmoot 2007). Besides, some claim that the presences of some filamentous bacteria (e.g. *M. Parvicella*, *Nocardia amarae* …) in the activated sludge system are the cause for foaming problems in the anaerobic digester (Pagilla *et al.*, 1997; Westlund *et al.*, 1998). Precisely all these uncertainty about the causes of biological foaming hinders the development of a mechanistic model to assess the biological foaming appearance.

Knowledge-based systems have proven to be appropriate tools to deal with complex processes like those involving microbiology-related problems in activated sludge systems (Comas *et al.*, 2003; Poch *et al.*, 2004). Specifically fuzzy logic has been successfully applied to a variety of systems. For instance, in Lardon *et al.* (2005) is applied to several AD operational imbalances and, in Carrasco *et al.* (2004) it is shown how a fuzzy system is able to control and diagnose acidification states in an anaerobic digester.

When building knowledge-based systems, the selection of input variables and the study of the data related to the problem under study is important in order to get a reliable system. For this reason, a previous variable selection was performed to a set of data from a pilot plant in order to find the most relevant input variables for the knowledge-based system developed. The knowledge gained with the variable selection together with the heuristic knowledge present in the literature led to the development of a knowledge-based AD risk model implemented in fuzzy logic to assess favorable conditions for biological foaming in simulation. The rationale behind this risk model was that the deterministic modelling of some WWTP simulation scenarios, although performing better regarding economic and environmental issues, can induce a higher risk of biological foaming.

The aim of this paper is to test the performance of the developed AD risk model with real data from a pilot plant. The paper is structured as follows; first the variable selection method is explained together with a brief summary of the AD risk model. Then the validation section illustrates and discusses the performance of the AD risk model validation with real data and, finally, some conclusions are drawn.

2. DEVELOPMENT OF THE AD RISK MODEL

To select the most relevant variables a wrapper approach with a hill-climbing elimination strategy (Kohavi and John, 1997) was used. The same methodology was used in Dalmau *et al.* (2007) in order to find the most relevant variables for acidogenic states in anaerobic digestion. Afterwards in Dalmau *et al.* (2008), the same approach was applied to biological foaming in AD which is, as commented above, a more challenging issue.

A home-made neural network toolbox for static models for use in MATLAB 5.3 or higher was used. Two layers were chosen in all the ANN architectures: a hidden layer of neurons with sigmoid transfer functions and an output layer with linear transfer functions for outputs. The initialization method was performed using the Nguyen-Widrow algorithm option, which initializes the weights with random values, later selecting their probability distributions to make all neurons active for the expected data ranges (Nguyen and Widrow, 1990). It also provides automatic data scaling and weights conversion. Bayesian regularisation is used to prevent over-fitting.

Figure 3 depicts the methodology used that starts with the ten times training of the reference ANN with all the variables. Its average Root Square Mean Error (RSME) is calculated and stored as the reference error. Next, one input variable is removed and a new ANN (ANN1 in figure 3) is trained ten times without it. This last step is repeated for each input variable ending up with *n* ANNs 1, one for each removed input variable with their related average RSME 1. Whenever a relevant variable is removed, the average RSME 1 of the related ANN 1 will increase with respect to the average reference error. On the other hand, whenever a non-relevant variable is removed the RSME 1 of the related ANN1 will decrease. Therefore, the variables which RSME 1 is higher than the reference error are selected as relevant variables.

Among relevant variables the one with the higher RSME 1 is selected first and a new ANN (ANN 2 this time) is trained ten times again using it as the only input. If the related average RSME (RSME 2) is higher than the average reference error no improvement is found, so the variable with the second higher average RSME 1 is selected and a new ANN 2 is trained (ten times as well) with both variables, and again, its average RSME 2 is compared with the reference. This iterative process is repeated until an average RSME 2 lower than the average reference RSME is obtained.

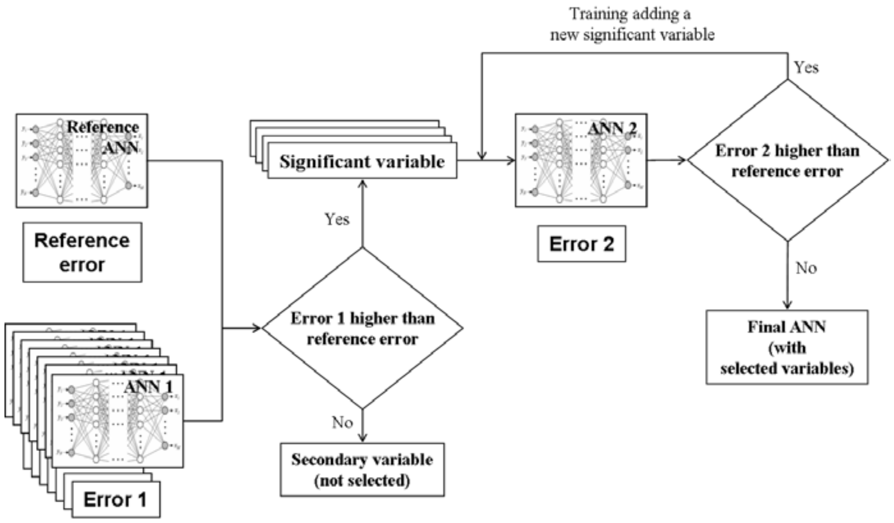

Figure 3. Methodology schema, based on Dalmau *et al.* (2007).

Experimental data used were obtained from a pilot plant from LBE of the INRA, France. Overall, a set of 8133 data was used for the variable selection. Among all variables a first selection was done based on the common variables which are available in real plants. Some others were not selected for instance, temperatures since it is usually constant so it will be difficult to extract information from its profile. Input variables involved in this study were: inflow rate and pH in the influent flow rate; volatile fatty acids concentration, total organic carbon and pH in the digester and, carbon dioxide and methane percentage in the gas phase. As output, biological foaming appearance (foaming index) in the digester was used, based on the heuristic knowledge provided by the experts. It was noticed that when foaming appeared in the digester high variations of the gas flow rate and pressure coincided due to the slug release of gas bubbles trapped inside the foam. It is important to point that even though foaming can be estimated this way; this is an approach to study variables influence or relation. This approach it cannot be used in simulation because biological foaming is not currently modelled so the results of the simulation cannot reflect its effects on the gas flowrate and pressure variations.

Eventually, as shown in figure 4, the variables with RSME higher than the reference error (i.e. relevant variables) were: total organic carbon in the digester, the carbon dioxide and methane percentage in the gas phase and, the inflow rate and the pH in the inflow rate. The relevance of gas-related variables (i.e. carbon dioxide and methane percentage) can be due to the approach taken to determine foaming. According to Zhao and Viraraghavan (2004) high carbon dioxide production is representative of poor digestion that may lead to foaming, but in general is representative of general process imbalance but no related to a specific cause. So, taking a look to the other variables, precisely total organic carbon in the digester and inflow rate, the results can be related to some statements present in the literature. In Massart *et al.* (2006) it is stated that inconsistent feeding in the digester is one of the causes for foaming. Feeding is related to Organic Loading Rate (OLR), which is related to the inflow rate and the amount of sludge feed to the digester (Metcalf and Eddy, 2003) related at the same time to the organic matter present in the digester (total organic carbon in the digester).

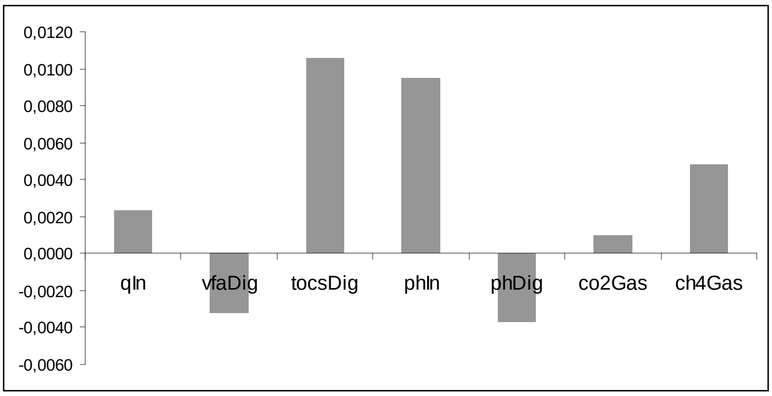

Figure 4. Difference between the RMSE and Reference error for each variable. From Dalmau *et al.* (2008).

3. AD RISK MODEL

To develop the risk model the relevant variables selected previously from real data were compared with the knowledge present in the experiences from the bibliography. As seen in the previous section, some coincidences were found. Finally, the combination of OLR and its variation were selected as inputs of the model. As a problem of biological origin, the presence of some filamentous bacteria (mainly *M. Parvicella*) in the anaerobic digester's inflow rate is also relevant regarding biological foaming so it was also taken into account in the AD Risk Model. This input is obtained from the risk model developed by Comas *et al.* (2008). This model used heuristic knowledge to evaluate simulation results and look for suitable conditions for the development of microbiology-related settling problems (i.e. bulking, foaming and rising sludge) in the AS system. More specifically, the AD risk model uses as input the risk of foaming related to *M. parvicella* which cause foaming in the AS (FAS risk) system as well. The basic knowledge base is presented in table 3.1. For a low FAS risk, as OLR and its variation (OLRvar) increase, the risk of foaming increases as well. Since the pilot plant treated diluted industrial distillery wastewater and was not sludge from an activated sludge system, the FAS risk was considered to be low in the validation step. Thus, the knowledge base for higher FAS risks is not presented here. However, further details on the AD Risk Model will appear in Dalmau *et al.* (2009).

Table 3.1. Knowledge base of the AD Risk Model for low FAS risk.

		OLR (kg VS·m^{-3}·d-1)				
		Very Low	**Low**	**Medium**	**High**	**Very High**
OLR var (%)	**Low**	Low	Low	Medium	Medium	High
	Medium	Low	Medium	Medium	High	High
	High	Medium	Medium	High	High	High

The AD risk model is used to assess the risk of biological foaming in AD simulation. As an example of the AD risk model performance, figure 5 shows a profile of the risk of biological foaming within the benchmark simulation model Nº2 (BSM2; Jeppsson *et al.*, 2007), however, it is out of the scope of this paper to discuss its performance. FAD risk stands for risk of biological foaming in AD. The x-axis contains one-year simulation time from July 1st. The seasonal effect of FAS (according to Hug *et al.*, 2006) can be noticed in the profile influencing the FAD risk. Although OLR is oscillating it remains in a constant range, however when OLRvar decreases (end of first summer period and middle winter) it is reflected in the FAD risk. Despite the AD risk model was developed and it can represent the dynamics of biological foaming in a simulated anaerobic digester, it was not validated yet with real data.

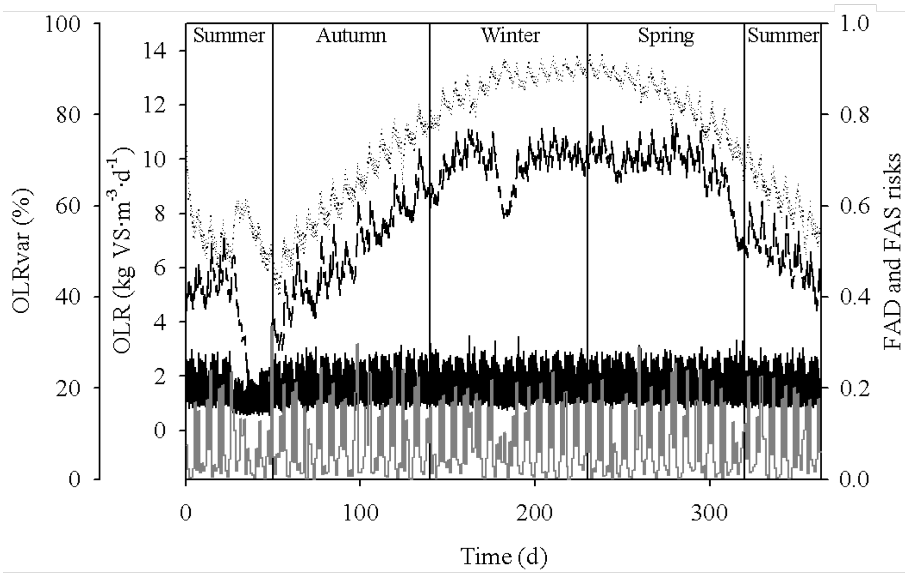

Figure 5. Simulated results of the FAD risk model for the open loop case for a one-year simulation (from July 1st). OLR (solid line); OLRvar (grey line); FAS risk (dotted black line); FAD risk (dashed line).

4. VALIDATION OF THE AD RISK MODEL

The AD risk model was developed to be implemented to BSM2, so previous to its validation with real data it was necessary to make some assumptions.

OLR and its daily variation have to be calculated from the variables measured in the pilot plant. OLR for the AD risk model is calculated as shown in Eq. 1.

$$OLR = \frac{VS}{HRT} \qquad \text{Eq.1}$$

where,
VS= Volatile Solids kg·L^{-1}
HRT= Hydraulic Retention Time (d)

HRT in days is obtained from Eq. 2

$$HRT = \frac{V}{qIn \cdot 24} \qquad \text{Eq. 2}$$

where,
V= pilot plant volume (1000 L.)
qIn= inflow rate in (L·h^{-1})

Since AD risk model OLR calculation is based on VS, it was necessary to transform the measured COD (tocsDig) into VS. According to Metcalf and Eddy (2003) for untreated wastewater the biological oxygen demand/total organic carbon ratio (BOD/TOC) is between 1.2 and 2.0 (1.6 was taken as the average of the rank), there is also a relation between BOD and chemical oxygen demand (COD) from 0.3 to 0.8 (0.55 was taken as the average of the rank). Therefore, putting together both ratios, COD can be expressed as a function of TOC (tocsDig in our case; Eq. 3).

$$COD = 2.9 * tocsDig \qquad \text{Eq. 3}$$

where,
tocsDig: total organic carbon in the digester ($mg \cdot L^{-1}$)
COD in mg COD $\cdot$ L^{-1}.

In Copp (2002) it is pointed that there is a relation between total suspended solids (TSS) and COD from particulate compounds (Eq. 4).

$$TSS = 0.75 \cdot COD_p \qquad \text{Eq. 4}$$

where,
COD_p in mg COD $\cdot$ L^{-1}.

A last assumption is made in order to simplify the conversion supposing that all the TSS can be accounted as VS. This way, we consider all the COD measured in the pilot plant can be degraded as it was VS. Thus, from Eq 3. and Eq 4. we get Eq. 5.

$$VS = \frac{2.175 \cdot tocsDig}{1000} \qquad \text{Eq. 5}$$

where,
VS in $kg \cdot L^{-1}$

Data gathered during approximately almost three months was used to validate the AD risk model. Figure 6 shows the profile for both the simulated biological foaming risk (SFR) from the AD risk model and the foaming index estimated from real data (FR).

From figure 6 some aspects can be pointed. First of all, reasonable good fitting is achieved (RMSE=0.06). Secondly, it becomes clear that there are two differentiated periods, approximately the first month and the last two months. The first period is marked for an apparent stability of the process with a good coincidence between SFR and FR (both showing low foaming risk), whereas the second shows much more oscillations and peaks revealing a probably more unstable period. In this last period, in some specific points (i.e. around days 33, 43 and 58) there are some divergences in which the model shows relatively high foaming risk when the real data show low risk of foaming. It is important to note that the inherent uncertainty of the mechanisms of foaming that hinders the development of mechanistic models cannot be included in the AD Risk Model. This can be the main reason behind the main differences in the validation results. Nevertheless, the general trends of the instability are indeed detected by the AD risk model allowing it to assess operational conditions of the anaerobic digester that can favour biological foaming.

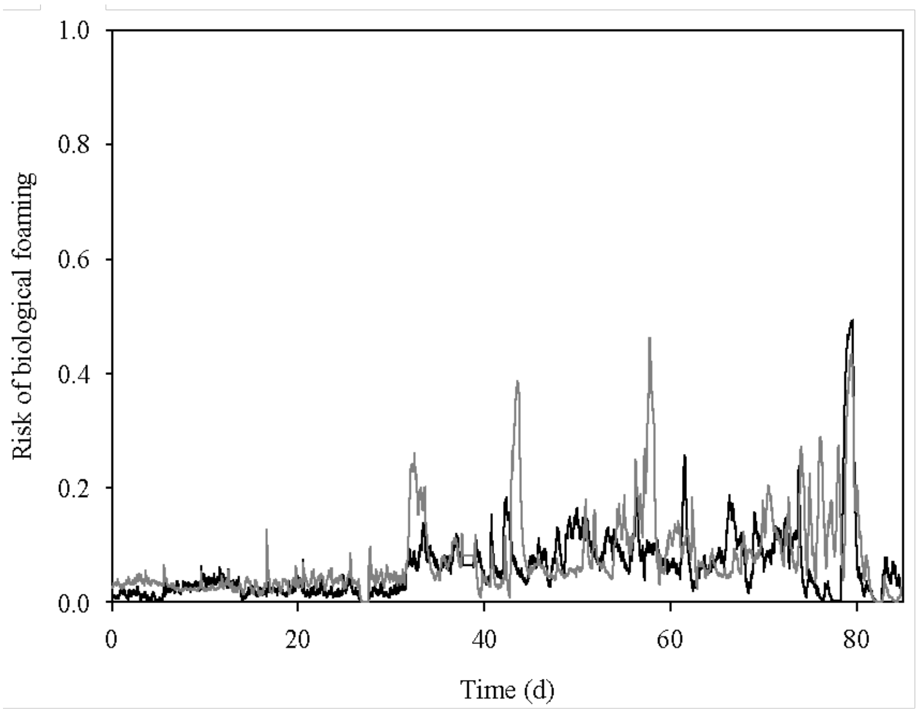

Figure 6. SFR (black line) versus FR (grey line).

5. CONCLUSIONS

The AD risk model has been validated using real data from a pilot plant. Real data has been adapted to the AD risk model and the results show that quite a good fitting of the data can be achieved, showing that the AD risk model is able to represent the general conditions of an anaerobic digester regarding biological foaming. However, further validation with real data from an anaerobic digester treating sludge from an activated sludge system would be of interest since it would allow to consider the effect of filamentous bacteria in the AD feed.

ACKNOWLEDGEMENTS

Authors wish to thank Dr. Cristian Trelea from the Institut National Agronomique Paris-Grignon for developing and providing us with the Neural Network toolbox. The work was also partially supported by the Spanish Ministry of Education and Sciences project DPI2006-15707-C02-01.

REFERENCES

Barjenbruch M and Kopplow O (2003) Enzymatic, mechanical and thermal pretreatment of surplus sludge. Adv Environ Res 7:715-720.

Carrasco E F, Rodríguez J, Puñal A, Roca E and Lema J M (2004) Diagnosis of acidification states in an anaerobic wastewater treatment plant using a fuzzy-based expert system. Control Eng Pract, 12:59-64.

Comas J, Rodríguez-Roda I, Poch M, Gernaey K V, Rosen C and Jeppsson U (2008) Risk assessment modelling of microbiology-related solids separation problems in activated sludge systems. Environ Modell Softw 23(10-11):1250-1261.

Comas J, Rodríguez-Roda I, Sánchez-Marrè M, Cortés U, Freixó A, Arráez J and Poch M (2003) A knowledge-based approach to the deflocculation problem: integrating on-line, off-line, and heuristic information. Wat Res, 37(10):2377-2387.

Copp J B (2002) The COST Simulation Benchmark – Description and Simulator Manual. Office for Official Publications of the European Communities, Luxembourg, ISBN 92-894-1658-0.

Dalmau J, Comas J, Rodríguez-Roda I, Pagilla K and Steyer J P (2009) Risk model development and simulation for foaming in anaerobic digestion. Wat Res, (Submitted).

Dalmau J, Comas J, Rodríguez-Roda I, Latrille E and Steyer J P (2008) A neural network approach to select the most relevant variables for foaming in anaerobic digestion. Proceedings of the International Congress on Environmental Modelling and Software. Barcelona, Spain, 7-10 July. ISBN: 978-84-7653-074-0

Dalmau J, Comas J, Rodríguez-Roda I, Latrille E and Steyer J P (2007) Using artificial neural networks for selecting relevant information in anaerobic digestion, Paper presented at 11th IWA Specialitst Conference on Anaerobic Digestion (AD11), Brisbane, Australia, September 23-27.

Elliot A and Mahmood T (2007) Pretreatment technologies for advancing anaerobic digestion of pulp and paper biotreatment residues. Wat Res, 41:4273-4286.

Hug T, Gujer W and Siegrist H (2006) Modelling seasonal dynamics of "*Microthrix parvicella*". Water Sci Technol, 54(1):189-198.

Jeppsson U, Pons M-N, Nopens I, Alex J, Copp J, Gernaey K V, Rosen C, Steyer JP and Vanrolleghem P A (2007) Benchmark simulation model no 2: general protocol and exploratory case studies. Water Sci Technol 56(8):67-78.

Kohavi R, John G H (1997) Wrappers for feature subset selection. Art Intell 97:273-324.

Lardon L, Puñal A, Martínez J A, Steyer J P (2005) Modular expert system for the diagnosis of operating conditions of industrial anaerobic digestion plants. Water Sci Technol 52(1-2):427-433.

Massart N, Bates R, Corning B and Neun G (2006) Design and operational considerations to avoid excessive anaerobic digester foaming. 79th Water Environmental Federation Technical Exhibition and Conference.

Metcalf and Eddy (2003) Wastewater engineering: Treatment and reuse. Fourth ed., McGraw-Hill, New York, USA.

Nguyen D; Widrow B (1990) Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. Proceedings of Int. Joint Conf. Neural Networks. Washington DC, USA, January 15-19, 21-26.

Pagilla K R, Kraney K C, Kido W H (1997) Causes and effects of foaming in anaerobic sludge digesters. Water Sci Technol, 36(6-7):463-470.

Poch M, Comas J, Rodríguez-Roda I, Sánchez-Marrè M, Cortés U (2004) Designing and building real environmental decision support systems. Environ Modell Softw 19(9):857-873.

Schafer P, Uhte W, Newman G (2006) Improved anaerobic digestion feed systems and concepts. Proceedings of the 80th Annual Water Environment Federation Technical Exhibition and Conference; Dallas, Texas, USA, Oct 21-25; Water Environment Federation: Alexandria, Virginia, 374-392.

Westlund A D, Hagland E and Rothman M (1998) Foaming in anaerobic digesters caused by *Microthrix parvicella.* Water Sci Technol 37(4-5), 51-55.

Zhao H W and Viraraghavan T (2004) Analysis of the performance of an anaerobic digestion system at the Regina wastewater treatment plant. Biores Technol, 95:301-307.