Jul 1st, 12:00 AM

A neural network approach for selecting the most relevant variables for foaming in anaerobic digestion

Jordi Dalmau

Joaquim Comas

Ignasi Rodríguez-Roda

Eric Latrille

Jean-Philippe Steyer

A neural network approach for selecting the most relevant variables for foaming in anaerobic digestion

Jordi Dalmau[a], Joaquim Comas[a], Ignasi Rodríguez-Roda[a], Eric Latrille[b] and Jean-Philippe Steyer[b]

[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: Activated sludge processes are complex biological systems in which organic matter and nutrients (nitrogen and phosphorous) are removed from wastewater. One of the most common alternatives for the treatment of waste from activated sludge systems is Anaerobic Digestion (AD). AD is even a more complex biological process. One of the most important problems which can appear in anaerobic digesters is foaming. In the literature there is not a big agreement on the foaming causes. Therefore, the aim of the paper is to apply a methodology based on artificial neural networks to determine the most relevant variables for foaming diagnosis in anaerobic digestion. To perform the study real data from a pilot plant located in the LBE in Narbonne were used. Results show inflow rate, total organic carbon and carbon dioxide percentage among the relevant variables which are, according to the literature, some of the factors which influence the presence of foaming. This methodology can be valuable when selecting probes to monitor AD processes or to select the most relevant variables to diagnose the state of the process when foaming is involved.

Keywords: Anaerobic digestion; artificial neural networks; data mining; diagnosis; foaming.

1. INTRODUCTION

1.1 Activated sludge systems

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.
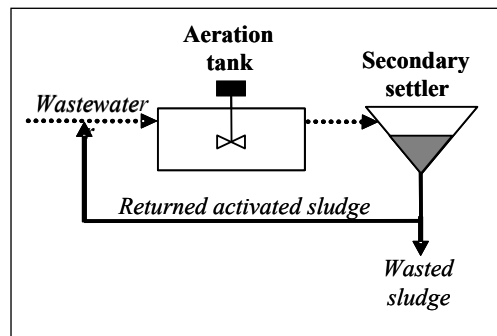
Figure 1. Activated Sludge System.

1.2 Anaerobic Digestion (AD)

The most common alternative for sludge treatment is AD. In this process the organic matter (e.g. sludge coming from activated sludge treatment) is biologically degraded in a digester in absence of oxygen. AD processes advantages are numerous since they provide a treatment for high loaded influents, 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 species are involved in the process.

1.3 Foaming

Within this complexity some bacteria can see its own growth promoted by certain conditions which can cause imbalances in the digester in the form of a thick foam blanket. This foaming can have many effects as: fouling of gas collection pipes, probes failure, problems with mixing devices, lose of digester efficiency... (Pagilla et al. [1997])

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 [2003]; Elliott [2007]). Besides, some claim that the presences of some filamentous bacteria in the activated sludge system are the cause for foaming problems in the anaerobic digester (Pagilla et al. [1997]; Westlund et al. [1998]).

1.4 Artificial Neural Networks (ANN)

When dealing with complex processes which their dynamics are not well known, ANNs are a useful tool to obtain an approximated model. Hence, in the present work they will be used to select the relevant variables and, at the same time, to finally obtain a black-box model for foaming diagnosis based on the most relevant variables. When dimensionality reduction is involved other techniques such as principal component analysis are easy to apply and they do not require a target data set as ANN does, nevertheless they run the risk of being suboptimal since they take no account of the target data (Bishop, [1994]).

The objective of this paper is to select the most relevant variables to control or prevent foaming in AD processes. Likewise, this work will highlight the variables most related to foaming in anaerobic digesters. Firstly, the used methodology will be presented followed by the results presenting the most relevant variables with some discussion to finish with conclusions.

2. METHODOLOGY

The methodology is a wrapper approach with a hill-climbing elimination strategy (Kohavi and John, [1997]). The same methodology was used in Dalmau et al. [2007] in which was applied in order to find the most relevant variables for acidogenic states in anaerobic digestion. In the current work, the same approach is applied to foaming in AD which is, as commented above, a more changeling issue since there is no full agreement on its causes.

Figure 2 depicts the procedure 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 1) 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.
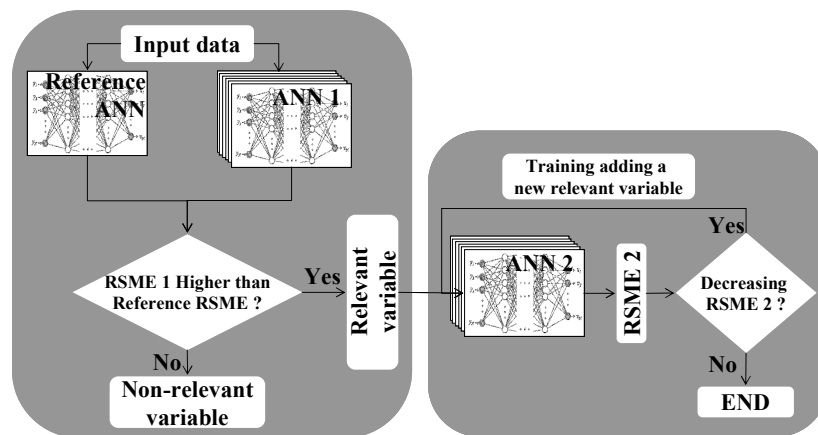

Figure 2. Methodology schema, based on Dalmau et al. [2007].

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.

2.1 Artificial Neural Network toolbox

A home made neural network toolbox for static models for use in MATLAB 5.3 or higher was used in this study. In all the ANN architectures two layers were used: a hidden layer of neurones with sigmoid transfer functions and an output layer with linear transfer functions for outputs. The initialisation method was performed using the Nguyen-Widrow algorithm option (Nguyen [1990]) was selected as initialisation method. This algorithm initialises the weights with random values, later selecting their probability distributions to make all neurones active for the expected data ranges. Automatic data scaling and weights conversion are included as well. Bayesian regularisation approach is used to prevent over-

fitting. Finally, training process is based on a Levenberg-Marquardt algorithm (Marquardt [1963] and Levenberg [1944]).

3. RESULTS

3.1 Data

Experimental data used were obtained from a pilot plant from LBE of the INRA, France. 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. Table 1 presents the input variables involved in this study. As output, foaming appearance 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. In order to have a suitable foaming index between 0 and 1 a fuzzy system was used. It is important to point that even though foaming can be determined likewise, this is an approach to study variables influence or relation.

Table 1. Input variables.

Inflow rate	Volatile fatty acids concentration in digester	Total organic carbon in digester	pH in the inflow rate	pH in the digester	Carbon dioxide percentage in the gas phase	Methane percentage in the gas phase
qIn	vfaDig	tocsDig	phIn	phDig	co2Gas	ch4Gas

Besides, the objective of the study is to determine the most relevant variables with the purpose of either saving in instrumentation or know which variables are most important to control to prevent foaming in the anaerobic digester.

3.2 Reference ANN and ANNs 1

The average RMSE of the reference ANN and the seven ANNs 1 are presented in Table 2.

Table 2. Reference ANN and ANNs1 average RSME

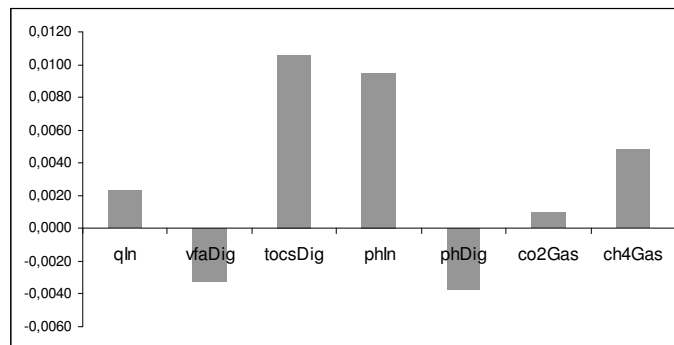| Reference	qIn	vfaDig	tocsDig	phIn	phDig	co2Gas	ch4Gas
0.11700	0.11934	0.11377	0.12764	0.12654	0.11328	0.11796	0.12179

Figure 3. Average RSME minus average reference RSME for each variable.

Figure 3 shows the differences between RSME of each variable and the reference RSME. As depicted 5 variables have RSME higher than the reference. Therefore, the selected as relevant variables were qIn; tocsDig; phIn; co2Gas and ch4Gas.

3.3 ANNs 2

To train ANNs 2 first tocsDig was selected. Being its average RSME 2 higher than the average reference RSME the second most relevant variable was added and ANN 2 was trained again. Table 3 summarizes the average RSME 2 for each ANN.

Table 3. Average RSME 2 for the relevant variables.

	tocsDig	tocsDig; phIn	tocsDig; phIn;ch4Gas	tocsDig;phIn; ch4Gas;qIn	tocsDig;phIn; ch4Gas;qIn; co2Gas
RSME 2	0.12666	0.12256	0.12141	0.12138	0.11944

In all cases RSME 2 were higher than the reference error. However, a test t performed comparing the averages between each RSME 2 and the RSME of the reference ANN revealed that in the last ANN 2 (last column in Table 3), the average RSME 2 was not significantly different than the reference RSME. This means that even though the average RSME 2 is not lower than the reference, by removing two variables from the whole input, similar results can be obtained. This can be important in terms of saving sensors and in terms of monitoring relevance.

The relevance of gas-related variables (i.e. co2Gas and ch4Gas) can be due to the approach taken to determine foaming (i.e. the fuzzy system used). So, taking a look to the other variables, precisely tocsDig and qIn, 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 (qIn) and the amount of sludge feed to the digester (Metcalf [2003]) related at the same time to the organic matter present in the digester (tocsDig). Finally, as shown, addition of co2Gas to the relevant variables reduces RSME 2. According to Zhao [2004] high CO_2 production is representative of poor digestion and may lead to foaming.

4. CONCLUSION

The presented methodology has been helpful to select the relevant variables related to foaming in anaerobic digesters. From seven variables two have shown to be non-relevant associated to foaming. This is indeed an indication of the complexity of this microbiology related problem and it shows that a lot of variables are more or less related to it. Nevertheless, when diagnosing or controlling the conditions to prevent foaming the selected variables are helpful indicators of what it is worth to control to avoid excessive foaming.

REFERENCES

Barjenbruch M. and O. Kopplow, Enzymatic, mechanical and thermal pre-treatment of surplus sludge, *Advances in Environmental Research, 7,* 715-720, 2003.

Bishop C. M., Neural networks and their applications, *Review of Scientific Instruments 65* (6), 1803-1832, 1994.

Dalmau J., J. Comas, I. Rodríguez-Roda, E. Latrille and J.P. Seteyer, 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, 2007.

Elliot A. and T. Mahmood, Pretreatment technologies for advancing anaerobic digestion of pulp and paper biotreatment residues. *Water Research 41*, 4273-4286, 2007.

Kohavi R. and G. H. John, Wrappers for feature subset selection, *Artificial Intelligence 97*, 273-324, 1997.

Levenberg K., A method for the solution of certain non-linear problems in least squares, *Quarterly Journal of Applied Mathematics 2*(2), 164–168, 1944.

Marquardt D.W., An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society of Industrial and Applied Mathematics 11*(2), 431–441, 1963.

Massart N., R. Bates, B. Corning and G. Neun, Design and operational considerations to avoid excessive anaerobic digester foaming, Paper presented at the Annual Water Environment Federation Technical Exhibition and Conference, Dallas, USA, October 21-25, 2006.

Metcalf and Eddy, *Wastewater engineering: Treatment and reuse*, McGraw-Hill, New York, USA, 2003.

Nguyen D. and B. Widrow, Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights, Paper presented at International Joint Conference on Neural Networks, San Diego, California, USA, June 17-21, 3, 21-26, 1990.
Pagilla K. R., K. C. Craney and W. H. Kido, Causes and effects of foaming in anaerobic sludge digesters, *Water Science and Technology*, 36(6-7), 463-470, 1997.

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

Schafer P., W. Uhte and G. Newman, Improved anaerobic digestion feed systems and concepts, *Proceedings of the Annual Water Environment Federation Technical Exhibition and Conference,* Dallas, USA, 2006.

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

Zhao H. and T. Viraraghavan, Analisys of the performance of an anaerobic digestion system at the Regina wastewater treatment plant, *Bioresource Technology, 95,* 301-307, 2004.