ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Data-driven modelling for filtration process in AnMBR technology

V. Sandoval-García, M.V. Ruano, A. Robles

CALAGUA Unitat Mixta UV-UPV, Departament d'Enginyeria Química, ETSE-UV, Universitat de València, Avinguda de la Universitats/n, 46100 Burjassot, Valencia, Spain. (e-mail: valeria.sandoval@uv.es, m.victoria.ruano@uv.es, angel.robles@uv.es)

**Abstract:** Membrane fouling is a major concern in anaerobic membrane bioreactors (AnMBRs). This study evaluates four machine learning models to predict fouling in an AnMBR system, which are based on fuzzy logic (FL), artificial neural network (ANN), random forest (RF), and partial least squares (PLS) techniques. The models were calibrated and validated using data from a semi-industrial AnMBR treating real wastewater. All models, particularly FL and ANN, performed with high average accuracy (r >0.96). The results highlight the ability of the models to be used for monitoring and control purposes aimed to optimise and improve the performance of AnMBR under realistic operating conditions.

**Keywords**: data-driven; fuzzy logic; artificial neural network; random forest; partial least squares.

## 1. INTRODUCTION

Further investigation of submerged anaerobic membrane bioreactor (AnMBR) technology is justified to improve understanding and optimising efficiency, particularly regarding membrane fouling. Mathematical modelling of filtration in AnMBR technology can provide relevant information on factors influencing membrane fouling (Mannina et al., 2011), assisting in the design, monitoring, control and optimisation of this technology (Ng and Kim, 2007). Models that focus on filtration process often rely on parameters that are difficult to measure on-line and require conventional laboratory equipment. Furthermore, some models are unable to reproduce the influence of operating stages such as relaxation or backwashing, which are common during membrane operation. However, due to the highly non-linear relationships between process inputs and outputs, empirical models normally only provide accurate results if the process dynamics are constrained within a defined linear zone. In this context, data-driven methodologies, such as fuzzy logic or neural networks, could serve as powerful tools for modelling filtration in AnMBRs. These approaches allow the incorporation of expert knowledge and facilitate the development of multiple-input-multiple-output process schemes, which are of special interest for optimising the performance of systems of this type.

The objective of this study was to assess and compare the effectiveness of four data-driven approaches for modelling the filtration process in AnMBR systems based on fuzzy logic (FL), artificial neural network (ANN), random forest (RF), and partial least squares (PLS) techniques. The study evaluated their performance not only in the short term but also in the long term for different purposes: development of soft-sensors, diagnosis tools, control systems, process optimisation, etc. Given that the impact of the main operating conditions on membrane fouling cannot be adequately assessed at the laboratory scale, the proposed models were validated using data from an AnMBR system equipped with industrial scale hollow-fibre (HF) membrane units. This AnMBR plant was operated with wastewater coming from the pre-treatment step of the "Conca del Carraixet" WWTP (Valencia, Spain). Therefore, the main disturbances occurring in full-scale plants were replicated.

## 2. MATERIALS AND METHODS

### 2.1 Semi-industrial AnMBR plant description

This study was carried out in a semi-industrial AnMBR plant fed with effluent from a full-scale WWTP pre-treatment. The AnMBR mainly consisted of an anaerobic reactor with a total volume of 1.3 $m^3$ connected to two membrane tanks each with a total volume of 0.8 $m^3$. Each membrane

tank contained a commercial ultrafiltration HF membrane system (PURON®, Koch Membrane Systems, 0.05 µm pore size, 30 $m^2$ total filtration area, and outside-in filtration). The membrane system was operated according to a combination of different stages (backwashing, ventilation, and degasification) taken from a basic filtration-relaxation cycle. Further details of this AnMBR system can be found in Robles *et al*. (2013).

### 2.2 Filtration process modelling

In this study, four data-driven approaches are proposed for modelling filtration in AnMBRs: (1) FL, (2) ANN, (3) RF, and (4) PLS. For the four models, the mixed liquor total solids (MLTS) concentration and the biogas recycling flow per bulk volume in the membrane tank (BRFv) were selected as input variables. Transmembrane flux (J) was also used as a model input variable for ANN, RF, and PLS models. The main characteristics of the models are shown in Table 1.

**Table 1** Characteristics of data-driven models

| Algorithms | Characteristics | Specification |
|---|---|---|
| **ANN** | Input layer neurons | 3 |
| | Output layer neurons | 1 |
| | Hidden layers | 1 |
| | Hidden layers neurons | 3 |
| | Activation function | Sigmoid function |
| **RF** | Number of features | 3 |
| | Number of trees | 100 |
| | Maximum depth | None |
| **PLS** | Output variable | Logarithmic transformation |
| **FL** | Linguistic labels | Low, Medium, and High |
| | Type of membership functions | Triangular- Medium label<br>Trapezoidal- Low and High labels |
| | Number of inference rules | 9 |
| | Defuzzification method | Sugeno's method |

For calibration purposes, the results from the application of a flux-step trials were used (Robles *et al*., 2013). This flux-step trials were conducted at different MLTS (from 18.5 to 28.5 g $L^{-1}$), BRF (from 5 to 10 $m^3$ $h^{-1}$), and J (from 9 to 16 LMH). On the other hand, short-term and long-term operating data were used for validation purposes.

## 3. RESULTS AND DISCUSSION

Figure 1 illustrates the calibration results for the four models evaluated, showing the comparison of predicted and experimental fouling rate data. The results demonstrate the great potential of the data-driven models to describe the filtration process in AnMBR systems. All the models have a Pearson product-moment correlation coefficient (r) value above 0.96. This is a clear indication that data-driven models can easily capture process dynamics from experimental observations in MBR systems. The main differences are observed in the prediction of boundary values. FL and ANN stand out as the most accurate models in predicting the fouling rate.

For instance, Figure 2 shows the validation results for the fuzzy logic model at a MLTS concentration of 21 g $L^{-1}$, BRF from around 5 to 10 $m^3$ $h^{-1}$, and a gross J ranging from 4 to 12 LMH. The correlation between the model prediction and the experimental data resulted in coefficient r values of around 0.7. This result supports a good performance of the fuzzy logic-

based model to reproduce TMP dynamics. Nonetheless, further evaluation of this model is still needed to validate its effectiveness across a wide range of operating conditions.

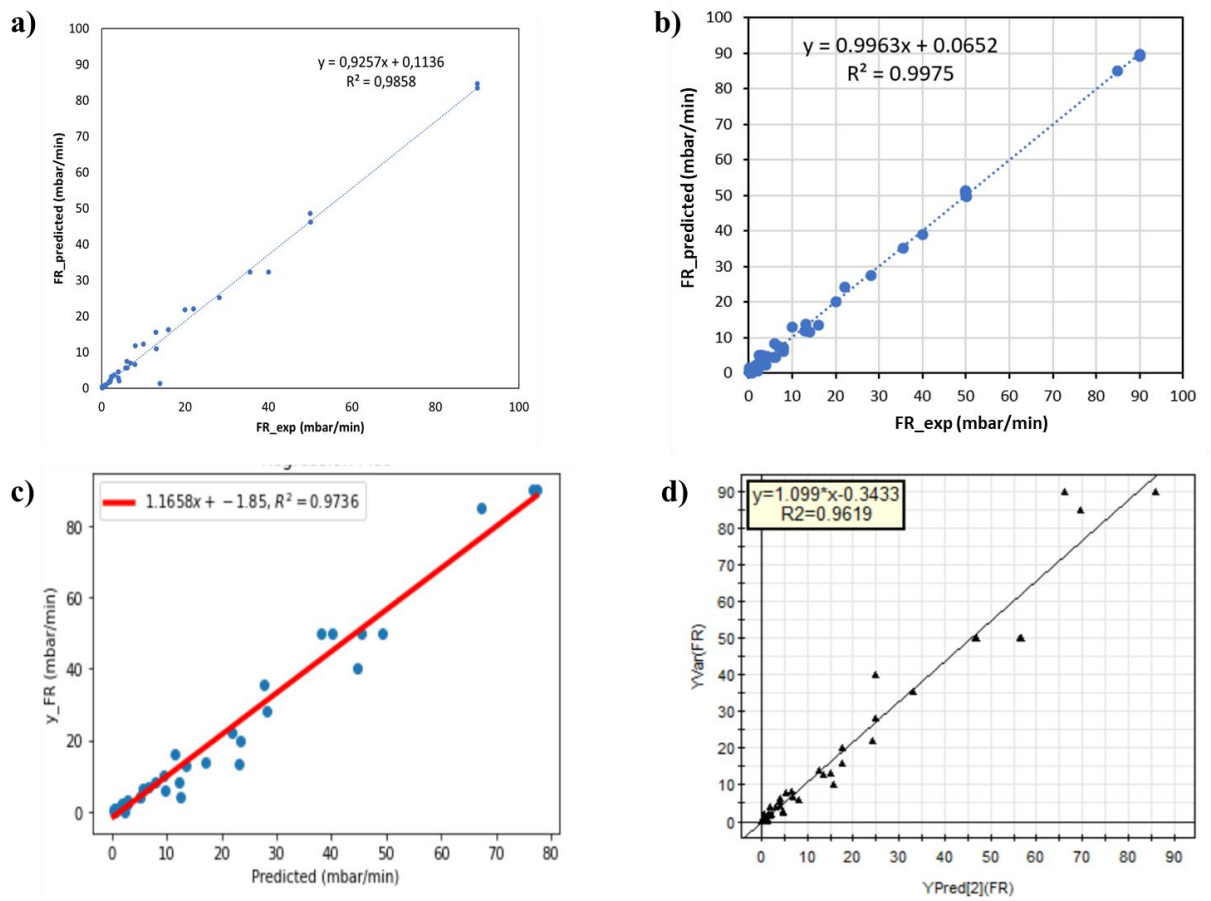


**Figure 1** Predicted and experimental fouling rate data for a) FL, b) ANN, b) RF and c) PLS models.

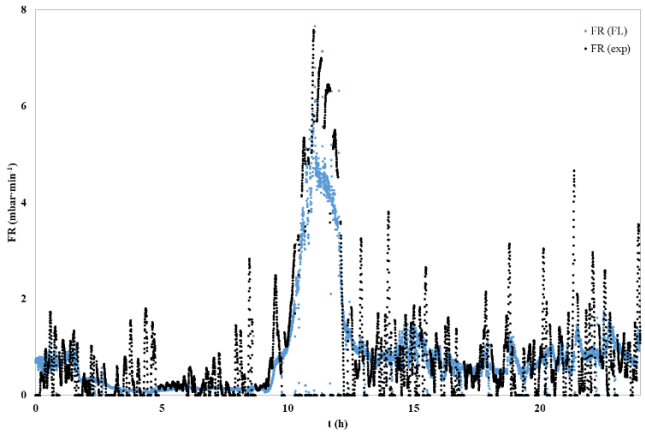


**Figure 2** Validation results for the FL model.

It is important to note that all models have been developed and validated using data from critical flux tests, which notably minimises the time required for data collection and model calibration. Nonetheless, despite the promising results obtained, it is important to highlight the simplifications inherent in each data-driven model, especially when compared to deterministic models. Therefore, when comparing and selecting a specific data-driven model, it is necessary to find a trade-off between complexity, simplicity and accuracy of each model.

## 4. CONCLUSIONS

Four data-driven models for predicting fouling in AnMBRs have been developed and validated. The results showed that the models demonstrated a great potential to predict TMP dynamics.

## ACKNOWLEDGEMENT

Grant PID2021-128426OA-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe, European Union.

Grant TED2021-132837A-I00 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR.

Grant ACIF/2021/384 funded by Generalitat Valenciana and by European Union.

Grant CIAICO/2023/165 funded by Conselleria de Educación, Cultura, Universidades y Empleo, Generalitat Valenciana.

## REFERENCES

Mannina G., Di Bella G. and Viviani G. 2011 An integrated model for biological and physical process simulation in membrane bioreactors (MBRs). *J. Membrane Sci.* **376**, 56-69.

Ng A.N.L. and Kim A.S. 2007 A mini-review of modeling studies on membrane bioreactor (MBR) treatment for municipal wastewaters. *Desalination* **212**, 261-281.

Robles A., Ruano M.V., Ribes J. and Ferrer J. 2013 Factors that affect the permeability of commercial hollow-fibre membranes in a submerged anaerobic MBR (HF-SAnMBR) system. *Water Res*. **47**, 1277-1288.