Real-time monitoring of membrane bioreactors with 2D-fluorescence data and statistically based models

C. F. Galinha, G. Carvalho, C. A. M. Portugal, G. Guglielmi, R. Oliveira, J. G. Crespo and M. A. M. Reis

ABSTRACT

The application of membrane bioreactors (MBR) for wastewater treatment is growing worldwide due to their compactness and high effluent quality. However, membrane fouling, mostly associated to biological products, can reduce MBR performance. Therefore, it is important to monitor MBRs as close to real-time as possible to accelerate control actions for maximal biological and membrane performance. 2D-fluorescence spectroscopy is a promising on-line tool to simultaneously monitor wastewater treatment efficiency and the formation of potential biological fouling agents. In this study, 2D-fluorescence data obtained from the wastewater and the permeate of a MBR was successfully modelled using projection to latent structures (PLS) to monitor variations in the influent and effluent total chemical oxygen demand (COD). Analysis of the results also indicated that humic acids and proteins highly contributed to the measured COD in both streams. Nevertheless, this approach was not valid for other performance parameters of the MBR system (such as influent and effluent ammonia and phosphorus), which is usually characterised through a high number of analytical and operating parameters. Principal component analysis (PCA) was thus used to find possible correlations between these parameters, in an attempt to reduce the analytical effort required for full MBR characterisation and to reduce the time frame necessary to obtain monitoring results. The 3 first principal components, capturing 57% of the variance, indicated and confirmed expected relationships between the assessed parameters. However, this approach alone could not provide robust enough correlations to enable the elimination of parameters for process description (PCA loadings ≥ 0.5).

Nevertheless, it is possible that the information captured by 2D-fluorescence spectroscopy could replace some of the analytical and operating parameters, since this technique was able to successfully describe influent and effluent total COD. It is thus proposed that combined modelling of 2D-fluorescence data and selected performance/operating parameters should be further explored for efficient MBR monitoring aiming at rapid process control.

Key words | 2D-fluorescence spectroscopy, membrane bioreactors, principal component analysis, project to latent structures

INTRODUCTION

Membrane bioreactors (MBR) are increasingly applied in wastewater treatment plants, mainly due to their small footprint and high effluent quality. The increasing demand for MBR technology requires the development of adequate monitoring and control techniques, particularly in view of the high operational costs associated to membrane fouling resulting from the adhesion of cells and cell products. Monitoring of MBR performance involves a high number of off-line and time-consuming analytical techniques, regarding the complexity of the media. Therefore, MBR technology would greatly benefit from real-time monitoring techniques that could be used to support immediate control actions.

2D-fluorescence spectroscopy is an on-line and non-destructive technique that can quickly provide information about the composition of complex biological media and
consequently be used as a real-time monitoring tool. Wastewater media contain high quantities of natural fluorophores, such as aminoacids (e.g. tyrosine, tryptophan and phenylalanine), vitamins, coenzymes and aromatic organic matter in general. Furthermore, extracellular polymeric substances (EPS) containing large amounts of proteins, are the major fouling agent of MBRs. Thus, fluorescence is a good candidate technique for MBR monitoring, able to capture fingerprinting information on the state of the biological media, including EPS.

Fluorescence spectroscopy, using selected excitation/emission wavelengths, was first explored by Li & Humphrey (1991) and Li et al. (1991) for monitoring cell growth and activity in biological reactors. Later, other studies have applied 2D-fluorescence spectroscopy to monitor water and wastewater treatment processes (Her et al. 2003; Lee et al. 2006; Kimura et al. 2009; Wang et al. 2009), and the need to extract deeper, quantitative information from 2D-fluorescence spectra obtained from high complex media guided other authors to use multivariate statistical tools (Wolf et al. 2001; Boehl et al. 2003; Surribas et al. 2006; Ganzlin et al. 2007; Teixeira et al. 2009).

In wastewater treatment plants (WWTPs), media are highly complex, generally composed by a wide variety of molecular species (fluorophores and non-fluorophores) which may promote mutual interference effects on the fluorescence signal. Consequently fluorescence excitation/emission matrices (EEMs) obtained from these systems contain highly embedded information. In this study, the information contained in fluorescence spectra was assessed using projection to latent structures (PLS) aiming at establishing correlations with performance parameters of a MBR.

PLS is a linear regression method that maximises the covariance between input data and output performance parameters. PLS models generate a set of latent variables that explain the maximum variance in the output variables, combining in a single step data decomposition and correlation with predicted outputs. PLS were used in this work to predict quality parameters (such as chemical oxygen demand, COD) of wastewater and permeate based only on 2D-fluorescence data. Moreover, the importance of specific areas of the fluorescence matrices was investigated for the prediction of each output.

In MBRs for wastewater treatment the relationships between operating and performance parameters are abundant and complex, essentially due to the interdependency between operating parameters, biological performance and membrane performance. This complexity was assessed in this work by principal component analysis (PCA) (Rencher 2002) of analytical parameters conventionally used to monitor MBRs for wastewater treatment.

PCA is often used to examine interrelationships between a large number of variables and to explain these variables in terms of their common underlying dimensions (Hair et al. 1998). In PCA, the input matrix is described as a linear correlation between scores and loadings that minimises the residuals. The best linear combination of variables is determined in order to capture the maximum variance in data. So, the first principal component (PC1) can be seen as the best summary of linear relationships present within data. The second principal component (PC2) is identified as orthogonal to the first one, and aims to find the best relationship for the remaining variance. The following PCs are all orthogonal between each other. The interpretation of the contribution of each variable in the PCA is quantified by the loadings matrix. The loadings represent the degree of correlation between the variables and the principal components. Therefore similar values represent variables with a high correspondence as well as higher values of loadings point to a more representative variable (Jackson 2003).

In this work, correlations across data were analysed based on loadings resultant from PCA applied to a comprehensive set of MBR operating and performance parameters.

The aim of this work was the development of a strategy for real-time monitoring of a MBR through the use of an online method based on 2D-fluorescence spectroscopy data and PLS modelling. Alternatively, this study also evaluated the reduction of the number of MBR monitoring parameters by multivariate data analysis that would enable the elimination of redundant analysis.

**METHODS**

**Membrane bioreactor**

The pilot scale MBR was operated to treat domestic wastewater and monitored with 2D-fluorescence spectroscopy. It was located at the wastewater treatment plant of Lavis, Italy and consisted of a biological anoxic/aerobic tank followed by a separate tank with a submerged hollow fibre system (ZeeWeed 500c, Zenon) with 0.04 µm membrane pore size (Figure 1). This pilot MBR was monitored with 2D-fluorescence for a period of 10 months, when it was operated under controlled permeate flux. During this period, operational changes were programmed and imposed in the permeate flux and solids retention time. Temperature changed due to seasons’ weather, hydraulic retention time
and dissolved oxygen (DO) changed due to other operating and control experiments.

2D-fluorescence spectroscopy

Fluorescence EEMs were acquired with a fluorescence spectrophotometer Varian Cary Eclipse coupled to a fluorescence optical fibre bundle probe. Fluorescence spectra were generated in a range of 250 to 700 nm (excitation) and 260 to 710 nm (emission), with an excitation wavelength incrementing step of 10 nm. Fluorescence spectra were obtained using excitation and emission slits of 10 nm and a scan speed of 3,000 nm/min.

Data collection

2D-fluorescence spectra were acquired in the wastewater feed, in the bulk activated sludge and in the permeate at the same time that samples were collected for further analysis of wastewater, permeate and mixed liquor (Table 1). Also transmembrane pressure (TMP), temperature and dissolved oxygen were measured on-line. All these data collected together with selected operating parameters—permeate flux, hydraulic retention time and sludge retention time—were used in this work to find correlations by multivariate analysis.

Multivariate data analysis

Before multivariate data analysis, all data was normalised by subtracting the respective average values and dividing by their standard deviations.

The 2D-fluorescence spectroscopy measurements were acquired and plotted in excitation-emission matrices (EEMs) where each value of fluorescence intensity corresponds to a pair of excitation/emission wavelengths, totalling 5,490 model input variables. The mathematical models were obtained through PLS regression, using EEMs of wastewater and permeate as model inputs and total COD in wastewater and permeate as outputs, respectively. Data from 146 observations obtained throughout the 10 months of operation were used for PLS modelling. These 146 observations were divided randomly into a training set (75% of the observations, which were used to calibrate the model) and a validation set (25% of the observations, which were used to validate the final model). The PLS models thus obtained are a linear correlation of the 5,490 fluorescence inputs to predict total COD in the wastewater and in the permeate accordingly. Model fitting to the experimental data was assessed by the training and validation correlation coefficients ($R^2$) and root mean square error of prediction (RMSEP), calculated as the squared root of the sum of the squared differences between predicted and experimental values.

PCA was applied to all data, except fluorescence matrices, and the loadings obtained for the first 3 PCs were analysed in search for correlations within data.

Both PLS models and PCA were implemented in Matlab (Matlab 2006) according to nPLS and Parafac functions (Andersson & Bro 2000), respectively.

RESULTS AND DISCUSSION

PLS models based on 2D-fluorescence spectroscopy

Multivariate linear models were obtained to predict total COD in permeate and total COD in wastewater using fluorescence data from permeate and wastewater, respectively. PLS models were performed using the values of emission intensity for each pair of excitation/emission wavelength as inputs (total of 5,490 inputs) and respective COD as output. Predicted results of training and testing data sub-sets were plotted against the experimental values obtained from the pilot MBR (Figures 2 and 3). The PLS model to predict total COD in the permeate was generated using 3 latent variables and has a good fitting for both training and validation sets ($R^2$ of 0.88 and 0.92, respectively) and RMSEP of 5.4 mg/L, corresponding to a mean error of 15%. The PLS model to predict total COD in wastewater made with 9 latent variables gave an overall...
lower correlation ($R^2$ of 0.95 and 0.60 for training and validation data, respectively) with a RMSEP of 175 mg/L corresponding to a mean error of 33%.

The establishment of these equations demonstrates that 2D-fluorescence spectroscopy is able to describe the variation of total COD in wastewater and permeate media. Furthermore, the coefficients of the multilinear regression were analysed in order to identify the 2D-fluorescence spectral regions that specifically correlate with total COD. Thus, if the model coefficients are regarded as the weight of each input (in this case each position in the fluorescence matrices) it is possible to identify the excitation/emission pairs that may have a higher contribution to the description of total COD. In Figures 4 and 5 these coefficients are plotted as a function of their position in the fluorescence spectrum.

The coefficients for total COD prediction in wastewater (Figure 4) show that COD is mostly expressed by variations in two spectral regions at approximately 280/340 nm and 360/430 nm of excitation/emission, where proteins and humic compounds have higher emission, respectively. Thus, fluorescence response due to the presence of proteins and humic compounds appears to be the most relevant to predict COD in wastewater. In addition, the two diagonal lines in fluorescence matrices, resultant from light scattering, which is commonly correlated with media

<table>
<thead>
<tr>
<th>Monitoring parameters and respective range of values assessed during MBR operation</th>
<th>Range</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process and Performance Parameters:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total chemical oxygen demand in wastewater (mg/L)</td>
<td>CODtw</td>
<td>151 1,862 524</td>
</tr>
<tr>
<td>Soluble chemical oxygen demand in wastewater obtained by sample flocculation (mg/L)</td>
<td>CODsw</td>
<td>17 410 108</td>
</tr>
<tr>
<td>Soluble chemical oxygen demand in wastewater obtained by sample filtration (mg/L)</td>
<td>CODfw</td>
<td>46 467 155</td>
</tr>
<tr>
<td>Ammonia in wastewater (mg N/L)</td>
<td>NH₄w</td>
<td>9.1 88.5 34.2</td>
</tr>
<tr>
<td>Nitrite in wastewater (mg N/L)</td>
<td>NO₂w</td>
<td>0.01 9.98 1.08</td>
</tr>
<tr>
<td>Nitrate in wastewater (mg N/L)</td>
<td>NO₃w</td>
<td>0.1 26.8 4.4</td>
</tr>
<tr>
<td>Organic nitrogen in wastewater (mg/L)</td>
<td>Norgw</td>
<td>0.1 89.0 19.6</td>
</tr>
<tr>
<td>Phosphate in wastewater (mg P/L)</td>
<td>PO₄w</td>
<td>0.10 7.35 3.26</td>
</tr>
<tr>
<td>Total phosphorus in wastewater (mg/L)</td>
<td>TPw</td>
<td>1.2 33.0 9.5</td>
</tr>
<tr>
<td>Total suspended solids in wastewater (mg/L)</td>
<td>TSSw</td>
<td>50 2,100 316</td>
</tr>
<tr>
<td>Volatile suspended solids in wastewater (mg/L)</td>
<td>VSSw</td>
<td>30 2,071 264</td>
</tr>
<tr>
<td>Soluble chemical oxygen demand in permeate (mg/L)</td>
<td>CODtp</td>
<td>10 77 36</td>
</tr>
<tr>
<td>Soluble chemical oxygen demand in permeate obtained by sample flocculation (mg/L)</td>
<td>CODsp</td>
<td>10 67 28</td>
</tr>
<tr>
<td>Ammonia in permeate (mg N/L)</td>
<td>NH₄p</td>
<td>0.1 19.3 2.4</td>
</tr>
<tr>
<td>Nitrite in permeate (mg N/L)</td>
<td>NO₂p</td>
<td>0.01 0.84 0.13</td>
</tr>
<tr>
<td>Nitrate in permeate (mg N/L)</td>
<td>NO₃p</td>
<td>0.1 37.2 13.6</td>
</tr>
<tr>
<td>Organic nitrogen in permeate (mg/L)</td>
<td>Norgp</td>
<td>0.1 6.0 1.4</td>
</tr>
<tr>
<td>Phosphate in permeate (mg P/L)</td>
<td>PO₄p</td>
<td>0.10 3.67 1.31</td>
</tr>
<tr>
<td>Total phosphorus in permeate (mg/L)</td>
<td>TPs</td>
<td>0.1 4.4 1.7</td>
</tr>
<tr>
<td>Total suspended solids in membrane reactor (mg/L)</td>
<td>TSSs</td>
<td>5,900 13,200 9,121</td>
</tr>
<tr>
<td>Volatile suspended solids in membrane reactor (mg/L)</td>
<td>VSSs</td>
<td>4,266 9,491 6,694</td>
</tr>
<tr>
<td>Mixed liquor suspended solids acquired on-line in the biological reactor (g/L)</td>
<td>MLSSbio</td>
<td>4.60 8.70 6.97</td>
</tr>
<tr>
<td>Transmembrane pressure (mbar)</td>
<td>TMP</td>
<td>100 378 195</td>
</tr>
<tr>
<td>Operating Parameters:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permeate flux (m³/h)</td>
<td>Q</td>
<td>0.80 3.00 1.80</td>
</tr>
<tr>
<td>Hydraulic retention time (h)</td>
<td>HRT</td>
<td>5.30 19.86 9.71</td>
</tr>
<tr>
<td>Dissolved oxygen (mg/L)</td>
<td>DO</td>
<td>0.10 7.37 1.64</td>
</tr>
<tr>
<td>Sludge wastage (m³/day)</td>
<td>Vslg/day</td>
<td>0.00 1.29 0.57</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>T</td>
<td>9.0 26.9 20.7</td>
</tr>
</tbody>
</table>
turbidity, suspended solid matter and high concentrations of solutes (Harms et al. 2002; Rinnan et al. 2005), also appear to be of high importance. This result is not a surprise when taking into account that some compounds contributing to total COD in domestic wastewaters (particulate COD) are associated with media turbidity.

In the permeate (Figure 5), total COD seems to be strongly described by the humic compounds region of fluorescence spectra and also by light scattering. This result was expected since proteins are generally larger than humic compounds and thus are more retained by the membrane; therefore, the permeate COD can be mostly predicted by fluorescence signal of humic compounds, as opposed to wastewater COD where proteins also have an important contribution. Light scattering in permeate EEMs is also lower than in wastewater (data not shown), likely due to the absence of turbidity and to reduced concentrations of solutes and colloids. However, its contribution to COD prediction is also important as revealed by the analysis of contributions given by the PLS coefficients (Figure 5).

Despite the fact that some areas of the fluorescence spectra have higher importance to predict total COD in permeate and in wastewater, the complete matrices have a role in modelling and it is not possible to arbitrary remove these inputs without compromising output prediction.

2D-fluorescence spectroscopy showed to be rich in information about media composition (enough to predict the variation of total COD) and therefore can be a useful tool in real-time monitoring of MBRs. However, the fluorescence EEMs could not provide sufficient information to predict other parameters, such as influent and effluent ammonia and phosphorus (results not shown). Therefore, the accurate quantitative monitoring of MBR performance will still require the acquisition of at least some process variables.

**Principal component analysis of operational and performance parameters**

In an attempt to reduce the analytical effort required for full characterisation and to reduce the time frame necessary to obtain monitoring results, PCA was used to study the complex correlations between a comprehensive set of operating and performance parameters (Table 1). Using PCA applied simultaneously to all data, it is possible to understand which parameters are strongly correlated and therefore reduce the number of parameters used to monitor MBRs performance.

In PCA, a new system of axes is generated based in the maximum variance captured in order to reduce redundancy and noise, and describe data more accurately. In this study, PCA was applied to all analytical data acquired together with the operating conditions used to obtain each observation. Loadings quantify the difference between the new system of axes and the original one. Therefore, the variables that are correlated have a similar value of loading or are in the same line that crosses the axes intersection, since the direction of the PC is identical. Otherwise, the uncorrelated variables have an orthogonal relationship and appear in a perpendicular direction to the considered variable. The strongest correlation is obtained when loadings have the value of one.

**Figure 6** illustrates the loadings of the PC1, PC2 and PC3 for operating and performance MBR data. PC1 captured 28.9% of variance, PC2 19.6% and PC3 8.5%, totalling 57% of variance for the first 3 principal components. The subsequent PCs captured progressively lower variance percentages, thus only the loadings of the first 3 PCs were analysed. Loadings of PC1 and PC2 show that transmembrane pressure (TMP) belongs to the same
cluster as ammonia (NH₄⁺), soluble COD by flocculation (CODsw) and soluble COD by filtration (CODfw) in wastewater, and therefore is positively correlated with these wastewater parameters. Also shows that ammonia in the permeate (NH₄⁻) decreases when the hydraulic retention time (HRT) increases. Loadings of PC2 and PC3 show that TMP is positively dependent of the permeate flux (Q) and the suspended solids in wastewater (total and volatile, respectively TSSw and VSSw). It is also noticeable, looking to the loading plots, that PC1 captures essentially the variation in permeate quality in function of the biomass concentration: permeate quality parameters (CODtp, CODsp, NO₃⁻, PO₄⁻ and TPp) are inversely correlated with the amount of sludge in the MBR (mixed liquor suspended solids in biological compartment, MLSSbio, and volatile and total suspended solids in sludge, VSSs and TSSs respectively). Additionally, PC2 captures variations due to the flux (note that both permeate flux, Q, and hydraulic retention time, HRT, were plotted since the MBR volume changed during the study).

It is also evident that permeate quality parameters vary inversely with temperature (T), which is consistent with the higher activity of sludge expected and observed at higher temperature (during the summer period). Moreover, since aeration was maintained constant, the increase of dissolved oxygen (DO) means that less oxygen is consumed and therefore lower sludge activity and lower COD and nutrients removal, as reflected by the inverse correlation of DO with T and positive correlation with permeate quality parameters.

Figure 4 | Coefficient values of the PLS model for prediction of total COD in wastewater plotted in function of the respective variable position in fluorescence matrices. Subscribers to the online version of Water Science and Technology can access the colour version of this figure from http://www.iwaponline.com/wst.

Figure 5 | Coefficient values of the PLS model for prediction of total COD in permeate plotted in function of the respective variable position in fluorescence matrices. Subscribers to the online version of Water Science and Technology can access the colour version of this figure from http://www.iwaponline.com/wst.
Other expected correlations are evident in the loadings plots such as the proximity between soluble COD by filtration and by flocculation in wastewater, or between MLSSbio, TSSs, VSSs and sludge wastage (Vslg/day) since the first three are different measurements of suspended solids and sludge wastage were performed in order to keep specific solid retention times.

These results confirm the complexity and abundance of correlations across data and show that process variables are not fully described by the first 3 PCs of the PCA, which only capture 57% of total variance. Additionally, loadings have low values ($\leq 0.5$), meaning that the monitoring parameters are not fully described by any of the three PCs. Therefore, and despite all the relationships found, it is not possible to identify strong correlations that could reduce the number of analytical parameters needed to make a complete characterisation of the MBR performance.

**CONCLUSIONS**

In this study, 2D-fluorescence spectroscopy data was combined with PLS modelling to predict performance parameters of an MBR, aiming at developing a strategy for real-time monitoring. Using only this technique, it was possible to describe total COD in both the influent wastewater and the permeate of a MBR. However, this approach was not successful in accurately predicting other performance parameters, suggesting that 2D-fluorescence spectroscopy cannot totally replace conventional MBR monitoring.

Alternatively, a set of MBR process and operating parameters were analysed by PCA in order to find correlations between them and thus reduce redundancy and analytical time necessary for monitoring. This multivariate analysis tool revealed some degree of correlation between certain parameters. However, these correlations were not strong enough (loadings $\leq 0.5$) to reduce the number of
parameters needed to describe the process. Additionally, the PCA possibly did not cover all the existing correlations, since only a low level of variance (57%) was captured using the most contributing PCs.

Overall, it was found that correlations across the MBR data are abundant and that relationships between operating parameters and performance variables are complex and interdependent. However, it is possible that the information captured by 2D-fluorescence spectroscopy could replace some of the analytical and operating parameters, since this technique was able to successfully describe some MBR influent and effluent quality indicators (total COD). It is thus proposed that combined modelling of 2D-fluorescence data and selected performance/operating parameters should be further explored for efficient MBR monitoring aiming at rapid process control.

ACKNOWLEDGEMENTS

The authors would like to thank the European Commission for its financial support through the EUROMBRA project (Contract No. 018480, FP 6th—Global Change and Ecosystems) and Fundação para a Ciência e Tecnologia (Portugal) for the project PTDC/EBB-EBI/098862/2008 and grants SFRH/BD/30253/2006 and SFRH/BPD/30800/2006.

REFERENCES


