

Activated sludge process monitoring through *in situ* near-infrared spectral analysis

A. M. A. Dias, I. Moita, R. Páscoa, M. M. Alves, J. A. Lopes and E. C. Ferreira

ABSTRACT

The application of near infrared (NIR) spectroscopy for industrial process monitoring is achieving increasing importance over the last twenty years. In fact, the real time monitoring capacity of NIR spectroscopy is a very important feature for process monitoring, prediction and control as it allows a fast evaluation of the state of the process. However, the application of NIR spectroscopy in wastewater treatment processes is still to be explored. Although some applications of the technique for wastewater monitoring have been reported in the literature, there is still a need for more investigation related with applications, limitations and advantages of the technique when compared with other methods.

An activated sludge reactor for aerobic treatment of a complex medium was monitored *in situ* with a NIR transreflectance probe and traditional chemical parameters analysed off-line. NIR spectrophotometric data measured at the feed, reactor and settler were coupled to principal component analysis (PCA) to infer about the ability of this monitoring system to detect changes in the feed influent. The analysis of the score plots resulting from PCA permitted to identify the moments at which the perturbations occurred and to follow the consequent instability induced in the reactor till the day where the system is recuperated. The promising results obtained, suggest the interest in more detailed studies on the feasibility of NIR spectroscopy as an alternative method for monitoring and control of wastewater treatment processes.

Key words | activated sludge process, chemometrics, near infrared spectroscopy (NIR) on-line monitoring, wastewater treatment

A. M. A. Dias
I. Moita
M. M. Alves
E. C. Ferreira

IBB-Institute for Biotechnology and Bioengineering,
Centre of Biological Engineering,
Universidade do Minho,
Campus de Gualtar,
4710-057 Braga,
Portugal
E-mail: adias@deb.uminho.pt;
isabelmoita@deb.uminho.pt;
madalena.alves@deb.uminho.pt;
ecferreira@deb.uminho.pt

R. Páscoa
J. A. Lopes

Requimte, Faculty of Pharmacy,
University of Porto,
4050-047 Porto,
Portugal
E-mail: joaolopes@ff.up.pt

INTRODUCTION

Activated sludge systems are among the most widely used secondary biological process treatments. It consists of inoculating a high filamentous and floc-forming bacteria concentration responsible for the oxidation of the organic matter in an aerated tank. Subsequently, the flocculated biomass is separated by means of their settling ability from the treated effluent in a settling tank. Part of the settled biomass is then returned to the aerated tank in order to maintain a constant biomass concentration. As most biological processes, aerobic systems are sensitive to sudden changes in

feedstock composition, which cause significant variability in the process conditions. So far, the control of such processes is most of the times achieved through manual sample extraction with off-line analysis of a few key process parameters such as total solids (TS), volatile solids (VS), chemical oxygen demand (COD), etc. This kind of analysis is time consuming, expensive, and can only provide a temporary view of the system's performance. Therefore, fast, simpler and inexpensive analysis techniques, that allow continuous and *in situ* monitoring, are needed. Near infrared spectroscopy (NIR) is

being proposed as a valuable candidate (Benson 1996). The low reflectivity and absorptivity in the NIR-range makes it possible to analyse samples that are strongly light scattering, such as opaque liquids and slurries (Hansson *et al.* 2003). The NIR region permits to record the response of the molecular bonds of certain chemical species to the IR radiation and consequently originate the characteristic spectrum of the sample being analysed.

Among the advantages of NIR spectroscopy, when compared to standard methods, are the facts that it is a non-destructive and non-invasive, fast, low maintenance cost technique that do not use reagents, do not produce residuals and allows the determination of several parameters simultaneously (Hansson *et al.* 2003; Pons *et al.* 2004; Uddin *et al.* 2006). The main disadvantages attributed to NIR technologies are the fact that although being considered a very flexible method, it is often affected by the operation conditions like temperature, agitation, aeration, dispersive light, etc. These interferences difficult the calibration process and reduce the quality of the results for quantitative assessments (Blanco *et al.* 2001). Moreover, it is not possible to obtain direct information from the technique by itself. In fact, due to the large amount of information inherent to each spectrum, the technique is always associated with chemometric tools that extract and report the most relevant information that can be taken from the spectral data. Chemometric tools allow correlating the spectral patterns with variations in the physical and chemical properties of the sample being analysed. This can be a problem when no skilled hand exists. However, specialized qualifications are mostly needed for models development. After, the analysis of the results given by those models is relatively simple and intuitive and no especial qualifications are needed to interpret the information.

Numerous applications are being given to NIR spectroscopic methods for monitoring, prediction and control of industrial processes (Heikka *et al.* 1997; Geladi & Forsström 2002; McGill *et al.* 2002; Lopes *et al.* 2004a). The applications include a broad range of areas ranging from food to petrochemical industries. The advantages that have been achieved by application of NIR techniques to those processes let us admit that similar good results could be obtained if the NIR spectroscopy is applied to environmental monitoring and correlation purposes. So far, the applications of the NIR technology to environmental processes are scarce, mostly

due to the great complexity of the large number of chemical species present and to the intrinsic composition changes in the matrix. To our knowledge only a few works were published reporting correlating methods between NIR spectral data and process parameters (Stephens & Walker 2002; Hansson *et al.* 2003; Holm-Nielsen *et al.* 2006)

The aim of this work is to conjugate NIR spectral data with chemometric tools to develop multivariable supervision models to monitor the operation status of an activated sludge reactor without using further analytical information.

METHODOLOGY

Process

The lab-scale plant is based on a 14 L activated sludge tank and 25 L total volume followed by a 2.5 L settler. An effluent with average inlet $\text{COD}_{\text{in}} = 600 \text{ mg/L}$ and $Q_{\text{in}} = 4.5 \text{ L/d}$ was fed to the reactor which was inoculated with biomass collected from another continuous aerobic reactor fed with a similar composition influent. In order to create system imbalances the influent charge was approximately doubled after five days of operation and then maintained around 950 mg/L in the following days. A schematic layout of the process plant is shown in Figure 1a. The efficient agitation of the system is guaranteed by the aeration process with an air diffuser that covers the bottom of the reactor. The bioreactor was equipped with a TFK 325 thermometer (WTW, Weilheim, Germany), one SensoLyt pH electrode connected to a 296 R/RS monitor (WTW), a TriOxmatic 690 dissolved oxygen probe (WTW) and a NIR spectrometer probe (OceanOptics, model 512).

The plant was monitored at three important points: at the influent, reactor and settler. Daily minimum volume samples were taken to measure VSS concentrations (reactor and settler) and COD concentrations (influent and settler). The analyses were done by weight differences and closed reflux colorimetric methods, respectively. Each sample was analysed in duplicate. In order to extensively monitor the evolution of the process inside the aerated reactor, the NIR probe was immersed in the reactor and NIR spectrum was acquired every 5 minutes. Simultaneously to each influent and settler sampling, a minimum of 5 spectra were obtained at the same monitored points.

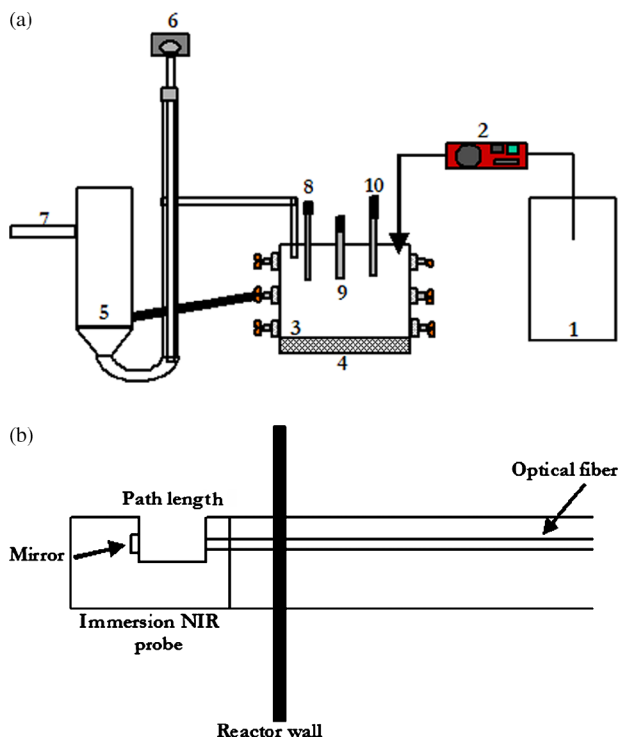


Figure 1 | (a) Schematic layout of the plant: (1) effluent, (2) feeding pump, (3) aerated tank, (4) aeration system, (5) settler, (6) air pump; (7) outlet, (8) dissolved oxygen probe; (9) pH probe, (10) NIR spectrometer probe. (b) Schematic layout of the NIR transreflectance probe.

NIR

The Ocean Optics NIR 512 model, a portable dispersive NIR equipment working in the range from 900 to 1,800 nm and including a PDA cooling detector, was used. A transreflectance probe (Ocean Optics/T300RT) is connected through optical fibers (OceanOptics/QP400-2-VISNIR) to a light source (Stellarnet/SL1) and to the NIR spectrometer which in turn is connected to a PC by a USB 2.0 cable. The beam coming from the light source passes through the sample as illustrated in Figure 1b. The optical path is twice the value of the mechanical gap of the transreflectance probe. It is adaptable and depends on the characteristics of the reactor content. In this work the optical path was equal to 1 cm. The returned beam is sent to the NIR 512 detector connected to the PC allowing the immediate spectra visualization and acquisition. Spectra are acquired using a programme from OceanOptics (OOIBase32/Ocean Optics). This software allows the configuration of certain parameters like the integration

time, average spectra, filter type (to avoid noise mostly when low integration times are used) and the temperature of the detector. All these parameters were previously optimized in order to improve the quality of the spectra acquired. When measuring for the first time, a reference spectrum is taken as in traditional spectroscopic methods. The reference spectrum is checked regularly and if changes occur, the above mentioned parameters may be adjusted in order to fix the deviations.

Data analysis and calculation

Principal Components Analysis (PCA) (Jackson 1980) is one of the most widely used chemometric tools for data compression and information extraction. The general objective of the application of the method is to describe the data using far fewer factors than original variables with no significant loss of information.

PCA finds combinations of variables, usually named *factors* that describe major trends in the data. For a given data matrix \mathbf{X} with m rows and n columns, PCA decomposes the data matrix \mathbf{X} as the sum of the outer product of vectors \mathbf{t}_i and \mathbf{p}_i plus a residual matrix \mathbf{E} :

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_k\mathbf{p}_k^T + \mathbf{E} \quad (1)$$

where k must be less than or equal to the smaller dimension of \mathbf{X} , i.e. $k \leq \min\{m, n\}$. The \mathbf{t}_i vectors are known as *scores* and contain information on how the *samples* relate to each other. The \mathbf{p}_i vectors are known as *loadings* and contain information on how the *variables* relate to each other. In matricial terms Equation 1 can be written as

$$\mathbf{X}(n \times p) = \mathbf{T}(n \times d)\mathbf{P}^T(d \times p) + \mathbf{E}(n \times p) \quad (2)$$

where n represents the number of objects (spectra), p represents the number of variables (wavelengths) and d the number of PC's. \mathbf{T} , \mathbf{L} and \mathbf{E} represent the scores, the loadings and the residuals matrix.

On a PCA analysis the $(\mathbf{t}_i, \mathbf{p}_i)$ pairs are arranged in descending order according to the amount of *variance* described by the pair. The first pair captures the greatest amount of variation in the data that it is possible to capture with a linear factor and each subsequent pair captures the

greatest possible amount of variance remaining after subtracting $t_i p_i^T$ from \mathbf{X} .

For the particular case of spectroscopic data, PCA identifies the major sources of correlated variance in the collection of spectra by defining a series of ranked factors and assigning each spectrum a score based on the relative contribution of each factor. The sources of variance, once identified, can aid in the visualization of the major data trends. In this way, the data collection can be reduced from a complicated multidimensional representation to a more easily visualized two or three-dimensional space (score plots) describing the main information present in the data (Lourenço *et al.* 2006).

The MatLab version 6.5 Release 13 (The Mathworks, Inc) was utilized to data treatment, calibration and validation of the chemometric models. The chemometric functions included in the PLS MatLab Toolbox (PLS Toolbox, Eigenvector Research, Inc) were used to generate the PCA model.

RESULTS AND DISCUSSION

Data pre-treatment

The collected spectra were pre-treated by first applying the standard normal variate method (Barnes *et al.* 1989) by subtracting raw data from the mean of each spectrum and

dividing by the standard deviation of the absorbencies of each sample. The second derivative was then applied to the data according to the method of Savitzky-Golay (Savitzky & Golay 1964). This procedure reduces the spectral noise and the effect of light dispersion due to the presence of particles as mentioned by Karlsson *et al.* (1995). The authors compared different pre-treatment methods and concluded that the Savitzky-Golay method is the most efficient, at least when dealing with spectral data pre-treatment. Figure 2 exemplifies the differences in the raw (a) and pre-treated (b) spectra for the data collected at the aerobic reactor during a week of operation. Pre-treated data was used for PCA model development.

Model development

The number of components used in a PCA model represents a measure of the data complexity and can be regarded as the number of independent underlying phenomena. There are several methods which can be used to establish the correct number of PC's in a PCA model. In the present case this number was assessed using a bootstrap strategy (re-sampling method). This method re-samples the spectra and builds several PCAs. Then the standard deviation of the eigenvalues (or captured variance) is estimated. Figure 3 represents the percentage of variance captured by each PC as a function of the number of PCs. The confidence interval

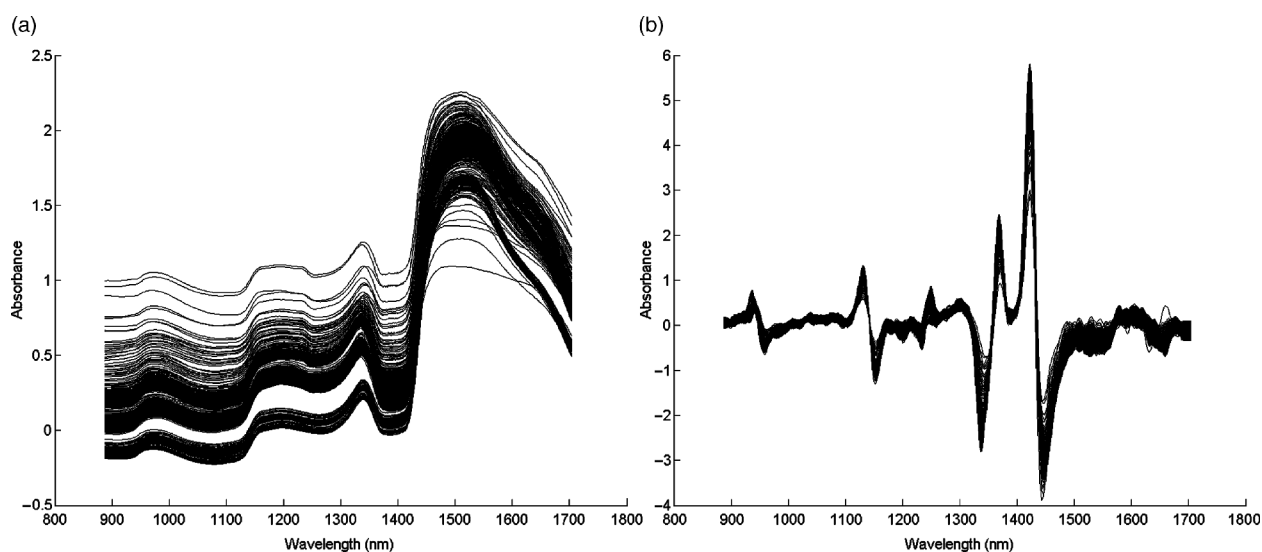


Figure 2 | Raw (a) and pre-treated (b) spectra for the data collected at the aerobic reactor during a week of operation.

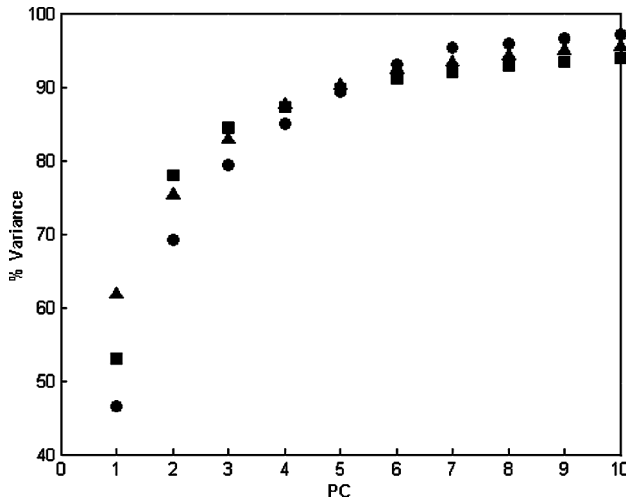


Figure 3 | Percentage captured variance for each PC in the PCA model of mean centred spectra of samples collected from the feed (●), the reactor (■) and the settler (▲).

thus obtained is used to verify statistical significance of each PC. In these data only 2 components appear to be statistically significant at a significance level equal to 0.05. This conclusion was the same for the spectra captured in the feed, reactor and settler.

The detection of anomalies by analysis of data as represented in Figure 2 can be a hard or even impossible task. In this case it is possible to distinguish groups of spectra, probably indicating similar operating conditions, and more isolated spectra most probably remarking disturbances on the system. However, on-line monitoring has to accent on more direct, practical and clear information in order to detect possible disturbances as soon and accurate as possible and to extract the maximum information that can in reality be 'hidden' from direct visual observation.

A more explicit way to analyse and detected changes in the measured data is through a score plot analysis. A score plot is any pair of score vectors plotted against each other, in which each sample spectrum appears as a data point and closely interrelated samples appear clustered together. The score plot of PC1 vs PC2 from the PCA model with two components generated from the pre-treated sample spectra for each of the monitored zones is shown in Figure 4.

The figures present the evolution of the spectra measured for a week of operation in the feed (a), in the reactor (b) and in the settler (c). All the samples taken on a same day of operation are represented by the same symbol

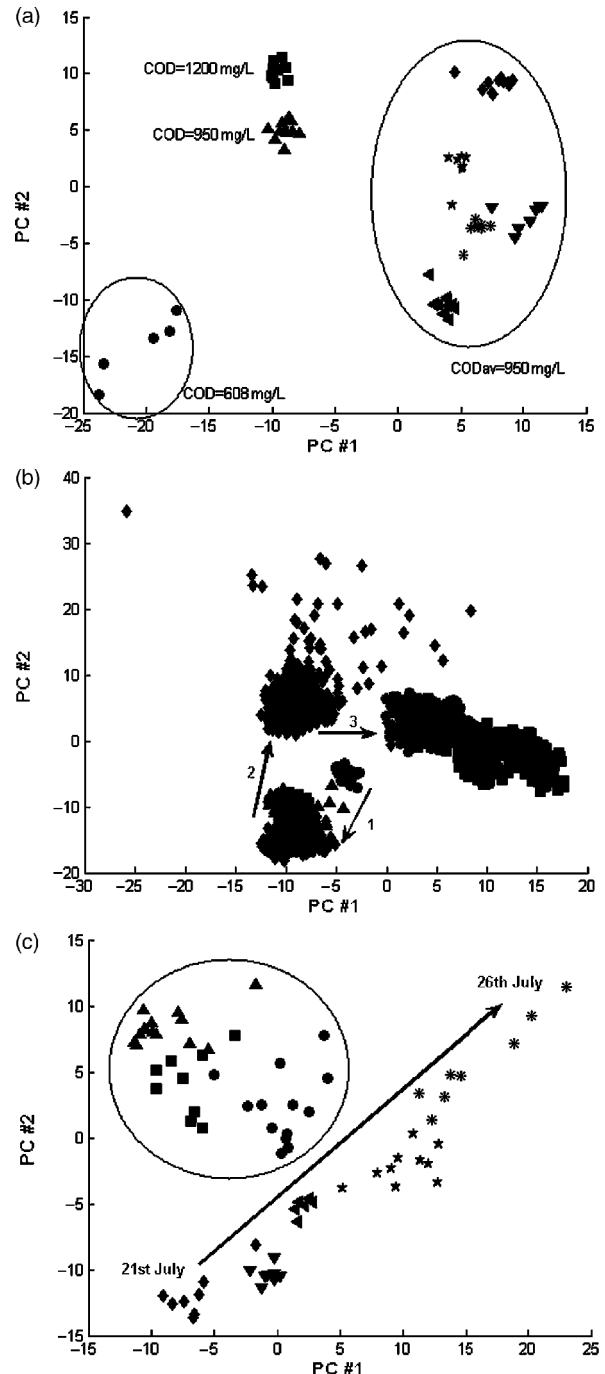


Figure 4 | Score-plot of the first two principal components from the PCA model of the NIR spectra from the feed (a), reactor (b) and settler (c) sample sets.

in the three figures. At the influent it is possible to perfectly distinguish three clusters. The first one identifies the feed samples before the increase of the COD (circles in Figure 4a). After the charge increment, a second cluster

appears including the samples corresponding to the higher COD fed to the system (squares in Figure 4a). Finally, the third cluster corresponds to the samples with an average COD equal to 950 mg/L corresponding to the days after the increment of the influent charge. According to the COD values, samples represented by triangles in the figure should also be located inside this third cluster instead of being closer to the samples with higher COD. It is believed that the reason for this behaviour has to do with the fact that on day 19, the feed with a $COD_{in} = 1,200$ mg/L was contaminated with microorganisms which promoted the degradation of the organic matter decreasing the COD_{in} that was monitored on day 20 (corresponding to the triangles in the figure). The samples located in the third cluster on the score plot diagram correspond to feed samples prepared on the corresponding monitoring day. On the reactor (Figure 4b), it is also possible to distinguish three clusters however, the 'frontiers' are not so clear in this case. Here, the large increase in COD_{in} is again easy to identify (1st arrow) but on the second cluster appears a third group of data that was not present on the influent analysis mentioned before. This state evolves to an intermediate state that conduces to a third data cluster where the samples correspond to the last days of operation are located, when the new stabilization point of the system appears. Following the arrows on the figure it is possible to identify the different states of operation inside the reactor but it is not possible to precisely identify the moments when the perturbations occur. This is easily understood having in mind the non-linear and slowly

response characteristics of a biological process like the one being studied. Finally, Figure 4c, refers to the monitoring process of the settler. Here, two major periods can be identified: one including the samples located inside the circle and the other including the samples that evolve along the arrow. The samples in the circle include data from day 19 (when the perturbation occurred) and day 20. At this point the settler was not yet affected by the increase in the influent charge. However, after day 21 the samples are located in a second cluster, meaning that a change was detected in the settler conditions. This observation can be easily understood knowing that the hydraulic residence time of the reactor was around 2 days.

During the days following the increase in the influent charge, the probe was immersed in the reactor for a continuous *in situ* monitoring in order to follow the effect of the perturbation on the process. As shown in Figure 5, the scores evolve along the plot originating clusters clearly indicating that changes are occurring in the reactor. They reflect changes due to variations in the organic compounds concentrations as well as due to the presence of particles (concentration, size, etc). In fact, NIR spectra reflect both the chemistry and physics of the influent stream. NIR absorption bands are originated by C–H, N–H and O–H bonds. However, because measurements are being made in transmittance mode and samples are slurries there is also the effect of light scattering that is originated by suspended particles. An increase in the particle number (or particle size) increases the internal diffuse reflectance phenomena

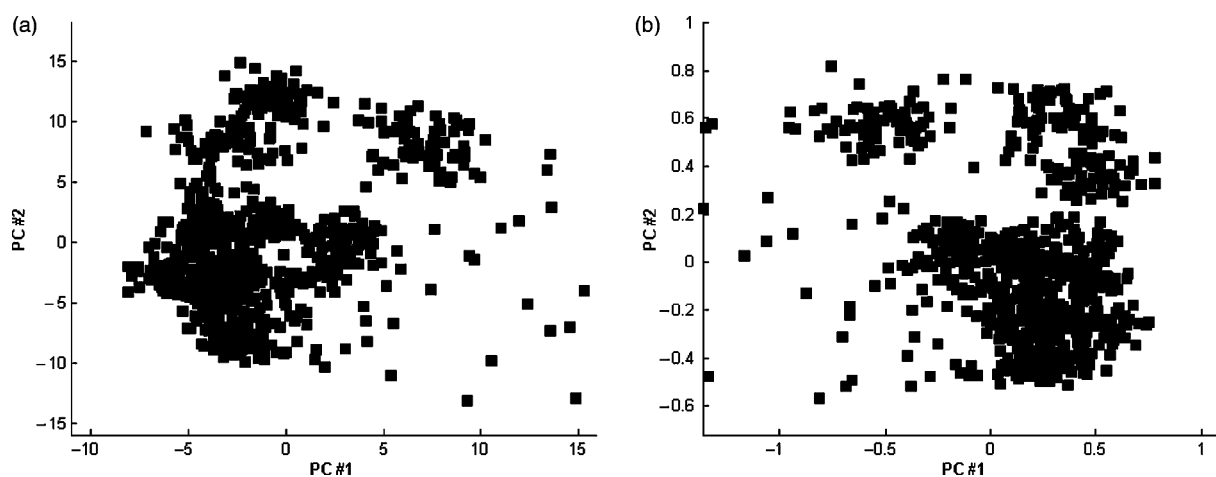


Figure 5 | Score plots obtained after a continuous monitoring period with the probe immersed in the reactor for days 20 (a) and 21 (b), after the increase in the influent charge.

which in practice increases the absorbance. As can be seen in Figure 6f, the concentration of the biomass in the reactor is increasing also promoting in some way the evolution of data along the score plot diagram.

An interesting way to monitor the state of the process along the different days of operation is by representing on the same score plot the data for the three monitoring points of the process. Figure 6 presents the sequence of the score

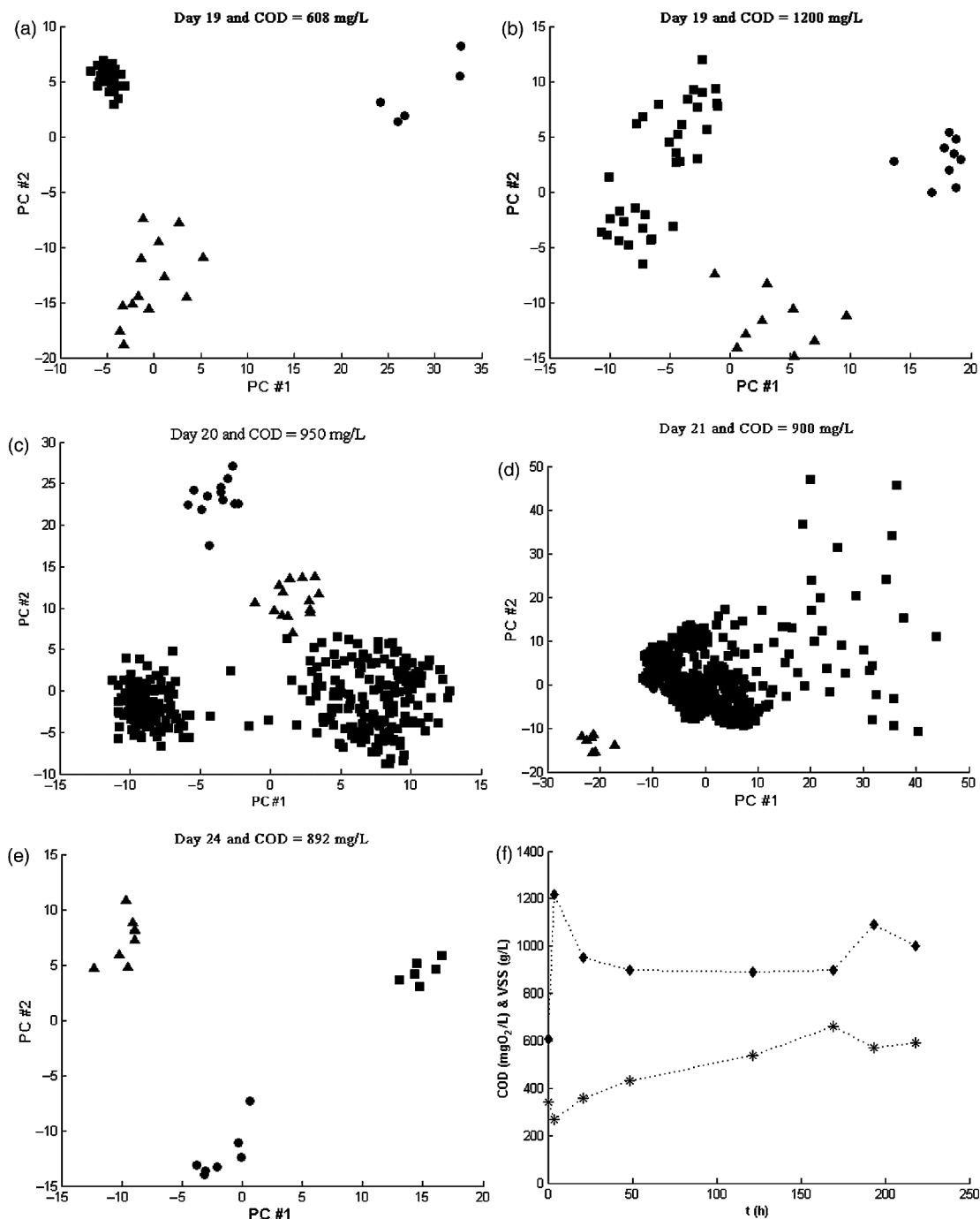


Figure 6 | Evolution of the state of an aerobic biological reactor monitored through the analysis of score plots including the three monitored points of the reactor: Feed (●), Reactor (■) and Settler (▲). Figure (f) represents the evolution of the COD in the feed (◆) and the VSS in the reactor (*).

plots obtained for different days of operation following the moment where the COD_{in} was largely increased.

From figure a to b it is possible to detect a first change on the reactor corresponding to the moment where the influent charge changed from 600 to 1,200 mg/L. One day after the perturbation, the reactor is still under unstable operating conditions or at least at conditions that deviate from the starting conditions. After five days of operation, the system is recovered and it is possible to assist again to the differentiation between the three monitored points of the process. It is interesting to notice that data from the influent or the settler moved along the score plot indicating that the new conditions are not equal to the starting conditions, i.e., the system evolved to a different equilibrium state.

CONCLUSIONS

The preliminary results presented in this work showed that NIR spectra, collected along a pilot installation of an aerobic biological wastewater treatment process, contain information that can be extracted and used for qualitative process monitoring. The descriptive capacity of PCA is evident in the score plot of a PCA model using 2 PCs applied to the NIR spectra of samples collected at the feed, the reactor and the settler of the pilot installation. It was possible to identify the moment when an induced perturbation occurred and to follow the evolution of the process until a new equilibrium state was reached. These results support the use of on-line monitoring methods in quality assessment of biological wastewater treatment processes monitoring and control. From these preliminary studies it can thus be concluded that NIR spectroscopy associated with simple PCA can be a valuable tool for wastewater quality monitoring.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the *Fundação para a Ciência e a Tecnologia* (Portugal) to A.M.A. Dias through a post-doctoral grant and to the project POCTI/AMB/60141/2001.

REFERENCES

- Barnes, R., Dhanoa, M. & Lister, S. 1989 Standard normal variate transformation and detrending of near infrared diffuse reflectance. *Appl. Spectrosc.* **43**, 772–777.
- Benson, I. B. 1996 Not just moisture: a review of some commercially successful near infrared applications. In: Davies, A. M. C. & Williams, P. (eds) *Near Infrared Spectroscopy: The Future Waves*. NIR Publications, West Sussex, UK, pp. 239–248.
- Blanco, M., Coello, J., Montoliu, I. & Romero, M. A. 2001 Orthogonal signal correction in near infrared calibration. *Anal. Chim. Acta* **434**, 125–132.
- Geladi, P. & Forsström, J. 2002 Monitoring, of a batch organic synthesis by near-infrared spectroscopy: modeling and interpretation of three-way data. *J. Chemom.* **16**, 329–338.
- Hansson, M., Nordberg, A. & Mathisen, B. 2003 On-line NIR monitoring during anaerobic treatment of municipal solid waste. *Water Sci. Tech.* **48**(4), 9–13.
- Heikka, R. A., Immonen, K. T., Minkkinen, P. O., Paatero, E. & Salmi, T. O. 1997 Determination of acid value, hydroxyl value and water content in reactions between dicarboxylic acids and diols using near-infrared spectroscopy and non-linear partial least squares regression. *Anal. Chim. Acta* **349**, 287–294.
- Holm-Nielsen, J. B., Dahl, C. K. & Esbensen, K. H. 2006 Representative sampling for process analytical characterization of heterogeneous bioreactor systems—a reference study of sampling issues in PAT. *Chemom. Intell. Lab. Syst.* **83**, 114–126.
- Jackson, J. E. 1980 Principal components and factor analysis: part 1—principal components. *J. Qual. Tech.* **12**, 201–218.
- Karlsson, M., Karlberg, B. & Olsson, R. J. O. 1995 Determination of nitrate in municipal waste water by UV spectroscopy. *Anal. Chim. Acta* **312**, 107–113.
- Lopes, J. A., Costa, P. F., Alves, T. P. & Menezes, J. C. 2004a Chemometrics in bioprocess engineering: process analytical technologies (PAT) applications. *Chemom. Intell. Lab. Syst.* **74**, 269–275.
- Lourenço, N. D., Chaves, C. L., Novais, J. M., Menezes, J. C., Pinheiro, H. M. & Diniz, D. 2006 UV spectra analysis for water quality monitoring in a fuel park wastewater treatment plant. *Chemosphere* **65**, 786–791.
- McGill, C. A., Nordon, A. & Littlejohn, D. 2002 Comparison of in-line NIR, Raman and UV-visible spectrometries, and out-line NMR spectrometry for the monitoring of an esterification reaction. *Analyst* **127**, 287–292.
- Pons, M., Bonté, S. & Potier, O. 2004 Spectral analysis and fingerprinting for biomedica characterization. *J. Biotechnol.* **113**, 211–230.
- Savitzky, A. & Golay, M. J. E. 1964 Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **36**, 1627–1635.
- Stephens, A. B. & Walker, P. N. 2002 Near-infrared spectroscopy as a tool for real-time determination of BOD₅ for single-source samples. *Trans. ASAE* **45**, 451–458.
- Uddin, M., Okazaki, E., Ahmad, M. U., Fukuda, Y. & Tanaka, M. 2006 NIR spectroscopy: a non-destructive fast technique to verify heat treatment of fish-meat gel. *Food Control* **17**, 660–664.