



## Characterization of activated sludge abnormalities by image analysis and chemometric techniques

D.P. Mesquita<sup>a</sup>, A.L. Amaral<sup>a,b</sup>, E.C. Ferreira<sup>a,\*</sup>

<sup>a</sup> IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

<sup>b</sup> Instituto Superior de Engenharia de Coimbra, Instituto Politécnico de Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

### ARTICLE INFO

#### Article history:

Received 31 December 2010

Received in revised form 29 April 2011

Accepted 31 May 2011

Available online 12 June 2011

#### Keywords:

Activated sludge

Image analysis

Morphology

Physiology

Chemometric techniques

### ABSTRACT

This work focuses on the use of chemometric techniques for identifying activated sludge process abnormalities. Chemometric methods combined with image analysis can improve activated sludge systems monitoring and minimize the need of analytical measurements. For that purpose data was collected from aggregated and filamentous biomass, biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria, and operational parameters. Principal component analysis (PCA) was subsequently applied to identify activated sludge abnormalities, allowing the identification of several disturbances, namely filamentous bulking, pinpoint flocs formation, and zoogal bulking as well as normal conditions by grouping the collected samples in corresponding clusters.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Activated sludge systems are very sensitive to unexpected changes in operating conditions, mainly in terms of organic load, nutrients and oxygen contents. The modification of the operating conditions mostly affects the solid–liquid separation, a stage that is often characterized by the sludge volume index (SVI). Problems related to the sludge settling ability were previously reported encompassing pinpoint flocs formation, filamentous bulking, dispersed growth, and zoogal or viscous bulking as the most common phenomena in activated sludge systems. Based on microscopy inspection, the biomass structure in activated sludge systems [1,2] has been shown to be related to the SVI, especially on filamentous bulking surveys [3–8].

Usually, pinpoint flocs phenomenon is related to the formation of small, weak and roughly spherical flocs, with low settling abilities. Consequently, pinpoint flocs occurrence originates a turbid and hardly settleable effluent presenting high suspended solids contents. Zoogal (or viscous) bulking occurs when floc-forming bacteria are present in large amounts producing large contents of extracellular material (namely exopolysaccharides) lowering the sludge settling properties. However, the most common sludge bulking problem reported in activated sludge systems is the filamentous bulking caused by an excessive growth of different types

of filamentous microorganisms [9–11]. The massive presence of filamentous microorganisms forms a large and open network, by linking the flocs, thus interfering with sludge compaction, settling and thickening. Several factors emerge in determining the community composition of filamentous bacteria including the operating conditions and the nature of the influent being treated. Consequently, the success of a given activated sludge system depends on the ecosystem balance among floc-forming bacteria, such as *Pseudomonas* spp., *Zoogal* spp., *Alcaligenes* spp., and *Achromobacter* spp., and filamentous bacteria, such as *Nocardia* spp., *Rhodococcus* spp., Type 1863 and *Microthrix* spp., where control strategies must be performed [12,13].

Image processing and analysis methodologies have been increasingly used for activated sludge characterization. Several studies [14–16] used automated image analysis to relate the sludge structure in biological systems with the sludge settling ability, since this characteristic is strongly influenced by floc structure and filamentous bacteria contents [17]. Recently, image analysis has been the basis of the assessment of biomass morphological changes using different types of microscopy techniques [5–7,18]. Furthermore, the combination of settling properties and image analysis data, such as the aggregated and filamentous bacteria contents and morphology (morphological data), and Gram-positive/Gram-negative and viable/damaged bacteria ratios (physiological data), may offer powerful information, as well as a mechanism to decide if a given intervention should be performed into the system.

Monitoring of an activated sludge system, may represent the determination of huge amounts of data, regarding the system biotic

\* Corresponding author. Tel.: +351 253604402; fax: +351 253678986.  
E-mail address: [ecferreira@deb.uminho.pt](mailto:ecferreira@deb.uminho.pt) (E.C. Ferreira).

and abiotic characterization. To make sense of such complexity, chemometric techniques are, quite often, indispensable. In fact, due to the large data contents provided by image analysis techniques, chemometric techniques are quite valuable to organize such information in order to better characterize the activated sludge system. As matter of fact, principal component analysis (PCA) and partial least squares (PLSs) have already proven to be important tools for the identification of operating conditions and the prediction of certain parameters [19–23].

PCA is a tool used for data compression and information extraction, finding combinations of variables describing major trends in the data [24]. The aim of PCA is to project the high dimensional space into a more visual low dimensional one, and by doing this find the key variables of the system. Thus, there will be a number of new principal components, which describe most of the variance in the process in a space of fewer dimensions than the original [25]. Recently, PCA techniques have been successfully applied for monitoring a wide range of wastewater treatment systems [26–29].

In the present work, a large amount of data, reflecting the operation of an activated sludge system in four different experimental conditions (pinpoint flocs formation, filamentous bulking, zoogeal bulking and normal conditions), was treated using chemometric methods. PCA was applied integrating image analysis parameters from aggregated and filamentous biomass characterization, biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria, alongside operational parameters. The potential of PCA was examined for the identification of each of the lab-scale activated sludge conditions.

## 2. Materials and methods

### 2.1. Activated sludge system description

Experimental data was obtained from a lab-scale activated sludge system based on a 17L aerated tank with suspended biomass, followed by a 2.5L cylindrical clarifier. The aerated tank was inoculated with activated sludge from a domestic wastewater treatment plant. The system was fed with a synthetic medium prepared with the following composition ( $\text{mg L}^{-1}$ ):  $\text{NaCH}_3\text{COO}\cdot 3\text{H}_2\text{O}$ , 2073;  $(\text{NH}_4)_2\text{SO}_4$ , 140;  $\text{MgSO}_4\cdot 7\text{H}_2\text{O}$ , 25;  $\text{KH}_2\text{PO}_4$ , 44;  $\text{K}_2\text{HPO}_4\cdot 2\text{H}_2\text{O}$ , 59;  $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ , 30;  $\text{FeCl}_3\cdot 6\text{H}_2\text{O}$ , 18.8;  $\text{NaHCO}_3$ , 105. A micronutrients solution was also added to the system to guarantee the biomass maintenance. The pH of the system was controlled with a pH meter and a control pump (Model BL 7916-BL 7917, Hanna Instruments, Woonsocket, RI, USA) with 0.01 M HCl solution. Complete mixing inside the reactor was guaranteed by supplying a continuous inflow of air bubbles through an air diffuser placed at the bottom of the reactor. An oxygen probe (TriOmatic 690, WTW, Weilheim, Germany) was used to measure the amount of dissolved oxygen. The concentration of dissolved oxygen was maintained around  $7 \text{ mg L}^{-1}$ . Sludge recirculation from the settler to the reactor was performed.

### 2.2. Off-line process monitoring

The system was regularly monitored for parameters as total suspended solids (TSSs), sludge volume index (SVI), chemical oxygen demand (COD), ammonium ( $\text{N-NH}_4^+$ ), nitrite ( $\text{N-NO}_2^-$ ) and nitrate ( $\text{N-NO}_3^-$ ) concentrations. The TSS measurements were conducted in accordance with the procedures described in Standard Methods [30]. The SVI was determined in a 1 L Imhoff cone, with the sludge height variation monitored for 30 min. Samples for COD, ammonium, nitrite and nitrates quantification were collected and filtered, from the feed and the reactor. COD was measured with Hach Lange COD cell tests (LCK 414 and LCK 514) on a spectrophotometer

(Hach Lange DR 5000). Nitrite was determined with Griess–Hosvay method, similar to the colorimetric method from Standard Methods [30]. Ammonium was determined according to Nessler's method [30]. Nitrate was determined by HPLC (Jasco, Tokyo, Japan) with automatic injection, UV detector (210 nm), and a Varian (Palo Alto, CA, USA) Metacarb 67H column operating at room temperature of  $60^\circ\text{C}$ . The eluent was a solution of sulfuric acid ( $0.005 \text{ mol L}^{-1}$ ) with a flow rate of  $0.7 \text{ mL min}^{-1}$  and a pressure between 70 and  $80 \text{ kg cm}^{-2}$ .

### 2.3. Staining procedures

For all staining procedures, a sterile solution of 0.85% NaCl was prepared.  $1.5 \mu\text{L}$  of each dye was put into 5 mL of the NaCl solution and the tube was wrapped with aluminum foil (staining solution). A volume of  $100 \mu\text{L}$  of undiluted biomass suspension was mixed with  $50 \mu\text{L}$  of staining solution and incubated in darkness for 15 min at room temperature. A preliminary experiment allowed concluding that this concentration was sufficient for staining the overall bacteria population. The bacteria population was then visualized through fluorescence microscopy.

The Live/Dead<sup>®</sup> BacLight<sup>™</sup> bacterial viability kit was used to differentiate viable and damaged bacteria [31]. The kit utilizes a mixture of SYTO<sup>®</sup> 9 green-fluorescent nucleic acid stain and a red-fluorescent nucleic acid stain, propidium iodide (PI). Viable bacteria are stained by SYTO<sup>®</sup> 9 and damaged bacteria are stained by PI.

The Live BacLight<sup>™</sup> bacterial Gram stain kit allows to easily classify bacteria as Gram-positive or Gram-negative without fixatives [32]. This kit utilizes a mixture of SYTO<sup>®</sup> 9 green-fluorescent nucleic acid stain and a red-fluorescent nucleic acid stain, hexidium iodide (HI). Gram-negative bacteria are stained by SYTO<sup>®</sup> 9 and Gram-positive bacteria are stained by HI.

### 2.4. Bright field image acquisition

The microbial community structure was observed by means of an Olympus BX51 optical microscope (Olympus, Tokyo, Japan), at  $100\times$  magnification, coupled with an Olympus DP25 camera (Olympus, Tokyo, Japan). Images were acquired at  $1360 \times 1024$  pixels and 8-bit format through the commercial software CellB (Olympus, Tokyo, Japan). Samples were taken from the aerated tank, and 3 slides per sample were used resulting in a total of 150 images ( $3 \times 50$  images/slide). A recalibrated micropipette with a sectioned tip at the end, with a large enough diameter to allow larger aggregates to flow, was used to deposit samples on the slides. For each slide, a volume of  $10 \mu\text{L}$  was covered with a  $20 \text{ mm} \times 20 \text{ mm}$  cover slip, for visualization and image acquisition. Images were acquired in the upper, middle and bottom of the slide in order to improve the representativeness of the microbial community in the system.

### 2.5. Fluorescence image acquisition

Slides with stained sludge samples ( $10 \mu\text{L}$  on each slide) were observed in an Olympus BX51 epifluorescence microscope (Olympus, Tokyo, Japan) at  $200\times$  magnification. Two filters were used, the first in the green wavelength range with an excitation bandpass of 470–490 nm and emission at 516 nm (long pass filter), and the second filter in the red wavelength range with an excitation bandpass of 530–550 nm and emission at 591 nm (long pass filter). Images were acquired at  $1360 \times 1024$  pixels, and 24-bit RGB format (8 bit red, 8 bit green and 8 bit blue channels) through the commercial software CellB (Olympus, Tokyo, Japan), and 2 slides per sample were used resulting in a total of 100 images ( $2 \times 50$  images/slide).



**Fig. 1.** Bright field image processing: (a) original image, (b) aggregates binary image and (c) filaments binary image.

## 2.6. Bright field image processing

The image analysis program used for the recognition of the aggregated and filamentous biomass in the grayscale images was adapted from a previous version of Amaral [33], developed in Matlab 7.3 (The Mathworks, Inc., Natick, USA). Fig. 1 shows an original grayscale image, and the final binary images of microbial aggregates and filamentous bacteria.

## 2.7. Fluorescence image processing

A new program was developed in Matlab 7.3 (The Mathworks, Inc., Natick, USA) to recognize and characterize the aggregated and filamentous biomass in the fluorescent images, comprising the extraction of the green channel from the original RGB image collected at 516 nm (green filter) and of the red channel from the original RGB image collected at 591 nm (red filter). For each of the above channels the program performed a background correction, image segmentation and aggregated and filamentous biomass recognition. The determination of the SYTO<sup>®</sup> 9-stained cells was obtained through the use of the green channel of the RGB image collected at 516 nm, whereas the determination of the propidium iodide and the hexidium iodide-stained bacteria was performed through the use of the red channel of the RGB image collected at 591 nm. The developed program also allowed for the calculation of a fluorescence-based intensity image, for both green and red channels, directly correlated with the fluorescence of the aggregated and filamentous biomass. These images were subsequently used, regarding the filamentous biomass, to accurately identify the Gram and viability status of overlapping SYTO<sup>®</sup> 9 and propidium or hexidium iodide stained cells. Fig. 2 presents the original, fluorescence based intensity and recognized aggregates and

filaments images in both the green filter (a–c) and in the red filter (d–f).

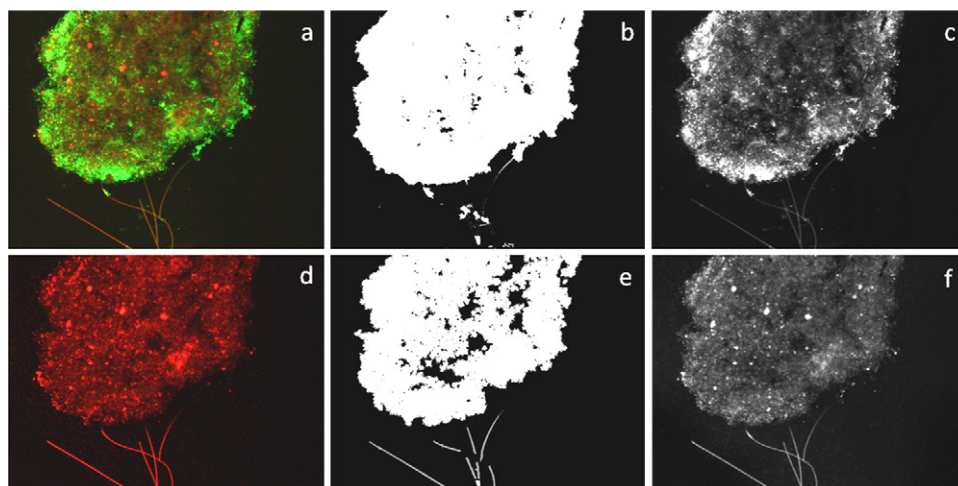
## 2.8. Morphological parameters determination

The aggregated and filamentous biomass contents and morphology were assessed by bright field microscopy, and biomass composition on Gram-positive/Gram-negative bacteria and viable/damaged bacteria by fluorescence microscopy. Table 1 presents the overall dataset of parameters used in this work. The morphological parameters definition can be found in Amaral [33] and Mesquita et al. [34]. The Area ratio (RelArea) is given as the ratio of the area of the holes in an aggregate and the area of the aggregate.

The information for the physiological parameters was obtained from the green (Gram-negative and viable bacteria) and red (Gram-positive and damaged bacteria) channels of the fluorescence images Gram negative/positive and viable/damaged aggregates and filaments were characterized both in terms of projected area and in mean intensity values. The ratios between gram positive/negative and viable/damaged bacteria, both in terms of projected area and mean intensity values, were next determined.

## 2.9. Principal component analysis (PCA)

Before PCA decomposition, variables were mean centered (adjusted to zero mean by subtracting the original mean of each column). Further, variables were normalized (adjusted to unit variance by dividing each column by its standard deviation) to give them equal influence on the model, since variables are expressed in different units and display substantially different absolute value ranges [35].



**Fig. 2.** Images from the green filter: (a) original image, (b) area detection image, (c) intensity image; and from the red filter: (d) original image, (e) area detection image and (f) intensity image.

**Table 1**  
Operational parameters, morphological parameters, and fluorescent parameters included in the dataset.

Variable	Name
<i>Operational parameters</i>	
COD <sup>a</sup>	Chemical oxygen demand
N–NO <sub>3</sub> <sup>–a</sup>	Nitrate
N–NH <sub>4</sub> <sup>+a</sup>	Ammonium
N–NO <sub>2</sub> <sup>–a</sup>	Nitrite
OLR	Organic loading rate
TSS	Total suspended solids
SVI	Sludge volume index
<i>Morphological parameters (from bright field microscopy)</i>	
Deq <sup>b</sup>	Equivalent diameter
P <sup>b</sup>	Perimeter
L <sup>b</sup>	Length
W <sup>b</sup>	Width
Ff <sup>b</sup>	Form factor
Conv <sup>b</sup>	Convexity
Comp <sup>b</sup>	Compactness
Round <sup>b</sup>	Roundness
Solid <sup>b</sup>	Solidity
Ext <sup>b</sup>	Extent
Ecc <sup>b</sup>	Eccentricity
Rob <sup>b</sup>	Robustness
LrgC <sup>b</sup>	Largest concavity
RelArea	Ratio between hole and object area
%Nb <sup>b</sup>	Aggregates number percentage
%Area <sup>b</sup>	Aggregates area percentage
Nb/Vol <sup>b</sup>	Aggregates number per volume
TA/Vol	Total aggregate area per volume
TL/Vol	Total filament length per volume
TL/TA	Total filament length per total aggregate area
TL/TSS	Total filament length per total suspended solids
<i>Physiological parameters (from epifluorescence microscopy)</i>	
G.AA.R; G.AA.G	Gram-positive aggregated bacteria area; Gram-negative aggregated bacteria area
G.INA.R; G.INA.G	Gram-positive aggregated bacteria intensity; Gram-negative aggregated bacteria intensity
G.AF.R; G.AF.G	Gram-positive filaments area; Gram-negative filaments area
G.INF.R; G.INF.G	Gram-positive filaments intensity; Gram-negative filaments intensity
LD.AA.R; LD.AA.G	Damaged aggregated bacteria area; viable aggregated bacteria area
LD.INA.R; LD.INA.G	Damaged aggregated bacteria intensity; viable aggregated bacteria intensity
LD.AF.R; GR.AF.G	Damaged filaments area; viable filaments area
LD.INF.R; LD.INF.G	Damaged filaments intensity; viable filaments intensity
G.AA.G/AA.R	Ratio between Gram-negative and Gram-positive aggregates area
G.INA.G/INA.R	Ratio between Gram-negative and Gram-positive aggregates intensity
G.AF.G/AF.R	Ratio between Gram-negative and Gram-positive filaments area
G.INF.G/INF.R	Ratio between Gram-negative and Gram-positive filaments intensity
LD.AA.G/AA.R	Ratio between viable and damaged aggregates area
LD.INA.G/INA.R	Ratio between viable and damaged aggregates intensity
LD.AF.G/AF.R	Ratio between viable and damaged filaments area
LD.INF.G/INF.R	Ratio between viable and damaged filaments intensity

<sup>a</sup> Determined inside the reactor.

<sup>b</sup> Determined for small (Deq < 25 μm), intermediate (25 μm < Deq < 250 μm) and large (Deq > 250 μm) aggregates.

PCA decomposes the data matrix **X** as the sum of the outer product of **T** (containing the scores) and **P** (containing the loadings) plus a residual matrix **E**:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (1)$$

Each of the principal components (PCs) captures as much as possible the variation which has not been explained by the former PCs, i.e., the first PC maximizes the covariance in the original data and the subsequent PCs maximize the covariance in the residual matrix

**Table 2**  
Percentage variance captured for each PC in the PCA model of each dataset.

PC	Eigenvalues	Variance (%)	Cumulative (%)
1st dataset – morphological parameters			
1	15.72	28.58	28.58
2	14.54	26.43	55.01
3	7.17	13.03	68.04
4	4.63	8.42	76.46
2nd dataset – morphological + physiological parameters			
1	18.33	23.20	23.2
2	17.62	22.30	45.5
3	8.56	10.84	56.34
4	6.06	7.68	64.02
5	4.07	5.15	69.17
6	2.73	3.45	72.61
7	2.14	2.71	75.34
8	1.87	2.37	77.70
3rd dataset – morphological + physiological + operational parameters			
1	20.89	24.01	24.01
2	18.23	20.95	44.96
3	9.36	10.76	55.72
4	6.76	7.77	63.49
5	4.15	4.77	68.27
6	2.89	3.33	71.59
7	2.52	2.89	74.49
8	1.98	2.27	76.77

ces which are left after extracting the former PCs. The dimensional reduction is based on the fact that the principal components are orthogonal and hence uncorrelated. It should be kept in mind that the principal components are linear combinations of the original variables. Therefore, they are abstract variables which are used to visualize latent structures and latent phenomena in the data. In this way, the original data is projected into a new coordinate system in which the objects are described by the scores and the variables by the loadings [36].

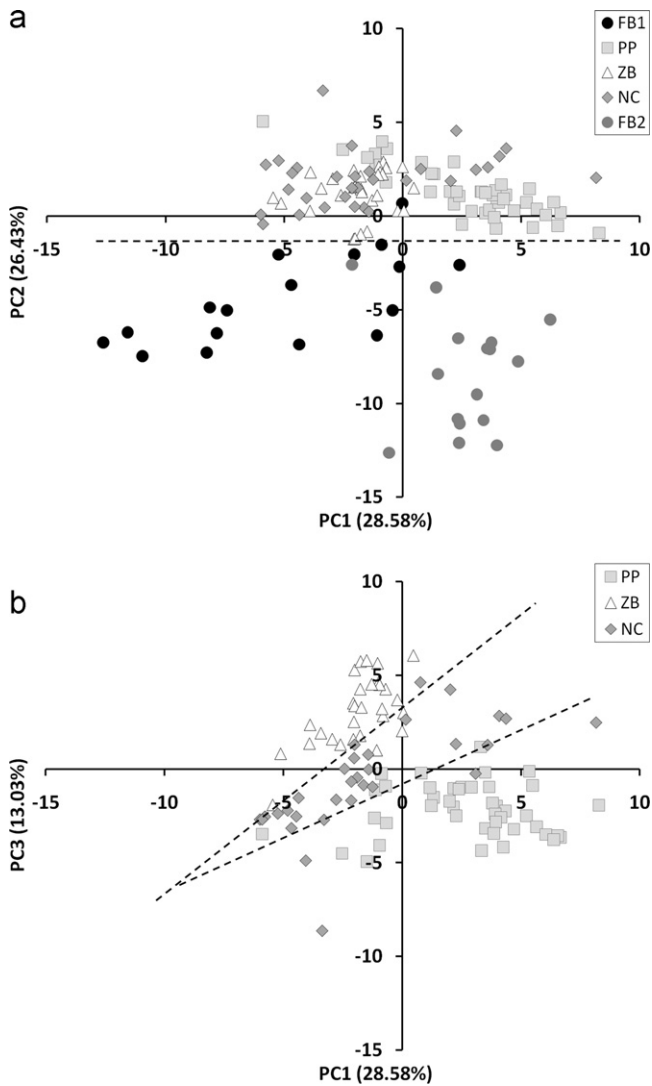
In this study, Matlab™ 7.3 (The Mathworks, Natick, MA) was used to perform PCA.

### 3. Results and discussion

A principal components analysis (PCA) was applied to the normalized data of the operation of the lab-scale activated sludge system. Experiments were conducted within the activated sludge system in order to observe four different operational conditions (filamentous bulking, zooglear bulking, pinpoint flocs formation and normal conditions). Furthermore, two distinct filamentous conditions were studied, the first one identified as FB1, and the second as FB2. The PCA study was performed with three different datasets. The first dataset used solely morphological parameters obtained from bright field microscopy, the second dataset combined morphological and physiological parameters and the third dataset combined all the information obtained from image analysis alongside the operational parameters. Table 2 presents the percentage of variance captured by each principal component (PC) as a function of the number of PCs. 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. In the present study the number of components was determined by the eigenvalue-one criterion (analysis of the eigenvalues of the covariance matrix), according to which only the PC's with eigenvalues greater than one are considered relevant [27].

The number of PCs presenting eigenvalues larger than 1 ascended to four (explaining 76.5% of the total variance) in the first dataset, and to eight in both the second and third datasets, corresponding to total variances of respectively 77.8% and 76.8%.

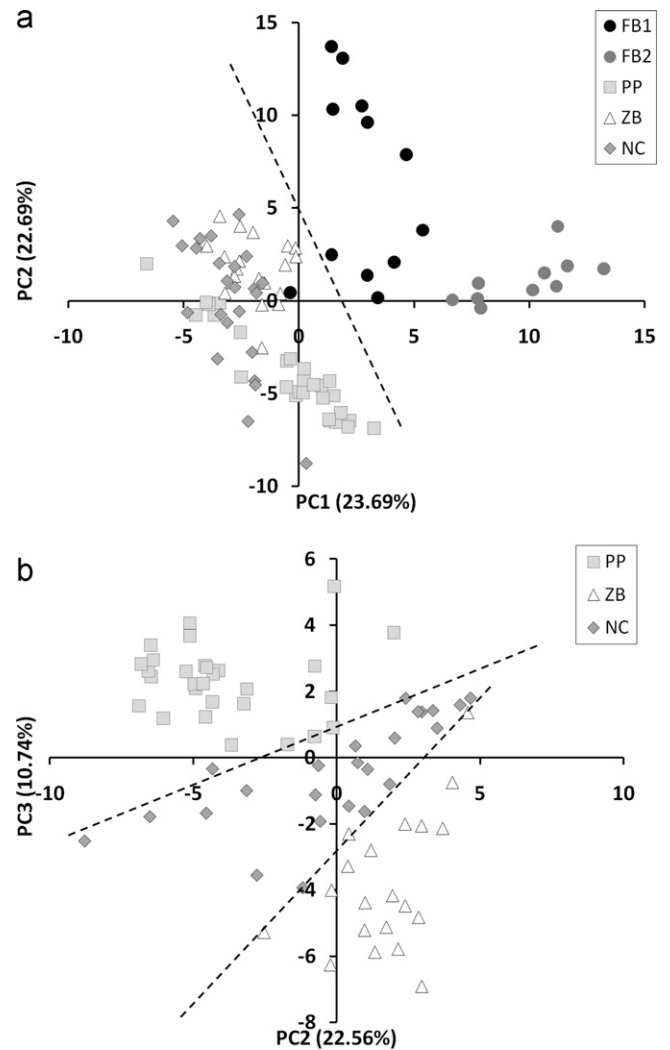
A more explicit way to analyze and detect changes or disturbances in the measured data is through a score plot analysis. A score plot is any pair of score vectors plotted against each other,



**Fig. 3.** PCA score plot of the PC1 vs. PC2 (a), PC1 vs. PC3 (b) for the recognition of all phenomena for the 1st dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zooglear bulking; NC – normal conditions.

in which each sample appears as a data point and closely interrelated samples appear clustered together [27]. For clarity purposes, only the most significant pair of each of the first three PCs score plot will be presented. Cluster zones for each condition were further determined for each significant PC pair and each condition was evaluated according to the percentage of samples correctly identified in the corresponding cluster. The determination of the value of each sample point in the new PC1, PC2 and PC3 variables was performed, followed by a linear transformation with the PCA loadings. Parameters averages, standard deviations (for normalization) and loadings are presented as supplementary material (Table S1), allowing the determination of each sample in the new PC spaces.

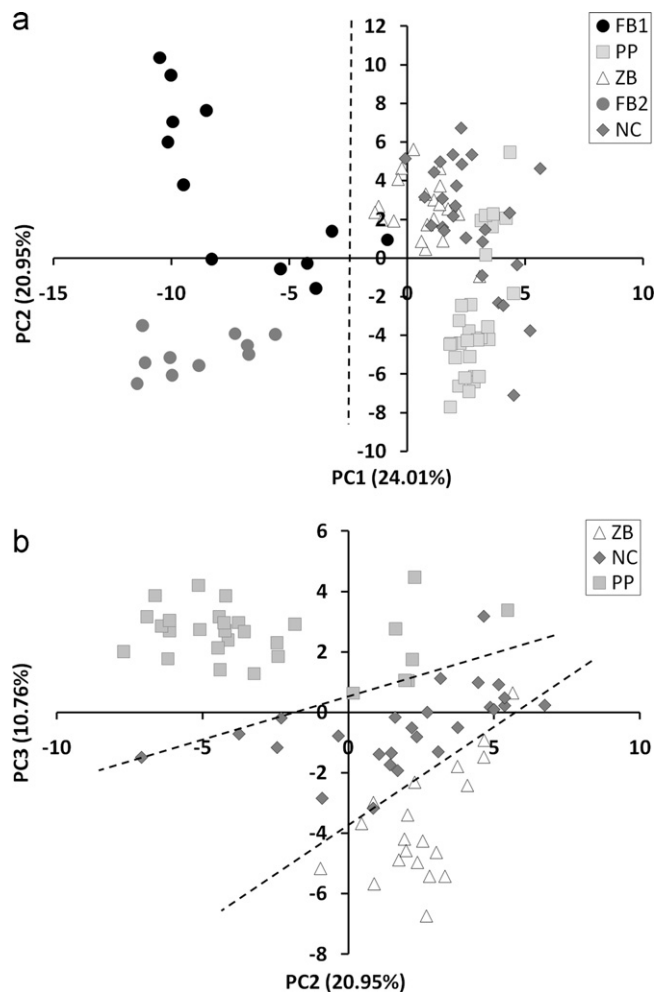
Fig. 3 shows the first three PCs capturing 68% of the total variance using the first dataset. PC1 vs. PC2 score plot (Fig. 3a) evidenced two filamentous bulking clusters, clearly distinct with respect to the remaining studied conditions. A threshold of  $-1.40$  on PC2 was established to separate filamentous bulking conditions (below the threshold limit) from the other disturbances (above the threshold limit). It was found that 97% (32 out of 33 samples) of the dataset was properly identified in the filamentous bulking cluster. All the remaining conditions were able to be separated in three different clusters, although only barely, by the use of the PC1 and PC3 score



**Fig. 4.** PCA score plot of the PC1 vs. PC2 (a), PC2 vs. PC3 (b) for the recognition of all phenomena for the 2nd dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zooglear bulking; NC – normal conditions.

plots (Fig. 3b). Three cluster zones were defined separating zooglear bulking, normal conditions and pinpoint flocs. The cluster zone representing zooglear bulking was defined by  $PC3 > 0.92 \times PC1 + 3.30$ , and a total of 25 out of 29 samples (86.2%) was correctly identified as zooglear bulking. The cluster zone representing pinpoint flocs formation was defined by  $PC3 < 0.69 \times PC1 - 0.90$ , and a total of 39 out of 44 samples (88.6%) was correctly identified as pinpoint flocs. Finally, the cluster zone representing normal conditions was defined as  $0.92 \times PC1 + 3.30 > PC3 > 0.69 \times PC1 - 0.90$ , with a total of 26 out of 31 samples (83.9%) being correctly identified as normal conditions. An estimated time of 1.5 h was required for sample preparation, image acquisition, and image processing regarding the first dataset.

The use of the second dataset representing 56.7% of the total variance for the first three PCs, is shown in Fig. 4. Although the total variance, explained by these PCs was smaller than the obtained by the use the first dataset, this can be explained by the larger original dataset. However, only a slight improvement regarding the individualization of the two filamentous bulking clusters by PC1 vs. PC2 score plot (Fig. 4a), as well as between the remaining conditions by PC2 and PC3 (Fig. 4b), could be attained. PC1 vs. PC2 score plot allowed to separate the filamentous bulking conditions from the remaining by defining a  $PC2 > -2.50 \times PC1 + 5.00$  filamentous bulking zone. A total of 21 out of 22 samples (95.5%) was correctly



**Fig. 5.** PCA score plot of the PC1 vs. PC2 (a), PC2 vs. PC3 (b) for the recognition of all phenomena for the 3rd dataset. FB1 – filamentous bulking; FB2 – filamentous bulking 2; PP – pinpoint flocs; ZB – zooglear bulking; NC – normal conditions.

identified as filamentous bulking. Regarding the PC2 vs. PC3 score plot three cluster zones were defined separating zooglear bulking, normal conditions and pinpoint flocs. The cluster zone representing zooglear bulking was defined by  $PC3 < 0.93 \times PC2 - 2.80$ , and a total of 18 out of 19 samples (94.7%) was correctly identified as zooglear bulking. The cluster zone representing pinpoint flocs formation was defined by  $PC3 > 0.36 \times PC2 + 1.00$ , and all of the 29 samples (100%) was correctly identified as pinpoint flocs. Finally, the cluster zone representing normal conditions was defined as  $0.36 \times PC2 + 1.00 > PC3 > 0.93 \times PC2 - 2.80$ , with a total of 23 out of 24 samples (95.8%) being correctly identified as normal conditions. An estimated time of 2.5 h was required for sample preparation, staining procedure, image acquisition and image processing regarding the second dataset.

In Fig. 5, corresponding to the third dataset, although the first three PCs (PC1, PC2 and PC3) only explained 55% of the total variance, again due to a larger original dataset, they allowed to further distinguish all studied conditions from each other. In fact, the presented PC1, PC2 and PC3 score plots reflected the five main clusters (including the two different filamentous bulking clusters) found in the data set. Once again, PC1 and PC2 allowed separating the filamentous bulking conditions from the rest, whereas the remaining conditions were able to be separated in different clusters by the use of the PC2 vs. PC3 score plot (Fig. 5b). A threshold of  $-2.50$  on PC1 was established to separate filamentous bulking conditions (below the threshold limit) from the

**Table 3**

Variables presenting larger principal component coefficients in the 1st dataset PCA.

Variables	PC1	Variables	PC2	Variables	PC3
Sol.int	0.238	% Area.int	0.223	Deq.int	0.301
Round.int	0.236	% Area.s	0.223	L.int	0.298
LrgC.int	0.233	W.s	0.208	W.int	0.291
Ext.int	0.229	% Nb.s	0.204	% Area.lg	0.268
Comp.int	0.227	Nb/Vol.s	0.203	P.int	0.260
Rob.int	0.222	% Nb.int	0.202	L.s	0.250
FF.int	0.220	FF.s	0.191	P.s	0.233
FF.lg	0.200	Deq.s	0.191	% Nb.lg	0.216
RelArea.lg	0.196	Comp.s	0.190	Nb/Vol.lg	0.205
Ecc.int	0.194	Round.s	0.185	Deq.s	0.193
Per.lg	0.191	Rob.lg	0.184	Conv.s	0.183

s – small aggregates (Deq < 25  $\mu$ m); int – intermediate aggregates (25  $\mu$ m < Deq < 250  $\mu$ m); lg – large aggregates (Deq > 250  $\mu$ m).

other disturbances (above the threshold limit). It was found that 95.8% (23 out of 24 samples) of the dataset was properly identified in the filamentous bulking cluster. PC1 vs. PC2 score plot allowed to separate the filamentous bulking conditions from the remaining by defining a  $PC2 > -2.50 \times PC1 + 5.00$  filamentous bulking zone. A total of 21 out of 22 samples (95.5%) was correctly identified as filamentous bulking. Regarding the PC2 vs. PC3 score plot three cluster zones were defined separating zooglear bulking, normal conditions and pinpoint flocs. The cluster zone representing zooglear bulking was defined by  $PC3 < 0.25 \times PC2 + 0.50$ , and a total of 18 out of 20 samples (90.0%) was correctly identified as zooglear bulking. The cluster zone representing pinpoint flocs formation was defined by  $PC3 > 0.67 \times PC2 - 3.80$ , and all of the 29 samples (100%) was correctly identified as pinpoint flocs. Finally, the cluster zone representing normal conditions was defined  $0.67 \times PC2 - 3.80 > PC3 > 0.25 \times PC2 + 0.50$ , with a total of 22 out of 24 samples (91.7%) being correctly identified as normal conditions. An estimated time of 12 h was required for sample preparation, staining, image acquisition, image processing and physical and chemical parameters determination, with TSS as the limiting factor.

Tables 3–5 show the variables presenting the larger principal component coefficients of each PC derived by PCA, regarding the first, second and third datasets, respectively.

The use of the first dataset (Table 3) revealed that PC2, the main component responsible for the separation of the filamentous bulking clusters from the remaining, was strongly influenced by the aggregated biomass size distribution and small aggregates characterization. PC1 and PC3, responsible for the separation of the remaining clusters, were strongly affected by the intermediate aggregates morphology and size characterization, respectively.

Analyzing Table 4, it was found that PC1, the main component responsible for the separation of the filamentous bulking clusters from the remaining, was strongly affected by the morphology of

**Table 4**

Variables presenting larger principal component coefficients in the 2nd dataset PCA.

Variables	PC1	Variables	PC2	Variables	PC3
Comp.s	0.200	Ext.int	0.225	Deq.int	0.276
% Area.s	0.198	Rob.int	0.223	W.int	0.258
% Area.int	0.198	FF.int	0.215	L.int	0.255
Round.s	0.198	Sol.int	0.215	% Area.lg	0.252
TL/Vol	0.197	LrgC.int	0.215	P.int	0.211
TL/TSS	0.196	FF.lg	0.204	% Nb.lg	0.207
FF.s	0.195	Conv.lg	0.203	L.s	0.207
W.s	0.186	P.lg	0.203	LD_AA.R	0.199
Ecc.s	0.181	Round.int	0.197	TA/Vol	0.195
LD_AF.G	0.176	RelArea.lg	0.197	Nb/Vol.lg	0.194
Ext.s	0.176	Comp.int	0.186	Per.s	0.190

s – small aggregates (Deq < 25  $\mu$ m); int – intermediate aggregates (25  $\mu$ m < Deq < 250  $\mu$ m); lg – large aggregates (Deq > 250  $\mu$ m).

**Table 5**  
Variables presenting larger principal component coefficients in the 3rd dataset PCA.

Variables	PC1	Variables	PC2	Variables	PC3
TL/Vol	0.200	Rob.int	0.210	Deq.int	0.247
TL/TSS	0.199	Ext.int	0.210	W.int	0.228
SVI	0.188	Conv.lg	0.208	LD.AF.R	0.228
Comp.s	0.185	P.lg	0.203	L.int	0.228
Round.s	0.184	FF.lg	0.202	% Area.lg	0.226
LD.AF.G	0.177	FF.int	0.199	TA/Vol	0.226
FF.s	0.176	Sol.lg	0.192	N-NO <sub>3</sub> <sup>-</sup> .r	0.203
% Area.int	0.175	Conv.int	0.192	% Nb.lg	0.189
% Area.s	0.174	Rob.lg	0.190	G.INA.G	0.184
Ecc.s	0.173	LrgC.int	0.190	P.int	0.184
ln(TL/TA)	0.173	RelArea.lg	0.188	Nb/Vol.lg	0.181

s – small aggregates (Deq < 25 μm); int – intermediate aggregates (25 μm < Deq < 250 μm); lg – large aggregates (Deq > 250 μm); r – measurement inside the reactor.

the small aggregates, by the aggregated biomass size distribution, and by the filamentous bacteria contents. PC2 and PC3, responsible for the separation of the remaining clusters, were strongly influenced by the morphology of intermediate aggregates, and by their size characterization and aggregated biomass size distribution, respectively. No significant individual contribution from the physiological parameters was noticed in the first three principal components, apart from a small contribution of the viability of aggregated and filamentous bacteria. However, an overall improvement on the separation of the conditions clusters was observed.

Considering the third dataset, Table 5 shows the variables presenting the larger principal component coefficients of each PC derived by PCA. It was found that PC1 was strongly affected by the filamentous bacteria contents and sludge settling ability (SVI), contributing for a clear distinction of the filamentous bulking phenomena. PC2 was strongly influenced by the morphology of large and intermediate aggregates, and PC3 by the intermediate aggregates size characterization. PC2, combined with PC3, was responsible for the distinction of the zooglear bulking conditions from the normal operating conditions, reflecting the importance of the intermediate and large aggregates morphology and size characterization for this purpose. PC3 was responsible mainly for the distinction of the pinpoint flocs conditions from the remaining conditions, emphasizing the importance of the intermediate aggregates size characterization for this purpose. Again, no significant individual contribution from the operational parameters was noticed in the first three principal components, apart from the sludge volume index and the nitrate concentration on the reactor. Furthermore, no significant improvement on the separation of the conditions clusters, regarding the second dataset, was observed.

Looking back at the obtained results, the performed PCA allowed identifying the different studied conditions. It was also found that the introduction of physiological parameters (regarding cell viability and Gram status) to the morphological parameters led to an improvement on the identification of the studied conditions by PCA, whereas the operational parameters seemed to contribute less to this purpose.

#### 4. Conclusions

This work evaluated the ability of image analysis information to monitor an activated sludge system. The performed PCA allowed the identification of filamentous bulking, pinpoint flocs, zooglear bulking and normal conditions from the use of morphological, physiological and operational datasets. Furthermore, this study demonstrated the usefulness of this methodology as a fast (between 1.5 and 2.5 h), simple (particularly bright field

analysis) and effective (recognition percentages above 84%) method for detecting deviating conditions in activated sludge, from the moment they start to occur. Therefore this methodology can be considered quite promising in full activated sludge wastewater treatment systems monitoring to evaluate the activated sludge status, and act accordingly in order to correct deviating conditions in real time.

#### Acknowledgements

The authors acknowledge the financial support to D.P. Mesquita through the grant SFRH/BD/32329/2006 and the project PTDC/EBB-EBI/103147/2008 provided by Fundação para a Ciência e a Tecnologia (Portugal).

#### Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.aca.2011.05.050](https://doi.org/10.1016/j.aca.2011.05.050).

#### References

- [1] J.C. Palm, D. Jenkins, D.S. Parker, *J. Water Pollut. Control Fed.* 52 (10) (1980) 2484.
- [2] C. Dagot, M.N. Pons, M. Casellas, G. Guibaud, P. Dollet, M. Baudu, *Water Sci. Technol.* 43 (3) (2001) 27.
- [3] S.E. Lee, B. Koopman, H. Bode, D. Jenkins, *Water Res.* 17 (10) (1983) 1421.
- [4] R. Jenné, E.N. Banadda, I.Y. Smets, J.F. Van Impe, *Water Sci. Technol.* 50 (7) (2004) 281.
- [5] R. Jenné, E.N. Banadda, I.Y. Smets, J. Deurincq, J.F. Van Impe, *Microsc. Microanal.* 13 (2007) 36.
- [6] A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta* 544 (2005) 246.
- [7] M. da Motta, M.N. Pons, N. Roche, *Water Sci. Technol.* 43 (7) (2001) 91.
- [8] M. da Motta, A.L. Amaral, M. Casellas, M.N. Pons, C. Dagot, N. Roche, E.C. Ferreira, H. Vivier, *IFAC Computer Applications in Biotechnology*, 2001, p. 427.
- [9] D.H. Eikelboom, *Process Control of Activated Sludge Plants by Microscopic Investigation*, 1st edition, IWA Publishing, London, 2000.
- [10] D. Jenkins, M.G. Richard, G. Daigger, *Manual on the Causes and Control of Activated Sludge Bulking, Foaming and Other Solids Separation Problems*, Lewis Publishing, Boca Raton, FL, 2003.
- [11] A.J. Schuler, D. Jassby, *Water Res.* 41 (2007) 4349.
- [12] A.M.P. Martins, K. Pagilla, J.J. Heijnen, M.C.M. van Loosdrecht, *Water Res.* 38 (2004) 793.
- [13] S. Rossetti, M.C. Tomei, P.H. Nielson, V. Tandoi, *FEMS Microbiol. Rev.* 29 (2005) 49.
- [14] S. Matsui, R. Yamamoto, *Water Sci. Technol.* 16 (1984) 69.
- [15] J.J. Ganczarzyk, *Water Sci. Technol.* 30 (1994) 87.
- [16] K. Grijspeerdt, W. Verstraete, *Water Res.* 31 (1997) 1126.
- [17] M. Sezgin, *Water Res.* 16 (1982) 83.
- [18] D.P. Mesquita, O. Dias, A.L. Amaral, E.C. Ferreira, *Microsc. Microanal.* 16 (2) (2010) 166.
- [19] K.P. Singh, A. Malik, D. Mohan, S. Sinha, V.K. Singh, *Anal. Chim. Acta* 532 (2005) 15.
- [20] K.P. Singh, N. Basant, A. Malik, G. Jain, *Anal. Chim. Acta* 658 (2010) 1.
- [21] D.S. Lee, M.W. Lee, S.H. Woo, Y.J. Kim, J.M. Park, *Process Biochem.* 41 (2006) 2050.
- [22] A.C. Sousa, M. Lucio, O. Neto, G. Marcone, A. Pereira, E. Dantas, W.D. Fragoso, M. Araújo, R. Galvão, *Anal. Chim. Acta* 588 (2007) 231.
- [23] S.H. Woo, C.O. Jeon, Y.S. Yun, H. Choi, C.S. Lee, D.S. Lee, *J. Hazard. Mater.* 161 (2009) 538.
- [24] B.M. Wise, N.B. Gallagher, R. Bro, J.M. Shaver, W. Windig, *PLS-Toolbox Version 3.5 for use with MATLAB™*, Eigenvector Research, Inc., 2004.
- [25] C. Rosen, G. Olsson, *Water Sci. Technol.* 37 (12) (1998) 197.
- [26] A.M.A. Dias, I. Moita, R. Páscoa, M.M. Alves, J.A. Lopes, E.C. Ferreira, *Water Sci. Technol.* 57 (10) (2008) 1643.
- [27] N.D. Lourenço, C.L. Chaves, J.M. Novais, J.C. Menezes, H.M. Pinheiro, D. Diniz, *Chemosphere* 65 (5) (2006) 786.
- [28] N.D. Lourenço, F. Paixão, H.M. Pinheiro, A. Sousa, *Appl. Spectrosc.* 64 (9) (2010) 1061.
- [29] J.C. Costa, M.M. Alves, E.C. Ferreira, *Bioresour. Technol.* 100 (2009) 1180.
- [30] APHA, AWWA, WPCF, *Standard Methods for the Examination of Water and Wastewater*, American Public Health Association, Washington, DC, 1989.
- [31] *Invitrogen Molecular Probes LIVE/DEAD BacLight™ Bacterial Viability Kits. Manuals and Product Inserts*, 2004, <http://probes.invitrogen.com/media/pis/mp07007.pdf>.
- [32] *Invitrogen Molecular Probes LIVE BacLight™ Bacterial Gram Kits. Manuals and Product Inserts*, 2001, <http://probes.invitrogen.com/media/pis/mp07008.pdf>.

- [33] A.L. Amaral. Image analysis in biotechnological processes: applications to wastewater treatment. PhD Thesis, Braga, Portugal, 2003, <http://hdl.handle.net/1822/4506>.
- [34] D.P. Mesquita, O. Dias, A.M.A. Dias, A.L. Amaral, E.C. Ferreira, *Anal. Chim. Acta* 642 (1–2) (2009) 94.
- [35] C. Rosen, A chemometric approach to process monitoring and control with applications to wastewater treatment operation, PhD Thesis, Lund Institute of Technology, Sweden, 2001.
- [36] P. Teppola, S.P. Mujunen, P. Minkkinen, T. Puijola, P. Pursiheimo, *Int. Lab. Syst.* 44 (1998) 307.