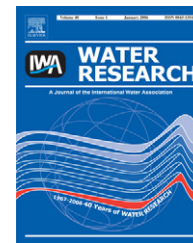


Available at [www.sciencedirect.com](http://www.sciencedirect.com)journal homepage: [www.elsevier.com/locate/watres](http://www.elsevier.com/locate/watres)

# Development of an image analysis procedure for identifying protozoa and metazoa typical of activated sludge system

Y.P. Ginoris<sup>a,\*</sup>, A.L. Amaral<sup>c</sup>, A. Nicolau<sup>b</sup>, M.A.Z. Coelho<sup>a</sup>, E.C. Ferreira<sup>b</sup>

<sup>a</sup>Departamento de Engenharia Bio química, Escola de Química/UFRJ, Centro de Tecnologia, E-113, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, CEP: 21941-900, Brazil

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

<sup>c</sup>Departamento de Tecnologia Química e Biológica—ESTIG, Instituto Politécnico de Bragança, Apartado 1038, 5301-854 Bragança, Portugal

## ARTICLE INFO

### Article history:

Received 13 October 2006

Received in revised form

25 January 2007

Accepted 5 February 2007

Available online 30 March 2007

### Keywords:

Protozoa

Metazoa

Image analysis

Pattern recognition.

## ABSTRACT

A procedure for the semi-automatic identification of the main protozoa and metazoa species present in the activated sludge of wastewater treatment plants was developed. This procedure was based on both image processing and multivariable statistical methodologies, leading to the use of the image analysis morphological descriptors by discriminant analysis and neural network techniques. The image analysis program written in Matlab has proved to be adequate in terms of protozoa and metazoa recognition, as well as for the operating conditions assessment.

© 2007 Elsevier Ltd. All rights reserved.

## 1. Introduction

A biological wastewater treatment plant (WWTP) can be considered an artificial ecosystem (Fried et al., 2000), consisting of abiotic and biotic components interacting. The abiotic components are represented by the plant and the sewage, whereas the biotic components comprise the decomposers (bacteria and fungi) that take energy for their growth from the dissolved organic matter and the oxygen in incoming wastewater and by protozoa and metazoa microfauna grazing on the decomposers (Madoni et al., 1993). The microfauna present in the aeration tank of an activated-sludge plant includes protozoa (flagellates, sarcodines and ciliates) and metazoa (rotifers, nematodes, tardigrades, gastrotrichs and oligochaetes) specimens.

Although the bacteria are generally prevalent in the aeration tank, high concentrations of protozoa in the tank normally indicates a good performance. The faunal species distribution and abundance have been pointed out as indicators of the water quality of the effluent issuing from an activated-sludge plant, providing a useful mechanism to evaluate and assess its performance. Several authors have already investigated the importance and role of the protozoa and metazoa community in the purification process of activated-sludge plants (Curds, 1973, 1975, 1982; Curds and Cockburn, 1970a, b; Curds and Vandyke, 1966). The correlation between the plant performance or operational conditions and the abundance of certain species has also been an object of study (Al-Shahwani and Horan, 1991; Esteban et al., 1991; Fried et al., 2000; Madoni 1984, 1994a, b, 2000; Madoni et al.,

\*Corresponding author. Tel.: +55 61 36316728; fax: +55 21 2562 7622.

E-mail address: [yovanka.perez@gmail.com](mailto:yovanka.perez@gmail.com) (Y.P. Ginoris).

0043-1354/\$ - see front matter © 2007 Elsevier Ltd. All rights reserved.

doi:10.1016/j.watres.2007.02.006

1981, 1993, 1996; Poole, 1984; Poole and Fry, 1980; Salvadó and Gracia, 1993; Curds and Cockburn, 1970b; Nicolau et al., 2001, 2005), which led to the development of a number of methodologies based on protozoa populations structure to assess the activated sludge plant performance, being the sludge biotic index (SBI) of Madoni (1994b) the most known method to accomplish this propose.

Species belonging to the genus *Opercularia* and *Trachelophyllum*, and to the species *Vorticella microstoma* are usually considered indicators of low effluent quality whereas *Aelosoma* sp., *Arcella* sp., *Carchesium* sp., *Epistylis* sp., *Euglypha* sp., *Euplotes* sp., order Monogononta, *Peranema* sp., *Trithigmostoma* sp. *Trochilia* sp., *Vorticella aquadulcis* and *Zoothamnium* sp., are regarded as indicators of a high quality of the treated effluent. Moreover, the organisms belonging to the Nematoda subclass, *Opercularia* sp. and *V. microstoma* are believed to dominate under poor aeration conditions (below 0.2–0.5 mg O<sub>2</sub> L<sup>-1</sup>), while *Aelosoma* sp., *Carchesium* sp., *Euglypha* sp., *Arcella* sp., Monogononta order, *Trochilia* sp., *V. aquadulcis* and *Zoothamnium* sp. are indicators of a satisfactory aeration (above 1–2 mg O<sub>2</sub> L<sup>-1</sup>). The nitrification process can be inferred by the occurrence in high densities of *Aelosoma* sp., *Arcella* sp., *Carchesium* sp., *Coleps* sp., *Epistylis* sp., *Euplotes* sp., *Trochilia* sp. and the Monogononta order. Furthermore, the presence of *Peranema* sp. and *V. microstoma* can be an indication of fresh sludge (few days), while *Aelosoma* sp., *Arcella* sp., *Euglypha* sp., and the Digononta and Monogononta orders have been pointed out as indicators of old sludge (20 days or more).

Usually, the identification and quantification of each protozoa and metazoa species is achieved by microscopic inspection and manual counting, requiring both time and high technical expertise. However, the technological advances and the decrease of computation costs gave the opportunity for new techniques such as image analysis to be used in routine classification and quantification of micro-organisms in an automated and non subjective manner overcoming some of the drawbacks of manual techniques. In image analysis, the inherent accuracy and precision of microscopy techniques and the speed of the hardware computation are combined thus reducing the human error factor.

Image analysis, in a wide sense, can be referred to both the strictly image analysis processes as well as to the overall processes of image capture, processing and analysis (Dougherty, 1994). This technique allows images improvement as well as automatic recognition and identification of patterns, such as arrangements or groups of elements that follow certain characteristics, resulting in a reduction of time and work. Indeed, image analysis has already proved to be a potentially alternative tool to overcome the drawbacks associated to the micro-organisms visual identification and quantification through the studies carried out by Amaral et al. (1999, 2004) and da Motta et al. (2001). In these studies, image analysis was used for identifying protozoa and metazoa commonly present in activate sludge treatment plants, and used straightforward to monitor the plant operational conditions in a way similar to the SBI developed by Madoni (1994a, b).

In the present work, an image analysis procedure, previously developed in Visilog by Amaral et al. (2004), was

extended to the recognition of other protozoa and metazoa species and adapted to the *Matlab* language. The morphological descriptors obtained after the image analysis were then processed using the discriminant analysis (DA) and neural network (NN) statistical multivariate techniques in order to identify the main protozoa and metazoa found in the WWTP activated sludge.

## 2. Materials and methods

### 2.1. Experimental survey

The protozoa and metazoa species studied in this work were collected from aeration basins of the urban section of Nancy (France) and Braga (Portugal) wastewater plants.

A total of 22 groups of protozoa and metazoa belonging to several species, genera, orders and sub-classes were included in the study and are presented in Table 1. In all cases the maximum period between the samples collection and the images acquisition did not exceed 3 h, and aeration was provided to the sludge during this period.

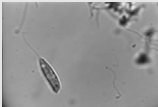
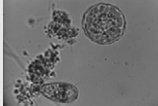
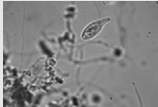
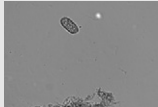
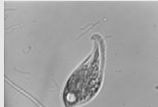
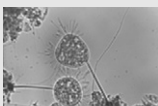
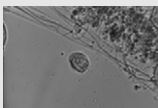
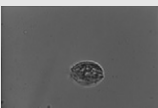
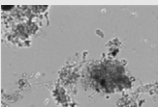

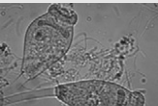

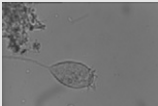
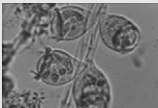
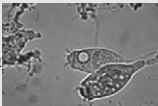





After the mixed liqueur collection, a drop of the samples was carefully deposited in a slide and covered with a cover slip (with addition of methyl-cellulose for the acquisitions between 2002 and 2006 in Portugal) for visualization and image acquisition using the bright field microscopic technique. The total magnification for visualizing and acquiring each protozoa and metazoa micro-organism was dependent on its size as follows: *Aelosoma* sp (25 and 100 times); Nematoda (100 and 250 times); Digononta, Monogonta, *Arcella* sp. and *Euglypha* sp. (250 and 400 times); *Aspidisca cicada*, *Carchesium* sp., *Epistylis* sp., *Euplotes* sp. *Litonotus* sp., *Coleps* sp., *Opercularia* sp., *Peranema* sp., Suctorina, *Trachellophyllum* sp., *Trithigmostoma* sp., *Trochilia* sp. *V. aquadulcis*, *V. microstoma*, *Vorticella* sp. and *Zoothamnium* sp. (400 times). The dimensions of metric units ( $\mu\text{m}$ ) were correlated with the corresponding pixel units using a micrometric slide.

Among the evaluated groups two different species of *Epistylis*, and *Trachellophyllum* were additionally analysed. Moreover, a group of micro-organisms with similar morphological characteristics of *Epistylis* sp. and *Opercularia* sp. was included due to the fact that when these organisms occur with the buccal apparel closed it is quite difficult to distinguish one group from the other. Finally, the frontal and lateral views of *Arcella* sp., *Aspidisca cicada* and *Trithigmostoma* sp. were also analysed, on cause of their axial lack of similitude.

### 2.2. Image acquisition

The image acquisition system used in Nancy was composed by a Leitz Dialux 20 optic microscope (Leitz, Wetzlar) coupled to a grey scale video camera Hitachi CCTV HV-720E(F) (Hitachi, Tokyo). The images were grabbed to the computer in 768 × 576 pixels and 8-bit format (256 grey levels) by a Matrox Meteor frame grabber (Matrox, Montreal) using the Visilog 5 commercial software (Noesis, S.A., les Ulis). In Braga, the acquisition system was composed by an optic microscope Zeiss Axioscop (Zeiss, Oberkochen) coupled to a Sony CCD ACV D5CE grey

**Table 1 – Protozoa and metazoa included in this study**

Protozoa	Flagellate	<i>Peranema</i>					
	Sarcodine	<i>Arcella and Euglypha</i>					
		Ciliate	Free Swimming	<i>Trachelophyllum</i>			
	Carnivorous		<i>Coleps</i>		<i>Litonotus</i>		
			Suctororia (sub-class)				
	Crawling		<i>Aspidisca cicada</i>		<i>Euplotes</i>		
					<i>Trochilia</i>		
	Sessile		<i>Carchesium</i>		<i>Epystilis</i>		
			<i>V. aquadulcis</i>		<i>V. microstoma</i>		
					<i>Zoothamnium</i>		
						<i>Opercularia</i>	
						<i>V. convallaria</i>	
	Metazoa	Rotifer	Digononta (order)		Monogononta (order)		
							
		Gastrotrichia	Nematoda (sub-class)				
Oligotrichia							
		<i>Aelosoma</i>					

scale video camera (Sony, Tokyo) and connected to a PC through the *Data Translation DT 3155* frame grabber (Data Translation, Marlboro), in order to convert the analogical voltage signal of the camera on an 8-bit digital  $768 \times 576$  pixels matrix. This digital representation was then acquired, exhibited in the computer screen and stored to the computer using the commercial software *Image-Pro® Plus* (Media Cybernetics, Silver Spring).

A smaller set of images was acquired during the present work using an acquisition system consisting of a *Leitz Laborlux S* optic microscope (Leitz, Wetzlar) coupled to a *Zeiss Axion Cam HR* video camera (Zeiss, Oberkochen). The images acquisition was performed in  $1300 \times 1030$  pixels and 8-bit format through the commercial software *Axion Vision 3.1* (Zeiss, Oberkochen). The digital images were stored in TIFF format and subsequently processed applying the image analysis technique.

### 2.3. Image analysis programme

The semi-automatic image analysis method, *ProtoRec v.4*, for the recognition and characterization of protozoa and metazoa groups was adapted from a previous programme developed by Amaral et al. (2004) and adapted to *MATLAB Image Processing Toolbox* (The Mathworks, Inc.).

The overall image processing and analysis programme consists of four major modules, namely, pre-treatment, segmentation, post-treatment and morphological descriptors determination, being each briefly explained in the following paragraphs.

### 2.4. Pre-treatment

The first stage of the *ProtoRec* programme consists in improving the original grey scale image (256 levels). The image is initially submitted to local histogram equalization in order to enhance the contrast of each region in the image, median filtering to perform noise reduction and *Bottom hat* filtering to emphasize the organism's borders. The final image

is then inverted for a better differentiation between the organism's borders and the background.

### 2.5. Segmentation

Previous to the segmentation step, a polygonal region of interest (ROI) is user defined manually around the selected organism using the mouse. Once defined the ROI, the image is segmented by thresholding the organism's borders, by a pre-defined threshold value resulting in objects assuming value 1 and the image background value 0. The available options for the threshold value are the manual threshold definition method or automatic methods applying either Otsu (Otsu, 1979) or Entropy (Russ, 1995) algorithms.

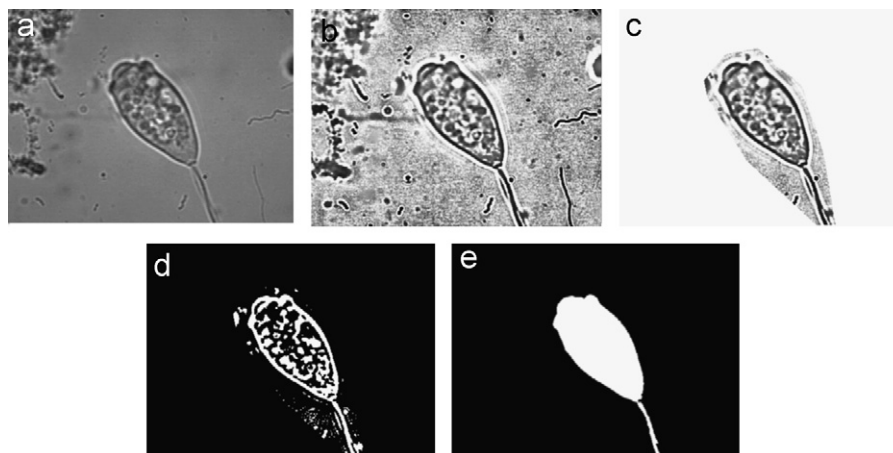
### 2.6. Post-treatment

In the subsequent stage, debris material (small artefacts and other materials that may interfere with the analysis) is eliminated by a series of morphological operations applied to the binary images including morphological closing, filling and opening operations. The main steps of the image analysis procedure are summarized in Fig. 1. Fig. 2 shows a schematic representation of the *ProtoRec* programme.

### 2.7. Morphological parameters determination

The determination of the protozoa and metazoa morphological parameters is performed in two stages. In the first stage, the parameters are computed to the whole organism's body including their external structures such as flagella, cilia, cirri and stalk. In the second stage, the parameters are determined for the organism's body core, i.e., after the removal of all external structures. The external structures removal process is performed after an empirical automatic determination of the number of erosions necessary to remove each of these structures.

Except when explicitly indicated, the morphological descriptors herein described were determined according to the *Matlab* built in functions (2004): surface (S); equivalent



**Fig. 1** – Main steps of the program: original image (a), pre-treated image (b), region of interest (c), binary image after segmentation (d), and final image (e).

diameter ( $D$ ); perimeter ( $P$ ); length ( $L$ ); width ( $W$ ); mean width ( $W_M = S/L$ ); Feret factor ( $FrF = L/W$ ); eccentricity ( $Ecc$ ); form factor ( $FF = P^2/(4\pi S)$ ); largest concavity index ( $LCI$ ) (Pons et al., 1997); robustness ( $Rob$ ) (Pons et al., 1997); concavity ratio ( $CR$ ) (Pons et al., 1997); convexity ( $Conv = P_{Conv}/P$ ), where  $P_{Conv}$  is the convex envelope perimeter; compactness ( $Comp = D/L$ ); solidity ( $Sol = S/S_{Conv}$ ), where  $S_{Conv}$  is the convex envelope surface; Euclidian distance map fractal dimension ( $D_{EDM}$ ) (Russ, 1995); mass and surface fractal dimensions ( $D_{BM}$  and  $D_{BS}$ , respectively) determined by the box counting algorithm (Obert et al., 1990); area vs. perimeter fractal dimension ( $D_{AvsP}$ ), which relates the areas with the perimeters of the whole set of the objects (Soddell and Seviour, 1994) and mass ratio fractal dimension ( $D_{MR}$ ) which determines the relation-

ship between the mass of an object enclosed within different radii (Soddell and Seviour, 1994).

Some descriptors were specifically designed for the protozoa and metazoa micro-organisms, such as the mean body width vs. body width ratio ( $W_M W_B = W_{MB}/W_B$ ), the mean stalk width vs. mean body width ratio ( $W_S W_{MB} = W_{Stk}/W_{MB}$ ) and the mean stalk width ( $W_{Stk} = S_{Stk}/L_{Stk}$ ), where  $S_{Stk}$  is the stalk surface and  $L_{Stk}$  the stalk length, given by

$$L_{Stk} = \frac{(P_{Stk}/2) + \sqrt{(P_{Stk}/2)^2 - 4A_{Stk}}}{2}, \tag{1}$$

where  $P_{Stk}$  is the stalk perimeter.

All of the descriptors were determined for the full protozoan and metazoan organism (including the external structures) as well as for the protozoan and metazoan body (without external structures), except for the mean stalk width and for  $W_S W_{MB}$ .

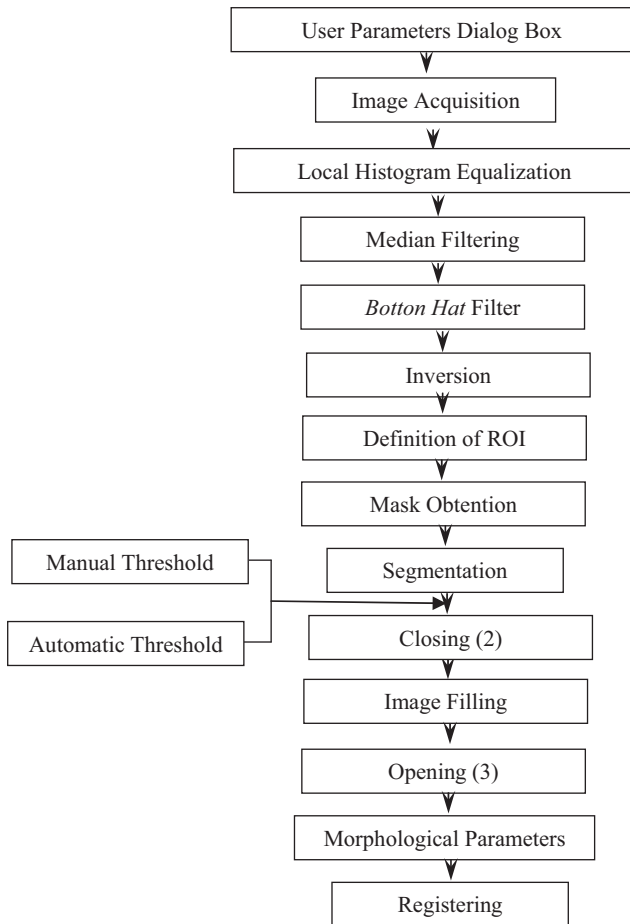


Fig. 2 – Schematic representation of ProtoRec programme.

### 2.8. Data organization

In the present work, the studied micro-organisms were split before the application of the multivariate statistical analysis in to two easily recognizable classes: stalked and non-stalked species. The separation is performed by the user, relying on the fact that the stalk is a well recognizable structure resulting in a simplified and faster multivariate analysis methodology.

Initially, a training set of each of the 22 micro-organisms were used for the determination of the discriminant functions and of the NN architecture. Regarding the stalked group two different analyses were performed: an analysis with the two *Epistylis* species as two different groups containing 10 groups and a second one with the two species represented in a single group performing a total of 9 groups. For the non-stalked set a total of 18 groups were analysed due to the fact that two different *Trachelophylum* species were studied and both front and side views (in separate groups) of *A. cicada*, *Arcella* and *Trithigmotoma* species were treated. For validation purposes a different set of individuals (test set) of each 22 micro-organisms was used with half of individual cells number of the training set and the same number of groups. The number of individual cells used in each case is presented in Table 2 and Table 3. The two values in the *Epistylis* column are reported to the cases where the two *Epistylis* species were analysed as a single group or two different groups, respectively.

In this work the protozoa and metazoa are represented by: *A. cicada* (acic), *Aelosoma* sp. (aelo), *Arcella* sp. (arce),

Table 2 – Number of individual cells present in the training set

acic	aelo	arce	carc	cole	digo	epis	ep/op	eugl	eupl	lito	mono
134	46	108	67	67	57	67/96	67	67	67	67	67
nema	oper	pera	suct	trac	trit	troc	vaqu	vcon	vmic	zoot	
37	47	67	38	86	78	46	67	67	67	67	

**Table 3 – Number of individual cells present in the test set**

acic	aelo	arce	carc	cole	digo	epis	ep/op	eugl	eupl	lito	mono
66	23	54	33	33	29	33/47	33	33	33	33	33
nema	oper	Pera	suct	trac	trit	troc	vaqu	vcon	vmic	zoot	
20	23	33	18	43	39	22	33	33	33	33	

*Carchesium* sp. (carc), *Coleps* sp. (cole), Digononta order (digo), *Epistylis* sp. (epis), *Euglypha* sp. (eugl), *Euplotes* sp. (eupl), *Litonotus* sp. (lito), Monogononta order (mono), Nematoda sub-class (nema), *Opercularia* sp. (oper), *Peranema* sp. (pera), Suctorina sub-class (suct), *Trachelophyllum* sp. (trac), *Trithigmostoma* sp. (trit), *Trochilia* sp. (troc), *V. aquadulcis* (vaqu), *V. convallaria* (vcon), *V. microstoma* (vmic) and *Zoothamnium* sp. (zoot). When it was not possible to determine if a given cell was an *Epistylis* or an *Opercularia* (closed buccal apparatus) the term ep/op was adopted.

The morphological descriptors determined in the last stage of the image analysis programme were subsequently processed using DA and NNs multivariate statistical techniques included in the *Matlab* statistical toolbox, in order to identify and classify each specie, genus, order or sub-class.

## 2.9. Data processing

DA and NN were used to identify each protozoan or metazoan organism. To speed up the identification process a parameters reduction methodology consisting in a joint Decision Trees and Correlation Analysis procedure was applied. The effect of parameters normalisation was also taken into account with the study of standard deviation and logarithmic based normalization techniques.

### 2.9.1. Discriminant analysis

This technique is used to determine which variables discriminate between two or more naturally occurring groups. In a manner similar to factor analysis and principal components analysis, DA defines new variables (discriminant functions) as linear combinations of the original descriptors, increasing the inter-class variability and obtaining an increase separation between the studied classes and/or groups. Additionally, this technique allows the groups or classes of data to be modelled in order to reclassify the given object with a minimum error and classify new objects using the new discriminant functions (Einax et al., 1997). The objects coordinates in the new discriminant functions space are determined with the parameters raw data for each object.

The DA was carried out with all the possible combinations between the micro-organisms for all the groups. In the validation process, in order to determine each group of micro-organisms, the combinations were analysed between each pair of groups. Firstly, the position of each one of the micro-organisms test in the new space of variables was determined and this position was associated with the average position of each class of micro-organisms. To each micro-organism was attributed the closest class (with the lowest difference between the position of the micro-organism and

the average position of the class), whenever the distance was not superior to a factor of the standard deviation of the class. Several values were tested for that factor between 0.25 and 5.0, being chosen the best: for the stalked group the best value was 5.0, while for the non-stalked group the best value was 2.75, for the totality of the classes. This procedure was applied for all of the possible combinations of micro-organisms classes.

### 2.9.2. Neural networks

The NN herein used was a feed forward NN (backpropagation algorithm) consisting of a 2-layer network with no hidden layers. Two training functions Levenberg–Marquardt optimization and resilient optimization (with backpropagation algorithm) were tested for the stalked micro-organisms whereas for the non-stalked only the resilient optimization was studied. The chosen learning function in both cases was the gradient descent algorithm and the mean squared error used as a performance (error) function with its goal set to zero. For each one of these configurations 100 tests with different initial values for a maximum of 500 iterations in each test were carried out.

### 2.9.3. Parameters reduction

The parameters reduction analysis was performed by a joint procedure of a Decision Tree to highlight the most important parameters and a Correlation Analysis to establish the parameters, which presented less variability among them and, therefore, discard duplicate parameters. Both techniques were carried out for the whole set of parameters determined for the stalked (39 parameters) as well as the non-stalked (54 parameters) micro-organisms, respectively.

### 2.9.4. Normalisation techniques

In order to normalise the results, two different approaches were studied: Logarithmic normalisation and Standard Deviation normalisation. Each of these procedures was applied to the stalked and non-stalked micro-organisms training and test data respectively. In the Logarithmic normalisation procedure the natural logarithm was computed for each parameter, whereas in the Standard Deviation normalisation, the average and standard deviation values were computed and the parameters values normalised.

## 3. Results and discussion

The effectiveness of the whole procedure was assessed by determining the following parameters: recognition percentage, estimated as the ratio between the number of

micro-organisms correctly classified in a given class and the total number of analysed micro-organisms in that class; Misclassification error (Misc. %), determined as the ratio between the total number of micro-organisms incorrectly classified in a given class and the total number of micro-organisms belonging to all the other classes; and Overall recognition performance, determined by multiplying the global recognition percentage by a factor 100-(misclassification error) and then dividing by 100.

The use of a joint Decision Tree and Correlation Analysis procedure in the parameters reduction methodology resulted in 28% and 30% reductions in terms of the initial parameters set for the stalked and non-stalked micro-organisms respectively. Therefore, only 28 of the initial 39 parameters determined for the stalked identification and 38 of the initial 54 parameters determined for the non-stalked micro-organisms identification were found to bear importance.

Regarding the study of the data processing techniques, the DA multivariate statistical technique with parameters reduction and logarithmic normalization was found the best data processing technique in order to assess the WWTP critical conditions. The results obtained for the individual identification of the stalked and non-stalked organisms after applying the DA technique with parameters reduction and logarithmic normalization are presented in Tables 4 and 5.

As observed in Table 4, the individual recognition percentages for all of the non-stalked organisms were above 80%, except for Monogononta with 78.8%. The misclassification error was found to be negligible (below 1%, except for Digononta and *Euglypha*) yielding good overall recognition performances. However, for the stalked organisms (Table 5) the individual recognition percentage were poor with values lower than 80% (except for *Suctorina* and *Opercularia*). In accordance, the misclassification error presented high values up to 5.5%, and the overall recognition performance did not usually surpass 80%.

**Table 4 – Global recognition percentages, misclassification error, and overall recognition performances for non-stalked micro-organisms**

	Rec. %	Misc. %	Overall %
acic	89.4	0	89.4
aelo	100	0	100
arce	94.4	0.9	93.6
cole	100	0	100
digo	82.8	1.7	81.3
eugl	100	1.7	98.3
eupl	84.8	0.9	84.1
lito	93.9	0.9	93.1
mono	78.8	0.9	78.1
nema	95.0	0	95.0
pera	100	0.2	99.8
trac	100	0	100
trit	82.1	0.4	81.7
troc	100	0.4	99.6
ni	0		

NI: not identified

**Table 5 – Global recognition percentages, misclassification error, and overall recognition performances for the stalked micro-organisms**

	Rec. %	Misc. %	Overall %
carc	54.5	3.6	52.6
epis	74.5	5.9	70.1
ep/op	54.5	4.7	52
oper	87.0	1.9	85.3
suct	94.4	0.4	94.1
vaqu	72.7	2.4	71.0
vcon	78.8	4.7	75.1
vmic	78.8	3.2	76.3
zoot	63.6	5.5	60.1
ni	0		

NI: not identified

**Table 6 – Global recognition percentages, misclassification error, and overall recognition performances for protozoa and metazoa groups**

	Rec. %	Misc %	Overall %
Flagellates	100	0.4	99.6
Ciliates	97.7	1.8	95.9
Sarcodines	97.7	1.7	96.2
Metazoa	97.1	0.3	96.9

**Table 7 – Global recognition percentages, misclassification error, and overall recognition performances for ciliated and non-ciliated protozoa**

	Rec. %	Misc. %	Overall %
Carnivorous	96.4	0.7	95.7
Crawling	91.9	0.3	91.6
Free swimming	100	0	100
Sessiles	99.6	0.2	99.4
Non-ciliated	98.2	2.3	95.9

The overall recognition performance, for the global stalked and non-stalked organisms, was dependent, as expected, on the individual values and was better for the non-stalked organisms, with a value of 85.6%, reflecting a global recognition percentage of 92.5% and global misclassification error of 7.5%. These results can be considered fairly good, all the more considering the complexity of the image acquisition of some microbial groups such as the crawling (movement on the flocs surfaces making difficult the identification of the external structures) and the free-swimming ciliates (high mobility and lack of axial similitude). On the opposite, the overall recognition performance for the identification of the stalked organism was lower (51.4%), reflecting a poor global recognition percentage (71.7%) and high misclassification percentage (28.3%). These results were considerably inferior with respect to the non-stalked organisms due to the high similarity in the projected shape of some species such as *V. microstoma* and

**Table 8 – Global recognition percentages, misclassification error, and overall recognition performances for the protozoa and metazoa groups, protozoan ciliates, final effluent quality, aeration, nitrification and sludge age assessment**

	Rec. %	Misc. %	Overall %
Groups	97.7	2.3	95.4
Ciliates	97.3	2.7	94.7
Effluent Quality	90.1	9.9	81.2
Aeration	92.4	7.6	85.4
Nitrification	90.1	9.9	81.2
Sludge Age	95.8	4.2	91.7

**Table 9 – Global recognition percentages, misclassification error, and overall recognition performances for the critical conditions assessment**

	Rec. %	Misc. %	Overall %
Low effluent quality	93.9	1.3	92.7
Low aeration	90.8	1.3	89.6
Fresh sludge	89.4	1.3	88.3

*V. aquadulcis*, and *V. convallaria*, *Carchesium* and *Zoothamnium*, hardly distinguishable from each other.

Analysing the main protozoa and metazoa groups identification (flagellates, ciliates, sarcodines, and metazoa) in Table 6 and the protozoan ciliates recognition (carnivorous, crawling, swimming, sessile and non-ciliated) in Table 7, the global recognition percentages were found to be higher than 95% (except for the crawling ciliates). The misclassification error was not significant (below 2% except for the non-ciliates) resulting in overall recognition performances higher than 95% (again except for the crawling ciliates). Furthermore, in the global protozoa and metazoa groups assessment, the overall recognition performance attained a quite high value of 95.4%, as well as the global ciliated protozoa groups identification (94.7%) presented in Table 8.

The values obtained for the overall plant operating conditions assessment are shown in Table 9 where the results can be considered fairly reasonable in the overall recognition performance for the effluent quality, aeration and nitrification evaluation (above 80%) and good for the sludge age determination (91.7%). Close related to the overall results, the assessment of critical conditions such as low effluent quality, low aeration, and fresh sludge shown in Table, attained also good overall recognition performance levels (around 90%).

#### 4. Conclusions

The use of a parameters reduction methodology resulted in 28% and 30% reductions in terms of the initial parameters set for the stalked and non-stalked micro-organisms, respectively. Regarding the study of the data processing techniques, the DA multivariate statistical technique with parameters

reduction and logarithmic normalization was found the best data processing technique in order to assess the WWTP critical conditions, and, therefore, comprise the results presented in this work.

The non-stalked organisms' identification obtained individual recognition percentages above 80% with insignificant misclassification errors (usually below 1%), therefore, yielding good overall recognition performances (above 80%). For the stalked organisms, however, the individual recognition percentage was poor with values generally lower than 80%, misclassification errors up to 5.5%, and overall recognition performances usually below 80%. Furthermore, a closer look to the non-stalked organisms' results reveals a slight misclassification problem between the Monogononta and Digononta assessment, whereas among the stalked organisms the *Suctorina* sp. was able to be correctly identified as well as *Opercularia* sp. at some extent.

Given the individual organisms results, the overall recognition performance for the global non-stalked organisms (85.6%) was naturally higher than for the stalked organisms (51.4%). The non-stalked recognition performance can be considered rather good, bearing in mind the crawling ciliates movement within the flocs and the free-swimming high mobility and lack of axial similitude. On the opposite, the overall recognition performance for the identification of the stalked organism was lower (51.4%) due to the high similarity in the projected shape of some species (*V. microstoma* and *V. aquadulcis*, and *V. convallaria*, *Carchesium* and *Zoothamnium*).

The main protozoa and metazoa groups (flagellates, ciliates, sarcodines, and metazoa) and protozoan ciliates (carnivorous, crawling, swimming, sessile and non-ciliated) recognition attained good overall recognition performances (above 95%, except for crawling ciliates), with global protozoa and metazoa groups assessment around 95%.

The overall recognition performance for the plant operating conditions assessment were quite fair for the effluent quality, aeration and nitrification evaluation (above 80%) and good on the sludge age determination (91.7%). Close related to the overall results, the assessment of critical conditions such as low effluent quality, low aeration, and fresh sludge attained also good overall recognition performance levels around 90%.

From these results it may be inferred that the current methodology can be used to predict at some extent, critical WWTP conditions in a feasible time period (within few hours) and without the need of specialized personnel in protozoology. The acquisition of the protozoa and metazoa images can be performed in a 1–2 h period for a total of 100–200 images, whilst the images treatment, data determination and analysis in 2–3 h.

#### Acknowledgments

The authors are grateful to the National Council of Scientific and Technological Development of Brazil (CNPq), the BI-EURAM III ALFA co-operation project (European Commission), and the POCI/AMB/57069/2004 project supported by the *Fundação para a Ciência e a Tecnologia* (Portugal). Data from Nancy plant made available by Prof. Maurício da Motta (UFPE, Recife, Brasil) is also acknowledged.



## REFERENCES

- Al-Shahwani, S.M., Horan, N.J., 1991. The use of protozoa to indicate changes in the performance of activated sludge plants. *Water Res.* 25 (6), 633–638.
- Amaral, A.L., da Motta, M., Pons, M.N., Vivier, H., Roche, N., Mota, M., Ferreira, E.C., 2004. Survey of Protozoa and Metazoa populations in wastewater treatment plants by image analysis and discriminant analysis. *Environmetrics* 15 (4), 381–390.
- Amaral, A.L., Baptiste, C., Pons, M.N., Nicolau, A., Lima, N., Ferreira, E.C., Mota, M., Vivier, H., 1999. Semi-automated recognition of protozoa by image analysis. *Biotechnol. Technol.* 13 (2), 111–118.
- Curds, C.R., 1973. The role of protozoa in the activated sludge process. *Am. Zool.* 13, 161–169.
- Curds, C.R., 1975. Protozoa. In: Curds, C.R., Hawkes, H.A. (Eds.), *Ecological Aspects of Used Water Treatment*. Academic Press, London, pp. 203–268.
- Curds, C.R., 1982. The ecology and role of protozoa in aerobic sewage treatment processes. *Ann. Rev. Microbiol.* 36, 27–46.
- Curds, C.R., Cockburn, A., 1970a. Protozoa in biological sewage treatment process-I. A survey of the protozoan fauna of British percolating filters and activated sludge plants. *Water Res.* 4 (3), 225–236.
- Curds, C.R., Cockburn, A., 1970b. Protozoa in biological sewage treatment process-II. Protozoa as indicators in the activated sludge process. *Water Res.* 4 (3), 237–249.
- Curds, C.R., Vandyke, J.M., 1966. The feeding habits and growth rates of some fresh-water ciliates found in activated sludge plants. *J. Appl. Ecol.* 3, 127–137.
- Da Motta, M., Pons, M.N., Vivier, H., Amaral, A.L., Ferreira, E.C., Mota, M., 2001. The study of protozoa population in wastewater treatment plants by image analysis. *Braz. J. Chem. Eng.* 18 (1), 103–111.
- Dougherty, E.R., 1994. *Digital image processing methods*. Marcel Dekker Inc.
- Einax, J.W., Zwanziger, H.W., Geiss, S., 1997. *Chemometrics in Environmental Analysis*. VCH, Weinheim.
- Esteban, G., Tellez, C., Bautista, L.M., 1991. Dynamics of ciliated protozoa communities in activated sludge process. *Water Res.* 25 (8), 967–972.
- Fried, J., Mayr, G., Berger, H., 2000. Monitoring protozoa and metazoa biofilm communities for assessing wastewater quality impact and reactor up-scaling effect. *Water Sci. Technol.* 41 (4–5), 309–316.
- MATLAB. Prentice-Hall, Upper Saddle River.
- Madoni, P., 1984. Estimation of the size of freshwater ciliate populations by sub-sampling technique. *Hydrobiologia* 111, 201–206.
- Madoni, P., 1994a. Microfauna biomass in activated sludge and biofilm. *Water Sci. Technol.* 29 (7), 63–66.
- Madoni, P., 1994b. A Sludge Biotic Index (SBI) for the evaluation of the biological performance of activated-sludge plants based on the microfauna analysis. *Water Res.* 28 (1), 67–75.
- Madoni, P., 2000. The allocation of the ciliate *Drepanomonas revoluta* to its correct functional group in evaluating the sludge biotic index. *Eur. J. Protistol.* 36 (4), 465–471.
- Madoni, P., Davoli, D., Chierici, E., 1993. Comparative analysis of the activated sludge microfauna in several sewage treatment works. *Water Res.* 27 (9), 1485–1491.
- Madoni, P., Davoli, D., Gorbi, G., 1996. Toxic effect of heavy metals on the activated sludge protozoan community. *Water Res.* 30 (1), 135–141.
- Madoni, P., Ghetti, P.F., 1981. The structure of ciliated protozoa communities in biological sewage-treatment plants. *Hydrobiologia* 83 (2), 207–215.
- Nicolau, A., Dias, N., Mota, M., Lima, N., 2001. Trends on the use of protozoa in the assessment of wastewater treatment. *Res. Microbiol.* 152 (7), 621–630.
- Nicolau, N., Martins, M.J., Mota, M., Lima, N., 2005. Effect of copper in the protistan community of activated sludge. *Chemosphere* 58 (5), 605–614.
- Obert, M., Pfeifer, P., Sernetz, M., 1990. Microbial growth patterns described by fractal geometry. *J. Bacteriol.* 172 (3), 1180–1185.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man, Cybern.* 9, 62–66.
- Pons, M.N., Vivier, H., Dodds, J., 1997. Particle shape characterization using morphological descriptors. *Part. Syst. Characterization* 14 (6), 272–277.
- Poole, J.E.P., 1984. A study of the relationship between the mixed liquor fauna and plant performance for a variety of activated sludge sewage treatment. *Water Res.* 18 (3), 281–287.
- Poole, J.E.P., Fry, J.C., 1980. A study of the protozoan and metazoan populations of three oxidation ditches. *Water Pollut. Control* 79 (1), 19–27.
- Russ, J.C., 1995. *The Image Processing Handbook*, second ed. CRC Press, Boca Raton, FL, USA.
- Soddell, J.A., Seviour, R.J., 1994. A Comparison of methods for determining the fractal dimensions of colonies of filamentous bacteria. *Binary* 6, 21–31.
- Salvadó, H., Gracia, M.P., 1993. Determination of organic loading rate of activated sludge plants based on protozoan analysis. *Water Res.* 27 (5), 891–895.