Anal Bioanal Chem (2008) 391:1321–1325
DOI 10.1007/s00216-008-1845-y

ORIGINAL PAPER

Stalked protozoa identification by image analysis and multivariable statistical techniques

A. L. Amaral · Y. P. Ginoris · A. Nicolau · M. A. Z. Coelho · E. C. Ferreira

Received: 15 October 2007 / Revised: 26 December 2007 / Accepted: 8 January 2008 / Published online: 8 March 2008

Abstract Protozoa are considered good indicators of the treatment quality in activated sludge systems as they are sensitive to physical, chemical and operational processes. Therefore, it is possible to correlate the predominance of certain species or groups and several operational parameters of the plant. This work presents a semiautomatic image analysis procedure for the recognition of the stalked protozoa species most frequently found in wastewater treatment plants by determining the geometrical, morphological and signature data and subsequent processing by discriminant analysis and neural network techniques. Geometrical descriptors were found to be responsible for the best identification ability and the identification of the crucial *Opercularia* and *Vorticella microstoma* microorganisms provided some degree of confidence to establish their presence in wastewater treatment plants.

Keywords Protozoa · Metazoa · Activated sludge · Image analysis · Multivariable statistical techniques

A. L. Amaral · A. Nicolau · E. C. Ferreira
IBB—Institute for Biotechnology and Bioengineering,
Centre of Biological Engineering, Universidade do Minho,
Campus de Gualtar,
4710-057 Braga, Portugal

A. L. Amaral (✉)
Departamento de Engenharia Química e Biológica,
Instituto Superior de Engenharia de Coimbra,
Instituto Politécnico de Coimbra, Rua Pedro Nunes,
Quinta da Nora,
3030-199 Coimbra, Portugal
e-mail: lpamaral@isec.pt

Y. P. Ginoris · M. A. Z. Coelho
Departamento de Engenharia Bioquímica,
Escola de Química/UFRJ, Centro de Tecnologia,
E-113, Cidade Universitária,
Ilha do Fundão, Rio de Janeiro, CEP 21941-900, Brazil

Introduction

The activated sludge process is a controlled aerobic biological wastewater treatment procedure relying on a biomass of bacteria, protozoa and metazoa to ensure the removal of organic matter and nutrients [1]. Among eukaryotes, protozoa prevail in well-performing plants, attaining densities higher than 10^6 microorganisms per millilitre in activated sludge systems [2]. Representatives of all the major taxa have been reported in several plants around the world [3], including flagellates, amoeba and, in particular, high numbers of ciliates consisting of free-swimming, crawling, carnivorous and stalked ciliates [2]. To date, only a few studies have focused on the significance of protozoa and metazoa in wastewater treatment plants (WWTP) as key organisms for improving the plant's final effluent quality [2]. Furthermore, these microorganisms act as excellent biological indicators and can be used to assess and predict the final effluent quality and overall plant performance [4]. Moreover, the plant's protozoa community structure rapidly changes as a response to different operating conditions, so regular plant monitoring is important for predicting day-to-day performance [5]. As a matter of fact, methods based on the protozoa population structure have already been used to assess WWTP performance, such as the sludge biotic index developed by Madoni [6].

However, to date, the time and labour-consuming manual activated sludge screening has not been implemented widely owing to its intrinsic drawbacks, leading to automatic image analysis being regarded as a promising tool for performing that task. Indeed, some studies have already been done using this technique combined with multivariable statistical analysis to perform the recognition of protozoa and metazoa commonly present in WWTP such as the works of Amaral et al. [7, 8], da Motta et al. [9] and Ginoris et al. [10–12].

Following those studies, in this work an image analysis procedure coupled with discriminant analysis and neural networks was used alongside a new set of signature descriptors in order to identify stalked protozoa.

Materials and methods

Experimental survey

The stalked protozoa species studied in this work were collected from aeration tanks of WWTP in Nancy (France) and Braga (Portugal) treating domestic and industrial effluents. A total of eight groups of protozoa belonging to several species, genera and subclasses were included in the study and are presented in Table 1. In all cases the maximum period between collection of the sample and image acquisition did not exceed 3 h, and aeration was provided to the sludge samples during this period.

Among the classes evaluated, two species of *Epistylis* were analysed. Moreover, an additional class of microorganisms (referred to as ep/op) with morphological characteristics similar to those of *Epystilis* sp. and *Opercularia* sp. was included owing to the fact that when these organisms occur with the buccal apparel closed it is quite difficult distinguish one class from the other.

Image acquisition

After the mixed liquor collection, a drop of the samples was deposited on a slide and covered with a cover slip (with addition of methylcellulose) for visualization and image acquisition using the bright field microscopic technique. The total magnification for visualizing and acquiring each protozoa class was ×400. As mentioned above, samples from two sites, Braga in Portugal and Nancy in France, were used. The image acquisition system used in both cases is detailed fully in [10, 11].

Image analysis program

The semiautomatic image analysis method for the recognition and characterization of protozoa and metazoa groups was adapted from a previous program developed by Amaral [7] in MATLAB (The Mathworks). The overall image processing and analysis program consists of four modules: pretreatment, segmentation, posttreatment and determination of geometrical, and morphological and signature descriptors (Fig. 1):

- Pretreatment: The first stage of the program consists in improving the original grey scale image by a local histogram equalization to enhance the image contrast, median filtering to perform noise reduction and bottom hat filtering to emphasize the organism's borders. The resulting images are then combined for better differentiation between the organism's borders and the background.
- Segmentation: First a polygonal region of interest is defined by the user around the selected organism and the organism's borders segmented by a predefined threshold. The threshold value options are the manual threshold definition method or automatic methods applying either Otsu [13] or Entropy [14] algorithms.
- Posttreatment: 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.
- Determination of geometrical, morphological and signature descriptors: The determination of the protozoa geometrical, morphological and signature descriptors is performed in two stages. In the first stage, the descriptors are computed for the whole organism's body, including their external structures such as cilia, cirri and stalk. In the second stage, the descriptors are determined for the organism's body core, after the removal of all external structures by an empirical automatic determination of the number of erosions necessary to remove each of these structures.

The geometrical descriptors (area, perimeter, length, width, average width, width ratio, average stalk width,

Table 1 Stalked protozoa and Suctoria species studied in this work

Carnivorous	Suctoria (subclass)
Stalked	*Carchesium* (genus)
	Epistylis (genus)
	Opercularia (genus)
	Vorticella aquadulcis
	V. convallaria
	V. microstoma
	Zoothamnium (genus)

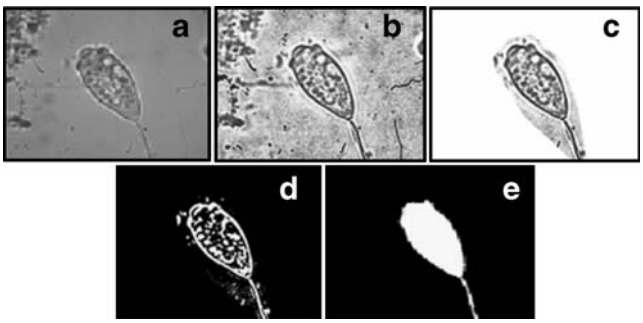

Fig. 1 Main steps of the program: original image (**a**); pretreated image (**b**); region of interest (**c**); binary image after segmentation (**d**); and final image (**e**)

average stalk/body width ratio and tentacle presence) as well as the morphological descriptors (Feret factor, eccentricity, form factor, largest concavity index, robustness, concavity ratio, convexity, compactness, solidity, Euclidian distance map fractal dimension, mass fractal dimension, surface fractal dimension, area versus perimeter fractal dimension and mass ratio fractal dimension) were determined as described in [7, 8].

The signature descriptors were determined upon the establishment of the microorganism's body signature as the frontier distances to its centroid throughout the 360° range (50 intervals) with the unitary value attributed to the first angle (0°). The 0° angle was set to correspond to the angle where the length parameter was measured (largest Feret diameter) and the absolute values were normalized to a maximum value of 1 at 0°. The length of the signature data was first normalized to 1,000 for all the microorganisms and was then sampled throughout 360° in 50 intervals. The signature descriptors were next computed as follows: maximum signature peak as the 75% percentile in the 180±35° signature range (intermediate maximum of the signature curve); minimum signature valley as the average between the 25% percentile in the 90±35° and 270±35° ranges (the two intermediate minima) and signature range as the difference between the 80% percentile in the 180±35° range and the average between the 20% percentile in the 90±35° and 270±35° ranges (difference between the intermediate maximum and minima).

Data organization

Initially, a training set of each of the eight microorganisms (around 67 individuals each) was used for the determination of the discriminant functions and of the neural network architecture with the two *Epistylis* species as two different groups in a total of ten groups (two *Epistylis* groups and the ep/op group). However, and for global identification purposes, the two different *Epistylis* species were treated as a single group for the outcome results. For validation purposes a different set of individuals (test set) was used with a third of the individual organisms of the training set.

Another study was performed on the different groups of stalked microorganisms present in this work. The Suctoria microorganisms were attributed to the carnivorous group, which represents a protozoa taxonomic division by itself. The *Vorticella* species were separated into a single group because of the fact that they are not colonial protozoa. Regarding colonial protozoa, two groups were formed on the basis of the width of their stalk (either small or large). The composition of the four groups (small stalk colonial, large stalk colonial, carnivorous and *Vorticella*) is presented in Table 2, as well as their dependence on effluent quality and aeration.

Table 2 Group division, effluent quality and aeration for the different protozoa stalked groups studied

		Effluent quality	Aeration
Small stalk colonial	*Carchesium*	Good	Good
	Zoothamnium	Good	Good
Large stalk colonial	*Epistylis*	Good	
	Opercularia	Mediocre	Mediocre
Carnivorous	Suctoria		
Vorticella	*V. aquadulcis*	Good	Good
	V. convallaria		
	V. microstoma	Mediocre	Mediocre

Data processing

To speed up the identification process a morphological descriptors reduction method consisting of a joint decision tree and correlation analysis procedure was applied. The effect of parameter normalization was also taken into account with the study of standard deviation and logarithm-based normalization techniques. Furthermore, and in order to study the influence of each set of descriptors (geometrical, morphological and signature) the standalone effect of the geometrical descriptors and the added value of the identification ability of the morphological (full and reduced) and signature descriptors were also determined.

The morphological descriptors reduction analysis was performed by a joint procedure of a decision tree to highlight the most important descriptors and a correlation analysis to establish the descriptors which had less variability among them and therefore to discard duplicates. Both techniques were carried out for the whole set of 39 descriptors ("morphological full") determined for the stalked microorganisms. This procedure resulted in a 30% reduction in terms of the initial descriptors set, with 28 ("morphological reduced") of the initial 39 descriptors being found to be of importance.

Table 3 Best recognition percentages for the different image analysis descriptors

	Recognition (%)	Misclassification (%)	Overall recognition (%)
Whole set	73.4	25.5	54.7
Geometrical	71.0	29.0	50.4
Geometrical and signature	72.7	27.3	52.9
Geometrical and morphological (full)	71.9	27.4	52.2
Geometrical and morphological (reduced)	71.7	28.3	51.4

Table 4 Recognition, misclassification and overall recognition percentages for the best overall results and the best *Opercularia* and *V. microstoma* ensemble

		Recognition (%)	Misclassification (%)	Overall recognition (%)
Best overall	*Opercularia*	87.0	2.7	84.6
	V. microstoma	69.7	3.6	67.2
	Global	73.4	25.5	54.7
Best ensemble	*Opercularia*	87.0	1.9	85.3
	V. microstoma	78.8	3.2	76.3
	Global	71.7	28.3	51.4

In order to normalize the results, two different approaches were studied: logarithmic normalization and standard deviation normalization. Each of these procedures was applied to the microorganism training and test data, respectively. In the logarithmic normalization procedure the natural logarithm was computed for each parameter, whereas in the standard deviation normalization the average and standard deviation values were computed for all the parameters for the ensemble of the microorganisms. The values of the parameters for each individual microorganism were then normalized by subtracting the average value of the parameter and dividing the result by the standard deviation value of the parameter.

Following the data organization, discriminant analysis [15] and neural networks [16] were used to identify each protozoan organism.

The discriminant analysis performed was of a linear type, i.e. the multivariate normal (MVN) density function used was a relative log posterior density function (D) with a pooled estimate of variance. The value of the MVN density function was therefore determined for each of the individual organisms regarding all the groups studied for both training and test sets. In the validation process, and in order to determine each microorganism group, the MVN density function value was determined for all the individual organisms in the test set and for each group. Each organism was then assigned to the group where it presented the highest MVN density function value (D) provided that

$$D < \left(\overline{D}_g - f\delta_g^D\right), \quad (1)$$

where $\overline{D}_g$ is the mean value of the MVN density function value for group g, δ_g^D is the standard deviation and f is a factor ranging from 0.25 to 20 in steps of 0.25. Microorganisms that did not fulfil the above condition were classified as not identified.

The programmed neural network was a two-layer (no hidden layers) feed forward neural network with a configuration of 15/10, a back-propagation algorithm and logistic sigmoidal activation functions. The gradient descent with momentum weight and bias learning function was the back-propagation learning function chosen, whereas the mean squared error was used as the performance (error) function and its goal was set to zero. The back-propagation training function was the Levenberg–Marquardt algorithm.

One hundred initial values for the neural network architecture were tested, and for each a maximum of 500 epochs were computed. In the validation process, the neural networks applied aimed to obtain an output value of 1 for the correct microorganism group and 0 for all the other groups. Therefore, each microorganism was attributed to the group with a single higher output value larger than 0.01, and microorganisms with more than a single maximum group output were classified as not identified.

Results and discussion

In this work the focus was, on one hand, on the study of a new image analysis set of descriptors (signature) further combined with the previously studied geometrical and morphological descriptors in order to describe each microorganism. On the other hand, a second goal was to determine the ability of this image analysis method to identify WWTP malfunctions by the screening of biological indicators.

The best overall recognition percentages obtained for the geometrical descriptors, geometrical and morphological descriptors, geometrical and signature descriptors and the whole set of descriptors are reported in Table 3.

From the analysis of Table 3 it is clear that the best overall recognition percentage (54.7% for the whole set by the discriminant analysis on the log-normalized data) did not attain very high values. This is due mainly to the fact that a much higher similitude exists between the morphologies of the stalked microorganisms than between the morphologies of the nonstalked microorganisms, and therefore it is not possible, at this time, to individually identify each stalked microorganism by this analysis.

Table 5 Recognition, misclassification and overall recognition percentages for the different studied stalked groups

	Recognition (%)	Misclassification (%)	Overall recognition (%)
Small stalk colonial	81.8	3.3	79.1
Large stalk colonial	91.3	5.5	86.3
Carnivorous	94.4	0.4	94.1
Vorticella	84.8	9.1	77.1

An in-depth analysis of Table 3 allows one to infer that the geometrical descriptors are responsible for the largest contribution (50.4% in 54.7%) in the identification process, accounting for more than 90% of the identification of the microorganisms. On the other hand, the signature and morphological descriptors only represent an increase in the identification of the microorganisms of, respectively, 2.5 and 1.8%. Such a result could be foreseen, given the above-mentioned high similitude between the morphologies of the stalked microorganisms and, therefore, the limited information that shape descriptors such as the morphological and signature descriptors could provide. Nevertheless, the new signature descriptors developed in this study proved to be better than the prior morphological descriptors in allowing the subtle shape differences of the microorganisms to be retrieved. Furthermore, in combination with all the previous work descriptors, it allowed for an improvement of 2.5% in overall recognition.

The results obtained for the *Opercularia* and *Vorticella microstoma* ensemble with the data processing method that yielded the best overall results are presented in Table 4. However, as these results were less than satisfactory, the most suitable data processing method for these two microorganisms was also studied and the results are presented in Table 4. These two stalked microorganisms were studied in particular given the fact that they can be seen as biological indicators of low effluent treatment quality and aeration problems, and are crucial to the full understanding of the WWTP. Analysing the results obtained, we found that *Opercularia* attained a fairly good overall recognition of 85.3%, whereas the 76.3% result for *V. microstoma* can only be seen as reasonable. Although far from perfect, these results may give some degree of confidence in order to accurately establish the presence of these two important WWTP biological indicators. These results were obtained for the geometrical and morphological (reduced) descriptors by discriminant analysis on log-normalized data, giving an overall recognition for the whole set of microorganisms of 51.4%, compared with 54.7% for the overall best results.

Regarding the study of stalked groups, the recognition, misclassification and overall recognition percentages for the four groups (small stalk colonial, large stalk colonial, carnivorous and *Vorticella*) are presented in Table 5. From the analysis of the results obtained it is possible to infer a quite good identification percentage for the carnivorous group (94.1%), and a good identification percentage for the large stalk colonial group (86.3%). These results are important given the fact that the carnivorous group represents a protozoa taxonomic division by itself and the large stalk colonial groups studied allow one to infer normal plant operating conditions. The less significant *Vorticella* and the small stalk colonial groups, however, only attained reasonable identification percentages.

Conclusions

The analysis of the different descriptors studied (geometrical, morphological and signature) obtained by image analysis allowed us to establish the relative contributions of each set of descriptors to the overall identification of microorganisms. It was found that the geometrical descriptors were, by far, the most important and only limited information could be drawn from shape descriptors such as the morphological and signature descriptors. Nevertheless, the new signature descriptors included in study this proved to be better than the previous morphological descriptors for identification purposes. Although the overall results were not satisfactory, the identification ability of the biological indicators *Opercularia* and *V. microstoma* resulted in a sharp increase in the usefulness of this method, resulting in some degree of confidence in accurately establishing their presence in WWTP. Regarding the stalked groups study, good identification percentages were obtained for the most significant carnivorous and large stalk colonial groups, whilst the less significant *Vorticella* and small stalk colonial groups only attained reasonable identification percentages.

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

References

1. Curds CR (1973) Am Zool 13:161–169
2. Madoni P (1994) Water Res 28:67–75
3. Salvadó H, Gracia MP, Amigó JM (1995) Water Res 29:1041–1050
4. Madoni P (2000) Eur J Protistol 36:465–471
5. Fried J, Mayr G, Berger H (2000) Water Sci Technol 41:309–316
6. Madoni P (1994) Water Sci Technol 29:63–66
7. Amaral AL (2003) Image analysis in biotechnological processes: applications to wastewater treatment. PhD thesis, Universidade do Minho, Braga
8. Amaral AL, da Motta M, Pons MN, Vivier H, Roche N, Mota M, Ferreira EC (2004) Environmetrics 15:381–390
9. da Motta M, Pons MN, Vivier H, Amaral AL, Ferreira EC, Mota M (2001) Braz J Chem Eng 18(1):103–111
10. Ginoris YP, Amaral AL, Nicolau A, Coelho MAZ, Ferreira EC (2007) Water Res 41:2581–2589
11. Ginoris YP, Amaral AL, Nicolau A, Coelho MAZ, Ferreira EC (2007) Anal Chim Acta 595:160–169
12. Ginoris YP, Amaral AL, Nicolau A, Coelho MAZ, Ferreira EC (2007) J Chemom 21:156–164
13. Otsu N (1979) A IEEE Trans Syst Man Cybernet 9:62–66
14. Russ CR (1995) The image processing handbook. CRC, Boca Raton
15. Einax JW, Zwazinger HW, Geiss S (1997) Chemometrics in environmental analysis. VCH, Weinheim
16. Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, Englewood Cliffs