Semi-automated recognition of protozoa by image analysis

A.L. Amaral[1], C. Baptiste[2], M.N. Pons[2], A. Nicolau[1], N. Lima[1], E.C. Ferreira[1], M. Mota[1],* &
H. Vivier[2]
[1]*Centro de Engenharia Biológica - IBQF, Universidade do Minho, 4700 Braga, Portugal*
[2]*Laboratoire des Sciences du Génie Chimique – ENSIC, Nancy, France*
Author for correspondence (Fax: +351.53.678986; E-mail: mmota@deb.uminho.pt)

Abstract

A programme was created to semi-automatically analyse protozoal digitised images. Principal Component Analysis technique was used for species identification. After data collection and mathematical treatment, a three-dimensional representation was generated and several protozoa (*Opercularia*, *Colpidium*, *Tetrahymena*, *Prorodon*, *Glaucoma* and *Trachelophyllum*) species could be positively identified.

Introduction

Protozoa are being used as biological indicators of wastewater treatment for a number of years (Curds & Cockburn 1970, Fernandez-Leborans & Novillo 1996, Lynn 1992, Macek 1998, Madoni *et al.* 1993, Salvadó *et al.* 1995, Yun-Fen *et al.* 1986). Being essential to the aerobic purification processes of residual waters, protozoa are present in the aeration tanks of wastewater treatment plants. Protozoa, and particularly ciliates, spontaneously colonise the aeration tanks, reaching densities of 10^7 per litre under optimal conditions. In normal operating conditions, about 10^6 protozoa per litre can be found. In turn, densities of 10^4 per litre or less are a clear sign of poor treatment. About 230 species of protozoa have been identified in the wastewater treatment plants – 33 flagellates, 25 rhizopodi, 6 actinopodi and 160 ciliates (Madoni 1994). Fortunately, only a small number of them occur frequently and most of these are ubiquitous appearing all over the world (Macek 1998; Madoni *et al.* 1993, Salvadó *et al.* 1995, Yun-Fen *et al.* 1986). Regarding the aeration tank as an artificial ecosystem, different functional groups may be identified taking in account the trophic relationships between protozoa or between protozoa and bacteria. Flagellates, free-swimming and sessile ciliates graze on free bacteria; crawling ciliates graze on flocculent bacteria; carnivorous ciliates predate on other protozoa. The predominance of some species or groups of protozoa may provide valuable information on the biological performance of the plant.

Some species representative of the main types of protozoa present in wastewater treatment plants (free-swimming ciliate *Colpidium* sp., crawling ciliate *Euplotes* sp., sessile ciliate *Vorticella convallaria* and carnivorous ciliate *Prorodon* sp.) are schematised in Figure 1.

One of the major drawbacks of using protozoa for wastewater treatment diagnostics is the need for protozoologists. Image analysis is, nowadays, a well-established complement of optical and electronic microscopy. It allows a routine classification and quantification of microorganisms in an automated and non-subjective manner. With the exponential increase of the processing capabilities of computers, as well as their price reduction, image analysis has become a routine in cellular biology studies. The most common applications are as diverse as the enumeration of bacteria in solid foods, *in situ* microscopy for on line fermentation monitoring, texture analysis of colonies, etc. (Vecht-Lifshitz *et al.* 1992). In this particular work, the morphology of protozoa species, present in wastewater treatment plants, is analysed in terms of
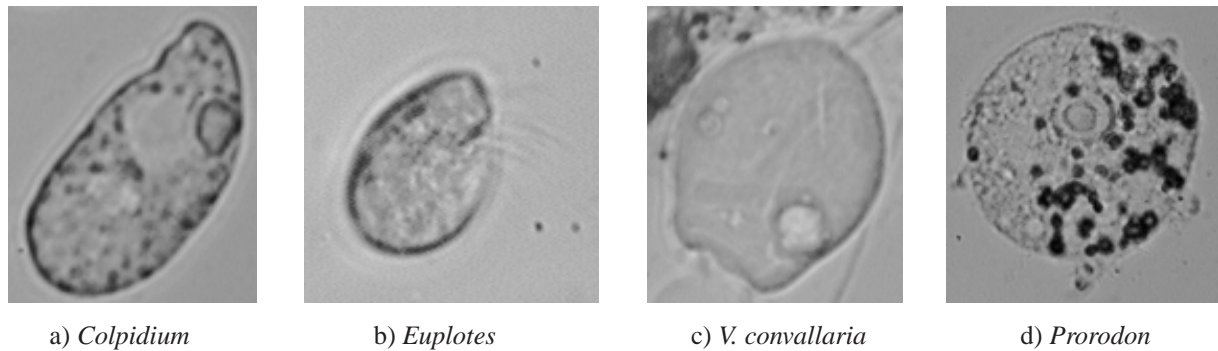
a) *Colpidium* b) *Euplotes* c) *V. convallaria* d) *Prorodon*

Fig. 1. Examples of protozoan present in wastewater treatment plants.

several mathematical morphology parameters, in order to automatically identify different protozoa groups.

Material and methods

Microorganisms and materials

The protozoa studied were obtained from the aeration tanks of the municipal wastewater treatment plant in Maxéville, Nancy (France) and belong to the four major groups of ciliates, in a total of 12 species: the free-swimming ciliates *Tetrahymena* sp., *Colpidium* sp., *Glaucoma* sp. and *Trachelophyllum* sp., the crawling ciliate *Euplotes* sp., the sessile ciliates *Opercularia* sp., *Vorticella microstoma*, *Vorticella convallaria*, *Epistylis* sp. and *Zoothamnium* sp. and the carnivorous ciliates *Litonotus* sp. and *Prorodon* sp.

The sessile protozoa *Opercularia* sp. and *Epistylis* sp. were identified without the stalk. The other sessile (*V. microstoma*, *V. convallaria* and *Zoothamnium* sp.) were divided in two groups each, according to the ability of the programme to identify the myonem.

The samples were kept in the laboratory for 24 h maximum, with constant aeration and temperature at 20 °C. Samples were viewed by phase contrast microscopy at a magnification of 400 with minimum lightning.

Image analysis

Image acquisition was achieved by a monochrome camera Hitachi CCTV (Tokyo, Japan), an acquisition board Matrox Meteor and the software VISILOG 5.1 from Noesis (les Ulis, France). The contrast value in the acquisition software was of 200% and the luminosity of 150%. With the magnification of 400, each pixel represents a length of 0.2899 μm, both in the

horizontal and in the vertical axis, and a total area of 0.0840 μm^2. The number of images acquired on 256 grey-levels was of 295, which were then analysed by the developed program (*ProtoRec*).

The commercial software XlStat from Thierry FHAMY (Paris, France) was used to perform the Principal Component Analysis (PCA) (Einax *et al.* 1997). The morphological descriptors chosen to be determined by the programme, for each individual protozoan, were the *Area* (A), the *Eccentricity* (E), the *Feret shape* (FS), the *Area-Perimeter shape* (APS) and the *Length* (L). The Feret shape and the Length are deduced from the distribution of the Feret diameters (Russ 1995). The Feret Shape is the ratio of the largest Feret diameter (which is the Length of the protozoan) to the Feret diameter at 90° of the largest one. The definitions of *APS* and *E* are:

$$APS = \frac{P_c^2}{4\pi A}$$

where P_C is the Crofton perimeter

$$E = \frac{(4\pi)^2 (M_{2x} - M_{2y})^2 + 4M_{2xy}^2}{A^2}$$

where M_{2x}, M_{2y} and M_{2xy} are the second order centred moments of the protozoan silhouette.

The programme developed

The software developed (*ProtoRec*) was programmed in VISILOG 5.1 allowing a semi-automated recognition of protozoa, using the high-level functions of VISILOG language. The programme code was then converted to the low-level language of VISILOG, the C related language COOL, which utilises functions of VISILOG's own library. A schematic representation of the programme is shown in Figure 2.
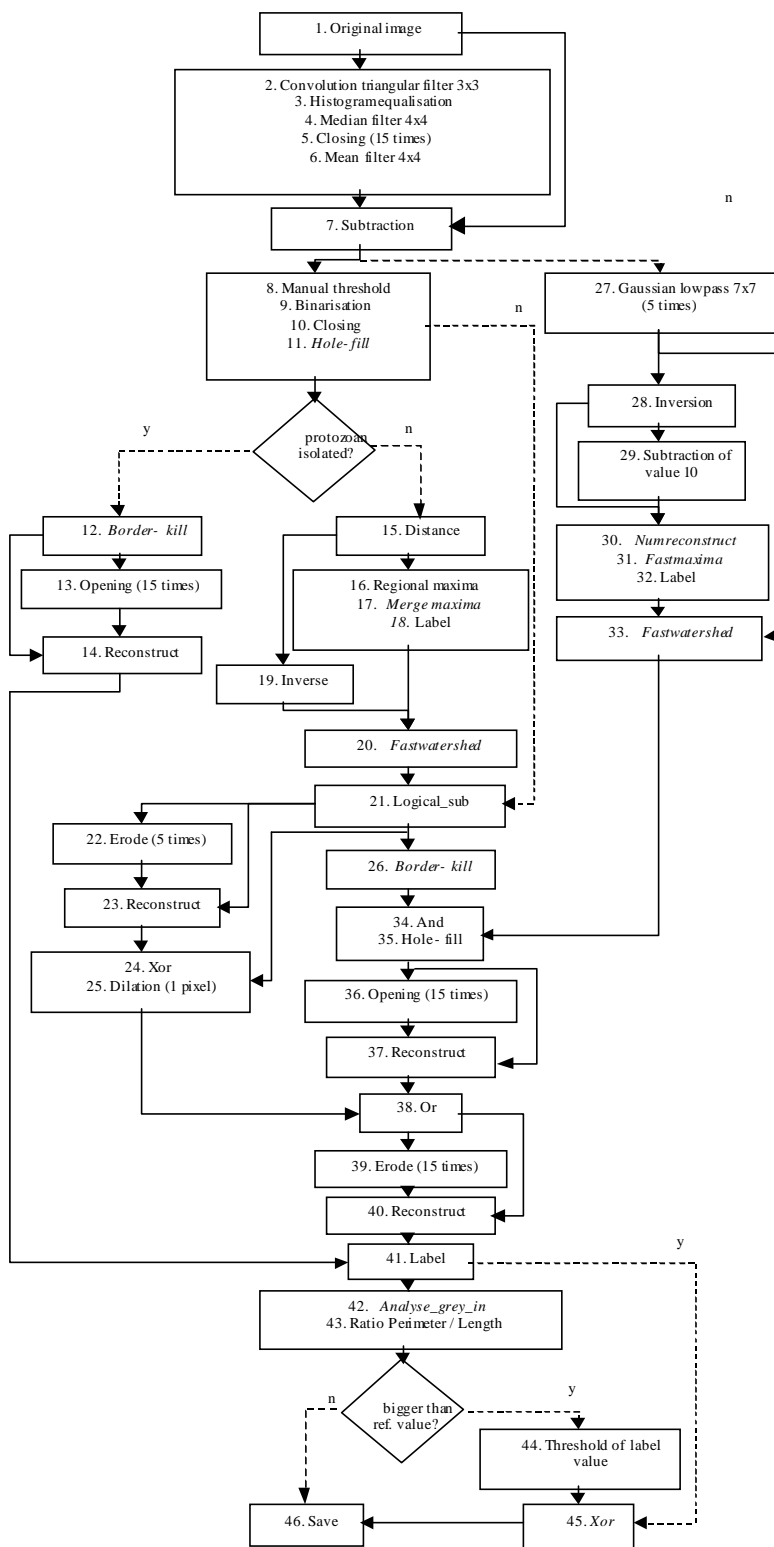
1. Original image

2. Convolution triangular filter 3x3
3. Histogram equalisation
4. Median filter 4x4
5. Closing (15 times)
6. Mean filter 4x4

7. Subtraction

n

8. Manual threshold
9. Binarisation
10. Closing
11. *Hole- fill*

27. Gaussian lowpass 7x7
(5 times)

n

protozoan isolated?

y   n

28. Inversion

29. Subtraction of value 10

12. *Border- kill*

15. Distance

30. *Numreconstruct*
31. *Fastmaxima*
32. Label

13. Opening (15 times)

16. Regional maxima
17. *Merge maxima*
18. Label

14. Reconstruct

33. *Fastwatershed*

19. Inverse

20. *Fastwatershed*

21. Logical_sub

22. Erode (5 times)

26. *Border- kill*

23. Reconstruct

34. And
35. Hole- fill

24. Xor
25. Dilation (1 pixel)

36. Opening (15 times)

37. Reconstruct

38. Or

39. Erode (15 times)

40. Reconstruct

41. Label

y

42. *Analyse_grey_in*
43. Ratio Perimeter / Length

bigger than ref. value?

n   y

44. Threshold of label value

46. Save

45. *Xor*

Fig. 2. Programme flowsheet.

The following stages were applied in first place: application of a convolution triangular filter, box size 3×3, to soften the image and eliminate the acquisition noise (point 2); histogram equalisation of the image, in order to enlarge the values range and enhance the contrast (point 3); application of a median filter with box size of 4×4, to soften the noise made by the previous operation (point 4); determination of the background image using the *closing* operation applied 15 consecutive times (point 5) and of a mean filter of size 30×30 to soften the background image (point 6) and finally subtraction of the background image from the image resultant of point 2, in order to eliminate the background and correct the pixel's values difference due to the non-uniform light (point 7).

In the end of this set of steps the 256 grey value input image of the protozoan becomes enhanced.

In a second stage, the programme recognises the objects (protozoa and a few flocs) by: binarisation using a manual threshold selection (point 8), with the objects acquiring the value 1 and the background the value 0 (point 9); *closing* operation applied once to close possible open object's contours (point 10) and finally inner area filling of the objects by the use of the VISILOG function *hole-fill*; this function calculates the inverse image and then performs its geodesic dilation (point 11).

Depending on the binary image obtained, the previous set of operations may lead to two different destinations: *isolated protozoa* or *non-isolated protozoa*. If the protozoa were isolated from the flocs, the latter will be removed by: elimination of the border-crossing objects, with the VISILOG function *border-kill* (point 12); elimination of the smaller objects (dirtiness present in the sludge), using the *opening* function applied 15 consecutive times (point 13) and reconstruction of the remaining objects with the VISILOG function *reconstruct* to the last image and the resultant image of point 12 (point 14). After these operations, the image is prepared to be analysed, which will be explained in detail ahead in items 41 to 47 of the text below.

In the case where protozoa are touching the flocs, the programme performs a different set of steps. The first procedure aims at the separation of the protozoa from the flocs and operates as follows: determination of the distance function of the binary image, with the VISILOG function *distance*; this function calculates the distance of each pixel to the nearest border (point 15); determination of regional maxima's image of the distance function, with a 2 grey level range (point 16), and numerical identification of each of the maxima obtained with the VISILOG functions *merge maxima* (Russ 1995) (point 17) and *label* (point 18); calculation of the inverse of the distance image (point 19) and utilisation of this image and the resultant image of point 18, for subsequent determination of the separation lines of the objects with the VISILOG function *fastwatershed* (Russ 1995) (point 20); isolation of the objects by eliminating the pixels which define the separation lines in the image resulting after completion of point 11, through the function *logical_sub* (point 21); elimination of the objects in contact with the image border line through the VISILOG function *border-kill* (point 26).

An image with the marks of the larger or more regular objects is then calculated: extreme smoothing of the image obtained in point 7, through the application of a linear filter of low gauss frequencies of box size 7×7, using the VISILOG function *lowpass_* 7×7, applied 5 consecutive times (point 27); inversion of the image (point 28) and subtraction of the value 10 (point 29) for later determination of the regional maxima enclosing a range of 10 levels of grey; numerical reconstruction of the inverse of the image, with the last image and the resultant image of point 28, with the VISILOG function *numreconstruct* (Russ 1995) (point 30); determination of the matrix of the maxima of the previous image, within a range of 10 levels of grey, and numerical identification of each one of the maxima by the VISILOG functions *fastmaxima* (point 31); and *label* (point 32); determination of the internal lines of separation in the objects through the VISILOG function *fastwatershed* (point 33); calculation of the image of the markers through the logical operation *and* of the resultant image from point 26 and the last image; this operation allows the internal encirclement of the objects by a closed line (point 34) and finally inner area filling by the VISILOG function *hole-fill* which concludes the isolation of protozoan from flocs (point 35).

The elimination of small flocs is achieved by using the operation *opening*, applied 15 consecutive times (point 36), and the reconstruction of the remaining objects through the VISILOG function *reconstruct* (point 37) to the last image and the resulting image from point 35.

Finally, it is necessary to recognise and bind, for each individual protozoan, its particular characteristics, such as cilia, flagella and myonem, cut off in the previous handling. This is achieved by the following operations: recognition of the cilia, flagella and

myonem with the sequence of the operations *erosion* (point 22), applied 5 consecutive times on the resultant image from point 21, reconstruction of the remaining objects by the function *reconstruct* (point 23) to the last image and the image resultant from point 21, and use of the *xor* logical operation (point 24) between this last image and the image from point 21; the image of the operation *erosion* followed by the reconstruction of the objects contains only objects with dimensions larger than 10 × 10 pixels; binding the cilia, flagella and myonem, that had been separated with the application of the *watershed* filter, to the protozoan. For this purpose, a *dilation* in 1 pixel size is used (point 25) and the operation *or* (point 38) between this last image and the image obtained from point 37. With this operation, the cilia, flagella and myonem that are found in 1 pixel distance of the protozoan are included in the individual characterisation.

As this last operation may reintroduce small particles of dirt present in the sludge (possessing one of the dimensions smaller than 11 pixels), it will be necessary to proceed to their elimination by using the operation *erosion* applied 15 consecutive times (point 39) and reconstruction of the remaining with the VISILOG function *reconstruct* (point 40), to the last image and the resultant image of point 38. The binary final image of the protozoan is thus achieved.

The programme is now ready to proceed to the image analysis stage in which the calculation of the parameters is performed and the remaining flocs eliminated. This is a two step procedure: determination through the VISILOG functions *label* (point 41) and *analyse_grey_in* (point 42) of the parameters *A*, *P*, *L*, *APS*, *FS* and *E*. These parameters are calculated from the protozoan projection on the plan of capture of the image; elimination of highly irregular objects (flocs), through the calculation of the ratio between the perimeter and the length (*P/L*) of each object (point 43) and matching with a reference value (6 for the case where the protozoan is isolated after the binarisation and 5 for the opposite case). Objects with higher values (flocs), and therefore more irregular, are discarded by the identification of the protozoan label value (point 44), binarisation of the image with the threshold set at the previous value (point 45), and the *xor* logical operation (point 46) with this last image and the resultant image from point 41.

Table 1. Eigenvalues and variability matrix.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Eigenvalue	2.0933	1.6520	0.9247	0.2578	0.0721
Variability %	0.4187	0.3304	0.1849	0.0516	0.0144
Cumulative %	0.4187	0.7491	0.9340	0.9856	1.0000

Results

The eigenvalues of the different axes are shown on Table 1. The variability taken into account for the different axes is also shown as well as the cumulative percentage, which is the basis for the selection of the first three axes.

The correlation circle is shown in Figure 3. As it can be seen the parameters Area and Length are correlated positively with the axes 1, 2 and 3, the parameter Feret Shape is correlated negatively with the axis 1, 2 and 3, and the parameters AP Shape and Eccentricity are correlated positively with the axis 1 and negatively with axis 2 and 3.

The co-ordinates (Co) of each protozoan (i) in the new axes (1, 2 and 3) are determined accordingly to the function:

$$\mathrm{Co}_i^1 = -0.0714\frac{\mathrm{FS}_i - \mu_{\mathrm{FS}}}{\sigma_{\mathrm{FS}}}$$
$$+ \quad 0.5623\frac{\mathrm{APS}_i - \mu_{\mathrm{APS}}}{\sigma_{\mathrm{APS}}} + 0.455\frac{\mathrm{E}_i - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}$$
$$+ \quad 0.3256\frac{\mathrm{A}_i - \mu_{\mathrm{A}}}{\sigma_{\mathrm{A}}} + 0.6047\frac{\mathrm{L}_i - \mu_{\mathrm{L}}}{\sigma_{\mathrm{L}}},$$
$$\mathrm{Co}_i^2 = -0.3195\frac{\mathrm{FS}_i - \mu_{\mathrm{FS}}}{\sigma_{\mathrm{FS}}}$$
$$+ \quad 0.3466\frac{\mathrm{APS}_i - \mu_{\mathrm{APS}}}{\sigma_{\mathrm{APS}}} + 0.3195\frac{\mathrm{E}_i - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}$$
$$+ \quad 0.6476\frac{\mathrm{A}_i - \mu_{\mathrm{A}}}{\sigma_{\mathrm{A}}} + 0.3177\frac{\mathrm{L}_i - \mu_{\mathrm{L}}}{\sigma_{\mathrm{L}}}$$

and

$$\mathrm{Co}_i^3 = -0.9413\frac{\mathrm{FS}_i - \mu_{\mathrm{FS}}}{\sigma_{\mathrm{FS}}}$$
$$+ \quad 0.1393\frac{\mathrm{APS}_i - \mu_{\mathrm{APS}}}{\sigma_{\mathrm{APS}}} + 0.0897\frac{\mathrm{E}_i - \mu_{\mathrm{E}}}{\sigma_{\mathrm{E}}}$$
$$+ \quad 0.2288\frac{\mathrm{A}_i - \mu_{\mathrm{A}}}{\sigma_{\mathrm{A}}} + 0.1850\frac{\mathrm{L}_i - \mu_{\mathrm{L}}}{\sigma_{\mathrm{L}}},$$

where μ is the average of the parameters for all the protozoa, σ is the standard variation and the nu-

116

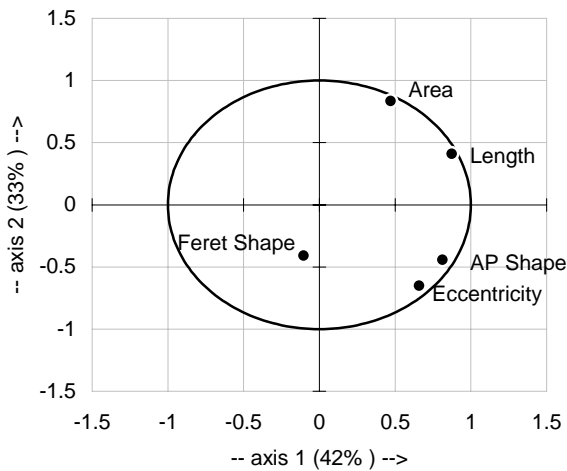Correlation's circle : axes 1 and 2 (75%)

Fig. 3. Correlation's circle.

Table 2. Values of the co-ordinates for each protozoa species in the Axis 1, Axis 2 and Axis 3.

Protozoan	Axis 1	Axis 2	Axis 3
Colpidium	−1.029	0.225	−0.374
*V. convallaria**	3.563	0.591	−0.074
V. convallaria	0.414	0.961	0.237
Epistylis	−0.057	0.836	0.112
Euplotes	−1.151	−0.446	−0.214
Glaucoma	−0.043	0.287	−0.470
Litonotus	0.480	−1.084	−0.084
*V. microstoma**	1.008	−1.407	−0.324
V. microstoma	−1.116	−0.334	0.052
Opercularia	0.843	2.038	0.466
Prorodon	−0.620	0.798	−0.456
Tetrahymena	−1.675	−0.929	−0.073
Trachelophyllum	−0.164	−2.090	0.583
*Zoothamnium**	2.546	0.320	0.734
Zoothamnium	0.142	1.119	0.185

* with stalk.

meric coefficients are values of the columns of the eigenvector matrix.

Table 2 summarises the results of each protozoan species, with respect to the axis determined by the PCA technique.

To improve the visualisation of the obtained data, the values and confidence interval ($t\sigma/\sqrt{n}$, where n is the number of samples and t the Student parameter) of these parameters, for each protozoan species, are

plotted in graphs defining the protozoan location zone. The representation of each protozoan species location in the 2D graph (Axis 1 and Axis 2) is shown in Figure 4 and the 3D graph (Axis 1, Axis 2 and Axis 3) is shown in Figure 5.

Conclusions

As may be seen from Figures 4 and 5, the introduction of a three-dimensional representation allows for reasonable separation of the different groups. As a matter of fact:

(1) The protozoan *Tetrahymena* sp., *Trachelophyllum* sp., *Colpidium* sp., *Prorodon* sp., *Glaucoma* sp. and *Opercularia* sp. are completely separated. The protozoan *V. microstoma* sp. with stalk, *V. convallaria* sp. with stalk, *Zoothamnium* sp. with stalk, *V. convallaria* sp., *Zoothamnium* sp., *V. microstoma* sp., *Epistylis* sp. and *Prorodon* sp. are reasonably well separated (Figure 5).

(2) The four defined functional groups of ciliated protozoa (free-swimming ciliates, crawling ciliates, sessile ciliates and carnivorous ciliates) can be correctly recognised and identified. The group of sessile ciliated protozoa appears relatively well isolated from the other groups especially when represented in the Axis 1 and Axis 2, occupying the positive quadrant of Axis 1 and Axis 2. The group of sessile ciliated protozoa recognised with stalk (*V. microstoma* with stalk, *Zoothamnium* sp. with stalk and *V. convallaria* with stalk), appear well isolated as compared with the remaining protozoa. These protozoa are also those that present higher values in the Axis 1. The *V. convallaria* with stalk and *Zoothamnium* with stalk groups appear superposed. However, as the presence of each or both groups as an indicator of good operating conditions, this feature is not a drawback of the method.

The group of the crawling ciliated protozoa (*Euplotes* sp.) appears in the minus quadrants of all the axis. The group of the free-swimming ciliated protozoa appears relatively well isolated especially when represented in the Axis 1 and Axis 2, occupying the minus quadrant of Axis 1.

The group of the carnivorous ciliated protozoa appears relatively well isolated especially when represented in the Axis 1 and Axis 2, as well as Axis 1 and Axis 3, with intermediate values, between the two groups mentioned above.

(3) *V. microstoma* sp. and *Opercularia* sp., indicators of a poor efficiency of a wastewater treatment
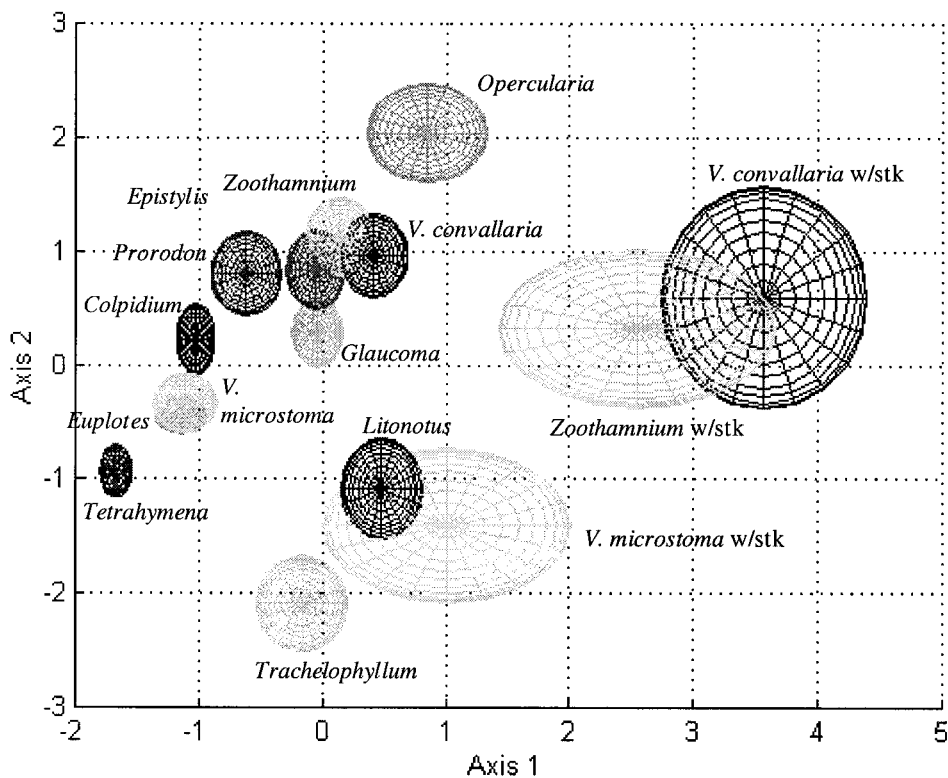
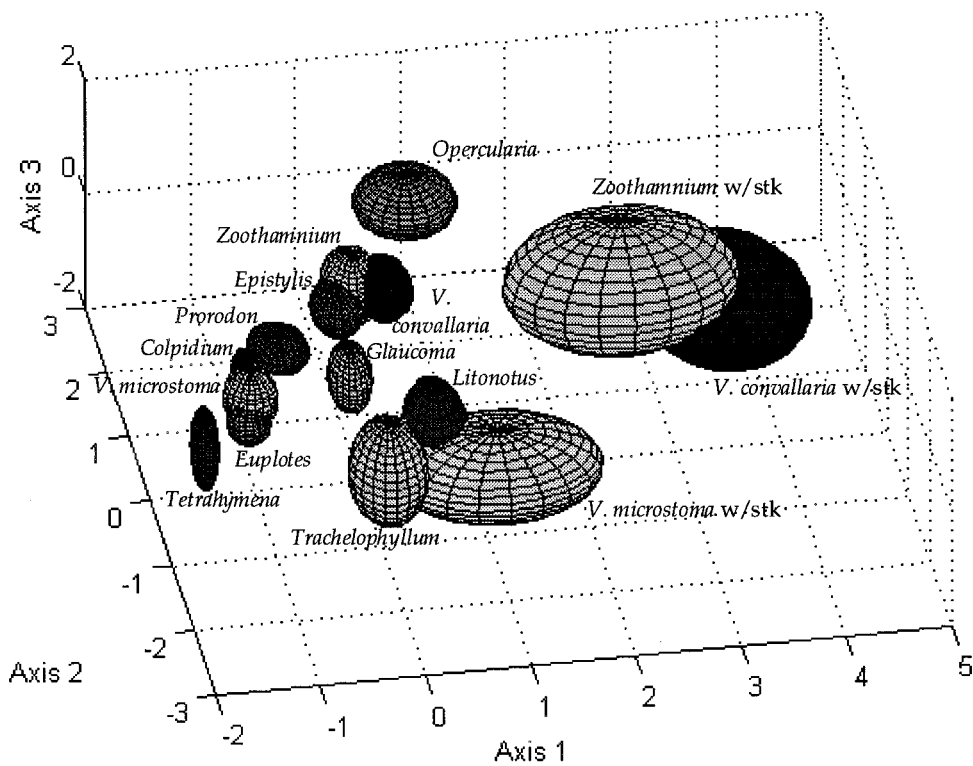Fig. 4. 2D representation of each protozoan species in the axis 1 and axis 2.

Fig. 5. 3D representation of each protozoan species in the 3 axes.

118

(Madoni 1994), are quite well isolated, thus allowing the determination of possible anomalies in the performance of the plant. It may be therefore concluded that this methodology allows already for an identification of anomaly indicators in wastewater treatments. The authors are also convinced that the introduction of another parameter, the mobility, will enhance strongly the resolution of this methodology. This is what is currently being done.

References

Chandwani RM, Pate ME, Thornhill NF, Horstmann B, Hoare M, Titchener-Hooker NJ (1997) *Bioprocess Eng.* **17**: 229–234.
Chen C, McAvoy TJ (1996) *Can. J. Chem. Eng.* **74**: 1010–1024.
Curds CR, Cockburn A (1970) *Water Res.* **4**: 237–249.
Einax JW, Zwanziger HW, Geiss S (1997) *Chemometrics in Environmental Analysis*. Weinheim: VCH.
Fenchel T (1987) *Ecology of Protozoa*. Berlin: Springer-Verlag.
Fernandez-Leborans G, Novillo A (1996) *Water Environ. Res.* **68**: 311–319.
Jahn TL, Bovee EC, Jahn FF (1979) *How to Know the Protozoa*. Dubuque, IO: Wm. C. Brown Company Publishers.
Lynn DH (1992) *J. Aquatic Ecosystem Health* **1**: 263–270.
Macek M (1989) *Int Revue Ges Hidro-Biology* **74**: 643–656.
Madoni P (1994) *Water Res.* **28**: 67–75.
Madoni P, Davdi D, Chieriici E (1993) *Water Res.* **27**: 1485–1491.
Pichon D, Vivier H, Pons MN (1992) In: Karim NM, Stephanopoulos G, eds. *Modeling and Control of Biotechnical Process*. London: Pergamon Press, pp. 307–309.
Pons MN, Mona H, Drouin JF, Vivier H (1995) In: Munack A, Schügerl K, eds. *Postprints of the 6th Int. Conf. on Computer Applications in Biotechnology*. London: Pergamon Press, pp. 189–194.
Pons MN, Wagner A, Vivier H, Marc A (1992) *Biotechnol. Bioeng.* **40**: 187–193.
Pons MN, Vivier H, Rémy JF, Dods JA (1993) *Biotechnol. Bioeng.* **42**: 1352–1359.
Russ JC (1995). *The Image Processing Handbook*. Boca Raton, FL: CRC Press.
Salvadó H, Gracia MP, Amigó JM (1995) *Water Res.* **29**: 1041–1050.
Sudo R, Aiba S (1984) *Adv. Biochem. Eng.* **29**: 117–141.
Vecht-Lifshitz SE, Ison AP (1992) *J. Biotechnol.* **23**: 1–18.
Yun-Fen S, Buikema AL, Vongue Jr WH, Pratt JR, Cairns Jr J (1986) *J. Protozool.* **33**: 146–151.