MORPHOLOGICAL CHARACTERISATION OF BIOMASS IN WASTEWATER TREATMENT USING PARTIAL LEAST SQUARES

A.L. Amaral, S. Rodrigues, M. Mota, E.C. Ferreira*
Centro de Engenharia Biológica, Universidade do Minho
4710-057 Braga, Portugal

Abstract

A wastewater treatment plant was followed in order to determine the biomass morphological properties and relate them with other measures as the Total Suspended Solids (TSS) or Sludge Volume Index (SVI). Image analysis was used to provide morphological data, which was subsequently treated by Partial Least Squares (PLS). The results showed very good correlations between observed and model predicted TSS but, considerably lower for SVI.

Key Words: Environmental Sciences, Activated Sludge, Image Analysis, Morphology, PLS.

1. Introduction

Activated sludge is a complex ecosystem constituted mainly by bacteria and protozoa. A good balance between the different species is necessary for an efficient pollution removal, a good settleability of the sludge in the final clarifier and a low level of suspended solids in the effluent. Bacteria agglomerate as flocs, mainly due to exopolymers they excrete and to filamentous bacteria that constitute the floc backbone. In absence of filamentous bacteria, small flocs are observed, that do not settle correctly (pinpoint bulking). When the filamentous bacteria are in excess, flocs do not settle correctly either (filamentous bulking). This over-development is generally due to either substrate limitation or excess [2].

Generally the amount of filamentous bacteria is estimated by manual counting under a microscope, which is a tedious and rather imprecise method. Some authors have tested automated image analysis methods to characterise the floc morphology and to relate it to settleability properties [5]. Automated image analysis seems indeed a good method to characterise quantitatively both flocs and filamentous bacteria. Reliable information of this type should enable to improve the daily operation of wastewater treatment plants. Such a method has been proposed by [7] and used to monitor bulking events in pilot plants [8].

In this work a wastewater treatment plant was followed for a period of three and a half months in order to determine the sludge morphological proprieties and relate them with the Total Suspended Solids (TSS) and Sludge Volume Index (SVI). Two image analysis programs were created providing the necessary morphological data, both in terms of flocs and filamentous bacteria, to be studied by the multivariate statistical technique Partial Least Squares (PLS).

2. Materials and Methods

2.1 Sampling of biomass

Activated sludge, the biomass analyzed in this work, were taken from a wastewater treatment plant treating domestic effluents in Braga municipality [9].

2.2 Physical measurements

Total Suspended Solids for the aeration tank was determined by weighting, as well as the diluted Sludge Volume Index, e.g. the volume in milliliters occupied by 1 g of a biomass suspension after 30 minutes settling [1]:

$$SVI = \frac{V_{30}f}{X} \qquad (1)$$

Where V_{30} is the 30 minutes settled sludge volume, X is the non-diluted suspended solids concentration and f is the dilution factor.

2.3 Image analysis

Image acquisition. Filament image acquisition (figure 1) was accomplished through phase contrast visualisation on a Zeiss Axioscop microscope (*Zeiss, Oberkochen*) with a 100x magnification, and a total of 30 images per sample. Flocs images were acquired through visualisation on an Olympus SZ 40 stereomicroscope (*Olympus, Tokyo*) with a 40x magnification, and a total of 20 images per sample.

* To whom correspondence should be sent.
Email: ecferreira@deb.uminho.pt, fax: +351 253 678 986.

All the images were digitised with the help of a CCD AVC D5CE Sony camera (*Sony, Tokyo*) and a DT 3155 Data Translation frame grabber *(Data Translation, Marlboro)*, with a 768 x 576 pixel size and 256 grey levels by the Image Pro Plus (*Media Cybernetics, Silver Spring*) software package.

The *Floc Program* and the *Filaments Program* were then created in *Matlab* 5.1 (*The Mathworks, Natick*) for determining the flocs and filaments morphological descriptors.
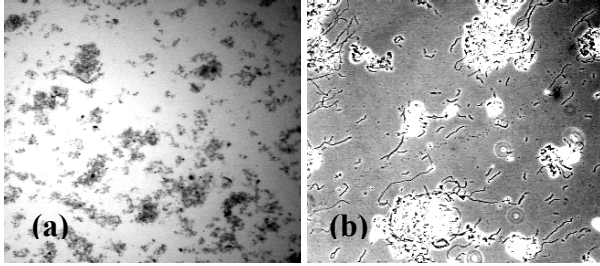
Figure 1 – Flocs (a) and filaments (b) images.

Filament program. This program was designed with the purpose of determining the filaments number, total length and average length. In this program the original image is first enhanced by a Mexican-hat filter, followed by a background homogenization, Wiener filtering and histogram equalization. After the pre-processing phase, the image was then segmented to a binary image in a 3-step procedure. In order to identify the filaments from the debris, skeletonisation and a 10-pixel length end-point removal was then performed. Finally, the required morphological parameters were determined and the results saved to file. The main steps of this program are as follows:

1. *Image acquisition*: The images names are loaded from a file and subsequently acquired as 576x768 pixel size images in 8 bit (256 levels) grayscale *tiff* format.

2. *Mexican hat*: An 8x8 pixel size Mexican hat filter is applied to the image in order to enhance the filaments on one hand and correct illumination problems on the other [10].

3. *Background homogenization*: The background is removed by a 5x5 pixel size grayscale bottom hat operation. This operation consists on subtracting the 5x5 pixel size gray closed image from the original image [10].

4. *Image smoothing*: The images resulting from points 2 and 3 are then smoothed by the Wiener filter. This low pass filter performs a 2-D adaptive noise-removal filtering using a pixel-wise adaptive Wiener method based on statistics estimated from a local neighborhood of each pixel. This filter first estimates the local mean μ and the local variance σ [6]:

$$\mu = \frac{1}{NM} \sum_{n_1,n_2 \in \eta} a(n_1,n_2) \qquad (2)$$

$$\sigma^2 = \frac{1}{NM} \sum_{n_1,n_2 \in \eta} a^2(n_1,n_2) - \mu^2 \qquad (3)$$

Where η is the N-by-M local neighborhood of each pixel in the image *a*. The filter then creates a pixel-wise Wiener filter *b* using these estimates [6]:

$$b(n_1,n_2) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2} \times (a(n_1,n_2) - \mu) \qquad (4)$$

where ν^2 is the noise variance, i.e., the average of all the local estimated variances.

5. *Image equalization*: The images resulting from point 4 are subsequently equalized by histogram equalization, in order to increase the image contrast. If a given image has T pixels using J gray levels there is a T/J pixels per gray level. An image is correctly contrasted if, for each gray level, the number of pixels is around that value. Thus, for each j gray level on the original image, the new k value in the resulting image is given by [10]:

$$k = \sum_{i=0}^{j} \frac{N_i}{T} \qquad (5)$$

where N_i is the ith intensity pixel number and T the total number of pixels. Finally, the pixel-to-pixel maximum of both images is retrieved into a new image.

6. *Segmentation*: This is a 3-step procedure. In the first step, the resulting image of point 5 is segmented into a binary image (objects: 1; background: 0), where the pixels above 0.2 are segmented to 0 and the values below 0.2 to 1 (filaments). In the second step a mask of the flocs in the image is obtained by a 10-pixel size gray closing applied to the original image (point 1), followed by a 10-pixel size average filter and then segmented to 0.9. The pixels above 0.9 are segmented to 1 (objects) and the values bellow 0.9 to 0 (background). Finally this binary image is subsequently filled (hole-fill), eroded (5 times) and dilated (6 times). The last step of this procedure consists on deleting debris inside the flocs present in the filaments image (first step), with the use of the flocs mask image (step 2). This is achieved by deleting all the objects of the filament image that appears on the flocs image.

7. *Image cleaning*: In this step the filaments binary image is first skeletonized, then the triple points of the skeleton are detected, dilated (2 times) and removed from the skeleton image. The end points of the resulting image are then determined and eroded in a 5-time iteration, corresponding to a 10-pixel length (or lower) branch deletion. Finally, the segments still present after the final iteration are used to reconstruct the skeleton image.

8. Determination of the morphological parameters: The morphological parameters were then determined from the final skeleton image. The first step consisted on labeling the different filaments and elimination of filaments (and mainly debris) shorter than 15 pixels. Afterwards the solidity of the filaments skeleton is

determined, as well as the major and minor axis length. The objects with both a solidity above 0.4 and a major axis length per minor axis length ratio bellow 5 are then discarded, allowing for the elimination of small flocs or clumps. The major axis length and minor axis length are computed respectively as the major and minor axis of the ellipse that has the same second-moments as the object, whereas the Solidity is given by [10]:

$$Sol = \frac{S}{S_C} \tag{6}$$

Where S is the area of the filaments skeleton and S_C is the area of the convex hull. The convex hull is defined as the smallest convex polygon that contains the object.

The main morphological parameters determined were the following:

Number of Filaments (Fil Nb): Number of end-points, divided by 2 and rounded upwards, after a 12-pixel length branch deletion.

Filament Length (L): Sum of the pixels in the skeleton multiplied by a factor of 1.207 to account for the different directions, divided by the number of filaments.

Total Length (TL): Sum of all the skeletons pixels from an image multiplied by a factor of 1.207.

9. *Saving the results*: The final binary image was saved in *tiff* format as well as the morphological parameters in ASCII mode.

Floc program. This program was created with the purpose of determining the morphological parameters of the floc sludge. In this program, the background from the original image is first subtracted and the resulting image is then segmented to a binary image. Next, the objects touching the border of the binary image are removed and the objects smaller than 5x5 pixels are discarded, in order to identify the flocs from the debris. Finally the required morphological parameters were determined and the results saved to file. The main steps of this program are as follows:

1. *Image acquisition*: The images names are loaded from a file and subsequently acquired as 576x768 pixel size images in 8 bit (256 levels) grayscale *tiff* format.

2. *Background image acquisition*: The background image is acquired as 576x768 pixel size image in 8 bit (256 levels) grayscale *tiff* format.

3. *Background homogenization*: The background is then removed by a pixel-to-pixel division of the original image by the background image. In order to maintain the original gray levels of the image a factor equal to the average background image is then multiplied.

4. *Segmentation*: The resulting image of point 3 is segmented into a binary image (objects: 1; background:

0), where the pixels above the 0.8 value are segmented to 0 and the values bellow 0.8 to 1 (flocs).

5. *Border cleaning*: In this step the objects touching the border of the binary image are removed.

6. *Determination of the morphological parameters*: The morphological parameters were then determined from the final binary image. The first step consisted on labeling the different flocs and elimination of flocs smaller than 5x5 pixels (mainly debris). The main morphological parameters determined were the following:

Floc Surface or Area (A): The floc surface is determined by the projected surface of the floc as the number of pixels belonging to that floc.

Total Area (TA): Sum of all the floc surfaces from an image.

Solidity: Defined as previous.

Extent: Defined as the ratio between the area of an object and the bounding box area [6]:

$$Ext = \frac{S}{S_{BoundingBox}} \tag{7}$$

Eccentricity: Defined by the following equation [4]:

$$Ecc = \frac{(4\pi)^2 (M_{2X} - M_{2Y})^2 + 4M_{2XY}^2}{S^2} \tag{8}$$

where the 2nd order moments are defined as:

$$M_{2X} = \frac{1}{S} \sum (x_i - M_{1X})^2 \tag{9}$$

$$M_{2Y} = \frac{1}{S} \sum (y_i - M_{1Y})^2 \tag{10}$$

$$M_{2XY} = \frac{1}{S} \sum (x_i - M_{1X})(y_i - M_{1Y}) \tag{11}$$

being (x_i, y_i) the coordinates of the objects pixels.

Convexity: Defined as the ratio between the convex perimeter and the perimeter of an object [10]:

$$Conv = \frac{P_{Conv}}{P} \tag{12}$$

Where the perimeter P is calculated from the 4-connected perimeter P_4 and 8-connected perimeter P_8 [4]:

$$P = \frac{0.414 P_4 + 0.586 P_8}{1.055} \tag{13}$$

being P_{Conv} the perimeter of the smaller convex polygon containing the object.

Form Factor: Defined as [10]:

$$FF = \frac{4\pi S}{P^2} \qquad (14)$$

Roundness: Defined as [4]:

$$Round = \frac{4\pi S}{P_C^2} \qquad (15)$$

7. *Saving the results*: The final binary image was saved in *tiff* format as well as the morphological parameters in ASCII mode.

2.4 Partial least squares (PLS) method

Simca 7.01 (Umetri, Umea) software package was used to perform PLS analysis from the dataset.

Partial Least Squares regression uses latent variables u_1, u_2, …, i. e., matrix U, for separately modelling the objects in the matrix of dependent data Y, and t_1, t_2, …, i. e., matrix T, for the objects in the matrix of dependent data X. The latent variables U and T are the basis of the regression model. The starting points are the centred matrices X and Y [3].

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m_y} \\ y_{21} & y_{22} & \cdots & y_{2m_y} \\ \cdots & \cdots & \ddots & \cdots \\ y_{n1} & y_{n2} & \cdots & y_{nm_y} \end{pmatrix} = Y \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m_x} \\ x_{21} & x_{22} & \cdots & x_{2m_x} \\ \cdots & \cdots & \ddots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm_x} \end{pmatrix} = X$$
$$\updownarrow \qquad\qquad\qquad \updownarrow$$
$$U \qquad\qquad\qquad T$$

where [3]:

$$U = A \bullet T + E \qquad (16)$$

In a manner similar to the methods of Principal Components Analysis and Factor Analysis, PLS extracts linear combinations of essential features which models the original data X and Y. Furthermore, PLS also models the dependence of the two data sets. This type of model is well suited for modeling and simulating environmental relationships or for multivariate calibration.

The goodness of fit of the PLS model is calculated as an error of the prediction. Using the so-called cross-validation test one can determine the number of significant vectors in U and T and also the error of the prediction. According to the cross-validation test one uses the matrix Y, deletes a certain number of values, and calculates the PLS model without these values. The omitted values are then predicted and compared with the original values. This procedure is repeated until all values have been omitted once. Therefore an error of prediction, in terms of its dependence on the number of latent variables, is determined. The predicted residual error sum of squares (*PRESS*) is also the parameter that limits the number of latent vectors u and t [3]:

$$PRESS(t_n) = \sum_{i=1}^{n} \sum_{j=1}^{m_y} e_{ij}^2 \qquad (17)$$

and:

$$\frac{PRESS(t_k)}{PRESS(t_{k+1})} > 1 \qquad (18)$$

If an additional latent vector does not improve the error of prediction it has no adequate effect on this model. Noise cannot be predicted and therefore a minimum value of *PRESS* can be reached.

The PLS modeling is very well suited for modeling of any relationships. One may simulate changes in X and then observe the changes in Y. It is, furthermore, possible to interpret the latent vectors t or u. The latent vectors have scores for each object, which can be used to display the objects. Another possibility is to compute the correlation between original features and the latent vectors to assess the kind of interacting features for both data sets.

3. Results

The main results obtained from the image analysis programs, as well as the Total Suspended Solids (TSS) and Sludge Volume Index (SVI) are summarized in figure 2.

PLS technique was then performed to this dataset with the TSS and SVI as Y variables and the morphological parameters as X variables. The results were obtained after discarding the days 34, 67, 80 and 94 for TSS and 67, 75, 80 and 82 for SVI, as suggested by a first PLS trial using all the dataset.

Results of the PLS technique are shown in figure 3. With 3 components the cumulative fraction of the variation of the Y's variables explained by the model attained a value of 0.767 and the cumulative fraction of the variation of X's and Y's variables a value of 0.665. For the TSS only these values were 0.903 and 0.879 respectively, while for the SVI were significantly lower: 0.632 and 0.459.

Regression coefficients (referred to scaled and centered data) between either TSS or SVI and X variables are explicited in figure 4. The highest regression coefficient between TSS and X variables occurred for the parameters total floc area and the ratio between filament mean length and floc mean area (L/A). For the SVI the highest value occurred for the ratio between total filaments length and total floc area (TL/TA).
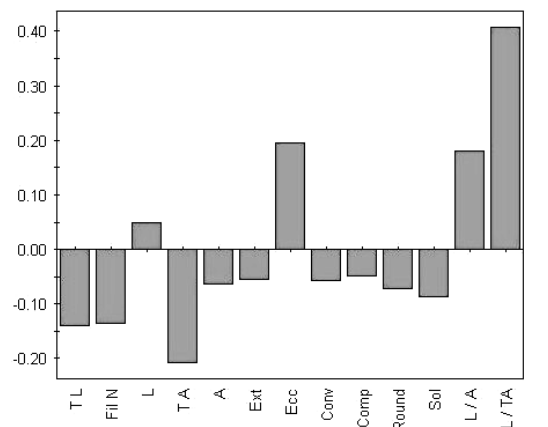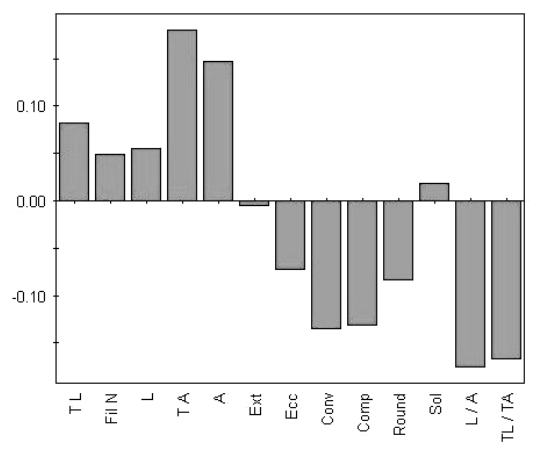
Correlations between observed and predicted (by the generated model) TSS and SVI values are shown in figure 5. The R-value for the correlation between the observed and predicted TSS was very good (0.952), while for the SVI was significantly lower (0.796).
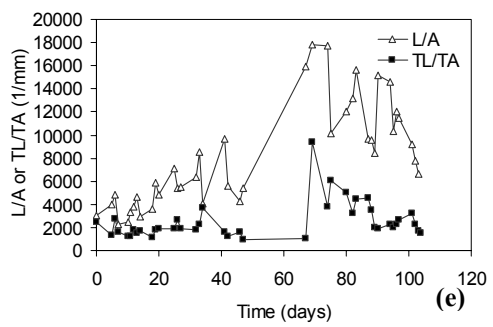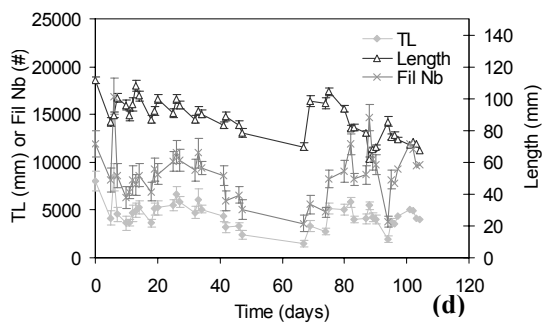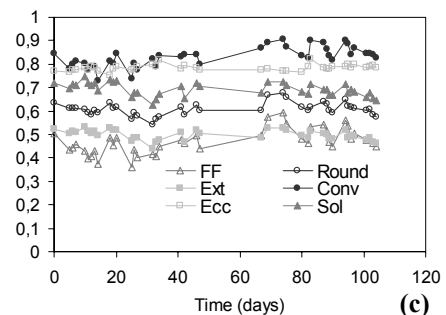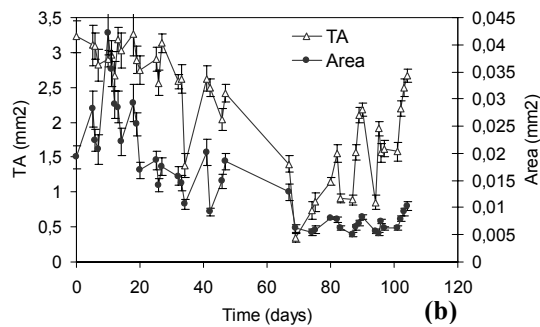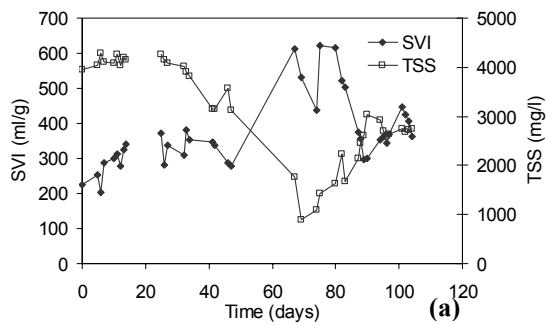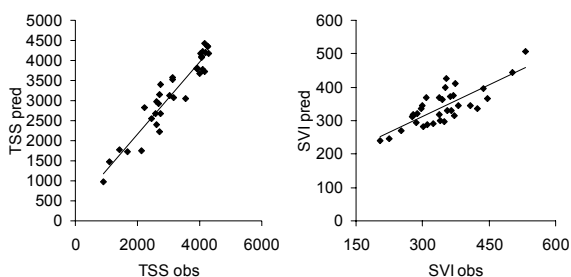
Figure 2 – Parameters time evolution: (a) SVI and TSS; (b) Mean and total floc area; (c) floc morphological parameters; (d) mean, total length and number of filaments; (e) L/A and TL/TA ratios.
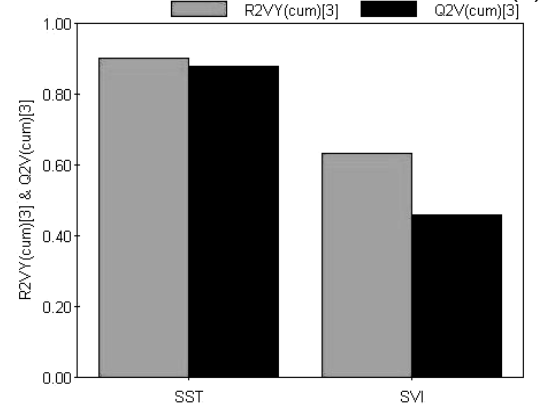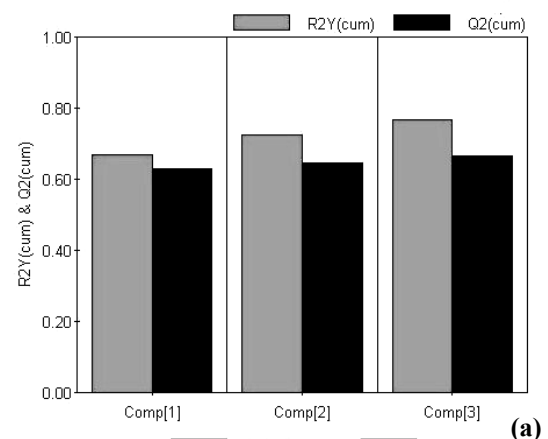
Figure 3 – Fit of the model as a function of (a) number of components and (b) each Y variable.

Figure 4 – Regression coefficients for (a) TSS and (b) SVI and the morphological parameters.

Figure 5 – Correlation between the observed and predicted TSS and SVI.

4. Discussion and Conclusions

During a wastewater treatment plant screening, filamentous bacteria and flocs were found to be in excess (high TSS values), leading to filamentous bulking problems. This ultimately led to a poor settling ability of the sludge, corroborated by the high SVI values (always superior to 200 ml/g).

Subsequent PLS analysis with 3 components showed that the cumulative fraction of the variation of X's and Y's variables obtained very good results (0.903 and 0.879) for the TSS while for the SVI the values were significantly lower: 0.632 and 0.459. These values indicate that the analysis with 3 components explained a fairly high variability of the dataset for the TSS, but not for the SVI. However, PLS analysis with higher number of components showed little increase on the explained variability.

The highest regression coefficients for the TSS occurred for both the total area, and the ratio between filament mean length and floc mean area (L/A), showing that these parameters are closely related to TSS. While the high positive correlation for the total area could be expected, the high negative correlation with L/A ratio may infer that TSS is higher for low filament/floc ratios in this range of TSS. Considering the SVI regression coefficients the highest values occurred for the ratio between total filaments length and total floc area (TL/TA), by far, the parameter more related to the SVI. These results showed the dependence of the SVI on the structure of the sludge and mainly on the filament/floc relationship, where a higher filamentous content of the biomass seems to increase SVI values. This conclusion could be also foreseeable.

Correlations between observed and predicted TSS and SVI were 0.952 and 0.796 respectively, which is a very good correlation for the TSS, but not as good for the SVI. This lower result is due probably on cause of very high SVI values and filaments contents, which can hinder the performance of the image analysis software on one hand, and only represents a small range of values in the expected relationship, on the other.

As a general conclusion the implementation of image analysis and PLS techniques was shown to provide useful information for better understanding of activated sludge processes. Furthermore, the obtained results allowed explaining the relationships between TSS and SVI and morphological parameters.

5. Acknowledgements

References

[1] L.S. Clesceri, A.E. Greensberg, R.R. Trussel, *Standard methods for the examination of water and wastewater* (Baltimore: American Public Health Association, 1989).

[2] D.H. Eikelboom, A. Andreadakis, K. Andreasen, Survey of filamentous populations in nutrient removal plants in four European countries, *Water Science and Technology*, *37*, 1998, 281-289.

[3] J.W. Einax, H.W. Zwanziger, S. Geiss, *Chemometrics in Environmental Analysis*, (Weinheim: VCH 1997).

[4] C.A. Glasbey, G.W. Horgan, *Image Analysis for the Biological Sciences*, (Chichester: John Wiley and Sons, 1995).

[5] K. Grijspeerdt, W. Verstraete, Image analysis to estimate the settleability and concentration of activated sludge, *Water Research*, *31*, 1997, 1126-1134.

[6] The MathWorks, *Image Processing Toolbox: User's Guide* (Natick: The MathWorks, Inc, 2001)

[7] M. da Motta, M.N. Pons, N. Roche, A.L. Amaral, E.C. Ferreira, M. Alves, M. Mota, H. Vivier, Automated monitoring of activated sludge using image analysis, *Proc. 1st World Congress IWA*, Paris, France, 2000, CD-ROM.

[8] M. da Motta, M.N. Pons, N. Roche, Automated monitoring of activated sludge in a pilot-plant using image analysis, *Water Science and Technology*, *47*(7), 2001, 91-96.

[9] S. Rodrigues, *Bulking filamentoso: análise, diagnóstico, soluções* (Braga: Industrial Training Report, in portuguese, 2000).

[10] C.R. Russ, *The Image Processing Handbook* (Boca Raton: CRC Press, 1995).