Correlation between sludge settleability and image analysis information using Partial Least Squares

D. P. Mesquita[1], O. Dias[1], A. M. A. Dias[1], A. L. Amaral[1,2], E. C. Ferreira[1,*]

[1]IBB-Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal ({daniela, odias, adias, ecferreira}@deb.uminho.pt)

[2]Instituto Superior de Engenharia de Coimbra, Instituto Politécnico de Coimbra, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal (lpamaral@isec.pt)

Keywords: Activated sludge, image analysis, SVI, partial least squares.

1   Introduction

In activated sludge systems, an adequate balance between the different types of bacteria is essential to ensure an efficient pollution removal, good sludge settling abilities and low suspended solids levels in the final effluent [4]. After the oxidation of the organic matter in the aerated tank, the flocculated biomass is separated from the treated effluent by means of their settling ability in the settling tank. The settling phase is considered a critical stage of the process in which filamentous bulking and deflocculation processes are the most common problems, causing the decrease of the sludge settling ability and effluent quality deterioration [12]. Poor settling biomass is normally attained by an improper aggregate's formation and filamentous bacteria proliferation, resulting in lower clarifier efficiency. Some malfunctions may occur within the activated sludge system such as pin-point flocs bulking, filamentous bulking, dispersed growth and zoogleal growth [5].

Image analysis procedures, based on microscopic acquisition, is nowadays considered to be a feasible method to characterize quantitatively aggregates and filamentous bacteria, and subsequently used to monitor bulking events in pilot plants. Bulking can be caused by both filamentous and non-filamentous factors, affecting in different ways the sludge settling ability which can be detected by image analysis methodologies [2, 6, 7, 9 and 10]. The combination of settling properties and the parameters obtained from image analysis, may offer powerful information enabling immediate interventions on the system. The macroscopic parameters can be subsequently related with the microscopic parameters using Partial Least Squares regression (PLS) multivariable statistical technique [2]. A close correlation between the filamentous bacteria per suspended solids ratio and the sludge volume index (SVI) was indeed achieved by Amaral and Ferreira [2] during filamentous bulking events. Encouraged by the success of image analysis over the last years in a broad range of different areas, the present work uses an automated image analysis method to characterize the activated sludge structure. Furthermore, this study aims to predict the sludge settling ability from the collected image analysis information using Partial Least Squares regression and a data set covering both good settleability and filamentous bulking periods.

2   Material and methods

2.1 Experimental Survey

The activated sludge samples analyzed in this work were collected from the aeration basins of six wastewater treatment plants (WWTP) treating domestic effluents in the North of Portugal. Microscopic observations were performed in 25 µL samples in order to estimate the morphological parameters of microbial aggregates and filamentous bacteria contents using digital image analysis. For each sample, total suspended solids (TSS) were determined by weight and further used to calculate the SVI [3] in a cylindrical column (10 L) for 30 min. Furthermore, previous results [2] were also included representing an earlier period of filamentous bulking conditions in one of the studied WWTP.

2.2 Image Acquisition

Around 200 images per sample were acquired in bright field microscopy to obtain representative information of the sludge. All the images were acquired in a Leitz Laborlux S optic microscope (Leitz, Wetzlar), with 100x magnification, coupled to a Zeiss AxioCam HRc (Zeiss, Oberkochen) camera. The image acquisition was performed in 1300x1030 pixels and 8-bit format through the Axion Vision 3.1 (Zeiss, Oberkochen) software. The image acquisition of the previous survey relied on 20 images per sample acquired in bright field microscopy in a SZ 4045TR-CTV Olympus stereo microscope (Olympus, Tokyo), with 40x magnification, coupled to a Sony CCD AVC–D5CE (Sony, Tokyo) grey scale video camera. The image acquisition was performed in 768×576 pixels and 8-bit format by a Data Translation DT 3155 (Data Translation, Marlboro) frame grabber using the commercial software package Image Pro Plus (Media Cybernetics, Silver Spring) [2].

2.3 Image Processing

The image processing and analysis program for aggregated and filamentous bacteria characterization, was developed in *Matlab 7.3* (The Mathworks, Inc., Natick) language, adapting a previous version developed by Amaral and Ferreira [2]. Primarily, the image processing step determined the binary images from the aggregated biomass and the protruding filamentous bacteria and thereafter, morphological parameters were determined. The main steps of the program comprise the image pre-treatment, segmentation, and debris elimination whereas the image analysis program is oriented to the aggregated and filamentous bacteria characterization and contents determination.

2.4 Image Analysis Parameters

Supported on the previous study of Amaral and Ferreira [2], 34 parameters were determined either directly from the image analysis program either in association with the sludge physical properties for a total of 400 000 aggregates. Total aggregates number per sludge volume (Nb/Vol), total aggregates area per sludge volume (TA/Vol), aggregates area (A), total filaments length per sludge volume (TL/Vol), filaments length (FL), aggregates length (L), aggregates perimeter (P) and aggregates equivalent diameter (Deq) were determined for all the samples collected. The morphological descriptors convexity (Conv), solidity (Solid), roundness (Round), and eccentricity (Ecc) were also determined by the image analysis methodology. Furthermore, the aggregates characterization was subsequently divided in two classes: intermediate aggregates between 0.025 and 0.25 mm, and large aggregates above 0.25 mm in equivalent diameter. The total filaments length per sludge volume (TL/Vol), filaments average length per aggregates average area ratio (FL/A), filaments average length per aggregates average equivalent diameter ratio (FL/Deq) and the total filaments length per total aggregated area ratio (TL/TA) were determined alongside the total filaments length per volatile suspended solids (TL/TSS) ratio characterizing the filaments dynamics. A more detailed description of each parameter can be found in [1].

2.5 Partial Least Squares

The collected data (5364 data points from 149 samples x 36 parameters) was correlated by Partial Least Squares regression (PLS), an iterative algorithm that extracts linear combinations of the essential features of the original X data while modeling the Y data dependence on the work set, being therefore well suited for multivariate calibration. PLS have been shown to be an efficient approach in monitoring complex processes since the high dimensional strongly cross-correlated data can be reduced to a much smaller and interpretable set of latent variables [8]. To perform the PLS analysis from the data set, *SIMCA 8.0* (Umetrics, Umëa) software package was used. *SIMCA* computes the variable importance in the projection (VIP) as the sum over all model dimensions (PLS components) of the variable influence. The parameters that are found to be the most important are the ones presenting a VIP value larger than 1. A more detailed description about this method can be found in [11].

3 Results and discussion

In order to apply the PLS regression, the measured SVI values were set as the Y variables while the image analysis parameters were set as the X variables. The X matrix contains representative data that can be classified, according to their physical and morphological meanings, in five main descriptor groups: free filamentous bacteria contents; free filamentous bacteria characterization; aggregates contents; aggregates size and aggregates morphology. The aggregates size and aggregates morphology groups were subdivided in 3 groups each representing the overall (average, avg), intermediate (int) and large (lrg) aggregates, resulting in a total of nine descriptor groups. For the aggregates contents and overall aggregates size, a more detailed analysis was performed, including data of intermediate aggregates (int), large aggregates (lrg) and overall (average, avg) values. For each descriptor group the parameters are described in Table 1.

Free filamentous bacteria contents	Free filamentous bacteria characterization	Aggregates contents
TL/TSS, TL/TA, TL/Vol	FL/Deq avg, FL avg, FL/A avg	Nb/Vol lrg, TSS, Nb/Vol int, Nb/Vol, TA/Vol
Aggregates size		
Overall	Intermediate aggregates	Large aggregates
A avg, Deq avg, L avg, P avg, % Area int, % Area lrg	P int, Deq int, L int	L lrg, P lrg, Deq lrg
Aggregates morphology		
Overall	Intermediate aggregates	Large aggregates
Conv avg, Ecc avg, Roun avg, Solid avg	Conv int, Ecc int, Roun int, Solid int	Conv lrg, Ecc lrg, Roun lrg, Solid lrg

Table 1 – Parameters of each studied descriptor group.

In a first approach, PLS regression was applied using an X matrix including all the data set collected from the image analysis as input parameters, to describe the SVI. According to the results the number of latent variables that allowed the best SVI model was 20, representing the cumulative fraction of the variation of the Y's and X's variables: R^2Y(cum) of 0.902, R^2X(cum) of 0.993 and Q^2(cum) of 0.812. No further improvement on the correlation was obtained by increasing the number of latent variables. Figure 1 presents the results of the PLS correlation for all the image analysis parameters included in the X matrix. The multiple correlation coefficient (goodness of fit) attained for the above PLS model was 0.948 (R^2 0.899). Furthermore, only 2 observations presented values of Y distances to model higher than 1, representing possible outliers.

The results of the variable importance in the projection (VIP), allowed for the identification of the most influent parameters to predict the SVI. Based on these results, a selection was carried out, regarding the parameters presenting VIP values higher than 1 to perform a second PLS study. The results showed that the TL/TSS was the most important parameter presenting a VIP value of 2.103, followed by the TL/TA and TL/Vol with VIP values of 1.994 and 1.864, respectively. All of these parameters represent free filamentous bacteria contents, thus clearly stating the importance of the determination of such descriptor group. It was also found that a total of 13 parameters presented a variable importance higher than 1 including: Convexity (1.444), Solidity (1.427) and Eccentricity (1.197) for the intermediate aggregates, Solidity (1.422), Convexity (1.396) and Eccentricity (1.197) for the overall aggregates, Length (1.272), Roundness (1.213), Perimeter (1.022) and Eccentricity (1.018) for the large aggregates.
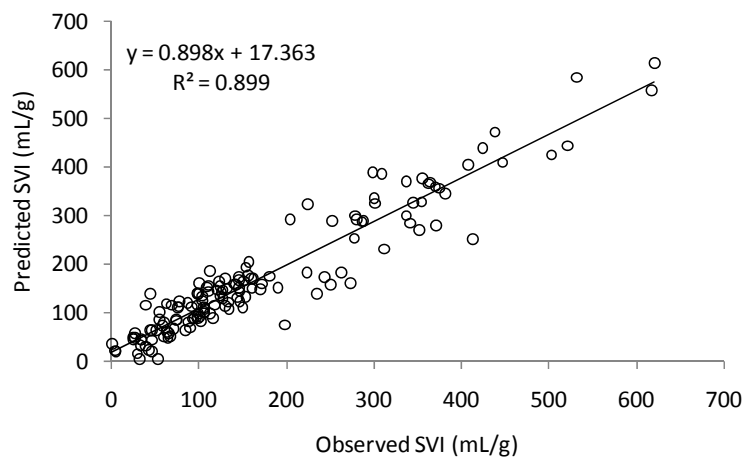
Figure 1 – Relationship between the predicted and observed SVI after PLS regression with all the data set.

Furthermore, for each descriptor group a PLS regression methodology was employed to determine the single best parameters. The parameters found to have more influence on the SVI prediction for each descriptor group were: TL/TSS (free filamentous bacteria contents group), FL avg (free filamentous bacteria characterization), Nb/Vol lrg (aggregates contents), L avg (overall aggregates size), P int (intermediate aggregates size), L lrg (large aggregates size), Solid avg (overall aggregates morphology), Solid int (intermediate aggregates morphology) and finally, Round lrg (large aggregates morphology). Based on these results, a third PLS study was performed, with the single best parameters of each of the nine descriptor groups.

Figure 2 presents the results of the PLS correlations for both the PLS regression for the parameters with VIP larger than 1 and the single best parameters of each descriptor group. When the PLS regression was performed with the single best parameters of each descriptor group a correlation coefficient of 0.914 (R^2 0.835) was achieved. Furthermore, 5 observations presented values of Y distances to model higher than 1, representing possible outliers. These results show that a reduction of 26 parameters on the data set (74% reduction) motivated a decrease of 0.034 (3.6%) on the correlation coefficient and therefore, on the prediction ability. Once again, the VIP values revealed the importance of the TL/TSS parameter (VIP 1.850) as already demonstrated in the full data set analysis, corroborating the findings of Amaral and Ferreira [2]. Apart from the free filamentous bacteria contents the descriptor groups that were found to contribute the most for the SVI prediction were the overall, intermediate (Solidity in both cases) and large (Roundness) aggregates morphology and with VIP values higher than 1.
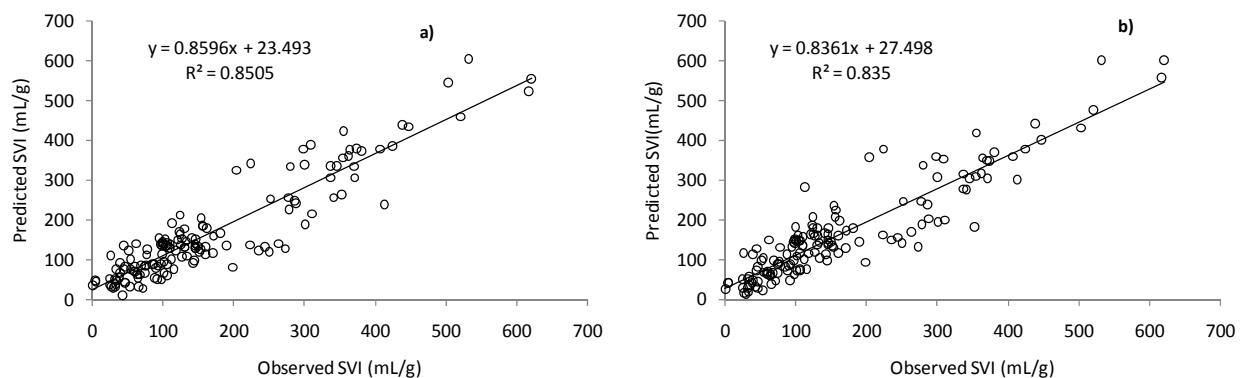

Figure 2 – Relationship between the predicted and observed SVI after PLS regression for the parameters with VIP larger than 1 (a) and the best parameter of each group (b).

Regarding the PLS regression for the parameters presenting a VIP higher than 1 on the full data set analysis a correlation coefficient of 0.922 (R^2 0.850) was achieved. Furthermore, only 3 observations presented values of Y distances to model higher than 1, representing possible outliers. These results show that a reduction of 22 parameters on the data set (63% reduction) originated a decrease of 0.026 (2.7%) on the correlation coefficient and therefore, on the prediction ability. As it had already occurred in the full data set analysis, the VIP results stated the importance of the free filamentous bacteria contents with TL/TSS, TL/TA and TL/Vol presenting the highest VIP values of the set (respectively 1.425, 1.359 and 1.218). In accordance to the findings of the descriptor groups best parameters analysis, 8 parameters representing the aggregates morphology (3 of overall aggregates, 3 of intermediate aggregates and 2 of large aggregates) and 2 parameters representing the large aggregates size were within the highest VIP parameters, following the free filamentous bacteria contents parameters.

Finally, a PLS analysis carried out solely with the 3 parameters of the free filamentous bacteria contents descriptor group obtained a correlation coefficient of 0.895 (R^2 0.801), thus explaining 95.3% of the correlation (0.948) obtained for the full data set analysis. Relating to the study of Amaral and Ferreira [2], in the course of this survey, it was possible to study a wider range of SVI data, comprising good and poor settleability properties of the sludge, whereas in the former study the SVI values were always higher than 250 mL/g corresponding to filamentous bulking. Therefore, and given the attained correlation coefficients, it seems reasonable to infer that the PLS regressions performed during this study were quite satisfactory. Furthermore, further analyses are already ongoing in order to establish non linear correlations.

4 Conclusion

The implementation of image analysis and PLS techniques has shown to provide important information for better understanding the behavior of activated sludge processes, and to predict, at some extent, the sludge volume index. Moreover, the obtained results allowed explaining the strong relationships between the sludge settling properties and the free filamentous bacteria contents, aggregates morphology and aggregates size, establishing relevant relationships between macroscopic and microscopic properties of the biological system.

5 References

[1] A. L. Amaral. *Image analysis in biotechnological processes: applications to wastewater treatment*. PhD Thesis, Braga, Portugal, 2003. (http://hdl.handle.net/1822/4506)

[2] A. L. Amaral, E. C. Ferreira. Activated sludge monitoring of a wastewater treatment plant using image analysis and partial least squares regression. *Analytica Chimica Acta* 544:246-253, 2005.

[3] APHA, AWWA, WPCF. *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington, DC, 1989.

[4] D. Jenkins, M. G. Richard, G. Daigger. *Manual on the causes and control of activated sludge bulking, foaming and other solids separation problems*. IWA publishing, 2003.

[5] G. Bitton. *Wastewater Microbiology*. Wiley/Liss, New York, 1994.

[6] M. da Motta, A. L. Amaral, M. Casellas, M. N. Pons, C. Dagot, N. Roche, E. C. Ferreira, H. Vivier. Characterisation of activated sludge by automated image analysis: validation on full-scale plants. *IFAC Computer Applications in Biotechnology*, 427-431, 2001.

[7] M. da Motta, M. N. Pons, N. Roche. Automated monitoring of activated sludge in a pilot plant using image analysis. *Water Science and Technology* 43(7):91-96, 2001.

[8] P. Teppola, S.-P. Mujunen, P. Minkkinen. Partial least squares modelling of an activated sludge plant: A case study. *Chemometrics and Intelligent Laboratory Systems* 38:197-208, 1997.

[9] R. Jenné, E. N. Banadda, I. Y. Smets, J. F. Van Impe. Monitoring activated sludge settling properties using image analysis. *Water Science and Technology* 50(7):281-285, 2004.

[10] R. Jenné, E. N. Banadda, I. Y. Smets, J. Deurinck, J. F. Van Impe. Detection of Filamentous Bulking Problems: Developing an Image Analysis System for Sludge Composition Monitoring. *Microscopy and Microanalysis*, 13: 36-41, 2007.

[11] Umetri AB. *User's Guide to SIMCA*. CD-ROM, 1998.

[12] X. Li, Y. Yuan. Settling velocities and permeabilities of microbial aggregates. *Water Research*, 36:3110-3120, 2002.