ACTIVATED SLUDGE IMAGE ANALYSIS DATA CLASSIFICATION: AN LS-SVM APPROACH

G. Gins, * **I.Y. Smets,** * **R. Jenné,** *
and J.F. Van Impe *

* *BioTeC, Department of Chemical Engineering*
Katholieke Universiteit Leuven
W. de Croylaan 46, B-3001 Heverlee, Belgium

Abstract: In this paper, a classifier is proposed and trained to distinguish between bulking and non-bulking situations in an activated sludge wastewater treatment plant, based on available image analysis information and with the goal of predicting and monitoring filamentous bulking. After selecting appropriate activated sludge parameters (filament length, floc fractal dimension and floc roundness), an LS-SVM approach is used to train a classification function. This classification function is shown to have a satisfactory performance after validation.

Keywords: classification, complex systems, image analysis, water pollution, waste treatment.

1. INTRODUCTION

One of the most often encountered problems when operating activated sludge wastewater treatment plants is filamentous bulking. Caused by the abundance of filamentous microorganisms, it precludes the proper aggregation and sedimentation of the biomass. This phenomenon has a negative influence on the performance of the wastewater treatment plant, in the worst case resulting in the escape of biomass into the environment. Even though filamentous bulking has been identified as a problem for a long time, it continues to hamper activated sludge wastewater treatment plants (Wanner, 1994).

To monitor the settleability of the activated sludge, Sludge Volume Index (SVI) [1] measurements are often performed. These measurements provide only macroscopic settling characteristics of the studied sludge. If more details on the composition of the sludge are desired, microscopic observation is required. Unfortunately, these microscopic observations are both time consuming and very subjective, varying greatly with the level of expertise of the analyst. As a result, little to no time for remedial actions is left once sedimentation problems are observed. Therefore, an automated procedure for the quick and objective analysis of activated sludge properties would be a major accomplishment in the battle against filamentous bulking.

Recent research has resulted in a well-performing image analysis algorithm (Jenné *et al.*, 2003, 2004) for the analysis of activated sludge properties. The goal of this work is to exploit the data obtained by this algorithm in predicting whether or not the studied sludge faces settling problems, thus acting

[1] The *Sludge Volume Index* is the volume in mL/g occupied by the activated sludge after 30 minutes of sedimentation. Bulking is said to occur when the SVI exceeds 150 mL/g.

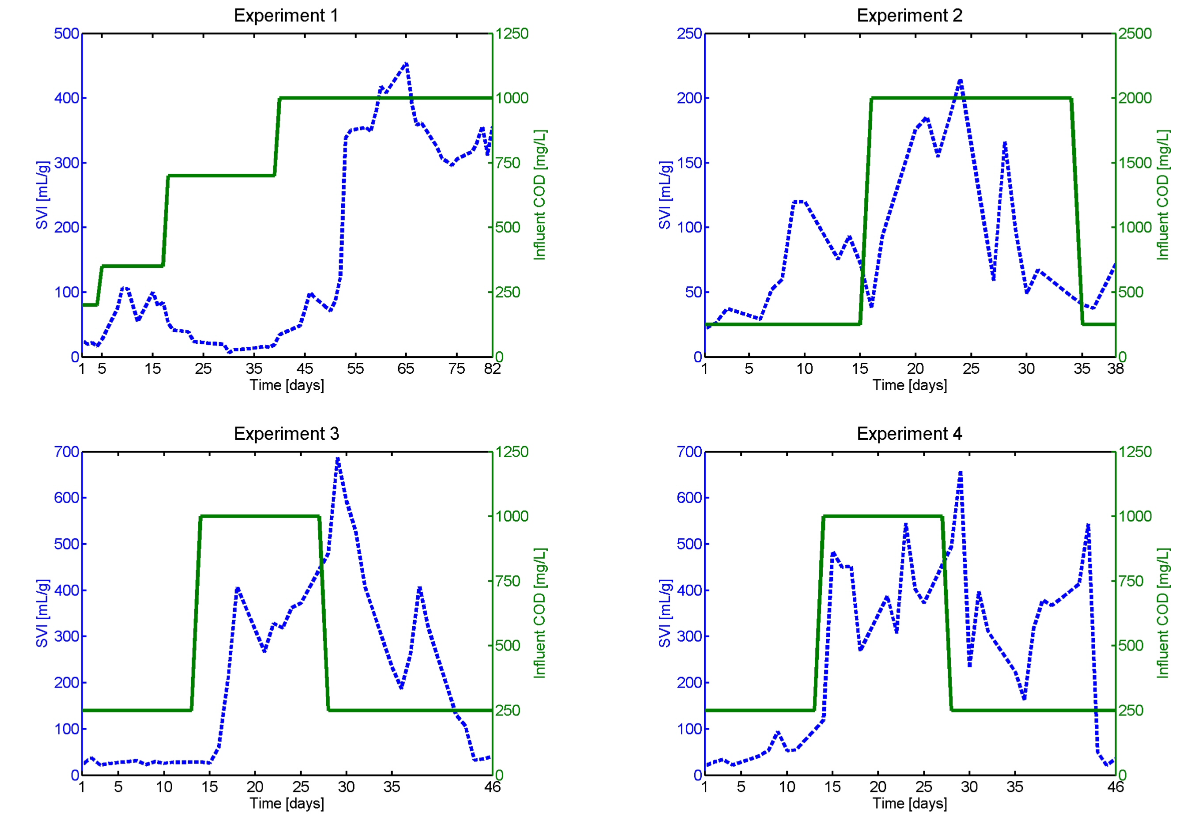

Fig. 1. SVI (- -) and COD loading (—) profiles during the four conducted experiments.

as a time saving and objective substitute for the classic SVI measurement.

2. MATERIALS AND METHODS

2.1 Experimental setup

A lab-scale activated sludge system was used to gather experimental data. The installation consists of an aeration tank with a capacity of 5 L, followed by a sedimentation tank with a volume of 3 L. From this sedimentation tank, the sludge is either recirculated to the aeration tank, or wasted. The sludge used in the experiments was obtained from the domestic wastewater treatment plant at Huldenberg (Belgium). A synthetic influent was fed to the lab-scale installation, with acetate acting as a carbon source in the first experiment and using glucose as substrate in the subsequent experiments.

To induce bulking, the Chemical Oxygen Demand (COD) loading was abruptly switched between a low (250 mg/L) and a high value (1000 or 2000 mg/L), except in the first experiment, where the COD loading was increased in smaller steps (from 200 to 1000 mg/L, passing through 350 and 700 mg/L). The exact COD loading profiles are given in Figure 1.

Daily measurements were performed on this system, such as the Sludge Volume Index (SVI) and Mixed Liquor Suspended Solids (MLSS), together with the Suspended Solids (SS) and COD of the effluent. In parallel, microscopic observations were performed to determine the sludge composition by means of image analysis. The observed SVI-profile is depicted in Figure 1.

2.2 Image acquisition equipment

Activated sludge images are acquired using an Olympus BX51 light microscope with a 10×10 magnification and equipped with a 3CCD color camera (Sony DXC-950P). The sample is subject to a phase contrast lighting in order to enhance the contrast between biomass and water. The images are sampled using the Carl Zeiss KS100 software, and compressed and stored in the JPG file format.

2.3 Image analysis and data processing software

In a next step, the stored images are processed and analyzed by means of the `MATLAB` *Image Processing Toolbox 3* (The Mathworks, Inc., Natick), according to the procedure described in Jenné *et al.* (2003). The processing and modelling of the

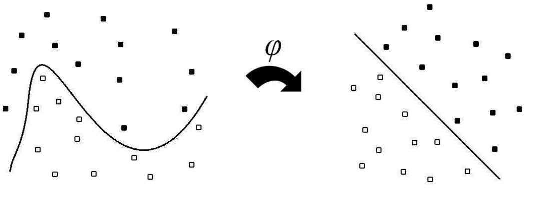

Fig. 2. Mapping of data points into the feature space, where a linear separator can be used.

resulting data, as described in Sections 4 and 5, is performed with the *LS-SVMlab 1.5* tool, a third-party `MATLAB` toolbox by Suykens *et al.* (2002).

3. LEAST SQUARES SUPPORT VECTOR MACHINES

Support vector machines (SVM) is a state-of-the-art method used for solving highly non-linear classification or modelling problems using linear techniques. The basis of the method is the mapping of all available data points to a feature space, thus transforming the problem into a simple linear problem. Least squares support vector machines (LS-SVM) express the training in terms of solving a linear set of equations instead of quadratic programming as for the standard SVM case.

3.1 Feature space mapping

The easiest way to define two classes of data points is using a simple linear separator. Unfortunately, very few classification problems can be solved this way. A possible solution is to map the data points $\{\mathbf{x}_k\}_{k=1}^N$ to a feature space, where a linear separator can be used, as illustrated in a qualitative way in Figure 2. This mapping is performed by the non-linear function $\varphi(\cdot)$, which is not explicitly known, but implicitly defined by satisfying the condition $\varphi(\mathbf{x}_i)\varphi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$, where $K(\cdot, \cdot)$ is called the kernel function. There are a few possible choices for the kernel function, but in the context of this paper the commonly used RBF kernel has been selected:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma^2}\right)$$

where σ is a parameter specifying the width of the kernel.

3.2 Classification function

Given a training set composed of N labelled data points $\{\mathbf{x}_k, y_k\}_{k=1}^N$, where $\mathbf{x}_k \in \mathbb{R}^m$ is the k-th input, and $y_k \in \mathbb{R}$ the corresponding class label ($y_k \in \{-1, 1\}$ in the binary case), a classification function is constructed as proposed by Suykens and Vandewalle (1999).

$$y(\mathbf{x}) = sign\left(\sum_{k=1}^{N} \alpha_k y_k K(\mathbf{x}, \mathbf{x}_k) + b\right)$$

where α_k are the support values and b is a real constant.

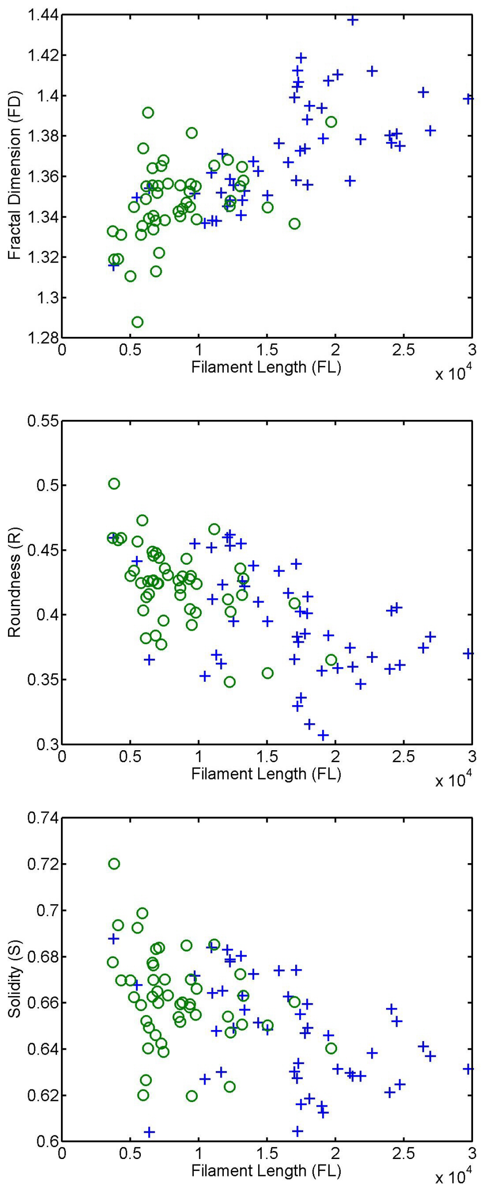

Fig. 3. Sample pairwise plots of the retained parameters. General bulking (+) and non-bulking (o) areas can easily be identified.

4. INPUT PARAMETER SELECTION

The image analysis algorithm offers 9 different variables for use as inputs to the classification function. Since some of these parameters are heavily correlated with others, the most relevant ones need to be selected before the classification function is trained.

The available parameters are divided into three categories. The first category consists of parameters related to the filament properties (filament length), the second category relates to the floc edges (convexity, form factor and fractal dimension) and the third to the general floc shape (aspect ratio, equivalent diameter, reduced radius of gyration, roundness and solidity). Next, the most significant parameters in the LS-SVM context for each category are selected within a Bayesian evidence framework. The retained variables are the filament length (FL), the fractal dimension (FD) and both the roundness (R) and solidity (S), which are all defined as in Jenné *et al.* (2004).

Visual inspection reveals that on pairwise plots of the retained parameters general bulking or non-bulking areas can easily be defined, as illustrated in Figure 3. This supports the assumption that the retained parameters are suitable for the construction of a classification function.

5. LS-SVM TRAINING

Because of their limited size, the four available data sets are combined into a training and a validation data set. For each possible arrangement of these data sets, the classification function is trained with a 10-fold cross validation. The number of misclassifications is used as a cost function, and the model parameters (the kernel parameter σ, the regularization parameter γ, the support values α_k and the constant b) are obtained through Bayesian optimization.

After training, the classification function is validated on the remaining data points, and the number of misclassifications is used as a performance measure.

Table 1. Performance issues when using the first experiment for training.

FL, FD, R			
Training experiments	Validation experiments	Misclass. during training	Misclass. during validation
1	2, 3, 4	2%	26%
2, 3, 4	1	13%	30%

FL, FD, S			
Training experiments	Validation experiments	Misclass. during training	Misclass. during validation
1	2, 3, 4	2%	26%
2, 3, 4	1	13%	49%

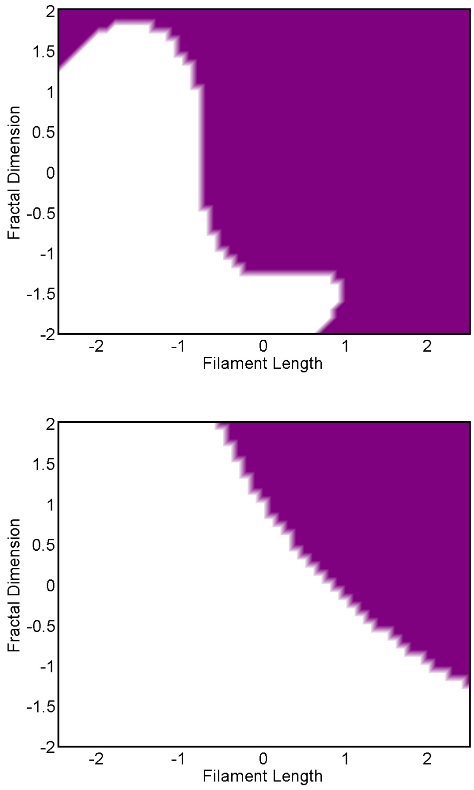

Fig. 4. Cut-through of the classification surface along $R = 0$, with the first experiment used as training data (top) and used as validation data (bottom). The clear difference between both views is indicative for the extrapolability problems with the first data set.

6. RESULTS

6.1 Data set selection

During the training of the classification function to distinguish between bulking and non-bulking situations in an activated sludge wastewater treatment plant, a large influence of the first experiment on the performance is noticed, with drastic differences between the cases where the first experiment is used as a training set and the cases where it is used as a validation set. This can be seen in Table 1, where the first data set's 2% misclassification rate when used for training jumps to 30% when the set is used for validation. When a cut-through of the classification boundary is made, the difference between both cases can clearly be observed, as illustrated in Figure 4. This significant difference in the classification function parameters leads to the conclusion that the first experiment is incompatible with the subsequent experiments. Because the difference between the first experiment on the one hand and the subsequent experiments on the other hand (different substrate and loading profiles) supports the in-

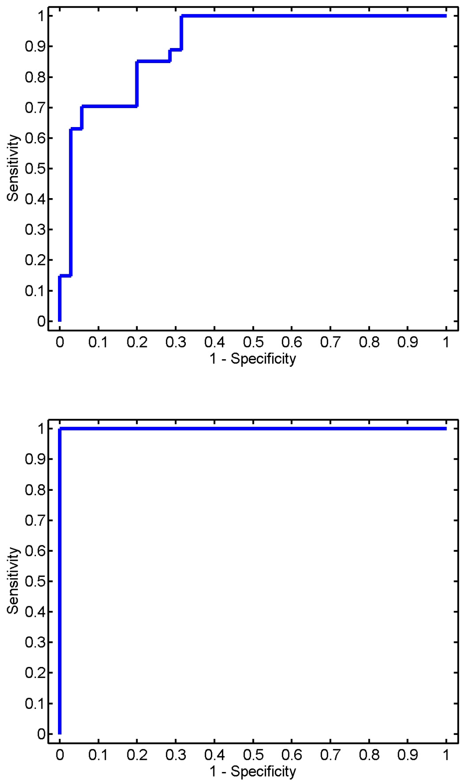

Fig. 5. ROC curve of the classification function trained on the second and fourth experiment for training (top) and validation (bottom), with an area of 0.92 and 1.00 respectively. These values indicate a well-performing classifier for the used data sets.

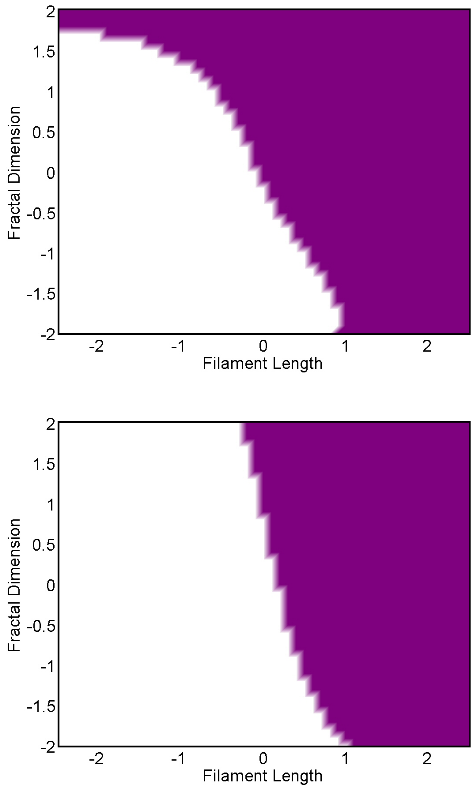

Fig. 6. Cut-through of the classification surface along $R = 0$ for the classifier trained (top) and validated (bottom) on the second and fourth experiment. Both cases result in similar views. Only the central area of the plots needs to be considered, as the edges of the view are less reliable due to extrapolation issues.

compatibility conclusion, the data from the first experiment are discarded, and training and validation is performed using the data from the second, third and fourth experiment only.

Table 2. Training and validation results after omitting the first experiment.

FL, FD, R			
Training experiments	Validation experiments	Misclass. during training	Misclass. during validation
2	3, 4	0%	32%
3	2, 4	0%	16%
4	2, 3	3%	42%
2, 3	4	8%	26%
2, 4	3	19%	3%
3, 4	2	9%	21%
FL, FD, S			
Training experiments	Validation experiments	Misclass. during training	Misclass. during validation
2	3, 4	17%	30%
3	2, 4	2%	18%
4	2, 3	17%	34%
2, 3	4	8%	21%
2, 4	3	17%	15%
3, 4	2	10%	21%

6.2 Classification results

After omitting the first experiment from the data, a new training is performed, with the results summarized in Table 2. These results also clearly demonstrate that, using the proposed approach, a good classifier to distinguish between bulking and non-bulking situations in an activated sludge wastewater treatment plant can be constructed. The best results are obtained when the classification function is trained on the second and fourth experiment and validated on the third experiment, using the roundness as floc parameter. Using the floc solidity instead of the roundness results in a classifier with a slightly worse performance. The good performance of this classifier can also be seen when the ROC-curves [2] are generated for both training and validation, with an area of 0.91 for the training and 1.00 for the validation, indicating a well-performing classifier

[2] A *Receiver Operator Characteristic* curve is a measure for the quality of a separator. The closer the spanned area is to 1, the better the classifier. If the area is 0.5, the classifier is worthless.

in both cases. These ROC-curves are illustrated in Figure 5.

Lastly, it is tested whether or not the best-case training and validation data sets are compatible. This is achieved by comparing the classifiers by making various cut-throughs of the classification surface, as also described in Section 6.1. The classification boundaries identified on the second and fourth experiment on the one hand and on the third experiment on the other hand show a very similar evolution, as illustrated in Figure 6. Only the central area of these plots needs to be considered, as the edges are less reliable due to extrapolation issues. This observation leads to the conclusion that the proposed grouping of experiments leads to data sets with a similar information content.

7. CONCLUSION

In this paper, a classifier for activated sludge image analysis data is constructed using an LS-SVM approach, in order to exploit these data and obtain an objective and time saving substitute for the classic SVI measurement. The selected image analysis parameters are the filament length, the floc fractal dimension, and either floc roundness or solidity. After a first training round, it is observed that the data from the first experiment are inconsistent with the data from the subsequent experiments, and training and validation are performed using data from these subsequent experiments only. After reviewing the results from this second classification training, it is observed that a valid classifier is obtained when the second and fourth experiment are used to compose the training data set, and the third experiment is used for validation. Another conclusion is that the best floc parameter is the floc roundness, with the solidity performing slightly worse. However, further experiments need to be conducted in order to confirm these conclusions.

ACKNOWLEDGEMENTS

Work supported in part by Project OT/03/30 of the Research Council of the Katholieke Universiteit Leuven and the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office. Ilse Smets and Annemie Geeraerd are postdoctoral fellows with the Fund for Scientific Research Flanders-Belgium (FWO-Vlaanderen). The scientific responsibility is assumed by its authors.

REFERENCES

Jenné, R., E.N. Banadda, N. Philips and J.F. Van Impe (2003). Image analysis as a monitoring tool for activated sludge properties in lab-scale installations. *Journal of Environmental Science and Health, Part A – Toxic/Hazardous Substances & Environmental Engineering* 38(**10**), 2009–2018.

Jenné, R., E.N. Banadda, I.Y. Smets, G. Gins, M. Mys and J.F. Van Impe (2004). Developing an early warning tool for filamentous bulking problems based on image analysis. In: G. Langergraber, S. Winkler, N. Fleischmann, A. Pressl and R. Haberl (Eds.), *Proceedings 2nd International IWA Conference on Automation in Water Quality Monitoring AutMoNet 2004*, 221–228, Vienna (Austria).

Suykens, J.A.K. and J. Vandewalle (1999). Least squares support vector machines. *Neural Processing Letters* **9**, 293–300.

Suykens, J.A.K., T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle (2002). *Least squares support vector machines.* World Scientific Publishing Company. Singapore.

Wanner, J. (1994). *Activated sludge bulking and foaming control.* Technomic Publishing Company. Lancaster.