Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Classification of Activated Sludge Settleability Using Linear and Nonlinear Classification Functions

Geert Gins * **Jef Vanlaer** * **Ilse Y. Smets** * **Jan F. Van Impe** *

* *BioTeC, Department of Chemical Engineering, Katholieke Universiteit Leuven, W. de Croylaan 46, B-3001 Leuven, Belgium.*
{geert.gins, jef.vanlaer, ilse.smets, jan.vanimpe}@ cit.kuleuven.be

Abstract: In this paper, two classifiers are proposed to distinguish between bulking and non-bulking situations in an activated sludge wastewater treatment plant, based on available image analysis information. The first classifier consists of a simple linear classification function, while the second classifier uses a highly nonlinear *least squares support vector machine* (LS-SVM) to distinguish between both situations. It is shown that the nonlinear LS-SVM classification function outperforms the linear classifier. Both exhibit identical misclassification rates, but fewer samples are located in the uncertainty area when using the nonlinear classifier. However, this better classification performance requires the identification of a substantial amount of model parameters, while the linear classifier is, except for the threshold values, parameterless.

1. INTRODUCTION

With the decreasing amounts of available freshwater and the ever increasing water pollution as a result of booming industrial development, there is a great need for efficiently working wastewater treatment systems. These systems are a crucial part of the water cycle, where wastewater is collected, purified, and -eventually- reused.

The most widely used wastewater treatment system is the activated sludge system, where a complex mixture of microorganisms is used to reduce the concentration of pollutants present in the wastewater to acceptable levels. The main problem in the operation of an activated sludge system is the occurrence of sedimentation failure [Wanner, 1994], which leads to an improper separation of the biomass and treated (purified) water. 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 and contamination of the environment. Ultimately, when the sedimentation problems persist, a complete loss of treatment capacity of the plant may occur [Wagner, 1984]. Therefore, the settleability of the activated sludge is closely monitored, and remedying actions are taken if a bad settleability is detected. However, these actions treat the symptoms of bulking without solving the underlying causes and are, therefore, non-preventive.

Because bulking is difficult to prevent, mathematical models capable of predicting the onset of bulking are required. Currently, apart from sludge settleability measurements, characterized by the *sludge volume index* (SVI) [1], occasional microscopic observations of the sludge composition are used to monitor the ratio of flocs to filaments. However, the on-site equipment for accurate microscopic observations is often not available. In addition, performing adequate microscopic observations is laborious and time-consuming. As a result, they are performed only periodically. Even when accurate microscopic observations are available, their interpretation is highly subjective, and heavily dependent on the operator's training level. Therefore, an image analysis procedure for the objective determination of various activated sludge characterization properties was developed in the K.U.Leuven/`BioTeC` research group [Cenens et al., 2002, Jenné et al., 2003, 2006].

For the modelling of activated sludge settleability using image analysis information with *black box* models, the potential of dynamic *autoregressive exogenous* (ARX) models is intensively investigated [Smets et al., 2006]. While the results during model identification are promising, the validation of these models on independent data sets often fails. Because the correct classification of digital sludge images in bulking and non-bulking situations is more important than the correct prediction of the actual SVI value for automated monitoring of an activated sludge wastewater treatment system, a classifier for digital activated sludge images is constructed in this paper.

In Section 2, the image analysis procedure used for the analysis of the activated sludge images is briefly described. Next, the experimental setup is detailed in Section 3, after which the data set used for the construction of the classifier is described in Section 4. The assessment of the quality of a classifier is explained in Section 5. A simple linear classification function is trained and validated in

⋆ Work supported in part by Projects OT/03/30 and EF/05/006 (Center-of-Excellence Optimization in Engineering) of the Research Council of the Katholieke Universiteit Leuven, and by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian Federal Science Policy Office. The scientific responsibility is assumed by its authors.

[1] The volume (in mL) occupied by 1 g of sludge after 30 minutes of sedimentation. Bulking occurs when the SVI exceeds 150 mL/g.

Section 6; a nonlinear classification function is constructed in Section 7. Finally, conclusions are drawn in Section 8.

2. IMAGE ANALYSIS PROCEDURE

A novel image analysis procedure for the objective determination of various activated sludge characterization properties has been developed in the K.U.Leuven/BioTeC research group [Cenens et al., 2002, Jenné et al., 2006].

After acquiring a digital activated sludge image using a microscope and digital camera, an *image segmentation* step separates the individual objects visible in the image from the background. Because the activated sludge images are obtained using *phase contrast* illumination, the intensity of the background is located between that of the flocs and the filaments. Hence, a single threshold procedure is incapable of separating flocs and filaments from the background in a single operation. Therefore, a series of image operations is performed on red and blue color layers of the digital image. In the resulting image, flocs and filaments are darker than the background, and they can be separated using a single threshold procedure.

Once the objects are extracted from the image, they are classified as flocs, filaments, or fragments using a series of size-, shape-, and intensity-related criteria in the *object recognition* step.

Finally, the morphological properties of the filaments and flocs visible in the image are characterized numerically using nine parameters. The aspect ratio AR, equivalent diameter D_{eq}, roundness R, form factor FF, convexity C, solidity S, reduced radius of gyration RG, and fractal dimension FD characterize the floc's size and shape, while the filament length FL provides a measure of the number of filamentous bacteria present in the sludge.

3. EXPERIMENTAL SETUP

Four experiments are conducted on a lab-scale continuous activated sludge system, consisting of a bioreactor vessel with a working volume of 5.5 L used as aeration tank and a custom-made glass cylindrical tank with a volume of 3 L as sedimentation tank. Using a peristaltic pump, the settled sludge is recirculated from the sedimentation to the aeration tank. A tap at the bottom of the sedimentation unit is used for periodic sludge wasting. The experiments have a duration of, respectively, 70, 38, 46, and 46 days.

At the start of each experiment, the aeration tank is filled with activated sludge obtained from the municipal wastewater treatment plant at Huldenberg (Belgium). The sludge is taken from the recycle of the plant, and diluted from 20 g/L to 3–5 g/L with clear effluent.

To ensure a sufficiently high oxygen concentration in the aeration tank, compressed air is supplied abundantly through aeration blocks. This results in a dissolved oxygen concentration of 5–7 mg O_2/L. In addition, this operation mixes the medium homogeneously, eliminating the need for any stirring devices.

A synthetic influent is used to feed the system at a flow rate of 5 L/day. This synthetic influent contains acetate as carbon source in the first experiment, and glucose in the subsequent experiments.

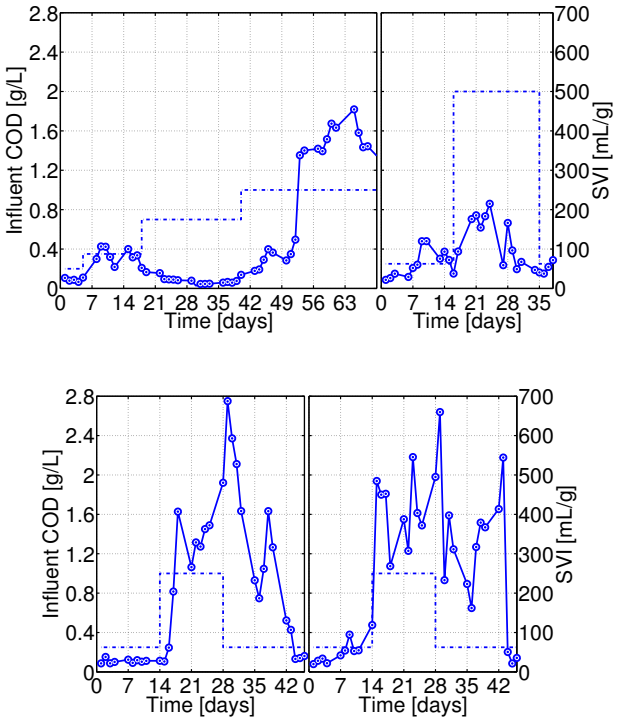


Fig. 1. Influent COD (–·–) and SVI (—) for the four experiments.

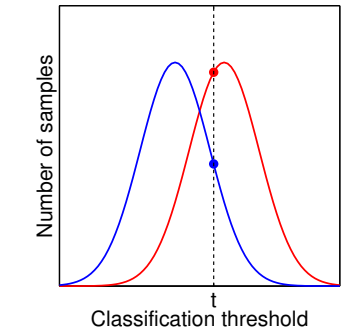
To induce bulking, the Chemical Oxygen Demand (COD) loading is abruptly switched between a low (250 mg/L) and a high value (1000 or 2000 mg/L), except in the first experiment, where the COD loading is increased in smaller steps (from 200 to 1000 mg/L, passing through 350 and 700 mg/L). The exact influent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

Using daily SVI measurements, the settleability of the activated sludge is monitored. The obtained SVI profiles are also provided in Figure 1. In parallel, microscopic observations are combined with the image analysis procedure to characterize the activated sludge composition.

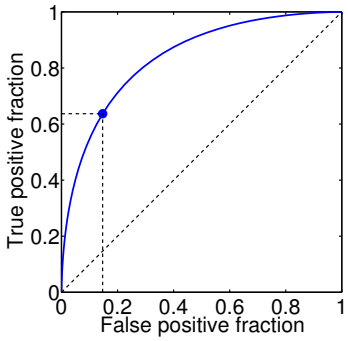
4. RESULTING DATA SET

After the completion of the lab-scale experiments, a data set of 200 data points is available, distributed over 4 sets of, respectively, 70, 38, 46, and 46 data points. These four experiments are used to construct training and validation data sets. Experiments 1, 2, and 3 compose the training set, and experiment 4 is used as validation data set.

For each data point, an SVI measurement is available, in addition to the nine parameters provided by the image analysis procedure. Based on the observations of Smets et al. [2006], where poor model validation results were observed, simple nonlinear transformations (i.e., $\sqrt{x}$, x^2, and $\ln(x)$) of these image analysis parameters are added to the input variable set. Additionally, all possible multiplicative and divisive combinations of the three most important of these nonlinear transformations (i.e., $\sqrt{FL}$, $\sqrt{D_{eq}}$, and $\ln(D_{eq})$) are included. This results in 52 transformations and/or combinations of image analysis parameters as possible classifier input variables.



(a)



(b)

Fig. 2. ROC curve construction.

5. CLASSIFIER QUALITY ASSESSMENT

The quality of a binary classifier is commonly evaluated through its *receiver operating characteristic* (ROC) curve [Green and Swets, 1966, Swets et al., 2000].

An ROC curve is constructed by varying a classification threshold t between $-\infty$ and $+\infty$, and comparing the *true positives fraction* (*TPF*, defined as the fraction of bulking images which are correctly classified), with the *false positives fraction* (*FPF*, the fraction of non-bulking sludge images which are misclassified as being bulking).

Initially, for a classification threshold value t of $-\infty$, *TPF* and *FPF* are both equal to one, as all digital sludge images are classified as bulking. When the threshold t is increased, as illustrated in Figure 2(a), *FPF* initially decreases faster than *TPF*, until both are zero. This leads to the ROC curve shown in Figure 2(b).

For two inseparable classes with identical histograms, *TPF* and *FPF* are always equal, resulting in an ROC curve equal to the diagonal of Figure 2(b). If both classes are perfectly separable, the *FPF* reaches zero while the *TPF* is still equal to one. This results in an ROC curve following the left and upper edges of the ROC plot.

Hence, the area under the ROC curve (*AuC*) numerically expresses the separation quality of a classifier. An *AuC* value of 0.5 corresponds with a worthless classification, whereas an *AuC* of 1 is obtained for a perfect separation.

Instead of *TPF* and *FPF*, the *sensitivity* and *specificity* are commonly used to express the classification performance for a given threshold t.

$$\text{sensitivity} \triangleq TPF \tag{1}$$

$$\text{specificity} \triangleq 1 - FPF \tag{2}$$

The optimal classification threshold t is commonly obtained from the ROC curve as the value of t for which sensitivity and specificity are equal. If a misclassification biased towards preventing either false positives or false negatives is required, the classification threshold t can be adjusted downwards or upwards, respectively.

6. LINEAR CLASSIFICATION FUNCTION

6.1 Description

The simplest method for the classification of sludge images corresponding with a bulking situation, is to impose a *single threshold* on the classification parameter x.

$$\hat{y} = \text{sign}\,(x - t) \tag{3}$$

All images for which the classification parameter x exceeds the specified threshold t yield a $\hat{y}$ of +1, and are classified as *bulking*. A value of -1 for $\hat{y}$ predicts a sludge image obtained from *non-bulking* sludge.

To decrease the misclassification rate, the single threshold is generalized to a *double-threshold* procedure.

$$\hat{y} = \text{sign}\Big(\text{sign}\,(x - t_l) + \text{sign}\,(x - t_u)\Big) \tag{4}$$

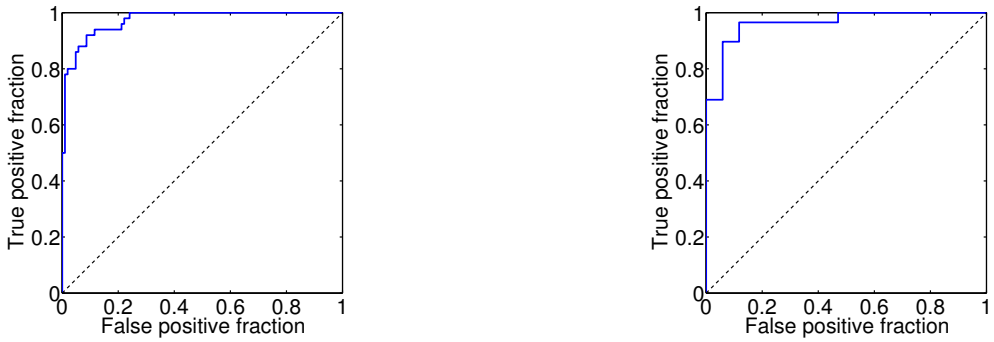
An activated sludge image with a classification parameter value x lower than the *lower threshold* t_l is identified as non-bulking, with a $\hat{y}$ value equal to -1. Similarly, a value of x higher than the *upper threshold* t_u leads to a $\hat{y}$ value of +1, and the classification of the image as bulking. If the classification parameter x is located in the uncertainty area between both thresholds t_l and t_u, $\hat{y}$ is exactly zero, and no classification is made.

As can be seen, the only parameter(s) in the linear classifier is/are the threshold values t, or t_l and t_u, for the single-threshold and double-threshold procedure, respectively.

6.2 Construction

For the selection of the optimal classification variable x, each of the available image analysis parameters is considered. After constructing the ROC curve for all potential classification variables, the corresponding *AuC* is computed. A maximal *AuC* value of 0.97, indicating a very good classifier performance, is obtained for $\sqrt{FL \cdot D_{eq}}$, which is therefore retained as the optimal classification variable x. The corresponding ROC curve is shown in Figure 3(a).

The single classification threshold t is obtained from Figure 3(a) as the value for which sensitivity and specificity are both equal to 91%. As listed in Table 1, this threshold value leads to a misclassification of 9% of the training images, with the *false positives* (i.e., the images obtained from non-bulking sludge which are classified as bulking) being more frequent than the *false negatives* (i.e., the bulking situations where a good settleability is predicted). Because the potential consequences of a missed bulking event are more severe than those of a misclassified good operation, as explained in Section 1, this is a valuable result.



(a) (b)

Fig. 3. ROC curves for the linear classification based on $\sqrt{FL \cdot D_{eq}}$ for (a) training ($AuC = 0.97$) and (b) validation ($AuC = 0.96$).

Table 1 also lists the training results for the double-threshold procedure. Similar to the determination of the optimal single threshold t, the ROC curve is used for the determination of the double-threshold values t_l and t_u. It can be seen that a sensitivity and specificity of 95% lead to a misclassification of approximately 5% of the images are misclassified. Fewer than 15% of the samples are located in the uncertainty area. Similar to the single-threshold results, the false positives are more frequent than the false negatives. Increasing the sensitivity and specificity to 99% results in a larger uncertainty area, in which 22% of the training batches are located. However, the misclassification rate drops to about 1%, with an equal number of false positives and false negatives.

It is clear that increasing sensitivity/specificity values lead to smaller misclassification rates, while the fraction of unclassified images, located in the uncertainty area, increases. Depending on the required operating performance, appropriate sensitivity and specificity values should be selected.

6.3 Validation

The linear classification function is validated on the 46 measurements of experiment 4, the validation data set. First, the ROC curve for the validation set is constructed (Figure 3(b)), and an AuC of 0.96 is obtained. This very high AuC value is nearly identical to the training AuC of 0.97 and indicates a very good separability of the bulking and non-bulking images in the validation set based on the $\sqrt{FL \cdot D_{eq}}$ parameter.

However, a high AuC is no guarantee for a good validation, as the selectivity and specificity for a specific threshold t might differ significantly between training and validation. Therefore, Table 1 compares the misclassification and uncertainty rates during training and validation for both single- and double-threshold classification. While the misclassification rates for the single threshold during validation are slightly higher than those observed during training, the difference is mainly caused by a higher false positive rate; the false negative rate for both training and validation is identical. For the double-threshold classification, the validation misclassification and uncertainty rates are nearly identical to the training results.

This leads to the conclusion that the simple linear classification function constructed in Section 6.2 is capable of accurately labelling the activated sludge images as either bulking or non-bulking, and is therefore suited for use in an automated activated sludge wastewater treatment monitoring system.

7. NONLINEAR CLASSIFICATION FUNCTION

7.1 Description

To further improve the quality of the linear classifier constructed in Section 6, the potential of a nonlinear classification function is investigated. Hereto, a *least-squares support vector machine* [LS-SVM, Suykens et al., 2002] structure is selected.

$$\hat{y} = \text{sign}\left(\sum_{n=1}^{N} \left(\alpha_n y_n K\left(\mathbf{x}, \mathbf{x}_n\right) + b\right) - t\right) \tag{5}$$

For a training set of N samples $\mathbf{x}_n$ with corresponding class labels y_n, those samples with nonzero α_n values are called the *support vectors*. The parameter b is a real constant.

An LS-SVM is capable of solving highly nonlinear classification problems using the *kernel function* K, for which the most common choice is the Gaussian (or radial basis function) kernel, with *kernel width* σ^2.

$$K\left(\mathbf{x}, \mathbf{x}_n\right) = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}_n||^2}{\sigma^2}\right) \tag{6}$$

Similar to the linear classifier, the misclassification rate of the LS-SVM can be decreased by using a double-threshold procedure.

In contrast with the linear classification function described in Section 6, the construction of an LS-SVM classifier requires the identification of the support values α_n, constant b and kernel width σ^2 in addition to the determination of the single- or double-threshold values, respectively t, and t_l and t_u. Furthermore, a *regularization parameter* γ must be identified for the penalization of misclassifications.

For the identification of these parameters, the LS-SVMlab MATLAB® toolbox [Suykens et al., 2002] is used. The optimal model and regularization parameters are obtained through optimization in a *Bayesian evidence framework* [Van Gestel et al., 2002].

Table 1. Linear classification results for training and validation.

Training			
Sensitivity & specificity	91%	95%	99%
Misclassification	14 (9.1%)	8 (5.2%)	2 (1.3%)
false positive	*9 (5.8%)*	*5 (3.2%)*	*1 (0.7%)*
false negative	*5 (3.2%)*	*3 (2.0%)*	*1 (0.7%)*
Uncertain	—	21 (13.6%)	34 (22.1%)
Validation			
Sensitivity & specificity	91%	95%	99%
Misclassification	6 (13.0%)	4 (8.7%)	1 (2.2%)
false positive	*5 (10.9%)*	*3 (6.5%)*	*1 (2.2%)*
false negative	*1 (2.2%)*	*1 (2.2%)*	*0 (0.0%)*
Uncertain	—	4 (8.7%)	10 (21.7%)

Table 2. LS-SVM classifier input variables.

FL	FL^2	$\sqrt{FL \cdot D_{eq}}$
$\sqrt{FL}$	$\sqrt{D_{eq}}$	$\sqrt{FL} \cdot \ln(D_{eq})$
$\ln(FL)$	$\ln(D_{eq})$	$\ln(D_{eq}) \cdot \sqrt{D_{eq}}/\sqrt{FL}$

7.2 Construction

In a first step, the optimal subset of the 52 available image analysis parameters and their nonlinear transformations and combinations is identified. Via *automatic relevance detection* in a Bayesian framework [Suykens et al., 2002], 47 input variables are retained.

This very high number of selected input variables is retained because an approximately equal importance is assigned to all available image analysis parameters and their simple nonlinear transformations and combinations. Therefore, the number of candidate input variables is reduced manually, and the variable selection is repeated. This results in an input set consisting of the variables listed in Table 2.

The optimal input variables are all derived from the filament length FL and/or the equivalent diameter D_{eq}. These parameters provide information on both the amount of filamentous bacteria present in the activated sludge and the shape of the sludge flocs, factors heavily influencing the sludge settleability. However, because these nonlinear transformations are arbitrarily chosen, a physical interpretation is impossible.

After identification of the LS-SVM parameters α_n, b, σ^2 and γ, an AuC of 0.99 is obtained, indicating a nearly perfect separation of the bulking and non-bulking images. The corresponding ROC curve is depicted in Figure 4(a).

Finally, the single and double classification thresholds t, t_l and t_u are determined, and the classifier performance is summarized in Table 3. For the single-threshold procedure, a sensitivity and specificity of 94% are obtained. This leads to a misclassification rate of 6% of the training data. As was the case with the linear classifier in Section 6, the false positives are approximately twice as frequent as the false negatives.

A selectivity and specificity of 95% is achieved using a double threshold. Here, 4% of the training images are located in the uncertainty area, and 5% are misclassified. Again, the false positives outnumber the false negatives, a valuable result. Increasing the selectivity and specificity to 99% decreased the misclassification rate to 1%. However, the uncertainty area expands, and contains 14% of the training images. As a result of the very small amount of misclassifications, the observed false positive and false negative rates are equal.

7.3 Validation

The obtained LS-SVM classifier is applied to the measurements of the validation set. For the validation data set, the AuC for the ROC curve given in Figure 4(b) is 0.97, only slightly lower than the training AuC of 0.99. This indicates a very good separability of the bulking and non-bulking activated sludge situations.

As shown in Table 3, the validation misclassification and uncertainty rates are identical to the training values for both single- and double-threshold classifiers, provided the limited resolution due to the relatively small data set size.

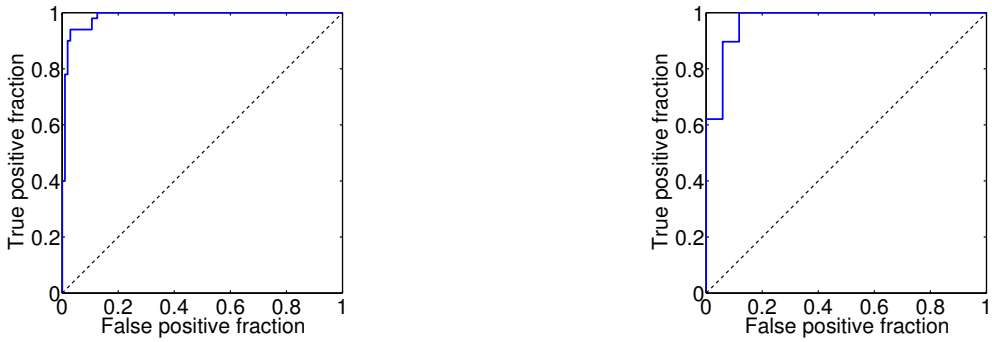
When comparing the training misclassification and uncertainty rates of the nonlinear LS-SVM classifier with those of the linear classifier (Table 1), it is evident that the nonlinear classifier provides similar misclassification rates. The number of uncertain classifications, however, is significantly reduced.

It is concluded that the identified LS-SVM classification function provides an accurate identification of bulking and non-bulking situations using nonlinear transformations and combinations of two image analysis parameters: the filament length FL and equivalent diameter D_{eq}.

8. CONCLUSIONS

In this paper, one linear and one nonlinear classifier for activated sludge digital images is constructed. Using image analysis parameters as inputs, these classifiers identify bulking and non-bulking sludge conditions.

The simple linear classifier provides a very good separation between both classes during training. Using a single-threshold classification, only 10% of the training samples are misclassified. With the introduction of an uncertainty area by means of a double-threshold classification, the misclassification rate decreases to 5% and 1% for, respectively, a sensitivity and specificity of 95% and 99%, with corresponding uncertainty rates of 14% and 22%. Of the misclassifications, the number of false positives (i.e., the classification of non-bulking sludge as bulking) is



(a) (b)

Fig. 4. ROC curves for the nonlinear classification for (a) training ($AuC = 0.99$) and (b) validation ($AuC = 0.97$).

Table 3. LS-SVM classification results for training and validation.

Training			
Sensitivity & specificity	94%	95%	99%
Misclassification	9 (5.8%)	8 (5.2%)	2 (1.3%)
false positive	*6 (3.9%)*	*5 (3.2%)*	*1 (0.7%)*
false negative	*3 (2.0%)*	*3 (2.0%)*	*1 (0.7%)*
Uncertain	—	6 (3.9%)	22 (14.3%)
Validation			
Sensitivity & specificity	94%	95%	99%
Misclassification	3 (6.5%)	2 (4.3%)	1 (2.2%)
false positive	*2 (4.3%)*	*2 (4.3%)*	*1 (2.2%)*
false negative	*1 (2.2%)*	*0 (0.0%)*	*0 (0.0%)*
Uncertain	—	2 (4.3%)	7 (15.2%)

higher than the number of false negatives (i.e., the bulking situations where a good settleability is nevertheless predicted). For use in an automated monitoring tool, this is a valuable result, as false negatives have potentially worse consequences.

The second classifier, using a nonlinear LS-SVM model structure with nine input variables, achieves near perfect classification during training. With a single threshold, a misclassification rate of 6% is achieved. With a double threshold, and 95% sensitivity/specificity, 5% of the training samples are misclassified, while 4% are located in the uncertainty area. Increasing the sensitivity/specificity to 99%, these values change to 1% misclassification and 22% uncertain classifications. Again, the false positives outnumber the false negatives.

It is therefore concluded that the nonlinear LS-SVM classification function outperforms the linear classifier. While both models exhibit identical misclassification rates in the double-threshold case, the LS-SVM achieves a lower uncertainty rate. However, this better classification performance requires the identification of a substantial amount of model parameters, while the linear classifier is, except for the threshold values, parameterless.

REFERENCES

C. Cenens, K.P. Van Beurden, R. Jenné, and J.F. Van Impe. On the development of a novel image analysis technique to distinguish between flocs and filaments in activated sludge images. *Water Science and Technology*, 46(1–2):381–387, 2002.

D.M. Green and J.M. Swets. *Signal detection theory and psychophysics*. John Wiley & Sons, New York (USA), 1966.

R. Jenné, E.N. Banadda, N. Philips, and J.F. Van Impe. 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 and Environmental Engineering*, A38(10): 2009–2018, 2003.

R. Jenné, E.N. Banadda, G. Gins, J. Deurinck, I.Y. Smets, A.H. Geeraerd, and J.F. Van Impe. The use of image analysis for sludge characterisation: studying the relation between floc shape and sludge settleability. *Water Science and Technology*, 54:167–174, 2006.

T. Van Gestel, J.A.K. Suykens, G. Lanckriet, A. Lambrechts, B. De Moor, and J. Vandewalle. A Bayesian framework for least squares support vector machines classifiers, Gaussian processes and fisher discriminant analysis. *Neural Computation*, 14(5):1115–1147, 2002.

I.Y. Smets, E.N. Banadda, J. Deurinck, N. Renders, and J.F. Van Impe. Dynamic modelling of filamentous bulking in lab-scale activated sludge processes. *Journal of Process Control*, 16:313–319, 2006.

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

J. Swets, R. Dawes, and J. Monahan. Better decisions through science. *Scientific American (October 2000)*, pages 82–86, 2000.

F. Wagner. Studies on the causes and prevention of bulking sludge in Germany. *Water Science and Technology*, 16:1–14, 1984.

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