Predicting the effluent quality of an industrial wastewater treatment plant by way of optical monitoring

Tomperi Jani[a*], Koivuranta Elisa[b], Leiviskä Kauko[a]

[a]Control Engineering, University of Oulu. Oulu, Finland.
[b]Fibre and Particle Engineering, University of Oulu. Oulu, Finland.
[*]Corresponding author: e-mail jani.tomperi@oulu.fi. University of Oulu, Control Engineering, P.O. Box 4300, FI-90014 University of Oulu, Finland.

Abstract
Wastewater samples taken from the aeration tank of a full-scale activated sludge plant were analyzed using an automatic optical monitoring device. Five variable selection methods were utilized to find the optimal subsets of input variables to develop predictive models for the important parameters of the wastewater treatment process efficiency and the quality of the effluent, including suspended solids, biochemical oxygen demand, chemical oxygen demand, total nitrogen and total phosphorus. The dependencies between the selected variables were also inspected. The study showed that the models based solely on the optical monitoring variables can be used to predict the level of the effluent quality parameters hours before the traditional sampling and analyses. Thus, predictive modelling based on the optical monitoring variables is a potential tool to be used assistance in a process control, keeping the process in a stable operating condition and avoiding environmental risks and economic losses.

Keywords: activated sludge process; cross-validation; digital imaging; modelling; pulp and paper industry; variable selection

1. Introduction
In pulp and paper industry, water is not only used as raw material but also for cooling and lubricating the machines. Thus huge amount of water flow through the processes and the amount of produced wastewater is large even though for economic and environmental reasons water is recycled as much as possible. For example, Finnish Forest Industries [1] states that nowadays in Finland from five to 50 cubic meters of water is used to produce one metric ton of pulp, and producing one metric ton of paper from seven to 15 cubic meters of water is consumed. The wastewaters of a pulp and paper industry are most commonly treated in a complex biological process, an activated sludge plant (ASP), where several physical, chemical and microbiological factors simultaneously affect the purification process. The operation of the treatment process depends on the condition of the biomass which is very sensitive to internal and external disturbances. A good bacterial balance enables the high pollution removal rate, low suspended solids in the effluent and a good settling properties of the sludge, which are considered as the key elements of an efficiently operating process. However, the quantity and quality of the incoming wastewater may vary drastically especially when wastewaters from other processes in the mill area (for example chemical processes) are led to the same wastewater treatment process. The major changes affect the condition of biomass detrimentally and because the recovery from the disturbances is slow the effects on the process operation and purification results are long-lasting. The disturbances in the bacterial balance will most often be shown as dysfunctional flocculation and settling properties which may have serious environmental and economic effects. [2]

The main goal of a wastewater treatment is to remove organic compounds, excessive nutrients and toxicants from the treated water that can be reused or discharged to waterways. Stringent limitations to effluent discharges defined by authorities and constantly rising operating costs force the wastewater treatment plant operation to be optimized. However, an advanced control requires accurate monitoring of the process. A modern activated sludge plant has a wide range of on-line and off-line measurements but the common process measurements do not include sufficient information on the special features of the wastewater or give an early-warning information on the process efficiency and the effluent quality. Biological oxygen demand (BOD) and chemical oxygen demand (COD) are measures of the amount of dissolved oxygen required to oxidize the organic substances in wastewater. Nitrogen (N) and phosphorus (P) are useful nutrients but as large amounts in discharged wastewaters they cause eutrophication. Together with suspended solids (SS) the above-mentioned parameters are often used to indicate the effluent quality and the efficiency of a wastewater treatment process. However, measuring SS, a sludge volume index (SVI), nutrients content, or amount of organic substances from the effluent only show the prevailing quality of the wastewater. On this account, there is a demand for using new on-line monitoring tools and methods for assessing the process and predicting the upcoming quality of the effluent. Optical monitoring of floc morphological characteristics is a potential tool to be used as assistance in process control as it gives fast objective information about the quality of wastewater and the state of the treatment process, reveals some of the reason for settling problems and combined to a predictive modelling shows the quality of effluent in advance hours before it is noticed by traditional process measurements. [3, 2, 4-8]

In this study, the results of the automatic optical monitoring of wastewater samples taken from a full-scale industrial active sludge process during a period over one year were utilized to develop predictive models for the effluent quality parameters (suspended solids, biological oxygen demand, chemical oxygen demand, nitrogen and phosphorus). Five variable selection methods were used for selecting the optimal subsets of input variables for each developed model. The study included also a short inspection of dependencies between the selected optical monitoring variables and the quality parameters.

2. Material and Methods

2.1 Data

The data for this study was collected from the activated sludge process of a Finnish pulp and paper mill that treats in addition to the pulp making processes wastewaters from two chemical processes located in the mill area. The average wastewater flow through the treatment process is around 30 000 m^3/day. The unit operations of the wastewater treatment process are intake, screening, preliminary settling, neutralization, aeration, secondary settling and discharge. Wastewater samples for the optical monitoring were taken from the aeration tank and analyzed on the same working day within a couple of hours after sampling. Hydraulic delay in the aeration tank is 20 hours at the average flow and in the secondary settling tank around 10 hours. The dataset included optical monitoring results and selected process measurements including the effluent quality parameters from a period of 13 consecutive months. The optical monitoring was carried out by a beforehand planed schedule which consisted of regular but sparse sampling periods and more frequent sampling periods when wastewater samples were monitored nearly daily. In overall, dataset included 54 measurement times.

Before modelling, the dataset including several different type of measurements was scaled between [-2, 2] using the nonlinear scaling method based on generalized moments, norms and skewness [9]. Before the scaling, the dataset was inspected and missing values were replaced with interpolation.

2.2 Optical monitoring and image analysis

To replace a laborious, slow and subjective method to study wastewater samples manually using a microscope, a small-scale automatic optical monitoring device was designed [10] and applied to analyze the wastewater samples taken from the aeration tank of the industrial ASP. Wastewater samples were diluted with deionized water at a ratio of 1:200 and pumped through a cuvette which was imaged with a high resolution charge-coupled device (CCD) camera. The CCD camera image sensor was 5.0 mm * 3.7 mm (1392 * 1040 pixels) with a pixel size of 3.6 μm * 3.6 μm. One video from a wastewater sample contained approximately 250-350 images.

The developed imaging system measures and analyses several morphological features of the flocs and filaments. Image processing and analyses methods and the mathematical formulas of the calculated size and shape parameters are presented in details in Koivuranta et al. [10]. For this study, the number of analyzed parameters were limited to the most suitable ones based on the preliminary laboratory studies carried out during the development period of the optical monitoring device. The parameters were calculated as an average of the values for individual objects over a single image. In addition to size parameters such as equivalent diameter, floc area and filament length, the calculated shape parameters included for example fractal dimension, form factor and roundness. The equivalent diameter is the diameter of a circle with an area equal to the object's area. The form factor is affected by the irregularity or roughness of the object's boundary and it is 1.0 for a perfect circle and below 1.0 for any other shape because objects with more irregular boundaries have a longer perimeter per surface area and therefore have smaller form factors. Roundness is defined as the ratio between the area of an object and the area of a circle with a diameter equal to the object's length. Roundness is also 1.0 for a perfect circle. The aspect ratio, which varies from 1.0 (for a perfect circle) to infinity, describes how elongated an object is. [11]

In the following results, the amount of filaments is presented as a ratio of total filament length and floc area, where the total filament length is the sum of filament length of all the individuals present in the image. The number of small objects was calculated based on the size distribution, where each object was assigned to a size category based on its equivalent diameter. The size distribution was calculated as the sum of the distributions of individual images. In the image analysis, the threshold value for floc area was 100 μm^2 because the boundaries of the smaller objects may not have been sharp enough due to the resolution of the camera. The limit value for the small objects was equivalent diameter of 25 μm.

2.3 Variable selection

Variable selection is a practical way to reduce the amount of variables available and to choose the optimal input variables of a predictive model from a large dataset. Input variables that include noise, are correlated to each other or have no significant relationship with the output variable increase the computational complexity and reduce the prediction result of the model. In model development, a good principle is to keep the amount of input variables decent. Using too many input variables increase the risk to develop an over-fitted model which has an excellent training result but is not usable with a new upcoming data. In this work, five variable selection methods (correlation based selection, forward selection, stepwise selection, a genetic algorithm (GA) and a successive projections algorithm (SPA)) were utilized to find the optimal subsets of input variables for modelling the SS, BOD, COD, total nitrogen and total phosphorus in the effluent.

Variable selection methods can be roughly grouped into wrapper and filter methods. In a filter method, variables are selected or deleted according to the formed ranking which is based on the correlation coefficients. Filter methods are very efficient but the developed model is seldom optimal. In a wrapper method, a subset of variables is assessed according to their usefulness to a given predictor. Wrapper methods wrap around an appropriate learning machine which is employed as the evaluation criterion, such as prediction or classification error. Wrappers often give better results but are slower than filters. Correlation-based selection and a successive projections algorithm are classified as filters and forward selection and a genetic algorithm are classified as wrappers. [12, 13] For a very large dataset one variable selection method can be used for the variable elimination before the final variable selection by another method. In Sorsa et al. [14] a successive projections algorithm was found to greatly improve the reliability of the genetic algorithm search. The more detailed backgrounds of the above mentioned variable selection methods are presented for example in Tomperi et al. [8].

2.4 Modelling

The quality of the developed model depends highly on the quality and length of the dataset. Data should include a sufficient number of samples and it should also be fully representative of the full spectrum of all possible conditions. For example, in environmental related processes the source dataset should encompass at least one full year of measured data to ensure all seasonal effects are included in data. In model development, efficient training and validation require long and representative subsets of data for both. Due to the small size of the dataset available in this study, a static split into the training and validation subsets of data was not an effective approach.

A cross-validation (leave one out (LOO), leave multiple out (LMO) or *k*-fold) is an efficient resampling method to predict the fit of a model for a validation set when dataset is small and an explicit validation set is not available. In cross-validation, the whole data set is used for training and validating the model by using part of the data for training and the rest of the data for validation and repeating this until the whole dataset is processed. Thus the largest possible subset can be used for both training and validation. In this study, a five-fold cross-validation was used. In *k*-fold cross-validation the original dataset is randomly partitioned into *k* subsets of equal size. One subset is used as a validation data for testing the model and the remaining *k*–1 subsamples are used as training data. The cross-validation process is repeated *k* times and each of the subsets is used only once as the validation data. A single estimation is then produced by combining these *k* results of the folds. Optimal *k* is reported to be between five and ten folds because statistical performance does not increase notably for larger values of *k*, and averaging over less than ten splits is computationally feasible. A multivariable linear regression (MLR) was utilized to predict the output variable as a linear combination of selected input variables and the performance of the model was evaluated by coefficient of determination (R^2) and Root Mean Square Error (RMSE). R^2 and RMSE values were calculated as an average of repeating the validation procedure five times for each model. [15, 16].

3. Results and Discussion

In this work, the results of the automatic optical monitoring of wastewater samples taken from the ASP located at a pulp mill were utilized in developing predictive models for the quality parameters of treated wastewater. The study included also an inspection of dependencies between the optical monitoring variables, the quality parameters and the selected process variables. The optimal subsets of input variables for the models were selected using five variable selection methods that did not take into account any deterministic models or chemical or biological knowledge about the activated sludge process but selections were performed based on mathematical ground only. Without presumptions the

methods are more generalizable and easily usable in different process cases but it is important to bear in mind that the results based solely on a mathematical analysis may not accurately correspond the actual situation in the wastewater treatment process. The complexity of a wastewater treatment process easily causes quasi-correlations. A high correlation between variables not always mean strong real-world causality and there may be also many hidden factors that affect in the real process but are not shown in the data analysis due to the limited amount of data or measurements and due to the analysis method.

Part of the data (34 samples) used in this study was comprehensively analyzed in Koivuranta et al. [5] where the image analysis results were compared with the effluent clarity during a three-month period of frequent sampling. The period of data collection included a maintenance stoppage at the wastewater treatment plant when the aeration tank was emptied but a small amount of activated sludge was left at the bottom of the tank to ensure a faster restart of the treatment process. Immediately after the restart a settling problem was discovered. The earlier study of Koivuranta et al. [5] suggested that based on the optical monitoring the settling problem was most likely caused by dispersed growth.

In Figure 1, the quality parameters of the effluent and the optical monitoring variables measured during 13 months are presented. The temperature, sludge age and influent flow are considered important operation conditions that affect the process efficiency and are also presented in Figure 1. The temperature and influent flow are not controllable or related to the season of the year as they are related to the operation of the processes in the mill area. The temperature of wastewater from pulp and paper industry is often very high but the temperature of the influent is stabilized in the early part of the treatment plant and kept as stable as possible to ensure the efficient function of the treatment process. Temperature drops basically only when the flow decreases or totally stops. In Figure 1, the restart after the above mentioned process maintenance stoppage is located at data point #30. Before the maintenance stoppage the quality of the treated wastewater was good and the suspended solids content was notably lower than immediately after the restart. In the ASP in question, the limit value for a high level of suspended solids content is 40 mg/l. Nitrogen content also increased drastically after the stoppage. The change in the treated water quality after the restart is clearly shown also in optical monitoring variables. The amount of filaments (filament length/floc area) was lower than before the stoppage and the area of the flocs was also lower but slowly increased after the restart, as did the roundness and form factor. Aspect ratio, on the contrary, was high and slowly decreased after the restart. Based on the optical monitoring the number of small particles was lower and the flocs were clearly larger and more regular in boundaries when the quality of treated wastewater was good.

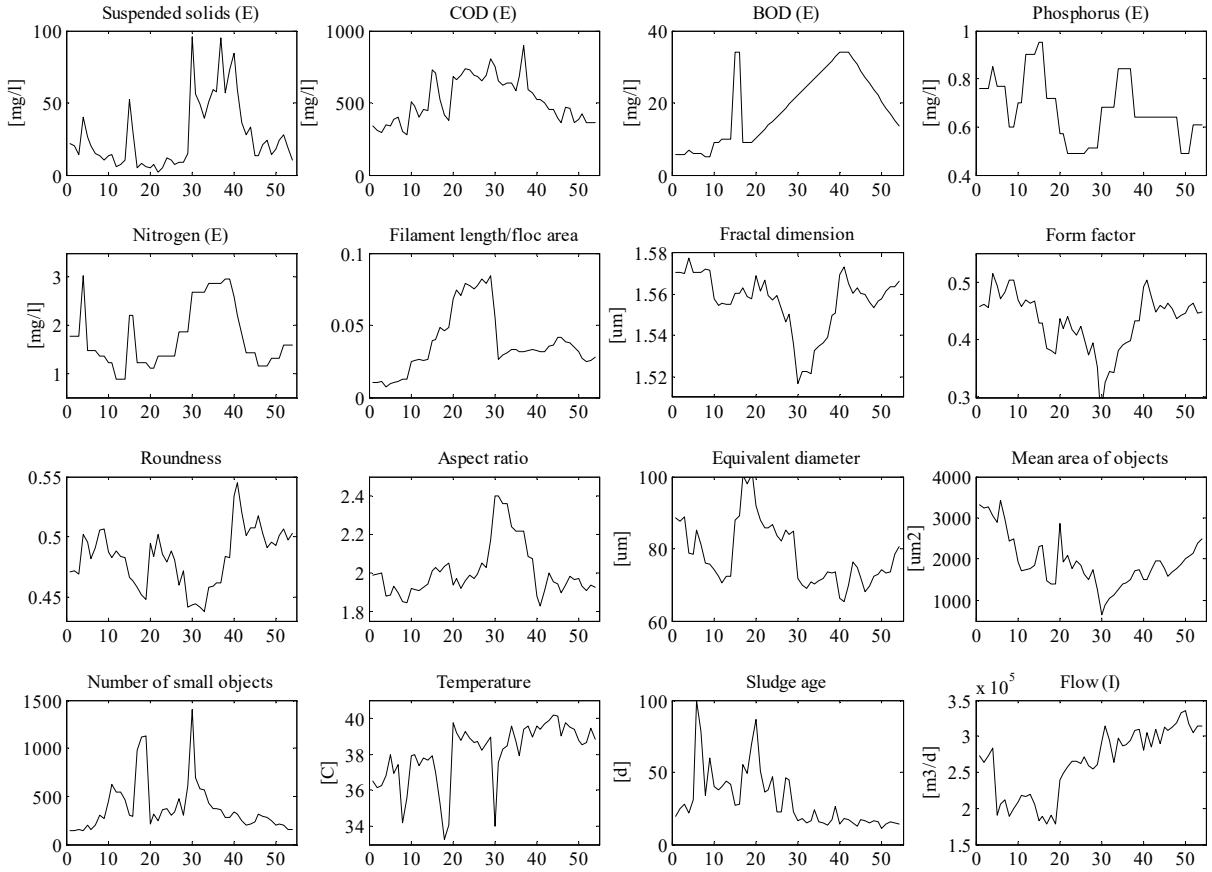

Figure 1. The quality parameters of the effluent, the optical monitoring variables and selected operating conditions of the industrial ASP during the study period.

In Table 1, the results of the linear correlation analysis for the above mentioned variables are presented. Only the correlation coefficients over |0.30| are listed. The results of the linear correlation analysis of the data from (supposedly) a nonlinear process may not be absolutely correct but they are approximate. COD and BOD have the most and the highest correlation coefficients to the optical monitoring variables yet none of the coefficients are relatively high. Suspended solids in effluent correlates with other quality parameters, phosphorus, nitrogen and BOD, which again correlates with COD. Suspended solids has a reasonable high negative correlation with the sludge age but the negative correlation coefficients with the optical monitoring variables, amount of filaments and equivalent diameter, are lower. BOD and COD, on the other hand, have more and higher correlations with the optical monitoring variables. Based on this data analysis, when the BOD and COD content in the effluent is high, fractal dimension and area of objects are low, and the aspect ratio and amount of filaments are high. Form factor, roundness and number of all objects have high correlation with COD and probably would have also with BOD if the amount of BOD data would have been greater and no interpolation would have needed in the data pretreatment stage. During a high flow the temperature of the wastewater increases and the sludge age decreases which is understandable in the industrial active sludge process. As the flow increases, the equivalent diameter of the objects decreases and when the temperature is higher, the roundness of measured flocs is higher.

In the previous studies, a municipal ASP were studied with the similar imaging system and it was noticed that the increasing flow decreased temperature [7, 8]. It was also discovered that the suspended solids and other quality parameters of treated wastewater had a positive correlation coefficient with the amount of filaments and number of small objects and negative correlation with roundness and fractal dimension among others. When the quality of the treated wastewater deteriorated the amount of filaments notably increased, the roundness of the flocs decreased and number of the objects increased. According to the data analyses the municipal process is more temperature related than the industrial. The differences between the results of the municipal and the industrial cases are most likely due to the nature of the processes. Based on the optical monitoring of the treatment process, in the industrial ASP the settling problem was caused by dispersed growth [5] and in the municipal ASP the poor settling was caused by filamentous bulking [6].

Table 1. The correlation coefficients of the effluent quality parameters, the optical monitoring variables and the selected operating condition parameters.

	SS	COD	BOD	P	N	Temp	Sludge age	Flow
Suspended solids			0.63	0.34	0.76		-0.61	0.45
COD			0.55					
BOD	0.63	0.55			0.45	0.64	-0.66	0.52
Phosphorus	0.34					-0.43		-0.33
Nitrogen	0.76		0.45				-0.36	
Filaments	-0.30	0.69	0.40	-0.48		0.35		
Fractal dimension		-0.58	-0.42					
Form factor		-0.66						
Roundness		-0.45		-0.35		0.45		
Aspect ratio		0.53	0.34		0.44			
Equivalent diameter	-0.52		-0.54			-0.36	0.56	-0.54
Area of all objects		-0.52	-0.49					
Number of small objects		0.53						-0.30
Temperature			0.64	-0.43			-0.48	0.63
Sludge age	-0.61		-0.66		-0.36	-0.48		-0.80
Flow	0.45		0.52	-0.33		0.63	-0.80	

The selected input variables for developing predictive models of treated wastewater quality parameters are presented in Table 2. Variables are listed in the order of importance (the order of selection) and all the selected variables were used as input variables in the developed models whose performances are presented in Table 3. The number of input variables in every model was decent, between two and five, which reduces the risk of overfitting. Variable selection showed that certain optical monitoring variables are important in modelling a particular quality parameter. For example, equivalent diameter, fractal dimension and aspect ratio were selected by almost every selection method as input variables to suspended solids model, mean area of objects and number of small objects are important in developing a model for BOD, and amount of filaments and form factor were selected by every selection method as input variables to COD model. Many methods also selected the identical or similar subsets of input variables to develop a certain model. For example, four out of five methods selected identical subsets for suspended solids model. Fractal dimension was found important input variable for suspended solids model and the area of flocs for the nitrogen model also in the earlier study [8] but in general the selected optical monitoring variables are different for each quality parameter in industrial and municipal wastewater treatment processes.

Table 2. Selected input variables for the effluent quality parameter models.

	BOD	COD	SS	N	P	Variables
Correlation	7, 6, 2, 3, 1	1, 3, 2, 7, 8	6, 5, 7, 1	5, 6, 7, 1	1, 4, 7, 8	1 Amount of filaments
						2 Fractal dimension
Stepwise	7, 8	1, 5, 3	6, 5, 2	5, 3	1, 4, 3	3 Form Factor
						4 Roundness
Forward	7, 8, 5, 2, 3	1, 5, 3, 8	6, 5, 2	5, 3, 1, 2, 7	1, 4, 3, 7	5 Aspect ratio
						6 Equivalent diameter
GA	2, 3, 5, 7, 8	1, 3, 5	2, 5, 6	1, 2, 3, 5, 7	3, 4, 5, 7	7 Mean area of objects
						8 Number of small obj.
SPA + GA	8, 7, 5, 2	1, 3	6, 5, 2	6, 5, 2	1, 4	

Overall, the performances of the best developed models for every quality parameter listed in Table 3 are acceptable taking into account that only the optical monitoring variables were selected as input variables and that the predicted parameters are measured from the effluent of the ASP. Based on the current data, the model development for the effluent phosphorus is the most challenging, yet the best results are still promising. The fitness of the best COD model is the highest of all (R^2=0.78) and can be considered good. The best BOD and SS models are also acceptable, yet none of the models can predict the exact values of the quality parameter. The regression coefficients of the best models developed are listed in Table 4 where x_0 is the bias and x_n are the selected input variables.

Further investigation showed that applying also some process measurements as input variables improves the model performance. For example, using fractal dimension, aspect ratio, equivalent diameter, effluent flow, influent suspended solids, and primary settling tank BOD and COD as input variables for developing a model for suspended solids in the effluent yields the fitness of R^2=0.81, RMSE=0.44. These input variables were selected using the genetic algorithm method from a larger dataset which consisted of in addition to the optical monitoring results tens of traditional process measurements from the automation system of the ASP. Again, selection was performed on the mathematical basis only. Also for COD in the effluent, a model of fitness R^2=0.89, RMSE=0.34 can be achieved using the amount of filaments, influent COD, influent suspended solids, influent conductivity, aspect ratio, primary settling tank pH and form factor as input variables. These models are very good and can be used in addition to predicting the level and changes of the quality parameter also predicting almost the exact values of the quality parameters. However, the amount of input variables is somewhat higher than in the earlier models which may increase the risk of developing an overfitted model.

Table 3. Performances of the developed models for BOD, COD, suspended solids, phosphorus and nitrogen.

	BOD		COD		SS		N		P	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Correlation analysis	0.43	0.97	0.66	0.60	0.54	0.68	0.51	0.78	0.51	0.80
Stepwise selection	0.61	0.80	0.76	0.51	0.67	0.58	0.60	0.71	0.54	0.77
Forward selection	0.71	0.69	0.78	0.49	0.67	0.58	0.69	0.63	0.58	0.74
Genetic algorithms	0.71	0.69	0.76	0.50	0.67	0.58	0.69	0.63	0.58	0.74
SPA +GA	0.69	0.71	0.60	0.65	0.67	0.58	0.61	0.70	0.49	0.82

Table 4. The coefficients of regression of the best developed models for the quality parameters of the effluent.

BOD (GA)	-1.48 x_0	1.02 x_2	0.43 x_3	1.19 x_5	-1.97 x_7	-1.11 x_8

COD (forward)	0.41 x_0	1.12 x_1	1.71 x_5	1.79 x_3	0.21 x_8	
SS (GA)	-0.55 x_0	-0.69 x_1	1.25 x_2	0.91 x_3		
P (GA)	-0.73 x_0	2.50 x_1	-1.42 x_2	0.71 x_3	-0.32 x_5	
N (forward)	-0.11 x_0	2.86 x_5	2.27 x_3	0.39 x_1	0.68 x_2	-0.42 x_7

The optical monitoring data including stable and non-stable condition of the activated sludge process gives a good starting point to develop a simple practical model that can predict the quality of treated wastewater in several conditions. In Figure 2 is presented the behavior of a model for suspended solids content in the effluent as scaled values. In Figure 3, Figure 4, Figure 5 and Figure 6 are also presented the behavior of the models for effluent BOD, COD, nitrogen and phosphorus, respectively. These models were developed using input variables selected by the genetic algorithm which in general yielded the most optimal input variables for the models. It can be seen that the models are able to predict the changes and the quality of treated wastewater in stable and non-stable conditions, although exact values are more or less challenging to predict.

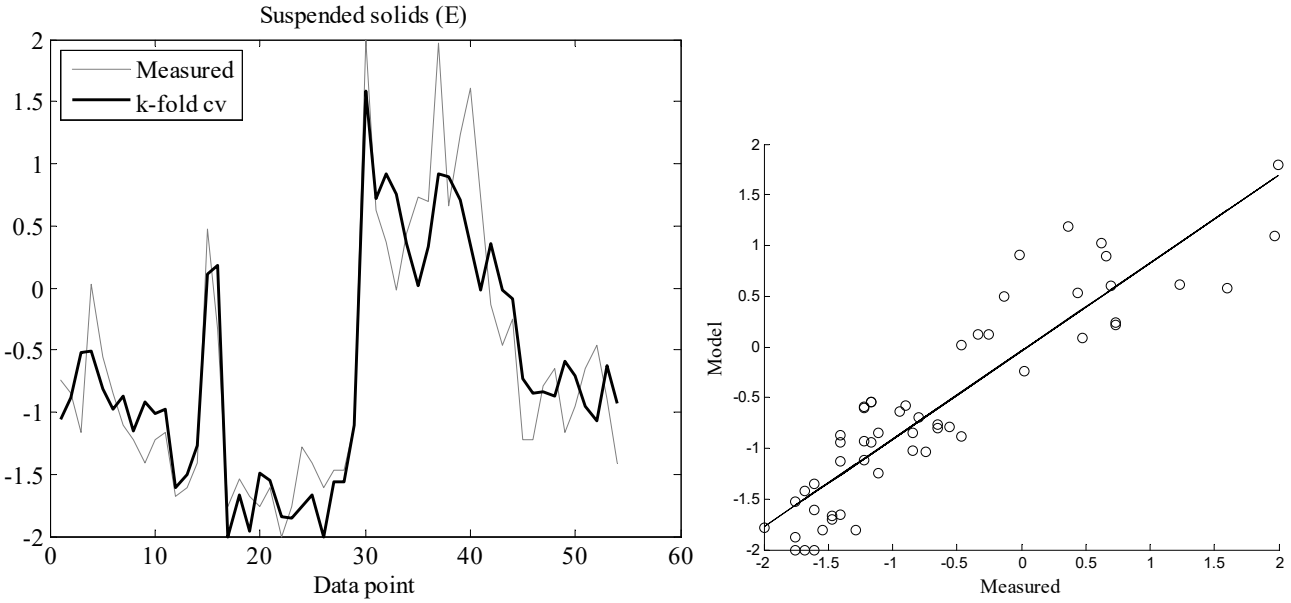

Figure 2. The modelled vs the original effluent suspended solids content as scaled values. The variables selected by the genetic algorithm method were used as the input variables of the model.

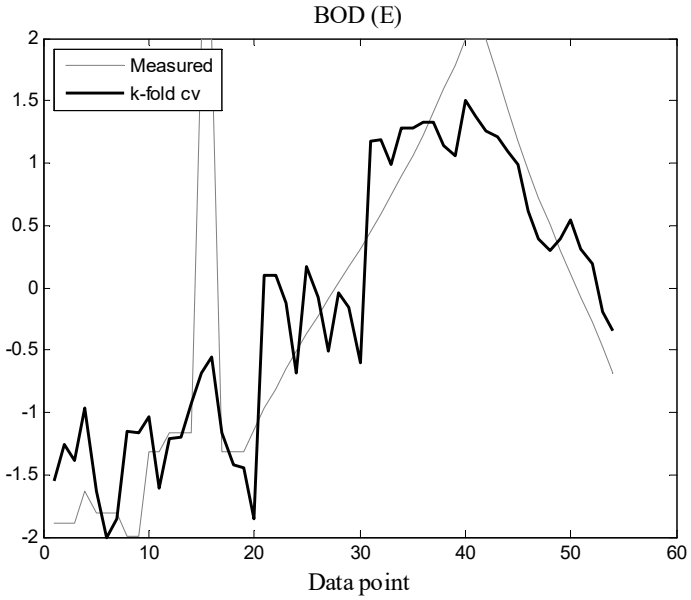

Figure 3. The modelled vs the original effluent BOD as scaled values. The variables selected by the genetic algorithm method were used as the input variables of the model.

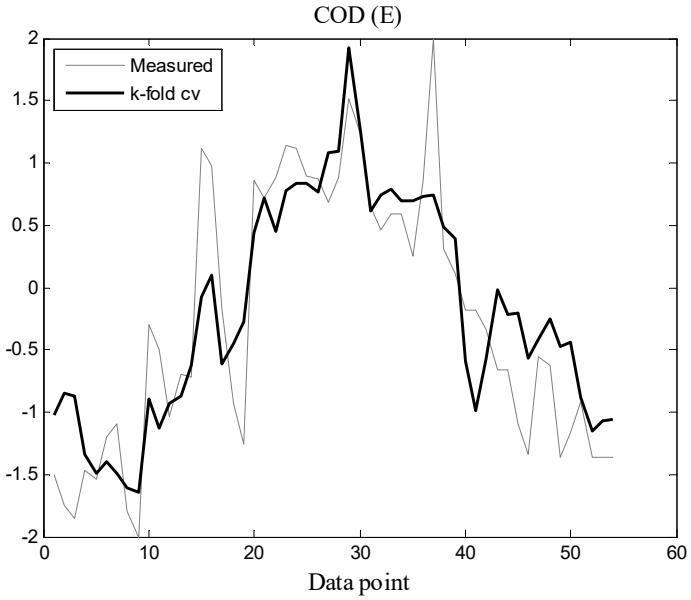

Figure 4. The modelled vs the original effluent COD as scaled values. The variables selected by the genetic algorithm method were used as the input variables of the model.

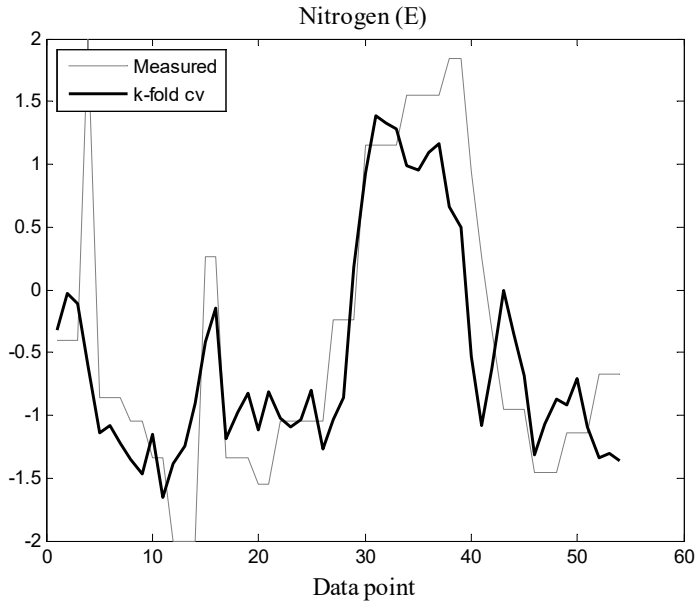

Figure 5. The modelled vs the original effluent nitrogen as scaled values. The variables selected by the genetic algorithm method were used as the input variables of the model.

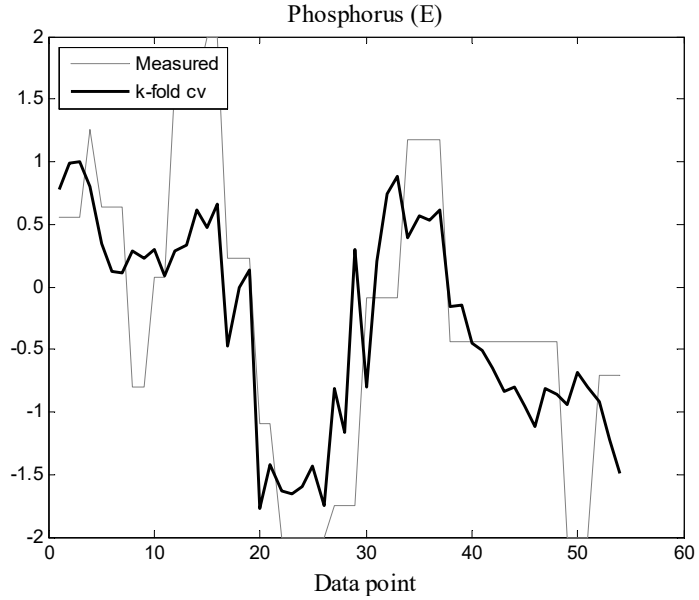

Figure 6. The modelled vs the original effluent phosphorus as scaled values. The variables selected by the genetic algorithm method were used as the input variables of the model.

5. Conclusion

In this study, the optical monitoring results of a full-size activated sludge process located in a pulp mill area were utilized to develop predictive models for five parameters that describe the quality of the effluent and the efficiency of the treatment process. Wastewater samples were taken from the aeration tank and imaged and analyzed using the novel automatic optical monitoring device. Five variable selection methods were utilized to find the optimal subsets of input variables for BOD, COD,

suspended solids, nitrogen and phosphorus content in the treated wastewater and a five-fold cross-validation was used to estimate the performances of the models. The dependencies between selected variables were also briefly studied.

The performances of the best developed models with a decent amount of input variables were promising to predict the upcoming quality of the effluent taking into account that only optical monitoring variables were used as input variables. It is also important to bear in mind that the variable selection results are based solely on a mathematical analysis and may not accurately correspond the reality in the process. Indirect effects between variables may occur in the process even though they are not revealed on a mathematical analysis. The modelling results could be improved by using a more complicated modelling method or including some additional information in the models as it was shown that adding some process measurements as input variables improved notably the fitness of the models. Nevertheless, the best developed models can be used for estimating the upcoming level of biological and chemical oxygen demand, suspended solids, nitrogen and phosphorus in the effluent of the industrial wastewater treatment plant in stable and non-stable condition. The comparison to the results of the study in a municipal wastewater treatment plant showed that even though some optical monitoring variables are found important in modelling a certain quality parameter, the results are not generalizable and every ASP requires an independent study and model development work. However, together with the expert knowledge the novel automatic optical monitoring method combined to the predictive modelling has potential to be used in process control and avoiding environmental risks by receiving information on the effluent quality hours before revealed by traditional process measurements.

Acknowledgements
This research was carried out as part of the Measurement, Monitoring and Environmental Efficiency Assessment (MMEA), the research programme of CLEEN Ltd. – Cluster for Energy and Environment. Aki Sorsa, D.Sc. (Tech.), is greatly acknowledged for assisting in variable selection issues.

References
[1] Finnish Forest Industries, http://www.forestindustries.fi, (accessed August 2016)
[2] G. Tchobanoglous, F.L. Burton, H.D. Stensel, Wastewater Engineering: Treatment and Reuse, 4th edition, 2003.
[3] M. Da Motta, M-N. Pons, N. Roche, H. Vivier, Characterisation of activated sludge by automated image analysis, Biochemical Engineering Journal, 9 (2001) 165-173.
[4] D.P. Mesquita, O. Dias, A.M.A. Dias, A.L. Amaral, E.C. Ferreira, Correlation between sludge settling ability and image analysis information using partial least squares, Analytica Chimica Acta, 642 (2009) 94–101.
[5] E. Koivuranta, J. Keskitalo, T. Stoor, J. Hattuniemi, M. Sarén, J. Niinimäki, A comparison between floc morphology and the effluent clarity at a full-scale activated sludge plant using optical monitoring, Environmental Technology, 35(13) (2014) 1605-1610, doi:10.1080/09593330.2013.875065.
[6] E. Koivuranta, T. Stoor, J. Hattuniemi, J. Niinimäki, On-line optical monitoring of activated sludge floc morphology, Journal of Water Process Engineering, 5 (2015) 28–34.
[7] J. Tomperi, E. Koivuranta, A. Kuokkanen, E. Juuso, K. Leiviskä, Real-time optical monitoring of the wastewater treatment process, Environmental Technology, 37 (2016) 344-351, doi:10.1080/09593330.2015.1069898.

[8] J. Tomperi, E. Koivuranta, A. Kuokkanen, K. Leiviskä, Modelling the effluent quality based on a real-time optical monitoring of the wastewater treatment process, Environmental Technology, (2016), doi:10.1080/09593330.2016.1181674.
[9] E. Juuso, Integration of intelligent systems in development of smart adaptive systems: linguistic equation approach. - Acta Universitatis Ouluensis. Series C, Technica 476, Oulu, Dissertation, (2013), http://urn.fi/urn:isbn:9789526202891.
[10] E. Koivuranta, J. Keskitalo, A. Haapala, T. Stoor, M. Sarén, J. Niinimäki, Optical monitoring of activated sludge flocs in bulking and non-bulking conditions, Environmental Technology, 34(5-8) (2013) 679–686.
[11] J. Russ, Computer-assisted Microscopy, Plenum Press, New York, 1990.
[12] M.A. Hall, Correlation-based feature selection for machine learning, The University of Waikato, New Zealand, Doctoral Thesis, 1999.
[13] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, The Journal of Machine Learning Research, 3 (2003) 1157-1182.
[14] A. Sorsa, K. Leiviskä, S. Santa-aho, M. Vippola, T. Lepistö, An Efficient Procedure for Identifying the Prediction Model Between Residual Stress and Barkhausen Noise, Journal of Nondestructive Evaluation, 32(4) (2013) 341-349.
[15] S. Arlot, A. Celisse, A survey of cross-validation procedures for model selection, Statistics Surveys, 4 (2010) 40–79.
[16] I.H. Witten, F. Eibem, M.A Hall, Data Mining, Practical Machine Learning Tools and Techniques, 3nd edition, Elsevier, Burlington, 2011.