(8)

Raport Badawczy

Research Report

RB/50/2016

Prognozowanie dopływu ścieków surowych do oczyszczalni i ładunku zanieczyszczeń w ściekach za pomocą sieci neuronowych i modeli operatorowych

L. Bogdan, J. Studziński,
I. Rojek, B. Szeląg

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Dr hab. inż. Jan Studziński

Warszawa 2016

PROGNOZOWANIE DOPŁYWU ŚCIEKÓW SUROWYCH DO OCZYSZCZALNI I ŁADUNKU ZANIECZYSZCZEŃ W ŚCIEKACH ZA POMOCĄ SIECI NEURONOWYCH I MODELI OPERATOROWYCH

Lucyna Bogdan, Jan Studziński, Izabela Rojek, Bartosz Szeląg

IBS PAN Warszawa 2016

Spis treści

Wprowadzenie

W raporcie zamieszczono trzy prace opracowane przez zespół autorów z Instytutu Badań Systemowych PAN, Politechniki Świętokrzyskiej w Kielcach i Uniwersytetu im. Kazimierza Wielkiego w Bydgoszczy.

Praca pierwsza pod tytułem *Prognozowanie wybranych wskaźników jakości ścieków na dopływie do oczyszczalni metodami Data Mining*, autorstwa Lucyny Bogdan i Jana Studzińskiego z IBS PAN oraz Bartosza Szeląga z Politechniki Świętokrzyskiej dotyczy modelowania wybranych wskaźników jakości ścieków w dopływie ścieków do oczyszczalni na podstawie ich pomiarów względnie na podstawie pomiarów przepływu ścieków.

Praca druga pod tytułem *Modelling mixed liquor suspended solid and substrate load on the basis of wastewater quality indices and operational parameters of the bioreactor: data mining approach*, autorstwa Izabeli Rojek z Uniwersytetu Kazimierza Wielkiego w Bydgoszczy oraz Jana Studzińskiego i Bartosza Szeląga, jest referatem zgłoszonym na międzynarodową konferencję pn. *Advances in Energy Systems and Environmental Engineering ASEE17*, organizowaną w dniach 2 – 5 lipca 2017 we Wrocławiu przez Politechnikę Wrocławską oraz University of New Mexico (USA) i Brunel University London (UK).

Praca trzecia pod tytułem *A Data Mining approach to the prediction of substrate load and mixed liquor suspended solid*, autorstwa Bartosza Szeląga I Jan Studzińskiego, jest artykułem złożonym do czasopisma *Polish Journal of Environmental Studies* w Olsztynie i znajdującym się obecnie na etapie recenzowania.

II. Modelling mixed liquor suspended solid and substrate load on the basis of wastewater quality indices and operational parameters of the bioreactor: data mining approach

Izabela Rojek, Jan Studziński, Bartosz Szeląg

Abstract. In the present study the Support Vector Method (SVM) is used to analyze the dependency between input variables (quantity and quality of wastewaters at the inflow and operational characteristics of an Activated Sludge Tank (AST) and a result of predicted mixed liquor suspended solid (MLSS) and substrate loads (F/M). Computations revealed, that the highest errors for MLSS are present if only the load of organic compounds susceptible to chemical degradation was included in the input data and lowest when load of coal, ammoniacal nitrogen, suspension and ASCh's operational characteristics were used as the input. Moreover, it appeared that indices of wastewater quality at the inflow to the treatment plant can be simulated on the basis of the measured discharge and temperature of wastewaters and in a result it is possible replacing these measured indices with modeled ones to simulated MLSS and F/M. The lowest errors of predicted substrate loads, computed on the basis of modeled using SVM indices values were obtained for coal, ammoniacal nitrogen, suspension loads and AST's operational characteristics - sludge temperature and pH and methanol dosage.

key words: support vector machines, data mining, mixed liquor suspended solid, wastewater treatment, substrate load, classification trees.

Introduction

It is a complex task to operate the wastewater treatment plant (WWTP). That requires maintaining an appropriate level of bioreactor technological parameters to ensure the required degree of contaminant reduction. Observations of the treatment facility operation showed that, due to variations in the quantity and quality of wastewater flowing into the WWTP, to be able to maintain constant process conditions, it is necessary to control on-line the operational parameters in the activated sludge tanks (Zealand and Russell 1984, Chua et al. 2000, Henze et al. 1999, 2002). That is aimed at maintaining constant age of the sludge, and appropriate ranges of substrate load and biomass content. Substrate load (F/M) describes the value of the load of compounds susceptible to biodegradation relative to biomass quantity in the activated sludge tank. In practice, substrate load can be treated as a factor that limits the performance of activated sludge (Jenkins et al. 2004, Barbusiński and Kościelniak 1995). The mode of the plant operation must be adjusted to the task of maintaining the required degree of nitrogen and phosphorus compounds reduction relative to the sedimentation unit size. The reduction degree is specified in the relevant regulations. Consequently, it necessary to maintain symbiosis between microorganisms to ensure good quality of the processes of removal of biogenic compounds. Wastewater treatment methods can be classified as low- and high-load systems (Henze et al. 2002, Mąkinia 2010, Medina and Neis 2007). As regards the former category, F/M value is not higher than 0.05 gBOD/g MLSS·d. Conversely, when aerobic sludge stabilisation takes place during the time, in which newly formed micro-organisms, stored in the cells, are oxidised and highly loaded, the sludge needs anaerobic stabilisation, and F/M value ranges 0.4 ÷ 1.5 gBOD/g MLSS·d. As regards typical systems for the removal of carbon, nitrogen and phosphorus compounds, it is recommended that F/M value should not exceed 0.10 gBOD/g MLSS·d and not be lower than 0.05 gBOD/g MLSS·d due to sedimentation problems caused by filamentous bacteria (Madoni et al. 1999, Santos et al. 2015, Eikelboom 2000). Therefore, substrate load must be kept within a specified range to maintain a required degree of reduction of biogenic compounds. Currently, the settings are usually corrected in real time, and also when the bioreactor malfunction occurs, which is not an optimal solution from both economic and environmental standpoint. Therefore, due to random character of inflow rate, and thus also biochemical oxygen demand, anomalous events may occur, in which it is necessary to correct the settings in different parts of the treatment facility to obtain the optimum range of F/M quantities. In view of the above, it is reasonable to try to predict, sufficiently in advance, the values of loads of biodegradable compounds. That allows the staff to increase the load of compounds susceptible to biodegradation by means of adding external supplemental carbon source (methanol, ethanol, leachate, etc.) to the treatment system, or by specifying adequate mixed liquor suspended solid. The latter is conditioned by the degree of recirculation, concentration of recirculated sludge, the amount of excess sludge, and others. To make predictions, statistical or physical-mathematical models are employed. The latter are based on the systems of differential equations that describe runoff (Zawilski and Brzezińska 2013, Gironás et al. 2010) and biochemical reactions that occur in wastewater flowing through the sewers (Hu et al. 2006, Huisman et al. 2000).

Due to the fact that the development of a physical model is very expensive, and numerous input data need to be collected, data mining methods are often used to model the treatment facility (Szeląg and Gawdzik 2016, 2017, Minsoo et al. 2016, Han et al. 2015) and its catchment area (Dellana and West 2009). Those methods are relatively quick to devise and the specialist knowledge of the physics of the phenomenon analysed is not required. At the so-called training stage, the model structure is built, which provides a basis for the prediction of a given parameter. The models include, among others, the following: Artificial Neural Networks (Rojek and Studziński 2014), Support Vector Machines (Guo et al. 2015), Random Forests (Verma et al. 2013), Boosted Trees (Kusiak et al. 2013), k-Nearest Neighbour, etc. For the sake of substrate load simulations, it is necessary to build, at the same time, three separate models to simulate wastewater inflow, biochemical oxygen demand and mixed liquor suspended solid. On the basis of literature review (Wei and Kusiak 2015), it can be concluded that the models developed for the simulation of inflow to WWTP over a time horizon from 1 to 7 days show satisfactory predictive abilities. As regards the simulations of mixed liquor suspended solid, it is necessary to perform costly investigations geared towards determining values of wastewater quality indices (including BOD, COD, TSS, and NH4) at the inflow to WWTP (Bagheri et al. 2016, Güçlü and Dursun 2010). Also, the measurements, often taken online, of the reactor technological parameters (degree of recirculation, recirculated sludge concentration, quantity of returned excess sludge, sludge temperature and pH, etc.) generate extra operating costs. As regards the determination of indices of pollution, biochemical oxygen demand determination is particularly problematic due to the time it takes, i.e. 5 days.

This study presents the methodology for substrate load modelling on the basis of statistical models developed to simulate inflow rate, biochemical oxygen demand and mixed liquor suspended solid. Models for BOD determination were devised on the basis of the results of inflow and wastewater temperature measurements, whereas the determination of the inflow rate was based exclusively on the results of measurements that preceded the modelled value. Due to high costs of continuous measurements of mixed liquor suspended solid, the impact of the load of contaminants, i.e. biogenic compounds (BOD, COD, SS, NH_4^+) and individual parameters of the bioreactor operation (pH, T_{sl}, DO, WAS, RAS), on the prediction of the parameter of concern was analysed in detail. That is important from the engineering standpoint, because it is possible to evaluate the effect of disregarding selected parameters on the results of mixed liquor suspended solid simulations. Additionally, due to technical problems related to the measurements of quality indices, the possibility of simulating those on the basis of flow values and inflowing wastewater temperature measurements was taken into consideration. From the practical perspective, it is an important fact as it ultimately allows the assessment of simulations of both mixed liquor suspended solid and substrate load on the basis of wastewater quality indices computed from basic parameters measured at the wastewater treatment facility.

The object of investigations

The investigations concerned the wastewater treatment plan (WWTP) located in the terrain of the commune of Sitkówka – Nowiny. The plant receives sanitary wastewater from the separate wastewater system of the city of Kielce, the commune of Sitkówka – Nowiny, and a part of the commune of Masłów. The design capacity of the treatment plant is 72,000 m^3/d, and it is capable of serving a population equivalent (P.E.) of 275,000. Wastewater delivered to the treatment plant is mechanically pretreated using step screens and aerated grit chambers, with separate grease traps (Fig 1). Next, wastewater is pumped to four primary settling tanks (O.1 – O.4), from which it is delivered to the biological unit (bioreactor with separate denitrification and nitrification chamber). After dephosphatation process, wastewater is conveyed to denitrification chamber for the removal of phosphorus compounds.

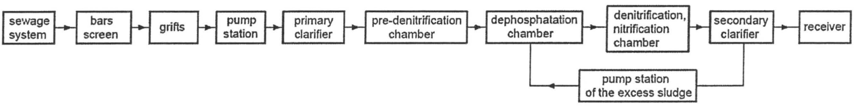

Fig. 1. Technological diagram of the Sitkówka – Nowiny treatment plant.

Then, wastewater together with activated sludge is transferred to four secondary settling tanks, from where after clarification, it flows to the receiving water, i.e. the river Bobrza.

Methodology

In this study, to compute substrate load, three separate statistical models were devised, namely those for predictions of inflow rate, biochemical oxygen demand and mixed liquor suspended solid As to wastewater inflow to the treatment plant, autoregressive model was developed, based on the results of the last measurements that were taken. Simulations of biochemical oxygen demand values were carried out on the basis of measured values of inflow, wastewater temperature, and their combinations from last measurements that preceded the modelled values.

Because of high costs of determining wastewater quality indices and the possibility of limiting the scope of investigations into biogenic compounds, the impact of organic and nitrogen compounds, and that of suspended solids on the error in prediction of mixed liquor suspended solid was thoroughly examined in the analyses. In the next stage, the impact of the reactor operational parameters, including oxygen concentration (DO), sludge temperature (T_{sl}) and pH, degree of recirculation (RAS), the amount of returned excess sludge (WAS) on the simulations of MLSS values was also analysed. As a result, the analyses mentioned above made it possible to specify the impact of independent variables on the simulation error for the parameter of concern. Ultimately, that also allowed a reduction in the scope of measurements of wastewater quality indices and activated sludge tank (AST) operational parameters, which lowered the costs of investigations.

Due to the fact that substrate load depends on biochemical oxygen demand load, the simulations also included the determination of the BOD prediction error.

Because of high cost of measurements of wastewater quality indices at the inflow to WWTP and technical problems related to obtaining series of constant resolution, simulations of quality indices of $C(t)_j$ were considered. That could be done on the basis of measurements of temperature and inflow rate of wastewater delivered to the treatment plant. The analyses quoted above are intended to develop a model form that would make it possible to predict mixed liquor suspended solid on the basis of flow functions $Q(t-i)_j$ describing values of quality indices $C(t)_j$, and also of operational parameters of the bioreactor. To identify variables in the models describing BOD, COD, NH_4^+ and SS, i.e. flow quantities and wastewater temperature from last measurements, the classification trees method was applied. Classification trees were used to determine the so-called predictors importance (IMP) that expresses the impact of independent variables on the parameter of concern, individual wastewater quality indices in this case. Next, the results of simulations, i.e. $C(t)_j$, and also $T_{sl}(t)$ and $Q(t)$, were substituted into the model expressed by formula (1) and substrate load was determined:

$$MLSS(t) = F\{Q(t), T_{sl}(t), C(t)_j, pH, WAS, RAS, DO\} \quad (1)$$

$$C(t)_j = f(Q(t-i)_j) \quad (1a)$$

$$T_{sl}(t) = f(T_{sl}(t-l)) \quad (1b)$$

$$Q(t) = f(Q(t-k)) \quad (1c)$$

Then, computational results were compared with those measured ones. That is important for practical applications, because it is also possible to predict values of quality indices on the basis of Q, T quantities, thus reducing the costs of carrying out continuous determinations. Additionally, from the technical standpoint, it is possible to simulate mixed liquor suspended solid, and thus to control, in appropriate advance, the values of WAS, RAS, DO, and pH. That could contribute to increase in AST operation efficiency and keep the processes taking place in WWTP within the range that ensures the required reduction of biogenic compounds.

In this paper, the possibility of using Support Vector Machines method to predict wastewater quality indices and other parameters was taken into consideration. To make the training process correct, and then to properly assess the performance of the statistical models applied, five-fold cross-validation was used. Simultaneously, the data were partitioned into the training set (75%), and the testing and validating set (25%). Prior to the start of the construction of models, data standardisation was performed by means of min-max transformation expressed by formula:

$$\overline{A_i} = \frac{A_i - \min A}{\max A - \min A} \tag{2}$$

where: $\overline{A_i}$ - normalized, with the min-max method, value of the *i*-th element of set A, A_i - value of the *i*-th element of set A recorded in measurements, max A – maximum value of a single element in the set of parameter A, min A - minimum value of a single element in the set of parameter A.

Support Vector Machines (SVM) cover a group of methods developed by Vapnik (1998) first exclusively for classification purposes, which expanded over time to include regression issues (SVR). For that reason, the dependence between the model output and input variables can be non-linear. As a result, in this method a non-linear transformation of N – dimensional space to K-dimensional feature space of a larger size is applied. In the SVM method, in the goal function definition, the error function with insensitivity threshold ε is used (Burges 1998):

$$L^{\varepsilon}(d, y(x)) = \begin{cases} 0 & gdy \ |d - f(x)| \le \varepsilon \\ |d - y(x)| - \varepsilon & gdy \ |d - f(x)| > \varepsilon \end{cases} \tag{3}$$

where: ε – assumed model accuracy, x – input vector, y(x) – value of the model output signal expressed by dependence (Burges 1998):

$$y(x)=\sum_{j=1}^{K} w_i \cdot \varphi_i(x)+b=w^T \cdot \varphi(x)+b \tag{4}$$

in which: $w = [w_1,\ldots, w_k]^T$ – transposed vector of weights, $\varphi(x) = [\varphi_1(x),\ldots, \varphi_k(x)]$ – vector of basis functions.

In the SVM method, the aim of training is to properly select a vector of weights *w*, number of neurons, basis functions and their parameters so that the error function in the form given below could be minimised:

$$E=\frac{1}{2}\sum_{i=1}^{p} L_\varepsilon(d_i, y(x_i)) \tag{5}$$

in which: p – number of training pairs (x_i,d_i).

Taking into account complementary variables ξ_i and ξ_i', the problem of SVM network training can be formulated as minimisation of the values of the network weights *w* and of variables ξ_i, ξ_i' written as follows (Burges 1998):

$$\min\{\phi(w,\xi,\xi')\}=\frac{1}{2}\cdot w^T w+C\left[\sum_{i=1}^{p}(\xi_i+\xi_i')\right] \tag{6}$$

under the following functional constraints:

$$\begin{cases} d_i - w^T\varphi(x)-b\le\varepsilon+\xi_i \\ w^T\varphi(x)+b-d_i\le\varepsilon+\xi_i \\ \xi_i\ge 0 \\ \xi_i'\ge 0 \end{cases} \tag{7}$$

Employing Lagrange multipliers (α_i) method, the system of equations described by dependences (7) and (8) can be converted to a dual problem and expressed in the form dependent on the so-called kernel function K (Burges 1998):

$$y(x) = \sum_{i=1}^{N_{sv}} (\alpha_i - \alpha_i^{'}) K(x, x_i) + w \tag{8}$$

in which: N_{sv} – number of support vectors dependent on C and ε. Linear, polynomial and Gaussian kernels are the most frequently used kernel functions. In this study, Gaussian kernels were employed and it was assumed ε = 0.01, whereas C was specified at the stage of analyses.

The evaluate the predictive abilities of the mathematical models developed for the study to compute inflow rate, sludge temperature, quality indices, mixed liquor suspended solid and substrate load, the following parameters were used:

- mean absolute error:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| y_{i,obs} - y_{i,pred} \right| \tag{9}$$

- mean relative error:

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_{i,obs} - y_{i,pred}}{y_{i,obs}} \right| \cdot 100\% \tag{10}$$

where: n – data set size, $y_{i,(obs,\ obl)}$ – measured and computed value of dependent variable, $\bar{y}_{obs\ ,pred}$ – arithmetic mean of measured-to-computed value of dependent variable.

Results

On the basis of the results of measurements of the quantity and quality of inflowing wastewater and bioreactor operational parameters, the range of parameter variation was identified. (Table 1). That is essential with respect to operation of the models used in this study because it provides information on the range of applicability of those models. The data in Table 1 show that the quantity and temperature of inflowing wastewater ranged vastly, which led to considerable differentiation in the content of biogenic compounds in the inflow. In the

Sitkówka-Nowiny treatment plant, high variation in the load of contaminants conveyed to AST contributed to substantial differentiation in mixed liquor suspended solid (1.19 ÷ 5.89 kg/m^3), degree of recirculation (44.6 ÷ 167.7%), or dissolved oxygen concentration (0.55 ÷ 2.78 mg/dm^3). The fact that methanol was periodically dosed into wastewater (maximum 4.56m^3) confirms a varied content of biodegradable carbon compounds in the inflow to the treatment plant, which affects, among others, the pattern of processes of removal of nitrogen compounds from wastewater. Based on the results of measurements of the wastewater quantity and quality and the bioreactor operation, statistical models for predicting mixed liquor suspended solid were developed.

The models accounted for the possibility of modelling a given parameter on the basis of the load of organic compounds, nitrogen compounds, and solids, and additionally AST parameters (RAS, WAS, pH, T_{sl}, DO). Table 2 presents the results of MLSS simulation using the SVM method. For statistical models (Table 2) devised with the SVM method, C value included in formula (6) ranged 5 ÷ 8.

The data in Table 2 indicate that errors in MLSS prediction based on organic compounds load expressed in terms of BOD and COD are equal to MAE = 0.792 kg/m^3 and MAPE = 20.10%. When the load of ammoniacal nitrogen and solids is accounted for in the models, mean absolute error and mean relative error decrease and amount to MAE = 0.733 kg/m^3, MAPE = 17.19% and also MAE = 0.700 kg/m^3, MAPE = 16.54%, respectively. Taking into account sludge parameters, namely temperature and pH in the model for MLSS simulations leads to a reduction in errors in mixed liquor suspended solid prediction to respective values: MAE = 0.672 kg/m^3, MAPE = 13.36%, and MAE = 0.617 kg/m^3, MAPE = 12.65%. Due to a decrease in BOD value at the inflow to WWTP, a drop in mixed liquor suspended solid in AST during abnormal events occurs. That, together with the necessity to ensure large reduction in pollutant loads make it necessary to dose methanol, being an external supplemental carbon source.

Table 1. Range of variation of parameters describing wastewater quantity (Q), quality (BOD, COD, SS, TN, NH4) and bioreactor operation (T_{os}, pH, DO, MLSS, RAS, WAS, m_{met}, F/M).

Variable	Minimun	Mean	Maximum
Q, m^3/d	32564	40698	86592
T_{in}, ^0C	8.4	16.6	20.9
T_{sl}, ^0C	10.0	15.9	23.0
pH	7.2	7.7	7.8
MLSS, kg/m^3	1.19	4.26	5.89
RAS, %	44.6	90.70	167.6
m_{met}, m^3	0.00	1.35	4.56
WAS, kg/d	3489	11123	19194
DO, mg/dm^3	0.55	2.56	5.78
F/M, gBOD/g s MLSS·d	0.03	0.07	0.15
BOD, mg/dm^3	127	309	557
COD, mg/dm^3	384	791	1250
SS, mg/dm^3	126	329	572
NH_4^+, mg/dm^3	24.4	7.8	65.9

Table 2. Parameters of fit of computations with SVM models for mixed liquor suspended solid to the results of measurements.

Variables	training		testing	
	MAE	MAPE	MAE	MAPE
Q, COD	0.821	20.772	0.845	20.98
Q, BOD, COD	0.776	19.678	0.792	20.10
Q, BOD, COD, NH_4^+	0.703	17.150	0.733	17.19
Q, BOD, COD, NH_4^+, SS	0.686	16.304	0.700	16.54
Q, BOD, COD, NH_4^+, SS, T_{sl}	0.669	12.950	0.672	13.36
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH	0.600	12.288	0.617	12.65
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}	0.557	11.970	0.580	12.01
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO	0.528	11.223	0.546	11.47
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO, RAS	0.469	10.832	0.472	11.05
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO, RAS, WAS	0.378	9.385	0.392	9.73

As a result of taking methanol into account in the statistical model, the values of MLSS prediction errors decrease to MAE = 0.580 kg/m^3 and MAPE = 12.01%. The data in Table 2 also indicate that when additional parameters, such as dissolved oxygen concentration, or a degree of recirculation are considered in the statistical model based on the load of organic compounds, nitrogen compounds, solids and reactor parameters (pH, T_{sl}, m_{met}) MLSS simulation errors are reduced to MAE = 0.546 kg/m^3, MAPE = 11.47% and MAE = 0.472 kg/m^3, MAPE = 11.05%, respectively. The lowest computational error values (MAE =

0.392 kg/m^3 and MAPE = 9.73%) for mixed liquor suspended solid in the treatment facility of concern are obtained when independent variables are loads of organic compounds, nitrogen compounds and suspended solids and also AST technological parameters (pH, T_{sl}, RAS, WAS, m_{met}, DO).

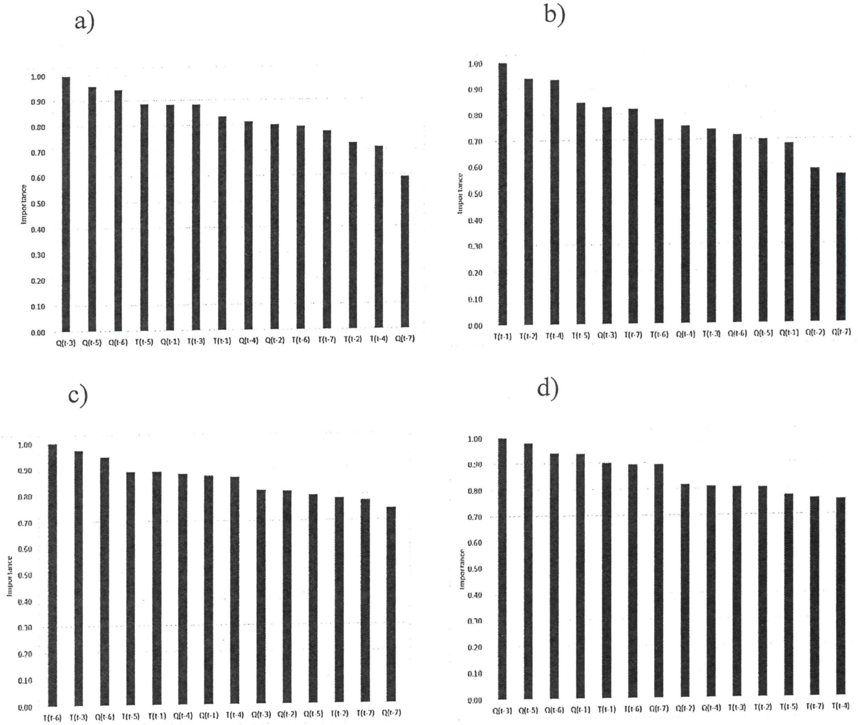

Fig. 2. Results of computations of the importance of predictors Q(t-i), T(t-j) for wastewater quality indices: a) BOD, b) COD, c) SS, d) NH_4^+.

To identify the structure of the model described by eqs. (1) and (2), the possibility of simulating wastewater quality indices on the basis of inflowing wastewater temperature and quantity was analysed. Additionally, the analyses also included autoregressive statistical models developed for the prediction of inflow Q(t) and temperature $T_{sl}(t)$ found in eq. (1). Results of computations of importance of independent variables, and the impact of individual predictors on the quality indices of concern and also Q, T_{sl} obtained with classification trees, are presented in Figs. 2 ÷ 5. The data in Figs. 2a ÷ 2d show that as regards the wastewater quality indices, values of importance of individual predictors are greater than IMP > 0.8 for variables that describe both temperature and inflow rate.

For instance, as to COD, IMP quantities reach the value equal to at least 0.8 for six independent variables, namely T(t-1), T(t-2), T(t-4), T(t-5), T(t-7) and Q(t-3). In the remaining cases (BOD, NH_4^+), the number of independent variables, for which importance value exceeds IMP = 0.8, is greater than for SS and amounts to ten variables for BOD and eleven for NH_4^+. Analyses carried out for inflow rate and sludge temperature (Table 3) demonstrated that values of importance of independent variables are the highest for the last two measurements preceding the modelled quantity and are equal to at least 0.80.

On the basis of the computations above, models were developed for predicting wastewater quality indices (BOD, COD, SS, NH_4^+). They were based on the last measured values of Q and T that preceded the modelled value. Additionally, statistical models were devised using the SVM method to predict daily inflow and sludge temperature. The results of the analyses are presented in Tables 4 and 5.

Table 3. Impact of individual independent variables on Q and T_{sl} values, expressed by means of predictor importance.

Q		T_{sl}	
Variable	Importance	Variable	Importance
Q(t-1)	1.00	T_{sl}(t-1)	1.00
Q(t-2)	0.80	T_{sl}(t-2)	0.96
Q(t-3)	0.67	T_{sl}(t-3)	0.85
Q(t-4)	0.53	T_{sl}(t-4)	0.72
Q(t-5)	0.52	T_{sl}(t-5)	0.63
Q(t-6)	0.50	T_{sl}(t-6)	0.62
Q(t-7)	0.49	T_{sl}(t-7)	0.48

Table 4. Parameters of fit of SVM method simulations to the results of measurements of wastewater quality indices (BOD, COD, SS, NH_4^+) on the basis of Q and T values.

Quality indices	training		testing	
	MAE	MAPE	MAE	MAPE
BOD	42.86	14.94	46.78	13.76
COD	89.77	12.26	90.26	13.02
NH_4^+	4.01	8.25	4.28	8.28
SS	39.02	13.76	41.74	17.66

Table 5. Parameters of fit of SVM method simulations to the results of measurements of

Q, T_{sl} values in the activated sludge.

Variables	training		testing	
	MAE	MAPE	MAE	MAPE
Q	2711	6.45	2588	6.34
T_{sl}	0.49	2.16	0.51	2.36

The data in Table 4 indicate the statistical models obtained with the SVM method, used for predictions of quality indices of concern, show satisfactory predictive abilities, which is confirmed by the computed error values. For instance, with respect to the model for do BOD simulations, the values of mean errors of prediction of the analysed index are MAE = 46.78 mg/dm^3 and MAPE = 13.76%. For SS and NH4, they are equal to MAE = 41.74 mg/dm^3, MAPE = 17.66%, and MAE = 4.28 mg/dm^3, MAPE = 8.28%, respectively. With respect to statistical models for Q, T_{sl} computations (Table 5), values of mean errors of prediction of the analysed variables are MAE = 2588 m^3/d, MAPE = 6.34% and MAE = 0.77^{0}C, MAPE = 5.19%. Using the computed values of Q and BOD, it was found that the mean absolute and relative errors in the simulation of the load of biochemical oxygen demand are MAE = 3119 kg/d and MAPE = 27.15%, whereas the mean value of BOD daily load is 12339 kg/d. As regards statistical models (Table 4 and 5) developed using the SVM method, quantity C ranged from 6 to 9.

The results of simulations presented above, i.e. quality indices (BOD, COD, SS, NH_4^+) and Q, T_{sl} were substituted into the model described by eq. (1). Mixed liquor suspended solid, and then substrate load were determined. Parameters of fit of MLSS and F/M computational results to measurements can be seen in Table 6. Figures 3 and 4 show the comparison of the results of simulations of the parameters analysed for selected computational variants (where: variant 1 – Q, BOD, COD, SS, NH4, pH, T_{sl}, RAS, WAS, m_{met}; variant 2 – Q, COD; variant 3 – Q, BOD, COD, SS, NH_4^+, pH, T_{sl}.) listed in Table 2.

Table 6. Comparison of the parameters of fit of MLSS and F/M computational results obtained on the basis of eq. (1) to measurements.

Variables	MLSS		F/M	
	MAE	MAPE	MAE	MAPE
Q, COD	0.835	20.73	0.0189	28.47
Q, BOD, COD	0.804	19.96	0.0188	28.22
Q, BOD, COD, NH_4^+	0.841	20.88	0.0196	30.32
Q, BOD, COD, NH_4^+, SS	0.832	20.66	0.0205	32.71
Q, BOD, COD, NH_4^+, SS, T_{sl}	0.792	19.66	0.0191	29.60
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH	0.931	23.11	0.0200	28.38
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}	0.821	20.38	0.0174	26.63
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO	0.718	17.82	0.0186	28.51
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO, RAS	0.560	13.90	0.0180	27.76
Q, BOD, COD, NH_4^+, SS, T_{sl}, pH, m_{met}, DO, RAS, WAS	0.439	10.89	0.0187	29.30

The data in Tables 6 make it possible to state that for the model described by eqs. (1), (1a), (1b), (1c), the values of errors in the sludge concentration predictions are greater compared with the situation when those input data (Table

2) directly represent wastewater quality indices (BOD, COD, NH_4^+, SS), flow, and also sludge temperature and pH measured at the instant t. In the case considered here (Table 6), as previously (Table 2), the lowest error values (MAE = 0.439 kg/m^3 and MAPE = 10.89%) in MLSS simulations were found for the variables including the loads of organic, nitrogen, and solids in the inflow to the WWTP, and also the bioreactor parameters.

The data in Table 6 demonstrate that the worst computational results concerning mixed liquor suspended solid were produced for the variables that express the loads of organic compounds, nitrogen compounds and solids in the inflow to the WWTP and also activated sludge parameters such as pH and T_{sl}. In this case, the values of errors in MLSS predictions were equal to MAE = 0.931 kg/m^3 and MAPE = 23.11%. It should be noted that minor (maximum 2 %) variation in the values of mean errors, i.e. MAE = 0.821 ÷ 0.841 kg/m^3 and MAPE = 20.38 ÷ 20.88% was found for statistical models (Table 6) for the simulations of mixed liquor suspended solid that were based only on the load of organic compounds susceptible to chemical degradation (MAE = 0.835 kg/m^3 and MAPE = 20.73%).

The other models showing low error variation were those developed on the basis of variables describing the loads of organic compounds, nitrogen compounds (MAE = 0.841 kg/m^3 and MAPE = 20.88%) and, additionally of solids (MAE = 0.832 kg/m^3 and MAPE = 20.66%), and also the reactor operational parameters such as pH, T_{sl} and the amount of methanol dosed (MAE = 0.821 kg/m^3 and MAPE = 20.38%).

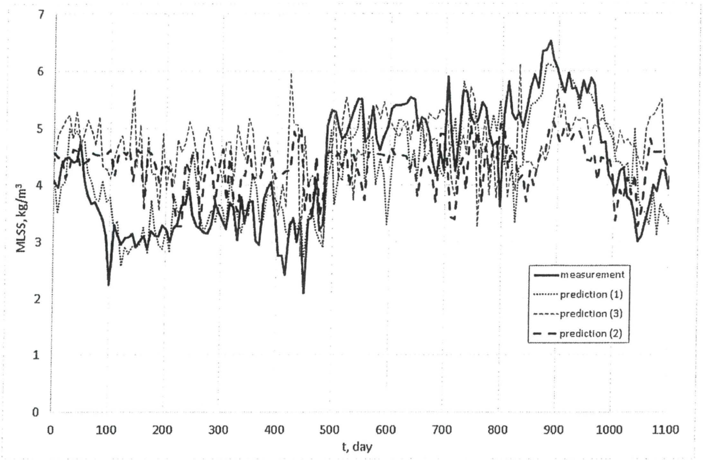

Fig. 8. Comparison of the results of mixed liquor suspended solid measurements and computations with the models 1, 2 and 3 developed using the SVM method for the period analysed.

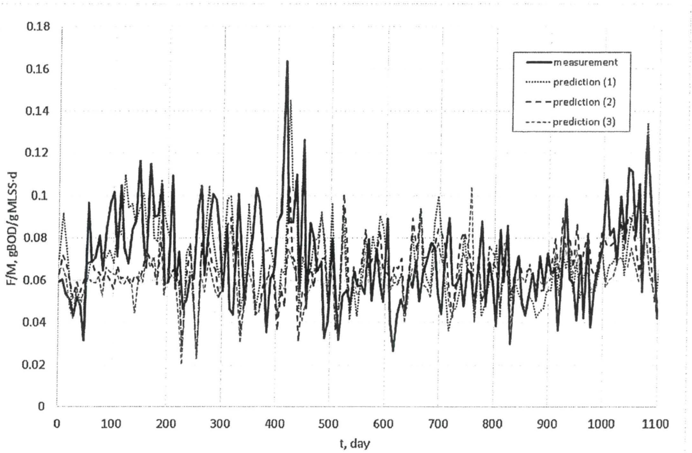

Fig. 9. Comparison of the results of substrate load measurements and computations with the models 1, 2 and 3 developed using the SVM method for the period analysed.

Additionally, the data listed in Table 6 show that the lowest error values in substrate load simulations (MAE = 0.0174 gBOD/g MLSS·d and MAPE = 26.63%) were observed for the data set for the variables describing the loads of BOD, COD, SS, NH4 and the parameters of the bioreactor operation, such as pH, T_{sl} and the quantity of the methanol dosed to provide an external source of carbon. Conversely, the highest values of errors in F/M prediction, equal to MAE = 0.0205 gBOD/g MLSS·d and MAPE = 32.71% were found when independent variables of the parameter of concern were loads of organic compounds, nitrogen compounds and solids. It should be noted that small (only 1 %) variation in the values of mean absolute error was observed for statistical models developed on the basis of variables describing the loads of organic compounds, nitrogen compounds and solids, and also the reactor operational parameters including pH, T_{sl}, m_{met}, RAS, WAS and DO (MAE = 0.0187 gBOD/g MLSS·d and MAPE = 29.3%). The same was true for the models based exclusively on COD load (MAE = 0.0189 gBOD/g s MLSS·d and MAPE = 28.47%). In the remaining statistical models, the variation in the mean absolute and relative errors in the data set was not wide, as values changed within a small range MAE = 0.018 ÷ 0.020 gBOD/g s MLSS·d and MAPE = 27.76 ÷ 30.22%.

Taking into consideration the values of errors in MLSS prediction, it is possible to state that the dependence expressed by eq. (1) can provide a basis for computations of the technological parameter of concern in practical applications. As regards statistical models developed to simulate substrate load, based on the loads of BOD, COD, SS and NH_4^+ and also the reactor operational parameters, the values of the mean absolute and relative errors ranged as follows: MAE = 0.0174 ÷ 0.0205 gBOD/g MLSS·d and MAPE = 26.63 ÷ 32.71%.

Conclusions

On the basis of computations performed for the study, it can be stated that the accuracy of prediction of mixed liquor suspended solid is substantially affected by both wastewater quality indices (BOD, COD, SS, NH_4^+) and also the bioreactor operational parameters. The analyses demonstrated that the highest values of mean errors in MLSS simulations were found for the statistical models devised using the SVM method, which were based on the load of biogenic compounds in the inflow to the treatment facility. Lower values of errors in mixed liquor suspended solid predictions were obtained for the models based on the load of inflow pollutants and also AST technological parameters such as RAS, WAS, pH, T_{sl}, DO, $m_{met.}$

Additionally, computations performed using the statistical models developed with the SVM method demonstrated it is possible to simulate the values of wastewater quality indices on the basis of quantities of wastewater flow and temperature obtained from last measurements that preceded the predicted value. From the practical standpoint, that is essential as it offers the possibility of making predictions, and thus controlling the bioreactor operational parameters in appropriate advance, which is necessary to optimise the operation of the treatment facility. The analyses conducted for the study show high potential of the derived dependence for mixed liquor suspended solid simulations. The dependence is based on the operational parameters (WAS, RAS, pH, DO, etc.), measured on-line, and on the values of wastewater quality indices (BOD, COD, SS, NH_4^+) obtained from computations using models relying on Q and T values. In this case, the highest values of errors in MLSS simulations were found for the model taking into account the load of biogenic compounds in the inflow to WWTP and the reactor parameters, namely pH and T_{sl}.

In practical applications, the statistical models developed for the prediction of mixed liquor suspended solid allow limiting the scope of measurements of quality indices and the bioreactor operational parameters. The models also permit to assess the influence of simplifications on the simulation results. A relevant practical aspect of the analyses conducted for the study is the possibility of making predictions of mixed liquor suspended solid, and consequently, of the substrate load of activated sludge on the basis of the determined flow functions that describe individual wastewater quality indices. That is of key importance, because in the study analyses, the control and optimisation of the values of F/M and MLSS in the WWTP can be based exclusively on the results of inflow and wastewater temperature measurements and AST operational parameters selected by the facility operator.

Substrate load computations performed for this study, based on the results of simulations flow rate, and the values of biochemical oxygen demand and mixed liquor suspended solid demonstrated that for the SVM–based models, extreme values of MAE and MAPE errors did not differ more than 17.81% and 22.83%, respectively. The results indicate a relatively narrow range of variation in the absolute error and relative substrate load, despite much differentiation in independent variables and predictive abilities of models for mixed liquor suspended solid simulations. The lowest error values were obtained for the model, in which the input data were BOD, COD, TSS and NH_4^+ loads and reactor parameters (pH, T_{sl}, m_{met}). Conversely, the highest error values were found for the model based exclusively on the load of biogenic compounds in the inflow to the WWTP. Considering prediction accuracy with respect to BOD load and MLSS value, it can be concluded that with respect to substrate load, the load of biodegradable carbon compounds was decisive for the error in the F/M value simulations.

References

Bagheri M., Mirbagheri S. A., Ehteshami M., Bagheri Z., Kamarkhani A. M. (2016). Analysis of variables affecting mixed liquor volatile suspended solids and prediction of effluent quality parameters in a real wastewater treatment plant. Desalination and Water Treatment. Vol. 57, No. 45. doi: 10.1080/19443994.2015.1125796.

Barbusiński K., Kościelniak H. (1995). Influence of substrate loading intensity on floc size in activated sludge process. Water Research. Vol. 29, pp. 1703 – 1710. doi:10.1016/0043-1354(94)00326-3.

Burges C. (1998) A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2: 121, doi:10.1023/A:1009715923555.

Dellana S.A., West D. (2009). Predictive modeling for wastewater applications: Linear and nonlinear approaches. Environmental Modelling and Software, Vol. 24, pp. 96-106. http://dx.doi.org/10.1016/j.envsoft.2008.06.002.

Eikelboom D.H. (2000). Process control of activated sludge plants by microscopic investigation. IWA Publishing, London, UK.

Gironás J., Roesner L.A., Rossman L. A., Davis J. (2010). A new applications manual for the Storm Water Management Model (SWMM). Environmental Modelling & Software. Vol. 25, pp. 813 – 814. doi: 10.1016/j.envsoft.2009.11.009.

Güçlü D., Dursun Ş. (2010). Artificial neural network modelling of a large-scale wastewater treatment plant operation. Bioprocess and Biosystems Engineering. Vol. 33, No. 9, pp. 1051 - 1058. doi:10.1007/s00449-010-0430-x.

Guo H., Jeong K., Lim J., Jo J., Kim Y.M., Park J.P., Kim J.H., Cho K.H. (2015). Prediction of effluent concentration in a wastewater treatment plant using machine learning models. Journal of Environmental Sciences. Vol. 32, No. 1, pp. 90 – 101. http://dx.doi.org/10.1016/j.jes.2015.01.007

Han H. G, Ying L., Guo Y. N., Qiao J.F. (2016). A soft computing method to predict sludge volume index based on a recurrent self – organizing neural network. Journal Applied Soft Computing. Vol. 38, No. C, pp. 477 – 486. http://dx.doi.org/10.1016/j.asoc.2015.09.051

Henze M., Gujer W., Mino T., Matsuo T., Wentzel M. C., Marais G.v.R., van Loosdrecht M.C.M. (1999). Actived Sludge Model No.2d, ASM2d, Water Science and Technology. Vol. 39, No. 1, pp. 165 – 182. doi:10.1016/S0273-1223(98)00829-4

Henze M., Harremoes P., Arvin E., LaCour Janses J. (2002). Wastewater Treatment. Biological and Chemical Processes. Springer – Verlag. Berlin. ISSN 1431-2492.

Jenkins D., Richard M.G., Daigger G.T. (2004). Manual on the causes and control of activated sludge bulking, foaming, and other solids separation problems, 3rd ed. Lewis Publisher, NY, USA, ISBN 9781566706476.

Kusiak A., Verma A., Wei X. (2013). A data – mining approach to predict influent quality. Environemntal Monitoring and Assessment. Vol. 185, pp. 2197 – 2210. doi: 10.1007/s10661-012-2701-2

Madoni P., Davoli D., Gibin G. (1999). Survey of filamentous microorganisms from bulking and foaming activated-sludge plants in Italy, Water Res., 34, 1767–1772. doi: 10.1016/S0043-1354(99)00352-8.

Mąkinia J. (2010). Mathematical Modeling and Computer Simulation of Actived Sludge Systems. IWA, London. doi: 10.2166/9781780401683

Verma A., Wei X., Kusiak A. (2013). Predicting the total suspended solids in wastewater: A data – mining approach. Engineering Applications of Artificial Intelligence. Vol. 26, No. 4, pp. 1366 – 1372. http://dx.doi.org/10.1016/j.engappai.2012.08.015

Wei X., Kusiak A. (2015). Short-term prediction of influent flow in wastewater treatment plant. Stochastic Environmental Research and Risk Assessment, Vol. 29, No. 1, pp. 241 – 249. doi:10.1007/s00477-014-0889-0.

Zawilski M., Brzezińska A. (2013). Areal rainfall intensity distribution over an urban area and its effect on a combined sewerage system. Urban Water Journal, Vol. 11, No. 7, pp. 532 - 542. doi: 10.1080/1573062X.2013.831909.

HU B., MATAR O.K., HEWITT G.F., ANGELI P., Population balance modelling of phase inversion in liquid–liquid pipeline flows, Chemical Engineering Science, 2006, Vol. 61, 4994–4997.

HUISMAN J.L., BURCKHARDT S., LARSEN T., KREBS P., GUJER W., Propagation of waves and dissolved compounds in a sewer, Journal of Environmental Engineering ASCE, 2000, Vol. 128, No. 1, 12–20

Zealand J., Russell M. (1984). Simple control equations using F:M ratio for operation of activated sludge aeration basins. Environmental Technology Letters. Vol. 5, No. 1 - 11, pp. 49 -56.

Chua H., Yu P.H., Sin S.N., Tan K.N. (2004). Effect of food:microorganism ratio in activated sludge foam control. Twenty-First Symposium on Biotechnology for Fuels and Chemicals Held May 2–6, 1999, in Fort Collins, Colorado, pp. 127 – 135.