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.

III. A data mining approach to the prediction of substrate load and mixed liquor suspended solid

Bartosz Szeląg, Jan Studziński

Introduction

The operation of the wastewater treatment plant (WWTP) is a very complicated process in which the bioreactor technological parameters must be maintained within an appropriate range so that the required effect of pollutant reduction could be achieved. Observations of the treatment facility operation, and of the processes that occur in the activated sludge made it possible to define the parameters for the sludge evaluation and design. According to the literature review, the key parameters include the food-to-mass ratio (F/M) and the activated sludge age (ASA) [1,2]. The food-to-mass ratio value should be treated as a factor that reduces the purifying effect of the activated sludge. This effect can be expressed as the load of the organic substance to be decomposed ($Q \cdot BOD_5$, where: Q – wastewater inflow and BOD_5 – Biochemical Oxygen Demand), which is delivered into the aeration tank (AT), and which needs to be removed using a certain amount of the sludge ($MLSS \cdot V$, where MLSS – mixed liquor suspended solids in AT having the volume of V). Depending on the task posed, the mode of operation of the treatment facility can be adjusted to produce simple or complex symbiosis of organisms. The processes in the facility can be designed in such a manner that they lead to either self-purification (i.e. the organic substance degradation by microorganisms), or to complete oxidation of the organic substance contained in the wastewater. Consequently, systems can be categorised as low and high food-to-mass ratio. In low food-to-mass ratio systems, F/M varies in the range 0.05-0.20 $gBOD_5/MLSS \cdot d$ the sludge undergoes aerobic treatment, in which compounds, newly created and stored in the microorganism cells, are oxidized. In high load systems, F/M value ranges $0.4 \div 1.5$ $gBOD_5/gMLSS \cdot d$ and the sludge needs to be subjected to anaerobic treatment. In the case of wastewater treatment systems intended to remove the compounds of organic, nitrogen and phosphorus from the wastewater, F/M should not exceed the value of 0.10 $gBOD_5/gMLSS \cdot d$ [3,4]. However, if F/M values are below 0.05 $gBOD_5/gMLSS \cdot d$, sludge sedimentation problems, caused by filamentous microorganisms, can arise [5,6].

Numerous investigations and tests performed at wastewater treatment plants [1,2] confirm the considerable impact of food-to-mass ratio on the activated sludge age. The latter determines the time microorganisms stay in the bioreactor. In practice, this time is calculated as the quotient of the amount of the excess sludge (WAS) removed from AT and the total amount of the sludge in the tank ($MLSS \cdot V$). For WWTP operation, it is important that the food-to-mass ratio would not be higher than the ability of microorganisms to metabolize the wastewater pollutants. On the other hand, food-to-mass ratio must not be too low because that can lead to the situation when endogenous respiration outperforms the catabolism of the external carbon sources, and consequently, to the biomass extinction. On the basis of the above observations, it can be concluded that to obtain the required effect of the reduction of biogenic compounds in the wastewater, the values of the activated sludge age and food-to-mass ratio should remain within the defined range. That, however, is not easy to achieve due to the fact that the wastewater inflow (Q) and biochemical oxygen demand (BOD_5) are stochastic in character. Abnormal events like heavy rainfalls, sudden inflow of wastewater with low carbon content into the sewage system, and others, make it difficult to keep the F/M values within the range that ensures the proper operation of the treatment facility [7,8]. In order to obtain high level of pollutant reduction, and to increase the efficiency of the WWTP operation, it is necessary to model Q and BOD_5 variables sufficiently in advance. That will offer the WWTP operator the possibility of specifying the right value of MLSS by regulating the rate of sludge recirculation, the concentration of recirculated sludge, the amount of the removed excess sludge, etc. In case when the permissible MLSS value ($2,5 \div 5,0$ kg/m^3) is exceeded, it is also possibile to use suplemental external carbon sources (methanol, ethanol, wastewater leachate, etc) to increase the load of the biodegradable compounds in the influent wastewater.

The literature review shows [9,10,11] that data mining methods are used to model the amount and quality of wastewater influent, and also the operation of bioreactors. These include, among others, Artificial Neural Networks, Support Vector Machines, Random Forests, Boosted Trees, and k-Nearest Neighbour [12,13,14]. In the methods, at the training stage, the model structure is formulated based on historical data on different input variables. The structure is decisive for the quality of model output. The analysis of the available literature [9,15,16] demonstrates that the modelling of wastewater influent by means of the data mining methods is relatively easy, and the statistical models show satisfactory predicting abilities, which is indicated by the values of mean absolute and relative errors. In order to determine the mixed liquor suspended solids (MLSS), the values of several parameters describing the wastewater quality (including BOD_5, COD, TSS, NH_4^+) and technological parameters of bioreactor operation (recirculation rate, recirculated sludge concentration, amount of excess sludge removed, sludge temperature and pH, etc.) have to be known [17]. In modern treatment facilities, the AT operational parameters are monitored on-line. Conversely, a majority of influent wastewater quality indicators are laboratory determined, which generates high costs. Addditionally, some technical problems concerning parameter determination can arise. Biochemical oxygen demand determination is particularly problematic as it takes as long as five days.

This paper presents the methodology of modelling the mixed liquor suspended solid and food-to-mass ratio. The statistical models for the predictions of the wastewater influent (Q), biochemical oxygen demand (BOD_5) and mixed liquor suspended solids (MLSS) were developed. Due to the fact that BOD_5 determination is difficult to perform, the possibility of predicting this wastewater quality index using the wastewater influent and the COD data was analysed. To determine MLSS, the measurements of BOD_5, COD, TSS, TN and NH_4^+, and also of the bioreactor operational parameters (recirculation rate, sludge pH and temperature, and amount of excess sludge removed) were applied. Because of high costs of determining wastewater quality indicators, the possibility of predicting those quantities on the basis of the flow rate recorded in the last measurements was taken in to account. The analyses performed for the paper made it possible to assess the impact of errors of wastewater quality prediction on the results of modelling of the food-to-mass ratio and mixed liquor suspended solids.

Object of investigation

The object investigated is the wastewater treatment plan (WWTP) located in the terrain of the commune of Sitkówka – Nowiny. The plant collects sanitary wastewater from the city of Kielce, the commune of Sitkówka – Nowiny, and partially also from 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. The wastewater inflow to WWTP is mechanically treated using bar screens and aerated grit chambers, with grease separators in first step. Next, wastewater is pumped to four primary clarifiers, from which it is delivered to the biological unit, i.e. bioreactor with separate denitrification and nitrification tanks (Fig. 1).

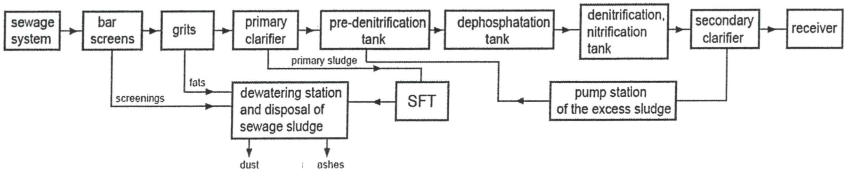

Figure 1. Technological diagram of the Sitkówka – Nowiny treatment plant.

In preliminary denitrification tanks, into which the activated sludge is recirculated, partial removal of nitrogen compounds occurs. Afterwards, wastewater is conveyed to dephosphatation tanks for the removal of phosphorus compounds. Then, wastewater together with activated sludge is transferred to four secondary clarifiers, from where after clarification, it flows to the receiving water, i.e. the river Bobrza.

Continuous monitoring conducted by the company Wodociągi Kieleckie Sp. z o.o. at the treatment plant since 2012 provides measurements of parameters describing influent wastewater quantity and quality, and also operational parameters of the aeration tanks.

Methodology

In the investigation statistical models for forecasting the activated sludge concentration and substrate load have been calculated. The modelling of MLSS and F/M variables has been done in several variants. At the first the possibility of MLSS prediction on the basis of the amount and quality of sewage inflow reaching the treatment plant and of technological parameters of the biological reactor was considered what can be written down in form of the following equation:

$$MLSS(t)_{pred} =$$
$$f(Q(t), BOD_5(t), COD(t), TN(t), NH_4^+(t), TSS(t), pH(t), T_{sl}(t), RAS(t), WAS(t)) \quad (1)$$

where: BOD_5 – biochemical oxygen demand, COD – chemical oxygen demand, TN – Nitrogen total, NH_4^+ - Ammonium-nitrogen, TSS – total sediment, Q – daily sewage inflow into treatment works, RAS – recirculation grade, WAS – amount of excessive sewage derived from the treatment works.

In other calculation variants the possibility of modelling and forecasting the variables BOD_5, COD, TN, NH_4^+, TSS as well as Q and T_{sl} on the basis of the sewage flow measurements has been considered. For the determining of BOD_5 value is very time and work consuming then the possibility of its prediction on the basis of only COD(t) and of COD(t) and Q(t) values has been additionally tested.

Statistical models to forecast the MLSS variable in which the values of Q, T_{sl} and of some sewage quality indicators are taken under consideration can be described with the following equation:

$$MLSS(t)_{pred} = f(Q(t-i), T_{sl}(t-k), C(t)_j, pH, RAS, WAS) \quad (2)$$

where: i, k – time delay between the forecasted value of the sewage inflow or of the sludge temperature and of the variable modelled; i = 1,2,3…m, m=7; k = 1,2,3…p = 7; $C(t)_{j,pred}$ –values of the sewage quality indicators calculated from the following relation:

$$C(t)_{j,pred} = f(Q(t-1), Q(t-2), \dots, Q(t-m))_j \quad (3)$$

and in the case of BOD_5, calculated from the relations:

$$BOD_{5,pred}(t) = f(Q(t), COD(t)) \quad (4)$$

$$BOD_{5,pred}(t) = f(COD(t)) \quad (5)$$

where: j – number of analysed sewage quality indicators (BOD_5, COD, TN, NH_4^+, TSS), j =5, m – time delay between the value of the quality indicator modelled and the value of the appropriate independent variable.

In the following the statistical models to predict the substrate load of the activated sludge have been calculated using the equation:

$$F/_M = \frac{Q(t)_{pred} \cdot BOD_{5,pred}(t)}{V_{KOC} \cdot MLSS(t)_{pred}} \tag{6}$$

where: $MLSS(t)_{pred}$ –concentration of the activated sludge forecasted by means of equations (1), (2), $BOD_{5,pred}(t)$ – values of biochemical oxygen demand calculated by means of the relations (3 - 5), $Q(t)_{pred}$ – sewage inflow into the treatment works predicted with the use of its earlier values Q(t-i).

The analyses presented above are intended to demonstrate the possibility of modelling F/M values on the basis of the data on wastewater quality, temperature and inflow. In everyday operation of the treatment facility, such analyses are important because they make it possible to reduce the number of variables which should be measured in the influent wastewater to determine the food-to-mass ratio. Additionally, the statistical models for the prediction of the food-to-mass ratio take into account the operational parameters of the bioreactor (recirculation rate) selected by the facility staff. As a result, the model facilitate the control of those parameters in advance so that the optimal operation of the facility could be ensured.

In this paper, MARS and ANN methods were applied to model the C(t), Q(t), $T_{sl}(t)$ and MLSS(t) variables. In subsequent simulations, the computational results for which the predictions that showed best fit to measurement data were used. Before the start of the modelling, the measurement data were standardized using the min-max transformation:

$$\overline{A_i} = \frac{A_i - minA}{maxA - minA} \tag{7}$$

where: $\overline{A_i}$ – normalized value of i-th element in set A, A_i – measured value of i-th element in set A, max A – maximum value of the elements in set A, min A – minimum value of the elements in set A.

The ANN methods are widely applied as they can be used to simulate linear and nonlinear processes, as well as to solve the tasks of optimization, classification and control [11,12,16]. The multilayer perceptron (MLP) is the most commonly used structure of neural network. In the MLP, the input signals are multiplied by weight values, and afterwards transferred to the neurons of the hidden layer. In the particluar neurons, the following–summation occurs.: $z_m=\Sigma x_n \cdot w_{nm}-b_j$, where m is the number of neurons in the network, $b_{1,k}$ – threshold or bias. The sums received are then transformed using a linear or nonlinear activation function and transferred to the output neurons. The optimal values of weights for individual neurons are determined by training.

With respect to the prediction of MLSS, activated sludge temperature and wastewater quality indicators (BOD$_5$, COD, TSS, TN, NH_4^+) recommendations for selecting the neural structure are not available. Consequently, Automatic Designer function of the STATISTICA program was used. Five hundred different neural networks were separately generated for the prediction of the quantities mentioned above and the parameters of computations fit to measurement data were given. As the optimal model the neural network has been accepted for which the error values (MAE and MAPE) have been smallest under remaining 499 models.

It was assumed that the minimum number of neurons in the hidden layer equal 5 neurons and maximum equal 20 neurons. In the hidden neuron layer and the output layer, the following activation functions were considered: hyperbolic tangent, logistic, sine, exponential and expotential. To make the training process correct, and then to properly assess the performance of the statistical models applied, the dataset was partitioned into the training set (75%), and the testing set (25%). Analysing the measurements there was decides that the training and testing sets will be consisted of 250 data concerning the sludge concentration, substrate load and the sewage quality parameters (BOD_5, COD, TN, NH_4^+, TSS). To calculate the models to forecast the daily sewage inflow and the activated sludge temperature the data sets consisting of 1250 measurements for each calculation have been used. In the calculation the training and testing sets have been chosen randomly. The neural network training was implemented using the Broyden – Fletcher – Goldfarb – Shanno algorithm.

The Multivariate Adaptive Regression Splines (MARS) method is one of numerous tools used for data exploration [18,19]. It constitutes an extension of the classical approach to predictors in regression models. In the classical approach, independent variables are treated uniformly, whereas in the MARS method, variation ranges of the input data of concern are divided into subranges in which independent variables can have different impacts on the process investigated. The boundaries of subranges are determined on the basis of threshold values (t). That means different weights or signs can be attributed to a variable in the model, depending on whether the variable in question is below or above the value of (t). The differentiation of independent variables into lower and higher than the threshold values (t_i) is performed using the following basis function:

$$h(X) = \alpha_i \cdot \left(max(0, X - t)\right) \qquad (8)$$

where: h(X) – vector of basis functions for individual variables (x_i) for which the condition:

$$x_i - t_i = \begin{cases} x_i - t_i; & for\ x_i > t_i \\ 0; & x_i \le t_i \end{cases} \qquad (9)$$

is fulfilled.

In the MARS method, the regression relation is a spline function obtained from a linear combination of the product of basis functions and weights:

$$f(X) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \cdot h_m(X) \quad (10)$$

where: $X=[x_1, x_2, \ldots, x_i]$ – vector of input data, α_m – values of weights, h_m – basis functions.

To determine the model parameters, a special algorithm was developed to search the observation space in order to compute the threshold values (nodes). The algorithm uses the recursive partitioning of the feature space and it comprises two stages that occur alternately until the stopping criterion is satisfied. The criterion constitutes the value of generalized error in fivefold cross-validation [20]. In the first stage of the algorithm, the model complexity is enhanced by adding basis functions until the maximum function number, set by the user, is reached. In the second stage of the algorithm, the procedure of elimination from the model (pruning) of the least important basis functions is started. Thus, independent variables whose removal causes the smallest decrease in predictive abilities of the model are eliminated.

To assess predictive abilities of the models employed to forecast the daily wastewater inflow, chemical and biological oxygen demands, total suspended solids, total nitrogen and ammonia nitrogen, activated sludge temperature and mixed liquor suspended solids, the following error formulas were applied:

- mean absolute error (MAE)

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| y_{i,obs} - y_{i,pred} \right| \quad (11)$$

- mean relative error (MPE)

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_{i,obs} - y_{i,pred}}{y_{i,obs}} \right| \cdot 100\% \quad (12)$$

where: $y_{i,obs,pred}$ – measured and calculated concentration values, respectively, n –dataset size.

Results

This study concerns the statistical models developed for the prediction of wastewater quantity, quality, and operational parameters of the bioreactor. It is therefore necessary to determine the range of variation of parameters (Table 1) in which those models can be employed.

The data in Table 1 show that the indicators of quantity and quality of the influent wastewater varied substantially. That led to changes in mixed liquor suspended solids and, consequently in the aeration tank food-to-mass ratio. For instance, in the time period of concern, the BOD_5 parameter ranged 127 ÷ 557 mg/dm^3, the wastewater inflow varied from 32564 m^3/d up to 86592 m^3/d and MLSS changed in the range of 1.19 ÷ 5.89 kg/m^3. A significant variation in the food-to-mass ratio values (0.03 ÷ 0.07 g BOD_5/gMLSS·d) substantiates the need for its modelling in order to improve the efficiency of the wastewater treatment facility operation.

In the MARS model, the algorithm for parameter estimation allows the removal of independent variables that have negligible effect on the dependent variable. Because of that this method has been used firstly and by means of it the independent variables concerning the sewage quality (BOD_5, COD, TN, NH_4, TSS), the daily sewage inflow into the treatment plant and the sludge temperature have been identified. Afterwards the statistical models to forecast the parameters included into equation (1) have been calculated using MARS and ANN methods.

First, based on the results of analyses carried out with the MARS method, variables were identified and statistical models for the prediction of daily wastewater inflow and the sludge temperature in the aeration tank were designed. The results of computations carried out with the methods employed in the study are presented in Table 2.

Table 1. Range of variation of parameters describing the wastewater inflow (Q), wastewater quality (BOD, COD, TSS, TN, NH_4^+) and the bioreactor operation (T_{sl}, pH, MLSS, RAS, WAS, F/M).

Variable	Minimum	Average	Maximum
Q, m^3/d	32564	40698	86592
T_{sl}, ^0C	10.0	15.9	23.0
pH	7.00	7.6	8.1
MLSS, kg/m^3	1.19	4.26	5.89
RAS, %	44.6	90.70	167.6
WAS, kg/d	3489	11123	19194
F/M, gBOD/gMLSS·d	0.03	0.07	0.13
BOD, mg/dm^3	127	309	557
COD, mg/dm^3	384	791	1250
TSS, mg/dm^3	126	329	572
TN, mg/dm^3	39.9	77.7	124.1
NH_4^+, mg/dm^3	24.4	49.31	65.9

Table 2. Parameters of fit of Q and T_{sl} computations with MARS and ANN methods to the results of measurements.

Variable	ANN				MARS			
	Training		Testing		Training		Testing	
	MAE, m^3/d	MAPE, %	MAE, m^3/d	MAPE, %	MAE, m^3/d	MAPE, %	MAE, m^3/d	MAPE, %
Q	2956	6.48	3037	7.24	3004	7.11	3050	7.95
T_{sl}	0.87	5.32	0.92	6.08	0.92	5.87	0.96	6.26

As regarding the models predicting Q(t) and $T_{sl}(t)$, it is sufficient to have the values Q(t-1) and $T_{sl}(t-1)$, and the number of the basic functions for which the results are best fitted to the measurements is 3, in both cases. Basing on this observation the ANN models to predict Q(t) and $T_{sl}(t)$ and using the values of the independent variables calculated by means of the MARS method have been performed. From the calculations done results that the smallest error values while predicting Q(t) have been received for the ANN model in which the number of neurons on the hidden layer was equal to 4 and as the activation function the tangent hyperbolic function was taken. There was identified also that the smallest error by the $T_{sl}(t)$ prediction has been got with the ANN model in which the neurons number on the hidden layer was equal to 3 and as the activation function the sinusoidal function was taken.

The results (Table 2) indicate that the ANN method produced lower values of mean absolute and relative errors than it was the case with the MARS method. For the first method, the values of Q prediction errors were MAE = $3037 m^3/d$ and MAPE = 7.24%, and for the second one MAE = $3050 m^3/d$ and

MAPE = 7.95%. As regards T_{sl} predictions, the respective values were MAE = 0.92°C, MAPE = 6.08% and MAE = 0.96°C, MAPE = 6.26%.

In the following the statistical models to forecast the activated sludge concentration have been prepared. It was done by means of MARS and ANN methods and under consideration of the variables included in equation (1) and concerning the amount and quality of the sewage as well as operational parameters of the biological reactor.

The next step of analysis involved modelling mixed liquor suspended solids with the MARS and ANN methods on the basis of the wastewater amount and quality indicators, and also the operational parameters of the bioreactor. The computation results are shown in Table 3. In the model obtained with the MARS method, the number of basic functions was equal to 14. Computations carried out by using the ANN method showed that among 500 generated neural networks structures, the one with 12 neurons in the hidden layer and with the exponential activation function of the hidden layer and linear activation function of the output layer produced the best results.

Table 3. Parameters of fit of mixed liquor suspended solids (MLSS) computations with MARS and ANN methods to the results of measurements.

Method	Training		Testing	
	MAE, kg/m^3	MAPE, %	MAE, kg/m^3	MAPE, %
MARS	0.47	12.52	0.52	13.03
ANN	0.38	10.07	0.46	11.81

The data in Table 3 indicate that the statistical model for MLSS prediction based on the ANN method has slightly better predictive abilities (lower MAE, MAPE) than the one developed using the MARS method. In the first case, values of mean absolute and relative errors were MAE = 0.46kg/m^3 and MAPE = 11.81% , in the second case MAE = 0.52kg/m^3 and MAPE = 13.03%.

Table 4. Parameters of fit of wastewater quality indicators computations with ANN and MARS methods to the results of measurements.

Wastewater quality indicators	ANN				MARS			
	Training		Testing		Training		Testing	
	MAE, mg/dm^3	MAPE, %	MAE, mg/dm^3	MAPE, %	MAE, mg/dm^3	MAPE, %	MAE, mg/dm^3	MAPE, %
BOD = f(Q(t-i))	33.24	13.09	38.94	13.66	45	16.2	47.04	17.2
BOD = f(COD(t))	52.58	18.28	54.93	19.84	62.14	22.34	62.89	22.95
BOD = f(COD(t),Q(t))	43.52	13.82	44.4	15.9	51.26	17.33	57.74	19.49
COD	96.04	14.82	110.46	14.97	107.7	15.22	119.52	16.17
SS	37.47	12.38	40.8	14.13	49.34	17.71	53.25	18.89
NH_4^+	2.82	5.59	3.05	6.23	4.25	8.67	4.35	8.94
TN	4.38	5.86	4.75	6.11	5.44	6.98	5.71	7.38

The models for MLSS(t) (eq. 1) prediction make it possible to simulate this technological parameter in time t with satisfactory accuracy, which is confirmed by computed values of MAE and MAPE (Table 3). However, for the bioreactor operation optimization, it is necessary to predict MLSS in advance, not just simulate MLSS at the instant of t. Earlier predictions allow pre-setting MLSS and the activated sludge recirculation rate. Additionally, the variables taken into account in the model were wastewater quality indicators, the measurements of which are costly and not always possible to take. Consequently, it was decided to simulate wastewater quality indicators (BOD_5, COD, TSS, TN, NH_4^+) in time t on the basis of wastewater inflow values obtained from previous measurements.

Table 5. Parameters describing the structures of the ANN models for the prediction of wastewater quality indicators.

Quality indicators	Number of neurons in the hidden layer	Activation function of the hidden layer	Activation function of the output layer
BOD = f(Q(t-i))	5	tanh	tanh
BOD = f(COD(t))	6	logistic	exp
BOD = f(COD(t),Q(t))	5	tanh	tanh
COD	4	exp	tanh
SS	5	logistic	exp
NH_4^+	6	tanh	exp
TN	4	exp	linear

Using the MARS method, independent variables Q(t-i) for wastewater quality indicators were identified, and consequently regression models were developed. On the basis of computations, the parameters of the neural network structures and of the fit of computation results, obtained for MARS and ANN methods, to measurement data were listed. They are presented in Tables 4 and 5. Simulation of the influent wastewater quality indicators based on inflow to WWTP, performed using the MARS method, showed that independent variables of wastewater quality indicators, i.e. BOD_5, COD, TSS are Q(t-1), Q(t-2) and Q(t-3) values. For the other cases, i.e. TN and NH_4^+, those variables are Q(t-1), Q(t-2) quantities. In the MARS-based models, the number of basis functions ranged 3 ÷ 6. In the ANN-based models, the number of neurons in the hidden layer varied from 4 up to 6, and the activation functions of the hidden and output layers was most often hyperbolic tangents (Table 4).

The results of computations (Table 4) show that with respect to wastewater quality indicators, the ANN method performed slightly better (lower value MAE, MAPE) than the MARS method. For instance, MARS-based model for COD prediction generated mean absolute and relative errors MAE = 119.52 mg/dm^3 and MAPE = 16.17%, whereas for the ANN-based model those were MAE = 110.46 mg/dm^3 and MAPE = 14.97%. As regards the BOD_5 prediction models, the lowest error values were found for the models in which input variables were the influent wastewater quantities (eq. 3) from the last three measurements. The error vales were as follows: MAE = 38.94 mg/dm^3 and MAPE = 13.66% for the ANN method, and MAE = 47.0 mg/dm^3 and MAPE = 17.20% for the MARS method. Conversely, the highest error values in the BOD_5 prediction were obtained for the models in which the input variables were the COD measurements (eq. 5). Then, the error values were MAE = 54.93 mg/dm^3

and MAPE = 19.84% for the ANN method, and MAE = 62.89 mg/dm^3 and MAPE = 22.95% for the MARS method.

The results of simulation of quality indicators (BOD_5, COD, TSS, TN, NH_4^+), the amount of wastewater (Q), and the bioreactor operation (T_{sl}, pH, RAS, WAS) were taken into account when designing the ANN-based model for MLSS(t) prediction. In the following the concentration of the activated sludge using formula (2) and under consideration of relations (3 – 5) has been determined. Afterwards the substrate load of the sludge by means of relation (6) has been calculated.

Basing on the computations, the parameters of fit of MLSS and F/M simulations to the measurement results were specified (Table 6). In addition, MLSS and F/M values measured at a week's interval and the ones computed for the period of concern were compared in Figs. 2 and 3. The comparison was made for the variants where BOD_5 was determined exclusively on the basis of Q(t-1) and COD(t).

Table 6. Parameters of fit of computations of the aeration tank technological parameters (MLSS and F/M) to the results of measurements.

Variables in the BOD models	MLSS		F/M	
	MAE, kg/m^3	MAPE, %	MAE, $gBOD_5/gMLSS \cdot d$	MAPE, %
Q(t-i)	0.49	11.95	0.013	19.64
COD(t)	0.57	14.16	0.016	24.71
COD(t),Q(t)	0.55	13.9	0.015	21.71

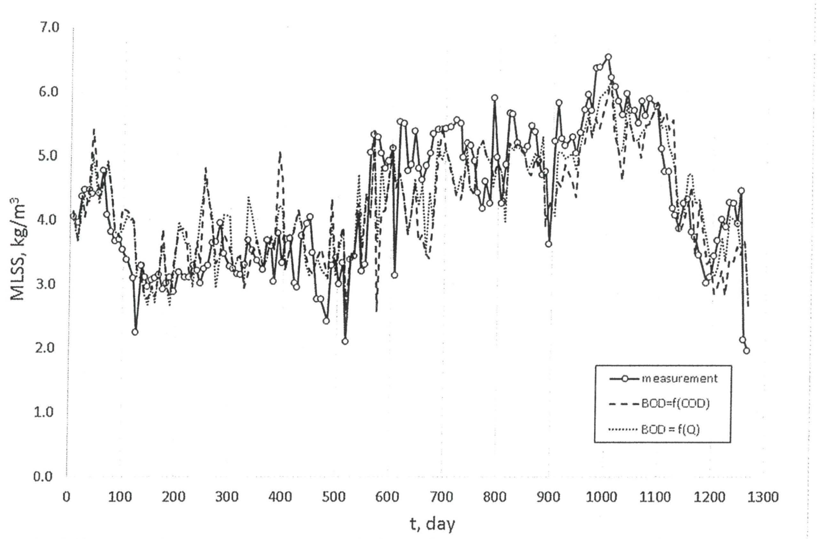

Figure 2. Comparison of the results of measurements and computations of mixed liquor suspended solids (MLSS) in the period of concern.

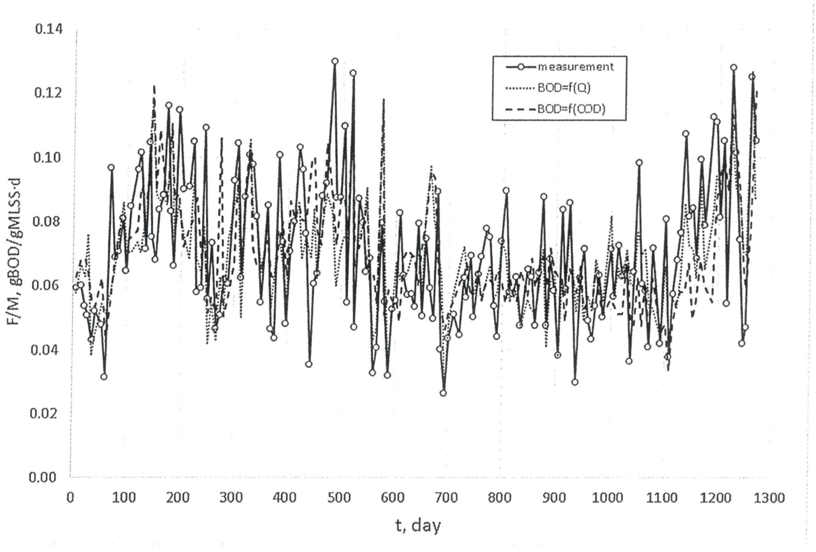

Figure 3. Comparison of the results of measurements and computations of the substrate load (F/M) in the period of concern.

The data presented in Table 6 show that the lowest values of errors in the prediction of the technological parmeters were found when the BOD_5 value was a function of only the inflow rate Q(t-1). In two other cases, the results of MLSS and F/M simulations did not differ much. For instance, for the F/M prediction model based on BOD_5 = f(Q(t-i)) (eq.3), mean error values were MAE = $0.027gBOD_5/gMLSS \cdot d$ and MAPE = 19.64%. For the model based on BOD_5 = f(COD(t)) (eq.5), errors were MAE = $0.033gBOD_5/gMLSS \cdot d$ and MAPE = 24.71%.

The computations performed for the study demonstrate that the data on influent wastewater flow rate obtained from the last measurements can be used to model the indicators of wastewater quality. That is confirmed by relevant prediction errors. Modelling provides a useful tool in practical applications. It allows predicting, in advance, the operational parameters of the aeration tanks. Their performance can be optimised using the variables measured online in the bioreactor, and the data on the flow rate of influent wastewater.

Conclusions

The modelling results show that the values of wastewater quality indicators, namely TN, NH_4^+ and also BOD_5, COD, TSS can be determined on the basis of wastewater inflow values obtained, respectively, from the last two and three measurements. In the cases considered, the ANN method produced lower errors in the prediction of wastewater quality indicators and MLSS than the MARS-based models. MLSS computations were performed using the models that describe wastewater quality indicators, determined on the basis of inflow rate and the bioreactor parameters. The results of simulations are in satisfactory congruence with measurement data. The results of simulations of wastewater quantity and quality, and also of MLSS were used to predict food-to-mass ratio. The models designed to that end produced computational results that were congruent with measurements. That was confirmed by the values of mean absolute and relative errors. In the operation of treatment facilities, modelling makes it possible to reduce the costs of measurements of biogenic compounds in the influent. Additionally, the bioreactor parameters (RAS, WAS, pH, T_{sl}), measured online, can be forecast and controlled, which is necessary to ensure an adequate degree of pollutant reduction. Also, the performance of the wastewater treatment plant can be enhanced due to the control of food-to-mass ratio.

References

[1] MĄKINIA J. Mathematical Modeling and Computer Simulation of Actived Sludge Systems. IWA, London 2010.

[2] HENZE M., HARREMOES P., ARVIN E., LACOUR JANSEN J. Wastewater Treatment. Biological and Chemical Processes. Springer – Verlag. Berlin 2002.

[3] BARBUSIŃSKI K., KOŚCIELNIAK H. Influence of substrate loading intensity on floc size in activated sludge process. Water Res. **29**, 1703, 1995.

[4] BEVER J., STEIN A., TEICHMANN H. Zaawansowane metody oczyszczania ścieków. Projprzem – EKO. Bydgoszcz, 1997.

[5] MADONI P., DAVOLI D., GIBIN G. (1999). Survey of filamentous microorganisms from bulking and foaming activated-sludge plants in Italy. Water Res. **34**, 1767, 1999.

[6] ŁOMOTOWSKI J., SZPINDOR A. Nowoczesne systemy oczyszczania ścieków. Wydawnictwo Arkady. Warszawa, 2002.

[7] ZEALAND J., RUSSELL M. Simple control equations using F:M ratio for operation of activated sludge aeration basins. Environmental Technology Letters **5**(1 – 11), 49, 1984.

[8] DYMACZEWSKI Z., JAROSZYŃSKI T., JEŻ – WALKOWIAK J., KAUFMAN – KOMOROWSKA M., MICHAŁKIEWICZ M., NIEDZIELSKI W., SOZAŃSKI M. Poradnik eksploatatora oczyszczalni ścieków, wydanie 3 rozszerzone. PZITS, Poznań, 2011.

[9] DELLANA S.A., WEST D. Predictive modeling for wastewater applications: Linear and nonlinear approaches. Environ. Model. Software **24**, 96, 2009.

[10] CHUCHRO M. Prediction of biogas production in separate digesters chambers from wastewater treatment plant. Pol. J. Environ. Stud. **21**(5A), 31, 2012.

[11] BARTKIEWICZ L., SZELĄG B., STUDZIŃSKI J. Ocena wpływu zmiennych wejściowych oraz struktury modelu sztucznej sieci neuronowej na prognozowanie dopływu ścieków komunalnych do oczyszczalni. Ochrona Srodowiska **38**(2), 29, 2016.

[12] RAVIKUMAR R., RENUKA K., SINDHU V., MALARMATHI K.B. Response Surface Methodology and Artificial Neural Network for Modeling and Optimization of Distillery Spent Wash Treatment Using Phormidium valderianum BDU 140441. Pol. J. Environ. Stud. **22**(4), 1143, 2013.

[13] BUIL H. M., DUONG H. T. G., NGUYEN C. D. Applying an Artificial Neural Network to Predict Coagulation Capacity of Reactive Dyeing Wastewater by Chitosan. Pol. J. Environ. Stud. **25**(2), 545, 2016.

[14] SZELĄG B., GAWDZIK J. Application of Selected Methods of Artificial Intelligence to Actived Sludge Settleability Predictions. Pol. J. Environ. Stud. **25**(4), 1709, 2016.

[15] SOBOTA J., SZETELA R. Jednowymiarowe modele natężenia i jakości ścieków dopływających do oczyszczalni komunalnych. Ochrona Środowiska **27**(1), 15-22, 2005.

[16] STUDZIŃSKI J., BARTKIEWICZ L., STACHURA M. Development of mathematical models for forecasting hydraulic loads of water and wastewater networks. EnviroInfo'2013: Environmental Informatics and Renewable Energies, 1, Aachen, 2013.

[17] GÜÇLÜ D., DURSUN Ş. (2010). Artificial neural network modelling of a large-scale wastewater treatment plant operation. Bioprocess Biosyst Eng **33**(9), 1051, 2010.

[18] SHARDA V. N., PRASHER S. O., PATEL R. M., OJASVI P. R., PRAKASH C. Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan microwatersheds with limited data. Hydrol. Sci. J. **53**(6), 1165, 2008.

[19] GUTIÈRREZ G., SCHNABEL Á. S., CONTADOR J. F. L. Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecol. Mod. **220**(24), 3630, 2009.

[20] FRIEDMAN J. Multivariate Adaptive Regression Splines, Annals of Statistics **19**, 1, 1991.