Pol. J. Environ. Stud. Vol. 26, No. 5 (2017), 2231-2238
DOI: *10.15244/pjoes/68448*

Original Research

A Data Mining Approach to the Prediction of Food-to-Mass Ratio and Mixed Liquor Suspended Solids

Bartosz Szeląg[1]*, Jan Studziński[2]

[1]Faculty of Environmental, Geomatic and Energy Engineering, Kielce University of Technology
Tysiąclecia Państwa Polskiego Av. 7, 25-314 Kielce, Poland
[2]Systems Research Institute PAN
Newelska Street 6, 01-447 Warsaw, Poland

Received: 14 November 2016
Accepted: 12 January 2017

Abstract

This paper presents methodology for constructing a statistical model to forecast food-to-mass ratio (F/M). In the model, wastewater inflow (Q), biochemical oxygen demand (BOD5) and mixed liquor suspended solids (MLSS) were modelled separately using artificial neural networks (ANN) and multivariate adaptive regression splines (MARS). To compute the value of MLSS, the quality indicators of influent wastewater and the operational parameters of the bioreactor were used. It was examined whether it is possible to predict wastewater quality indicators that determine the values of F/M and MLSS on the basis of the wastewater inflow to the treatment plant. Computations performed demonstrated that ANN predictions of MLSS and F/M showed smaller errors than those obtained using the MARS method. Moreover, all developed models of wastewater quality indicators were considered as satisfactory.

Keywords: wastewater treatment, artificial neural network, data mining, multivariate adaptive regression spline, food-to-mass ratio

Introduction

The operation of a wastewater treatment plant (WWTP) is a very complicated process in which the bioreactor's technological parameters must be maintained within an appropriate range so that the required effect of pollutant reduction can 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 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 F/M 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 is wastewater inflow and BOD_5 is biochemical oxygen demand), which is delivered into the aeration tank (AT) and which needs to be removed using a certain amount of sludge ($MLSS \cdot V_{AT}$, where MLSS is mixed-liquor suspended solids in AT having a volume of V). Depending on the task posed, the

*e-mail: bszelag@tu.kielce.pl

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 having low and high F/M ratios. In low F/M systems, F/M varies in the range 0.05-0.20 $gBOD_5$/MLSS·d and 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÷1.5 $gBOD_5$/gMLSS·d and the sludge needs to be subjected to anaerobic treatment. In the case of wastewater treatment systems intended to remove organic compounds, nitrogen, and phosphorus from the wastewater, F/M should not exceed the value of 0.10 $gBOD_5$/gMLSS·d [3-4]. However, if F/M values are below 0.05 $gBOD_5$/gMLSS·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 F/M 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 sludge in the tank ($MLSS \cdot V_{AT}$). For WWTP operation, it is important that the F/M ratio would not be higher than the ability of microorganisms to metabolize the wastewater pollutants. On the other hand, the F/M ratio must not be too low because that can lead to a situation when endogenous respiration outperforms the catabolism of the external carbon sources, and consequently, to 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 F/M 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 a 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. This 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. To increase F/M value, it is possible to use supplemental external carbon sources, i.e., methanol, ethanol, wastewater leachate, and others.

The literature review [9-11] shows 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 [12-14]. 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 wastewater quality (including BOD_5, COD, TSS, NH_4^+) and technological parameters of the bioreactor operation (recirculation rate, recirculated sludge concentration, amount of excess sludge removed, sludge temperature, pH, etc.) have to be known [17]. In modern treatment facilities, the AT operational parameters are monitored online. Conversely, a majority of influent wastewater quality indicators are laboratory determined, which generates high costs. Additionally, 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 F/M ratios. 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 into account. The analyses performed for this paper made it possible to assess the impact of errors of wastewater quality prediction on the results of modelling of the F/M ratio and MLSS.

Object of Investigation

The object investigated is the WWTP located in 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 influent wastewater is mechanically pre-treated using bar screens and aerated grit chambers, with grease separators. Next, wastewater is pumped to four primary clarifiers, from which it is delivered to the biological unit (i.e., a bioreactor with separate denitrification and nitrification tanks, Fig. 1. In preliminary denitrification tanks, into which the activated

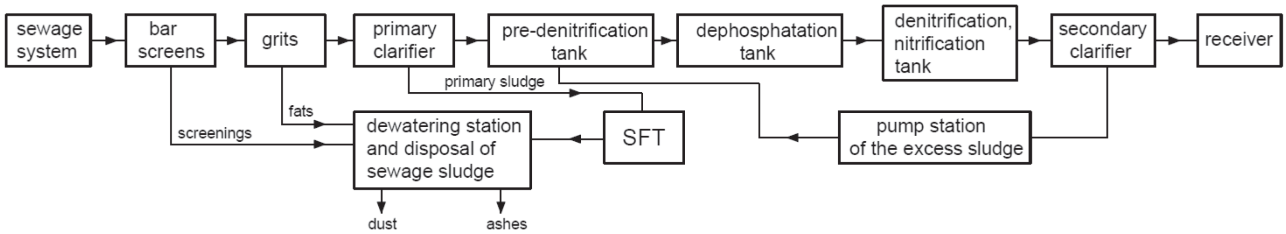

Fig. 1. Technological diagram of the Sitkówka-Nowiny treatment plant.

sludge is recirculated, partial removal of nitrogen compounds occurs. Afterward, 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 which after clarification it flows to the receiving water, i.e., the Bobrza River. 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 this study, statistical models were developed for predicting MLSS and F/M ratio. MLSS and F/M simulations were performed using a few variants. In the first variant, the MLSS prediction was considered, based on the influent quantity and quality, and also the bioreactor technological parameters. The general formula of this model variant can be written as follows:

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

…where BOD_5 is biological oxygen demand, COD is chemical oxygen demand, TN is total nitrogen, NH_4^+ is ammonia nitrogen, TSS is total suspended solids, Q is daily wastewater inflow, RAS is return activated sludge, and WAS is waste activated sludge.

For subsequent analyses, due to the problems related to the measurements of wastewater quality indicators and the necessity of ensuring high efficiency of the aeration tank in case of the measurement system failure, the possibility of substituting the actual values of BOD_5, COD, TN, NH_4^+, TSS, and also Q and T_{sl} with the results of the computations of those values was taken into account. It was analysed whether the wastewater quality indicators mentioned above could be predicted on the basis of flow data. Additionally, because BOD_5 value determination is time-consuming and troublesome, the prediction of this indicator was based exclusively on COD(t), and also COD(t) and Q(t). For selected wastewater quality indicators, the statistical models for MLSS value prediction, in which Q and T_{sl} are modelled, could be expressed with a general formula:

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

…where i and k are the time shift between the predicted value of activated sludge flow/temperature at the instant t, and the successive independent variables of the modelled quantity; i = 1, 2, 3…m, m = 7; k = 1, 2, 3… p = 7; and $C(t)_{j,pred}$ is computed values of the wastewater quality indicators of concern based on the following dependence:

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

…and, additionally, for the value of BOD_5:

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

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

…where j is the number of wastewater quality indicators analysed (BOD_5, COD, TN, NH_4^+, TSS), j = 5, and m is the time shift between the modelled value of the selected quality indicator and the last independent variable of the given indicator-predicted magnitude.

The next stage involved the development of the statistical models for determining F/M of activated sludge based on the following equation:

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

…where $MLSS(t)_{pred}$ is MLSS predicted on the basis of Eqs. (1) and (2); $BOD_{5,pred}(t)$ is modelled values of biochemical oxygen demand based on dependences (3÷5), and $Q(t)_{pred}$ is predicted inflow to WWT based on the value of 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 that should be measured in the influent wastewater to determine the F/M ratio. Additionally, the statistical models for the prediction of the F/M ratio take into account the operational parameters of the bioreactor (recirculation rate) selected by the facility staff. As a result, the model facilitates 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:

$$\bar{A}_i = \frac{A_i - minA}{maxA - minA} \quad (7)$$

…where $\bar{A}_i$ is normalized value of i-th element in set A, A_i is measured value of i-th element in set A, max A is maximum value of the elements in set A, and min A is 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 afterward transferred to the neurons of the hidden layer. In the individual neurons, the summation occurs. 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, the automatic designer function of the STATISTICA program was used. Five-hundred different neural networks were generated for the prediction of each of the quantities mentioned above and the parameters of computations fit to measurement data were given. The optimal model that was selected was this neural network for which the computed error values (MAE, MAPE) were the lowest among all 500 ANN models. It was assumed that the minimum number of neurons in the hidden layer was equal to five neurons and maximum was equal to 20 neurons. In the hidden neuron layer and the output layer, the following activation functions were considered: hyperbolic tangent, logistic, sine, and exponential. To make the training, it was partitioned into the training set (75%), and the testing set (25%). Based on the measurement results, it was found that the training set and the testing set each comprise 250 values process correct, and then to properly assess the performance of the statistical models applied, the dataset of MLSS, F/M, and wastewater quality indicators (BOD_5, COD, TN, NH_4^+, and TSS). The datasets, each including 1,250 values, provided a basis for the development of the models for the prediction of daily wastewater inflow and activated sludge temperature. In the computations, the data for the training set and the testing set were selected 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:

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

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

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

is fulfilled.

In the MARS method, the regression relationship 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, …, x_i] – vector of input data, α_m is values of weights, and h_m is 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 five-fold 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 it started. Thus, independent variables whose removal causes the smallest decrease in predictive abilities of the model are eliminated.

The following error formulas were applied 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:

Table 1. Range of variation of parameters describing wastewater inflow (Q), wastewater quality (BOD_5, COD, TSS, TN, NH_4^+), and bioreactor operation (T_{sl}, pH, MLSS, RAS, WAS, F/M).

Variable	Minimum	Average	Maximum
Q (m^3/d)	32,564	40,698	86,592
T_{sl} (ºC)	10.0	15.9	23.0
pH	7.00	7.6	8.1
MLSS (kg/m^3)	1.97	4.26	6.59
RAS (%)	44.6	90.70	167.6
WAS (kg/d)	3,489	11,123	19,194
F/M ($gBOD_5$/gMLSS·d)	0.03	0.07	0.13
BOD_5 (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

– 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 (MAPE)

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

… where $y_{i,obs},_{pred}$ is measured and calculated concentration values, respectively, and n is 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 MLSS and, consequently F/M in the aeration tank. For instance, in the time period of concern, the BOD_5 parameter ranged 127÷557 mg/dm^3, the wastewater inflow varied from 32,564 m^3/d up to 86,592 m^3/d, and MLSS changed in the range of 1.97÷6.59 kg/m^3. A significant variation in the F/M 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.

As in the MARS model, the algorithm for parameter estimation allows for the removal of independent variables that have a negligible effect on the dependent variable. Therefore, using the MARS method, first the independent variables of the modelled wastewater quality indicators (BOD_5, COD, TN, NH_4, and TSS) daily inflow to WWTP and the activated sludge temperature were identified. Next, the statistical models, based on MARS and ANN methods, were developed for the prediction of the above-mentioned variables found in Eq. (1).

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. As regards the model predicting Q(t) and T_{sl}(t), it is sufficient to have the values Q(t-1) and T_{sl}(t-1), and in both cases the number of basis functions for which the results are best fitted to measurements is three. On that basis, ANN models for the prediction of Q(t) and T_{sl}(t) were developed. The models relied on the values of independent variables obtained using the MARS method. The computations performed with the ANN method show that the lowest values of errors of Q(t) prediction where obtained when the number of neurons in the hidden layer was four, and the activation function took the form of hyperbolic tangent. Additionally, the lowest T_{sl}(t) prediction error was found for the model, in which the number of neurons in the hidden layer was five, and the activation function was a sine dependence. The results (Table 2) indicate that the ANN method produced lower values of mean absolute and relative errors than was the case with the MARS method. For the first method, the values of Q prediction errors were MAE = 3,037m^3/d and MAPE = 7.24%, and for the second one MAE = 3,050m^3/d and MAPE = 7.95%. As regards T_{sl} predictions, the respective values were MAE = 0.92ºC, MAPE = 6.08%, and MAE = 0.96ºC, MAPE = 6.26%.

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	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
	(m^3/d)	(%)	(m^3/d)	(%)	(m^3/d)	(%)	(m^3/d)	(%)
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

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

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

The next stage involved the development of the statistical models, based on MARS and ANN methods, for the prediction of MLSS. The models relied on the independent variables found in Eq. (1), which included the wastewater amount and quality indicators, and also the bioreactor operational parameters. The computation results are shown in Table 3. In the model obtained with the MARS method, the number of basis 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. The data in Table 3 indicate that the statistical model for the 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 = 0.46 kg/m^3 and MAPE = 11.81%, and in the second case MAE = 0.52 kg/m^3 and MAPE = 13.03%.

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 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 RAS. 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. 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 the inflow to the WWTP, performed using the MARS method, showed that independent variables of wastewater quality indicators, i.e. BOD_5, COD, and 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) and 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 were most often hyperbolic tangents (Table 4).

Table 4. 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_5 = f(Q(t-i))	5	tanh	tanh
BOD_5 = f(COD(t))	6	logistic	exp
BOD_5 = f(COD(t),Q(t))	5	tanh	tanh
COD	4	exp	tanh
TSS	5	logistic	exp
NH_4^+	6	tanh	exp
TN	4	exp	linear

The results of computations (Table 5) show that with respect to wastewater quality indicators, the ANN method performed slightly better (lower values of MAE, MAPE) than the MARS method. For instance, the 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 values 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 next stage, MLSS was determined on the basis of Formula (2), while taking into account dependences (3-5). The next step involved F/M computations using Formula (6). Based 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

Table 5. Parameters of fit of wastewater quality indicator computations with ANN and MARS methods to the results of measurements.

Wastewater quality indicators	ANN				MARS			
	Training		Testing		Training		Testing	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
	(mg/dm³)	(%)	(mg/dm³)	(%)	(mg/dm³)	(%)	(mg/dm³)	(%)
BOD_5 = f(Q(t-i))	33.24	13.09	38.94	13.66	45	16.2	47.04	17.2
BOD_5 = f(COD(t))	52.58	18.28	54.93	19.84	62.14	22.34	62.89	22.95
BOD_5 = 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
TSS	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

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).

The data presented in Table 6 show that the lowest values of errors in the prediction of the technological parameters 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·d and MAPE = 19.64%. For the model based on BOD_5 = f(COD(t)) (Eq.5), errors were MAE = 0.033gBOD_5/gMLSS·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.

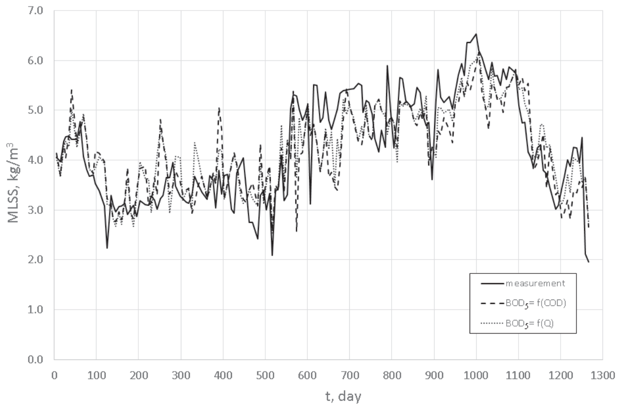

Fig. 2. Comparison of the results of measurements and computations of mixed-liquor suspended solids (MLSS) in the period of concern.

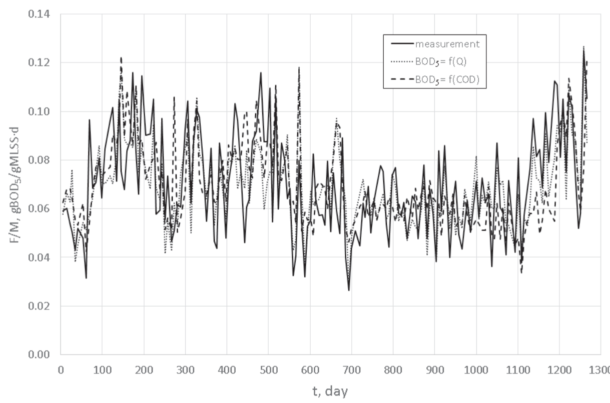

Fig. 3. Comparison of the results of measurements and computations of the food-to-mass ratio (F/M) in the period of concern.

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	MAPE	MAE	MAPE
	(kg/m³)	(%)	($gBOD_5$/ gMLSS·d)	(%)
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

Conclusions

The modelling results show that the values of wastewater quality indicators, namely TN, NH_4^+, and also BOD_5, COD, and 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 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 F/M 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 F/M ratio.

References

1. MĄKINIA J. Mathematical Modeling and Computer Simulation of Activated 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. 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, **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. 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**.