Assessment of the Effect of Wastewater Quantity and Quality, and Sludge Parameters on Predictive Abilities of Non-Linear Models for Activated Sludge Settleability Predictions

Bartosz Szeląg, Jarosław Gawdzik*

Faculty of Environmental, Geomatic and Energy Engineering, Kielce University of Technology,
Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland

Abstract

This paper discusses the possibility of applying the three "black-box" methods to sludge settleability predictions. Additionally, the impact of the load of biogenic compounds and parameters of activated sludge on predictive abilities of the devised mathematical models is analysed in the paper. To conduct analyses we relied on the results of measurements of wastewater quantity and quality, and of the bioreactor operational parameters, taken on continuous basis at the Sitkówka-Nowiny treatment plant in 2012-16. The analyses conducted for the study indicate that the lowest values of errors in activated sludge settleability predictions for the wastewater treatment plant of concern were obtained for input data on the load of biogenic compounds at the inflow, microorganism culture environment, and activated sludge concentration.

Keywords: activated sludge settleability, SVM, k-NN, MARS methods

Introduction

Computer software (GPS-X, WEST, STOAT, BioWin, SIMBA) is widely used to assess the operation of wastewater treatment plants and variability of technological parameters of activated sludge and wastewater. However, to calibrate physical models it is necessary to collect huge amounts of data on the processes taking place in selected units of the treatment plant. Due to a large number of parameters and interactions between them, the calibration of models tends to be complex and requires the use of algorithms to automate this process [1-2]. On account of this, the so-called "black-box" methods are also employed to model the processes occurring in the wastewater treatment plant (biogas production, electric power consumption). In those methods, the structure of the statistical model is generated at the training stage, for which knowledge of the physics of the phenomenon is not required. To model treatment plant operation, artificial neural networks (ANN) are most frequently used. The methods of support vector machines (SVM), k-nearest neighbour (k-NN), and regression trees or multivariate adaptive regression splines (MARS) are employed far less frequently. The literature review [3-4] indicates that except for MARS, the methods mentioned above have been used to predict the amount and quality of wastewater

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

conveyed to the treatment facility, and the efficiency of removal of biogenic compounds contained in wastewater. Additionally, the k-NN and also the SVM methods have been employed to model the quantity of biogas, sludge volume index, and electric power demand [5-6].

One of the crucial parameters that affect the efficiency of wastewater treatment using the activated sludge process is the assessment of the sludge settling ability. Deterioration of sludge settleability can lead to an increase in the content of sludge and carbon in wastewater at the secondary settler outlet above the levels permissible for effluent. In addition, poor sludge settleability can cause problems with surplus sludge treatment. To prevent such difficulties, it is important to develop mathematical models that are capable of predicting the parameters of activated sludge settleability. The literature review [6-7] indicates that the value of sludge volume index usually provides the basis for assessing sludge settleability. To model the index, different types of neural networks are used, including hierarchical, probabilistic, cellular, and hybrid ones, which combine the models mentioned above with networks used for classification. A parameter decisive for the value of the sludge index is activated sludge settleability, which is expressed as activated sludge volume after 30-minute settling in the Imhoff cone. It is not by chance that this parameter has become a basis for evaluating settling ability of activated sludge in many treatment facilities. Due to biochemical processes occurring in activated sludge, and also to its metastable character (also in measurements), the sludge index, which is inherently an intensive variable, is not always an optimal data source for modelling the operation of the wastewater treatment plant. An alternative approach based on an extensive variable, expressed by means of activated sludge settleability, seems a justifiable option. In view of the above, it is reasonable to develop mathematical models to predict the technological parameter of concern. As the literature review [6-8] does not provide data on sludge settleability predictions, the effect of the quantity and quality of inflowing wastewater and operational parameters of aeration tanks on the value of sludge settleability cannot have been analysed.

This paper discusses the possibility of applying the three "black-box" methods to sludge settleability predictions. Additionally, we analyse the impact of the load of biogenic compounds and parameters of activated sludge on predictive abilities of the devised mathematical models. To conduct analyses, we relied on the results of measurements of wastewater quantity and quality, and of the bioreactor operational parameters taken on a continuous basis at the Sitkówka-Nowiny treatment plant in 2012-16.

The Object of Investigations and the Material

The investigations concerned the urban wastewater treatment plant 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 in 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 pre-treated using step screens and aerated grit chambers, with separate grease traps. Next, wastewater is pumped to four primary settling tanks, from which primary sludge is delivered to the secondary sludge pumping station. Having left the mechanical unit of the treatment plant, wastewater is delivered into the biological unit. The treatment proceeds according to the modified (multiphase denitrification) three-stage Bardenpho process. In preliminary denitrification tanks, to which activated sludge is recycled, nitrogen compounds are partially removed from wastewater. Wastewater is then directed to dephosphatation tanks to remove phosphorus compounds. A bioreactor constitutes the core component of the treatment plant. In the bioreactor's separate denitrification and nitrification tanks, contaminants are finally removed from wastewater. From the aeration tank, wastewater together with activated sludge is transferred to four secondary settling tanks, 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 inflowing wastewater quantity and quality, and also operational parameters of the aeration tanks.

Methodology

To compute activated sludge settleability, the parameters describing the quantity and quality of inflowing wastewater were used. Those include the following: concentration of chemical oxygen demand (COD), total suspended solids (SS), total phosphorus (TP), total nitrogen (TN), ammonia nitrogen (NH$_4^+$), and five-day biochemical oxygen demand (BOD). In the analyses, the bioreactor operational parameters – namely temperature, pH, and activated sludge concentration – were also taken into account.

A number of combinations of the parameters mentioned above were taken into consideration. First, we analysed the possibility of forecasting the technological parameter of concern on the basis of the loads of biogenic compounds (BOD, COD, TN, TP), and also ammonia nitrogen and total suspended solids in inflowing wastewater. Also, an attempt was made to find out for which biogenic substances the generated statistical model would show the best predictive abilities with respect to activated sludge settleability. Then, on the basis of the results of analyses obtained in the first stage, the best mathematical model for the determination of the settleability (SE) value was selected. The model accounted for the data describing the microorganism culture environment, which include pH and sludge temperature. In the final stage of

computations, in the best model obtained on the basis of previous analyses, activated sludge concentration was taken into account. In addition, the possibility of applying substrate load to the determination of activated sludge settleability was considered. In the simulations, the possibility of forecasting SE value on the basis of the input data on wastewater quantity and bioreactor operational parameters was taken into consideration. The aim of the analyses mentioned above was to determine the impact of individual input data on the predictive abilities of the developed statistical models for activated sludge settleability predictions. Another goal was to identify the method for which prediction error values with respect to the technological parameter of concern were the lowest.

With input data presented above and appropriate assumptions made, statistical models for activated sludge settleability prediction were developed using the SVM, k-NN, and MARS methods. To make the training process appropriate, and then to properly assess the performance of a statistical model, the data were partitioned into the training set (75%) and the validating and testing set (25%). Prior to the start of the construction of mathematical models, input and output data normalization was performed by means of normal form transformation [9]:

$$\overline{A}_i = \frac{A_i - \min A}{\max A - \min A}$$ (1)

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

Support vector machines (SVM) cover a group of methods developed by [10] 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 much larger size is applied. In this study, the support vector regression method with the radial kernel function was applied to predict activated sludge settleability. The radial kernel function was used to minimise the function of the following form:

$$\sum_{i=1}^{m} \frac{c}{m} |y_i - f(x_i)|_\varepsilon + \frac{1}{2} \cdot \|f\|_k^2$$ (2)

...where $|y_i - f(x_i)|_\varepsilon = \max\{0, |y_i - f(x_i) - \varepsilon\}$, ε is permissible error, $\|f\|$ is norm f in Hilbert space, c is constant selected by the user (depending on value ε [11]), m is the size of the training set, and $f(x_i)$ is value of function f(x) at point x_i, described by equation:

$$f(x) = \sum_{i=1}^{N_{sv}} (\alpha_i - \alpha_i^{'}) \cdot K(x, x_i) + w_0$$ (3)

...in which w_0 is deviation; N_{sv} is the number of support vectors, dependent on C; ε, α_i, and $\alpha_i^{'}$ are Lagrange multipliers, and $K(x,x_i)$ is kernel function with radial basis functions [11].

The nearest neighbour method (k-NN) is one of the simplest non-parametric methods, and like those already mentioned can be applied to classification and regression problems [12]. In this case, the dependent variable prediction is expressed by formula:

$$\hat{y} = \frac{1}{K} \cdot \sum_{i=1}^{N} y_i J(x_i, x_j)$$ (4)

...in which x_i is one of K number of nearest neighbours of x_j when distance $d(x_i, x_j)$ belongs to the smallest distances between observations from set $Z_N = \{(x_1, y_1), \ldots, (x_n, y_n)\} \in R^{m+1}$ where: $x_i = (x_{1,i}, \ldots, x_{p,i})$ is the i-th vector of independent variables with m number of coordinates, y_i is the i-th dependent variable, N is the number of observations, and $J(x_i, x_j)$ is a function of the form:

$$J(x_i, x_j) = \begin{cases} 1, & \text{when } x_i \text{ is one of K of nearest neighbours of } x_j \\ 0, & \text{otherwise} \end{cases}$$ (5)

In the computations shown above, Euclidean (employed here) and Mahalanobis distances are most frequently used. The number of neighbours (K) was established by trial and error, seeking such a value of K, for which the model devised would show the best predictive abilities.

The MARS method is one of the numerous data-mining tools used for solving regression problems [13-16]. The method is an extension of the classic approach to input data in the regression model. In addition to overall account of predictors, as is the case in the classic regression model, in the MARS method the variation ranges of individual independent variables are divided into three intervals, in which the predictors of concern may have a different effect on the analysed phenomenon. Interval limits are established on the basis of threshold values (t), which means that depending on whether a variable is below or above value t, in the model the variable of concern may have a different weight or sign. As regards individual input data values, to differentiate between those being smaller or greater than the threshold value t_i, it is necessary to use the basis function described by dependence:

$$h(X) = \alpha_i \cdot (\max(0, X - t))$$ (6)

...where h(X) is vector of basis functions for individual variables (x_i), for which the following condition is satisfied:

$$x_i - t_i = \begin{cases} x_i - t_i; & \text{for} \quad x_i > t_i \\ 0; & \text{for} \quad x_i \leq t_i \end{cases} \tag{7}$$

In the MARS method, regression dependence has a form of a spline function, which is a linear combination of the product of basis functions and individual weights, and which can be written as follows:

$$f(X) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \cdot h_m(X) \tag{8}$$

…where $X = [x_1, x_2, …, x_i]$ is vector of input data, α_m is values of weights, and h_m is basis functions.

To estimate model parameters, a special algorithm was designed [17] to search the space of observations in order to specify threshold values (nodes). Using selected nodes, basis functions are created, which together with appropriate weights constitute a foundation for the description of the analysed phenomenon. It should be noted that so far this method has not been used to model the operation of the wastewater treatment plant.

To evaluate predictive abilities of mathematical models and regression dependences (2), (3) and (7) for computing activated sludge settleability, 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 absolute percentage 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 is data set size, $y_{i,(obs, \ pred)}$ is measured and calculated value of activated sludge settleability, and $\overline{y}_{i, obs, pred}$ is arithmetic mean of measured/calculated sludge settleability.

Results and Discussion

First, input data were relied upon to establish the range of variation in values for which the mathematical models developed for the study could be used. Table 1 lists mean values and range of variation (minimum and maximum values) of analysed bioreactor parameters, and also the quantity and quality of wastewater flowing to the treatment plant. In the time period investigated (2012-16), total nitrogen (TN) concentration ranged

Table 1. Range of variation (medium, minimum, maximum) in values of parameters describing the inflowing wastewater quality and parameters of bioreactor operation.

Variable	Range	Mean
Q (m³/d)	32,564-86,592	41,584
T (ºC)	10.6-23.0	15.9
pH	7.2-7.8	7.5
BOD (mg/dm³)	127-557	320
COD (mg/dm³)	384-1250	790
SS (mg/dm³)	45-168	91.4
TP (mg/dm³)	4.3-12.6	7.8
TN (mg/dm³)	39.91-124.09	77.6
NH4 (mg/dm³)	24.40-65.90	48.86
MLSS (kg/dm³)	1.91÷6.59	3.57
F/M (g BOD5/g smo·d)	0.057-0.375	0.151
SE (cm³/dm³)	250-980	785

Q (1) is daily wastewater inflow to the treatment plant, T (2) is temperature of activated sludge,
pH (3) is concentration of hydrogen cations, BOD (4) is concentration of five-day biochemical oxygen demand, COD (5) is chemical oxygen demand, SS (6) is suspended solids, TP (7) is concentration of total phosphorus, TN (8) and NH$_4$ (9) are concentrations of total and ammonia nitrogen, MLSS (10) is concentration of activated sludge in the aeration tanks, and F/M (11) food-to-mass ratio from the formula Q·BOD/(MLSS·V$_K$), where V$_K$ is active volume of aeration tanks and SE is activated sludge settleability.

39.91-124.09 mg/dm³, that of total phosphorus ranged 43-12.6 mg/dm³, and the values of BOD and COD parameters were 127-557 mg/dm³ and 384-1,250 mg/dm³, respectively. Sludge temperature and pH varied from 10.6-23.0ºC and from 7.2-7.8, respectively. Activated sludge concentration in the aeration tanks was between 1.91 kg/m³ and 6.59 kg/m³, and the mean value was 3.57 kg/m³. The daily amount of wastewater flowing into the treatment plant was 32,564-86,592 m³/d. Substantial changes in BOD, Q, and MLSS led to considerable differentiation in the food-to-mass ratio (0.058-0.375 gBOD/g smo·d).

Additionally, from the data in Table 1 it can be seen that in the period analysed, activated sludge settleability ranged 250-980 cm³/dm³, whereas the mean value of this parameter (SE) was 785 cm³/dm³.

On the basis of the inflowing wastewater quantity and physico-chemical parameters, and also the bioreactor operational parameters (Table 1), mathematical models for activated sludge settleability modelling were developed using the support vector machines (SVM), multivariate adaptive regression splines (MARS), and k-nearest neighbour (k-NN) methods. Table 2 shows the values of parameters (MAE, MAPE) that describe the fit of

computational results to measurements of a technological parameter. In the statistical models (Table 2), the value of C ranged 9-12 and $\varepsilon = 0.001$ for the SVM method and value K was from 5 to 8 for k-NN method, whereas for MARS the number of basis functions varied 2-7.

The calculations conducted for the study (Table 2) indicate that for the loads of biogenic compounds taken into consideration, the lowest values of settleability prediction errors were obtained when the model input was five-day biochemical oxygen demand. Conversely, the worst results of sludge settleability simulations were found for total phosphorus load. In addition, the analyses (Table 2) indicate that sludge settleability predictions with SVM, k-NN, and MARS methods on the basis of the data on BOD and COD loads did not differ considerably from the results obtained on the basis of BOD loads. With respect to the methods examined in the study, the best results of activated sludge settleability predictions were produced by MARS, for which the error values of SE prediction on the basis of BOD were MAE = 101.4 cm³/dm³ and MAPE = 13.8%. Conversely, the worst results of SE value calculations were found for the model developed with the k-NN method (MAE = 143.8 cm³/dm³ and MAPE = 20.6%). The data in Table 2 show that taking into account the total nitrogen load in the model designed on the basis of BOD load leads to a slight improvement in predictive abilities of statistical models. In the methods analysed in the study, the best results of sludge settleability calculations were achieved using the model based on the MARS method (MAE = 91.7 cm³/dm³ and MAPE = 14.5%), and the worst for the k-NN-based model, for which error

values are MAE = 130.1 cm³/dm³ and MAPE = 18.8%. The SVM method produced slightly worse results of SE simulations, namely MAE = 103.7 cm³/dm³ and MAPE = 15.4%, compared with the MARS method. The analysis of the data from Table 2 shows that in the statistical model based on BOD and total nitrogen loads, the addition of another independent variable, i.e., ammonia nitrogen, leads to a decrease in the absolute and relative error values in activated sludge settleability prediction. For instance, in both MARS and SVM methods, the values of absolute error (MAE) compared with the previous model (BOD, TN) were lower by 10%, i.e., those decreased from 91.7 cm³/dm³ to 83.1 cm³/dm³ and from 103.7 cm³/dm³ to 94.2 cm³/dm³, respectively.

In addition, analysis results indicate that when the total phosphorus load is introduced into the existing model (BOD, TN, NH_4^+), the predictive abilities of the statistical model are improved, which is shown by the MAE and MAPE error values (Table 2). In this case, as in the previous ones, the lowest values of error in sludge settleability prediction were obtained using the MARS method (MAE = 69 cm³/dm³ and MAPE = 9.4%), and the worst using the k-NN method (MAE = 109.6 cm³/dm³ and MAPE = 16.5%). Other analyses demonstrated that the load of total suspended solids or sludge pH do not significantly affect activated sludge sedimentation, which is shown by values of errors in settleability predictions. As regards suspended solids, that may result from the fact that a major portion of solids is removed in the primary settling tank. For the case of pH, it may be due to the low volatility index, which was

Table 2. Fit parameters (MAE, MAPE) for mathematical models obtained with support vector machine (SVM), k-nearest neighbour (k-NN), and MARS methods.

Variables	SVM				k-NN				MARS			
	Training		Test, validation		Training		Test, validation		Training		Test, validation	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
1+4*	106.5	14.1	114.6	16.9	131.5	17.5	139.8	20.6	94.1	13.5	101.4	15.9
1+5	113.1	15.0	122.7	18.2	141.1	18.0	150.4	21.3	103.2	15.01	111.8	17.7
1+5+4*	112.5	14.8	116.1	17.3	131.6	17.5	140.1	20.7	98.7	14.2	103.9	16.5
1+8	117.9	15.9	127.6	18.9	143.2	18.2	151.8	21.7	101.1	14.8	109.5	17.3
1+7	123.0	16.7	132.8	19.7	145.5	19	153.6	22.4	102.2	15.7	111.8	17.7
1+4+8*	96.5	13.2	103.7	15.4	123.1	15.7	130.8	18.8	85.6	13.2	91.7	14.5
1+4+8+9*	85.8	12.1	92.2	13.7	113.8	15.1	121.6	17.9	76.1	10.7	81.1	12.3
1+4+8+9+7*	72.5	10.5	85.2	12.0	104.2	13.8	109.6	16.5	58.9	8.1	69.0	9.4
1+4+8+9+7+6	82.1	11.6	88.5	13.1	105.8	14.1	112.3	16.8	62.2	8.6	71.8	9.9
1+4+8+9+7+3	83.9	11.9	90.1	13.3	111.0	14.7	109.4	17.6	60.4	7.9	66.1	9.5
1+4+8+9+7+2*	62.0	8.2	70.2	10.1	84.7	11.1	89.2	13.6	42.2	7.6	45.2	8.0
1+4+8+9+7+2+10*	50.1	6.9	50.6	8.2	65.7	8.2	70.3	10.2	27.1	3.9	30.6	4.0
1+4+8+9+7+2+11	73.1	9.9	63.2	11.9	71.3	9.1	77.3	11.5	35.1	4.7	39.1	5.4
1+2+10	74.3	10.0	81.1	12.0	82.2	11.1	88.4	13.8	58.2	8.2	64.1	9.9

estimated on the level of approximately 0.6. Although the presented range of variation (Table 1) correspond to approximately four times the growth of the activity of oxonium ions, in industrial practice often only slightly different from the actual measurement accuracy in the aeration tanks. When in the statistical model, which accounts for BOD, TN, NH4, and TP, another parameter is introduced (in the previous case, pH was analysed) which describes microorganism culture environment, i.e., sludge temperature and values of activated sludge settleability prediction are reduced. In this case, just like with previous models, the best simulation fit-to-measurement data was obtained using the MARS method (MAE = 45.2 cm^3/dm^3 and MAPE = 8.0%). Much worse SE prediction values were found for SVM (MAE = 70.2 cm^3/dm^3 and MAPE = 10.4%) and k-NN (MAE = 89.2 cm^3/dm^3 and MAPE = 13.6%) methods.

The data listed in Table 2 shows that statistical models, which account for independent variables such as loads of organic, nitrogen, phosphorus compounds, and also sludge parameters (i.e., its concentration and temperature), show better predictive abilities with respect to activated sludge settleability. That is confirmed by values of errors in computation fit-to-measurement data. As regards the methods under consideration (SVM, k-NN, MARS), the lowest values of error in SE prediction were found for MARS (MAE = 30.6 cm^3/dm^3 and MAPE = 4.0%). In the remaining models based on SVM and k-NN, the values of error in simulation fit to measurement data (Table 2) are MAE = 50.6 cm^3/dm^3 and MAPE = 7.0%, and also MAE = 70.3 cm^3/dm^3 and MAPE = 10.2%, respectively. Additionally, on the basis of computation results presented in Table 2 it can be seen that mathematical models relying on data describing the quality of wastewater flowing

to the treatment plant and sludge parameters (pH and concentration) show better predictive abilities with respect to SE than models developed on the basis of, among others, substrate load. The determined values of absolute errors in sludge settleability prediction for food-to-mass ratio are as follows: MAE = 39.1 cm^3/dm^3 for the MARS method, MAE = 77.3 cm^3/dm^3 for the k-NN-based model, and MAE = 80.2 cm^3/dm^3 for SVM. Values of MAE errors in mathematical models based on food-to-mass ratio F/M are greater by 23%, 10%, and 37%, respectively, from those calculated for models based on activated sludge concentration.

Fit parameters (Table 2) indicate that statistical models developed on the basis of input data on activated sludge parameters (temperature and concentration) possess predictive abilities that are similar to those of models determined for the loads of biogenic compounds (BOD, TN, NH$_4^+$, TP) flowing to the treatment plant. For instance, for the MARS model based on input data concerning sludge concentration and temperature, the value of absolute error in SE prediction was 64.1 cm^3/dm^3. That is only 8% greater than the error value computed for data on the load of biogenic compounds flowing to the treatment plant.

Fig. 1 presents the results of computations in a graphic form. In the figure, the values of activated sludge settleability, measured and obtained from simulations, are compared. The simulations were based on mathematical models for which the values of error in SE prediction were the lowest.

The analyses showed that the mathematical models for predicting activated sludge settleability on the basis of variables that describe inflowing wastewater quantity and quality, and also bioreactor operational parameters (or
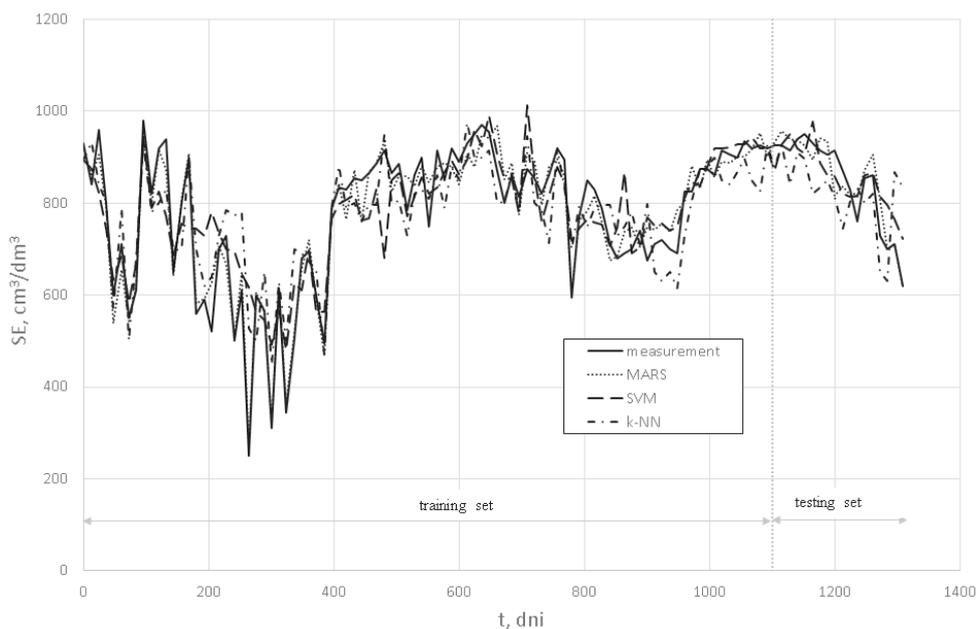

Fig. 1. Comparison of the results of measurements and computations of activated sludge settleability using the SVM, k-NN, and MARS methods.

combinations of those) possess diverse predictive abilities. Another finding is that SVM, k-NN, and MARS methods can be used to model activated sludge settleability.

Conclusions

The activated sludge process is often used in modern biological wastewater treatment plants in which carbon, nitrogen, and phosphorus are removed. In this technology, flock forms the fundamental structural component. To achieve a high degree of contaminant load reduction in the treatment process, it is necessary to produce flock that has appropriate settling properties that can be identified due to the value of sludge settleability. It is therefore necessary to model settleability and to identify parameters that are decisive for settleability variation.

To predict settleability in this study we used black-box methods such as support vectors, k-nearest neighbour, and MARS. With respect to the methods considered, the best results of sludge settleability computations were obtained using the MARS method, whereas those produced by k-NN were much worse. Computations show that models for predicting activated sludge settleability developed on the basis of inflowing wastewater quantity and quality, bioreactor operational parameters, and also combinations of those, possess diverse predictive abilities. This is indicated by values of errors specifying simulation fit-to-measurement data. The greatest errors in activated sludge settleability predictions were found for independent variables that describe only the load of biogenic compounds in inflowing wastewater. Taking into account, in statistical models, factors including microorganism culture environment (activated sludge temperature) and parameters of bioreactor operation (concentration of activated sludge) contributed to the improvement in simulation results. The analyses conducted for the study indicate that the lowest values of errors in activated sludge settleability predictions for the wastewater treatment plant of concern were obtained for input data on the load of biogenic compounds at the inflow, microorganism culture environment, and activated sludge concentrations.

References

1. ROMANOWICZ R.J., KICZKO A., NAPIÓRKOWSKI J.J. Stochastic Transfer Function Simulator of a 1-D Flow Routing. Publs. Inst. Geophys. Pol. Acad. Sc., E-**10** (406), 151, **2008**.
2. KICZKO A., ROMANOWICZ R., OSUCH M., KARAMUZ E. Maximising the usefulness of flood risk assessment for the River Vistula in Warsaw. Natural Hazards and Earth System Sciences, **13**, 3443, **2013**.
3. KUSIAK A., ZENG Y., ZHANG Z. Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach. Engineering Applications of Artificial Intelligence **26** (4), 1643, **2013**.
4. VERMA A., WEI X., KUSIAK A. Predicting the total suspended solids in wastewater: A data-mining approach. Engineering Applications of Artificial Intelligence, **26** (4), 1366, **2013**.
5. WEI X., KUSIAK A. Optimization of Biogas Production Process in a Wastewater Treatment Plant. Proceedings of the 2012 Industrial and Systems Engineering Research Conference **2012**.
6. HAN H., LI Y., GUO Y., QIAO J. A soft computing method to predict sludge volume index based on a recurrent self organizing neural network. Applied Soft Computing. **38**, 477, **2016**.
7. LOU I., ZHAO Y. Sludge Bulking Prediction Using Principle Component Regression and Artificial Neural Network. Mathematical Problems in Engineering. **2012**, 1, **2012**.
8. GIOKAS D.L., DAIGGER G.T., SPERLING M., KIM Y., PARASKEVAS P.A. Comparison and evaluation of empirical zone settling velocity parameters based on sludge volume index using a unified settling characteristics database. Water Research. **37** (16), 3821, **2006**.
9. RUTKOWSKI L. Artificial intelligence methods and techniques Warszawa, PWN, **2006** [In Polish].
10. VAPNIK V. Statistical Learning Theory. John Wiley and Sons. New York, **1998**.
11. BURGES C., A tutorial on support vector machines for pattern recognition (Knowledge discovery and data mining, Usama Fayyad). Kluwer Academic Publishers, Boston. Manufactured in The Netherlands, 1, **2000**.
12. PIOTROWSKI A., NAPIORKOWSKI J., ROWIŃSKI P. Flash-flood forecasting by means of neural networks and nearest neighbour approach – a comparative study. Nonlinear Processes Geophysics, **13**, 443, **2006**.
13. ZHANG W., GOH A.T.C. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. Geoscience Frontiers. **7** (1), 45, **2016**.
14. 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. Hydrological Sciences Journal des Sciences Hydrologiques, **53** (6), 1165, **2008**.
15. DE VEAUX R.D., PSICHOGIOS D.C., UNGAR L.H. A Comparison of two nonparametric estimation schemes: MARS and neural networks. Computers & Chemical Engineering **17** (8), 819, **1993**.
16. GUTIÉRREZ G., SCHNABEL Á., S., CONTADOR J. F. L. Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. Ecological Modelling **220** (24), 3630, **2009**.
17. FRIEDMAN J. Multivariate Adaptive Regression Splines, Annals of Statistics, **19**, 1, **1991**.