Pol. J. Environ. Stud. Vol. 28, No. 5 (2019), 3105-3113
DOI: *10.15244/pjoes/94050* ONLINE PUBLICATION DATE: 2019-05-07

Original Research

Using Data Mining to Predict Sludge and Filamentous Microorganism Sedimentation

Krzysztof Chmielowski[1]*, Adam Czekański[2], Aleksandra Leśniańska[3]

[1]University of Agriculture in Cracow, Department of Sanitary Engineering and Water Management, Kraków, Poland
[2]Sanockie Przedsiębiorstwo Gospodarki Komunalnej Sp. z o.o., Sanok, Poland
[3]EkoWodrol Sp. z o.o., Koszalin, Poland

Received: 30 May 2018
Accepted: 7 August 2018

Abstract

This study attempted to develop statistical regression models for predicting the settleability of activated sludge based on the quality of incoming sewage and on the identified dominant filamentous species. As part of the analyses conducted for the purpose, classification models are presented that enable identification of the respective filamentous microorganisms, based on the working parameters of the bioreactor and the quality of the influent. The study calculations demonstrated that the modeling methods based on artificial neural networks, random forests, and boost trees can be applied for the identification of filamentous microorganisms *Microthrix parvicella*, *Nostocoida* sp., and *Thiotrix* sp. in activated sludge chambers in the STP located in Sitkówka-Nowiny. The best predictive capacity, covering identification of the above-mentioned filamentous bacterial species in activated sludge chambers, was observed for statistical models obtained by the random forest method.

Keywords: activated sludge, hybrid models, settleability

Introduction

Considering the stochasticity of the volume and quality of the incoming sewage, the operating parameters of reactors need to be controlled within a certain range so that the performance of the respective units of a sewage treatment plant (STP) can be optimized and the desirable pollution reduction effect achieved [1, 2]. This requires prediction of various biochemical processes taking place in the reactor, using physical or statistical method for the purpose. The former have the advantage that the qualitative variability of sewage at the discharge from the reactor and the biological reactor's operating parameters is based on differential equation systems. However, for the calibration of physical models, one needs detailed information about the reaction path in the respective units, which requires continuous high-resolution measurements of a number of qualitative parameters of the sewage at the inlet, discharge and inside the reactor, leading to considerable problems in the experimental phase. Moreover, due to the number of model parameters and strong interactions between them, their calibration may often be difficult, as confirmed in multiple

*e-mail: k.chmielowski@ur.krakow.pl

reports [3, 4]. In view of the above, statistical models are also used in which a model structure is generated in the training phase, based on which the parameters of interest will be predicted. In this category, multiple regression is the simplest method and the more complex methods include artificial neural network, support vectors, genetic programming, regression trees and others. Since the processes occurring in activated sludge are very complex, statistical models are used in the simulation of its settling properties, as reported in [5-7]. It is worth noting that continuous control of the parameters determining the sludge settling process is essential to the course of the sewage treatment process. Their deterioration may result in increasingly high concentrations of carbon compounds and suspended solids in the sewage at the discharge of the secondary tank, exceeding their permissible levels. Therefore, it is necessary to create mathematical models enabling the parameters of the activated sludge settling process to be predicted. A literature survey [8-11] indicates that the settleability of sludge is assessed based on the sludge index. On the other hand, the previously used models failed to account for the fact that both the quality and the working parameters of the activated sludge chamber tend to affect the presence of filamentous microorganisms, which are known to be essential to the course of the activated sludge settling process. The presence of the filamentous species in the activated sludge is normal: filamentous colonies form a skeleton structure that promotes the attachment of bacterial colonies constituting the sludge mass. Moreover, the presence of filaments considerably reduces the susceptibility of flocs to fragmentation under vigorous mixing conditions and favors the formation of larger aggregates that settle more readily. On the other hand, particularly excessive microbial growth is detrimental to the flocs structure and their settleability. From the technological point of view, the value of the sludge index is determined by its settleability, defined as the sludge volume after settling in the Imhoff cone for 30 minutes. It is worth noting that the parameter is used in many STPs for the assessment of sludge settleability. This is due to the fact that, considering the various biochemical processes occurring in place in the activated sludge and its metastable nature, the sludge index that is predominantly used in practice (the intensive quantity) is not always an optimum source of data for simulating STP performance. Therefore, it seems correct to use an alternative approach, based on the extensive quantity, which means that creating models for simulating the process parameter is justified. The literature survey indicates that the issue of predicting sludge settleability has so far been of interest only to [5, 12], although the effect on the value of SE for filamentous microorganisms was not described in these reports.

For the reasons stated above, this study attempted to develop statistical regression models for predicting the settleability of activated sludge based on the quality of incoming sewage and on the identified dominant filamentous species. As part of the analyses conducted for the purpose, classification models are presented that enable identification of the respective filamentous microorganisms based on the working parameters of the bioreactor and the quality of the influent.

Test Facility

The tests were carried out in the municipal STP located in the commune of Sitkówka-Nowiny, which handles sewage from the separate sewage system of the city of Kielce, Sitkówka-Nowiny and part of the commune of Masłów. There are many factories located in the area, which results in a considerable variation of the pollution load in the influent since industrial wastewater contributes more than 20% of the total pollution load. Nominal capacity of the test facility is 72,000 m^3/d for a load of 275,000 PE. The influent is first handled mechanically (stepped screens, aerated sand trap, primary settlement tank), and then biologically in a reactor with separate de-nitrification and nitrification chambers (BARDENPHO system).

After being mechanically treated, sewage with the activated sludge is clarified in secondary tanks and discharged to the receiving water body: the Bobrza River. The parameters of the sewage are controlled by the municipal water supply company, Wodociągi Kieleckie Sp. z o.o., by measuring the quality of the influent and the working parameters of activated sludge. The sludge biocenosis is also determined as, for instance, its filamentous microorganisms are identified.

Methodology

The analyses aimed to develop a hybrid model simulating activated sludge settleability (SE). In this model the process parameter of interest is found by three

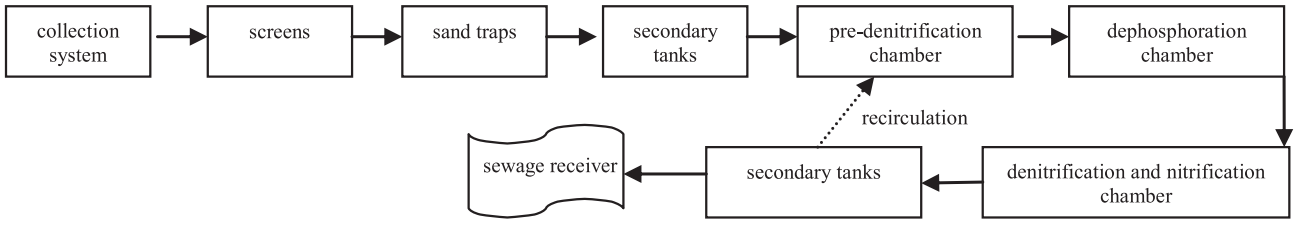

Fig. 1. A simplified block diagram of the STP in Sitkówka-Nowiny.

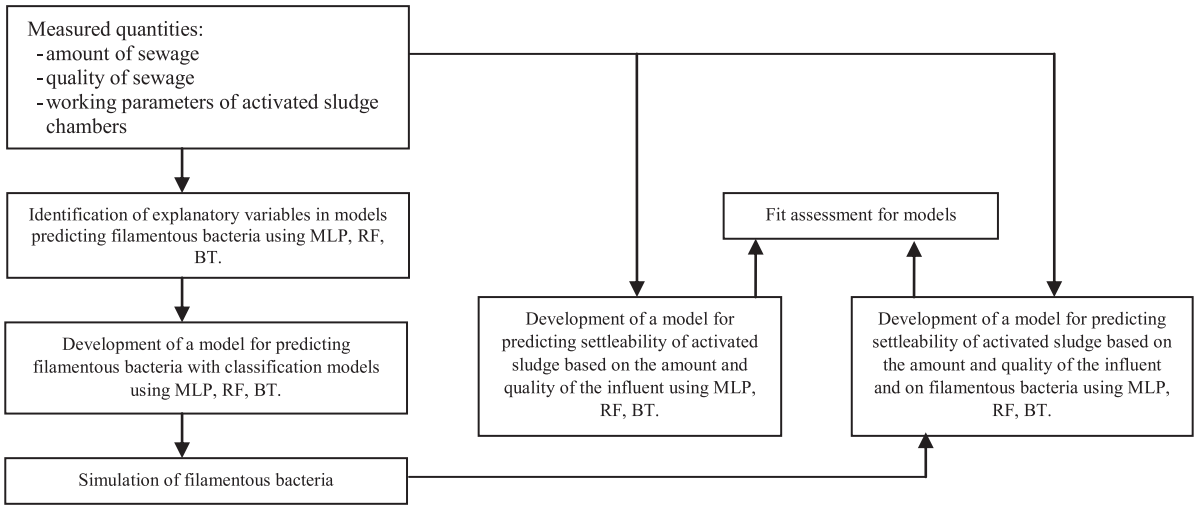

Fig. 2. Calculating activated sludge settleability.

data mining methods, based on input data describing the amount (Q), quality of the influent (BOD, COD, NH_4, TN, TSS) and the presence of various filamentous microbial species in activated sludge chambers. It was assumed that the filamentous bacteria taken into consideration in the mathematical model for simulating the value of SE would be predicted using classification models in which the exogenous variables would comprise the values of the parameters describing the sewage quality and the working parameters of the activated sludge chambers, such as: pH, sludge temperature (T_{sl}), oxygen concentration in the nitrification chamber (DO), substrate load (F/M), dosage of methanol, and PIX. To identify the variables explaining the presence of the respective filamentous microorganisms in the activated sludge, the values of importance (IMP) were found for the considered predictors by means of classification trees; the algorithm for the measurement of the value of IMP is described in detail in [13]. The calculation diagram for the present method for predicting the settleability of sludge is shown in Fig. 2.

The filamentous bacteria were identified by the method proposed by [14], in which the microorganisms are divided into dominant and subordinate species. Molecular methods offer an alternative solution: they enable quantitative and qualitative identification of microbial species but have a serious disadvantage of being costly, therefore their applicability in the daily operation of STPs is limited. A quantitative assessment of the prevalence of the respective filamentous species in the sludge by the conventional method is very complex, since there may be a multiple number of species in it in comparable quantities, especially if the STP handles industrial wastewater.

In this case, several filamentous bacterial species may be present and be regarded as dominant, which is commonly assumed in such considerations. In this study, the conventional method based on the observation of multiple activated sludge preparations enabled classification of the bacteria as either dominant or subordinate, and the former type was only used in the statistical model.

The input and output data were normalized by min-max transformation [15] before proceeding to the development of mathematical models. Moreover, to avoid an excessive fitting of simulation results to the measurement data, a 5-fold cross-validation was performed as follows: the dataset was divided into N subsets, the training dataset and the testing dataset were isolated and N results of models were averaged in order to obtain a single result. In this paper, methods based on artificial neural networks, random forests and boost trees were used for predicting the settleability of activated sludge and identification of microorganisms.

Artificial neural networks (ANN) are used for the simulation of linear and non-linear processes, optimization, classification and control [16, 17]. MLP (multi-layer perceptron) is a structure which has many applications in the simulation of processes in environmental engineering and other fields. In MLP, the input signals (x_i) are multiplied by weight inputs (w_{ij}) before being transmitted to the neurons of the hidden layer and there, in single neurons, summing takes place. The sums are subjected to linear or non-linear transformation using the activation function and then transferred to the output neurons. Since there is no hint at the choice of the ANN structure for modeling the value of SE, STATISTICA software and the automatic designer function were used in this paper. At the step of analyses, 200 different neural networks were generated for the input data and their parameters of fitting between the calculation results and the measurement data were established. In these considerations, it was assumed that, in the models of interest, the number of neurons in the hidden layer varied between 2 and 12 and the authors considered the linear, exponential, sinus, and hyperbolic tangent functions for the hidden neural layer, and the linear function in the output layer.

The random forest algorithm (RF), originally developed by [18] for classification purposes, has also been applied to regression problems. In the RF method, according to the classic bootstrap principle, random samples are first drawn with replacement from the testing dataset, based on which training datasets (n) are determined for trees. Next, n structures of tress are generated using a suitable algorithm, thus constructing a forest. Based on the obtained forest of n trees, prediction is established for every tree by finding the arithmetic mean of individual predictions as the result of the whole model. As part of the analyses (settleability and identification of filamentous microorganisms), the number of trees taken into account when generating the forest varied between 100 and 500.

Boosting trees (BT) are the implementation of the stochastic gradient boosting method, as used in classification- and regression-related problems [19]. The main idea of the method is to create a series of decision trees, such that each following tree will be used for determining the residues generated by the previous one.

Model Assessment Criteria

The predictive capacity of the models used for predicting activated sludge settleability was assessed using the following:

- mean absolute error (MAE):

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

- mean percentage error (MPE):

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_{i,obs} - y_{i,pred}}{y_{i,obs}} \right| \cdot 100\% \quad (2)$$

…where $y_{i,obs,cal}$ – suitably measured and calculated values of sewage quality indicators, n – number of components in dataset; in the case of interest, the dataset is composed of 160 values.

The predictive capacity of the obtained classification models for identifying various filamentous microorganisms was assessed using the following parameters:

- sensitivity:

$$SENS = 100 \cdot \frac{TP}{TP + FN} \quad (3)$$

- specificity:

$$SPEC = 100 \cdot \frac{TN}{FP + TN} \quad (4)$$

- prediction error:

$$R_z^2 = 100 \cdot \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

…where FP, FN, TP, TN – results of classification (Table 1).

Results and Discussion

Based on the amount, quality and operating parameters of the activated sludge chambers, variability range was determined for these parameters (Table 1). The variability range is very important in the aspect of applicability of the models being developed because it indicates the range of their usability. The data in Table 1 indicate that the values of the sewage quality indicators (BOD, COD, TN, TP, TSS, $N\text{-}NH_4$) varied in a broad range, affecting significantly the activated sludge in terms of concentration and settleability and, as a result, also its substrate load. The process data in this research paper were the basis of theoretical considerations of which the aim was to develop models for predicting such parameters of the activated sludge as concentration, substrate load, sludge index, and quality indicators [5, 6, 20].

Microbiological tests on the samples drawn from the activated sludge chamber in the test period identified five filamentous species: *Microthrix parvicella, Thiothrix* sp., *Nostocoida* sp., *Beggiatoa* sp. and *Sphaerotilus natans*. The species 1-3 were dominant either individually or in combinations, and 4-5 were the subordinate species (Table 2). Another observation was that not all five filamentous species were found in the activated sludge at a time. Observations of the sludge samples indicated the simultaneous presence of the species *Thiothrix* sp., *Sphaerotilus natans*, *Microthrix parvicella* and *Beggiatoa* sp., potentially suggesting low oxygen content (Table 2).

The microbiological tests of the activated sludge indicate that the bacteria *Microthrix parvicella* was present as the dominant species either alone or as one of two or three dominant species, in combination with the filamentous species *Thiothrix* sp. or *Nostocoida* sp. Moreover, the analyses showed that in the test period, *Microthrix parvicella* was typically detected in the activated sludge (either alone or as the dominant species

Table 1. Example of classification table.

Classification		Predicted decision	
		Positive	Negative
Observed decisions	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Table 2. Range of variability (mean, minimum, maximum) of the values of the qualitative indicators of the incoming sewage and the operating parameters of the activated sludge chamber.

Variable	Range	Mean
Q; m^3/d	32564÷86592	41584
T; ºC	10.6÷23.0	15.9
pH	7.2÷7.8	7.6
DO	0.55÷5.78	2.56
BOD; mg/dm^3	127÷557	320
COD; mg/dm^3	384÷1250	790
TSS, mg/dm^3	45÷168	91.4
TP; mg/dm^3	4.3÷12.6	7.8
TN; mg/dm^3	39.91÷124.09	77.6
$N\text{-}NH_4$, mg/dm^3	24.40÷65.90	48.86
MLSS; kg/m^3	1.19÷5.89	3.57
F/M, $gBZT_5/gMLSS \cdot d$	0.028÷0.172	0.07
methanol, m^3/d	0.0÷4.56	1.35
PIX, m^3/d	0.00÷1.93	0.80
SE, cm^3/dm^3	250÷980	785

in 30% of the test samples), in combination with *Thiotrix* sp. in 25% of the test samples and with *Nostocoida* sp. in 45% of the test samples.

In the next step, the classification trees method was used to develop statistical models for predicting the dominant filamentous bacterial species (*Microthrix parvicella, Nostocoida* sp., *Thiotrix* sp.). It enabled identification of the explanatory variables (quality of the sewage and operating parameters of the reactor), which are decisive for their presence; the results of these analyses are shown in Figs 3(a-c). The analysis indicate that the presence of *Microthrix parvicella*

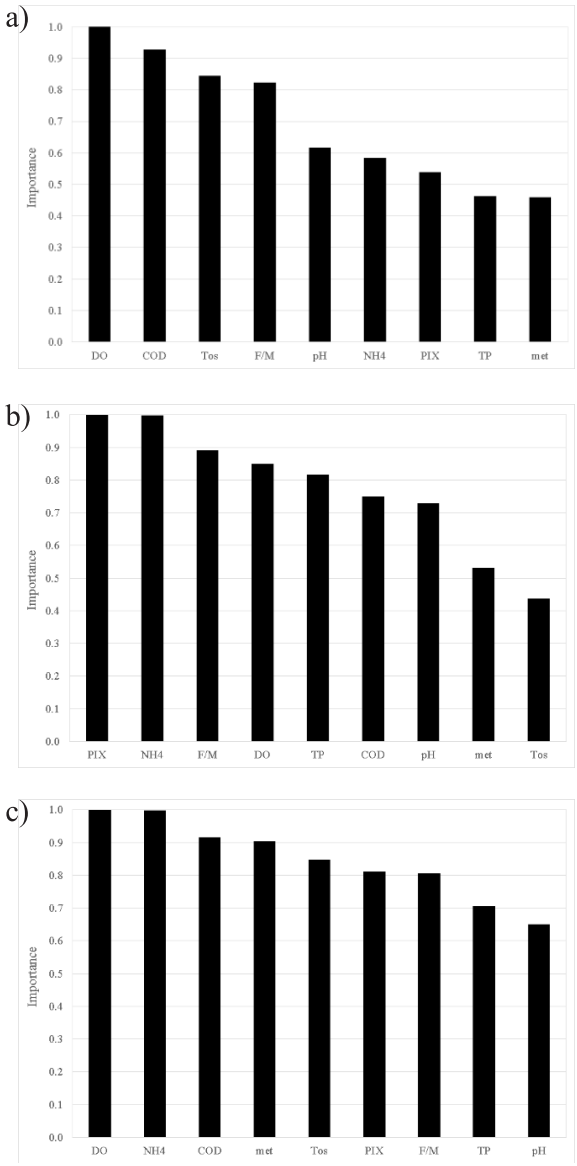

Fig. 3. Numerical values of the importance of explanatory variables for the presence of filamentous bacteria: a) *Microthrix parvicella*, b) *Thiotrix* sp., c) *Nostocoida* sp.

Table 3. Possible combinations of the filamentous bacterial species in the test period (+/-): presence/absence of the respective filamentous bacterial species in activated sludge.

Variant	*Microthrix parvicella*	*Thiothrix* sp.	*Sphaerotilus natans*	*Beggiatoa* sp.	*Nostocoida* sp.
1	+	-	-	-	+
2	+	+	+	-	+
3	+	+	-	-	-
4	+	-	+	-	-
5	+	+	+	-	-
6	+	+	-	-	+
7	+	-	-	-	+
8	+	+	+	+	-
where: (+/-) – presence/absence of the respective filamentous bacterial species in activated sludge.					

Table 4. Parameters of fitting (SPEC, SENS, R_z^2) of classification of the respective filamentous bacterial species by ANN, RF and BT.

Parameter	*Microthrix parvicella*			*Thiotrix* sp.			*Nostocoida* sp.		
	ANN	BT	RF	ANN	BT	RF	ANN	BT	RF
SPEC	0.91	0.90	0.88	0.81	0.87	0.90	0.89	0.90	0.89
SENS	0.62	0.86	0.86	0.69	0.84	0.51	0.67	0.87	0.77
R_z^2	0.86	0.89	0.87	0.77	0.86	0.67	0.76	0.88	0.80

in the activated sludge in the test facility was determined predominantly by oxygen content, as suggested by the maximum value of IMP = 1.0. A slightly lower effect was observed for COD (IMP = 0.91), temperature of activated sludge (IMP = 0.88), and substrate load (IMP = 0.86). The values of importance for pH, NH_4, TP and dosage of PIX and methanol are below 0.80.

Moreover, the calculations indicate that the dosage of PIX and the value for NH_4 have the decisive effect on the presence of *Thiotrix* sp. (IMP = 1.0), whereas a less pronounced effect was observed for the load (IMP = 0.89), oxygen content in the nitrification chamber (IMP = 0.85) and total phosphorus content (IMP = 0.82). In the other cases (COD, pH, T_{sl}, methanol) the values of importance were lower than IMP = 0.80. The analyses indicate that the decisive effect on the presence of *Nostocoida* sp. was observed for oxygen content (IMP = 1.0) and for the content of organic substances and nitrogen compounds in the influent, as confirmed by the calculated values of importance for NH_4 and COD, which were IMP = 0.99 and IMP = 0.92, respectively, as well as the dosage of methanol and PIX, sludge temperature and substrate load, for which the value of importance was IMP = 0.90; IMP = 0.81; IM = 0.85 and IMP = 0.80. These results of calculations for the respective filamentous species are confirmed by the analytical results, reported by [21, 22], who investigated sludge bulking in municipal STPs, caused by the presence of filamentous bacteria.

In view of the above calculation results, classification models were developed using the methods of ANN, RF and BT for identification of filamentous microorganisms, based on variables for which IMP≥0.80. Table 4 shows the values of parameters of fitting (SPEC, SENS, R_z^2) between the results of measurements and calculations for validation datasets (test datasets).

The data in Table 4 indicate that the methods of ANN, RF and BT can be applied in predicting the presence of filamentous microorganisms in activated sludge chambers, as confirmed by the calculated values of the fitting parameters SPEC, SENS and R_z^2. Among the three methods (ANN, BT, RF), the statistical model obtained using the boost tree (BT) showed the best predictive capacity for *Microthrix parvicella*. The presence of *Microthrix parvicella* was correctly identified in 90% of cases and its absence in the activated sludge in 86% of cases. The model based on artificial neural networks (ANN) showed a much worse predictive capacity in respect to the analyzed bacteria. The results of calculations indicate that ANN enabled correct classification in 62% of cases in which *Microthrix parvicella* was not identified in the activated sludge.

Considering the developed statistical models (Table 3) for the analysis of the presence of *Thiotrix* sp. in the activated sludge, the poorest predictions were provided by the random forest method. The best fitting between the results of calculation (classification of *Thiotrix* sp.) and measurements was obtained in the BT-based statistical model. In the mathematical models based on RF and BT, 90% and 87%, respectively, of occurrences were correctly identified and the presence of *Thiotrix* sp. was detected in the activated sludge. A much worse fitting between the results of calculations and measurements was obtained in the classification of cases, including occurrences when *Thiotrix* sp. was not detected in the activated sludge chamber: the models based on RF and BT correctly identified 51% and 84% of occurrences. Moreover, the data in Table 3 indicate that among the considered methods (ANN, BT, RF), the least error in the classification of *Nostocoida* sp. was obtained by the BT-based mathematical model. The model enabled correct classification in 90% of occurrences when the presence of *Nostocoida* sp. in the activated sludge chamber was detected and 87% occurrences in which the bacteria was not found in the activated sludge.

Taking into account the fact that the developed statistical models for the analysis of the occurrence of filamentous bacteria have a satisfactory predicting capacity, it seems justified to propose a hybrid model concept (based on classification and regression), establishing the impact of interactions between the biocenosis species found in the activated sludge on its settleability. Therefore, mathematical models were created to predict the settleability of the sludge: first, the pollution load (COD, BOD, NH_4, TN, TP, TSS) of the influent and then the bacterial flora, including the filamentous species, were taken into account. Table 4 shows the values of mean relative error (MRE) and mean percentage error (MAPE) of the results of measurements for the calculation of settleability after a 5-fold cross-validation for the validation dataset. Additionally, Fig. 5 shows a comparison between measured values of SE and those obtained using the statistical model, resulting in the lowest MAE and MAPE.

Table 5. Fitting parameters (MAE, MAPE) for the results of measurements of settleability by RF, BT, ANN for a model based on the influent sewage quality (I) and the presence of the filamentous species (II).

Model	RF		BT		ANN	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
I	101.9	15.88	102.37	16.89	86.04	14.06
II	85.73	14.03	43.6	6.39	17.82	2.362

The data in Table 5 indicate that the lowest errors were obtained for models provided by the ANN method and the highest values of the mean relative error and mean percentage error resulting from the RF method. The analyses showed that the lowest errors (MAE = 17.82 cm^3/dm^3 and MAPE = 2.36%) were obtained for the ANN model, in which the input data were the sewage quality and the identified filamentous microorganisms; much higher values of errors (MAE = 86.04 cm^3/dm^3 and MAPE = 14.06%) were obtained for the model covering the quality of the influent only. The values of errors in the SE prediction, resulting from the model in which the filamentous microorganisms were covered as well, are much lower than those obtained by [5], who developed a mathematical model using the MARS method based on the measurements of the amount and quality of the influent sewage and the bioreactor's operating parameters, and who obtained MAE = 9.9 cm^3/dm^3. The simulation results obtained in the study indicate that both the quality of sewage and filamentous microorganisms have a significant impact on the course of the activated sludge sedimentation process, as confirmed by the values of errors of results match to the measurements (Table 5) and studies carried out by Belanche et al. [23]. The relationships observed in this study (strong correlation between the settleability of activated sludge and the presence of individual groups of filamentous bacteria) are confirmed by the study of Bezak-Mazur et al. [24], who – based on the results of measurements of operational parameters of a bioreactor (concentration of activated sludge, dissolved oxygen concentration, temperature, pH) and the microbial composition of the sludge (bacteria *Microthrix parvicella, Thiothrix* sp., *Nostocoida* sp., *Beggiatoa* sp., *Sphaerotilus natans*) – developed a qualitative model by a logistic regression method to assess how individual microorganisms and interactions between them affect the process of activated sludge sedimentation. In the models presented in the paper, it is not possible to clearly assess the effect of the considered filamentous bacteria on the settleability of activated sludge based on the determined parameters of the model, which is a disadvantage of many data mining methods. In order for this relationship to be established, it is necessary to perform additional calculations for various boundary conditions, which is labor-intensive and time-consuming.

When analyzing the obtained results of analyses (Table 5) it may be concluded that the same range of variability of SE forecast errors in the model taking into account only the quality of sewage on the inflow to the treatment plant was obtained by Szeląg and Gawdzik [5], who – by using different data mining methods for independent variables including the measurements of

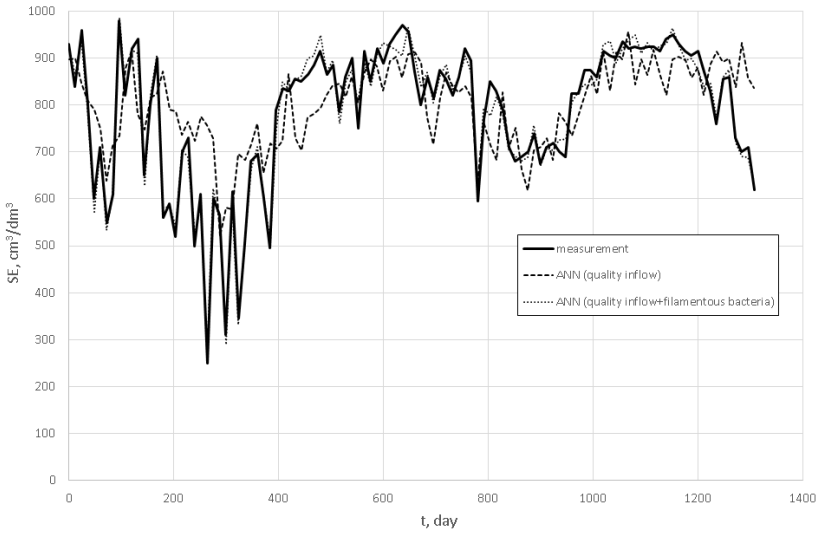

Fig. 4. Comparison of the results of measurements and calculations of activated sludge settleability as obtained by the artificial neural networks method.

BOD, SS, TP, TN – obtained the value of MAE error of the SE forecast equal to 82.1 cm^3/dm^3 when using the SVM method and 62.2 cm^3/dm^3 when using the MARS method. The SE forecast error values (Table 5) obtained in this study, based solely on the data covering the quality of sewage on the inflow to the STP, are larger than those obtained in the models given by Szeląg and Gawdzik [5], where independent variables, except for the quality of sewage, were the operational parameters of the biological reactor, such as sludge temperature, concentration of activated sludge, and substrate load of the sludge. This fact confirms the significant impact of operational parameters of the bioreactor on the course of the sedimentation process of the sludge, which is also confirmed by the analyses conducted by Bagherii et al. [7], Cortés et al. [10], and Szeląg et al. [5].

Fig. 4 presents comparison of the results of measurements and calculations of activated sludge settleability, obtained by the artificial neural networks method. The measured SE values ranged from 250 to 980 cm^3/dm^3.

It is worth noting that the inclusion of microbial species in the BT method results in lower MAE and MAPE, which decreased from 102.37 cm^3/dm^3 to 43.60 cm^3/dm^3 and from 16.89% to 6.39%, respectively. In the RF method, the values of the percentage error and relative error decreased from 101.90 cm^3/dm^3 to 85.73 cm^3/dm^3 and from 15.88% to 14.03%, respectively. Based on these analyses, it can be stated that the developed classification models for identifying filamentous bacteria and the obtained regression model for modeling activated sludge settleability have a satisfactory predictive capacity.

Conclusions

The study calculations demonstrated that the modeling methods based on artificial neural networks, random forests, and boost trees can be applied for the identification of filamentous microorganisms *Microthrix parvicella*, *Nostocoida* sp., and *Thiotrix* sp. in activated sludge chambers in the STP located in Sitkówka-Nowiny. The best predictive capacity, covering identification of the above-mentioned filamentous bacterial species in activated sludge chambers, was observed for statistical models obtained by the random forest method. In the case of *Microthrix parvicella, Nostocoida* sp. and *Thiotrix* sp., the poorest classification capacity was observed for the models based on the artificial neural networks and the random forests, respectively. Moreover, it was concluded that much lower values of errors of prediction of activated sludge settleability are obtained by statistical models with the inclusion of the filamentous microorganisms. It is worth noting that, within the methods of interest, the lowest errors of settleability prediction were obtained for the artificial neural networks and the highest for the model developed by the random forest method. The study demonstrated that the sludge settleability prediction model presented in this paper, in which the filamentous microorganisms are identified based on sewage quality indicators and bioreactor operating parameters using the classification model, can be applied in practice.

Conflict of Interest

The authors declare no conflict of interest.

References

1. FLORES-ALSINAA X., ARNELLA M., AMERLINCKD Y., COROMINAS L., GERNAEY K., GUOF L., LINDBLOMA E., NOPENS I., PORROD J., SHAWI A., SNIP L., VANROLLEGHEM P., JEPPSSON U. Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/ operational strategies in WWTPs. Science of the Total Environment, 466, **2014**.
2. COROMINASA L., LARSEN H., FLORES-ALSINAA X., VANROLLEGHEM P. Including Life Cycle Assessment for decision-making in controlling wastewater nutrient removal systems. Journal of Environmental Management, **128**, 759, **2013**.
3. KICZKO A., SZELĄG B., KOZIOŁ A., KRUKOWSKI M., KUBRAK E., KUBRAK J., ROMANOWICZ R. Optimal Capacity of a Stormwater Reservoir for Flood Peak Reduction.J. Hydrol. Eng., **23** (4), **2018**.
4. FLORES-ALSINAA X., RODRI´GUEZ-RODAA I., SINB G., GERNAEY K. Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty. Water Research, **42**, 4485, **2008**.
5. SZELĄG B., GAWDZIK J. Assessment of the Effect of Wastewater Quantity and Quality, and Sludge Parameters on Predictive Abilities of Non-Linear Models for Activated Sludge Settleability Predictions. Pol. J. Environ. Stud. Vol. **26** (1), 315, **2017**.
6. SZELĄG B., SIWICKI P. Application of the selected classification models to the analysis of the settling capacity of the activated sludge – case study. E3S Web of Conferences **17**, 00089. **2017**.
7. BAGHERII M., MIRBAGHERI S.A., BAGHERI Z., KAMARKHANI A.M. Modeling and optimization for a real wastewater treatment plant using hybrid artificial neural networks – genetic algorithm approach. Process Saf Environ. **95**, 12, **2015**.
8. HENZE, M., HARREMOES, P., COUR JANSEN, J. LA, ARVIN, E. Wastewater Treatment Biological and Chemical Processes, **2002**.
9. BARTOSZEWSKI K., BICZ W., DYMACZEWSKI BARTOSZEWSKI, K., BICZ, W., DYMACZEWSKI, Z., JAROSZYŃSKI, T., KUJAWA, K., LEMAŃSKI, J., ŁOMOTOWSKI, J., NALBERCZYŃSKI, A., NIEDZIELSKI, W., OLESZKIEWICZ, J., SAWICKI, M., SOZAŃSKI, M., URBANIAK, A., WASILEWSKI M. Guide to the operator of the sewage treatment plant, Polish Association of Sanitary Engineers and Technicians, Poznań, **2011**.
10. CORTÉS U., MARTINEZ J., COMAS M., SÁNCHEZ–MARRÉA M., RODRIDUEZ I.A conceptual model to facilitate knowledge sharing for bulking solving in

wastewater treatment plant. AI Communications, **16**, 279, **2003**.
11. ANDRZEJCZAK O., LIWARSKA–BIZUKOJĆ E. The effect of the pollutant load on the actived sludge flocks morphology. Gas Water and Sanitary Engineering, **12**, 480, **2014**.
12. SZELĄG B., GAWDZIK A., GAWDZIK A. Application of selected methods of black box for modelling the settleability process in wastewater treatment plant. ECOL CHEM ENG. **24** (1), 119, **2017**.
13. GATNAR E. A multi-model approach in issues of discrimination and regression. PWN Publisher, Warsaw, **2012**.
14. EIKELBOOM D.H., VAN BUIJSEN H.J.J. Manual of microscopic examination of activated sludge. Edition: "Seidel - Przywecki", Szczecin, **1999**.
15. RUTKOWSKI L. Metody i techniki sztucznej inteligencji. PWN, Warszawa **2006**.
16. LAI K., LIM S., THE P., YEAP K. An Artificial Neural Network Approach to Predicting Electrostatic Separation Performance for Food Waste Recovery. PJOES **26** (4), 1921, **2017**.
17. WĄSIK E., CHMIELOWSKI K., KACZOR G., CUPAK A. Stability Monitoring of the Nitrification Process: Multivariate Statistical Analysis. PJOES **27** (5), 1, **2018**.
18. BREIMAN L. Random forest. Journal Machine Learning. **45** (1), 5, **2000**.
19. FRIEDMAN J. Stochastic gradient boosting. Computational Statis-tics and Data Analysis, **38** (4), 367, **2002**.
20. SZELĄG B., STUDZIŃSKI J. A data mining approach to the prediction of food-to-mass ratio and mixed liquor suspended solids. Pol. J. Environ. Stud. **26** (5), 2231, **2017**.
21. COMAS J., DZEROSKI S., GIBERT K., RODA I., SÀNCHEZ–MARRÈ M. Knowledge discovery by means of inductive methods in wastewater treatment plant. AI Communication **14**, 45, **2001**.
22. DEEPNARAIN N., KUMARI S., RAMJITH J., MAHOMEDF., TANDOI V., PILLAY K., BUX F. A logistic model for the remediation of filamentous bulking in a biological nutrient removal wastewater treatment plant.Water Science and Technology. **72** (3), 391, **2015**.
23. BELANCHE L., VALDE J., COMAS J., RODA I., POCH M. Prediction of the bulking phenomenon in wastewater treatment plants. ArtificialIntelligence in Engineering **14**, 307, **2000**.
24. BEZAK-MAZUR E., STOIŃSKA R., SZELĄG B. Evaluation of the impact of operational parameters and particular filamentous bacteria on activated sludge volume index - a case study. Rocznik Ochrona Środowiska **18**, 480, **2016**.