Desalination and Water Treatment
www.deswater.com
doi: 10.5004/dwt.2020.25439

186 (2020) 134–143
May

Simulation of the influence of wastewater quality indicators and operating parameters of a bioreactor on the variability of nitrogen in outflow and bulking of sludge: data mining approach

Krzysztof Barbusiński[a,*], Bartosz Szeląg[b], Jan Studziński[c]

[a]*Silesian University of Technology, Konarskiego 18 Street, 44–100 Gliwice, Poland, Tel. +32-237-11-94; email: krzysztof.barbusinski@polsl.pl (K. Barbusiński)*
[b]*Kielce University of Technology, Tysiąclecia Państwa Polskiego 7 Av., 25-314 Kielce, Poland, Tel. +41-342-47-35; email: bszelag@tu.kielce.pl (B. Szeląg)*
[c]*Systems Research Institute PAN, Newelska 6 Street, 00-001 Warszawa, Poland, Tel. +22-381-02-75; email: Jan.Studzinski@ibspan.waw.pl (J. Studziński)*

Received 18 September 2019; Accepted 5 January 2020

ABSTRACT

A mathematical model to simulate and study the interaction between the content of nitrogen compounds in the outflow of a wastewater treatment plant and the efficiency of activated sludge sedimentation in its secondary clarifiers is presented in this paper. The goal of the model is to control the biological reactor settings (e.g. total nitrogen (TN) and sedimentation properties of the sludge) in case of discontinuity of quality indicators at wastewater inflow. Such an approach has not been applied so far. Continuity of calculated numerical values of model-dependent variables (TN, sludge volume index) was obtained by replacing the independent variables with the results of calculations obtained using the statistical models. Data mining methods were used to simulate the content of nitrogen compounds at the wastewater outlet. To simulate active sludge sedimentation a classification model based on the logistic regression method was used. According to the obtained results the proposed comprehensive model of simulation, referring to TN and activated sludge sedimentation, enables optimal selection of bioreactor settings.

Keywords: Wastewater treatment plant; Sludge volume index; Total nitrogen; Simulation and control; Neural networks

1. Introduction

Maintaining a high efficiency of a wastewater treatment plant (WWTP) is a complex task. The process requires constant adjustment of the set values in the bioreactor. To select optimal set values in the bioreactor and to minimize the operating costs, mathematical (physical and statistical) models describing the technological process of treatment are widely applied. Physical models activated sludge model (ASM) are based on systems of differential equations to describe the unit WWTP processes (e.g. primary settling tank, activated sludge chambers, secondary settling tank). In statistical models, so-called parametric models are used to simulate processes, where the determined parameters of the model have no physical interpretation and usually cannot be identified with the physicochemical parameters measured at the WWTP. The group of statistical models includes the method of artificial neural networks (ANN), regression trees, random forests, genetic programming, etc. [1–3].

The literature review [4–6] shows that physical models are usually used to stabilize a biological reactor (to control the wastewater quality indicators, that is, BOD, COD – biochemical and chemical oxygen demands, TN – total nitrogen, TP – total phosphorus, TSS – total suspended solids, at the outlet).

* Corresponding author.

Presented at the 14th Conference on Microcontaminants in Human Environment, 4–6 September 2019, Czestochowa, Poland

Statistical models, on the other hand, can be used both to stabilize a plant's operation and to control a plant by adjusting the reactor's settings to the required parameters. Currently, the efforts are limited to the control of selected indicators of wastewater quality at the outlet from the WWTP. Generally, there are no problems to obtain optimum values for biochemical oxygen demand, chemical oxygen demand, and TSS concentration. On the other hand, the nitrogen compounds are difficult to control [7–9]. The models developed so far mostly ignore the phenomenon of active sludge bulking. Proper sedimentation of sludge in secondary clarifiers is a key parameter to achieve the high effects of wastewater treatment and sludge dewatering. Moreover, in many cases, the authors of the developed models ignore the influence and disturbances of partial processes in technological objects [10,11]. The role of secondary clarifiers in the mentioned processes has been the subject of many papers and has been almost fully explained [12–14]. On the other hand, the study of the interaction between sludge bulking and nitrogen content at the WWTP outlet is more complex and depends on a greater number of parameters [15,16]. This problem has been discussed by Comas et al. [17], Flores-Alsina et al. [18], who analyzed the impact of activated sludge bulking on the results of a biological reactor optimization (Ludzack–Ettinger). However, due to the fact that this approach was based on the ASM model, its application to on-line control tasks was limited.

An important limitation of the created models is the continuity of the independent variables. It is a reason for many troubles. Therefore, the use and implementation of currently developed parametric models for simulation, optimization, and control of WWTP operation was limited. With the limited numbers of independent variables, it is not possible to select the optimal settings to reach a high efficiency of the plant [19,20]. Therefore, a new approach, as presented in this paper, seems to be a required solution.

The paper presents a statistical model for the simulation of TN concentration at the outlet from WWTP, with a basic assumption of discontinuities in the measurements of wastewater quality at the inflow. In this model, in comparison to the others, the interactions and influence of sedimentation of activated sludge in the secondary clarifier on the continuity of wastewater treatment processes in the bioreactor were also included. The proposed model could describe a simulation of TN removal from wastewater, and could also optimize a whole treatment process.

2. Object of study

The analyses were performed in the Sitkówka–Nowiny WWTP located close to the city of Kielce in central Poland. The nominal capacity of the facility is 75,000 m^3/d, which corresponds to a load of 275,000 equivalent residents. The plant is based on a mechanical-biological system, supported by chemical processes, if necessary. In the tested plant, the incoming wastewater is mechanically treated on grids, an aerated grit, and a primary clarifier. Then, it is transported to a biological block (to the activated sludge chamber) operating in a modified BARDENPHO system (with a pre-denitrification chamber). The treated wastewater flows into a secondary clarifier, it is discharged into the Bobrza River.

3. Methodology used

A mathematical model of a biological reactor (activated sludge chambers) with a secondary settling tank was applied using statistical methods (Fig. 1). In the proposed methodology, in comparison to the other papers [17,18], it is important to correctly determine a model for wastewater quality simulation (referring to TN) at the plant outlet and activated sludge sedimentation (classification model, e.g. based on logistic regression method to assess activated sludge bulking).

On the other hand, due to the large number of independent variables describing the analyzed processes and technical problems at the stage of variables measurement, the Fisher–Snedecor test was used to optimize and limit the number of variables taken into account in the modeling. To simulate the processes, the application of data mining methods was proposed in the paper. In this approach, several different methods were considered to optimally select them for the forecast of the studied phenomena. To obtain continuity of forecasts concerning sludge sedimentation (activated sludge bulking) and TN concentration in the bioreactor, it was proposed to calculate wastewater quality indicators at the inlet of the plant using mathematical modeling. In this modeling, the inflow concentration data were used. Therefore a continuous control and optimization of the wastewater treatment process, at the changing amount and quality of inflow wastewater, relating to weather conditions, was possible. The proposed model could be described as follows:

- selection of independent variables for models designed to simulate sludge volume index (SVI) (bulking of activated sludge) and TN,
- determination of models to forecast SVI and TN_{eff} values and to optimize a model,
- construction of statistical models to simulate wastewater quality indicators at wastewater inflow, to assume in models for SVI, TN_{eff} simulation; determination of these models is carried out with the use of data mining methods based on measured values of raw wastewater inflow,
- correction and selection of settings in the biological reactor.

The paper presents an example of determining the model for SVI and TN_{eff} simulation at the tested Sitkówka–Nowiny plant.

3.1. Model to forecast nitrogen outlet concentration

The literature review [3,7,21–23] shows that TN content in wastewater in the WWTP outflow is modeled (with the use of statistical methods) based on values of selected indicators of wastewater quality at the inflow and operating parameters of the reactor. Analyses carried out by many researchers and laboratory studies have shown that the key influence on the process of nitrogen removal in the bioreactor has, on the one hand, the number of organic compounds entering the treatment plant (BOD, COD), as well as the concentration of nitrogen forms (NO_2, NO_3, $N–NH_4$). The process of biological nitrogen removal in a biological reactor also depends on the air temperature, which affects the kinetics of biochemical processes in activated sludge [3–7]. During the operation of a

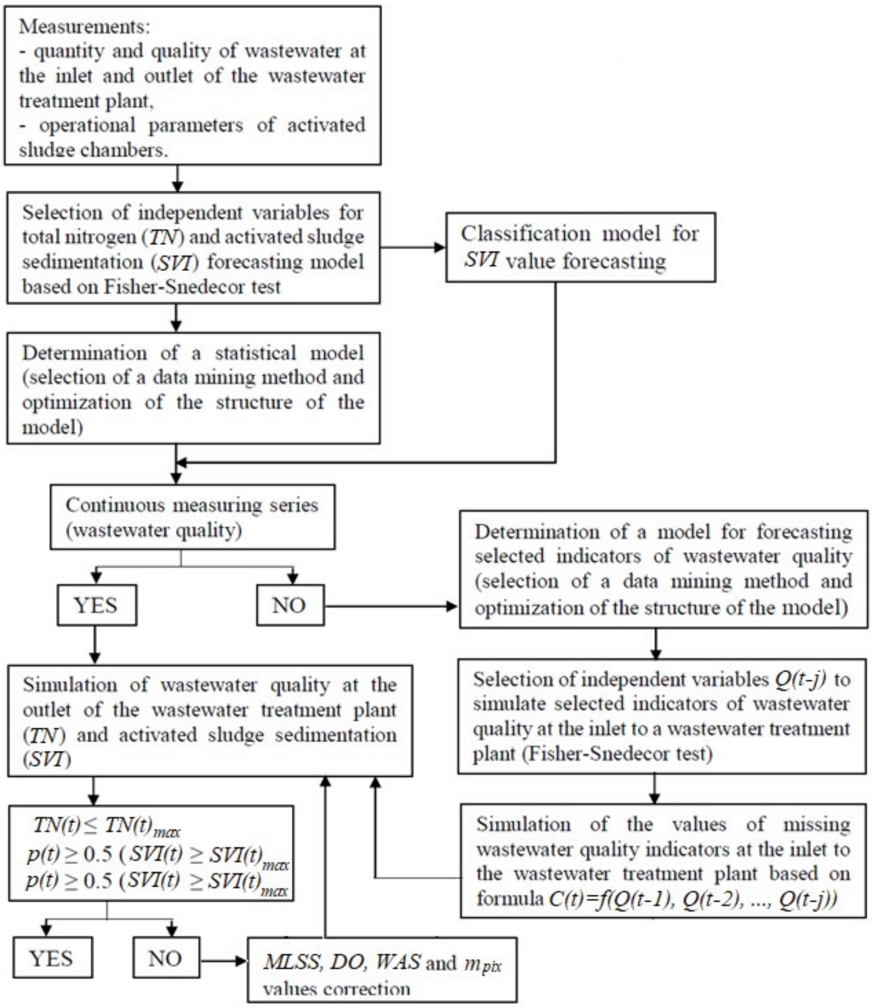

Fig. 1. Calculation scheme to determine a model for simulation, optimization, and control of WWTP operation (nitrogen removal, sedimentation of sludge in the secondary clarifier) in case of discontinuity of selected wastewater quality indicators.

WWTP, the nitrogen concentration in the treated wastewater, depending on the season and raw wastewater inflow, can be regulated by the amount of air entering the activated sludge, which translates into bacterial flora activity and biomass concentration [21–23]. The number of independent variables included in the models results from complex process physics and can be described by a general relation [3,7,22]:

$$\mathrm{TN}(t)_{\mathrm{eff}} = f\left(Q(t), C(t)_k, x_1, x_2, \ldots, x_m\right) \tag{1}$$

where $Q(t)$ – wastewater inflow to the treatment plant, $C(t)_k$ – k values of selected wastewater quality indicators at the inflow to the object (organic compounds and various forms of nitrogen), $x_{1,2,m}$ – numerical values of settings in the biological reactor (concentration of oxygen in nitrification chambers, the concentration of activated sludge, amount of excessive sediment, degree of recirculation).

Due to the different composition of bacterial flora in the activated sludge, the potential correlation of independent variables included in the model and interactions between them, not all the variables in Eq. (1) have the same effect on the quality of wastewater and some of them could be neglected. For this purpose, some practical methods [24–26] are used to reduce the number of variables in the model. Therefore, the Fisher–Snedecor test has been used, to reject independent variables having an overlooked influence on the studied phenomenon [15,16]. Thus, the independent variables included in the analyses in Eq. (1), for which the value of the test probability obtained from the calculations will be greater than $\alpha = 0.05$, could be removed from further analyses; their impact on the modeling results in comparison with other variables was considered negligible.

3.2. Simulation of activated sludge sedimentation in a secondary clarifier

The phenomenon of activated sludge sedimentation in a secondary clarifier is very complex and is affected by a number of factors including the quantity and quality of wastewater, hydraulic parameters of the clarifier, operating parameters of the biological reactor, meteorological conditions. These relations are local in nature and may vary depending on the quality of sewage, weather and climate conditions and the technological solution adopted [15,16]. The analyses carried out showed that the quality of wastewater has a significant impact on the sedimentation capacity of activated sludge, which translates into the amount of nutrients taken up by activated sludge flocs. In case of a deficiency of organic compounds, problems with sludge bulking may occur [27,28]. Undoubtedly, temperature plays an important role in sedimentation, which is closely related to kinetics and biochemical transformations taking place in the activated sludge. Oxygen concentration has a significant influence on the correction of the sedimentation capacity of the sludge because oxygen is necessary for the life processes of the microorganisms forming the sludge [29,30]. Therefore, no universal model of activated sludge sedimentation has been developed so far. Hence, the following relation [27–30] could be presented:

$$\mathrm{SVI} = f\left(T_{\mathrm{KOC}}, Q, C(t)_k, x_1, x_2, \ldots, x_m\right) \tag{2}$$

where T_{KOC}, Q, $C(t)_k$, x_1, x_2, x_m – as indicated above.

The examined process shows strong non-linearity, therefore the data mining methods of the black box type are used to its simulation. It could happen that the designated models do not provide satisfactory results of the simulation, then it may be necessary to change the type of output signals from the model from continuous to binary data. In the case of sludge sedimentation modeling such a change is possible, because the limit values of the SVI [30] enable us to achieve the sludge sedimentation capacity. In the systems with integrated removal of organic compounds, nitrogen, and phosphorus, the limit SVI value is equal to 150 cm^3/g and after exceeding it, activated sludge bulking occurs. For binary data simulation a model of logit regression can be used, in which the transformation of dependent variable space (SVI values) to probability space (in the range of 0–1) is performed. It can be shown that by substituting relation (4) to Eq. (3) it is also possible to simulate sludge sedimentation in case of discontinuity in measurements of wastewater quality indicators. According to Szeląg et al. [29], the sedimentation of sludge can be correctly predicted using a logit model. For the WWTP Sitkówka–Nowiny the model could be described as:

$$\begin{aligned}\ln\left(\frac{p}{1-p}\right) &= 0.02 \cdot \frac{\mathrm{BOD}_5}{\mathrm{TN}} + 0.32 \cdot \frac{\mathrm{BOD}_5}{\mathrm{TP}} + 0.0012 \cdot L_{\mathrm{N-NH_4}} \\ &\quad - 0.37 \cdot T_K + 14.38 - 1.36 \cdot X_{\mathrm{OC}} \\ &\quad - 1.76 \cdot m_{\mathrm{PIX}} - 1.18 \cdot \mathrm{DO}\end{aligned} \tag{3}$$

where: Dissolved oxygen (DO) – concentration of oxygen in the nitrification chamber (mg/L), X_{OC} – activated sludge concentration (kg/m^3), m_{PIX} – daily dose of chemical coagulant PIX (m^3/d), T_{KOC} – temperature in the activated sludge chambers (°C), $L_{\mathrm{N-NH_4}}$ – load of ammonium nitrogen (kg/d), TN – total nitrogen content (mg/L), TP – total phosphorus content (mg/L), BOD_5 – biochemical oxygen demand (mg/L).

The relation was obtained assuming that sludge bulking occurs when SVI = 150 cm^3/g, and $p = 0.50$, respectively. The limit value SVI = 150 cm^3/g = $\mathrm{SVI}_{\mathrm{lim}}$ is a typical value for a WWTP with integrated carbon, nitrogen and phosphorus removal technology [31]. The presented relations between active sludge sedimentation in the secondary clarifier and independent variables of the model are confirmed in several works on the real plants and using mathematical modeling of activated sludge bulking [15,17,18].

The presented above has already been a basis for the analysis of the reliability of WWTP operation in relation to activated sludge bulking [29]. However, that approach was simplified and there were limited possibilities of parallel control of the outlet wastewater quality and activated sludge bulking in the secondary clarifier. The presented model could be recognized as a significant development of that approach. Moreover, it could allow control and optimization of the continuous operation of the plant, taking into account both the quality of outlet wastewater (e.g. total nitrogen) and the bulking of activated sludge in the secondary clarifier.

3.3. Simulation of discontinuity of wastewater quality indicators in the model as a basis of SVI, TN forecasting

Besides the selected variables for process models (i.e. sludge sedimentation and TN removal), the laboratory analyses costs and their duration are also of significant importance. These models should guarantee high reliability of the plant operation even when the monitoring system is temporarily out of order. Therefore, the approach proposed by Szeląg et al. [32], Lubos et al. [33], Ahnert et al. [34] could be applied. The proposed model can be described as a general relationship Eq. (4):

$$C(t)_k = f\left(Q(t-1), Q(t-2), \ldots, Q(t-j)\right) \tag{4}$$

where: $Q(t{-}1, t{-}2, t{-}j)$ – values of inflow to the WWTP measured in the days preceding the model value of wastewater quality indicator k, determined based on the Fisher–Snedecor test. Thus, the values of activated sludge sedimentation (SVI) and TN concentration at the outlet from the treatment plant ($\mathrm{TN}_{\mathrm{eff}}$) can be determined by calculating the missing wastewater quality indicators using Eq. (4).

3.4. Data mining method for simulation and determination of the model structure

Apart from the selection of independent variables for the model, the proper selection of the type of statistical model is the key factor in the modeling process. The literature review [35–37] shows that artificial neural networks of the perceptron multilayer type (MLP) are commonly used to simulate the quality of wastewater at the inlet and outlet of WWTP. In case of unsatisfactory results of the simulation, one of the modifications of neural networks [Adaptive neuro fuzzy inference system (ANFIS), recurrent artificial neural network (RANN)], leading to the development of the model structure, maybe then applied. The cascade neural networks (CNN) method could be pointed out, where an additional connection between adjacent neural layers improves the predictive capabilities of the model in comparison to the MLP [38]. The disadvantage of the mentioned methods (MLP, ANFIS, CNN, RANN) is the lack of a global minimum of the defined target function [39] and the presence of many local minima.

This disadvantage is lacking in the method of support vectors machine (SVM), where the *N*-dimensional space of independent variables is transformed into *K*-dimensional linear space by means of Kernel function [40]. Apart from the methods mentioned above, the methods of regression trees are also used to simulate processes in technological systems, for example, the method of boosted trees (BT) [41]. The structure of the model obtained by this method is less complex than in the case of MLP, SVM, RANN, ANFIS methods and in many cases the obtained results are not worse (in some cases even better) than those determined by more complex methods.

The paper presents mathematical models for simulation of TN removal using the SVM, CNN and BT methods. By application of the CNN model, the number of neurons in the range (j–$2 \cdot j$+1) was selected to prevent the model from overturning [42]. The Broyden–Fletcher–Goldfarb–Shanno method [43] was used to determine the weight values for neurons. When optimizing the network structure for the assumed number of neurons, different activation functions (linear, exponential, hyperbolic tangent, sinus) were considered in the hidden and output layers to obtain the best possible matching of the calculation results to the measurements expressed by the correlation coefficient (R) and mean absolute (MAE) and mean relative (MAPE) errors. In the model determined by the SVM method, the Gauss function [39] was assumed. During the model determination, iterative steps were carried out (using the method of successive substitutions) on the ranges of values for C (capacitance), ε (insensitivity threshold), γ (Kernel function), until satisfactory MAE and MAPE values were obtained. In the BT method, the optimization of the model structure was based on minimizing the calculation errors for the number of trees in the model less than 200 [41].

4. Results

Based on the collected data at the Sitkówka–Nowiny WWTP in the period 2013–2017 the ranges of variability of selected technological parameters of operation of activated sludge chambers and the quantity and quality of wastewater at the outflow and inflow were determined (Table 1).

These changes have a significant impact on the values of settings in the bioreactor (concentration of oxygen in nitrification chambers, amount of dosed PIX, the concentration of activated sludge, amount of discharged excessive sludge), which in turn affects the quality of wastewater at the outlet from the WWTP (TN content) [44]. A wide range of changes in SVI values may indicate activated sludge bulking due to temperature decrease, as confirmed in the other experiments by Bayo et al. [30]. It was due to the reduced concentration of organic compounds in the inlet and the changing concentrations of oxygen in the activated sludge chambers, which is in accordance with the results shown, for example, by Comas et al. [17]. A wide range of changes of the SVI values at the Sitkówka–Nowiny treatment plant resulted also from the fact that in the analyzed period the composition of bacterial flora in the activated sludge, in which filamentous bacteria also occurred, changed. Based on the data in Table 1, it can be concluded that some operating problems could have occurred while the measuring period. This is confirmed by the fact that the acceptable TN_{eff} values in treated wastewater (TN_{eff} > 10 mg/L) were exceeded and the values of the SVI increased, to indicate the occurrence of active sludge bulking.

Therefore, it is necessary to develop mathematical models to simulate the above mentioned processes in order to improve the efficiency of the WWTP. Following the calculation scheme presented in Fig. 1, initial mathematical models for the simulation of wastewater quality indicators (BOD_5, TN, TP, N–NH_4) were determined. First of all, independent variables of particular quality indicators were determined and mathematical forecast models using SVM, CNN and BT methods were developed [26,29]. The calculations have shown that BOD_5 values can be predicted using the raw sewage inflow measurements $Q(t–1)$ to $Q(t–7)$. The same measurements can also be used to determine the content of TN and TP in the sewage inflow to the treatment plant. Probability values determined by Fisher–Snedecor test are

Table 1
Summary of numerical values of variables describing the quantity and quality of wastewater at the inflow and outflow from the WWTP and the operating parameters of the biological reactor

Indicators	Winter				Spring, summer, autumn			
	Min.	Mean	Max.	Standard deviation	Min.	Mean	Max.	Standard deviation
Q, m^3/d	29.952	39.364	88.986	6.563	3.0125	41.842	94.772	8.559
BOD_5, mgO_2/L	151	290	489	81.83	132	340	557	81.2
N–NH_4, mg/L	28	48.9	62	5.68	22	54.52	66.9	7.13
TN, mg/L	56.2	82.01	95.16	8.42	39.9	95.15	124.1	11.58
TN_{eff}, mg/L	3.6	7.02	17.8	2.38	6.26	8.89	13.92	1.43
TP_{eff}, mg/L	3.1	7.22	12.1	1.44	3.5	7.83	12.6	1.65
T_{KOC}, °C	10	11.9	13.5	0.8	11.3	17.8	23	3.1
DO, mg/L	1.8	2.85	3.25	0.8	1.51	2.2	3.25	0.65
X_{OC}, mg/L	2.85	4.95	6.54	0.84	2.15	4.11	5.28	0.95
WAS, kg MLSS/d	12.69	15.35	18.35	3.51	10.02	12.35	17.25	3.77
SVI, cm^3/g	154	198	291	35	90	138	200	37
m_{PIX}, m^3/d	0	0.81	1.75	0.27	0	0.84	1.82	0.28

given in Szeląg et al. [29]. Table 2 presents the calculated values of the measures of the adjustment of the calculation results to the measurements (MAE, MAPE, *R*). Fig. 2 also presents a comparison of data obtained from simulation calculations with measurements. The results for the best model for which the obtained error values were the least significant are presented.

The average values of the forecast errors of the selected quality indicators obtained by using the model determined by the SVM method are higher than the results obtained by the CNN method by no more than 15% (taking into account the values of MAPE). The highest values of errors were obtained for the BT method. The results of calculations and data in Table 2 are confirmed by curves plotted in Fig. 2 showing the variability of wastewater quality indicators (BOD_5, TN, N–NH_4, TP) measured and calculated using the CNN method.

The obtained results of the simulation are confirmed by the analyses of Ahnert et al. [34] and Szeląg et al. [25,29]. Thus, the quality of wastewater could be determined by the diversified degree of dilution of wastewater in the wastewater system.

The simulation results (see Table 2) confirm the possibility of introducing simplifications in the simulation of wastewater quality limiting the number of independent variables taken into account in the analyses [45,46], which could lead to lower costs of determining wastewater quality indicators, according to the laboratory analyses costs. Similar results of BOD_5 simulation were obtained by Dogan et al. [45], who included a number of other quality indicators in their calculations (COD, TSS, TP, TN), to confirm that the cost of performing measurements of these indicators would be much higher than these of flow. Slightly better results (about 5%) of simulation of the indicators (TN, N–NH_4, TP) were obtained by Minsoo et al. [47], who used modified the *k*-nearest neighbor's method in their calculations.

Assuming all the mentioned above conclusions (see Table 2), the next step in the calculation scheme (Fig. 1), that is, determination of the model for the forecast of TN, was proceeded with the assumption that discontinuities may occur in the measurement series concerning the wastewater quality at the inlet to the facility. Using the Fisher–Snedecor test, it was shown that at the level of statistical significance

Table 2
Values of measures of matching calculation results to measurements for selected wastewater quality indicators

Parameters	Inputs	CNN			SVM			BT		
	Number	MAE	MAPE	*R*	MAE	MAPE	*R*	MAE	MAPE	*R*
		mg/L	%	–	mg/L	%	–	mg/L	%	–
BOD_5	9	31	9.7	0.9	40	13	0.9	58	19	0.81
TN	9	3.8	4	0.9	4	4.7	0.9	11	12	0.7
N–NH_4	8	2.8	5.3	0.9	3	5.5	0.9	6.1	11	0.8
TP	8	0.6	7.2	0.9	0.6	8.1	0.9	1	14	0.8

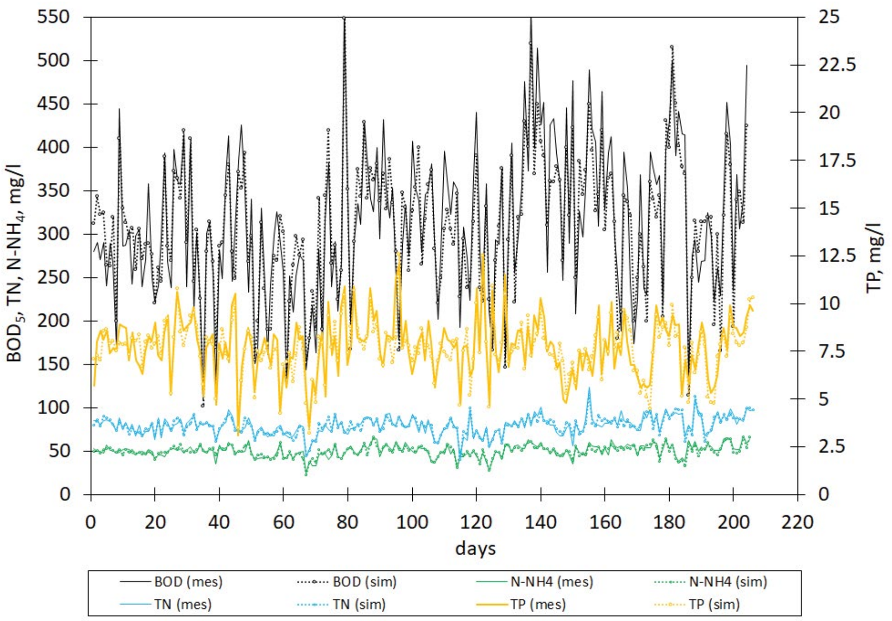

Fig. 2. Comparison of calculation results with measurements of selected wastewater quality indicators (BOD_5, TN, N–NH_4, TP) for the Sitkówka–Nowiny WWTP in the period covered by the measurements.

p = 0.05 the model for TN_{eff} forecast can be expressed as a general relation (5):

$$TN_{eff} = f(T_K, DO, WAS, RAS, X_{OC}, Q, BOD_{5(in)}, TN_{(in)}, N-NH_{4(in)}) \quad (5)$$

The correctness of Eq. (5) is confirmed by the studies of the other authors who used data mining methods to simulate TN values at the outlet from a WWTP. Analyses performed by Lee et al. [21], Mirbagheri et al. [48] confirm the influence of the values of wastewater quality indicators (BOD_5, TN, $N–NH_4$) and operational parameters of the bioreactor [mixed liquor suspended solid (MLSS), DO] on the TN content in the outlet. Relation (5) is also confirmed by the calculations of Corominas et al. [10] carried out using the ASM2d model to emphasize the physical nature of the phenomenon. Based on the general relation (5), the models for simulation of TN_{eff} values were determined, where the variables are continuous (option I) or there are gaps in the input data to the model (option II). In the option I, the statistical models for the sewage quality forecast are determined on the basis of the CNN, SVM and BT methods. In option II, the sewage quality forecast is based on general relations (2) and (5), where the values of quality indicators are modeled using the CNN and SVM methods. Thus, in option II, TN at the outlet is forecasted using hybrid models determined by combinations of the methods CNN + CNN, CNN + SVM, etc. The results of the study are presented in Table 3.

It can be concluded that the lowest error values for the TN_{eff} indicator simulation were found for the CNN method, while the highest error values were obtained using the model determined by the BT method (Table 3). The hybrid approach, that is, the combinations of CNN + CNN and CNN + SVM models, confirm the equivalent results of the TN_{eff} simulation. To get additional information, Figs. 3 and 4 present a comparison of simulation results and measurements in the period covered by the research. From the curves shown in Fig. 4 it can be concluded that the values of TN calculated with the CNN + SVM model can be significantly overestimated (even up to 4–7 mg/L), which can lead to erroneous information about exceeding the value of the calculated wastewater quality index. In practice, this may result in an overvaluation of the selected settings of the biological reactor. In case of a hybrid model of CNN + CNN type, a smaller differentiation of calculation results and TN value measurements was found, leading to optimal selection of biological reactor settings.

In order to evaluate the predictive abilities of the obtained models, they were compared with the simulation results obtained by the other authors (Table 4).

However, due to the limited range of results, different ranges of variability of TN_{eff} values at the outlet, different measures of matching the calculation results to the measurements presented in the works of other authors [3,7,21,27,48], in the present paper correlation coefficients were used to assess the models. This approach is correct and commonly used in comparative analysis of statistical models obtained for various objects, what was presented in numerous papers [2,49–51].

By comparing the results of TN_{eff} variable simulation described by Eqs. (2) and (5) obtained in the paper, it can be concluded that these results are not worse than the results obtained by the other authors who did not include the discontinuities of input data in their mathematical models. This is a significant advantage of the method given in the paper, since it gives the possibility to control the operation of a bioreactor when there are no sufficient data of wastewater quality indicators at the inlet to the WWTP. The obtained results confirm the fact that artificial neural networks are a valuable tool allowing to simulate with high accuracy the concentration of TN at the outlet from WWTP.

At the same time, comparing the R-values obtained by the other authors with the values obtained in this study, it can be stated that the modification of the neural network model of the MLP (by introduction of additional connections between subsequent layers) has a significant impact on the improvement of predictive capabilities of the model. This fact is confirmed by the analyses carried out by the other authors [52,53] dealing with the simulation of technological processes (Table 4).

Table 3
Comparison of the measures of adjustment of calculation results to the measurements data of TN_{eff} using the methods CNN, SVM, BT, and their combinations

Method	Validation		
	MAE	MAPE	R
	mg/L	%	–
CNN	0.59	7.58	0.95
SVM	0.71	9.52	0.91
BT	0.84	10.96	0.82
CNN + CNN	1.07	13.79	0.86
CNN + SVM	1.07	13.12	0.78

Table 4
Characteristics of exemplary models for TN_{eff} values prediction determined by other authors

Source	R	Independent variables	Method
Clara [7]	0.955	s_i, WAS, Q, E_p	AGF
Luo et al. [23]	0.938	TN, SE, DO, $N–NH_4$	ANN+FR
Lee et al. [21]	0.920	Q, T, pH, COD, TN, TP, X_{OC}, DO, SVI	ANN+PLS
Mirbagheri et al. [48]	0.882	Q, BOD_5, COD, TP, TDS, $N–NH_4$, HT, pH, X_{OC}	ANN
Hongbin et al. [3]	0.748	Q, BOD_5, COD, TSS, TN, TP	ANFIS+GA

where: GA – genetic algorithm, FR – fuzzy rough, PLS – partial least square, AGF – adaptative gradient with fuzzy variable selection, E_p – Energy Pump, SE – Settleability, HT – Hydraulic time, s_i – fraction of the COD.

In order to make practical use of the examined models, an example of their application to improve the efficiency of the Sitkówka–Nowiny facility is presented below. Three calculation options were considered:

- *Option I*: the current state of the plant operation,
- *Option II*: control of the reactor settings at continuous measurements of the inlet parameters
- *Option III*: in case of discontinuities in the measurement data.

The control is based on the assumption that the TN_{eff} concentration in the outlet does not exceed 10 mg/L. Due to the fact that the values of selected settings for each individual plant differ from each other [18,20,54], it is not possible to develop general guidelines for the selection of these settings. Therefore, the set-point values are iteratively selected: their initial values are assumed and the TN_{eff} content at the outlet and the active sludge sedimentation are calculated based on the relations (2), (5), and (6); then the calculated values are compared with the limit values of these variables (TN_{eff} = 10 mg/L; p = 0.50, which corresponds to SVI = 150 cm^3/g) and the preset values are adjusted so that the calculated values are equal to TN_{obi} < 10 mg/L and p < 0.5. The results of the calculations are shown in Figs. 5 and 6.

Based on the variability of the curves shown in Figs. 5 and 6, it can be concluded that the limit values for TN_{eff} during the winter were exceeded during the test period. Moreover, in the winter period, the activated sludge bulking was observed. The results of the calculations obtained in the paper are confirmed by the results of numerical experiments carried out by Comas et al. [17]. Based on the determined statistical model for modeling of activated sludge bulking using the fuzzy set theory and using continuous (annual) time series of wastewater quality indicators (BOD, TN, TP, N–NH_4) and reactor operating parameters [DO, MLSS, waste activated sludge

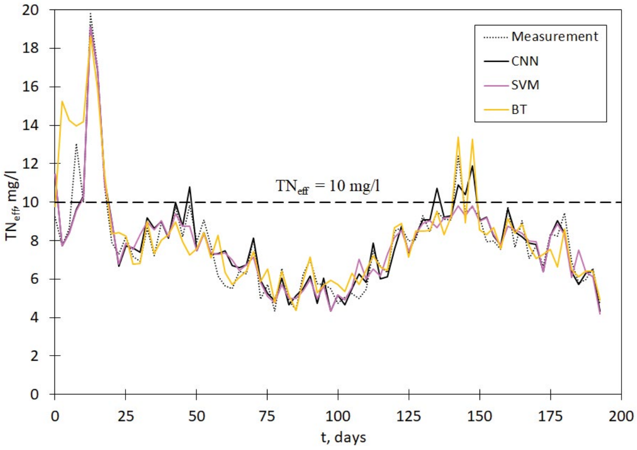

Fig. 3. Comparison of TN value measurements and calculation results obtained with CNN, SVM and BT models during the measurement period.

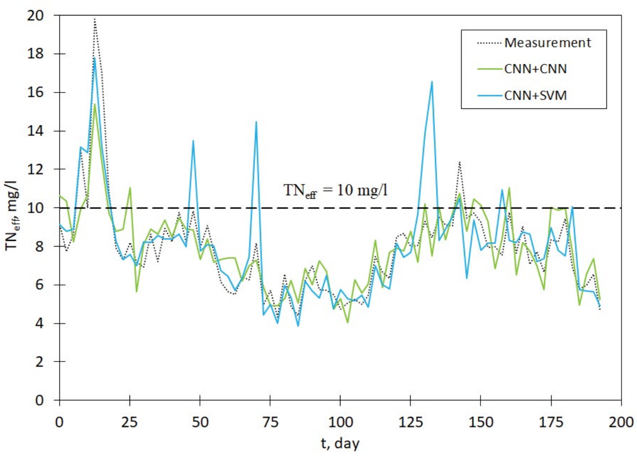

Fig. 4. Comparison of the TN value measurements and calculation results obtained with the CNN + CNN and CNN + SVM models during the measurement period.

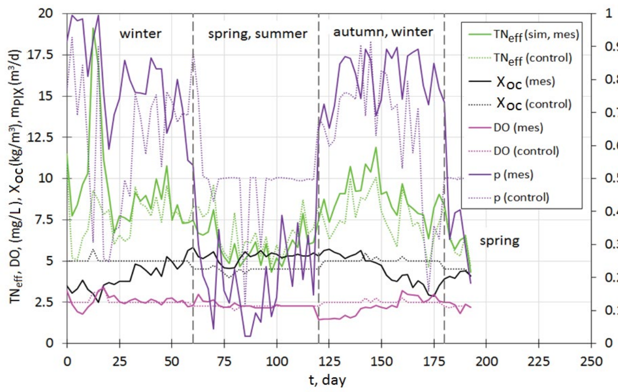

Fig. 5. Results of TN_{eff} and p calculations for the current state of the process and after its optimization and for the settings in the bioreactor with the assumption of continuity of the collected data.

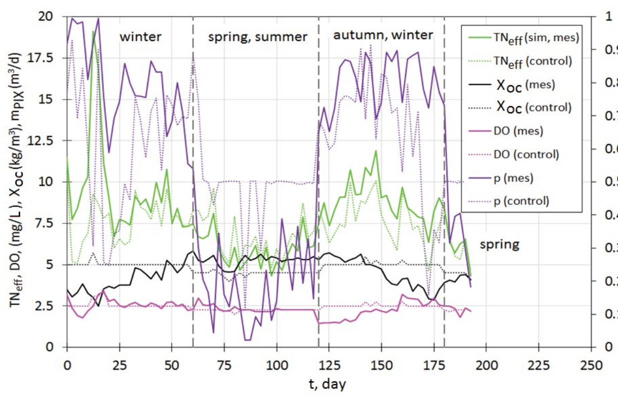

Fig. 6. Results of TN_{eff} and p calculations for the existing state of the process and after its optimization and the setting values in the bioreactor assuming discontinuity of the collected data.

(WAS), return activated sludge (RAS)], they simulated the reliability of WWTP operation. Their results showed problems with the sedimentation of activated sludge (bulking) in winter and autumn periods, which was confirmed by the measurements and calculation experiments carried out by Bayo et al. [30], Flores-Alsina et al. [18]. It is worth noting that the exceeding of TN content in treated wastewater was accompanied in the secondary clarifier by the phenomenon of bulking of activated sludge, which is confirmed by values of $p > 0.5$. [29], who for WWTP (Ludzack–Ettinger system) developed the ASM1 model to forecast BOD, COD, TSS, TN values and simultaneously performed simulations of sedimentation using Comas et al. [17]. In the spring and summer, it was found that TN_{eff} values are much lower than the maximum permissible values (TN_{eff} = 10 mg/L) and that in the episodes shown in Figs. 5 and 6 there were no problems with the activated sludge bulking.

Thus, using the mathematical models proposed in the paper to forecast the TN_{eff} value, it is possible to improve the efficiency of WWTP. Using the developed models (with consideration of two cases of measurement data in the form of continuous and discontinuous time series) it is possible to eliminate TN_{eff} exceedances, which has a key impact on the quality of wastewater discharge. It was achieved by increasing the concentration of sludge in activated sludge chambers and increasing the concentration of oxygen in nitrification chambers. It should be noted that for data collected in a discontinuous manner, the TN_{eff} values are lower than the results obtained when the data are recorded continuously (Figs. 5 and 6).

5. Conclusion

The analyses carried out in the study showed that it is possible to simulate the quality of wastewater at the outlet from WWTP using the SVM, CNN, and BT methods. The best results of the simulation were obtained with the CNN method, and the highest values of errors in the forecast of wastewater quality indicators were obtained with the use of the model determined with the BT method. Moreover, the obtained results confirm the possibility of modeling the TN content at the outlet from the WWTP using hybrid models. The concept of the hybrid model is based on the fact that the values of independent variables (the inlet parameters) are modeled with the use of statistical models based on wastewater flowing into the plant. This approach allows us to reduce the costs of continuous measurements of wastewater quality indicators at the inlet to the treatment plant and to simplify the calculations in WWTP plants supporting the operating routine.

It is possible to continuously simulate and control the operation of a WWTP in the absence of continuous measurements of raw wastewater quality indicators constituting independent variables in the process models. The model developed in the paper gives the possibility to analyze the influence of variable quantity and quality of incoming wastewater, weather conditions and operating parameters of a biological reactor on the quality of wastewater at the outlet from the WWTP and on the process of sedimentation of sludge in secondary clarifiers. The observation that it is possible to control the operation of a biological reactor in the absence of continuous measurements of key indicators of wastewater quality could also be recognized as apparent innovation in relation to the models developed by the other authors. It could help to increase the efficiency of a WWTP and to improve its reliability in comparison with the classical solutions, which assume continuous measurements of wastewater quality at the inlet to the WWTP.

References

[1] Y.S. Hong, R. Bhamidimarri, Evolutionary self-organising modelling of a municipal wastewater treatment plant, Water Res., 37 (2003) 1199–1212.

[2] A. Kusiak, A. Verma, X. Wei, A data-mining approach to predict influent quality, Environ. Monit. Assess., 185 (2013) 2197–2210.

[3] L. Hongbin, H. Mingzhi, Y. ChangKyoo, A fuzzy neural network-based soft sensor for modeling nutrient removal mechanism in a full-scale wastewater treatment system, Desal. Water Treat., 51 (2014) 6184–6193.

[4] B. Béraud, J.P. Steyer, C. Lemoine, E. Latrille, G. Manic, C. Printemps-Vacquier, Towards a global multi objective optimization of wastewater treatment plant based on modeling and genetic algorithms, Water Sci. Technol., 56 (2007) 109–116.

[5] J. Alex, L. Benedetti, J. Copp, K.V. Gernaey, U. Jeppsson, I. Nopens, M.N. Pons, L. Rieger, Ch. Rosen, J.P. Steyer, P.A. Vanrolleghem, S. Winkler, Benchmark Simulation Model No., 1 (BSM1). Technical Report, Department of Industrial Electrical Engineering and Automation, Lund University, LUTEDX/(TEIE7229)/1-62/(2008).

[6] J.F. Canete, P.D. Saz-Orozco, R. Baratti, M. Mulas, A. Ruano, A. Garcia-Cerezo, Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network, Expert Syst. Appl., 63 (2016) 8–19.

[7] N. Clara, Neural networks complemented with genetic algorithms and fuzzy systems for predicting nitrogenous effluent variables in wastewater treatment plant, WSEAS Trans. Syst., 6 (2008) 695–705.

[8] J. Drewnowski, M. Zmarzły, The Use of Mathematical Models for Diagnosis of Activated Sludge Systems in WWTP, E3S Web Conf., 22 (2017), https://doi.org/10.1051/e3sconf/20172200037.

[9] M. Ebrahimi, E.L. Gerber, T.D. Rockaway, Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis, J. Environ. Manage., 193 (2017) 234–246.

[10] L. Corominas, H.F. Larsen, X.F. Alsina, P.A. Vanrolleghem, Including life cycle assessment for decision-making in controlling wastewater nutrient removal systems, J. Environ. Manage., 128 (2013) 759–767.

[11] B. Szeląg, J. Studziński, A data mining approach to the prediction of food-to-mass ratio and mixed liquor suspended solids, Pol. J. Environ. Stud., 26 (2017) 2231–2238.

[12] A. Luciano, P. Viotti, G. Mancini, V. Torretta, An integrated wastewater treatment system using a BAS reactor with biomass attached to tubular supports, J. Environ. Manage., 113 (2012) 51–60.

[13] S. Conserva, F. Tatti, V. Torretta, N. Ferronato, P. Viotti, An integrated approach to the biological reactor–sedimentation tank system, Resources, 94 (2019) 1–19.

[14] M. Al-Sammarraee, A. Chan, S.M. Salim, U.S. Mahabaleswar, Large-eddy simulations of particle sedimentation in a longitudinal sedimentation basin of a water treatment plant. Part I: particle settling performance, J. Chem. Eng., 152 (2009) 315–32.

[15] U. Cortés, M. Martínez, J. Comas, M. Sànchez-Marrè, I. Rodríguez-Roda, A conceptual model to facilitate knowledge sharing for bulking solving in wastewater treatment plant, AI Commun., 16 (2006) 279–289.

[16] A.M.P. Martins, J.J. Heijnen, M.C.M. van Loosdrecht, Bulking sludge in biological nutrient removal systems, Biotechnol. Bioeng., 86 (2004) 125–135.

[17] J. Comas, I.R. Roda, K.V. Gernaey, C. Rosen, U. Jeppsson, M. Poch, Risk assessment modelling of microbiology-related solids separation problems in activated sludge systems, Environ. Modell. Software, 23 (2008) 1250–1261.
[18] X. Flores-Alsina, J. Comas, I.R. Roda, M. Poch, K.V. Gernaey, U. Jeppsson, Evaluation of plant-wide WWTP control strategies including the effects of filamentous bulking sludge, Water Sci. Technol., 60 (2009) 2093–2103.
[19] A.C. Avella, T. Görner, J. Yvon, P. Chappe, P. Guinot-Thomas, P. Donato, A combined approach for a better understanding of wastewater treatment plants operation: statistical analysis of monitoring database and sludge physico-chemical characterization, Water Res., 45 (2011) 981–992.
[20] A. Asadi, A. Verma, K. Yang, Wastewater treatment aeration process optimization: a data mining approach, J. Environ. Manage., 203 (2016) 1–10.
[21] D.S. Lee, M.W. Lee, S.H. Woo, Y.J. Kim, J.M. Park, Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant, Process Biochem., 41 (2006) 2050–2057.
[22] H.W. Lee, M.W. Lee, J.M. Park, Multi-scale extension of PLS algorithm for advanced on-line process monitoring, Chemom. Intell. Lab. Syst., 98 (2009) 201–212.
[23] F. Luo, R. Yu, Y. Xu, Y. Li, Effluent Quality Prediction of Wastewater Treatment Plant Based on Fuzzy-Rough Sets and Artificial Neural Networks, Sixth International Conference on Fuzzy Systems and Knowledge Discovery, IEEE, Tianjin, China, 2009, pp. 47–51.
[24] A. Kusiak, Z. Zhang, Short-horizon Prediction of Wind Power: A Data-driven Approach, IEEE Trans. Energy Convers., 25 (2010) 1112–1122.
[25] B. Szeląg, K. Barbusiński, J. Studziński, Activated sludge process modelling using selected machine learning techniques, Desal Water Treat., 117 (2018) 78–87.
[26] L. Breiman, Random Forests, J. Mach. Learn., 45 (2000) 5–32.
[27] E. Kowalska, E. Paturej, M. Zielińska, Use of *Lecane inermis* for control of sludge bulking caused by the *Haliscomenobacter* genus, Desal Water Treat., 57 (2016) 10916–10923.
[28] I. Lou, Y. Zhao, Sludge bulking prediction using principle component regression and artificial neural network, Math. Probl. Eng., 2012 (2012) 1–17.
[29] B. Szeląg, K. Barbusiński, J. Studziński, Application of the model of sludge volume index forecasting to assess reliability and improvement of wastewater treatment plant operating conditions, Desal Water Treat., 140 (2019) 132–143.
[30] J. Bayo, J.M. Angosto, J. Serrano-Aniorte, Evaluation of physicochemical parameters influencing bulking episodes in a municipal wastewater treatment plant, Water Pollution VIII: Modell. Monit. Manage., 95 (2006) 531–542.
[31] M. Henze, P. Harremoes, E. Arvin, J. Lacour, Wastewater Treatment, Biological and Chemical Processes, Springer-Verlag, Berlin, 2002.
[32] B. Szeląg, L. Bartkiewicz, J. Studziński, Black-box forecasting of selected indicator values for influent wastewater quality in municipal treatment plant, Environ. Prot., 38 (2016) 39–46 (in Polish).
[33] J. Lubos, T. Kaletova, M. Sedmakova, P. Balazova, A. Cervenanska, Comparison of service characteristics of two town's WWTP, J. Ecol. Eng., 18 (2017) 61–67.
[34] M. Ahnert, C. Marx, P. Krebs, V. Kuehn, A black-box model for generation of site-specific WWTP influent quality data based on plant routine data, Water Sci. Technol., 74 (2016) 2978–2986.
[35] L. Belanche, J. Valdes, J. Comas, I. Rodriguez – Roda, M. Poch, Prediction of the bulking phenomenon in wastewater treatment plants, Artif. Intell. Eng., 14 (2000) 307–317.
[36] S. Venkadesh, G. Hoogenboom, W. Potter, R. McClendon, A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks, Appl. Soft Comput., 13 (2013) 2253–2260.
[37] P. Kundu, A. Debsarkar, S. Mukherjee, S. Kumar, Artificial neural network modelling in biological removal of organic carbon and nitrogen for the treatment of slaughterhouse wastewater in a batch reactor, Environ. Technol., 35 (2014) 1296–1306.
[38] G. Capizzi, G.L. Sciutto, P. Monforte, C. Napoli, Cascade feed forward neural network based model for air pollutants evaluation of single monitoring stations in urban areas, Int. J. Electron. Telecommun., 61 (2015) 327–332.
[39] L. Rutkowski, Artificial Intelligence Methods and Techniques: Computational Intelligence, PWN, Warsaw, 2006 (in Polish).
[40] C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, U. Fayyad, Knowledge Discovery and Data Mining, Kluwer, 1998, pp. 1–43.
[41] J.H. Friedman, Stochastic gradient boosted, Comput. Stat. Data Anal., 38 (2002) 367–378.
[42] S.G. Setti, R.N. Rao, Artificial neural network approach for prediction of stress–strain curve of near b titanium alloy, Rare Met., 33 (2014) 249–257.
[43] V. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.
[44] K. Barbusiński, H. Kościelniak, Influence of substrate loading intensity on floc size in activated sludge process, Water Res., 29 (1995) 1703–1710.
[45] E. Dogan, A. Ates, E.C. Yilmaz, B. Eren, Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand, Environ. Prog., 27 (2008) 439–446.
[46] H.Z. Abyaneh, Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters, J. Environ. Health Sci., 12 (2014) 1–8.
[47] K. Minsoo, K. Yejin, K. Hyosoo, P. Wenhua, K. Changwon, Evaluation of the *k*-nearest neighbour method for forecasting the influent characteristics of wastewater treatment plant, Front. Environ. Sci. Eng., 10 (2016) 299–310.
[48] S.A. Mirbagheri, M. Bagheri, S. Boudaghpour, M. Ehteshami, Z. Bagheri, Performance evaluation and modeling of a submerged membrane bioreactor treating combined municipal and industrial wastewater using radial basis function artificial neural networks, J. Environ. Health Sci., 13 (2015) 13–17.
[49] M. Häck, M. Köhne, Estimation of wastewater process parameters using neural networks, Water Sci. Technol., 33 (1996) 101–115.
[50] A. Kusiak, X. Wei, Prediction of methane production in wastewater treatment facility: a data-mining approach, Ann. Oper. Res., 216 (2014) 71–81.
[51] A. Verma, X. Wei, A. Kusiak, Predicting the total suspended solids in wastewater: a data-mining approach, Eng. Appl. Artif. Intell., 26 (2012) 1366–1372.
[52] L.I.L. Fanjun, Q. Junfei, Z. Wei, A Fast Growing Cascade Neural Network for BOD Estimation, Proceedings of the 34th Chinese Control Conference (CCC), Hangzhou, 2015, pp. 3417–3422.
[53] F. Li, J. Qiao, H. Han, C. Yang, A self – organizing cascade neural network with random weights for nonlinear system modeling, Appl. Soft Comput., 42 (2016) 184–193.
[54] A. Kusiak, X. Wei, Optimization of the activated sludge process, J. Energy Eng., 139 (2013) 12–17.