

*Original Research*

# Uncertainty of Forecast and Control of Activated Sludge Sedimentation Capacity

**Bartosz Szela<sup>1\*</sup>, Jan Studziński<sup>2</sup>, Krzysztof Chmielowski<sup>3</sup>, Ewa Dacewicz<sup>3</sup>,  
Piotr Siwicki<sup>4</sup>, Dorota Mirosław-Świątek<sup>4</sup>, Justyna Lisowska<sup>1</sup>**

<sup>1</sup>Faculty of Environmental, Geomatic and Energy Engineering, Kielce University of Technology, Kielce, Poland

<sup>2</sup>Systems Research Institute PAN, Warsaw, Poland

<sup>3</sup>University of Agriculture in Cracow, Department of Sanitary Engineering and Water Management, Krakow, Poland

<sup>4</sup>Warsaw University of Life Sciences, University Faculty of Civil and Environmental Engineering, Warsaw, Poland

*Received: 11 January 2019*

*Accepted: 24 April 2019*

## Abstract

Our paper presents an analysis of the effect of data selection uncertainty for the teaching and testing sets in the black box model (multilayer perceptron type of artificial neural networks) using the bootstrap method on the accuracy of forecast and control of activated sludge sedimentation (SE) and the sludge volume index (SVI). The calculations show that sludge sedimentation, and hence also the sludge volume index, can be predicted based on the wastewater quality indicators and biological reactor operating parameters. The presented analyses also confirmed the significant influence of the neural network model structure on the uncertainty of estimating biological reactor operating parameters (mixed liquor suspended solids, concentration of oxygen) which, in practical considerations, leads to problems of continuous control of the sludge sedimentation capacity.

**Keywords:** activated sludge sedimentation, sludge volume index, uncertainty of forecast, neural network, mathematical model

## Introduction

Due to the complexity of processes occurring in wastewater treatment plants, their operation is a very complicated task. In order to achieve satisfying operational effectiveness, the maintenance of operating conditions of the treatment process must be assured on the assumed level [1]. This requires control of operating parameters of the plant bioreactor. One of

the crucial parameters influencing the operation of a wastewater treatment plant (WWTP) is the sludge volume index (SVI). It is calculated as the quotient between the sludge sedimentation capacity (SE) in an Imhoff funnel over the time of 30 minutes and the sludge concentration (mixed liquor suspended solids – MLSS). Since the sludge volume index is not always the sole basis for analyzing settling properties of sludge, the sedimentation capacity also seems helpful. Obtaining an appropriate range of SVI is crucial for the operation of the WWTP. Deterioration of the sludge settling parameters negatively affects the sewage clarifying process, resulting, in consequence, in decreased quality of wastewater on the outflow from the clarifiers,

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

which causes problems with the sludge dewatering [1]. To avoid this, a number of mathematical models of regression and classification using artificial neural networks (ANN), fuzzy logic (FL), and support vector machines (SVM) have been developed [2-5]. Results of modeling obtained with the aforementioned methods are characterized by uncertainty. This is due to the parameter estimation errors in the model structure or input data measurement errors. The problem of doubtful results by the WWTP modeling has been discussed in papers concerning the calibration of ASM models and their use for control of wastewater treatment processes; however, the settling capacity of the activated sludge was not included. In the case of WWTP, the SVI forecast uncertainty can have an essential impact on the exactness of prediction of bioreactor technological parameters (mixed liquor suspended solids, recirculation grade, oxygen concentration in nitrification chambers, pH, amount of coagulant dosed, amount of external carbon supplied). At the operating stage of the WWTP, appropriate decisions made by the technologist affect the maintenance of optimum operating parameters of the biological reactor. This aspect is crucial for obtaining high efficiency of the treatment plant operation and reducing the amount of pollutant load introduced into the environment.

Having regard to the above, this paper presents the model for forecasting settling capacity of activated sludge (sedimentation capacity and sludge volume index) developed using the artificial neural networks. The effect of ANN structure uncertainty on the predictive ability

of the models to forecast the sedimentation capacity and sludge volume index was analyzed in detail.

## Research Object

Our research object was the municipal wastewater treatment plant of 72,000 m<sup>3</sup>/d nominal capacity situated in the Sitkówka-Nowiny municipality, which receives wastewater from the sewerage system of Kielce and Sitkówka-Nowiny municipality and partially from Masłów municipality. Wastewater is preliminarily mechanically purified (grit chambers, aerated sand traps with separated fat removal device, preliminary clarifiers), and then it is directed into the biological part. Biological reactors consist of separated dephosphatation, denitrification and nitrification chambers in which the organic pollutants and nitrogen and phosphorus compounds are removed. After the biological treatment, wastewater flows into four secondary clarifiers, from which – after separation from activated sludge – it is drained into the Bobrza River.

## Methods

### Independent Variables Describing Sedimentation of Activated Sludge

In the studied treatment plant, the measurements of the quantity and quality of wastewater (biological

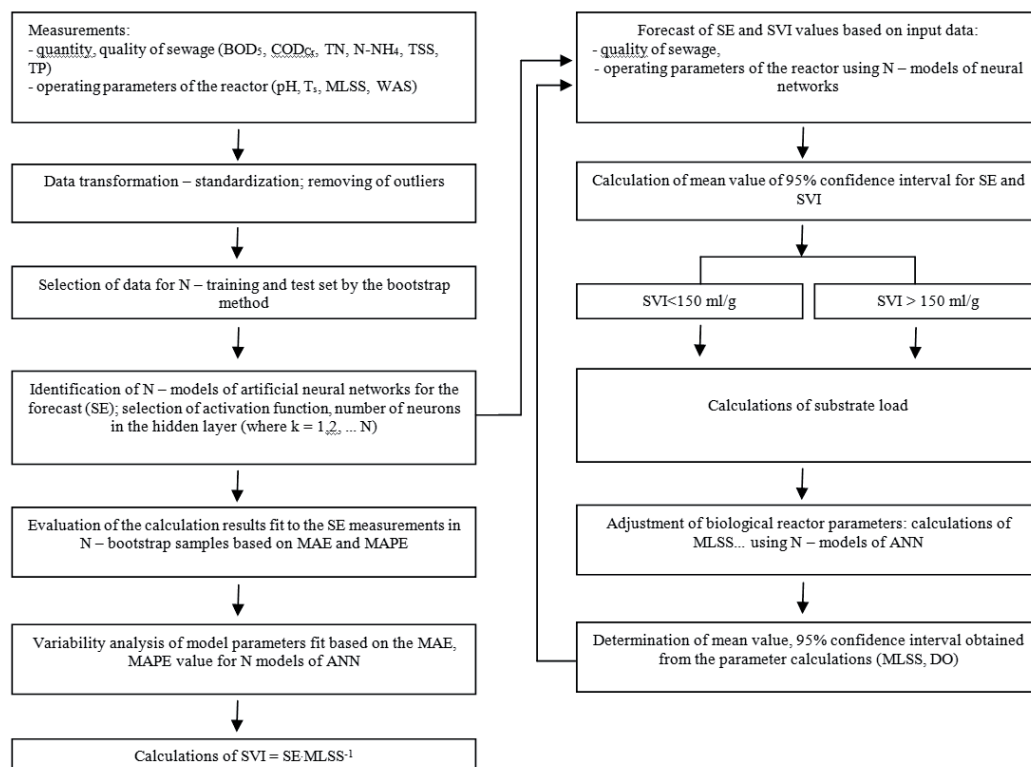


Fig. 1. Calculation scheme for SE, SVI by the artificial neural networks taking into consideration the uncertainty of the model structure.

oxygen demand BOD<sub>5</sub>, chemical oxygen demand COD, total nitrogen TN, ammonium nitrogen N-NH<sub>4</sub><sup>+</sup>, total phosphorus TP, total suspended solids TSS), as well as the bioreactor operating parameters (temperature – T<sub>st</sub>, pH, mixed liquor suspended solids – MLSS, oxygen concentration in nitrification chambers – DO, sludge volume index – SVI, sedimentation capacity – SE) are conducted under continuous monitoring.

BOD<sub>5</sub> was determined according to PN-EN 1899-1: 2002 and PN-EN 1899-2: 2002, whereas COD was according to PN-ISO 6060: 2006. The concentration of total nitrogen was carried out according to PN-73/C-04576.14, whereas ammonium nitrate used PN-ISO 7150-1: 2002. Total phosphorus was determined according to the PN-EN ISO 6878: 2006 standard. The content of total suspended solids was determined in accordance with PN-EN 872: 2007. Measurements of dissolved oxygen concentration were made in accordance with PN-EN 5814: 2013-04.

As part of the conducted analyses, the mathematical models for forecasting the sludge sedimentation capacity (SE) were developed, where explanatory variables comprised the above-mentioned indicators of the quantity and quality of wastewater, as well as the operating parameters of the reactor [3, 6] measured in the period 2015÷2017. At the same time, the sludge volume index was determined by using the relationship between the sedimentation capacity (SE), which is measured by the assessment of the sediment volume in Imhoff funnel after 30 minutes of sedimentation:

$$SVI = SE \cdot MLSS^{-1} \quad (1)$$

The sludge sedimentation capacity is an important technological parameter, which in practical considerations also constitutes the basis for optimizing the operation of technological objects [7-9].

The concept of a model for simulation, sludge sedimentation capacity and volume sludge index within the conducted calculations, a multilayer perceptron (MLP) class of artificial neural networks method was used in order to forecast the sludge sedimentation capacity (SE), whereas the evaluation of the effect of the model structure uncertainty on the simulation results was based on the bootstrap method for the selection of data for the training and test set [10] methodology used for predicting and controlling sedimentation of activated sludge (i.e., both sedimentation capacity and sludge volume index), taking into account the uncertainty of the model structure as presented in Fig. 1. A STATISTICA program was used to build a model of a neural network allowing for the simulation of sludge sedimentation capacity. Details related to model identification are described below.

In the course of the analyses conducted, based on the computational diagram in Fig. 1, first of all, a selection of data for learning and testing sets was made using the bootstrap method. Using the obtained learning and test data sets, models for simulation of activated sludge

sedimentation (SE) were developed which formed the basis for determining the activated sludge volume index (SVI) and the confidence interval. In addition, based on the results of measurements of the quantity and quality of the wastewater inflow as well as of operational parameters of the bioreactor, the determination of substrate load, which is the basis for the evaluation of the course of organic compounds, nitrogen and phosphorus removal processes, was envisaged. If the calculated SVI values exceeded the limit value (150 ml/g) and/or the substrate load value was too high (i.e., above 0.20 kgBOD<sub>5</sub>/kgMLSS-d – when problems with the nitrification process arise), the settings were corrected (MLSS, DO).

### Uncertainty in the Neural Network Model

We used the bootstrap method to evaluate the effect of uncertainty of the model structure on the calculations of the activated sludge sedimentation. Bootstrap sampling requires the N – fold draw with return from the data set A = (x<sub>n</sub>, y<sub>n</sub>) of values (x, y). In the paper, each A<sub>N</sub> and A<sub>N</sub> bootstrap set was first used for training and then for testing the model  $\widehat{y}_N(x) = f(x, \widehat{w}_N)$ . In this case, based on the values (x, y) N – samples (training and test set) are generated for n – SE values. On this basis mean values are generated for each from and to p SE values based on the formula:

$$\overline{y}_{boot}(x) = f(x, \{\widehat{w}_N, j = 1, 2, 3, \dots, N\}) = \frac{\sum_{i=1}^N f(x; w_N)}{N} \quad (2)$$

### Analysis of the Model's Sensitivity to the Prediction of Sedimentation of Activated Sludge

In this paper the sensitivity was also analyzed to determine the effect of various biological reactor parameters (MLSS, DO, T<sub>st</sub>) and wastewater quality indicators (BOD<sub>5</sub>, COD, TN, TP, N-NH<sub>4</sub>) on the results of the activated sludge sedimentation (SE) simulations with the ANN model, wherein the model structure is subject to uncertainty. For this purpose, the relationship is used in the form of:

$$\eta_{i,boot} = 100 \cdot \left(\frac{\Delta y}{y}\right)_{i,boot} = f\left(\frac{\Delta x_i}{x_i}\right) \quad (3)$$

...where  $\left(\frac{\Delta y}{y}\right)_{i,boot}$  is the bootstrap estimator (e.g., arithmetic mean) obtained on the basis of the results of calculations for the individual i-th explanatory variables using N of ANN models. In the paper for the assumed values of x<sub>i</sub> and x<sub>i</sub> + Δx<sub>i</sub>, N calculations of SE were performed using N ANN models and on this basis N Δy values were determined. Using the obtained N – results of SE calculations for individual  $\left(\frac{\Delta x_i}{x_i}\right)_j$  values, mean η<sub>i</sub> was determined based on formula (1) and 95% confidence interval was assessed based on the relationship:

$$P \left\{ \overline{\eta}_{i\text{boot}}(x_i) - \frac{u_{\alpha,\text{boot}} \cdot \sigma_{\text{boot}}}{\sqrt{N}} < m < \overline{\eta}_{i\text{boot}}(x_i) + \frac{u_{\alpha,\text{boot}} \cdot \sigma_{\text{boot}}}{\sqrt{N}} \right\} = 1 - \alpha \quad (4)$$

...where:  $\sigma_{\text{boot}}$  is standard deviation and  $u_{\alpha,\text{boot}}$  is the value of the random variable  $U_0$  standardized with normal distribution  $N(0;1)$ .

### Artificial Neural Networks

The multilayer perceptron neural networks used in the study, are now often used to model nonlinear relationships. At the learning stage of the model, based on the results of the measurements in set  $A = (x_n, y_n)$  ( $i = 1, 2, \dots, n_p$ ) for the assumed number of neurons in the hidden layer and the activation function (Fig. 2), the weight values ( $w_{ij}$ ) are estimated and the obtained results of the simulations ( $\mu_y(x)$ ) and measurements ( $y$ ) are associated by the relationship:

$$y - \mu_y(x) = \varepsilon(x) \quad (5)$$

...where the resulting relationship  $\mu_y(x)$  has an implicit character, so only by additional calculations is it possible to determine the effect of individual variables  $x_i$  on the result of simulation  $\mu_y(x)$ , whereas the value  $\varepsilon(x)$  is the so-called white noise described by the Gaussian distribution  $N(0; \sigma_\varepsilon^2)$ .

The thus-obtained model (provided that the structure and a learning method are suitably selected) is usually characterized by satisfactory predictive abilities where the weight values ( $w_{ij}$ ) (Fig. 2) set at the learning stage based on the function in the form:

$$E = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^n (\widehat{y}_n - y_n)^2 \quad (6)$$

...which in these cases reaches a global minimum.

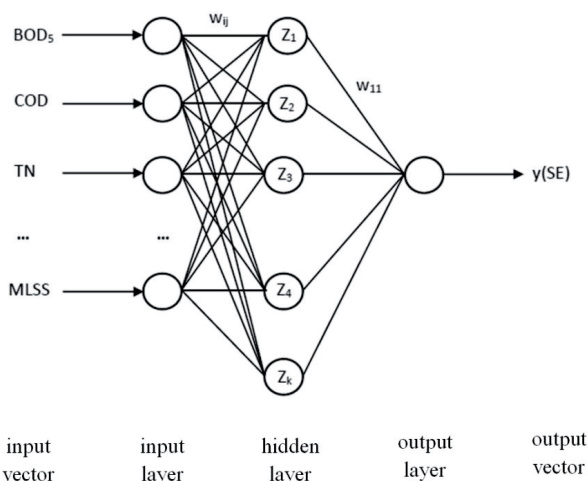


Fig. 2. Scheme of the MLP artificial neural network for the forecast of SE value.

In many cases, obtaining a global minimum, and consequently a model with adequate predictive capabilities, is limited, which may result from an inadequate model structure, which has a significant impact on the prediction of the weight values ( $w_{ij}$ ) in the model. In addition, from a probabilistic point of view, the data  $A = (x_n, y_n)$  used for learning and determining of weights can, at the random selection stage, get into both sets (teaching and test) which can cause problems with the estimation of model parameters and influence the simulation results.

### Optimizing Neural Network Structure (Case Study)

In this study, the set of data ( $x_n, y_n$ ) obtained based on the results of measurements included 130 values, whereby a set of independent data, including  $n = 19$  values, was used to evaluate the prediction ability of the obtained mathematical models. The analyzes assumed that the training set accounted for 80%, while the test set was 20% of data. In the calculation,  $N = 300$  bootstrap samples were taken [9]. On this basis,  $N$  models of artificial neural networks were developed for forecasting the SE and SVI values. For this purpose, a multilayer perceptron method (MLP) was used for the assumed number of neurons in the hidden layer  $S \div 2 \cdot S + 1$ , where:  $S$  is the number of inputs [3, 11] and hyperbolic tangent activation function, artificial neural networks models were developed to predict the sedimentation capacity of activated sludge using the STATISITCA program. On this basis the parameters for the calculation results fit to the measurements were determined, i.e., mean absolute percentage error (MAPE) and mean absolute error (MAE) for training and test sets in bootstrap samples. Based on the results obtained, box-and-whisker plots were constructed, showing the variability of the MAE and MAPE values in the bootstrap samples according to the number of neurons.

### Control, Optimization of Sedimentation of Activated Sludge (SE, SVI) on the Example of a Selected Wastewater Treatment Plant

In the next step, calculations of SE values for 110 data including wastewater quantity and quality, as well as biological reactor operation parameters, were performed on the basis of the calculation scheme for the obtained 300 ANN models shown in Fig. 1. In each of the points ( $i = 1, 2, 3 \dots j = 110$ ) the compatibility of the calculated SE values' distribution with the normal distribution was analyzed using the Kolmogorov-Smirnov test in order to determine the 95% confidence interval based on formula (4). In the next step, for the independent data ( $N = 19$  points) the SE values were calculated using the obtained 300 ANN models, and on this basis the  $SVI = SE \cdot MLSS^{-1}$  was determined. In the case when the determined SVI value exceeded  $SVI_{lim} = 150$  ml/g, the reactor operation parameters

Table 1. Variability range of parameters describing sewage quantity (Q) and quality, and parameters of biological reactor.

Variables	Units	Minimum	Mean	Maximum	Standard deviation
Q	m <sup>3</sup> /d	32564	40698	86592	8088
T <sub>os</sub>	°C	10.0	15.9	23.0	3.58
pH	-	7.2	7.6	7.8	0.20
MLSS	kg/m <sup>3</sup>	2.01	4.26	6.52	1.04
DO	mg/dm <sup>3</sup>	0.55	2.56	5.78	1.03
BOD <sub>5</sub>	mgO <sub>2</sub> /dm <sup>3</sup>	127	309	557	86.01
COD	mgO <sub>2</sub> /dm <sup>3</sup>	384	791	1250	173.98
TP	mg/dm <sup>3</sup>	4,3	7,8	12,6	0,46
N-NH <sub>4</sub> <sup>+</sup>	mg/dm <sup>3</sup>	24.4	7.8	65.9	7.14
TN	mg/dm <sup>3</sup>	33.91	77.73	124.09	10.62
SVI	ml/g	95	186	320	37
SE	cm <sup>3</sup> /dm <sup>3</sup>	250	790	980	141

(MLSS, DO) were adjusted in such a way that the relationship  $SVI < SVI_{lim}$  was fulfilled. In addition, the substrate load (F/M) was taken into consideration in the analyses to evaluate the reactor's operation correctness, i.e., considering the processes of organic compounds as well as nitrogen and phosphorus removal.

In the calculations, it was assumed that the SVI value was adjusted by appropriate modification of the oxygen concentration in the nitrification chamber (in the paper an increase by 10% from the baseline was assumed), and then by the concentration of activated sludge, so that the modeled SVI value was 150 ml/g. The adopted solution is a frequently used approach in engineering practice, as confirmed by research in this field [13, 14]. Based on the obtained results of the MLSS calculations, the mean MLSS value and the 95% confidence interval for the respective DO values were calculated for each considered point (N = 19).

## Results and Discussion

### Artificial Neural Network Model – Multilayer Perceptron

The variability ranges of the results of measurements of the quantity and quality of wastewater entering the treatment plant as well as the biological reactor parameters were determined and are presented in Table 1. These data show that there was a high variability of pollutant loads entering the activated sludge chambers, which contributed to a considerable variation in the operating parameters of the biological reactor. The box-and-whisker plots (Fig. 3) show the range of variability (mean, standard deviation, 95% confidence interval) for the MAE and MAPE values in the obtained bootstrap samples for different neural

network structures, respectively for training and test sets.

Based on the data shown in Fig. 3, we found that the number of neurons in the hidden layer does not significantly affect the variability and value of the obtained absolute and percentage error for the activated sludge sedimentation capacity forecast. The mean MAE value in the training and test set varies in the small range, i.e. 30.2÷31.9 cm<sup>3</sup>/dm<sup>3</sup> and 32.3÷34.1 cm<sup>3</sup>/dm<sup>3</sup>, while the 95% confidence interval in the above cases ranges within 10.8÷13.2 cm<sup>3</sup>/dm<sup>3</sup> and 12.1÷14.2 cm<sup>3</sup>/dm<sup>3</sup>. The range of changes in the minimum and maximum values for individual numbers of neurons in the bootstrap samples indicates the effect of the uncertainty of the identification of weight values in the model structure on the mathematical model ability to predict the activated sludge sedimentation capacity. In order to supplement the simulation sedimentation results in Fig. 4, the variability of the SVI and SE values in the period covered by the calculations is shown. Based on the data shown in Fig. 4, it can be concluded that the vast majority of the results of the sedimentation and sludge volume index measurements for the training and test set falls within the scope of the stochastic solution and 95% of the data is within the range of the confidence interval. At the same time, based on the results of calculations presented in Fig. 4 we found that in the considered period of 2015-2017 there were problems of activated sludge bulking, as confirmed both by the sludge volume index and substrate load values.

The mean values of errors (MAE) obtained in the study for prognoses of SE values based on an MLP model are slightly higher than those determined by Szelağ and Gawdzik [5]. Based on the measurements of Q, T, BOD, TN, TP, N-NH<sub>4</sub> and MLSS acquired at the BADENPHO treatment plant and using the MARS

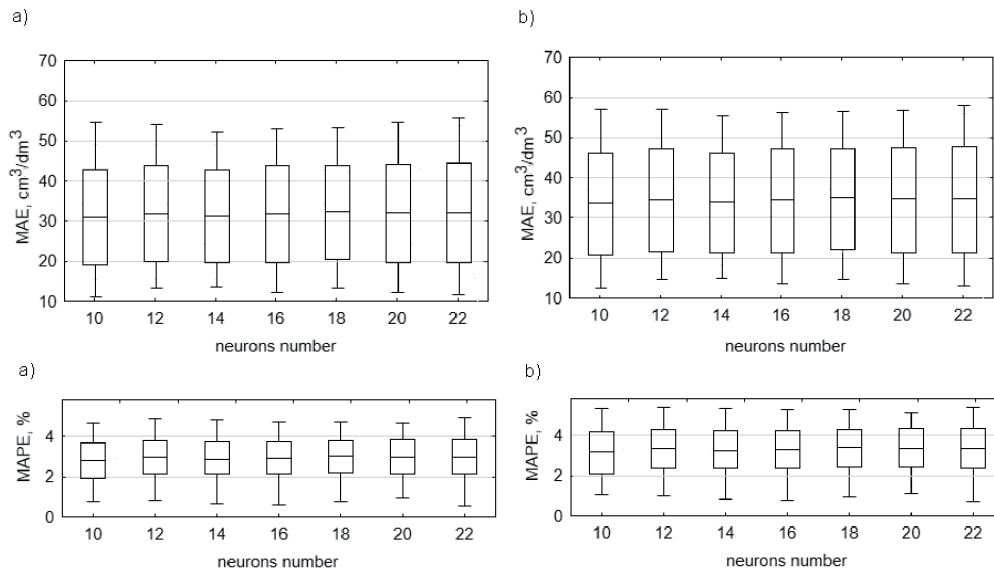


Fig. 3. Effect of neuron number in the network hidden layer on the variability of MAE and MAPE error values for the sludge sedimentation capacity forecast in the bootstrap samples for the sets: a) training, b) test.

(multivariate adaptive regression spline) method, they obtained a model for which IEA = 27.1 cm<sup>3</sup>/dm<sup>3</sup>. Similar results of SE value simulation (Fig. 3) were obtained either by Szeląg and Gawdzik [6], who on the basis of temperature value in the bioreactor (EvuPerl system) and SE value taken on time (t-1) in the measurement preceding the simulated value calculated at the moment (t), developed an MLP model and obtained MAE = 36.63 cm<sup>3</sup>/dm<sup>3</sup>. The fact that precipitation of activated sludge is an important technological parameter is confirmed by the work [13] in which Bagheri and others developed models using the MLP method to simulate the quality of wastewater at the treatment plant

outlet. Their simulations showed a significant influence of SE parameter on the values of BOD<sub>5</sub>, COD, TN, etc.

### Sensitivity Analysis of the ANN Model

The conducted model sensitivity analysis (Fig. 5) provides important information, as it is possible to clearly evaluate the effect of the considered input data on the result of the simulation, i.e., the activated sludge sedimentation capacity. Moreover, based on the conducted analyses (Figs 3 and 4), the effect of the ANN model structure uncertainty on the SE change can be determined when the value of the considered

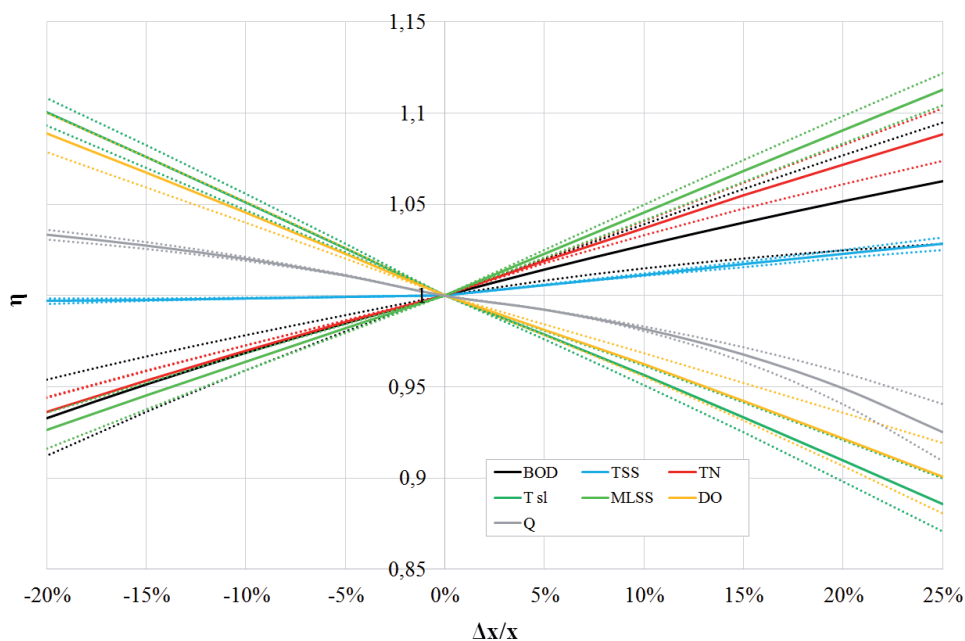


Fig. 4. Results of sensitivity analysis describing the relationships  $\eta_i = f(\Delta x_i \cdot x_{i-1})$ .

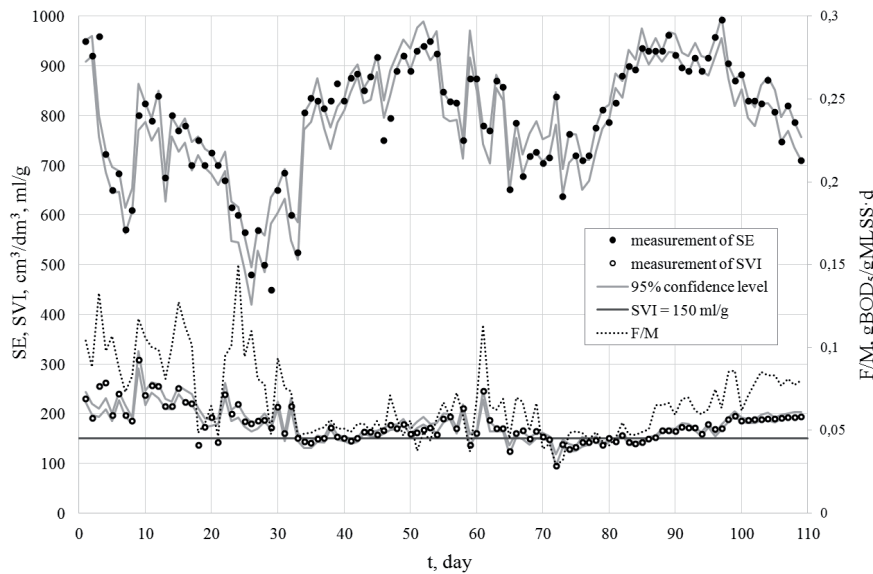


Fig. 5. Summary of measurement results and the stochastic solution of SE and SVI in the analyzed period (2015-2017) at the Sitkówka-Nowiny treatment plant.

variable increases by  $\Delta x_i$  with respect to the input value ( $x_i$ ). Due to the large number of explanatory variables, examples of sensitivity analysis are presented (BOD<sub>5</sub>, TSS, TN, T<sub>sl</sub>, MLSS, DO, Q). The data presented in Fig. 5 show an increase in the values of quality indicators, i.e., BOD<sub>5</sub>, TSS, TN, TP. The increase in the value of MLSS leads to an increase in SE; nevertheless, a detailed analysis of the values of  $\eta = f(\text{MLSS})$  in Fig. 5 shows that the SVI value decreases. On the other hand, the increase in temperature (T<sub>sl</sub>), wastewater inflow to the treatment plant (Q) and the oxygen concentration in nitrification chambers (DO) leads to a decrease in the value of SE. These relationships were

found by Bagheri et al. [14] and by Szeląg and Gawdzik [5], who used artificial intelligence methods to estimate the sedimentation of activated sludge in municipal wastewater treatment plants. This approach, as well as confirmation of the results obtained in the paper, can be found in the works of Han et al. [2], Lou and Zhao [3] and Liu et al. [15]. In addition, based on the obtained  $\eta = f(x_i)$  curves (Fig. 5), we found that the uncertainty of the MLP model at the control stage (maintenance of SVI within an appropriate range of SVI < 150 ml/g) affects the results of calculations of operating parameters of the reactor, leading to ambiguous solutions for the operation of the facility.

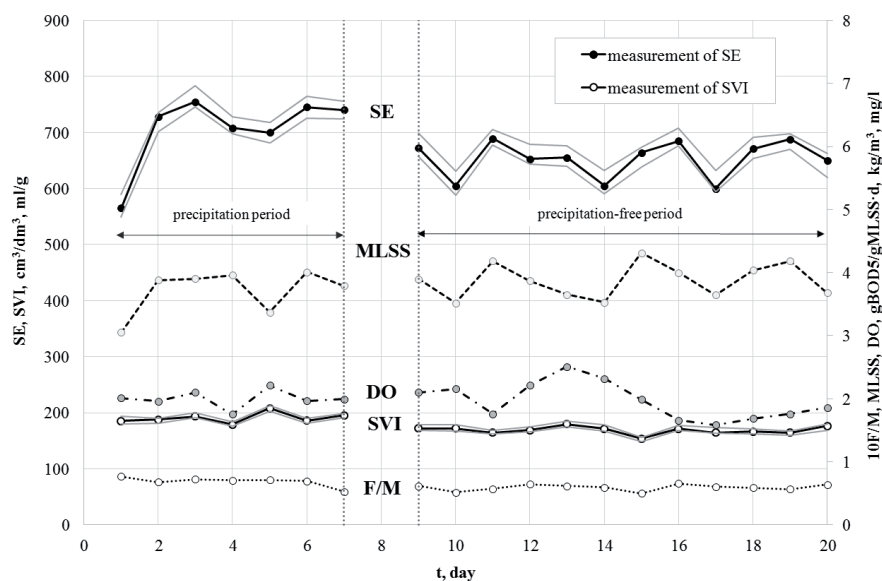


Fig. 6. Summary of calculation results and measurements of SE and SVI, and biological reactor operating parameters (MLSS, DO) for the validation set.

### Control of Sedimentation of Activated Sludge during Rainfall and Rainless Periods

Data shown in Fig. 6 for the precipitation and precipitation-free periods indicate that the measured SE and SVI values, including validation data set, are within the range of the probabilistic solution and within the defined confidence interval, and the obtained values indicate problems with the sedimentation of activated sludge ( $SVI > 150$  ml/g). Based on the data shown in Figure 6, we found that in the precipitation period there was an increase in the substrate load of the activated sludge chambers and in the precipitation-free period the F/M value was lower (i.e., on average by 16%) than in the precipitation period. At the same time, during the precipitation period the mean value of the activated sludge sedimentation for the analyzed events was higher by 8% than in the precipitation-free period. In view of the above results, the possibility of correcting the operating parameters of the reactor, as discussed in the methodology, was analyzed; the simulation results are shown in Fig. 7.

Based on the data presented in Fig. 7, we found that the increase in the DO value by 10% and the corresponding increase in MLSS leads to an improvement in sedimentation capacity of activated sludge as the modeled value of sludge volume index did not exceed 150 ml/g. The calculated results showed that the predicted expected SE values in the precipitation and precipitation-free periods (Fig. 7), where  $SVI = 150$  ml/g, proved to be higher than the SE values measured for the actual condition ( $SVI > 150$  ml/g) by  $18 \div 125$   $\text{cm}^3/\text{dm}^3$  (on average by  $94$   $\text{cm}^3/\text{dm}^3$ ) and  $3 \div 64$  (on average by  $24$   $\text{cm}^3/\text{dm}^3$ ). At the same time, based on the calculation results (Fig. 6), we found that in the precipitation and precipitation-free periods the MLSS values are higher

than the values for the actual condition ( $SVI > 150$  ml/g) respectively by  $0.99 \div 2.10$   $\text{kg}/\text{m}^3$  (on average by  $1.61$   $\text{kg}/\text{m}^3$ ) and  $0.31 \div 1.28$   $\text{kg}/\text{m}^3$  (on average by  $0.61$   $\text{kg}/\text{m}^3$ ).

It is worth noting that the MLSS values obtained in four cases in the precipitation period are higher than  $5.00$   $\text{kg}/\text{m}^3$ , which is assumed as a maximum value in biological systems of carbon, nitrogen and phosphorus removal. In addition, the F/M values obtained from the calculations (Fig. 7) are within the optimum range (i.e., about  $0.05$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$ ), indicating no problems with the removal of C, N, and P compounds from wastewater and significant improvement of sedimentation capacity of sludge with respect to the actual conditions, as supported by the calculated values of  $SVI = 150$  ml/g. It is also worth noting that the adjustment of the reactor operating parameters (MLSS, DO) in the precipitation and precipitation-free period caused the equalization of the biological reactor substrate load with the organic pollutants; during the precipitation and precipitation-free periods the mean F/M values do not differ by more than 2% and do not exceed  $0.05$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$ . Based on the data presented in Figures 6 and 7, it can be stated that the increase in MLSS and DO values in the precipitation and precipitation-free period caused a decrease in the substrate load of the biological reactor and sludge volume index by  $0.0157 \div 0.0275$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$  and  $35.10 \div 57.72$  ml/g (on average by  $0.0226$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$  and  $40.89$  ml/g), and  $0.0081 \div 0.0159$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$  and  $2.32 \div 30.3$  ml/g (on average by  $0.008$   $\text{gBOD}_5/\text{gMLSS}\cdot\text{d}$  and  $19.53$  ml/g).

In addition, the data presented in Figs 6 and 7 are confirmed in the above calculations, which indicates the uncertain effect of the ANN model on the results of calculations of the biological reactor operating parameters at the control stage.

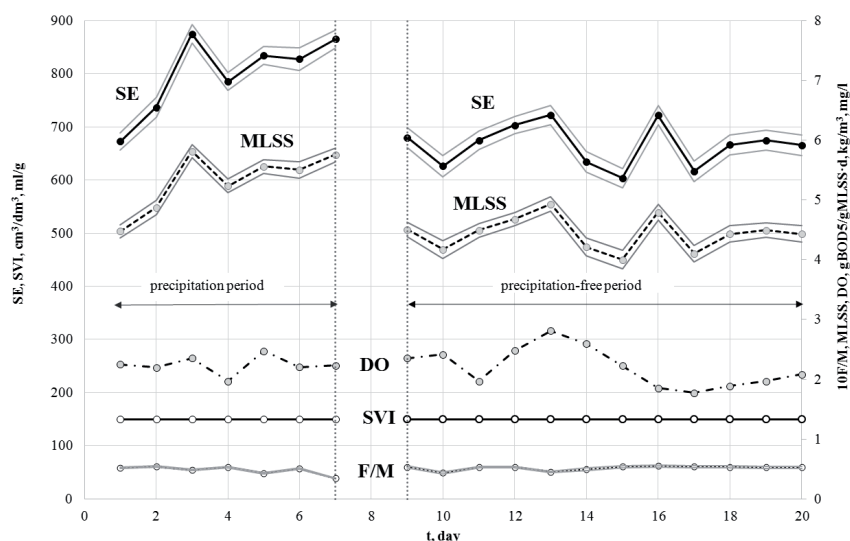


Fig. 7. Summary of calculation results of operating parameters (DO, MLSS) and sludge sedimentation capacity (SE), with the  $SVI = 150$  ml/g.



## Conclusions

Based on the conducted analyzes, we found that the uncertainty of the MLP model structure affects the abilities of the developed models for the prediction of the activated sludge settling (sedimentation capacity and sludge volume index), as confirmed by the ranges of the calculated MAE and MAPE errors in the bootstrap samples. These results are also confirmed by the sensitivity analysis performed within this study. In addition, the calculations show that sludge sedimentation, and hence also sludge volume index, can be predicted based on the wastewater quality indicators (BOD<sub>5</sub>, COD, TSS, TN, TP, N-NH<sub>4</sub>) and biological reactor operating parameters (pH, T<sub>sl</sub>, MLSS, DO). The model presented in this paper is a development of other research in this field, as it contains parallel modeling of two important technological parameters (sedimentation capacity and sludge volume index), which are used by the technologist to evaluate the activated sludge settling in the wastewater treatment plant.

The presented analyses also confirmed the significant influence of the neural network model structure on the uncertainty of estimating biological reactor operating parameters (mixed liquor suspended solids, concentration of oxygen), which in practical considerations leads to problems of continuous control of the sludge sedimentation capacity.

Having regard for the obtained results, further analyzes are needed to develop the methodology for controlling the biological reactor parameters, taking into account the uncertainty of the developed mathematical model.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- 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. Exploratory manual of the wastewater treatment plant. Poznań: Polskie Zrzeszenie Inżynierów i Techników Sanitarnych. **2011** [In Polish].
- 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**.
- LOU I., ZHAO Y. Sludge Bulking Prediction Using Principle Component Regression and Artificial Neural Network. *Mathematical Problems in Engineering*, **2012**, 1, **2012**.
- BEZAK-MAZUR E., STOJŃSKA R., SZELAĞ B. Ocena wpływu parametrów operacyjnych i występowania bakterii nitkowatych na objętościowy indeks osadu czynnego – studium przypadku [Evaluation of the influence of operational parameters and occurrence of bacteria on volumetric sludge index - case study]. *Rocznik Ochrona Środowiska*, **18**, 487, **2016**.
- SZELAĞ 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. *Polish Journal of Environmental Studies*, **26** (1), 315, **2017**.
- SZELAĞ B., GAWDZIK J. Application of Selected Methods of Artificial Intelligence to Activated Sludge Settleability Predictions. *Polish Journal of Environmental Studies*, **25** (4), 1709, **2016**.
- SZELAĞ B., BARBUSIŃSKI K., STUDZIŃSKI J. Application of the model of sludge volume index forecasting to assess reliability and improvement of wastewater treatment plant operating conditions. *Desalination and Water Treatment*, 1-12, **2019** [In Press].
- AMANATIDOU E., SAMIOTIS G., TRIKOILIDOU E., PEKRIDIS G., NIKOLAOS TAOUSANIDIS. Evaluating sedimentation problems in activated sludge treatment plants operating at complete sludge retention time *Water Research Journal*, **69**, 20, **2015**.
- KOWALSKA E., PATUREJ E., ZIELIŃSKA M. Use of *Lecane inermis* for control of sludge bulking caused by the *Haloscomenobacter* genus. *Desalination and Water Treatment*, **57** (23), **2016**.
- SRIVASTAV R.K., SUDHEER K.P., CHAUBEY I. A simplified approach to quantifying predictive and parametric uncertainty in artificial neural network hydrologic models. *Water Resources Research*, **43** (10), 1, **2007**.
- SRIVASTAV R., SRINIVASAN K., SUDHEER K.P. Simulation-optimization framework for multi-site multi-season hybrid stochastic streamflow modeling. *Journal of Hydrology*, **542**, 506, **2016**.
- SZELAĞ B., STUDZIŃSKI J. A Data Mining Approach to the Prediction of Food – to – Mass Ratio and Mixed Liquor Suspended Solids. *Polish Journal of Environmental Studies*, **26**, 5, **2017**.
- COMAS J., RODRIGUEZ-RODA I., SANCHEZ-MARREB M., CORTESB U., FREIXOC A., ARRAEZC J., POCHA M.A. knowledge-based approach to the deflocculation problem: integrating on-line, off-line, and heuristic information. *Water Research*, **37**, 2377, **2003**.
- BAGHERI 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 Safety and Environmental Protection*, **95**, 12, **2015**.
- LIU Y., GUO J., WANG Q., HUANG D. Prediction of filamentous sludge bulking using a state – based Gaussian process regression model. *Scientific Reports*, **6** (31303), **2016**.