Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2021.27674

232 (2021) 421–432
August

Assessment of wastewater quality indicators for wastewater treatment influent using an advanced logistic regression model

Bartosz Szeląg[a,*], Mohamad-Javad Mehrani[b], Jakub Drewnowski[b,*], Monika Majewska[a], Grzegorz Łagód[c], Sheena Kumari[d], Faizal Bux[d]

[a]*Faculty of Environmental, Geomatic and Energy Engineering, Kielce University of Technology, Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, Poland, emails: bszelag@tu.kielce.pl (B. Szeląg), majewska.monika1995@gmail.com (M. Majewska)*
[b]*Faculty of Civil and Environmental Engineering, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland, emails: jdrewnow@pg.edu.pl (J. Drewnowski), mohammad.mehrani@pg.edu.pl (M.-J. Mehrani)*
[c]*Faculty of Environmental Engineering, Lublin University of Technology, Nadbystrzycka 40B, 20-618 Lublin, Poland, email: g.lagod@pollub.pl*
[d]*Institute for Water and Wastewater Technology, Durban University of Technology, Durban 4000, South Africa, emails: sheenak1@dut.ac.za (S. Kumari), faizalb@dut.ac.za (F. Bux)*

Received 11 March 2021; Accepted 3 August 2021

A B S T R A C T

Influent quality indicators play a significant role in wastewater treatment plant performance due to their correlation with reactor operations and effluent quality. However, selecting a specific/best parameter indicator for predicting influent wastewater quality is one of the challenges in wastewater treatment. This study, therefore, focused on determining suitable variables as influent quality indicators. For this purpose, a logistic regression model involving different inflow parameters from two wastewater treatment plants in Poland was used to identify the best wastewater parameter as a suitable indicator for operational monitoring, process control and simulation purpose. The results showed that the model is flexible enough to simultaneously predict two or three effective wastewater quality indicators. Furthermore, the sensitivity analysis results showed a strong nonlinear relationship between the complex values of total nitrogen, total phosphorus and suspended solids.

Keywords: Wastewater treatment plants; Influent quality; Logistic regression; Sensitivity analysis; Quality indicators

1. Introduction

Quality of the influent wastewater is an essential factor affecting wastewater treatment plant (WWTP) performance, operation and control. The variability of selected wastewater quality over time determines the optimal selection of settings and operation of the bioreactors, which has a significant impact on final effluent quality and thus allows decreasing the number of probes by reducing the energy to aeration [1–4]. Finding the relationships between independent variables like influent or effluent can be helpful for technical and modeling purposes using machine learning methods [5–12] and statistical models [7,9,13–15]. In addition, few studies reported using a logistic regression model to predict pipe failures in the water supply network, with high accuracy [16–18]. Avila et al. [15] showed the possibility of using logistic regression to simulate the water quality in the Oreti River, South of New Zealand. These analyses also confirmed the calculations performed by Thoe et al. [19] based on the case study on the water quality of Santa Monica Beach. Similarly, Saha and Pal [20] studied the physical hazards of wetlands in the Atreyee Basin in India and

* Corresponding authors.

Presented at the 1st International Conference Strategies toward Green Deal Implementation – Water and Raw Materials (ICGreenDeal2020), 14–16 December 2020, held online by the Mineral and Energy Economy Research Institute, Polish Academy of Sciences

Bangladesh. In another study, logistic regression was used to determine the likelihood of groundwater pollution [11]. Twarakavi and Kaluarachchi [21] determined the probability of heavy metals occurrence in groundwater using the case study of Sumas-Blaine Aquifer in Washington State. Lee et al. [11] developed a logistic regression model for forecasting the water quality in a housing well on the Lanyang Plain in Japan. The earlier analyses of this problem have already been undertaken by Yamijala et al. [16] based on the case study of a city in the state of Texas in the United States of America (USA). For instance, based on the models developed by Dogan et al. [22] for predicting biological oxygen demand (BOD_5), the effects of total suspended solids (TSS), total phosphorus (TP) and chemical oxygen demand (COD) on the calculation results cannot be easily determined. Additional simulation calculations are necessary. It is similar to the models developed by Szeląg et al. [23], who demonstrated the possibility of modeling and identifying ranges of the variability of selected values of wastewater quality indicators using cascade neural networks, and support vector machines based on flow rate.

To simulate the quality of wastewater, Verma et al. [24] used the artificial neural networks (ANN) method, based on the results of carbonaceous biochemical oxygen demand or the values of flow rate and total suspension. Mair and El-Kadi [25] found the logistic regression (LR) model is flexible enough for a vulnerability study and creating separate groups of wells and introducing some variables as a suitable indicator of groundwater contamination. Maniquiz et al. [26], using linear regression models, showed that the pollutant load could be raised due to heavy rainfall. Moreover, Thoe et al. [19] created and compared five models for predicting the water quality (enterococci and fecal coliform concentrations) based on workability and accuracy. They found that the classification tree, ANN and LR methods are the three best models. Another study conducted by Avila et al. [15] evaluated the performance of various statistical models, including multinomial logistic regression for real-time water quality prediction.

In this study, a logistic regression model was developed to identify specific indicator values of wastewater quality in the inflow of WWTPs. The real data from two WWTPs located in the north and south of Poland were prepared. Total nitrogen (TN), TP, and TSS were selected as independent variables and applied to the model based on the influent ammonium-nitrogen (NH_4–N) values. Furthermore, a sensitivity analysis was carried out for assessing the impact of selected independent variables on the calculation of the results.

2. Methodology

2.1. Wastewater treatment plants

The measurement data of two Polish WWTPs located in the southern part of Poland (Rzeszow) and in the northern part (Gdansk-Wschod), shown in Fig. 1, are included in this study.

The municipal WWTP in Rzeszow (50°06′ and 22°03′) is located on the right bank of the Wislok River and has been operating since 1988. It was designed for a flow of Q = 62,500 m^3/d and 400,000 population equivalent (PE). The WWTP currently operates in a sanitary wastewater network with a total length of 785.6 km and 94 wastewater pumping stations. The difference in land ordinates within the city between the highest (384 m above sea level) and the lowest point (197 m above sea level) is 183 m. The area from which wastewater flows into the WWTP covers approximately 126.6 km^2. The average annual temperature in the catchment area is 7.5°C, while the amount of precipitation is around 615 mm/y. The warmest month is July with an average temperature of 18.6°C, with the highest rainfall of 89 mm.

The municipal WWTP Gdansk-Wschod (54°23′ and 18°28′) receive approximately 96,000 m^3/d of wastewater flows. After the renovation in 2012, its capacity increased to 120,000 m^3/d and 860,000 PE. Wastewater flows through a sewerage system, which works in a gravity-pumping system. The wastewater from neighboring communes, that is, Kolbuda, Żukowo, Gmina Pruszcz Gdański and Miasto Sopot, also flows into the WWTP, which gives a catchment area of 300.71 km^2. The difference between the highest

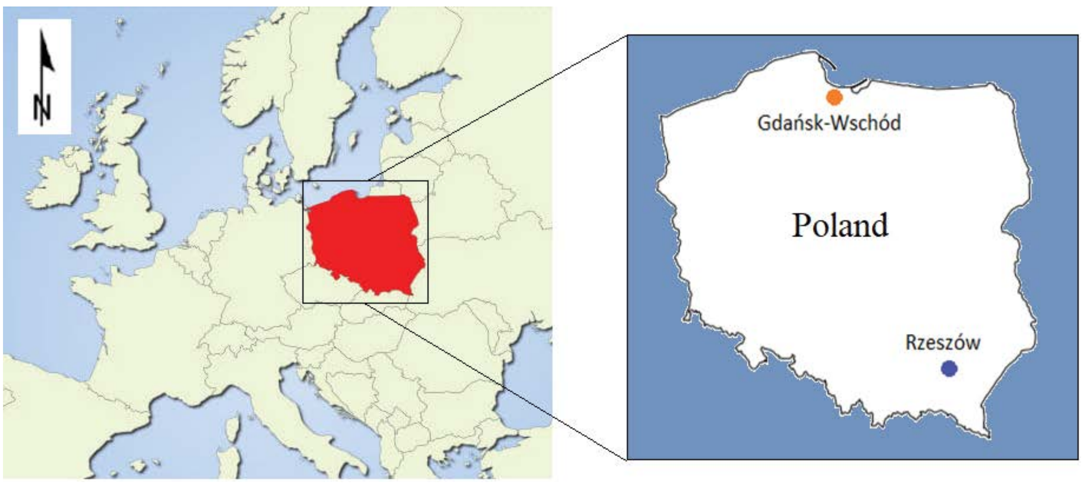

Fig. 1. Location of the Gdansk-Wschod and Rzeszow WWTPs on the map of Poland.

(180.1 m above sea level) and the lowest point (3 m below sea level) within the city is 183 m. The area from which wastewater flows into the WWTP is serviced by a wastewater network with a length of about 1,311 km. The treated wastewater is discharged via a 2.5 km pressure pipeline to Gdansk Bay. The average annual temperature in the catchment area is 6.7°C, while the precipitation is around 541 mm/y. The warmest month is July, with an average temperature of 16.2°C, with the highest rainfall of 70 mm.

Table 1 presents the statistical analysis of the measured indicators of raw data for both treatment plants.

In the period under consideration, the hydraulic load for the Gdansk-Wschod WWTP was on average at 86,649 m^3/d, while for Rzeszow it was 38,572 m^3/d. The wastewater flowing into the WWTPs was characterized by heterogeneous qualitative composition.

2.2. New concept of the algorithm to identify the wastewater quality indicators of WWTPs influent

Based on the logit model described above, a new concept was proposed for identifying the quality of wastewater at the inflow to the treatment plant (Fig. 2). At the first stage of analysis, models are used to identify wastewater quality indicators (TN, TP, TSS).

In the next stage, depending on the results of calculations (probability values of p), individual conditions were checked to determine the ranges of variability in the wastewater quality indicators, respectively. The algorithm

Table 1
Data values measured at the influent of the Gdansk-Wschod and Rzeszow WWTPs

Indicators of pollution		Min.	Max.	Average	Median	Standard deviation
Total nitrogen (g/m^3)	Gdansk-Wschod	28.0	108.0	79.7	80.0	9.3
	Rzeszow	21.3	99.0	69.7	71.0	11.6
Total phosphorus (g/m^3)	Gdansk-Wschod	5.5	19.0	12.8	12.8	2.0
	Rzeszow	3.4	37.5	12.4	12.3	3.3
Total suspension (g/m^3)	Gdansk-Wschod	109.0	924.0	449.8	438.0	110.8
	Rzeszow	80.0	1,140.0	430.1	430.0	106.3
Ammonium nitrogen (g/m^3)	Gdansk-Wschod	22.1	82.9	56.4	56.5	6.7
	Rzeszow	13.7	80.0	54.4	56.0	10.1

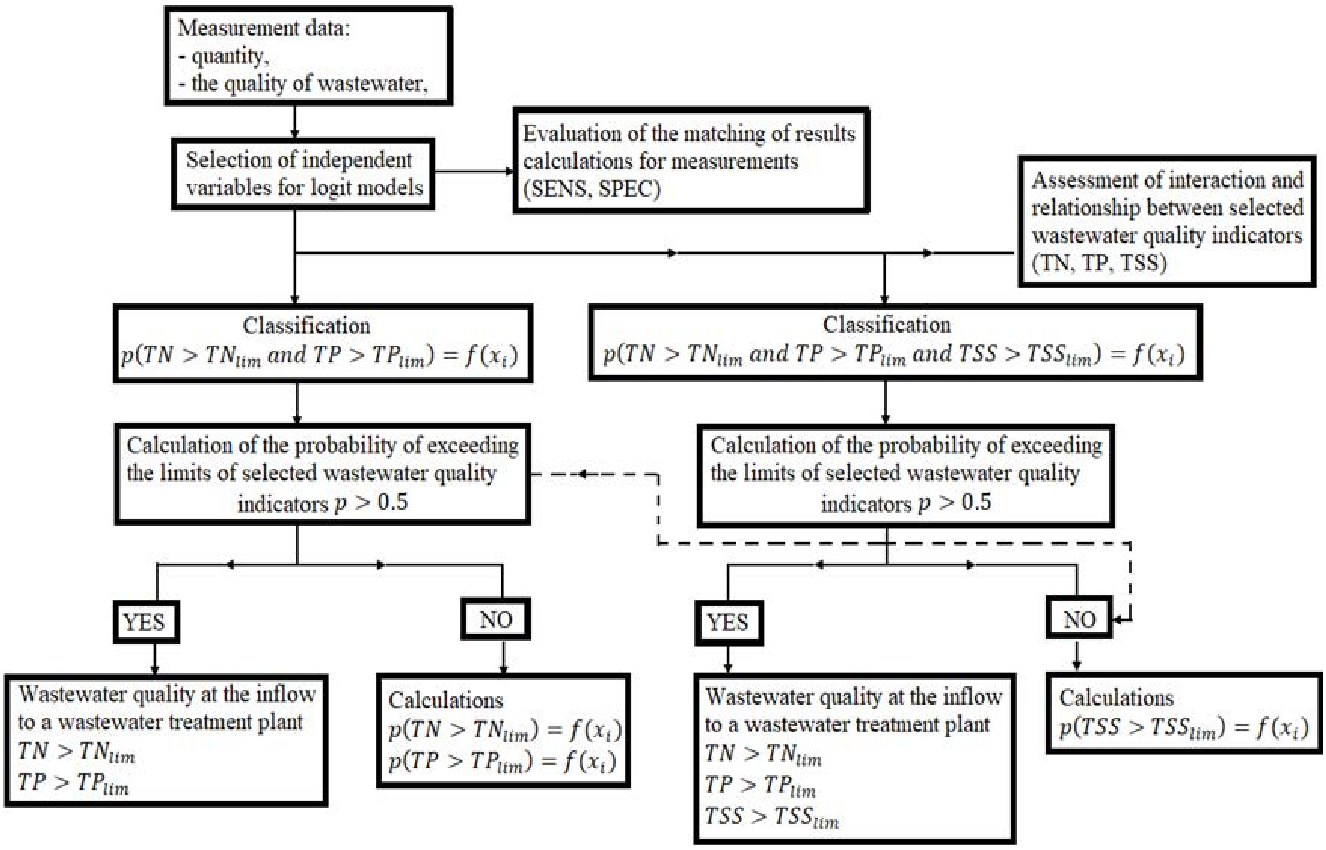

Fig. 2. Algorithm for identifying the influent wastewater quality using a logistic model.

of the system shown in Fig. 2 has the ability to analytically describe the relationship between the selected indicators and independent variables. The solution adopted in the work allows assessing the impact of inclusion in the calculation models of subsequent dependent variables (TN, TP, TSS, TN and TP and TSS), which has not been analyzed by the researchers involved in the simulation of wastewater quality indicators. The adopted approach also allows for a detailed analysis of the impact of selected independent variables on the modeled quality indicators. This is also very important because it allows supervised search for similarities in data groups, taking into account the variability of the values of selected independent variables. This is a significant simplification in the methods used to identify similarities in multidimensional data sets (HCA – hierarchical cluster analysis, K-NN – K-nearest neighbors or Kohonen neural networks, ANN – artificial neural network) so far. In the proposed approach, there is no need to implement complex numerical algorithms and the obtained results have a physical interpretation, which is not limited to the above-mentioned statistical methods. These analyses are supplemented by a sensitivity analysis, in accordance with Eqs. (2)–(4), which allows assessing the impact of various ranges of selected independent variables on the modeled values of wastewater quality indicators. The above-mentioned aspect may have practical significance because it can be used at the bioreactor modeling and optimization stage, limiting the pollution load of technological objects.

2.3. Logistic regression

Logistic regression model, also called binary regression, is used to analyze binominal data. Compared to the methods commonly used in classification issues such as the linear regression model, artificial neural networks, regression trees, etc. The logistic model allows probability modeling, which is a significant advantage [27–30]. This method can be used for modeling a variety of scientific data like complex phenomena in ecology, hydrology, water, water supply network failure [16,17] and wastewater modeling [18,31–34]. The analyses performed showed that the logit model was used to identify the quality of surface water effluents in flowing [15] and stagnant waters [20], underground water [11,21] as well as perform the analysis of the relationship between the activity of the society and the concentration of heavy metals [35]. The logistic regression model in its general form can be described by the following relationship:

$$p = \frac{1}{1+\exp\left(-\left(\sum_{i=1}^{j}\alpha_i x_i + \alpha_0\right)\right)} \quad (1)$$

where p is the probability of exceeding of limit values – other independent variables included in the model $i = 1, 2, 3, \ldots, j$.

On the basis of the definition of the local sensitivity coefficient given by Petersen et al. [36] it can be concluded that there is the following relationship:

$$S_{\mathrm{xi}} = \frac{\partial p}{\partial x_i}\cdot\frac{x_i}{p} \quad (2)$$

Knowing that there is a dependence of characters:

$$\frac{\partial p}{\partial x_i} = \alpha_i \cdot p \cdot (1-p) \quad (3)$$

which can be written the following equation for the value of the sensitivity coefficient:

$$S_{\mathrm{xi}} = \alpha_i \cdot p \cdot (1-p)\cdot\frac{x_i}{p} = \alpha_i \cdot x_i \cdot (1-p) \quad (4)$$

On the basis of the above-mentioned relationship, the values of the sensitivity factor S_{xi} in the surroundings of the value x_i can be determined. In addition, while analyzing Eq. (4), it can be stated that when determining the value of S_{xi}, the probability value of p can be taken into account, which can be related to the numerical quantities of the modeled dependent variable $p = f(y)$. On this basis, it can be concluded that the value of $p_1 > p_2$ the calculated values of the sensitivity coefficient will differ, and the higher the probability value is, the smaller the S_{xi} values.

In order to assess the predictive abilities of the logit model, the SENS – sensitivity, expresses the correct case classification when ($y > y_{\mathrm{lim}}$) and SPEC – specificity, expresses the correct case classification when ($y < y_{\mathrm{lim}}$) was used, which was discussed in detail by Harrell [37].

In this work, the logistic regression model tries to identify the suitable range of variability of single wastewater quality indicators (TSS, TN, TP), based on the results of quality measurements at the inlet of Gdansk-Wschod and Rzeszow WWTPs. Logit models also can be used to classify the variability of two indicators in set (TSS, TN), (TN, TP), and three (TSS, TN, TP) wastewater quality indices simultaneously. On the basis of the measurement data for the Gdansk-Wschod WWTP, reference models were determined, while, based on the data from Rzeszow WWTP, the possibility of identifying selected indicators for the previously given WWTP was checked.

2.4. Classification for the variability of wastewater quality indicators

In the classic approach, the logistic regression model was used to identify the numerical values of individual dependent variables. For example, the limit value of total nitrogen at the influent of WWTPs as ($\mathrm{TN_{lim}}$) is expressed by means of two independent variables:

$$p\left(\mathrm{TN} = \mathrm{TN_{lim}}\right) = \frac{1}{\exp\left(-\left(\alpha_1 x_1 + \alpha_2 x_2 + \alpha_0\right)\right)} \quad (5)$$

However, considering that the value of $\mathrm{TN} > \mathrm{TN_{lim}}$ and $\mathrm{TSS} > \mathrm{TSS_{lim}}$, the above-mentioned criteria can be separated, and the logistic regression model can be written as:

$$p\left(\mathrm{TN} = \mathrm{TN_{lim}} \text{ and } \mathrm{TSS} = \mathrm{TSS_{lim}}\right) = \frac{1}{\exp\left(-\left(\alpha_{11}x_1 + \alpha_{21}x_2 + \cdots + \alpha_{i1}x_i + \alpha_{01}\right)\right)} \quad (6)$$

Moreover, assuming from the literature data [31,38] that in Eq. (5), the value of TN_{lim} is equivalent to the probability value, $p = 0.50$ which can be written as follows:

$$\alpha_1 x_1 + \alpha_2 x_2 + \alpha_0 + 0.693 > 0 \quad (7)$$

However, assuming for Eq. (6) the value of $p = 0.5$ equivalent to TN = TN_{lim} and TSS = TSS_{lim}. It can be written $p > 0.5$ for TN > TN_{lim} and TSS > TSS_{lim}.

$$\alpha_{11} x_1 + \alpha_{21} x_2 + \cdots + \alpha_{i1} x_i + \alpha_{01} + 0.693 > 0 \quad (8)$$

In addition, $p < 0.5$ for the values of TSS < TSS_{lim} and TN > TN_{lim} or vice versa. Following the methodology described above, the logit model may include subsequent values of wastewater quality indicators constituting the basis for the identification of multidimensional classes of quality variability.

2.5. Assessment of the interactions and relationships between the selected wastewater quality indicators (TN, TP, TSS)

In the classic approach, logistic regression is usually used to identify the classes of the variability of individual dependent variables (wastewater quality, wastewater pouring, sludge bulking, etc.). The adopted model identifies in parallel the ranges of variability in two (TN and TP) and even three (TN, TP, TSS) indicators. Therefore, the differentiation of the obtained p(TN) or p(TN, TP) values for independent variables (x_j) is a measure of the interaction between them. The above-mentioned dependence can be written as follows:

$$p(\text{TN},\text{TP}) = f\left(p(\text{TN}), x_1, x_2, \ldots, x_j\right) \quad (9)$$

In the theoretical case, that is, when a single variable is analyzed, the relationship above Eq. (9) is linear and as p(TN) = p(TN,TP) and may indicate that TP has no influence on the identified quality class expressed by TN, TP. For the assumed values of x_j, the obtained values have the form of $p(\text{TN}) \neq p(\text{TN,TP}) = f(x_1, x_2, \ldots, x_j)$. The adopted solution can thus be used to analyze the impact of selected independent variables on the relationship between the modeled wastewater quality indicators or multidimensionality of their variability, that is, TN and TP or TN, TP, TSS. This is important from the point of identification for processing that occurs in flowing wastewater and the possibility of modernization by designing systems depending on local conditions to ensure the quality of wastewater at the inflow of WWTPs and has a large impact on the optimization of processes at the WWTP.

3. Results and discussion

3.1. Logistic regression models

The measured data for the Gdansk-Wschod WWTP (320 samples) and Rzeszow WWTP (900 samples), including COD, NH_4–N, TN, TSS, TP, phosphate-phosphorus (PO_4–P), pH, and Q were considered in this study. The range of each value variation was determined as TN and TP and on the same TN, TP, and TN, TP, TSS constituting the basis for building a logit model for both plants (Table 2). In the analyses, due to the broadest scope of laboratory tests at the Gdansk-Wschod WWTP, it was selected as the so-called reference model (including independent variables that determine the best fit). This allowed the assessment of the possibility of simulating the wastewater quality at the Gdansk-Wschod WWTP, considering the limited range for measurement data. In parallel, the model for the Rzeszow WWTP was made based on the independent variables found in the previous case. The analysis results (empirical coefficients α, standard deviations and measures of matching the simulation results to measurements) are given in Table 3.

Based on the data in Table 2, it can be observed that the limit value for TN in the case of Rzeszow WWTP is lower than Gdansk-Wschod WWTP, while an inverse relation was found for the TP. For two and three dependent variables included in the models in parallel, the established limit values are conditioned by the predictive abilities of the models. The analysis of those values shown in Table 2 shows that the order of indicators magnitude is similar and the differences in the values of the analyzed wastewater quality indicators do not exceed 16%.

Due to the fact that only two WWTPs were adopted in the considerations, the possibilities of analyzing the impact of the catchment and meteorological characteristics on their values of the physical and geographical characteristics are limited. However, further study is required, which may constitute a basis to the generalization of the relationship given above. In Table 3, it can be stated that among the considered models, the greatest problem with identifying the wastewater quality indicators occurs in relation to TP. This statement applies primarily to the Gdansk-Wschod WWTP and for independent variables specified for the Rzeszow WWTP. The values of the measures matching the calculation related to TP are SPEC = 58.53%–69.90% (concerning 900 data obtained from measurements in 527 and 629 events, the calculations were consistent with the measurements) and SENS = 69.70%–80.61% (concerning 900 data obtained from measurements in 628 and 726 events), the calculations were consistent with the measurements). On the basis of Table 3, it was found that the values of the measures matching the simulation results are not lower than SENS = 69.09% and SPEC = 71.25%. This way, the results confirm the applicability of the new concept logistic models for forecasting wastewater quality indicators.

Table 2
Range of selected wastewater quality indicators adopted for determining logit models

Gdansk-Wschod WWTP		Rzeszow WWTP	
Model	Value	Model	Value
TN	80	TN	76
TP	9.4	TP	11.2
TN/TP	80/9.4	TN/TP	76/11.2
TN/TP/TSS	80/9.4/396	TN/TP/TSS	76/11.2/411

TN = TN_{lim}, TP = TP_{lim}, TSS = TSS_{lim}.

Table 3
List of determined empirical coefficients (α_i), and standard deviations in the logistic models for wastewater quality indicators prediction

	(Gd)wz			(Gd)nwz			Rz		
	Var.	α_i	Std. Dev.	Var.	α_i	Std. Dev.	Var.	α_i	Std. Dev.
TN	NH_4	0.196	0.028	NH_4	0.191	0.029	NH_4	0.27	0.018
	PO_4	0.287	0.107	TSS	0.006	0.002	TSS	0.004	0.007
	TSS	0.008	0.001	Inter.	–14.6	1.813	Inter.	–19.51	1.248
	Inter.	–16.76	2.04						
	SENS = 71.52%; SPEC = 77.9%			SENS = 69.70%; SPEC = 75.56%			SENS = 90.21%; SPEC = 72.97%		
TP	PO_4	0.544	0.116	COD	0.002	0.0009	COD	0.004	0.001
	TSS	0.008	0.001	NH_4	0.017	0.0029	NH_4	0.08	0.011
	Q	–0.00008	0.0000	TSS	0.004	0.002	TSS	0.011	0.001
	Inter.	–0.925	0.085	Inter.	–4.772	1.161	Inter.	–12.787	0.85
	SENS = 71.66%; SPEC = 71.25%			SENS = 71.51%; SPEC = 58.83%			SENS = 78.60%; SPEC = 78.81%		
TN, TP	NH_4	0.149	0.025	NH_4	0.147	0.025	NH_4	0.261	0.018
	PO_4	0.432	0.113	TSS	0.007	0.002	TSS	0.005	0.001
	TSS	0.009	0.001	Inter.	–12.46	1.63	Inter.	–19.29	1.231
	Inter.	–15.879	1.975						
	SENS = 75.08%; SPEC = 78.56%			SENS = 72.53%; SPEC = 67.90%			SENS = 79.60%; SPEC = 80.81%		
TN, TP, TSS	COD	0.005	0.001	COD	0.005	0.001	COD	0.008	0.007
	NH_4	0.128	0.026	NH_4	0.093	0.022	NH_4	0.15	0.014
	pH	–3.813	0.026	Inter.	–10.07	1.42	Inter.	–16.642	1.076
	Inter.	17.146	7.70						
	SENS = 78.46%; SPEC = 69.09%			SENS = 80.61%; SPEC = 63.67%			SENS = 90.08%; SPEC = 69.92%		

TN(Gd)wz – a logit model for modeling TN at the inflow to the Gdansk-Wschod WWTP; TN(Rz) – a logit model for modeling TN at the inflow to the Rzeszow WWTP; TN(Gd)nwz – a logit model for modeling TN at the inflow to the Gdansk-Wschod WWTP, for independent variables identical to the model for the Rzeszow WWTP; TNTP(Gd,Rz) – a logit model for modeling the value of TN > TN_{lim} and TP > TP_{lim} at the inflow of the Gdansk-Wschod and Rzeszow WWTPs; TNTPTSS(Gd,Rz) – a logit model for modeling the value of TN > TN_{lim} and TP > TP_{lim} and TSS > TSS_{lim} at the inflow to the Gdansk-Wschod and Rzeszow WWTPs.

3.2. Impact of selected independent variables on the content of TSS, TN and TP in the wastewater quality indicators of WWTPs influent

The results presented in Table 3 have indicated the non-linear relationships between the TN and NH_4–N. Total nitrogen includes all forms of nitrogen, that is, NH_4–N, nitrite–nitrogen (N–NO_2), nitrate–nitrogen (N–NO_3) and organic nitrogen (N_{org}). These relationships were also previously demonstrated by analyzing the data on WWTP influent in Jiangsu Province (China), obtaining a high correlation coefficient [39]. The simulation results also showed an influence of the COD and TSS values on TN. This is also confirmed by the analyses carried out by Kim et al. [40], using the 24-h measurements from a WWTP in Busan, South Korea to simulate TN. Zou et al. [39] showed a non-linear relationship TN = f(COD), analyzing the data from a Jiangsu WWTP obtaining R = 0.14. Additionally, the relationship of TP = f(COD) has been confirmed by Zou et al. [39] obtaining R = 0.32. Ebrahimi et al. [41], showed non-linear relationships TP = f(TSS), obtaining correlation coefficients at the level of $R \approx 0.22$. Ebrahimi et al. [41], when conducting analyses for a single treatment plant also showed a significant impact of BOD and TSS values on the simulation of TP results. Their reports find theoretical confirmation in models describing the kinetics of organic compound's transformation in wastewater systems. This dependence is local and is depended on the hydraulic conditions (filling, retention time), size of the wastewater network, etc [42]. Due to the fact that the BOD_5 measurement is time-consuming (5 d), this indicator has been omitted in the calculation models. The models in which the BOD_5 value is included have limited practical application. The results obtained in this way confirmed that identification of TP variability is possible to a limited extent (COD, N–NH_4, TSS), which simplifies the results of computational experiments. Ebrahimi et al. [41] showed the relationship TP = f(BOD, TSS, COD) and TP = f(BOD, TSS, TN). The results of analysis [43] based on the data from Solumstrad WWTP in Norway confirmed the effect of NH_4–N on TP by obtaining R^2 = 0.79. Despite the fact that they showed the relationship TP = f(Q, pH, NH_4–N), it should be borne in mind that the WWTP analyzed in Norway was much smaller (PE = 130,000) than the WWTP considered in work.

The wastewater quality can be shaped by fewer factors than in the treatment plants considered in this work. The results of research Ansari et al. [44] performed for the WWTP in Kuala Lumpur with a size of 100,000 PE showed that the forecast of N compounds in the inflow is possible using modified neural networks (ANFIS + GA or PO) based on the inflow and precipitation depth. However, such relationships are usually possible for facilities of a limited size (PE value), where the processes of providing the wastewater quality are conditioned by the dilution of the wastewater. This was confirmed by the results of [45,46], who demonstrated the relationship COD = $f(Q)$ by analyzing several treatment plants in Germany and Belgium (167 sites). While analyzing the determined values of coefficients in logit models, it can be stated that the characteristics of the catchment area (length of the wastewater network, the surface of the drained area, the load of the WWTP – PE) have an impact on the quality of wastewater flowing into the wastewater treatment plant. For the Rzeszow WWTP in this study, the obtained coefficients in logit models for forecasting individual indicators of wastewater quality were higher than for the Gdansk-Wschod WWTP facility. It may indicate that in the case of smaller facilities (PE = 400,000 – Rzeszow WWTP), the modeled TN and TP values show lesser sensitivity to the values of selected indicators (COD, TSS). For the TN and TP forecast models applied simultaneously, an inverse relationship was found – for the Gdansk-Wschod WWTP, the α_i values in logit models were higher than for Rzeszow. The impact of the size of the treatment plant (facility load using PE) on the model parameters for the COD, total Kjeldahl nitrogen (TKN), TP forecasts was demonstrated by Langergraber et al. [42], by adopting second-order Fourier models in their study. The impact of geographical location and thus the size of the facility on the relationship between TN = f(TP) was demonstrated by the statistical analyses pertaining to the operation of wastewater treatment plants in China located in 31 provinces, as described in the work of Sun et al. [47]. Their dependencies showed increased content of TN and TP for the objects in the north (Beijing) and North-East of China (Gansu, Xinjang etc.). Similar relationships were found in relation to the relationship TN = $f(NH_4–N)$. The results of the analysis conducted by Sun et al. [47] for the relationship TN = f(COD) showed increased values for the WWTPs located in the South-East (Gansu, Xinjang, Qinghai etc.). However, the analyses conducted did not take into account the size of the treatment plant. In the case of the models for simultaneously identifying the range of variability of several wastewater quality indicators, similar analyses have not been conducted so far. This is an interesting aspect, which enables to study the impact of interactions between the selected wastewater quality indicators (TN and TP, TN as well as TP and TSS, etc.). This gives an opportunity to generalize the results obtained and to explain the variability of wastewater quality at the inflow to facilities in a global perspective (several indicators at the same time) rather than locally. This is important from the point of view of optimizing the operation of WWTP (modeling energy consumption, wastewater quality, reducing greenhouse gas emissions), because quality identification gives the opportunity to control the operation of the facility and the selection of optimal settingsThe solution adopted in the work is a significant simplification and the algorithm developed (Fig. 2) allows the identification of the values of individual wastewater quality indicators. Considering the above-mentioned remarks, it can be stated that many authors showed correlations between selected quality indicators of influent wastewater, however, due to the models used, they did not provide analytical relationships. Therefore, they did not indicate how the content of TN is affected quantitatively by NH_4–N and COD.

3.3. Quantitative analysis of the impact of selected independent variables on the wastewater quality

The interaction between TN, TP and TSS was checked, and the impact on the variability of the probability of a single value and simultaneously the values of two and three wastewater quality indicators was determined (Fig. 3). On the basis of the set coefficients for logit models (Table 3) and the equation describing the sensitivity coefficient, the effect of independent variable values (NH_4–N) and probability (p) on the variability of S_{xi} values was determined (Fig. 4).

Based on Fig. 3, it can be observed that both the NH_4–N and TSS values have a significant impact on the probability of exceeding TN. For example, for TSS = 500 mg/L, the increase in the NH_4–N value from 50 to 55 mg/L leads to an increase in the $p(TN_{lim})$ value from 0.28 to 0.55 for the Gdansk-Wschod case. In the case of Rzeszow, an identical change in the value of the NH_4–N results in an increase in $p(TN_{lim}$ = 76 mg/L) from 0.028 to 0.083. On the other hand, increasing the TSS value from 250 to 500 mg/L for NH_4–N leads to an increase in the value of $p(TN_{lim})$ from 0.048 to 0.280. For the Rzeszow WWTP, an identical increase in the TSS value leads to an increase in the probability of exceeding TN_{lim} from 0.001 to 0.0280. Additionally, it was found that for TN > 63 mg/L the value of $p(TN_{lim})_{TSS = 500}$ is greater than $p(TN_{lim})_{TSS = 1000}$. In turn, for TN > 64 mg/L the value of $p(TN_{lim})_{TSS = 250}$ is greater than $p(TN_{lim})_{TSS = 500}$. The results indicate a strongly differentiated impact of both TSS and NH_4–N on the probability of exceeding TN_{lim}, depending on the wastewater influent. According to Fig. 4, describing the probability of exceeding TN_{lim} and TP_{lim}, it can be stated that for Rzeszow WWTP, the increase in the value of independent variables (TSS) from 0.05 to 0.50 has a lesser impact than the change in the TSS, PO_4–P for the Gdansk-Wschod WWTP.

While analyzing the determined curves (Fig. 4) for the average values (0.50 percentile) of independent variables (Table 3) in the TN = 80 mg/L and TP = 10.1 mg/L model (Gd) with the probability value $p(TN_{lim}$ and $TP_{lim})$ are greater than those obtained with the simplified model, that is, $p = f(N–NH_4$, TSS). On the other hand, the inverse relationship was obtained when 0.05 percentiles were substituted in the designated models, that is, TNTP(Gd)wz and TNTP(Gd)nw.

3.4. Interaction analysis and the relationship between TN, TP, TSS in logit models

On the basis of the obtained logit models and determined empirical coefficients (Table 2), the curves p(TN,TP) = $f(p$(TN), NH_4–N, PO_4–P) – for the WWTP Gdansk-Wschod (Fig. 5a) and Rzeszow WWTP (Fig. 5b) were drawn. The values of p(TN) and p(TN,TP) were determined for the assumed values TSS = 285–855 mg/L and PO_4–P = 6.94–8.88 mg/L and for $N–NH_4$ = 20–84 mg/L. For example, subsequent p(TN) values for TSS = 285 mg/L and PO_4–P = 6.94 mg/L were calculated for increasing

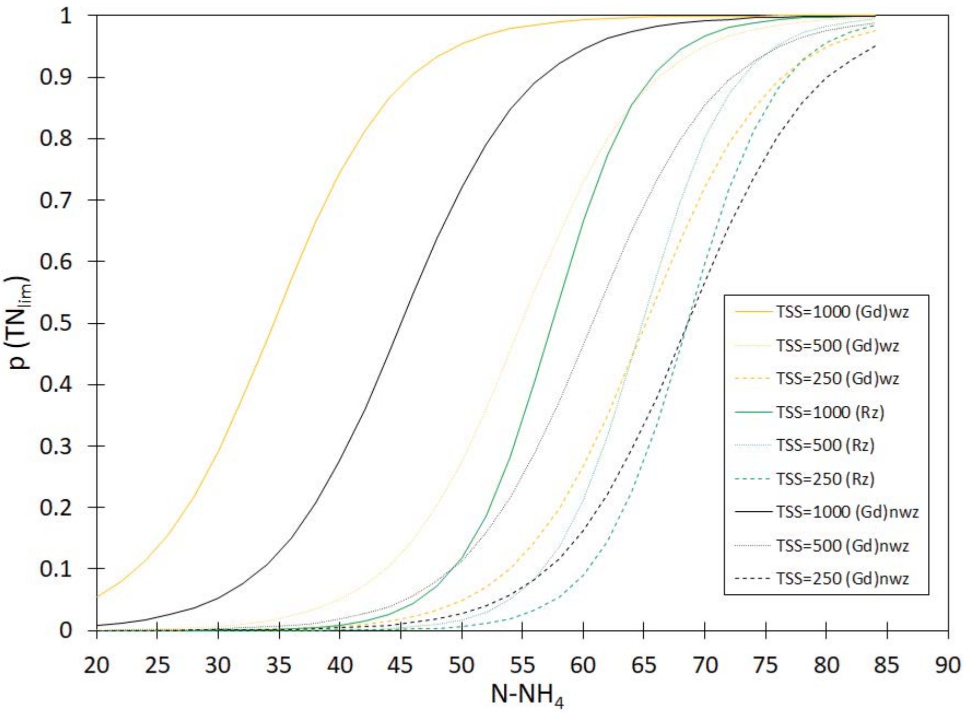

Fig. 3. Effect of NH_4–N and TSS for Gdansk-Wschod and Rzeszow WWTPs on the probability of exceeding TN_{lim}.

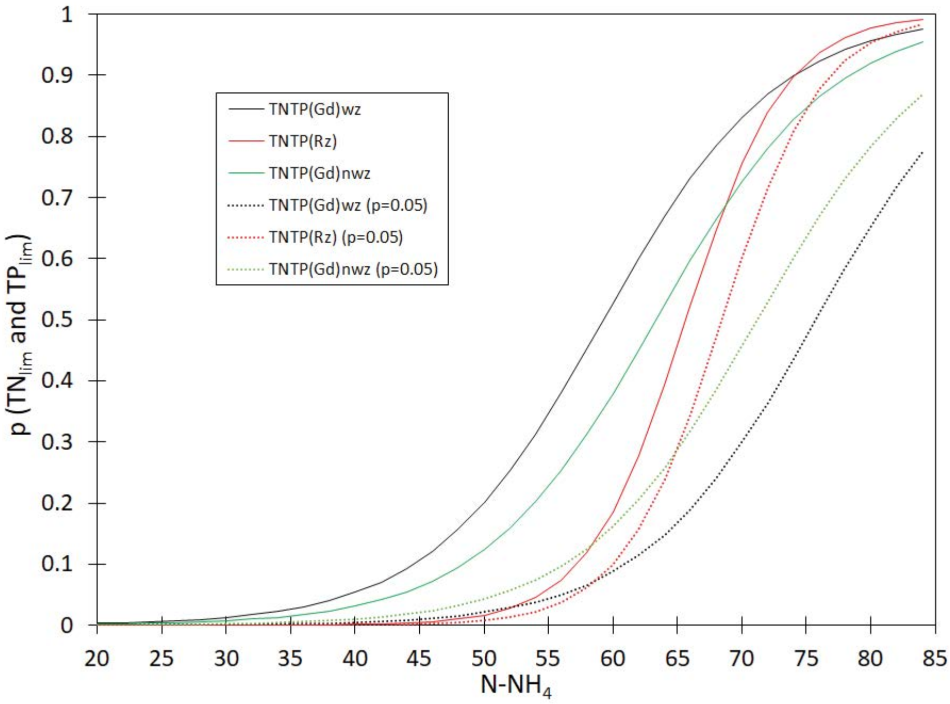

Fig. 4. Impact of the NH_4–N value for the Gdansk-Wschod and Rzeszow plants on the probability of exceeding TN_{lim} and TP_{lim} for average values and 0.05 percentiles of independent variables.

NH_4–N values from 20 mg/L with the constant step equal NH_4–N = 2 mg/L. The *p*-values (TN, TP) were determined identically. For changing NH_4–N values, variability of *p*(TN) and *p*(TN,TP) values was obtained – Fig. 5a and b, which allowed determining the curves *p*(TN,TP) = *f*(*p*(TN), NH_4–N, PO_4–P). On the basis of the above-mentioned assumptions, the curves describing the relationship *p*(TN, TP, TSS) = *f*(*p*(TP), NH_4–N, COD) and *p*(TN) = *f*(*p*(TP), NH_4–N) were determined, which are presented in Fig. 6 for Rzeszow WWTP.

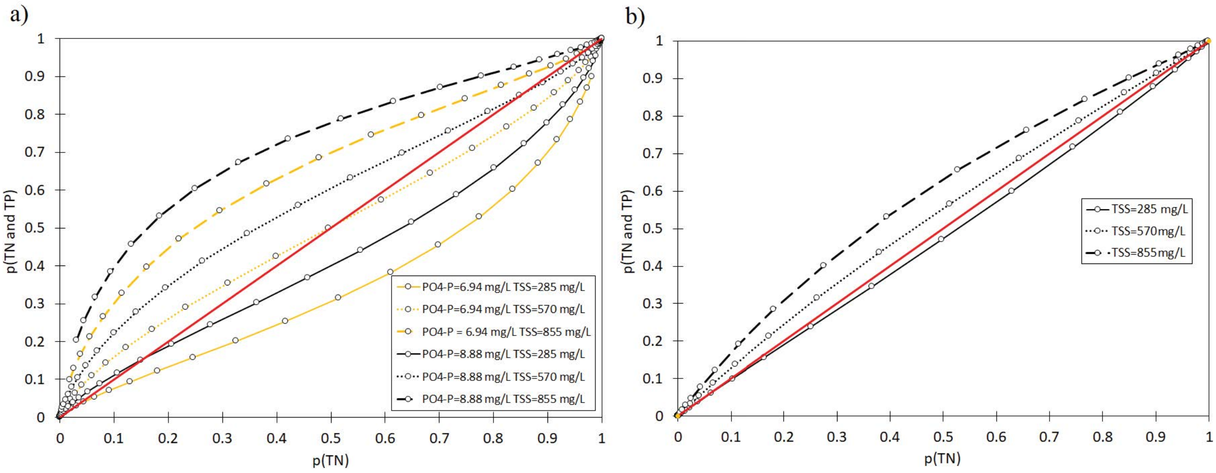

Fig. 5. Relationship of p(TN,TP) and p(TN) for WWTPs: (a) Gdansk-Wschod and (b) Rzeszow.

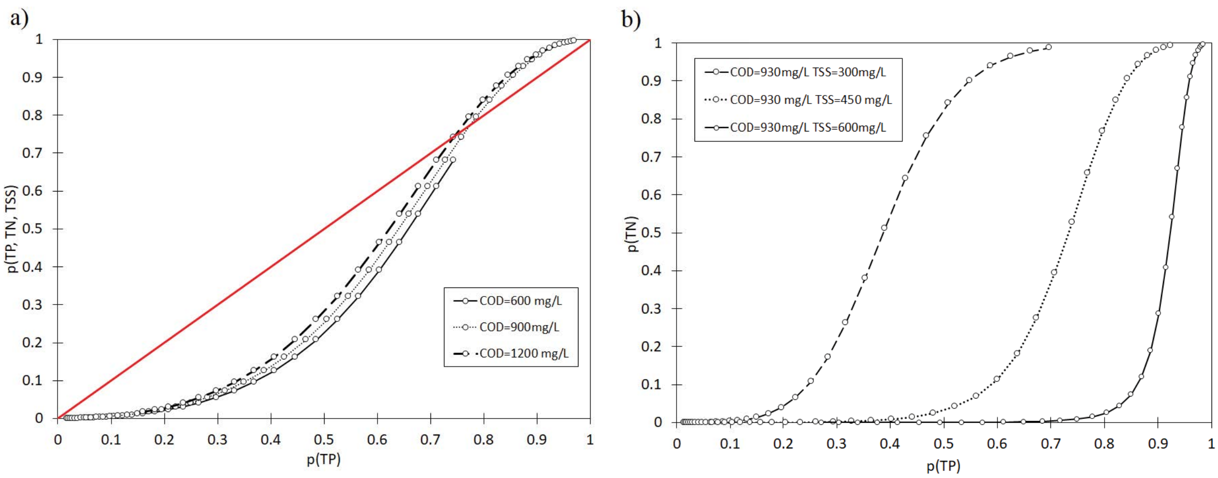

Fig. 6. Impact of p(TP) and COD values on: (a) p(TP,TN,TSS) and (b) p(TN) for the Rzeszow WWTP.

On the basis of Fig. 5, for the Rzeszow WWTP, the relationship of p(TN,TP) and p(TN) is almost similar for a linear one, provided that the TSS value varies within the appropriate range – in such a way as to obtain the compatibility of p data (TN) = f(TN, TP). An increase in the TSS value for the Gdansk-Wschod WWTP of p(TN) = f(NH_4–N, TSS) indicates increasingly stronger nonlinearities.

Strongly linear relationships were found between p(TN) and p(TN,TP) in which the increase in TSS for p(TN) = f(NH_4–N) has a significant impact on the relationship p(TN) = f(p(TN,TP)). Assuming PO_4–P = 6.94 mg/L and TSS = 285 mg/L for p(TN) = 0.18 and p(TN) = 0.42, the values of p(TN,TP) are 0.12 and 0.25, respectively. In turn, for PO_4–P = 6.94 mg/L and TSS = 855 mg/L, the values of p(TN,TP) are 0.41 and 0.64. Only for the mean values TSS = 570 mg/L and PO_4–P = 6.94 mg/L and the changing NH_4–N value – the relation p(TN) = f(p(TN,TP)) is close to linear, which confirms the obtained course curves. The non-linear relationships, as well as strong interactions between dependent variables (TN, TP, TSS), are also confirmed by the curves in Fig. 6a. The complex relationships between TN and TP are confirmed in Fig. 6b. The obtained curves indicate that COD and TSS significantly impact the relationship between the analyzed wastewater quality indicators. The lower the concentration in the TSS wastewater, the stronger the relationship between TN and TP.

The results obtained above are of significant importance from the point of view of building regression models for forecasting wastewater quality. They indicate the need to create several models in parallel to simulate the selected wastewater quality indicators. This approach is confirmed in the analyses where the first step in building the regression model was the analysis aimed at separating similar data groups. For this purpose, clustering methods were used, such as HCA, method K-NN. This is an important stage in creating models because it allows identifying local relationships between the selected indicators of wastewater quality. The methods listed above are complex algorithms that require numerical implementation.

Conversely, the methodology proposed in this study is much simpler. Secondly, the dependencies obtained are of analytical nature, expressed by means of empirical models, with the help of which the impact of the examined independent variables on the simulation results can be analyzed. Using such methods as HCA, K-NN, neural networks, etc., the analysis of the impact of selected wastewater quality indicators allocation to separate data classes that are similar to each other is limited.

3.5. Sensitivity analysis

A sensitivity analysis was performed in order to supplement the analysis results. Therefore, based on the developed logit models (Table 3) and Eq. (4), its variability was determined on the example for TN value in the Gdansk-Wschod plant data. The range of NH_4–N = 40–60 mg/L was used for calculations, while the values of independent variables included in the equation were adopted in the range PO_4–P = 4.5, 10.0, and 12.0 mg/L and TSS = 300, 600, and 750 mg/L, respectively. The results of the probability of exceeding TN_{lim} and S_{xi} are presented in Fig. 7.

On the basis of Fig. 7, it can be observed that the boundary values PO_4–P and TSS have a great impact on the sensitivity model. In most cases, *p*-value got higher with the value moderately increase of NH_4–N, and the model's sensitivity decreases. For example, an increase in the concentration of NH_4–N on the inflow to the treatment plant from 40 to 46 mg/L (PO_4–P = 10 mg/L and TSS = 600 mg/L), caused to decrease in the value of S_{xi} from 6.09 to 4.67 (approx. 24%). An identical change in the value of NH_4–N for PO_4–P = 4.5 mg/L and TSS = 300 mg/L leads to an increase in S_{xi} from 7.98 to 8.86.

On the basis of the determined logistic regression models and the derived sensitivity coefficient [Eq. (4)], the impact of the interaction between TN and TP as well as TN, TP and TSS on the relative change in the value of $S_{xi}(TN_{lim})$ was analyzed. For this purpose, set values of PO_4–P and TSS, the values of sensitivity coefficients were calculated considering two (TN and TP) and three (TN, TP and TSS) indicators as the basis for classification. The calculation results are presented in Fig. 8. The values of S_{xi}(TNTP) compared to S_{xi}(TN) depending on NH_4–N are smaller at Gdansk-Wschod WWTP by 25% (PO_4–P = 4.5 mg/L and TSS = 300 mg/L). With an increase in the NH_4–N value from 40 to 60 mg/L, it was found that the values of S_{xi}(TNTP) are smaller than S_{xi}(TN) by 12%. A similar relation, taking into account the variability of NH_4–N was also found for S_{xi}(TNTP) when P–PO_4 = 6.9 mg/L and TSS = 450 mg/L.

In the range of NH_4–N = 40 - 50 mg/L, it was found that the values S_{xi}(TNTP) and S_{xi}(TNTPTSS) for P–PO_4 = 10 mg/L and TSS = 600 mg/L and P–PO_4, respectively 6.9 mg/L and TSS = 450 mg/L are lower than S_{xi}(TN) by approx. 43%–51%. A further increase in NH_4–N caused a decrease in the relative difference between the values of S_{xi}(TNTP) and S_{xi}(TNTPTSS) and S_{xi}(TN). In the considered variants, it was found that an increase in the NH_4–N value of S_{xi}(TNTPTSS) for (P–PO_4 = 10 mg/L and TSS = 600 mg/L) is 2.8 times higher, compared to S_{xi}(TN).

4. Conclusions

A new concept of logistic regression model provided a suitable range of TN, TP, and TSS as suitable indicators for WWTPs influent quality in two evaluated plants. The developed logistic regression model approach in this study is flexible to use as a wide range of indicators, but machine learning methods limit the possibilities of analyzing the impact of selected independent variables on the results of

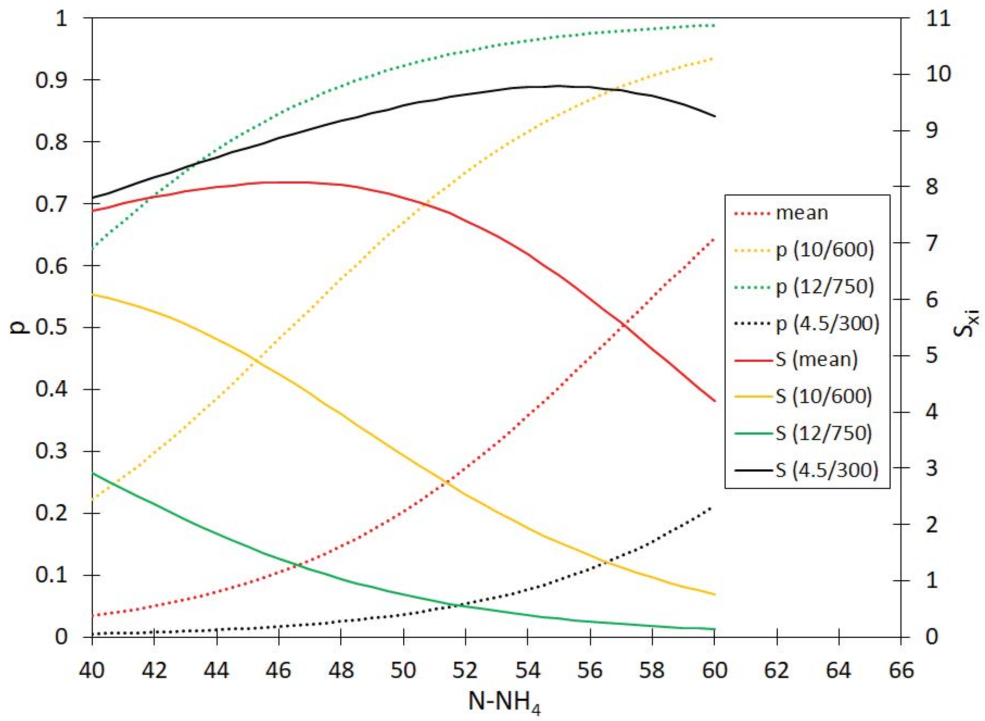

Fig. 7. Impact of the NH_4–N value for the Gdansk-Wschod plant on the probability of exceeding the TN_{lim} and values of the coefficients sensitivities (S_{xi}).

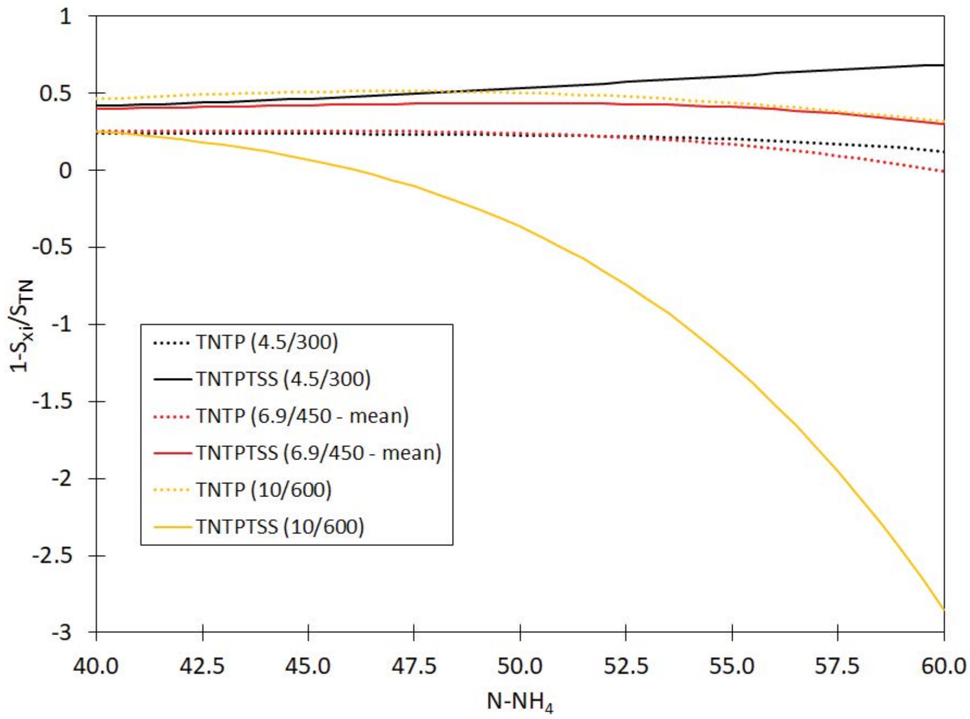

Fig. 8. Impact of NH_4–N values on the quotient $(1 - S_{xi})/S_{TN}$ for the TNTP(Gd)wz and TNTPTSS(Gd)wz models for the assumed values of (PO_4–P/TSS).

calculations. This method was found as a useful tool, and can be extended for prediction beyond the current case studies. Although the dependencies obtained in the paper are presented for only two WWTPs, the obtained results provide the opportunity for further research in the area and ultimately build universal models. In addition, the results of the sensitivity analysis showed the occurrence of strong nonlinearities between the values of TN, and TP as well as TN, TP, and TSS, which is confirmed by relative changes in the values of the calculated sensitivity coefficients.

Acknowledgments

This work was financially supported by the internal grant at Lublin University of Technology no. FD-20/IS-6/021 and by the National Science Centre as a result of the research project no. 2017/26/D/ST8/00967.

References

[1] F. Hernández-del-Olmo, E. Gaudioso, R. Dormido, N. Duro, Energy and environmental efficiency for the n-ammonia removal process in wastewater treatment plants by means of reinforcement learning, Energies, 9 (2016) 755, doi: 10.3390/en9090755.
[2] Z.J. Zhang, A. Kusiak, Y.H. Zeng, X.P. Wei, Modeling and optimization of a wastewater pumping system with data-mining methods, Appl. Energy, 164 (2016) 303–311.
[3] D. Torregrossa, J. Hansen, F. Hernández-Sancho, A. Cornelissen, G. Schutz, U. Leopold, A data-driven methodology to support pump performance analysis and energy efficiency optimization in wastewater treatment plants, Appl. Energy, 208 (2017) 1430–1440.
[4] J.-J. Zhu, L. Kang, P.R. Anderson, Predicting influent biochemical oxygen demand: balancing energy demand and risk management, Water Res., 128 (2018) 304–313.
[5] T.Y. Pai, P.Y. Yang, S.C. Wang, M.H. Lo, C.F. Chiang, J.L. Kuo, H.H. Chu, H.C. Su, L.F. Yu, H.C. Hu, Y.H. Chang, Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality, Appl. Math. Modell., 35 (2011) 3674–3684.
[6] H. Guo, K. Jeong, J. Lim, J. Jo, Y.M. Kim, J. Pyo Park, J.H. Kim, K.H. Cho, Prediction of effluent concentration in a wastewater treatment plant using machine learning models, J. Environ. Sci. (China), 32 (2015) 90–101.
[7] X. Liu, Y. Chen, H. Ge, P. Fazio, G. Chen, Numerical investigation for thermal performance of exterior walls of residential buildings with moisture transfer in hot summer and cold winter zone of China, Energy Build., 93 (2015) 259–268.
[8] K. Gibert, J. Izquierdo, M. Sànchez-Marrè, S.H. Hamilton, I. Rodríguez-Roda, G. Holmes, Which method to use? an assessment of data mining methods in Environmental Data Science, Environ. Modell. Software, 110 (2018) 3–27.
[9] F. Harrou, A. Dairi, Y. Sun, M. Senouci, Statistical monitoring of a wastewater treatment plant: a case study, J. Environ. Manage., 223 (2018) 807–814.
[10] B. Khalil, J. Adamowski, A. Abdin, A. Elsaadi, A statistical approach for the estimation of water quality characteristics of ungauged streams/watersheds under stationary conditions, J. Hydrol., 569 (2019) 106–116.
[11] J.J. Lee, C.S. Jang, C.W. Liu, C.P. Liang, S.W. Wang, Determining the probability of arsenic in groundwater using a parsimonious model, Environ. Sci. Technol., 43 (2009) 6662–6668.
[12] X. Wang, H. Ratnaweera, J.A. Holm, V. Olsbu, Statistical monitoring and dynamic simulation of a wastewater treatment plant: a combined approach to achieve model predictive control, J. Environ. Manage., 193 (2017) 1–7.
[13] A. Kusiak, X. Wei, Prediction of methane production in wastewater treatment facility: a data-mining approach, Ann. Oper. Res., 216 (2014) 71–81.
[14] A. Asadi, A. Verma, K. Yang, B. Mejabi, Wastewater treatment aeration process optimization: a data mining approach, J. Environ. Manage., 203 (2017) 630–639.

[15] R. Avila, B. Horn, E. Moriarty, R. Hodson, E. Moltchanova, Evaluating statistical model performance in water quality prediction, J. Environ. Manage., 206 (2018) 910–919.
[16] S. Yamijala, S.D. Guikema, K. Brumbelow, Statistical models for the analysis of water distribution system pipe break data, Reliab. Eng. Syst. Saf., 94 (2009) 282–293.
[17] Y. Kleiner, B. Rajani, Comparison of four models to rank failure likelihood of individual pipes, J. Hydroinf., 14 (2012) 659–681.
[18] A. Robles-Velasco, P. Cortés, J. Muñuzuri, L. Onieva, Prediction of pipe failures in water supply networks using logistic regression and support vector classification, Reliab. Eng. Syst. Saf., 196 (2020) 106754, doi: 10.1016/j.ress.2019.106754.
[19] W. Thoe, M. Gold, A. Griesbach, M. Grimmer, M.L. Taggart, A.B. Boehm, Predicting water quality at Santa Monica Beach: evaluation of five different models for public notification of unsafe swimming conditions, Water Res., 67 (2014) 105–117.
[20] T.K. Saha, S. Pal, Exploring physical wetland vulnerability of Atreyee river basin in India and Bangladesh using logistic regression and fuzzy logic approaches, Ecol. Indic., 98 (2019) 251–265.
[21] N.K.C. Twarakavi, J.J. Kaluarachchi, Aquifer vulnerability assessment to heavy metals using ordinal logistic regression, Ground Water, 43 (2005) 200–214.
[22] 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.
[23] 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) 143–154.
[24] A. Verma, X. Wei, A. Kusiak, Predicting the total suspended solids in wastewater: a data-mining approach, Eng. Appl. Artif. Intell., 26 (2013) 1366–1372.
[25] A. Mair, A.I. El-Kadi, Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA, J. Contam. Hydrol., 153 (2013) 1–23.
[26] M.C. Maniquiz, S. Lee, L.H. Kim, Multiple linear regression models of urban runoff pollutant load and event mean concentration considering rainfall variables, J. Environ. Sci., 22 (2010) 946–952.
[27] S. Dreiseitl, L. Ohno-Machado, Logistic regression and artificial neural network classification models: a methodology review, J. Biomed. Inf., 35 (2002) 352–359.
[28] W. Yoo, B.A. Ference, M.L. Cote, A. Schwartz, A comparison of logistic regression, logic regression, classification tree, and random forests to identify effective gene-gene and gene-environmental interactions, Int. J. Appl. Sci. Technol., 2 (2012) 268.
[29] Y. Lee, J.A. Nelder, Y. Pawitan, Generalized Linear Models with Random Effects, 2nd ed., Chapman and Hall, Boca Raton, 2018.
[30] A. Witteveen, G.F. Nane, I.M.H. Vliegen, S. Siesling, M.J. IJzerman, Comparison of logistic regression and bayesian networks for risk prediction of breast cancer recurrence, Med. Decis. Mak., 38 (2018) 822–833.
[31] B. Peeters, R. Dewil, I.Y. Smets, Improved process control of an industrial sludge centrifuge-dryer installation through binary logistic regression modeling of the fouling issues, J. Process Control, 22 (2012) 1387–1396.
[32] N. Deepnarain, S. Kumari, J. Ramjith, F.M. Swalaha, V. Tandoi, K. Pillay, F. Bux, A logistic model for the remediation of filamentous bulking in a biological nutrient removal wastewater treatment plant, Water Sci. Technol., 72 (2015) 391–405.
[33] B. Szeląg, K. Barbusiński, J. Studziński, Activated sludge process modelling using selected machine learning techniques, Desal. Water Treat., 117 (2018) 78–87.
[34] B. Szelag, R. Suligowski, J. Studziński, F. De Paola, Application of logistic regression to simulate the influence of rainfall genesis on storm overflow operations: a probabilistic approach, Hydrol. Earth Syst. Sci., 24 (2020) 595–614.
[35] Y.P. Lin, B.Y. Cheng, H.J. Chu, T.K. Chang, H.L. Yu, Assessing how heavy metal pollution and human activity are related by using logistic regression and kriging methods, Geoderma, 163 (2011) 275–282.
[36] B. Petersen, K. Gernaey, M. Henze, P.A. Vanrolleghem, Evaluation of an ASM1 model calibration procedure on a municipal-industrial wastewater treatment plant, J. Hydroinf., 4 (2002) 15–38.
[37] F.E. Harrell, Regression Modeling Strategies, Springer International Publishing, Cham, 2015.
[38] B. Szeląg, J. Drewnowski, G. Łagód, D. Majerek, E. Dacewicz, F. Fatone, Soft sensor application in identification of the activated sludge bulking considering the technological and economical aspects of smart systems functioning, Sensors (Switzerland), 20 (2020) 1941, doi: 10.3390/s20071941.
[39] L. Zou, H. Li, S. Wang, K. Zheng, Y. Wang, G. Du, J. Li, Characteristic and correlation analysis of influent and energy consumption of wastewater treatment plants in Taihu Basin, Front. Environ. Sci. Eng., 13 (2019) 83.
[40] M. Kim, Y. Kim, H. Kim, W. Piao, C. Kim, Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant, Front. Environ. Sci. Eng., 10 (2016) 299–310.
[41] 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.
[42] G. Langergraber, J. Alex, N. Weissenbacher, D. Woerner, M. Ahnert, T. Frehmann, N. Halft, L. Hobus, M. Plattes, V. Spering, S. Winkler, Generation of diurnal variation for influent data for dynamic simulation, Water Sci. Technol., 57 (2008) 1483–1486.
[43] X. Wang, K. Kvaal, H. Ratnaweera, Characterization of influent wastewater with periodic variation and snow melting effect in cold climate area, Comput. Chem. Eng., 106 (2017) 202–211.
[44] M. Ansari, F. Othman, A. El-Shafie, Optimized fuzzy inference system to enhance prediction accuracy for influent characteristics of a sewage treatment plant, Sci. Total Environ., 722 (2020) 137878, doi: 10.1016/j.scitotenv.2020.137878.
[45] 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.
[46] D. Rousseau, F. Verdonck, O. Moerman, R. Carrette, C. Thoeye, J. Meirlaen, P.A. Vanrolleghem, Development of a risk assessment based technique for design/retrofitting of WWTPs, Water Sci. Technol., 43 (2001) 287–294.
[47] Y. Sun, Z. Chen, G. Wu, Q. Wu, F. Zhang, Z. Niu, H.Y. Hu, Characteristics of water quality of municipal wastewater treatment plants in China: Implications for resources utilization and management, J. Cleaner Prod., 131 (2016) 1–9.