

Lidia Bartkiewicz, Bartosz Szelaąg, Jan Studziński

## Ocena wpływu zmiennych wejściowych oraz struktury modelu sztucznej sieci neuronowej na prognozowanie dopływu ścieków komunalnych do oczyszczalni

W Polsce od wielu lat są prowadzone badania nad fizykalnymi modelami oczyszczalni ścieków, opisującymi jednoznacznie zależność między ilością i jakością dopływających ścieków a parametrami działania obiektów technologicznych oczyszczalni. Zależności te dają możliwość fizykalnej interpretacji zmiennych objaśniających [1–3], jednak ze względu na liczne interakcje występujące między parametrami modelu fizykalnego, a przede wszystkim ze względu na ich znaczą liczbę, występują poważne problemy z kalibracją takich modeli, co jest przyczyną ich ograniczonego zastosowania. W takim przypadku do modelowania parametrów oczyszczalni ścieków oraz prognozowania ilości i jakości ścieków dopływających do oczyszczalni można stosować modele tzw. czarnej skrzynki (modele parametryczne). W modelach tych, na podstawie danych wejściowych i wyjściowych, struktura modelu neuronowego jest zakładana arbitralnie i testowana po etapie uczenia, bez znajomości fizyki rozpatrywanego zjawiska oraz bez fizykalnej interpretacji parametrów modelu.

Z doniesień naukowych wynika, że do modelowania ilości i jakości ścieków za pomocą modeli parametrycznych dotychczas były stosowane różne rodzaje modeli w postaci sztucznych sieci neuronowych, falek, zbiorów rozmytych, lasów losowych czy szeregów czasowych, w tym ostatnim przypadku z użyciem takich metod, jak metoda najmniejszej sumy kwadratów Kalmana, metoda uogólnionej sumy kwadratów Clarke'a czy metoda największej wiarygodności [4–12]. Na etapie opracowywania modeli fizykalnych czy modeli parametrycznych do prognozowania ilości i jakości ścieków na dopływie do oczyszczalni wykonywane są analizy mające na celu optymalizację liczby zmiennych objaśniających rozpatrywane zjawisko. Efektem tych analiz jest identyfikacja zmiennych wejściowych istotnie wpływających na wynik prognozy oraz eliminacja zmiennych, których wpływ jest pomijalny. W ramach tych analiz określa się ponadto wpływ struktury opracowywanego modelu matematycznego na dokładność predykcji w celu uniknięcia jego nadmiernej rozbudowy, która zwykle prowadzi do wydłużenia czasu obliczeń oraz może skutkować

uzyskiwaniem błędnych wyników symulacji, co w przypadku oczyszczalni ścieków może być przyczyną podejmowania błędnych decyzji operatorskich.

W rozważaniach praktycznych optymalizacja zarówno liczby zmiennych wejściowych, jak i samej struktury modelu jest bardzo istotna. Pozwala bowiem znacznie ograniczyć liczbę zmiennych wejściowych oraz koszty prowadzonych badań. Do tego celu wykorzystuje się powszechnie współczynnik korelacji [13] i drzewa klasyfikacyjne, pozwalające zidentyfikować zmienne mające największy wpływ na analizowane zjawisko. Ponadto do oceny wpływu danych wejściowych czy też struktury modelu na wynik obliczeń można wykorzystać metody wariacyjne [14, 15]. Ich zastosowanie wymaga jednak zastosowania odpowiednich algorytmów numerycznych i znacznych nakładów obliczeniowych, co w rozważaniach inżynierskich stwarza liczne problemy z ich wykorzystaniem.

W niniejszym artykule przedstawiono propozycję zastosowania modelu regresji logistycznej do oceny wpływu zmiennych objaśniających na jakość modelu. Uwzględniono takie zmienne, jak wysokość opadów atmosferycznych i dobowy dopływ ścieków do oczyszczalni, odpowiednio opóźnione w stosunku do wartości prognozowanej. Zbadano także wpływ parametrów struktury rozpatrywanego modelu na dokładność predykcji tworzonych modeli matematycznych. W analizach wykorzystano trzyletnie wyniki pomiarów wysokości opadów atmosferycznych oraz dobowych ilości ścieków dopływających do miejskich oczyszczalni w Rzeszowie i Kielcach.

### Obszar badań

Rzeszów jest aglomeracją miejsko-przemysłową zamieszkaną przez ponad 170 tys. osób. Miasto ma system kanalizacji rozdzielczej, jedynie w części Starego Miasta około 15% stanowi sieć ogólnospławna. Poziom wód gruntowych na terenie miasta silnie zależy od stanów wody w Wisłoku i jego dopływach, zbierających wody opadowe z Pogorza Dynowskiego. Ścieki komunalne odprowadzane są do oczyszczalni o przepustowości nominalnej równej 62 500 m<sup>3</sup>/d.

Kielce są głównym ośrodkiem przemysłowo-handlowym województwa świętokrzyskiego, zamieszkałym przez około 200 tys. osób. W granicach zlewni miejskiego systemu kanalizacyjnego znajduje się około 83% zabudowanych terenów miasta, z których ścieki odprowadzane są do oczyszczalni o nominalnej przepustowości równej 72 000 m<sup>3</sup>/d, zlokalizowanej w odległości 10 km od miasta.

Dr inż. L. Bartkiewicz: Politechnika Świętokrzyska w Kielcach, Wydział Inżynierii Środowiska, Geomatyki i Energetyki, Katedra Technologii Wody i Ścieków, al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce

Dr inż. B. Szelaąg: Politechnika Świętokrzyska w Kielcach, Wydział Inżynierii Środowiska, Geomatyki i Energetyki, Katedra Geotechniki, Geomatyki i Gospodarki Odpadami, al. Tysiąclecia Państwa Polskiego 7, 25-314 Kielce, [bszelag@tu.kielce.pl](mailto:bszelag@tu.kielce.pl)

Dr hab. inż. J. Studziński: Instytut Badań Systemowych Polskiej Akademii Nauk, Centrum Zastosowań Informatyki w Inżynierii Środowiska, ul. Newelska 6, 01-447 Warszawa, [jan.studzinski@ibspan.waw.pl](mailto:jan.studzinski@ibspan.waw.pl)

## Metody

W pierwszej kolejności wyznaczono zakres zmienności (wartość średnia, minimalna i maksymalna) danych wejściowych (całkowita wysokość opadów deszczu, dobowo ilość ścieków), które stanowiły podstawę badanych modeli parametrycznych. Do modelowania dobowej ilości ścieków dopływających do oczyszczalni wykorzystano trójwarstwowe sieci neuronowe typu MLP (multi layer perceptron). Mając na uwadze brak konkretnych wytycznych dotyczących wyboru optymalnej architektury do identyfikacji najlepszych sieci neuronowych prognozujących dopływ ścieków do oczyszczalni, wykorzystano opcję Automatycznego Projektanta w programie Statistica. W badaniach symulacyjnych w neuronach warstwy ukrytej i warstwy wyjściowej zastosowano następujące funkcje aktywacji – liniową, hiperboliczną, sinusoidalną, tangensa hiperbolicznego, wykładniczą, eksponentialną i logistyczną. Uczenie sieci neuronowych przeprowadzono metodą Broydena-Flechera-Goldfarba-Shanno [16]. Optymalną strukturę modelu sieci neuronowej, najlepiej prognozującego dopływ ścieków do oczyszczalni, określono na etapie analiz wykonanych z użyciem stu różnych struktur sieci neuronowych. Aby uzyskać najlepszą prognozę, w prowadzonych analizach badano również wpływ na wyniki modelowania opóźnień, przyjmowanych w przypadku dobowej wysokości opadu deszczu i dobowego dopływu ścieków komunalnych do oczyszczalni. W celu przeprowadzenia prawidłowego procesu uczenia, a następnie właściwej oceny działania otrzymanego modelu, dokonano podziału danych na trzy zbiory – uczący (50% danych), testowy (25%) i walidacyjny (25%). Przed przystąpieniem do obliczeń dane wejściowe i wyjściowe poddano normalizacji za pomocą przekształcenia w postaci [16]:

$$\bar{A}_i = \frac{A_i - A_{\min}}{A_{\max} - A_{\min}} \quad (1)$$

w której:

$\bar{A}_i$  – znormalizowana wartość  $i$ -tego elementu zbioru  $A$  metodą min-maks

$A_i$  – wartość  $i$ -tego elementu zbioru  $A$  zarejestrowana w czasie pomiarów

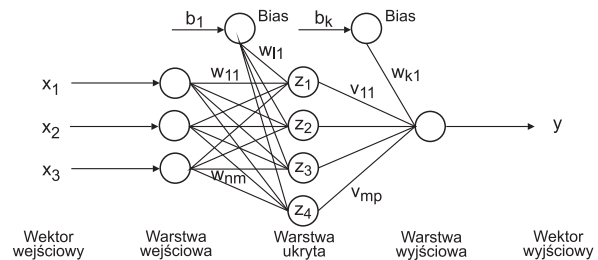
$A_{\max}$  – wartość maksymalna pojedynczego elementu w zbiorze  $A$

$A_{\min}$  – wartość minimalna pojedynczego elementu w zbiorze  $A$

Sztuczne sieci neuronowe (SSN) mają bardzo szerokie zastosowanie, ponieważ mogą być wykorzystane do modelowania różnych zjawisk. Wykazały to liczne badania potwierdzające możliwość wykorzystania sieci neuronowych do symulacji procesów liniowych i nieliniowych, optymalizacji, klasyfikacji i sterowania [16–18]. Opracowano wiele różnych struktur sieci neuronowych dopasowanych do różnych modelowanych zjawisk. Jedną z najpowszechniej stosowanych struktur jest sieć wielowarstwowa jednokierunkowa, znana pod nazwą perceptron wielowarstwowy (MLP). Wyniki badań potwierdzają, że sieć neuronowa już tylko z trzema warstwami jest w stanie z dużą dokładnością rozwiązać większość zadań występujących w praktyce. W sieciach typu MLP (rys. 1) sygnały wejściowe ( $x_k$ ) są przemnażane przez wartości wag ( $w_{ik}$ ), a następnie przekazywane do neuronów warstwy ukrytej, w której w poszczególnych neuronach odbywa się sumowanie zgodnie z zależnością:

$$z_i = \sum_k^m x_k w_{ik} - b_i, \quad i = 1, \dots, n \quad (2)$$

w której  $m$  oraz  $n$  oznaczają liczby neuronów w poprzedzającej i aktualnej warstwie sieci, zaś  $b_i$  oznacza próg (tzw. bias). Uzyskane sumy ( $z_i$ ) są poddawane transformacji za pomocą liniowej lub nieliniowej funkcji aktywacji ( $f$ ) i są przekazywane do neuronów wyjściowych. W warstwie wyjściowej neuron wykonuje identyczne zadania, jak w warstwie ukrytej, a jeśli jest pozbawiony odpowiedniej funkcji aktywacji, odpowiada to przyjęciu liniowej funkcji regresji i może on wówczas tylko przybliżać zależności liniowe. Ustalenie optymalnych wartości wag ( $w_{ik}$ ) przy poszczególnych neuronach odbywa się w procesie uczenia sieci.



Rys. 1. Schemat wielowarstwowej sieci neuronowej  
Fig. 1. Flow diagram of a multilayer neural network

## Kryteria oceny modeli

Do oceny zdolności predykcyjnej modeli prognozujących dobowy dopływ ścieków komunalnych do oczyszczalni w Rzeszowie i Kielcach zastosowano następujące miary:

– średni bezwzględny błąd procentowy (MAPE):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{i,\text{obs}} - y_{i,\text{progn}}}{y_{i,\text{obs}}} \right| 100 \quad (3)$$

– średni błąd bezwzględny (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n \left| y_{i,\text{obs}} - y_{i,\text{progn}} \right| \quad (4)$$

– współczynnik korelacji (R):

$$R = \frac{\sum_{i=1}^n (y_{i,\text{obs}} - \bar{y}_{\text{obs}})(y_{i,\text{progn}} - \bar{y}_{\text{progn}})}{\sqrt{\sum_{i=1}^n (y_{i,\text{obs}} - \bar{y}_{\text{obs}})^2} \sqrt{\sum_{i=1}^n (y_{i,\text{progn}} - \bar{y}_{\text{progn}})^2}} \quad (5)$$

w których:

$n$  – liczebność zbioru danych

$y_{i,\text{obs}/i,\text{progn}}$  – zmierzona/prognozowana ilość ścieków dopływających do oczyszczalni,  $\text{m}^3/\text{d}$

$\bar{y}_{\text{obs}/\text{progn}}$  – średnia arytmetyczna zmierzonej/prognozowanej ilości ścieków dopływających do oczyszczalni,  $\text{m}^3/\text{d}$

## Regresja logistyczna

Model regresji logistycznej [19], zwany też dwumianowym modelem logitowym, jest najczęściej stosowany w analizie danych w postaci binarnej i może przedstawiać wartości prawdopodobieństwa wystąpienia lub niewystąpienia pewnego zdarzenia, ponieważ przyjmuje wartości w przedziale od 0 do 1. Model logitowy jest przede wszystkim powszechnie wykorzystywany zarówno w naukach ekonomicznych, społecznych i medycznych, jak również w inżynierii rzecznej [20–22], geotechnice [23],

geomorfologii [24, 25], ekologii [26] i mikrobiologii [27]. Na uwagę zasługuje fakt, że model ten nie był dotychczas stosowany do oceny wpływu parametrów zmiennych objaśniających i struktury modelu na jego zdolności predykcyjne. Model regresji logistycznej stanowi szczególnie przypadek uogólnionego modelu regresji liniowej, który można zapisać w postaci wzoru:

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_j X_j \quad (6)$$

w którym:

$g$  – funkcja wiążąca zależność średniej wartości zmiennej objaśnianej  $\mu = E(Y|X_1=x_1, X_2=x_2, X_j=x_j)$  z liniową kombinacją predyktorów

$\beta_0$  – wyraz wolny

$\beta_1, \beta_2, \dots, \beta_j$  – współczynniki regresji

$X_j$  – zmienne niezależne (dobowa wysokość opadów deszczu, dobowa ilość ścieków dopływająca do oczyszczalni, parametry struktury modelu matematycznego)

W dalszych analizach przyjęto, że zmienne niezależne mają charakter binarny i gdy są ujęte w modelu, wówczas przyjmują wartość równą 1, a gdy są pominięte – wartość 0. W badanym modelu regresji logistycznej funkcja wiążąca, nazywana logitem, ma postać:

$$g(p) = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \quad (7)$$

w której:

$$p = P(Y|X_1=x_1, X_2=x_2, \dots, X_j=x_j) = \frac{\exp\left(\beta_0 + \sum_{i=1}^j \beta_i x_i\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^j \beta_i x_i\right)} \quad (8)$$

Parametry modelu  $\beta_0, \beta_1, \dots, \beta_j$  oblicza się metodą największej wiarygodności, maksymalizując logarytm funkcji wiarygodności względem parametrów modelu za pomocą metody Newtona–Raphsona. Zaletą regresji logistycznej jest możliwość interpretacji parametrów funkcji  $\exp(\beta_i)$  i do tego celu wykorzystuje się pojęcie tzw. szansy, definiowanej jako iloraz prawdopodobieństwa wystąpienia danego zdarzenia ( $p$ ) i prawdopodobieństwa jego niewystąpienia ( $1-p$ ):

$$\frac{p}{1-p} = \theta(x_1, x_2, \dots, x_k) = \exp\left(\beta_0 + \sum_{i=1}^j \beta_i x_i\right) \quad (9)$$

Z kolei wpływ przyrostu wartości zmiennych niezależnych o wartość  $\Delta x_i$  ( $i=1, 2, \dots, j$ ) na szansę wystąpienia zjawiska można obliczyć, wyznaczając tzw. iloraz szans opisany zależnością:

$$OR = \frac{\theta(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_k + \Delta x_k)}{\theta(x_1, x_2, \dots, x_k)} = \exp\left(\sum_{i=1}^m \beta_i \Delta x_i\right) \quad (10)$$

Jeżeli  $x_i$  ( $i=1, 2, \dots, k$ ) jest zmienną zero-jedynkową, to  $\exp(\beta_i)$  jest równy ilorazowi szans w przypadku grupy, w której  $x_i=1$  oraz grupy, w której  $x_i=0$ , przy pozostałych zmiennych jednakowych. Jeżeli natomiast zmienna ta jest zmienną rzeczywistą, to iloraz szans  $\exp(\beta_i)$  mówi, jak zmieni się szansa, jeżeli zmienna  $x_i$  wzrośnie o  $\Delta x_i$ , przy ustalonych wartościach pozostałych zmiennych. W praktyce, im większa jest wartość OR, tym większa jest wrażliwość rozpatrywanej zmiennej niezależnej ( $x_i$ ) na binarną zmienną zależną.

Jakość modelu, tzn. stopień dopasowania wyników obliczeń do pomiarów w modelu regresji logistycznej, określono za pomocą współczynników korelacji  $R^2$  McFaddena i  $R^2$  Cox-Snella [28] wyrażonych następująco:

$$R_{McFadden}^2 = 1 - \frac{\ln L_{UR}}{\ln L_R} \quad (11)$$

$$R_{Cox-Snell}^2 = 1 - \exp\left(\frac{(-2 \ln L_{UR}) - (-2 \ln L_R)}{n}\right) \quad (12)$$

Poza stopniem dopasowania oceniano także zdolność predykcyjną modelu regresji logistycznej (trafność prognoz) na podstawie czułości (SENS), swoistości (SPEC) i błędu zliczeniowego ( $R_z^2$ ), opisanych zależnościami:

$$SENS = 100 \frac{TP}{TP + FN} \quad (13)$$

$$SPEC = 100 \frac{TN}{FP + TN} \quad (14)$$

$$R_z^2 = 100 \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

w których: TP, TN, FP, FN – wyniki klasyfikacji w oparciu o model regresji logistycznej (tab. 1).

Tabela 1. Przykładowa tabela klasyfikacyjna  
Table 1. Exemplary classification table

Klasyfikacja		Decyzje prognozowane	
		pozytywna	negatywna
Decyzje zaobserwowane	pozytywna	prawdziwie pozytywna (TP)	falszywie negatywna (FN)
	negatywna	falszywie pozytywna (FP)	prawdziwie negatywna (TN)

Kryterium oceny wpływu poszczególnych zmiennych objaśniających i struktury modelu na jego zdolności predykcyjne stanowiła wartość współczynnika korelacji ( $R$ ), obliczona w przypadku modeli wyznaczonych za pomocą sieci neuronowych. Jeżeli obliczona wartość współczynnika korelacji była większa od zadanej wartości  $R_{gr}$ , warunkującej zadowalające dopasowanie wyników obliczeń do pomiarów, to przyjmowano w modelu regresji logistycznej wartość równą 1, a w przeciwnym przypadku wartość 0. W badanym modelu logitowym wynikiem obliczeń nie był współczynnik korelacji  $R$ , lecz prawdopodobieństwo przewyższenia zadanej wartości  $R_{gr}$ . W niniejszych analizach przyjęto, że jeśli  $p(x_i) > 0,50$ , to wartość rozpatrywanego współczynnika  $R$  w przypadku kombinacji danych wejściowych oraz założonej struktury modelu sieci neuronowej będzie co najmniej równa  $R_{gr}$ .

## Wyniki obliczeń i dyskusja

Na podstawie wykonanych analiz stwierdzono, że w rozpatrywanym czasie (2005–2008) dobowy dopływ ścieków do oczyszczalni w Rzeszowie zmieniał się od 24 306 m<sup>3</sup>/d w okresach bezdeszczowych do 73 640 m<sup>3</sup>/d w czasie intensywnych deszczów, co wynikało ze zmiennej dobowej wysokości opadów w zakresie 0÷59,2 mm. W przypadku oczyszczalni ścieków w Kielcach, dobowy dopływ ścieków zmieniał się od 12 784 m<sup>3</sup>/d do 94 722 m<sup>3</sup>/d, przy zmiennej dobowej wysokości opadów w zakresie 0÷46,2 mm [29].

Dysponując wynikami dobowych pomiarów dopływu ścieków do oczyszczalni, wyznaczono szeregi czasowe wysokości opadów  $P(t)$  i dopływu ścieków  $Q(t)$  z jedno- i dwudobowym opóźnieniem, a następnie dane wejściowe i wyjściowe poddano normalizacji. Na podstawie spreparowanych danych pomiarowych wyznaczono modele sieci neuronowych dotyczące dopływu ścieków do oczyszczalni w Kielcach i Rzeszowie. Podczas obliczeń przeanalizowano 40 kombinacji zmiennych wejściowych, na podstawie których zbudowano trójwarstwowe sieci neuronowe typu MLP. W tabeli 2 zamieszczono statystyczne miary dopasowania (R, MAE, MAPE) modeli najlepiej prognozujących dobową ilość ścieków dopływających do obu oczyszczalni.

Na podstawie wykonanych obliczeń stwierdzono, że znacznie lepszymi zdolnościami prognozy dobowego dopływu ścieków do oczyszczalni charakteryzuje się model matematyczny opracowany w oparciu o wartości  $Q(t-1)$  niż model wykonany na podstawie pomierzonej dobowej wysokości opadów w poprzedniej dobie. W pierwszym przypadku, tj.  $Q(t-1)$ , wartości współczynnika korelacji oraz średniego błędu bezwzględnego i średniego bezwzględnego błędu procentowego w przypadku oczyszczalni w Kielcach wynosiły  $R=0,74$ ,  $MAE=3600\text{ m}^3/\text{d}$ ,  $MAPE=8,15\%$ , natomiast w przypadku oczyszczalni w Rzeszowie –  $R=0,74$ ,  $MAE=2786\text{ m}^3/\text{d}$ ,  $MAPE=7,20\%$ . W drugim przypadku, tj.  $P(t-1)$ , wartości parametrów dopasowania były odpowiednio równe  $R=0,42$ ,  $MAE=5675\text{ m}^3/\text{d}$ ,  $MAPE=13,19\%$  (oczyszczalnia w Kielcach) i oraz  $R=0,50$ ,  $MAE=4165\text{ m}^3/\text{d}$ ,  $MAPE=10,89\%$  (oczyszczalnia w Rzeszowie). Uwzględnienie w modelach wyznaczonych na podstawie jedynie pojedynczej zmiennej objaśniającej  $Q(t-1)$  lub  $P(t-1)$  dodatkowego sygnału

wejściowego w postaci  $Q(t-2)$ , bądź w drugim przypadku  $P(t-2)$ , prowadziło do nieznacznej poprawy zdolności predykcyjnych modeli, na co wskazują wyznaczone wartości R, MAE i MAPE. Uwzględnienie w modelu dodatkowego sygnału wejściowego  $Q(t-3)$  także nie miało znaczącego wpływu na wyniki uzyskanych obliczeń.

Analizując otrzymane wyniki można stwierdzić, że w przypadku obu miast modele wyznaczone na podstawie dwóch zmiennych  $Q(t-1)$  i  $P(t-1)$  charakteryzują się znacznie lepszymi zdolnościami prognozowania dobowej ilości ścieków dopływających do oczyszczalni niż modele bazujące tylko na jednej zmiennej. W przypadku oczyszczalni ścieków w Kielcach i Rzeszowie uwzględnienie dodatkowej zmiennej  $P(t-1)$  w modelu wyznaczonym na podstawie  $Q(t-1)$  prowadziło do zmniejszenia wartości MAE odpowiednio o około 2% i 3%. Ponadto ujęcie nowej zmiennej  $P(t-1)$  w modelach otrzymanych na podstawie  $Q(t-1)$  i  $Q(t-2)$ , a także na podstawie  $Q(t-1)$ ,  $Q(t-2)$  i  $Q(t-3)$ , prowadziło do znacznej poprawy uzyskanych wyników, na co wskazują wartości R, MAE i MAPE. Z przeprowadzonych analiz wynika, że w przypadku obu oczyszczalni najlepszymi zdolnościami predykcyjnymi dobowego dopływu ścieków charakteryzowały się modele wyznaczone na podstawie pomiarów wysokości opadów z dwóch poprzednich dób oraz na podstawie pomiarów dopływów ścieków z trzech poprzednich dób. Przy takiej strukturze modelu, obliczone wartości współczynnika korelacji oraz średniego błędu bezwzględnego i średniego bezwzględnego błędu procentowego wynosiły odpowiednio  $R=0,83$ ,  $MAE=2482\text{ m}^3/\text{d}$ ,  $MAPE=6,47\%$  (oczyszczalnia w Rzeszowie) oraz  $R=0,81$ ,  $MAE=3333\text{ m}^3/\text{d}$ ,  $MAPE=8,10\%$  (oczyszczalnia w Kielcach).

Tabela 2. Statystyczne miary dopasowania sieci SSN prognozujących dobowy dopływ ścieków do oczyszczalni w Kielcach i Rzeszowie

Table 2. Statistical adjustment measures for ANN models forecasting the daily sewage inflow into the treatment plant in Kielce and Rzeszow

Dane	Uczenie			Test – walidacja		
	R	MAE $\text{m}^3/\text{d}$	MAPE %	R	MAE $\text{m}^3/\text{d}$	MAPE %
oczyszczalnia ścieków w Kielcach						
Q(t-1)	0,73	3331	7,27	0,74	3600	8,15
Q(t-1), Q(t-2)	0,72	3209	7,34	0,75	3569	8,55
Q(t-1), Q(t-2), Q(t-3)	0,73	3320	7,86	0,76	3583	8,58
P(t-1)	0,30	5389	12,69	0,42	5675	13,19
P(t-1), P(t-2)	0,35	5249	12,33	0,49	5549	12,86
P(t-1), Q(t-1)	0,74	3141	7,18	0,80	3507	8,41
P(t-1), Q(t-1), Q(t-2)		3145	7,20	0,79	3500	8,53
Q(t-1), Q(t-2), Q(t-3), P(t-1)	0,76	2973	6,77	0,77	3485	8,54
Q(t-1), Q(t-2), Q(t-3), P(t-1), P(t-2)	0,75	3009	6,86	0,81	3333	8,10
oczyszczalnia ścieków w Rzeszowie						
Q(t-1)	0,73	3107	8,31	0,74	2786	7,20
Q(t-1), Q(t-2)	0,74	3067	8,05		2720	7,04
Q(t-1), Q(t-2), Q(t-3)	0,77	2883	7,81	0,75	2703	7,02
P(t-1)	0,50	4339	11,76	0,58	4165	10,89
P(t-1), P(t-2)	0,53	4229	11,45	0,56	4312	11,33
P(t-1), Q(t-1)	0,77	3038	8,14	0,79	2705	7,06
P(t-1), Q(t-1), Q(t-2)	0,75	3009	8,03	0,81	2686	6,15
Q(t-1), Q(t-2), Q(t-3), P(t-1)	0,81	2814	7,62	0,82	2545	6,65
Q(t-1), Q(t-2), Q(t-3), P(t-1), P(t-2)		2779	7,50	0,83	2482	6,47

Na podstawie wyników symulacji, w tabeli 3 zamieszczono parametry definiujące struktury wyznaczonych sieci neuronowych, prognozujących dobowy dopływ ścieków do oczyszczalni w Rzeszowie i Kielcach. Można zauważyć, że liczba neuronów ukrytych zmieniała się od 3 do 10 (oczyszczalnia w Kielcach) oraz od 3 do 8 (oczyszczalnia w Rzeszowie). W modelach dotyczących oczyszczalni w Kielcach funkcję aktywacji w warstwie ukrytej najczęściej stanowiła zależność wykładnicza, a w warstwie wyjściowej funkcja liniowa, natomiast w przypadku oczyszczalni w Rzeszowie w obu warstwach najczęściej występowała funkcja tangensa hiperbolicznego. Jednak w tym przypadku trudniej było sformułować ogólny wniosek, wskazujący na wpływ określonych funkcji na jakość modelu, ponieważ wyniki obliczeń nie były tak jednoznaczne, jak poprzednio i jednocześnie wpływ doboru tych funkcji na wyniki obliczeń był mniejszy niż wpływ liczby neuronów w warstwie ukrytej sieci. Można jedynie zauważyć, że w modelach sieci uznanych za poprawne, funkcjami aktywacji na wejściach do warstwy ukrytej były funkcje wykładnicza, logarytmiczna oraz tangens hiperboliczny, natomiast na wejściach do warstwy wyjściowej – liniowa, sinusoidalna oraz tangens hiperboliczny. W celu wizualizacji otrzymanych obliczeń na rysunku 2 zamieszczono porównanie wyników pomiarów i prognoz dobowego dopływu ścieków do analizowanych oczyszczalni, otrzymanych za pomocą sieci neuronowych o najmniejszych wartościach MAE i MAPE.

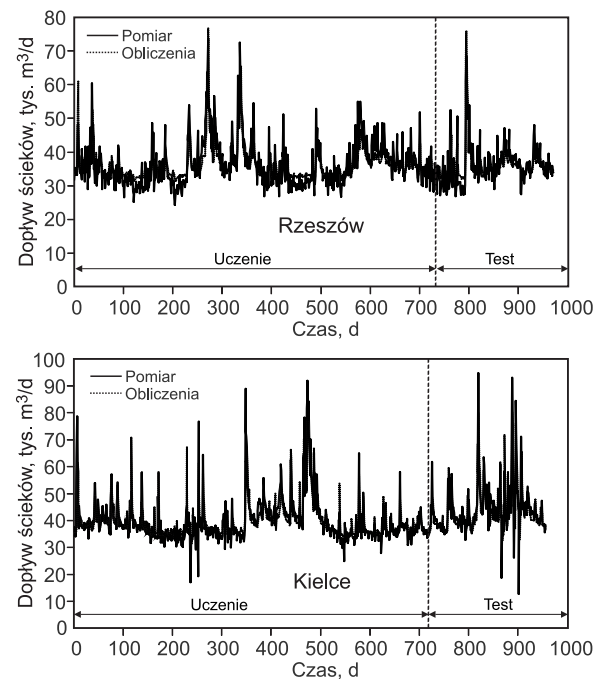
Na podstawie wykonanych obliczeń z użyciem sieci neuronowych opracowano model regresji logistycznej, pozwalający w rozpatrywanym przypadku ocenić wpływ poszczególnych zmiennych objaśniających oraz struktury sieci na zdolność predykcyjną wyznaczonych modeli sieci neuronowych. W analizie tej przyjęto zależność  $R=R_{gr}=0,70\div 0,75$ , stanowiącą podstawę transformacji obliczonych wartości współczynnika korelacji ( $R$ ) do postaci zmiennej zero-jedynkowej. Model regresji logistycznej w prowadzonych rozważaniach opisano następująco:

$$p = \frac{\exp[F(Q, h, P) + F(N, \theta_1, \theta_2)]}{1 + \exp[F(Q, h, P) + F(N, \theta_1, \theta_2)]} \quad (16)$$

w którym:

$F(Q, h, P)$  – funkcja opisująca liniową kombinację między binarnymi wartościami danych wejściowych  $Q(t-1)$ ,  $Q(t-2)$ ,  $Q(t-3)$ ,  $P(t-1)$ ,  $P(t-2)$

$F(N, \theta_1, \theta_2)$  – funkcja definiująca liniową kombinację między binarnymi wartościami zmiennych opisujących strukturę modelu



Rys. 2. Porównanie wyników obliczeń i pomiarów dobowego dopływu ścieków do oczyszczalni w Rzeszowie i Kielcach  
Fig. 2. Comparison of the calculation results and daily sewage inflow into the treatment plant in Rzeszow and Kielce

$N$  – liczba sygnałów wejściowych do modelu sieci  
 $\theta_1, \theta_2$  – funkcje aktywacji na wejściach odpowiednio do warstwy ukrytej i warstwy wyjściowej sieci (1 – wykładnicza, 2 – logistyczna, 3 – tangens hiperboliczny, 4 – liniowa, 5 – sinusoidalna).

W tabeli 4 zestawiono wartości obliczonych parametrów  $\beta_i$  (ujętych w równaniu regresji logistycznej) i odpowiadające im wartości prawdopodobieństw testowych ( $p$ ), a także parametry dopasowania danych pomiarowych do wyników obliczeń ( $SPEC$ ,  $SENS$ ,  $R_z^2$ ,  $R_{McFadden}^2$  i  $R_{Cox-Snell}^2$ ). Na podstawie tych wyników stwierdzono, że na prawdopodobieństwo przewyższenia wartości  $R_{gr}=0,75$  i  $R_{gr}=0,70$  miały wpływ zarówno zmienne objaśniające, jak i struktura wyznaczonego modelu sieci (liczba neuronów oraz funkcje aktywacji neuronów w warstwie ukrytej oraz funkcje aktywacji między warstwami), na co wskazują otrzymane wartości prawdopodobieństw testowych obliczonych wartości parametrów  $\beta_i$ . W przypadku  $R_{gr}=0,70$  wartości parametrów  $\beta_i$  występujących przy zmiennych opisujących

Tabela 3. Zestawienie parametrów opisujących strukturę modeli SSN w przypadku oczyszczalni ścieków w Kielcach i Rzeszowie  
Table 3. Summary of parameters describing the ANN model structure for the sewage treatment plants in Kielce and Rzeszow

Dane	Oczyszczalnia ścieków w Kielcach			Oczyszczalnia ścieków w Rzeszowie		
	liczba neuronów ukrytych	funkcja aktywacji		liczba neuronów ukrytych	funkcja aktywacji	
		wejście	wyjście		wejście	wyjście
$Q(t-1)$	3	lin.	lin.	5	log	tanh
$Q(t-1), Q(t-2)$	7	log	sin	8	wykl.	log
$Q(t-1), Q(t-2), Q(t-3)$	8	wykl.	lin	4	tanh	wykl.
$P(t-1)$	6	tanh	log	3		tanh
$P(t-1), P(t-2)$	5	wykl.	sin	5	4	sin
$P(t-1), Q(t-1)$	9	tanh	lin.	4		log
$P(t-1), Q(t-1), Q(t-2)$	10	wykl.	log		sin	log
$Q(t-1), Q(t-2), Q(t-3), P(t-1)$	9	log	sin	7	log	tanh
$Q(t-1), Q(t-2), Q(t-3), P(t-1), P(t-2)$	10	wykl.	lin.	8	tanh	

Tabela 4. Wartości liczbowe parametrów ( $\beta_i$ ) i prawdopodobieństw testowych ( $p$ ) zmiennych wejściowych oraz parametrów opisujących strukturę sieci neuronowej w modelach regresji logistycznej  
 Table 4. Parameter ( $\beta_i$ ) and test probability ( $p$ ) values of input variables and of ANN structure parameters in the calculated logistic regression models

Zmienna	Oczyszczalnia ścieków w Rzeszowie				Oczyszczalnia ścieków w Kielcach				
	R=0,75		R=0,70		R=0,75		R=0,70		
	$\beta_i$	p	$\beta_i$	p	$\beta_i$	p	$\beta_i$	p	
$\beta_0$	-5,529	0,001	-5,256	<0,0001	-8,732	<0,0001	-8,052	<0,0001	
Q(t-3)	-0,004	0,042	-0,021		-0,186	0,034	-0,390		
Q(t-2)	0,851	<0,0001	1,121		0,133	0,017	0,548		
Q(t-1)	5,042		6,125		7,481	8,540			
P(t-1)	0,751	0,427	0,002		0,525	<0,0001	0,215		0,003
P(t-2)	0,089	0,040	0,114		0,003	0,185	0,183		<0,0001
act <sub>ent</sub> (1)	1,032	0,043	0,756	0,012	0,422	0,044	0,368	0,003	
act <sub>ent</sub> (2)	1,164	0,042	0,811	0,023	0,401	0,037	0,351	0,001	
act <sub>ent</sub> (3)	1,369	0,040	0,921	0,034	0,307	0,021	0,021	0,002	
act <sub>ent</sub> (4)	0,251	0,038	0,201	0,025	0,229	0,031	0,466	0,001	
act <sub>ent</sub> (5)	0,624	0,040	0,457	0,037	0,413	0,028	0,160	0,003	
act <sub>ex</sub> (1)	1,231	0,030	0,712	0,022	1,540	0,009	0,800	<0,0001	
act <sub>ex</sub> (2)	1,862	<0,0001	0,821	<0,0001	1,424	0,048	0,655		
act <sub>ex</sub> (3)	1,993		1,072		1,656	0,037	0,596	0,001	
act <sub>ex</sub> (4)	1,764		0,812		0,007	0,644	0,002		
act <sub>ex</sub> (5)	1,601		0,914		1,283	0,021			
Liczba neuronów	0,035	0,010	0,002	0,030	0,021	0,025	0,004	0,003	
SENS	-	92,92%	-	88,15%	-	90,18%	-	85,21%	
SPEC		85,98%		82,23%		85,28%		91,94%	
R <sub>z</sub> <sup>2</sup>		89,52%		84,65%		88,42%		84,92%	
AIC		6296		5432		6400		3551	
R <sup>2</sup> <sub>McFadden</sub>		0,585		0,552		0,562		0,468	
R <sup>2</sup> <sub>Cox-Snell</sub>		0,532		0,531		0,512		0,474	

strukturę sieci neuronowej były znacznie mniejsze niż w przypadku  $R_{gr}=0,75$ , co oznacza, że przy takim założeniu maleje ich wpływ na prawdopodobieństwo przekroczenia granicznej wartości współczynnika korelacji (R). Wartości parametrów  $\beta_i$  i prawdopodobieństw testowych (p) wyznaczano w zakresie liczby neuronów od 2 do 12. Funkcje aktywacji w warstwach sieci przyjmowano dowolnie ze zbioru funkcji rozpatrywanych. Można zauważyć, że w przypadku korelacji przedziałowej, malejąca wartość współczynnika korelacji oznaczała jednocześnie gorsze zdolności predykcyjne modelu, malejący wpływ jego struktury i rosnący wpływ zmiennych objaśniających na wyniki obliczeń.

W następnej kolejności obliczono wartości ilorazów szans (OR) wskazujących, który z analizowanych parametrów sieci miał największy wpływ na prawdopodobieństwo przewyższenia założonej wartości  $R_{gr}$ , warunkującej wymaganą zgodność wyników pomiarów i obliczeń uzyskanych za pomocą modeli sieci neuronowej (tab. 5). Przy estymacji parametrów w modelu regresji logistycznej uwzględniono wyłącznie parametry istotne statystycznie na przyjętym poziomie istotności  $\alpha=0,05$ . W analizach tych do oceny istotności parametrów zastosowano test Chi-kwadrat. Na podstawie uzyskanych wyników stwierdzono,

że w przypadku, gdy wartość współczynnika korelacji modelu sieci neuronowej była na poziomie  $R_{gr}=0,75$ , wówczas dobierając funkcję aktywacji w warstwie neuronów wyjściowych modelu najlepiej będzie zastosować postać tangensa hiperbolicznego. Spośród wszystkich rozpatrywanych funkcji ta postać miała największy wpływ na prawdopodobieństwo przekroczenia założonej wartości  $R_{gr}$ , co potwierdziły wartości wyznaczonego ilorazu szans w przypadku obu oczyszczalni, wynoszące odpowiednio  $OR=7,338$  (Rzeszów) i  $OR=5,237$  (Kielce). W przypadku doboru funkcji aktywacji w warstwie neuronów ukrytych najlepiej będzie zastosować także funkcję tangensa hiperbolicznego (wartość ilorazu szans  $OR=3,931$  – Rzeszów) lub funkcję wykładniczą ( $OR=1,522$  – Kielce).

W oparciu o uzyskane wartości ilorazu szans (tab. 5) stwierdzono, że spośród rozpatrywanych zmiennych wejściowych największy wpływ na prawdopodobieństwo przewyższenia przyjętej wartości  $R_{gr}=0,75$  i na zdolności prognostyczne modelu miały dobowy dopływ ścieków do oczyszczalni i dobową wysokość opadów zmierzone w poprzedniej dobie. W rozpatrywanych przypadkach obliczone wartości ilorazów szans w przypadku oczyszczalni ścieków w Rzeszowie i Kielcach były wówczas równe odpowiednio  $OR=154,779$  i  $OR=1773,417$  (dopływ ścieków)

Tabela 5. Wartości ilorazu szans (OR) modeli regresji logistycznej  
Table 5. Odds ratio (OR) values for logistic regression models

Zmienna	Iloraz szans (OR)			
	Oczyszczalnia ścieków w Rzeszowie		Oczyszczalnia ścieków w Kielcach	
	R=0,75	R=0,70	R=0,75	R=0,70
Q(t-3)	0,996	0,979	0,830	0,677
Q(t-2)	2,119	3,068	1,143	1,730
Q(t-1)	154,779	457,145	1773,417	5116,019
P(t-1)	2,342	1,533	1,690	1,240
P(t-2)	1,093	1,121	1,203	1,200
act <sub>ent</sub> (1)	2,801	2,130	1,522	1,445
act <sub>ent</sub> (2)	3,728	2,250	1,493	1,420
act <sub>ent</sub> (3)	3,931	2,512	1,359	1,021
act <sub>ent</sub> (4)	1,285	1,223	1,349	1,234
act <sub>ent</sub> (5)	1,866	1,579	1,511	1,174
act <sub>ex</sub> (1)	3,425	2,038	4,663	2,226
act <sub>ex</sub> (2)	6,437	2,273	4,155	1,925
act <sub>ex</sub> (3)	7,338	2,921	5,237	1,815
act <sub>ex</sub> (4)	5,836	2,252		1,903
act <sub>ex</sub> (5)	4,958	2,494	3,607	
Liczba neuronów	1,036	1,002	1,021	1,004

oraz  $OR=2,342$  i  $OR=1,690$  (wysokość opadów). Znacznie mniejszy wpływ na dokładność predykcji modelu miały dobowy dopływ ścieków do oczyszczalni i wysokość opadów zmierzone dwie doby wcześniej. Podobne wyniki otrzymano przy wartości współczynnika korelacji na poziomie  $R_{gr}=0,70$ . Największy wpływ na prawdopodobieństwo przewyższenia tej wartości, a więc także na zdolności predykcyjne modelu, miał dobowy dopływ ścieków do oczyszczalni zmierzony z jednodobowym opóźnieniem ( $OR=457,145$  – Rzeszów i  $OR=5116,09$  – Kielce), a znacznie mniejszy wpływ miał dopływ ścieków zmierzony z dwudobowym opóźnieniem ( $OR=3,068$  – Rzeszów i  $OR=1,730$  – Kielce). Podobnie, wraz ze zwiększaniem opóźnienia, malał wpływ wysokości opadów na wyniki obliczeń (przy jednodobowym opóźnieniu  $OR=1,533$  – Rzeszów i  $OR=1,240$  – Kielce, natomiast przy dwudobowym opóźnieniu  $OR=1,121$  – Rzeszów i  $OR=1,200$  – Kielce).

Oceniając wpływ doboru funkcji aktywacji na wyniki obliczeń (przy założonej wartości  $R_{gr}=0,70$ ) stwierdzono, że w tym przypadku wartości ilorazów szans były znacznie mniejsze niż gdy  $R_{gr}=0,75$ , co oznaczało, że wpływ struktury modelu był mniejszy na prawdopodobieństwo przekroczenia granicznej wartości współczynnika korelacji. Na podstawie uzyskanych wyników (tab. 5) stwierdzono, że w przypadku, gdy wartość współczynnika korelacji modelu sieci neuronowej zastosowanej do prognozowania dobowej ilości ścieków ustaliła się na poziomie  $R_{gr}=0,75$ , to dobierając funkcję aktywacji w warstwie neuronów wejściowych modelu najlepiej będzie zastosować postać tangensa hiperbolicznego. Z kolei w przypadku doboru funkcji aktywacji w warstwie neuronów ukrytych, najlepiej będzie zastosować funkcję tangensa hiperbolicznego ( $OR=3,931$  – Rzeszów) lub funkcję wykładniczą ( $OR=1,522$  – Kielce).

Na podstawie przeprowadzonych analiz wykazano, że model logitowy może być z powodzeniem zastosowany do oceny wpływu poszczególnych zmiennych objaśniających i wybranych parametrów struktury modelu sieci neuronowej na dokładność predykcji modelu. W analizowanych przypadkach stwierdzono, że na prawdopodobieństwo przekroczenia założonej granicznej wartości  $R_{gr}$  istotny wpływ miały zarówno sygnały wejściowe modelu, jak i parametry sieci neuronowej dotyczące użytych funkcji aktywacji. W analizowanych przypadkach nie stwierdzono natomiast istotnego wpływu liczby neuronów w warstwie ukrytej na zdolności predykcyjne opracowanych modeli sieci neuronowych, co potwierdzają również rezultaty dopasowania wyników modelowania do pomiarów w poszczególnych kombinacjach zmiennych objaśniających.

## Podsumowanie

W optymalizacji działania poszczególnych elementów oczyszczalni celowe jest prognozowanie z odpowiednim wyprzedzeniem ilości i jakości dopływających ścieków. Ze względu na wysokie koszty prowadzenia takich kompleksowych badań należy uwzględnić przede wszystkim te zmienne, które mają najbardziej istotny wpływ na przebieg procesów oczyszczania ścieków.

Na podstawie wykonanych obliczeń stwierdzono, że model regresji logistycznej można z powodzeniem zastosować do oceny wpływu zarówno zmiennych wejściowych, jak i struktury modeli sieci neuronowych na ich zdolności predykcyjne. Biorąc pod uwagę uzyskane wyniki należy uznać za celowe prowadzenie dalszych badań dotyczących zastosowania modelu regresji logistycznej do opracowania wytycznych odnośnie doboru także innych zmiennych wejściowych w modelach sieci neuronowych stosowanych do prognozowania dopływu ścieków do oczyszczalni, takich jak przepustowość obiektu, lokalizacja i liczba urządzeń pomiarowych (deszczomierze), stan wody w odbiorniku, poziom wód gruntowych itp. W powyższych analizach modele logitowe opracowano na podstawie wyników prognozowania dobowego dopływu ścieków do dwóch oczyszczalni z wykorzystaniem licznych sieci neuronowych oraz przedstawiono ogólnie wiążące wskazówki doboru zmiennych objaśniających, funkcji aktywacji i liczby neuronów w warstwie ukrytej, jakimi należy kierować się przy tworzeniu modelu sieci neuronowej w przypadku dowolnej oczyszczalni, co jest bardzo istotne na etapie planowania eksperymentu.

Wykonane obliczenia za pomocą sieci neuronowych wykazały, że w przypadku oczyszczalni ścieków w Rzeszowie i Kielcach najlepsze prognozy dobowego dopływu ścieków uzyskano uwzględniając w modelach wysokość opadów atmosferycznych zmierzonych w dwóch poprzednich dobach oraz dopływ ścieków z trzech poprzednich dób. Z wykonanych analiz wynika jednocześnie, że znacznie lepsze wyniki obliczeń otrzymano na podstawie dobowego dopływu ścieków do oczyszczalni niż całkowitej wysokości opadów atmosferycznych.

Przeprowadzona analiza wykazała, że mimo iż badane oczyszczalnie ścieków i zlewnie znacznie się różniły, co potwierdzają wartości parametru  $\beta_1$ , to jednak uzyskane wyniki pozwalają na ich uogólnienie oraz na odpowiedni dobór zarówno zmiennych objaśniających, jak i parametrów struktury modelu, w celu uzyskania wymaganej dokładności prognozy dobowej ilości ścieków dopływających do badanego obiektu.

## LITERATURA

1. R. W. SZETELA: Model dynamiczny oczyszczalni ścieków z osadem czynnym. *Prace Naukowe Instytutu Inżynierii Ochrony Środowiska Politechniki Wrocławskiej*, seria: Monografie, Wrocław 1990.
2. J. STUZIŃSKI, L. BARTKIEWICZ, M. STACHURA: Development of mathematical models for forecasting hydraulic loads of water and wastewater networks. *EnviroInfo 2013: Environmental Informatics and Renewable Energies*, Shaker Verlag, Aachen 2013.
3. M. STACHURA: Komputerowa symulacja i optymalizacja modelu oczyszczalni ścieków. *Prace Naukowe IBS PAN*, seria: Badania Systemowe, t. 59, Warszawa 2008.
4. S. A. DELLANA, D. WEST: Predictive modeling for wastewater applications: Linear and nonlinear approaches. *Environmental Modelling and Software* 2009, Vol. 24, pp. 96–106.
5. A. G. EL-DIN, D. W. SMITH: Modelling approach for high flow rate in wastewater treatment operation. *Journal of Environmental Engineering and Science* 2002, Vol. 1, No. 4, pp. 275–291.
6. F. J. FERNANDEZ, A. SECO, J. FERRER, M. A. RODRIGO: Use of neurofuzzy networks to improve wastewater flow-rate forecasting. *Environmental Modelling and Software* 2009, Vol. 24, pp. 686–693.
7. X. WEI, A. KUSIAK, H. R. SADAT: Prediction of influent flow rate: Data-mining approach. *Journal of Energy Engineering* 2012, Vol. 19, No. 2, pp. 118–123.
8. A. BOROWA, M. A. BRDYŚ, K. MAZUR: Modeling of wastewater treatment plant for monitoring and control purposes by state-space wavelet networks. *International Journal of Computers, Communications & Control* 2007, Vol. 2, No. 2, pp. 121–131.
9. M. CHUCHRO: Struktura szeregu czasowego natężenia dopływu ścieków do oczyszczalni. Rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków 2009.
10. L. BARTKIEWICZ, J. STUZIŃSKI: Mathematical modeling of the hydraulic load of communal wastewater networks. In: G. K. JANSSENS, K. RAMAKERS, A. CARIS [Eds.]: *Modeling and Simulation*, EUROSIS-ETI, Hasselt (Belgium) 2010, pp. 156–160.
11. J. STUZIŃSKI, L. BARTKIEWICZ: Metody i programy wspomagające rozwiązywanie zadań modelowania i identyfikacji systemów. *Instal* 2009, nr 4a, ss. 59–64.
12. L. BARTKIEWICZ, M. STACHURA, J. STUZIŃSKI: Modelowanie obciążenia hydraulicznego miejskiego systemu wodociągowo-kanalizacyjnego. W: J. ŁOMOTOWSKI [red.]: *Wody opadowe a zjawiska ekstremalne*. Wydawnictwo Seidel-Przywecki, Warszawa 2011, ss. 75–82.
13. M. VILIMEK: An artificial neural network approach and sensitivity analysis in predicting skeletal muscle forces. *Acta of Bioengineering and Biomechanics* 2014, Vol. 16, No. 3, pp. 119–127.
14. A. KICZKO, R. J. ROMANOWICZ, M. OSUCH, E. KARAMUZ. Maximising the usefulness of flood risk assessment for the River Vistula in Warsaw. *Natural Hazards and Earth System Sciences* 2013, Vol. 13, pp. 3443–3455.
15. G. ARCHER, A. SALTELLI, I. M. SOBOL: Sensitivity measures, Anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation* 1997, Vol. 58, No. 2, pp. 99–120.
16. L. RUTKOWSKI: *Flexible Neuro-fuzzy Systems: Structures, Learning and Performance Evaluation*. Kluwer Academic Publisher, Boston 2004.
17. D. RUTKOWSKA, M. PILIŃSKI, L. RUTKOWSKI: *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. Wydawnictwo Naukowe PWN, Warszawa 1997.
18. D. RAHA: Exploring Artificial Neural Networks (ANN) modelling for a biological nutrient removal (BNR) sewage treatment Plant (STP) to forecast effluent suspended solids. *Indian Chemical Engineer* 2007, Vol. 49, No. 3, pp. 205–220.
19. D. MCFADDEN: Conditional logit analysis of qualitative choice behavior. In: P. ZAREMBKA [Ed.]: *Frontiers in Econometrics*. Academic Press, New York 1973, pp. 105–142.
20. B. PRADHAN: Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *Journal of Spatial Hydrology* 2009, Vol. 9, No. 2, pp. 1–18.
21. X. Y. SUN, C. J. THOMPSON, B. F. W. CROKE: Using a logistic regression model to delineate channel network in southeast Australia. Proc. of 19<sup>th</sup> International Congress on Modelling and Simulation, Perth (Australia) 2011.
22. T. HEYER, J. STAMM: Levee reliability analysis using logistic regression models – abilities, limitations and practical considerations. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* 2013, Vol. 7, No. 2, pp. 77–87.
23. F. INGELMO, M. J. MOLINA, J. M. PAZ, F. VISCONTI: Soil saturated hydraulic conductivity assessment from expert evaluation of field characteristics using an ordered logistic regression model. *Soil & Tillage Research* 2011, Vol. 115–16, pp. 27–38.
24. L. AYLEW, H. YAMAGISHI: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains. Central Japan. *Geomorphology* 2005, Vol. 65, No. 1–2, pp. 15–31.
25. S. KAWAGOE, S. KAZAMA, P. R. SARUKKALIGE: Probabilistic modelling of rainfall induced landslide hazard assessment. *Hydrology and Earth System Sciences* 2010, Vol. 14, No. 6, pp. 1047–1061.
26. P. A. GRANIERO, J. S. PRICE: Distribution of bog and heath in a Newfoundland blanket bog complex: Topographic limits on the hydrological processes governing blanket bog development. *Hydrology Earth System Sciences* 1999, Vol. 3, No. 2, pp. 223–231.
27. M. DANCZUK, J. ŁOMOTOWSKI: Application of the microwave energy to the hygienization of sewage sludge. *Environment Protection Engineering* 2010, Vol. 4, No. 36, pp. 77–85.
28. F. HARRELL: *Regression Modeling Strategies with Application to Linear Models, Logistic Regression, and Survival Analysis*. Springer Verlag, New York 2001.
29. Szeregi opadowe z Lotniskowej Stacji Meteorologicznej Rzeszów. IMGW, Rzeszów 2005–2008.

**Bartkiewicz, L., Szeląg, B., Studziński, J. Impact Assessment of Input Variables and ANN Model Structure on Forecasting Wastewater Inflow into Sewage Treatment Plants. *Ochrona Środowiska* 2016, Vol. 38, No. 2, pp. 29–36.**

**Abstract:** Due to a stochastic nature of sewage inflow into a treatment plant the inflow amount and its quality are highly variable which has a significant impact on the plant technological objects operation. Hence, sewage inflow forecasting would be desirable as it allows for mitigating the impact of abnormal events that might lead to major plant installation disruption. This paper presents the results of a raw sewage inflow modeling using

Artificial Neural Networks (ANNs). Results of the three-year measurements of precipitation rates and sewage treatment plant inflow in Rzeszów and Kielce were used in the analyses. To assess the impact of exogenous variables on the model quality the logistic regression method was applied. The variables considered were the precipitation rate and daily sewage inflow, which were appropriately delayed in relation to the forecasted inflow values. Impact of the model structure parameters on accuracy of the mathematical model forecasts was also investigated.

**Keywords:** Sewage treatment plant, wastewater inflow, modeling, forecasting, ANN, correlation coefficient.