

Bartosz Szeląg, Jan Studziński, Krzysztof Chmielowski, Aleksandra Leśniańska, Izabela Rojek

Prognozowanie ilości ścieków dopływających do oczyszczalni za pomocą sztucznych sieci neuronowych z wykorzystaniem liniowej analizy dyskryminacyjnej

Eksploatacja oczyszczalni ścieków jest złożonym zadaniem, wymagającym zapewnienia odpowiedniego przebiegu wielu procesów technologicznych w celu uzyskania założonego stopnia usuwania zanieczyszczeń. Wynika to ze stochastycznego charakteru dopływu ścieków do oczyszczalni, na co wpływa zmienna w czasie ilość wody pobieranej z sieci wodociągowej, a także (w niektórych przypadkach) wysokość opadów atmosferycznych, powodujących określoną nierównomierność dopływu ścieków do oczyszczalni, jak również stan wody w odbiorniku i poziom wód gruntowych. W celu zapewnienia poprawnej eksploatacji oczyszczalni ścieków niezbędna jest odpowiednio wczesna informacja o ilości i jakości dopływających ścieków, co pozwala na przyjęcie z wyprzedzeniem właściwych wartości parametrów pracy poszczególnych reaktorów. Prognoza ilości ścieków, które dopłyną w ciągu najbliższej godziny, doby lub tygodnia odgrywa ważną rolę w eksploatacji oczyszczalni, umożliwiając identyfikowanie zdarzeń nagłych, które mogą prowadzić do zaburzeń w działaniu różnych obiektów, co z kolei może przekładać się na podejmowanie nieodpowiednich decyzji przez technologa obiektu. Z uwagi na to, że proces powrotu danego obiektu oczyszczalni do właściwego stanu funkcjonowania jest zwykle czasochłonny i kosztowny, podjęto liczne prace badawcze w kraju i za granicą mające na celu opracowanie modeli matematycznych (fizykalnych i probabilistycznych) prognozujących ilość i jakość ścieków dopływających do oczyszczalni.

Na podstawie dokonanego przeglądu literatury można stwierdzić, że do prognozowania ilości ścieków dopływających do oczyszczalni stosuje się przede wszystkim liczne

metody uczenia maszynowego, takie jak modele szeregów czasowych, sztuczne sieci neuronowe, analizę falkową czy logikę rozmytą [1, 2]. Wpływ poszczególnych zmiennych niezależnych na zdolność predykcyjną wyznaczanych modeli analizuje się zwykle przez proste usuwanie lub dodawanie konkretnego sygnału wejściowego w modelu i porównywanie wartości współczynnika korelacji między wartościami obliczonymi i uzyskanymi z pomiarów [3, 4]. Z doniesień literaturowych wynika również, że dotychczas nie podjęto próby bardziej formalnej oceny wpływu poszczególnych sygnałów wejściowych modelu na prognozę ilości i jakości ścieków dopływających do oczyszczalni. Właściwy dobór sygnałów wejściowych i struktury modelu jest przy tym istotny z punktu widzenia uzyskania wiarygodnej prognozy.

W niniejszym artykule zastosowano sztuczne sieci neuronowe do prognozowania ilości ścieków dopływających do oczyszczalni na podstawie 3-letnich ciągów obserwacyjnych obejmujących przepływ ścieków, wysokość opadów atmosferycznych, stany wody w odbiorniku i ilość wody tłoczonej do sieci. Do oceny wpływu poszczególnych sygnałów wejściowych i struktury modelu na dopasowanie wyników obliczeń do danych pomiarowych zastosowano drzewa klasyfikacyjne C&RT, na podstawie których eliminowano zmienne mające pomijalny wpływ na dokładność prognozy. Do oceny zdolności predykcyjnych opracowanych modeli prognostycznych zastosowano również liniową analizę dyskryminacyjną, badając i porównując w ten sposób trzy metody analizy struktury modeli.

Obiekt i metody badań

Obiektem badań była sieć kanalizacyjna i oczyszczalnia ścieków w Rzeszowie. Miasto ma zasadniczo system kanalizacji rozdzielczej, jednak w jego niektórych rejonach brakuje kanalizacji deszczowej. Ponad 50% długości kolektorów to kanały betonowe, a wiek około 65% sieci kanalizacyjnej przekracza 20 lat. Poziom wód gruntowych na terenie miasta istotnie zależy od stanów wody w Wisłoku oraz jego dopływach [5–8]. W rezultacie, przy wysokich stanach wód gruntowych (lokalnie 0,3 m p.p.t.) i dużej ilości wód opadowych, obserwuje się wzrost wypełnień kolektorów kanalizacyjnych. Ścieki z terenu Rzeszowa i okolicznych miejscowości odprowadzane są systemem kanalizacji sanitarnej do oczyszczalni mechaniczno-biologicznej o przepustowości nominalnej 72 tys. m³/d.

Dr inż. B. Szeląg: 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

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

Dr hab. inż. K. Chmielowski: Uniwersytet Rolniczy im. Hugona Kołłątaja w Krakowie, Wydział Inżynierii Środowiska i Geodezji, Katedra Inżynierii Sanitarnej i Gospodarki Wodnej, al. Adama Mickiewicza 24/28, 30-059 Kraków

Mgr inż. A. Leśniańska: Politechnika Koszalińska, Wydział Inżynierii Łądowej, Środowiska i Geodezji, Zakład Sieci i Instalacji Sanitarnych, ul. Śniadeckich 2, 75-453 Koszalin

Dr inż. I. Rojek: Uniwersytet Kazimierza Wielkiego w Bydgoszczy, Wydział Matematyki, Fizyki i Techniki, Zakład Systemów Baz Danych i Inteligencji Obliczeniowej, ul. Mikołaja Kopernika 1, 85-074 Bydgoszcz

Sztuczne sieci neuronowe (SSN)

Sztuczne sieci neuronowe mają szerokie zastosowanie do modelowania różnorodnych zjawisk w inżynierii środowiska [9–11]. Opracowano dotychczas wiele struktur SSN dopasowanych do różnych przypadków obliczeń, jednak jedną z najpowszechniej stosowanych struktur jest sieć jednokierunkowa wielowarstwowa, znana jako perceptron wielowarstwowy (MLP). Badania potwierdziły, że SSN z trzema warstwami ukrytymi może z dużą dokładnością rozwiązać większość zadań występujących w praktyce [12].

W przeprowadzonych obliczeniach do budowy modeli prognostycznych opartych na SSN wykorzystano program Statistica. Sygnałami wejściowymi były kombinacje wysokości opadów (P), dobowej ilości wody tłoczzonej do sieci (Q_w), stanu wody w odbiorniku (h) i dobowej ilości ścieków dopływających do oczyszczalni (Q), mierzonej w chwilach (t-j), przy czym t oznacza czas wyrażony w dobach, zaś j – opóźnienie wyrażone także w dobach. Celem obliczeń było znalezienie takiego zestawu sygnałów wejściowych (P, Q_w , h, Q) i takich wartości ich opóźnień (j), aby otrzymać najlepszą prognozę ilości ścieków dopływających do oczyszczalni w danej chwili (t). W trakcie badań przeanalizowano 500 różnych kombinacji parametrów sieci (liczba i rodzaj zmiennych niezależnych, opóźnienia w pomiarach, liczba neuronów w warstwie ukrytej, rodzaj funkcji aktywacji), na podstawie których zbudowano łącznie 5 tys. sztucznych sieci neuronowych typu MLP.

Przy modelowaniu wykorzystano opcję Automatycznego Projektanta (Statistica), co pozwoliło na zbudowanie i przeanalizowanie sieci bardzo zróżnicowanych ze względu na liczbę warstw ukrytych, liczbę neuronów w poszczególnych warstwach i rodzaj funkcji aktywacji między poszczególnymi warstwami. Rozpatrywano przy tym funkcje: liniową, hiperboliczną, sinus, tangens hiperboliczny, wykładniczą, eksponentyjalną i logistyczną. W celu prawidłowego przeprowadzenia procesu uczenia i właściwej oceny działania sieci dokonano podziału zbioru danych pomiarowych na trzy zbiory (uczący – 50%, walidacyjny – 25% i testowy – 25%). Do uczenia sieci neuronowych zastosowano algorytm Broydena-Flechera-Goldfarba-Shanno [12].

Kryteria oceny modeli

Do oceny zdolności predykcyjnej modeli prognozujących dopływ ścieków do oczyszczalni komunalnej w Rzeszowie zastosowano następujące miary dopasowania wyników obliczeń do danych pomiarowych:

– średni błąd procentowy:

$$MPE = \frac{1}{N} \sum_{i=1}^N \frac{(y_{t,i} - y_{n,i})}{y_{t,i}} 100 \quad (1)$$

– maksymalny błąd względny:

$$\text{maks. PE} = \text{maks.} \left[\frac{(y_{t,i} - y_{n,i})}{y_{t,i}} 100 \right] \quad (2)$$

– minimalny błąd względny:

$$\text{min. PE} = \text{min.} \left[\frac{(y_{t,i} - y_{n,i})}{y_{t,i}} 100 \right] \quad (3)$$

– względny średniokwadratowy błąd resztowy:

$$WBR = \frac{1}{\bar{y}_t} \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{t,i} - y_{n,i})^2} \quad (4)$$

– współczynnik korelacji:

$$R = \frac{\sum_{i=1}^N (y_{t,i} - \bar{y}_t)(y_{n,i} - \bar{y}_n)}{\sqrt{\sum_{i=1}^N (y_{t,i} - \bar{y}_t)^2} \sqrt{\sum_{i=1}^N (y_{n,i} - \bar{y}_n)^2}} \quad (5)$$

w których:

N – liczebność zbioru danych

$y_{t,n}$ – wartość zmierzona, obliczona

$\bar{y}_{t,n}$ – wartość średnia z wartości zmierzonych, obliczonych

Do oceny zdolności predykcyjnych przyjętych modeli prognostycznych zastosowano również tak zwane kryterium informacyjne, na podstawie którego ocenia się statystyczną poprawność dopasowania modelu do danych pomiarowych, uwzględniając złożoność modelu związaną z liczbą parametrów występujących w jego strukturze. Wzięto pod uwagę kryterium AIC (Akaike information criterion) opisane zależnością w postaci:

$$AIC = [2,83788771 + \ln(\text{SSE}/N)]N + 2L_p \quad (6)$$

będącą uproszczoną postacią ogólnej formuły [13]:

$$AIC = [\ln(2\pi \cdot \text{SSE}/N)]N + 2L_p \quad (7)$$

w której SSE oznacza sumę kwadratów różnic wartości zmierzonych i obliczonych, natomiast L_p – liczbę estymowanych parametrów w strukturze modelu, odpowiadającą liczbie uwzględnionych predyktorów. Kryterium AIC jest wygodnym i powszechnie stosowanym wskaźnikiem poprawnego doboru struktury modelu, wyznaczanego jedną z metod najmniejszej sumy kwadratów [14, 15].

Ocena wrażliwości modelu

Do oceny wpływu rozpatrywanych zmiennych niezależnych na dokładność predykcyjną sztucznej sieci neuronowej wykorzystano drzewa klasyfikacyjne (C&RT) [16]. Pozwoliło to na ustalenie rankingu zmiennych niezależnych i w rezultacie współczynników ważności przyjętych predyktorów ($Q(t-i)$, $P(t-i)$, $h(t-i)$, $Q_w(t-i)$) oraz ustalenie właściwej struktury sieci MLP (liczba neuronów w warstwie ukrytej, rodzaj funkcji aktywacji w warstwie ukrytej i wyjściowej). Takie podejście umożliwia wyeliminowanie zmiennych niezależnych mających pomijalny wpływ na zmienną zależną. Szczegółowy opis algorytmu wyznaczania rankingu predyktorów w metodzie C&RT omówiono w pracy [17].

Liniowa analiza dyskryminacyjna

Liniowa analiza dyskryminacyjna (linear discriminant analysis – LDA) jest metodą statystyczną pozwalającą dzielić badany zbiór przypadków (danych pomiarowych) na grupy na podstawie analizy zbioru zmiennych niezależnych (predyktorów). Autorem koncepcyjnych i matematycznych podwalin tej metody był Ronald Fisher [18], przy czym analizę dyskryminacyjną można stosować w wielu dziedzinach nauki i praktyki społecznej [19]. Psycholodzy mogą ją stosować do selekcji pracowników lub rekrutacji studentów, można ją także wykorzystywać w zagadnieniach ekonomicznych do oceny bankructwa spółek akcyjnych i firm oraz klasyfikacji regionów ze względu na ich użytkowanie rolnicze [20]. Analiza dyskryminacyjna może być także stosowana do prognozowania stanu technicznego kanałów [21] oraz modelowania sedymentacji osadu czynnego [6]. W metodzie tej poszukuje się równań

hiperpłaszczyzn, zwanych funkcjami dyskryminacyjnymi, odseparowujących wydzielone grupy przypadków i opisanych zależnościami [18]:

$$D_{kj} = \alpha_0 + \alpha_1 x_{1kj} + \dots + \alpha_p x_{pkj} \quad (8)$$

w której:

D_{kj} – wartość funkcji dyskryminacyjnej w k -tym przypadku w j -tej grupie ($k=1, \dots, N$; $j=1, \dots, g$; g – liczba grup)

α_i – współczynniki funkcji dyskryminacyjnej

x_{ikj} – wartość i -tej zmiennej niezależnej w k -tym przypadku w j -tej grupie ($i=1, \dots, p$)

p – liczba zmiennych niezależnych (predyktorów)

Do doboru zmiennych niezależnych w równaniu (8) zastosowano algorytm krokowy regresji z programu Statistica (forward stepwise regression algorithm), w którym poszczególne zmienne są dobierane sukcesywnie w sposób automatyczny. Przed przystąpieniem do estymowania parametrów strukturalnych (α_i) przeprowadza się podział obserwacji zmiennych niezależnych na co najmniej dwie podgrupy, to jest zmiennych determinujących przyjęcie przez zmienną zależną wartości równej zero oraz zmiennych determinujących przyjęcie przez zmienną zależną wartości równej jeden. Parametry strukturalne modelu opisanego równaniem (8) wyznacza się zgodnie z metodą podaną na przykład w pracach [17, 22]. Podstawę oceny zdolności predykcyjnej modelu stanowi tak zwana wrażliwość (względnie czułość) modelu, co zostało szczegółowo opisane w pracach [5, 6]. Po określeniu wartości parametrów α_i można – na podstawie obliczonych wartości funkcji dyskryminacyjnych D_{kj} – dokonać klasyfikacji przypadków do odpowiednich grup określanych według zmiennych x_i . Dany przypadek zostaje zaklasyfikowany do grupy, w której wartość funkcji dyskryminacyjnej osiąga maksimum.

W niniejszej pracy zastosowano liniową analizę dyskryminacyjną pozwalającą ocenić zdolność predykcyjną budowanej sztucznej sieci neuronowej na podstawie przyjętych sygnałów wejściowych i przyjętej struktury SSN. Ze względu na dużą liczbę zmiennych niezależnych (x_i) wykorzystano do analiz wyniki obliczeń otrzymane za pomocą

drzew C&RT, co pozwoliło na wyeliminowanie zmiennych mających pomijalny wpływ na wynik prognozy uzyskany z zastosowaniem SSN. Uwzględniając wyniki obliczeń wykonane wcześniej [5], przyjęto trzy funkcje dyskryminacyjne odpowiadające różnym wartościom współczynnika korelacji (R) obliczonego w procesie walidacji modelu:

$$D_1 = f(R \leq 0,6) \quad (9)$$

$$D_2 = f(0,6 < R < 0,8) \quad (10)$$

$$D_3 = f(R \geq 0,8) \quad (11)$$

W przyjętych przedziałach wartości R i odpowiednio dobranych zmiennych niezależnych określono parametry modelu dyskryminacyjnego. Należy podkreślić, że zmienna zależna badana w modelu dyskryminacyjnym nie ma charakteru ciągłego i jest binarna, ponieważ gdy zmienna niezależna jest ujęta w modelu, to zmienna zależna przyjmuje wartość 1 (w przeciwnym wypadku 0).

Wyniki obliczeń

Prognozowanie ilości ścieków

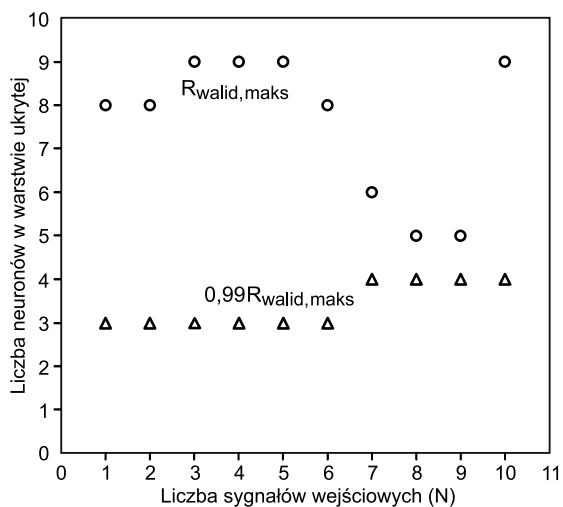
Na podstawie przyjętych założeń, za pomocą sieci MLP opracowano prognozę ilości ścieków dopływających do oczyszczalni. W tabeli 1 zamieszczono miary dopasowania (R , AIC, MPE, WBR, maxPE, minPE) wartości prognozowanych do wartości zmierzonych. Z przeprowadzonych analiz wynika, że najlepsze efekty prognozowania ilości ścieków dopływających do oczyszczalni uzyskano za pomocą sieci MLP z trzema wejściami, to jest zmierzoną ilością dopływających ścieków, wysokością opadów atmosferycznych oraz stanem wody w odbiorniku, ustalonych z 1-dobowym opóźnieniem. Na podstawie wartości zebranych w tabeli 1 można wnioskować, że wzrost liczby zmiennych (x_i) objaśniających ilość ścieków dopływających do oczyszczalni ma istotny – chociaż nie zawsze pozytywny – wpływ na dokładność prognozy za pomocą sieci MLP, ponieważ oprócz ich liczby istotną rolę odgrywa również rodzaj zmiennej.

Tabela 1. Statystyczne miary dopasowania SSN prognozujących ilość ścieków dopływających do oczyszczalni
Table 1. Statistical adjustment measures for ANN forecasting of sewage inflow into the treatment plant

Dane wejściowe	Współczynnik korelacji (R)		AIC	MPE, %	WBR	PE, %	
	uczenie	walidacja				min.	maks.
Q(t-1)	0,803	0,748	9617	-11,11	0,128	-56,17	0,05
Q(t-1), h(t-1)	0,713	0,769	9637	-10,50	0,128	-44,27	-0,07
Q(t-1), P(t-1), h(t-1)	0,774	0,815	9492	-7,07	0,115	-42,65	-0,07
Q(t-1), h(t-1, t-2), P(t-1)	0,758	0,810	9551	-10,85	0,112	-45,29	0,05
Q(t-1, t-2), h(t-1, t-2), P(t-1)	0,744	0,823	9701	-8,39	0,129	-40,78	-0,09
Q(t-1, t-2), h(t-1, t-2), P(t-1, t-2)	0,682	0,825	9729	-8,09	0,132	-62,23	-0,04
Q(t-1, t-2), h(t-1, t-2, t-3), P(t-1, t-2)	0,731	0,830	9678	10,31	0,133	-34,23	0,04
Q(t-1, t-2), h(t-1, t-2, t-3), P(t-1, t-2), Q _w (t-1)	0,712	0,831	9683	6,45	0,130	-40,68	-0,31
Q(t-1, t-2), h(t-1, t-2, t-3), P(t-1, t-2), Q _w (t-1, t-2)	0,738	0,830	9647	6,90	0,124	-37,68	-0,41
Q(t-1, t-2), h(t-1, t-2, t-3), P(t-1, t-2), Q _w (t-1, t-2, t-3)	0,771	0,834	9734	-5,72	0,121	-33,75	0,04

Przykładowo, przy jednym sygnale wejściowym (pojedyncza zmienna niezależna – x_1) wartość współczynnika korelacji (R) zbioru walidacyjnego wynosiła 0,748, podczas gdy w przypadku 3 oraz 4 i 5 sygnałów wejściowych jego wartości były równe odpowiednio 0,815 oraz 0,810 i 0,823. Wartość średniego błędu procentowego (MPE) zmieniła się w zakresie od –11,11% do 10,31%, minimalne i maksymalne wartości błędu względnego (maks. PE, min. PE) wynosiły odpowiednio od –62,23% do –33,75% oraz od –0,41% do 0,05%, natomiast wartość WBR zmieniła się w zakresie 0,112÷0,133. Funkcjami aktywacji w warstwach ukrytej i wyjściowej sieci MLP były tangens hiperboliczny, liniowa, sinus oraz eksponentialna i logistyczna.

Na podstawie otrzymanych wyników obliczeń sporządzono wykres (rys. 1) obrazujący wpływ liczby wejść (zmiennych niezależnych) do SSN na liczbę neuronów w warstwie ukrytej, otrzymaną przy maksymalnych wartościach współczynnika korelacji w przypadku danych ze zbioru walidacyjnego ($R_{\text{walid, maks}}$) i wartości współczynnika korelacji $R_{\text{walid}}=0,99R_{\text{walid, maks}}$. Analiza danych wykazała stosunkowo małe zróżnicowanie liczby neuronów w uzyskanych modelach. Przykładowo, w przypadku trzech zmiennych niezależnych (liczba sygnałów wejściowych do modelu $N=3$) liczba neuronów w warstwie ukrytej wynosiła 9, a w przypadku 10 sygnałów wejściowych także 9, przy czym korelacja ta mocno zależała nie tyle od liczby wejść, ile od ich rodzaju (przy $N=8$ i $N=9$ liczba neuronów w warstwie ukrytej wynosiła 5). Ostatecznie można stwierdzić, że ilość ścieków dopływających do oczyszczalni można prognozować z zadowalającą dokładnością za pomocą SSN zawierających od 3 do 5 neuronów w warstwie ukrytej, co znacznie upraszcza strukturę modelu.

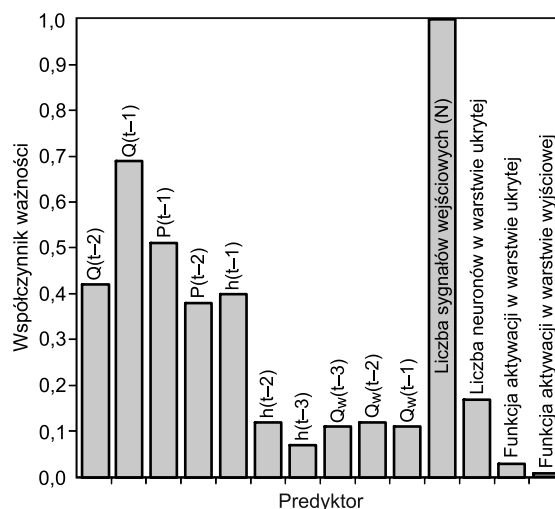


Rys. 1. Wpływ liczby sygnałów wejściowych do sieci na liczbę neuronów w warstwie ukrytej ($R_{\text{walid, maks}}$ – maksymalna wartość R uzyskana przy N niezależnych zmiennych wejściowych)

Fig. 1. Effect of number of inputs to the network on the number of neurons in the hidden layer ($R_{\text{walid, maks}}$ – maximum R value obtained for N independent input variables)

Wrażliwość modeli

Wykorzystując uzyskane wyniki zbadano wpływ poszczególnych zmiennych i ich opóźnień oraz struktury SSN (liczba sygnałów wejściowych, liczba neuronów w warstwie ukrytej, funkcje aktywacji) na dokładność wyznaczonej prognozy. Z rysunku 2 wynika, że największy wpływ na dokładność prognozy ilości ścieków dopływających do oczyszczalni miała liczba zmiennych niezależnych (liczba



Rys. 2. Wpływ poszczególnych predyktorów na wartość współczynnika ważności w procesie walidacji

Fig. 2. Influence of individual predictors on the importance coefficient value in the validation process

sygnałów wejściowych modelu – N). Ponadto stwierdzono, że zarówno liczba neuronów w warstwie ukrytej, jak i w szczególności wybór funkcji aktywacji miały niewielki wpływ na wyniki obliczeń, co uzasadniały wartości współczynnika ważności omawianych predyktorów.

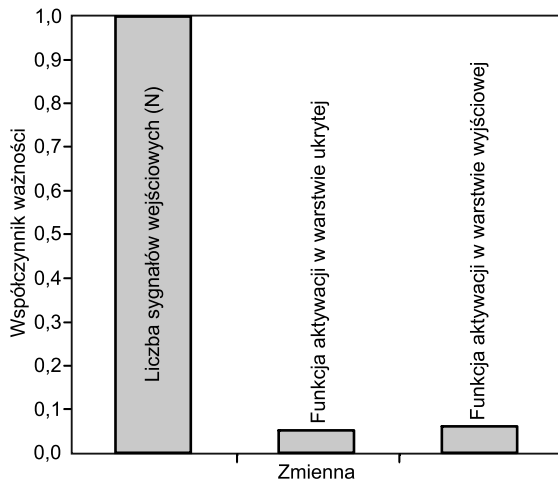
Stosunkowo mały wpływ na ilość ścieków dopływających do oczyszczalni miał stan wody w odbiorniku przyjęty z 2- i 3-dobowym opóźnieniem, jak również dobową ilość wody tłoczona do sieci ($Q_w(t-1) \approx Q_w(t-3)$). Wartości ważności tych predyktorów wynosiły jedynie od 0,107 do 0,113. Ta ostatnia konkluzja, chociaż znajduje potwierdzenie w pracach [5–10], kłóci się jednak z wynikami podanymi w tabeli 1, z której wynika, że uwzględnienie w modelu kolejno zmiennych $Q_w(t-1)$, $Q_w(t-2)$ oraz $Q_w(t-3)$ zwiększyło dokładność predykcji modelu, sądząc po malejącej wartości WBR (od 0,130 do 0,121) i rosnącej wartości R w procesie walidacji (od 0,831 do 0,834). Stan wody w odbiorniku przyjęty z 1-dobowym opóźnieniem oraz wysokość opadów atmosferycznych, a także ilość ścieków dopływających do oczyszczalni, ustalone z 1- i 2-dobowym opóźnieniem, miały już istotny wpływ na dopływ ścieków, gdyż ważności rozpatrywanych predyktorów plasowały się w przedziale wartości od 0,37 do 0,69.

Reasumując powyższe obserwacje można stwierdzić, że stosunkowo najlepszym modelem była sieć MLP z trzema neuronami na warstwie ukrytej i trzema sygnałami wejściowymi (P, h, Q) mającymi 1-dobowe opóźnienie. Taki model miał odpowiednio duże wartości R zarówno w procesie uczenia, jak i walidacji sieci, stosunkowo małą wartość WBR i również małą, a więc wskazaną wartość współczynnika AIC.

Liniowa analiza dyskryminacyjna

Budując modele prognostyczne ilości ścieków dopływających do oczyszczalni stwierdzono, że największy wpływ na liczbę neuronów w warstwie ukrytej sieci MLP miała liczba sygnałów wejściowych do modelu, natomiast dobór funkcji aktywacji w warstwie ukrytej i w warstwie wyjściowej na liczbę tych neuronów miał praktycznie nieistotny wpływ, gdyż wartości ważności tych predyktorów były mniejsze od 0,10 (rys. 3).

Mając na uwadze wyniki obliczeń opracowano liniowe modele dyskryminacyjne do opisu zależności (9)–(11).



Rys. 3. Wpływ liczby sygnałów wejściowych oraz dobranych funkcji aktywacji w warstwie ukrytej i wyjściowej na wartość współczynnika ważności

Fig. 3. Effect of the number of inputs and selected activation functions in the hidden and output layer on the importance coefficient value

W początkowym etapie uwzględniono takie zmienne niezależne (sygnały wejściowe), jak $Q(t-1)$, $Q(t-2)$, $P(t-1)$, $P(t-2)$, $h(t-1)$, $h(t-2)$, $h(t-3)$ oraz różne liczby neuronów w warstwie ukrytej. Liniowa analiza dyskryminacyjna wykazała, że liczba neuronów w warstwie ukrytej oraz stan wody w odbiorniku podany z 2- i 3-dobowym opóźnieniem ($h(t-2)$ i $h(t-3)$) nie były statystycznie istotne na przyjętym poziomie ufności. Z tego względu dokonano modyfikacji modelu i danych wejściowych i sformułowano następujące równania:

$$D_1 = 0,21Q(t-1) + 0,09Q(t-2) + 0,69P(t-1) + 0,70P(t-2) + 0,02h(t-1) - 2,57 \quad (12)$$

$$D_2 = 2,49Q(t-1) + 2,39Q(t-2) + 3,89P(t-1) + 2,73P(t-2) + 1,57h(t-1) - 3,08 \quad (13)$$

$$D_3 = 3,64Q(t-1) + 3,79Q(t-2) + 5,90P(t-1) + 2,82P(t-2) + 4,27h(t-1) - 9,45 \quad (14)$$

Przeprowadzone obliczenia wykazały, że uzyskany model dyskryminacyjny charakteryzował się dobrą zdolnością do poprawnej klasyfikacji współczynnika korelacji wyznaczonego w procesie walidacji na podstawie zmiennych ujętych w modelu – w pierwszej, drugiej i trzeciej grupie pomiarów otrzymano zgodność danych pomiarowych z wynikami obliczeń równą odpowiednio 86,3%, 79,7% i 79,6%, co potwierdziło możliwość stosowania opracowanych modeli w rozważaniach praktycznych.

Wnioski

◆ Dobową ilość ścieków dopływających do oczyszczalni w Rzeszowie można prognozować z zadowalającą dokładnością za pomocą sztucznych sieci neuronowych z sygnałami wejściowymi mającymi 1-dobowe opóźnienie, co potwierdziły wyznaczone miary dopasowania wyników obliczeń do pomiarów.

◆ Największy wpływ na dokładność prognozy ilości ścieków dopływających do oczyszczalni miała liczba sygnałów wejściowych w SSN. Przeprowadzone obliczenia wykazały również, że liczba neuronów w warstwie ukrytej oraz zastosowane funkcje aktywacji miały nieznaczny wpływ na dokładność prognozy.

◆ Spośród rozpatrywanych sygnałów wejściowych największy wpływ na prognozę ilości ścieków dopływających do oczyszczalni miały wysokość opadów i zmierzona ilość ścieków z 1- i 2-dobowym opóźnieniem oraz stan wody w odbiorniku podany z 1-dobowym opóźnieniem.

◆ Wszystkie analizowane sposoby wyznaczania struktury sztucznej sieci neuronowej poprawnie prognozującej ilość ścieków dopływających do oczyszczalni (1 – klasyczne obliczenia symulacyjne, w wyniku których (po wielu próbach) otrzymuje się najlepszy model, 2 – wykorzystanie analizy wrażliwości modeli, za pomocą której wyznacza się wartości współczynników ważności poszczególnych predyktorów, 3 – zastosowanie liniowej analizy dyskryminacyjnej, która również pozwala uzyskać wskazówki, jakie sygnały wejściowe i z jakimi opóźnieniami umożliwiają wyznaczenie najlepszego modelu) dały podobne wyniki, przy czym pierwszy z nich był najbardziej czasochłonny, z uwagi na uciążliwą metodę polegającą na powtarzaniu tych samych czynności aż do osiągnięcia spodziewanego wyniku.

◆ Wykonane obliczenia wykazały celowość stosowania liniowej analizy dyskryminacyjnej do wyznaczenia struktury sieci MLP wykorzystywanej w prognozowaniu ilości ścieków dopływających do oczyszczalni.

LITERATURA

1. A. DELLANA, D. WEST: Predictive modeling for wastewater applications: Linear and nonlinear approaches. *Environmental Modelling and Software* 2009, Vol. 24, pp. 96–106.
2. 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.
3. J. STUDZIŃSKI, 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.
4. L. BARTKIEWICZ, M. STACHURA, J. STUDZIŃSKI: Modelowanie obciążenia hydraulicznego miejskiego systemu wodociągowo-kanalizacyjnego. W: J. ŁOMOTOWSKI [red.]: *Wody opadowe a zjawiska ekstremalne*. Wydawnictwo Seidel-Przywecki, Warszawa 2011, ss. 75–82.
5. L. BARTKIEWICZ, B. SZELĄG, J. STUDZIŃSKI: Ocena wpływu zmiennych wejściowych oraz struktury modelu sztucznej sieci neuronowej na prognozowanie dopływu ścieków komunalnych do oczyszczalni (Impact assessment of input variables and ANN model structure on forecasting wastewater inflow into sewage treatment plants). *Ochrona Środowiska* 2016, vol. 38, nr 2, ss. 29–36.
6. B. SZELĄG, P. SIWICKI: Application of the selected classification models to the analysis of the settling capacity of the activated sludge – case study. *E3S Web of Conferences* 17, Kielce (Poland) 2017.
7. B. SZELĄG, L. BARTKIEWICZ, J. STUDZIŃSKI, K. BARBUSIŃSKI: Evaluation of the impact of explanatory variables on the accuracy of prediction of daily inflow to the sewage treatment plant by selected models nonlinear. *Archives of Environmental Protection* 2017, Vol. 43, No. 3, pp. 74–81.
8. B. SZELĄG, K. BARBUSIŃSKI, J. STUDZIŃSKI, L. BARTKIEWICZ: Prediction of wastewater quality indicators at the inflow to the wastewater treatment plant using data mining methods. *E3S Web of Conferences* 22, Wrocław (Poland) 2017.
9. M. CHUCHRO: *Struktura szeregu czasowego natężenia dopływu ścieków do oczyszczalni*. Rozprawa doktorska, Akademia Górniczo-Hutnicza, Kraków 2009.
10. 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.

11. H. Z. ABYANEH: Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. *Journal Environmental Health Science & Engineering* 2014, Vol. 12, pp. 1–8.
12. L. RUTKOWSKI: Metody i techniki sztucznej inteligencji. Wydawnictwo Naukowe PWN, Warszawa 2006.
13. H. AKAIKE: A new look at the statistical model identification. *IEEE Transaction on Automatic Control* 1974, Vol. 19, No. 6, pp. 716–723.
14. K. W. HIPEL, A. I. McLEOD: Time Series Modelling of Water Resources and Environmental Systems. Series: Developments in Water Science 45. Elsevier Science B.V., Amsterdam 1994.
15. K. MAŃCZAK, Z. NAHORSKI: Komputerowa identyfikacja obiektów dynamicznych. PWN, Warszawa 1983.
16. L. BREIMAN, J.H. FRIEDMAN, R.A. OLSHEN, C.J. STONE: Classification and Regression Trees (Wadsworth Statistics/Probability). Chapman & Hall, Boca Raton 1984.
17. E. GATNAR: Podejście wielomodelowe w zagadnieniach dyskryminacji i regresji. Wydawnictwo Naukowe PWN, Warszawa 2012.
18. R.A. FISHER: The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 1936, Vol. 7, No. 2, pp. 179–188.
19. M.J. NORUŠIS and SPSS Inc.: SPSS for Windows Base System Users Guide. Release 6.1. Prentice Hall, Englewood Cliffs 1994.
20. M. HAMROL, J. CHODAKOWSKI: Prognozowanie zagrożenia finansowego przedsiębiorstwa. Wartość predykcyjna polskich modeli analizy dyskryminacyjnej. *Badania Operacyjne i Decyzje* 2008, t. 3, ss. 17–32.
21. D.H. TRAN, A.W.M. NG, B.J.C. PERERA, S. BURN, P. DAVIS: Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal* 2006, Vol. 3, No. 6, pp. 175–184.
22. B. PRUSAK: Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw. Difin, Warszawa 2005.

Szelaĝ, B., Studzinski, J., Chmielowski, K., Lesnianska, A., Rojek, I. Forecasting the Sewage Inflow into a Treatment Plant Using Artificial Neural Networks and Linear Discriminant Analysis. *Ochrona Srodowiska* 2018, Vol. 40, No. 4, pp. 9–14.

Abstract: The paper presents the results of forecasting the sewage inflow into the municipal wastewater treatment plant in Rzeszow using multilayer perceptron neural networks. For the forecast model, the following independent variables were adopted: the measured inflow volume to the treatment plant from the previous days, the water level in the Wislok River (effluent receiver), the total daily precipitation and the daily water inflow into the network. The calculations led to conclusions that variables substantially affecting the prognostic capacity of the forecast model included the water level in the Wislok River, the volume of precipitation and the sewage inflow to the facility from the previous days. Additionally, the impact of individual

structural parameters of the model based on artificial neural networks on forecasting results was analyzed. The research conducted with the use of classification trees demonstrated that number of neurons in the hidden layer was influenced by the number of inputs to the model, while the type of activation function in the hidden and output layer was of minor importance which was confirmed by the data of prognostic value. The applicability of a linear discriminant analysis for assessment of prognostic ability of the constructed forecast models was also investigated. The results obtained demonstrated that the linear discriminant model might be an interesting assessment tool to select variables for the forecast model of sewage inflow to a treatment plant.

Keywords: Wastewater inflow, sewage treatment plant, water level, daily precipitation, artificial neural networks (ANN), discriminant model, mean square error, mean percentage error, relative error, residual relative error, importance coefficient.