

8

Raport Badawczy
Research Report

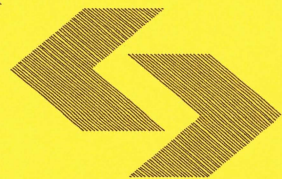
RB/50/2016

**Prognozowanie dopływu ścieków
surowych do oczyszczalni i
ładunku zanieczyszczeń w ściekach
za pomocą sieci neuronowych
i modeli operatorowych**

**L. Bogdan, J. Studziński,
I. Rojek, B. Szela**

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Dr hab. inż. Jan Studziński

Warszawa 2016

**PROGNOZOWANIE DOPŁYWU ŚCIEKÓW SUROWYCH DO
OCZYSZCZALNI I ŁADUNKU ZANIECZYSZCZEŃ W ŚCIEKACH ZA
POMOCA SIECI NEURONOWYCH I MODELI OPERATOROWYCH**

Lucyna Bogdan, Jan Studziński, Izabela Rojek, Bartosz Szeląg

IBS PAN Warszawa 2016

Spis treści

Wprowadzenie

- I. Prognozowanie wybranych wskaźników jakości ścieków na dopływie do oczyszczalni metodami Data Mining
- II. Modelling mixed liquor suspended solid and substrate load on the basis of wastewater quality indices and operational parameters of the bioreactor: data mining approach
- III. A Data Mining approach to the prediction of substrate load and mixed liquor suspended solid

Wprowadzenie

W raporcie zamieszczono trzy prace opracowane przez zespół autorów z Instytutu Badań Systemowych PAN, Politechniki Świętokrzyskiej w Kielcach i Uniwersytetu im. Kazimierza Wielkiego w Bydgoszczy.

Praca pierwsza pod tytułem *Prognozowanie wybranych wskaźników jakości ścieków na dopływie do oczyszczalni metodami Data Mining*, autorstwa Lucyny Bogdan i Jana Studzińskiego z IBS PAN oraz Bartosza Szelağa z Politechniki Świętokrzyskiej dotyczy modelowania wybranych wskaźników jakości ścieków w dopływie ścieków do oczyszczalni na podstawie ich pomiarów względnie na podstawie pomiarów przepływu ścieków.

Praca druga pod tytułem *Modelling mixed liquor suspended solid and substrate load on the basis of wastewater quality indices and operational parameters of the bioreactor: data mining approach*, autorstwa Izabeli Rojek z Uniwersytetu Kazimierza Wielkiego w Bydgoszczy oraz Jana Studzińskiego i Bartosza Szelağa, jest referatem zgłoszonym na międzynarodową konferencję pn. *Advances in Energy Systems and Environmental Engineering ASEE17*, organizowaną w dniach 2 – 5 lipca 2017 we Wrocławiu przez Politechnikę Wrocławską oraz University of New Mexico (USA) i Brunel University London (UK).

Praca trzecia pod tytułem *A Data Mining approach to the prediction of substrate load and mixed liquor suspended solid*, autorstwa Bartosza Szelağa i Jana Studzińskiego, jest artykułem złożonym do czasopisma *Polish Journal of Environmental Studies* w Olsztynie i znajdującym się obecnie na etapie recenzowania.

I. PROGNOZOWANIE WYBRANYCH WSKAŹNIKÓW JAKOŚCI ŚCIEKÓW NA DOPŁYWIE DO OCZYSZCZALNI METODAMI DATA MINING

Lucyna Bogdan, Jan Studziński, Bartosz Szeląg

Eksploatacja oczyszczalni ścieków jest złożonym zadaniem, wymaga bowiem zapewnienia wysokiego i względnie stabilnego stopnia oczyszczania ścieków pomimo dużej zmienności ilości i jakości ścieków surowych dopływających systemem kanalizacyjnym. Dlatego w celu utrzymania wymaganego efektu redukcji zanieczyszczeń w ściekach bardzo istotne jest opracowanie narzędzi umożliwiających przewidywanie zmian zarówno ilości jak i jakości ścieków dopływających, co umożliwi ewentualne działania wyprzedzające związane z korektą parametrów technologicznych w urządzeniach oczyszczających ścieki, głównie reaktorach biologicznych.

W nowoczesnych oczyszczalniach wartości wskaźników opisujących jakość ścieków mierzone są przy pomocy systemów pomiarowych on-line z założonym krokiem czasowym lub są określane na drodze badań laboratoryjnych. Jednakże w obu przypadkach mogą wystąpić problemy z uzyskaniem stałej rozdzielczości w szeregu czasowym danych pomiarowych. W praktyce prowadzi to do braków danych stanowiących wejście do modeli matematycznych umożliwiających optymalizację działania poszczególnych obiektów oczyszczalni. W związku z tym, w celu uzyskania wysokiej efektywności działania oczyszczalni ścieków celowe jest opracowywanie niezależnych systemów predykcji jakości oraz ilości ścieków dopływających.

Kluczowymi wskaźnikami jakości ścieków będącymi jednocześnie podstawą do oceny efektywności pracy oczyszczalni są: biochemical oxygen demand (BOD_5), chemical oxygen demand (COD), total suspended solids (TSS), total nitrogen (TN) and total phosphorus (TP). Do prognozy ilości i jakości ścieków w obrębie rozdzielczych i ogólnospławnych sieci

kanalizacyjnych opracowano fizyczne modele matematyczne bazujące na układach równań różniczkowych, ale ze względu na konieczność zgromadzenia szczegółowych informacji o zlewni i warunkach meteorologicznych mogą wystąpić problemy z ich kalibracją. Dlatego też, do modelowania ilości i jakości ścieków spływających ze zlewni do oczyszczalni stosuje się metody data mining. Na podstawie danych literaturowych (Abyaneh 2014, Delana i West, 2009; El-Din i Smith, 2002) można stwierdzić, że do symulacji stężenia zawiesiny ogólnej, azotu i fosforu ogólnego w ściekach dopływających do oczyszczalni można wykorzystać modele typu ARIMA jak i sztuczne sieci neuronowe. Verma i in. do prognozy stężenia zawiesiny ogólnej zastosowali szereg modeli, takich jak metoda wektorów nośnych (SVM), lasów losowych (RF), MARS (multivariate adaptive regression spline) i k-najbliższego sąsiada (k-NN) uzyskując tylko nieznaczne różnice w wartościach symulowanych i pomierzonych. Ponadto, analizy wykonane przez Kim i in. (2014) potwierdziły możliwość wykorzystania metody k-NN do predykcji wskaźników jakości ścieków na dopływie do oczyszczalni ścieków. Poza wymienionymi wyżej metodami do modelowania działania oczyszczalni (optymalizacja procesu nitrifikacji, denitrifikacji, defosfatacji, usuwania związków węgla itp.) stosowane są modele hybrydowe stanowiące połączenie dwóch modeli statystycznych (klasyfikacyjnego i regresyjnego lub kilku o charakterze regresyjnym).

W prezentowanej publikacji przedstawiono propozycję modelowania (przy pomocy trzech metod data mining) wartości wybranych wskaźników jakości ścieków (BOD_5 , COD, TSS, TN, TP) na dopływie do oczyszczalni w oparciu o ostatnie wyniki ich pomiarów poprzedzające prognozowaną wielkość oraz na podstawie zmierzonych wartości dobowego dopływu do obiektu. Mając powyższe na względzie, w prezentowanej pracy przedstawiono koncepcję niezależnych systemów predykcji jakości ścieków na dopływie do oczyszczalni na wypadek uszkodzenia systemów pomiarowych wskaźników zanieczyszczeń.

Do analiz wykorzystano 3-letnie dobowe szeregi czasowe natężenia przepływu i wskaźników jakości ścieków pomierzonych na dopływie do oczyszczalni ścieków w Rzeszowie.

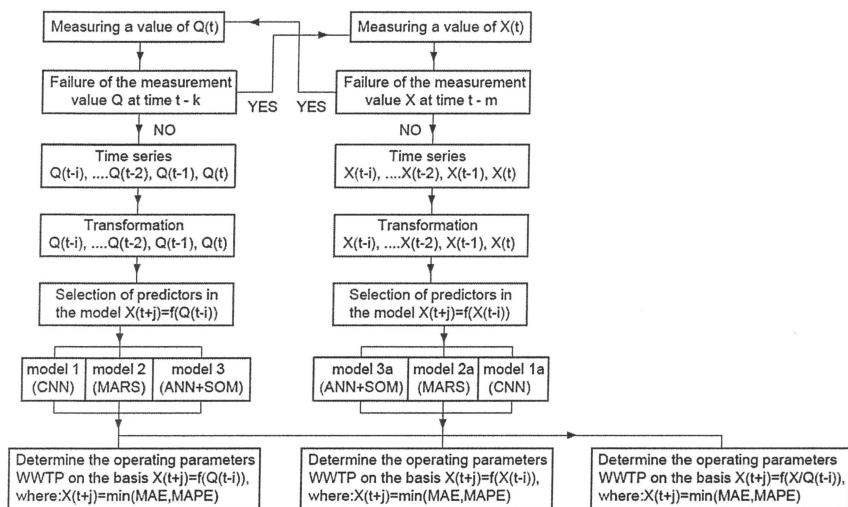
Obiekt badań

Rzeszów jest aglomeracją miejsko-przemysłową położoną w południowo-wschodniej części Polski. Posiada głównie system kanalizacji rozdzielczej (sanitarnej i deszczowej). Ścieki komunalne z terenu miasta i okolicznych miejscowości odprowadzane są do oczyszczalni mechaniczno-biologicznej o nominalnej przepustowości 62.500 m³/d. W ramach monitoringu ciągłego parametry opisujące ilość i jakość ścieków na dopływie i odpływie mierzone są od 2013 roku z dobową rozdzielczością.

Metodyka

W pracy przeprowadzono symulację dla dwóch przypadków. W pierwszym z nich rozpatrzono możliwość predykcji wskaźników jakości ścieków (BOD₅, COD, TSS, TN i TP) na podstawie ich ostatnich wyników pomiarów. W drugim przypadku przeanalizowano możliwość prognozowania wartości wymienionych wyżej wskaźników zanieczyszczeń w ściekach na podstawie wyników pomiarów natężenia przepływu na dopływie do oczyszczalni ścieków. Rozwiązanie to ma szereg zalet, bowiem w przypadku awarii systemu pomiarowego on-line, alternatywnie istnieje możliwość wykonania predykcji wybranych wskaźników jakości ścieków na dopływie do oczyszczalni na podstawie pomierzonych przepływów. W warunkach eksploatacji oczyszczalni pozwala to na uzyskanie ciągłości prognoz jakości ścieków na dopływie, a co za tym idzie daje także możliwość stabilnej pracy bioreaktorów przez odpowiedni dobór parametrów technologicznych (np. stężeń

tlenu i osadu czynnego, ilości dawkowanego PIX-u, stopnia recyrkulacji itp.).
 Koncepcję opisanego wyżej systemu przedstawiono na rys. 1.



Rys. 1. Schemat niezależnego systemu pomiarowo – predykcyjnego wskaźników jakości ścieków na dopływie do OŚ w celu predykcji nastaw obiektów technologicznych.

gdzie: i – przesunięcie czasowe między wartością prognozowaną a uwzględnioną w modelu do predykcji wybranego wskaźnika jakości ścieków; j – wyprzedzenie czasowe do przodu (w pracy przyjęto $j = 1$) z jakim jest prognozowany analizowany wskaźnik jakości; k, m – okres trwania uszkodzenia, w którym nie ma dobowych wartości wyników pomiarów Q, X .

Na tej podstawie opracowano modele statystyczne do prognozy wskaźników jakości ścieków przy zastosowaniu metody MARS i kaskadowych sieci neuronowych. Ponadto, wykonano model hybrydowy stanowiący połączenie sieci neuronowej typu perceptron wielowarstwowy i

samoorganizującej się sieci Kohonena. W celu zapewnienia prawidłowego procesu uczenia i właściwej oceny działania modelu matematycznego, wykonano 5-krotną cross-walidację, przy czym uprzednio dokonano podziału danych na zbiór uczący (75%), testowy i walidacyjny (25%). Przed przystąpieniem do wykonania modeli statystycznych dane wejściowe i wyjściowe poddano standaryzacji za pomocą przekształcenia min - max (Rutkowski 2006).

Metoda MARS jest jedną z wielu metod wykorzystywanych do rozwiązywania zagadnień o charakterze regresyjnym (Zhang and Goh, 2016; Sharda et. al., 2008) i jest rozszerzeniem klasycznego ujęcia danych wejściowych w dotychczas opracowanych modelach matematycznych. Poza całościowym ujęciem zmiennych objaśniających, jak ma to miejsce w klasycznym modelu regresji, w metodzie MARS zakresy zmienności branych pod uwagę predyktorów dzielone są na przedziały, gdzie analizowane zmienne mogą mieć różny wpływ na rozpatrywane zjawisko. Granice przedziałów określone są na podstawie wartości progowych (t_i). Rozróżnienie wartości analizowanych predyktorów na mniejsze i większe od wartości progowych t_i realizuje się na podstawie funkcji bazowej postaci:

$$h(X) = \alpha_i \cdot (\max(0, X - t)) \quad (1)$$

gdzie: $h(X)$ – wektor funkcji bazowych dla poszczególnych zmiennych (x_i), dla których spełniony jest następujący warunek:

$$x_i - t_i = \begin{cases} x_i - t_i; & \text{dla } x_i > t_i \\ 0; & \text{dla } x_i \leq t_i \end{cases} \quad (2)$$

W metodzie MARS zależność regresyjną stanowi funkcja sklejana otrzymana z kombinacji liniowej iloczynu funkcji bazowych z odpowiednimi wagami, którą można zapisać:

$$f(X) = \alpha_0 + \sum_{m=1}^M \alpha_m \cdot h_m(X) \quad (3)$$

gdzie: $X = [x_1, x_2, \dots, x_i]$ – wektor danych wejściowych, α_m – wartości wag, h_m – funkcje bazowe.

Do obliczenia wartości parametrów modelu, Friedman (1991) opracował specjalny algorytm umożliwiający przeszukiwanie obserwacji w celu ustalenia wartości progowych. Algorytm ten daje możliwość eliminowania zmiennych mających pomijalny wpływ na wynik symulacji.

Sztuczne sieci neuronowe

Jedną z najpowszechniej stosowanych struktur jest sieć wielowarstwowa jednokierunkowa znana pod nazwą perceptron wielowarstwowy, która była wielokrotnie wykorzystywana do modelowania ilości, jakości ścieków dopływających do oczyszczalni, jak również działania obiektów technologicznych. Jedną z modyfikacji ANN typu perceptron wielowarstwowy jest sieć kaskadowa (CNN) zawierająca dodatkowe połączenia wag z wejścia do każdej warstwy i z każdej warstwy do kolejnej. Poza klasycznymi sieciami neuronowymi z odpowiednio zmodyfikowaną strukturą opracowywane są modele hybrydowe stanowiące połączenie modelu klasyfikacyjnego z modelem regresyjnym, jak na przykład sieć neuronowa typu perceptron wielowarstwowy. Najpopularniejszym typem sieci określanej mianem samoorganizującej się jest sieć Kohonena (1982). Na etapie uczenia samoorganizujących się sieci neuronowych definiowana jest ilość neuronów tworzących warstwę topologiczną i ilość epok uczenia.

W prezentowanej pracy opracowano modele sieci neuronowej kaskadowej, gdzie analizowano CNN z dwoma i trzema warstwami ukrytymi. W poszczególnych warstwach ukrytych rozpatrywano różne funkcje aktywacji (tangens-hiperboliczny, liniowa, sinusoidalna, eksponentialna, logistyczna,

sigmoidalna), a w warstwie wyjściowej rozpatrzono funkcję liniową. Ponadto, w obliczeniach przyjęto, że liczba neuronów w kolejnych warstwach ukrytych jest identyczna. Optymalną liczbę neuronów i wartości wag zarówno w sieci MLP jak i CNN ustalono generując za każdym razem 500 różnych modeli dla założonych zmiennych objaśniających i wybierając wyuczoną sieć, dla której uzyskano najmniejsze wartości błędu bezwzględnego i względnego prognozy wartości analizowanych wskaźników jakości ścieków. W celu uzyskania modeli MLP jak i CNN o odpowiednich zdolnościach uogólniających na etapie uczenia zastosowano algorytm regularyzacji bayesowskiej. W rozpatrywanym przypadku jako klasyfikator w przypadku modelu hybrydowego zastosowano samoorganizujące się sieci neuronowe typu Kohonena o warstwie topologicznej 2x2, podczas gdy liczba epok na etapie uczenia wynosiła 1000.

W powyższych analizach w celu oceny zdolności predykcyjnej modeli prognozujących wartości BOD₅, COD, TSS, TN, TP zastosowano:

- błąd średni (MAE)

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_{i,obs} - y_{i,pred}| \quad (4)$$

- procentowy błąd procentowy (MPE)

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_{i,obs} - y_{i,pred}}{y_{i,obs}} \right| \cdot 100\% \quad (5)$$

- średni błąd kwadratowy (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_{i,obs} - y_{i,pred})^2} \quad (6)$$

gdzie: $y_{i,obs,obl}$ – odpowiednio pomierzone i obliczone wartości wskaźników jakości ścieków, n – liczebność zbioru, w rozpatrywanym przypadku zbior danych obejmuje 1100 wartości w szeregu czasowym.

Wyniki

Na podstawie zgromadzonych wyników pomiarów jakości i ilości ścieków dopływających do oczyszczalni ustalono zakres ich zmienności (tab. 1). Analiza danych z tab. 1, pozwala stwierdzić, że wartości BOD₅, COD, TSS, TN i TP zmieniały się w znacznym zakresie.

Tab. 1. Zakres zmienności wartości parametrów opisujących ilość i jakość ścieków dopływających do oczyszczalni w okresie 2013-2015.

Parametr	Minimalna	Maksymalna	Średnia
BOD ₅ , mg/dm ³	38.1	788.0	374.0
COD, mg/dm ³	159.0	2510.0	927.1
TSS, mg/dm ³	80.0	1140.0	430.0
TN, mg/dm ³	21.3	99.0	69.7
TP, mg/dm ³	3.4	37.5	12.45
Q, m ³ /d	26973	66773	38658

Ponieważ wymienione wskaźniki jakości ścieków stanowią dane wejściowe do modeli opisujących kinetykę przemian związków węgla, azotu i fosforu w reaktorach biologicznych, a ich wartości zmieniają się w znacznym zakresie, dlatego zachodzi potrzeba ich prognozowania. W obrębie analizowanych w niniejszej pracy metod jedynie w modelu MARS opracowany algorytm estymacji parametrów daje możliwość eliminacji predyktorów mających pomijalny wpływ na zmienną objaśnianą. Dlatego w pierwszej

kolejności wykonano symulację wybranych wskaźników jakości ścieków dopływających do oczyszczalni przy zastosowaniu tej metody. Na podstawie zmiennych objaśniających ustalonych przy wykorzystaniu metody MARS wykonano prognozę wartości BOD₅, COD, TSS, TN i TP pozostałymi metodami. Do tego celu wykorzystano wartości parametrów jakości ścieków zmierzonych w ostatnich wykonanych pomiarach. Następnie przeprowadzono symulację wskaźników jakości ścieków na dopływie do oczyszczalni w oparciu o pomiary natężenia przepływu.

Wykonane analizy wykazały, że do obliczenia wartości COD, TN i TP celowe jest uwzględnienie dwóch ostatnich pomiarów ($i=1,2$) rozpatrywanych wskaźników poprzedzających prognozowaną wielkość. Do predykcji TSS konieczne jest ujęcie wartości od TSS($t-1$) do TSS($t-4$), z kolei do prognozy BOD₅ wystarczą wartości analizowanego wskaźnika takie jak BOD($t-1$) i BOD($t-3$). Na podstawie wykonanych analiz stwierdzono, że do prognozy COD, TSS, TN i TP można wykorzystać wartości Q($t-1$) i Q($t-2$), z kolei do obliczenia BOD₅ celowe jest uwzględnienie wartości Q($t-1$), Q($t-2$) i Q($t-3$).

Uzyskane wyniki analiz metodą MARS znajdują potwierdzenie w badaniach wykonanych przez Szelağa et. al. (2016) metodą drzew klasyfikacyjnych. W tab. 2 i 3 zestawiono parametry opisujące strukturę uzyskanych modeli matematycznych. Ponadto, w tab. 4-6 zamieszczono wielkości parametrów dopasowania (MAE i MAPE) wyników pomiarów do obliczeń metodą MARS, CNN i ANN+SOM wybranych wskaźników jakości ścieków dla zbioru testowego.

Na podstawie danych z tab. 2 można stwierdzić, że w większości otrzymanych modeli statystycznych metodą kaskadowych sieci neuronowych liczba warstw ukrytych była równa 3, a występująca w nich liczba neuronów wynosiła od 2 do 6. W przypadku modeli do predykcji wskaźników jakości w oparciu o X($t-i$) w warstwie ukrytej najlepsze wyniki obliczeń uzyskano dla

funkcji tangensa-hiperboliczny, natomiast stosując do modelowania stężenia związków organicznych, azotu i fosforu najmniejsze wartości błędów MAE i MAPE otrzymano dla funkcji logistycznej. Dla hybrydowych sieci neuronowych (tab. 3) liczba neuronów w warstwie ukrytej w otrzymanych modelach statystycznych do predykcji wskaźników jakości w wydzielonych klasach zmieniała się w zakresie 2-6, a funkcję aktywacji najczęściej stanowiła zależność tangensa-hiperbolicznego.

Tab. 2. Parametry otrzymanych modeli w oparciu o kaskadowe sieci neuronowe do prognozy wybranych wskaźników jakości ścieków.

Parametr	X(t-i)			Q(t-i)		
	l. neuron	l. warstw	f. aktywacji	l. neuron	l. warstw	f. aktywacji
BOD ₅	5	3	tanh	4	2	log
COD	3	3	tanh	3	3	tanh
TSS	3	2	tanh	2	3	log
TN	6	3	log	6	3	log
TP	4	3	tanh	5	3	tanh

Na podstawie danych zamieszczonych w tab. 3 - 5 można stwierdzić, że lepszymi zdolnościami predykcyjnymi zmiennych niezależnych (mniejsze wartości błędu MAE i MAPE) charakteryzują się w obu przypadkach (MARS, CNN) modele opracowane na podstawie wartości wskaźników jakości ścieków. Z kolei, modele statystyczne uzyskane metodą ANN+SOM do obliczenia BOD₅, TSS i TN na podstawie Q(t-i) charakteryzują się lepszymi zdolnościami predykcyjnymi wartości wskaźników jakości ścieków.

Tab. 3. Parametry otrzymanych modeli w oparciu o hybrydowe sieci neuronowe (ANN+SOM) do prognozy wybranych wskaźników jakości ścieków.

Parametry	X(t-i)		Q(t-i)	
	l. neuron	f. aktywacji	l. neuron	f. aktywacji
BOD ₅	3 ÷ 6	tanh, lin	3 ÷ 6	lin, sigm, exp
COD	2 ÷ 6	exp, lin	2 ÷ 5	exp, sigm
TSS	3 ÷ 4	exp, tahn	2 ÷ 5	simg, tanh
TN	2 ÷ 5	tanh, exp	3 ÷ 6	exp, tanh
TP	3 ÷ 6	exp, sigm	3 ÷ 5	simg, tanh

Tab. 4. Wartości parametrów dopasowania (MAE, MAPE) wyników obliczeń wskaźników jakości ścieków metodą MARS do pomiarów wyznaczonych w oparciu o wartości X(t-i), Q(t-i) określone w ostatnich wykonanych pomiarach.

Parametry	X(t-i)			Q(t-i)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
BOD ₅	41.12	60.31	12.39	48.1	62.48	14.8
COD	117.54	149.28	14.06	133.02	166.28	15.89
TSS	59.01	91.04	15.9	71.88	99.31	20.07
TN	6.07	8.96	9.64	7.04	8.88	11.43
TP	1.61	2.57	13.97	2.16	2.97	19.03

Tab. 5. Wartości parametrów dopasowania (MAE, MAPE) wyników obliczeń wskaźników jakości ścieków metodą CNN do pomiarów wyznaczonych w oparciu o $X(t-i)$, $Q(t-i)$ ustalone w poprzednich pomiarach.

Parametry	X(t-i)			Q(t-i)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
BOD ₅	33.07	45.73	9.67	37.29	49.5	11.29
COD	99.56	124.45	10.26	112.45	151.81	13.14
TSS	52	77.65	13.89	60.9	88.15	16.61
TN	5.45	6.22	8.53	5.76	7.06	8.96
TP	1.4	1.74	12.19	1.76	2.05	15.54

Tab. 6. Wartości dopasowania (MAE, MAPE) wyników obliczeń wybranych wskaźników jakości ścieków za pomocą modelu hybrydowego (ANN+SOM) do wyników pomiarów

Parametry	X(t-i)			Q(t-i)		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE
BOD ₅	30.16	40.15	8.75	27.93	45.02	7.62
COD	93.29	116.61	11.34	92.50	124.88	10.93
TSS	49.05	70.12	13.09	50.89	80.45	14.62
TN	5.00	5.56	7.83	4.38	6.26	6.42
TP	1.72	1.53	15.11	1.79	1.83	15.75

Mając na uwadze czas wykonywania pojedynczego oznaczenia BOD₅ i uzyskane wyniki symulacji wskaźników jakości ścieków stwierdzono, że

modele autoregresyjne uwzględniające wartości BOD(t-i) nie mogą być wykorzystane w rozważaniach praktycznych. Ponadto, wielkości błędów dla otrzymanych modeli na podstawie wartości BOD(t-i) i Q(t-i) zmieniają się odpowiednio w zakresach MAE=30,16-41,12 mg/dm³ i MAPE=8,75-12,39% oraz MAE=27,93-48,10 mg/dm³ i MAPE=7,62-14,80%.

W przypadku modeli do predykcji COD uzyskanych na podstawie COD(t-i) oraz wyłączenie Q(t-i) wartości błędów prognozy wskaźnika zmieniały się w zakresach MAE=93,29-117,54 mg/dm³ i MAPE=10,26-14,06% oraz MAE=92,50-133,02 mg/dm³ i MAPE=10,93-15,89%. Dla modeli do prognozy TSS uwzględniających TSS(t-i) i tylko Q(t-i) wielkości błędów symulacji analizowanego wskaźnika jakości ścieków wynosiły odpowiednio MAE=49,05-59,01 mg/dm³ i MAPE=13,09-15,90% oraz MAE=50,89-71,88 mg/dm³ i MAPE=14,62-20,07%.

Rozpatrując modele do predykcji TN opracowane w oparciu o TN(t-i) i wyłącznie Q(t-i) stwierdzono, że wartości błędu bezwzględnego i względnego są równe MAE=5,00-6,07 mg/dm³ i MAPE=7,83-9,64% oraz MAE=4,38-7,04 mg/dm³; MAPE=6,42-11,42%. Ponadto, z danych w tab. 3-5 wynika, że dla modeli do prognozy TP uzyskanych na podstawie TP(t-i) i tylko Q(t-i) błędy predykcji wskaźnika zmieniają się w zakresach MAE=1,40-1,72 mg/dm³ i MAPE=12,19-15,11% oraz MAE=1,76-2,16 mg/dm³ i MAPE=15,54-19,03%.

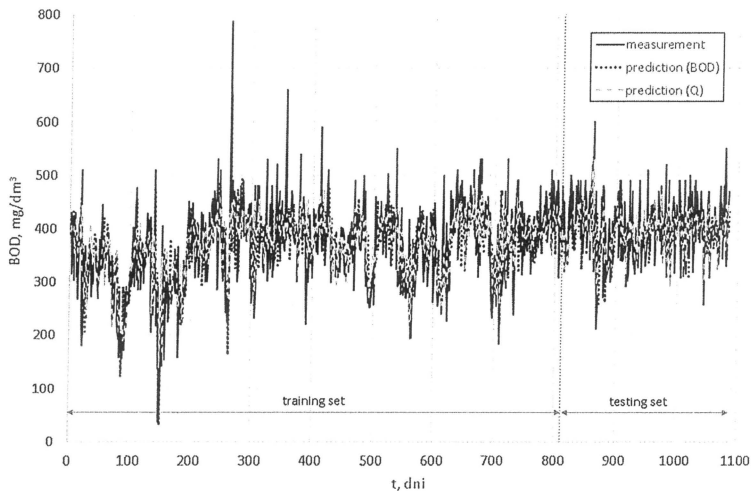
Z wykonanych prognoz (tab. 3-5) wartości wskaźników jakości ścieków (BOD₅, COD, TSS, TN i TP) stwierdzono, że największe błędy predykcji wymienionych wyżej parametrów uzyskano metodą MARS. Natomiast, najlepszą zgodność wyników obliczeń z pomiarami wybranych wskaźników jakości ścieków (BOD₅, COD, TSS, TN) otrzymano metodą ANN+SOM. Jedynie w przypadku TP najmniejsze wartości błędów prognozy wskaźnika otrzymano metodą kaskadowych sieci neuronowych. Na podstawie wykonanych obliczeń metodą ANN+SOM stwierdzono, że na wypadek awarii sond

pomiarowych prognozowane wartości COD, TSS, TN i TP w oparciu o $Q(t-i)$ różnią się nie więcej niż 5% od wyników otrzymanych za pomocą modeli, gdzie zmienne objaśniające stanowiły wskaźniki jakości ścieków. W przypadku wielkości BOD_5 stwierdzono podobną zależność jak wyżej.

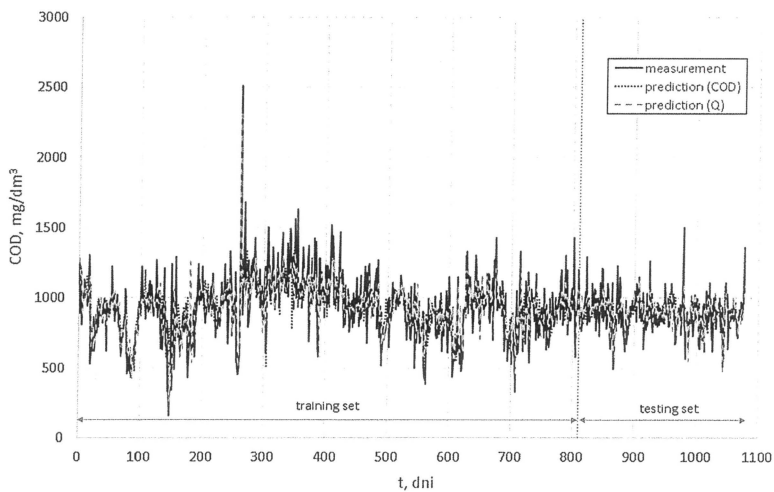
Porównując uzyskane wartości błędów (RMSE) predykcji BOD_5 metodą ANN+SOM (tab. 6) można stwierdzić, że są one znacznie większe niż otrzymane przez Abyaneh (2014) ($RMSE = 25,10 \text{ mg/dm}^3$) metodą ANN, który do prognozy rozpatrywanego wskaźnika jakości wykorzystał temperaturę, pH dopływających ścieków oraz stężenie TSS. Obliczone wartości błędów RMSE prognozy wielkości COD metodą ANN+SOM na podstawie $COD(t-i)$ oraz $Q(t-i)$ wykazują, że są one wielokrotnie większe niż otrzymane przez Abyaneh (2014) ($RMSE = 49,4 \text{ mg/dm}^3$) metodą ANN i Minsoo et. al. (2016).

W ostatnim przypadku, autorzy stosując metodę k-najbliższych sąsiadów stwierdzili, że błędy prognozy COD w przypadku prognozy bezdeszczowej i deszczowej zmieniają się odpowiednio w zakresach $RMSE = 4-5 \text{ mg/dm}^3$ i $RMSE = 6-8 \text{ mg/dm}^3$.

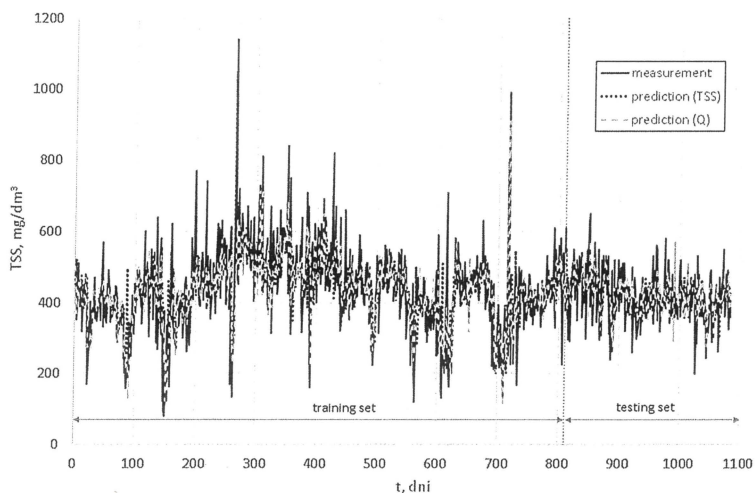
Na podstawie wyznaczonych wartości błędów (MAE) prognozy zawiesiny ogólnej stwierdzono, że są one większe niż uzyskane przez Verma i Kusiaka (2013) ($RMSE = 28,81 \text{ mg/dm}^3$), którzy do symulacji metodą ANN wartości TSS wykorzystali wyniki pomiarów $CBOD_5$ oraz przepływów. Znacznie większe wielkości błędów otrzymali oni metodą k-najbliższego sąsiada wykorzystując do prognozy TSS wartości $TSS(t-i)$, uzyskując wartość błędu równą $MAE = 62,15 \text{ mg/dm}^3$. Wyniki obliczeń wartości pozostałych rozpatrywanych wskaźników (TN, TP) metodami (MARS, CNN, ANN+SOM) wykazały, że są one większe niż otrzymane przez innych badaczy (Minsoo et. al., 2016).



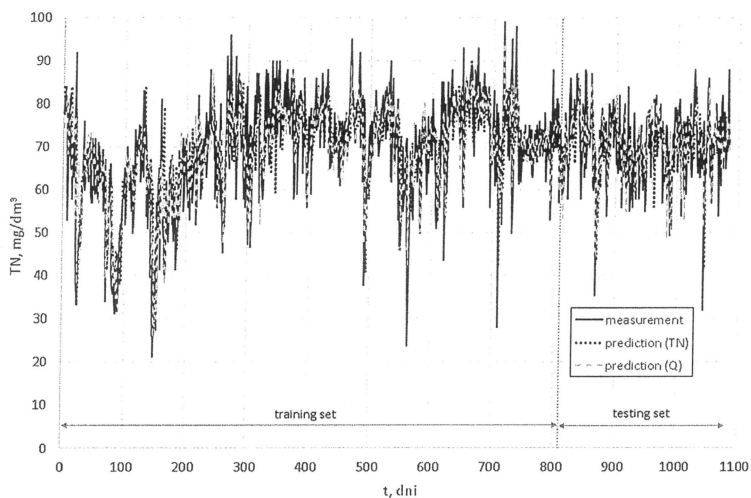
Rys. 2. Porównanie wyników obliczeń modelami hybrydowymi i pomiarów BOD₅ w rozpatrywanym okresie.



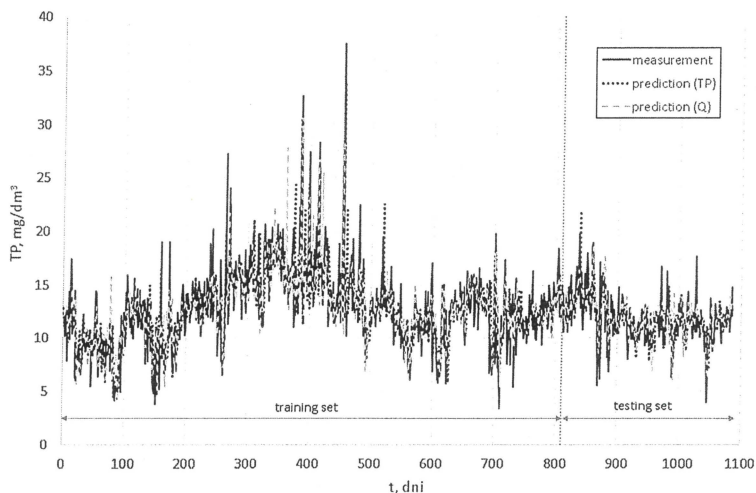
Rys. 3. Porównanie wyników obliczeń modelami hybrydowymi i pomiarów COD w rozpatrywanym okresie.



Rys. 4. Porównanie wyników obliczeń modelami hybrydowymi i pomiarów TSS w rozpatrywanym okresie.



Rys. 5. Porównanie wyników obliczeń modelami hybrydowymi i pomiarów TN w rozpatrywanym okresie.



Rys. 6. Porównanie wyników obliczeń modelami CNN i pomiarów TP w rozpatrywanym okresie.

Mając na względzie powyższe uwagi, można stwierdzić, że uzyskane modele statystyczne charakteryzowały się gorszymi zdolnościami predykcyjnymi niż uzyskane przez innych badaczy, bowiem nie uwzględniały wszystkich predyktorów. Jednakże, w analizach wykonanych przez Abyaneh (2014), Verma i Kusiak (2013), Minsoo i inni (2016) ilość zmiennych jest znacznie większa niż w opracowanych modelach przez co wymagają prowadzenia ciągłego wyników pomiarów. W przypadku modernizacji lub awarii układu pomiarowego wskaźników jakości ścieków może być to niemożliwe, dlatego też przedstawione wyżej podejście może stanowić okresowe rozwiązanie, które w znaczący sposób warunkuje uzyskanie wymaganego efektu redukcji zanieczyszczeń w oczyszczanych ściekach.

Podsumowanie

W publikacji przedstawiono możliwość modelowania wartości wybranych wskaźników (BOD_5 , COD, TSS, TN, TP) jakości ścieków na dopływie do oczyszczalni na podstawie wartości wskaźników pomierzonych w ostatnich wykonanych pomiarach, ale również wyłącznie w oparciu o wartości natężenia przepływu. Wykonane obliczenia wykazały, że mniejsze wartości błędów prognozy stężenia biochemicznego i chemicznego zapotrzebowania na tlen, azotu, fosforu i zawiesiny otrzymano za pomocą modelu statystycznego opracowanego metodą kaskadowych sieci neuronowych niż metodą MARS. W obrębie rozpatrywanych metod w większości rozpatrywanych przypadków najlepszymi zdolnościami predykcji wskaźników jakości ścieków charakteryzowały się modele hybrydowe, stanowiące połączenie sieci neuronowej Kohonena i perceptrona wielowarstwowego.

Z przeprowadzonych obliczeń wynika, że nieznacznie lepsze rezultaty symulacji rozpatrywanych wskaźników jakości otrzymano zarówno metodą MARS jak i CNN w przypadku gdy zmienne objaśniające rozpatrywane parametry stanowiły odpowiednio wielkości wskaźników jakości ścieków aniżeli wartości przepływów. W przypadku modeli hybrydowych wartości błędów prognozy analizowanych wskaźników jakości ścieków były mniejsze niż dla modeli uzyskanych metodą MARS i CNN. Jedynie dla modelu do obliczenia TP przy pomocy $Q(t-i)$ mniejsze wartości błędu MAE i MAPE otrzymano za pomocą modelu bazującego na CNN niż ANN+SOM.

Pozwoliło to na opracowanie koncepcji systemu predykcji jakości ścieków, w którym prognozy wybranych wskaźników wykonywane są niezależnie, na podstawie stężenia związków organicznych, zawiesiny, azotu i fosforu oraz oddzielnie na podstawie pomierzonych wielkości natężenia przepływu. Pozwala to na uzyskanie ciągłości prognoz jakości dopływających

ścieków nawet w przypadku uszkodzenia przepływomierza lub aparatury do pomiaru jakości ścieków.

Wykazano, że w przypadku analizowanej oczyszczalni ścieków możliwe jest zapewnienie ciągłości prognoz wartości wybranych wskaźników zanieczyszczeń na podstawie opracowanych dwóch niezależnych systemów do predykcji jakości ścieków. Może to zapewnić niezawodne funkcjonowanie obiektów oczyszczalni w sytuacji awarii systemu monitoringu jakości ścieków na dopływie lub jego modernizacji.

Literatura

- Rutkowski L. (2006). Artificial intelligence methods and techniques (in Polish). Warszawa, PWN, 2006.
- Friedman J.H.: (2002). Stochastic gradient boosted. *Computational Statistics and Data Analysis*. Vol. 38, No. 4, 367-378.
- Friedman J.: (1991). Multivariate Adaptive Regression Splines. *Annals of Statistics*. Vol. 19, 1-141.
- Zhang W., Goh A.T.C.: (2016). Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*. Vol. 7, No. 1, 45-52.
- De Veaux R.D., Psychogios D.C., Ungar L.H.: (1993). A Comparison of two nonparametric estimation schemes: MARS and neural networks. *Computers & Chemical Engineering*. Vol. 17, No. 8, 819-837.
- Gutiérrez G., Schnabel Á.S., Contador J.F.L.: (2009). Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies. *Ecological Modelling*. Vol. 220, No. 24, 3630-3637.

- Verma A., Wei X., Kusiak A.: (2013). Predicting the total suspended solids in wastewater: A data-mining approach. *Engineering Applications of Artificial Intelligence*. Vol. 26, No. 4, 1366-1372.
- Minsoo K., Yejin K., Hyosoo K., Wenhua P., Changwon K.: (2016). Evaluation of the k – nearest neighbour method for forecasting the influent characteristics of wastewater treatment plant. *Frontiers of Environmental Science & Engineering*. Vol. 10, No. 2, 299-310.
- Solaimany-Aminabad M., Maleki A., Hadi M.: (2013). Application of artificial neural network (ANN) for prediction of water treatment plant influent characteristics. *Journal of Advance in Environmental Health Research*. Vol. 1, No. 2, 89-100.
- Dellana S.A., West D.: (2009). Predictive modeling for wastewater applications: Linear and nonlinear approaches. *Environmental Modelling and Software*. Vol. 24, 96-106.
- Sobota J., Szetela R.: (2005). One-dimensional models of quantity and quality of wastewater flowing to the urban treatment plant (in Polish). *Ochrona Środowiska*. Vol. 27, No. 1, 15-22.
- Szeląg B., Bartkiewicz L., Studziński J. (2016). Prognozowanie wybranych wskaźników zanieczyszczeń na dopływie do oczyszczalni metodami czarnej skrzynki. *Ochrona Środowiska*. (w druku).
- Szeląg B., Gawdzik J. (2016). Application of selected methods artificial intelligence to activated sludge settleability predictions. *Polish Journal of Environmental Studies*. Vol. 25, No. 4, pp. 1709-1714.



