



Zastosowanie wybranych modeli nieliniowych do prognozy ilości osadu nadmiernego

*Jarosław Gawdzik, Bartosz Szelaǳ,
Elżbieta Bezak-Mazur, Renata Stoińska
Politechnika Świętokrzyska*

1. Wstęp

Podczas procesów oczyszczania ścieków metodą osadu czynnego powstają produkty uboczne – odpady takie, jak skratki, piasek, a także tłuszcze oraz osady. Pośród nich największą ilość, czyli około 2-3% objętości ścieków, stanowią osady ściekowe. Osady ściekowe to organiczno-mineralna faza stała, wyodrębniona ze ścieków. W procesie sedymentacji w osadnikach wstępnych otrzymuje się osad wstępny. Osad wtórny wydzielany jest w osadnikach wtórnych. Osady te zawraca się do obiegu oczyszczania ścieków jako osady recykulowane lub usuwa z obiegu do dalszej przeróbki jako osad nadmierny (Gawdzik i.in. 2015). Przystępując do optymalizacji pracy układu oczyszczania ścieków należy w szczególności zwrócić uwagę na regulację wielkości recykulatów, które mogą zasadniczo wpływać na efektywność procesów biologicznej defosfatacji i denitryfikacji. Część osadów wtórnych nie jest już potrzebna (osad nadmierny, tzn. dodatkowy osad lub zużyty osad czynny). Ich ilość zależy od wieku osadu, wykorzystania substratów takich jak metanol, strącania fosforu, biologicznego usuwania fosforu oraz wcześniejszego oczyszczania, takiego jak biologiczna filtracja. Strącanie chemiczne fosforu prowadzi do wzrostu zawartości substancji nieorganicznych w osadach ściekowych (Rydzziński 2012, Bień i in. 2001). Stosowane są zróżnicowane metody ustalania ilości niezbędnego do usunięcia osadu nadmiernego, np: metoda stałego stężenia osadu czynnego, metoda stałego obciążenia osadu ładunkiem BZT₅, metoda stałego wieku osadu czy

też metoda utrzymania stałej jakości osadu czynnego (Bieł i in. 2001). Wobec powyższego wysoce celowe jest modelowanie dobowej ilości osadów nadmiernych w dynamicznym układzie jakim jest biologiczny reaktor podwyższonego usuwania związków biogenych.

Do określenia ilości osadu nadmiernego na etapie eksploatacji obiektu czy też oceny redukcji związków biogenych zawartych w ściekach można zastosować model fizyczny (Henze i in. 2000, Serón i in. 2011) bądź model bazujący w oparciu o wyniki pomiarów na istniejących obiektach (ATV – A 131P, Tao i in. 2013). Jednakże, ze względu na złożoność modeli fizycznych mogą występować problemy ze znalezieniem jednego zestawu parametrów warunkujących zadowalające rozwiązanie, na co wskazują liczne publikacje (Martin i in. 2010, Kiczko i in. 2013). Natomiast, zależności ustalone w oparciu o wyniki badań na innych oczyszczalniach mogą być stosowane wyłącznie na podobnych obiektach, gdzie sposób oczyszczania ścieków pod względem technologicznym jest zbliżony. Poza aspektami technologicznymi istotne jest również aby charakterystyki zlewni (Licznar i Szela 2014) oraz wskaźniki jakości dopływających ścieków były porównywalne. Dlatego też w celu oceny funkcjonowania oczyszczalni ścieków sporządza się również modele tzw. czarnej skrzynki. W celu optymalizacji działania oczyszczalni stosuje się również modele hybrydowe, w których ilość i jakość ścieków stanowiąca dane wejściowe do modelu fizycznego jest modelowana przy pomocy modeli statystycznych. Jednakże, jak wykazały liczne prace (Raha 2007, Studzinski i in. 2013) dokładność prognozy parametrów na dopływie do OŚ jest ograniczona przez co w przypadkach tych jedynym rozwiązaniem może być opracowanie modelu statystycznego. Jednym z najczęściej stosowanych modeli czarnej skrzynki są sztuczne sieci neuronowe (Delana i in. 2009), przy czym ze względu na rozwój technik obliczeniowych opracowano inne metody, takie jak na przykład wektorów nośnych, lasów losowych, drzew wzmacnianych (Wei i Kusiak 2015), k – najbliższego sąsiada, które były wielokrotnie wykorzystywane przy prognozowaniu procesów zachodzących na oczyszczalni.

W artykule przedstawiono zastosowanie i wyniki obliczeń ilości osadu nadmiernego przy pomocy metody wektorów nośnych, k – najbliższego sąsiada oraz drzew wzmacnianych. Mając na względzie powyższe uwagi uzasadnione jest opracowanie modeli matematycznych do prognozowania ilości osadu nadmiernego w dowolnym horyzoncie czasowym.

2. Opis obiektu badań

Obiektem prowadzonych analiz była oczyszczalnia ścieków komunalnych zlokalizowana na terenie gminy Sitkówka – Nowiny, do której doprowadzane są ścieki sanitarne z rozdzielczego sytemu kanalizacji miasta Kielce, gminy Sitkówka – Nowiny i części gminy Masłów. Nominalna przepustowość oczyszczalni wynosi 72.000 m³/d przy obciążeniu 275.000 RLM.

Ścieki dopływające do oczyszczalni są podczyszczane mechanicznie na kratkach schodkowych i piaskownikach napowietrzanych z wydzielonym usuwaniem tłuszczów. W następnej kolejności ścieki tłoczone są do czterech osadników wstępnych. Po części mechanicznej oczyszczalni ścieki przepływają do części biologicznej. W komorach denitryfikacji wstępnej zachodzi częściowe usuwanie związków azotu ze ścieków. Następnie ścieki kierowane są do komór defosfatacji, gdzie ma miejsce usuwanie związków fosforu metodami chemicznymi i biologicznymi. Najistotniejszy element oczyszczalni stanowi reaktor biologiczny z wydzielonymi komorami denitryfikacji i nitryfikacji, w którym zachodzi ostateczne usuwanie zanieczyszczeń ze ścieków. Ścieki wraz z osadem czynnym przepływają do czterech osadników wtórnych, skąd po sklarowaniu odpływają do odbiornika – rzeki Bobrzy.

W ramach monitoringu ciągłego prowadzonego przez przedsiębiorstwo Wodociągi Kieleckie Sp. z o.o. od 2012 roku na terenie oczyszczalni określana jest ilość i jakość ścieków dopływających do OŚ oraz mierzone są kluczowe parametry technologiczne reaktora biologicznego.

3. Metodyka

W ramach prowadzonych analiz do obliczeń 1-7 dniowej ilości osadu nadmiernego (m_{os}) powstającego w trakcie oczyszczania ścieków w oparciu o dane wejściowe (Tao i.in. 2013) opisujące ilość (dobowy dopływ ścieków), jakość ścieków (BZT₅, zawiesina, $N_{całk}$, P_{og}) oraz parametry działania reaktora biologicznego (stężenie i temperatura osadu, recyrkulacja zewnątrz, ilość dawkowanego metanolu i PIX) zastosowano metodę wektorów nośnych, najbliższego sąsiada oraz drzew wzmocnianych. Na uwagę zasługuje fakt, że w obliczeniach tych ujęto wartości większości rozpatrywanych wyżej parametrów z co najmniej dobowym opóźnieniem w odniesieniu do wartości prognozowanej. W przypadku

stężeń związków biogenych w tym również zawiesiny w modelu ujęto ich wartości z poprzedniej doby i tygodnia. Ze względu na to, że nie dysponowano wynikami pomiarów jakości ścieków dopływających bezpośrednio do komór osadu czynnego (stężenie BZT₅, zawiesin, fosforu), nie było możliwe oszacowanie ilości osadów nadmiernych powstających na skutek rozkładu węgla z zależności empirycznej przytoczonej w wytycznej ATV – A 131 P (2001). Nie było tym samym możliwe oszacowanie ilości powstających osadów na drodze strącania chemicznego (Heidrich i Witkowski 2005) a co za tym idzie wyznaczenia całkowitego strumienia generowanych osadów.

Przed przystąpieniem do analiz wartości liczbowe analizowanych poddano normalizacji przy pomocy zależności (1):

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

gdzie:

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

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

A – wartość maksymalna pojedynczego elementu w zbiorze parametru A , min

A – wartość minimalna pojedynczego elementu w zbiorze parametru A .

Metoda wektorów nośnych (Vapnik 1998) została pierwotnie stosowana do rozwiązywania problemów dyskryminacyjnych, jednakże z tego powodu, że okazała się efektywna i szybka obliczeniowo, została ona zaimplementowana do zagadnień regresyjnych (Vapnik 1998). Z tego powodu, że związek między zmienną zależną Y a zmiennymi niezależnymi x może mieć charakter nieliniowy, w metodzie wektorów nośnych (SVM) stosuje się nieliniową transformację $\phi: X^L \rightarrow Z$ obserwacji ze zbioru uczącego do przestrzeni o dużo większym wymiarze, w której poszukuje się funkcji liniowej.

Metoda najbliższego sąsiada (kNN) należy do jednych z najprostszych metod nieparametrycznych i podobnie jak wyżej wymienione może być stosowana w zagadnieniach klasyfikacyjnych i regresyjnych (Piotrowski et al. 2006). Prognozę wartości zmiennej objaśnianej w niniejszym przypadku określa się ze wzoru (2):

$$\hat{y} = \frac{1}{K} \cdot \sum_{i=1}^N y_i J(x_i, x_j) \quad (2)$$

w którym:

x_i jest jednym z K najbliższych sąsiadów x_j gdy odległość $d(x_i, x_j)$ należy do najmniejszych odległości między obserwacjami, N – ilość obserwacji, $J(x_i, x_j)$ – funkcja postaci:

$$J(x_i, x_j) = \begin{cases} 1, & \text{gdy } x_i \text{ jest jednym z } K \text{ najbliższych sąsiadów } x_j \\ 0, & \text{w przeciwnym przypadku} \end{cases}$$

Najczęściej w powyższych obliczeniach wykorzystuje się odległości Euklidesa i Mahalanobisa.

Algorytm drzew wzmacnianych (BT) stanowi połączenie metody drzew regresyjnych i metody wzmacniania (Friedman 2001, 2002). Jest to jedna z najskuteczniejszych metod modelowania pozwalająca na odwołanie złożonych zależności z dużą szybkością i odpornością na niską jakość danych. Drzewa wzmacniane stanowią zespół prostych modeli (drzew regresyjnych) przy czym każdy model jest tworzony na próbie, w której większe wagi są przypisane poszczególnym obserwacjom, dla których dotychczasowe modele miały największy błąd.

Do oceny zdolności predykcyjnej uzyskanych modeli do obliczeń ilości osadu nadmiernego zastosowano następujące parametry (3 i 4):

- średni błąd bezwzględny:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |y_{i,obs} - y_{i,pred}| \quad (3)$$

- średni błąd względny:

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_{i,obs} - y_{i,pred}}{y_{i,obs}} \right| \cdot 100\% \quad (4)$$

gdzie:

n – liczebność zbioru danych,

$y_{i,obs}$ – pomierzona ilość osadu nadmiernego,

$y_{i,obl}$ – obliczona ilość osadu,

$\bar{y}_{obs, pred}$ – średnia arytmetyczna pomierzonej/obliczonej ilości osadu nadmiernego.

W rozważaniach praktycznych analizując modele matematyczne, najlepszymi zdolnościami predykcyjnymi charakteryzuje się ten w przypadku, którego uzyskano najmniejsze wartości średniego błędu bezwzględnego i względnego.

4. Wyniki

W celu wyznaczenia zakresu stosowalności opracowanych modeli matematycznych w tabeli 1 zamieszczono wartości średnie i zmienność rozpatrywanych parametrów funkcjonowania reaktora biologicznego, jakości i ilości dopływających ścieków do oczyszczalni, natomiast w tabeli 2 zestawiono ilość powstających osadów nadmiernych w okresie 1-7 dni.

Na podstawie tabeli 1 można stwierdzić, że rozpatrywane zmienne w okresie badawczym wykazywały znaczne zróżnicowanie co prowadziło do zmienności dobowej ilości osadu nadmiernego w zakresie 340-2122 m³, podczas gdy wartość średnia wynosiła 1336 m³.

W analizowanym okresie (2012-2015) stężenie azotu całkowitego na dopływie zmieniało się w znacznym zakresie (39,91-124,09 mg/dm³), fosforu ogólnego (4,3-12,6 mg/dm³), BZT₅ (127-557 mg O₂/dm³), ChZT (384-1250 mg O₂/dm³). W analizowanym okresie, dobowa ilość ścieków dopływających do OŚ i stężenie osadu czynnego wynosiły odpowiednio 32564-86592 m³/d i 1,19-5,89 kg/m³. Tak znaczny zakres zmian jakości i ilości ścieków na dopływie i parametrów działania reaktora biologicznego prowadził do znacznej zmienności obciążenia substratowego (0,062-0,312 g BZT₅/g s.m.o·d).

Natomiast, na podstawie danych w tabeli 2 wynika, że zarówno dobowa jak i tygodniowa ilość osadów nadmiernych powstająca na rozpatrywanym obiekcie zmienia się w znacznym zakresie co dodatkowo potwierdza potrzebę jej modelowania. Znaczna zmienność ilości osadów nadmiernych generowana w poszczególnych okresach (1-7 dni) potwierdza znaczne zróżnicowanie ilości i jakości ścieków dopływających do oczyszczalni. Przykładowo, dobowa ilość powstających osadów nadmiernych zmienia się w zakresie 340-2122 m³, natomiast tygodniowa wynosi od 3597 m³ do 14573 m³. W oparciu o opisane wyżej zmienne objaśniające ilość powstającego osadu nadmiernego w okresie (t = 1-7 dni) opracowano modele matematyczne, dla których parametry opisujące dopasowanie wyników obliczeń do pomiarów zamieszczono w tabeli 3 i 4.

Tabela 1. Zestawienie zakresu i wartości średnich analizowanych parametrów opisujących ilość, jakość ścieków na dopływie oraz parametrów działania reaktora biologicznego

Table 1. Summary of the range and mean values of the analyzed parameters describing the quantity, quality of wastewater influent and performance of the biological reactor

Zmienna	Zakres	Średnia
T_{os} , °C	10-23	15,0
Q ; m ³ /d	32564-86592	42004
$N_{całk}$; mg/dm ³	39,91-124,09	77,6
P_{og} ; mg/dm ³	4,3-12,6	7,8
BZT ₅ ; mg O ₂ /dm ³	127-557	320
ChZT; mg O ₂ /dm ³	384-1250	790
s_{zaw} , mg/dm ³	126-572	330
F/M ; gBZT ₅ /g .sm.o .d	0,062-0,312	0,138
S_{os} ; kg/m ³	1,19-5,89	3,57
% rec	44,59-167,55	91,40
m_{met} ; m ³	0,00-4,56	1,28
m_{PIX} ; m ³	0,00-7,93	0,77

gdzie:

Q – dobowe natężenie przepływu,

BZT₅ – pięciodniowe biochemiczne zapotrzebowanie na tlen,

ChZT – chemiczne zapotrzebowanie na tlen,

$N_{całk}$ – stężenie azotu całkowitego,

P_{og} – stężenie fosforu ogólnego,

s_{zaw} – stężenie zawiesiny ogólnej,

T_{os} – temperatura osadu,

S_{os} – stężenie osadu czynnego,

% rec – stopień recykulacji,

$m_{met,PIX}$ – ilość dawkowanego metanolu, PIX-u,

F/M – obciążenie substratowe obliczane jako $(Q \cdot BZT_5) \cdot (V \cdot S_{os})^{-1}$.

Tabela 2. Zestawienie zmienności ilości (m_0) powstających osadów nadmiernych w okresie $t = 1-7$ doby

Table 2. Summary of variation amount (m_0) of excess sludge generated in the period $t = 1-7$ days

Okres, doby	Zakres, m^3	Średnia, m^3
t = 1	340-2122	1336
t = 2	443-4240	2654
t = 3	728-6331	3980
t = 4	1553-8402	5329
t = 5	2301-10450	6680
t = 6	2869-12511	8020
t = 7	3597-14573	9358

Tabela 3. Zestawienie wartości parametrów dopasowania (MAE, MAPE) uzyskanych modeli matematycznych (SVM, k – NN) do prognozy ilości osadu nadmiernego w okresie $t = 1-7$ dni

Table 3. Summary of fitting parameters (MAE, MAPE) obtained mathematical models (SVM, k – NN) to forecast the amount of excess sludge in period $t = 1-7$ days

Okres, doby	SVM				k-NN			
	Uczenie		test, walidacja		Uczenie		test, walidacja	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
t = 1	160	14,70	204	16,45	206	18,11	215	19,66
t = 2	294	14,11	370	14,73	378	15,60	428	15,80
t = 3	421	13,31	634	14,46	529	14,97	717	15,78
t = 4	531	11,51	680	14,29	658	15,16	758	18,26
t = 5	628	10,59	814	13,43	801	14,38	934	17,69
t = 6	758	10,65	948	12,66	942	13,89	1060	16,02
t = 7	928	11,25	1083	12,44	1096	15,15	1151	16,25

Na podstawie przeprowadzonych analiz (tabela 3, 4) można stwierdzić, że w rozpatrywanych okresach, dla których wykonywane były obliczenia ilości osadów nadmiernych wartości średnich błędów bezwzględnych (MAE) zwiększały się, natomiast wartości błędów względnych nieznacznie zmniejszały się. Z przeprowadzonych obliczeń (tabela 3, 4) wynika, że najlepszymi zdolnościami predykcyjnymi do-

wej ilości osadu charakteryzuje się model opracowany w oparciu o metodę wektorów nośnych, dla którego wartości błędów w obrębie rozpatrywanych metod (SVM, k-NN, BT) w zbiorze testowym są najmniejsze i wynoszą $MAE = 204 \text{ m}^3$ i $MAPE = 16,45\%$.

Tabela 4. Zestawienie wartości parametrów dopasowania (MAE, MAPE) uzyskanego modelu matematycznego (BT) do prognozy ilości osadu nadmiernego w okresie $t = 1-7$ dni

Table 4. Summary of fitting parameters (MAE, MAPE) obtained mathematical model (BT) to forecast the amount of excess sludge in period $t = 1-7$ days

Okres, doby	BT			
	uczenie		test, walidacja	
	MAE	MAPE	MAE	MAPE
t = 1	166	15,57	236	34,86
t = 2	196	8,67	208	8,66
t = 3	407	17,06	409	12,09
t = 4	345	7,61	357	7,41
t = 5	584	10,25	675	11,26
t = 6	1016	11,99	1222	16,13
t = 7	1012	15,05	1284	16,63

Nieznacznie gorsze wyniki (tabela 3) uzyskano modelem wykonanym przy pomocy metody k – najbliższego sąsiada, w którego przypadku średni błąd prognozy wartości m_0 w zbiorze testowym wynosił 215 m^3 , a błąd względny był równy $19,66\%$. Zdecydowanie najgorsze rezultaty obliczeń dobowej ilości osadów nadmiernych otrzymano za pomocą drzew wzmacnianych, dla których wartości MAE i MAPE w zbiorze testowym spośród rozpatrywanych metod (SVM, k-NN, BT) są największe i wynoszą 236 m^3 i $34,86\%$.

W przypadku prognozy ilości osadów generowanych w przeciągu dwóch dni najmniejsze wartości błędów dla zbioru testowego, a co za tym idzie najlepszy model uzyskano metodą drzew wzmacnianych ($MAE = 208 \text{ m}^3$ i $MAPE = 8,66\%$). Natomiast, najgorsze wyniki symulacji otrzymano przy pomocy modelu bazującego na metodzie k – najbliższego sąsiada, dla którego wartości błędów w zbiorze testowym w obrębie analizowanych metod (SVM, k-NN, BT) są największe i wynoszą odpowiednio $MAE = 428 \text{ m}^3$ i $MAPE = 15,8\%$. Podobnie jak w powyższym

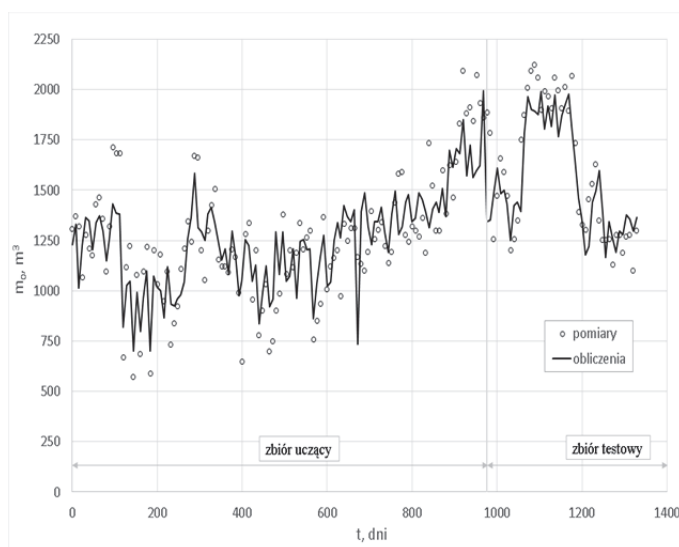
przypadku (2 dobowa prognoza osadu nadmiernego) najdokładniejsze wyniki obliczeń dla 3 dniowego okresu czasu uzyskuje się metodą drzew wzmacnianych ($MAE = 409 \text{ m}^3$ i $MAPE = 12,09\%$) a najgorsze metodą k – najbliższego sąsiada ($MAE = 717 \text{ m}^3$ i $MAPE = 15,78\%$).

Obliczenia przeprowadzone dla 4 i 5-cio dniowej prognozy ilości osadu nadmiernego wykazały, że najdokładniejsze modele zapewniające również najmniejsze błędy predykcji otrzymano metodą drzew wzmacnianych, w przypadku których dla okresu 4-dniowego uzyskano $MAE = 357 \text{ m}^3$ i $MAPE = 7,41\%$ a dla 5 – cio dniowego $MAE = 675 \text{ m}^3$ i $MAPE = 11,26\%$. Znacznie gorsze wyniki obliczeń (tabela 3, 4) 4 i 5-cio dniowej ilości osadów otrzymano metodą wektorów nośnych, ponieważ wartości błędów predykcji analizowanego parametru działania układu technologicznego w zbiorze testowym są równe $MAE = 680 \text{ m}^3$ i $MAPE = 14,29\%$ w przypadku prognozy 4-dniowej, z kolei dla 5-dniowej wynoszą $MAE = 814 \text{ m}^3$ i $MAPE = 13,43\%$.

W przypadku obliczeń ilości osadów generowanych w okresie 6-7 dni najmniejsze wartości błędów predykcji osadu nadmiernego w obrębie rozpatrywanych metod (SVM, RF, BT) uzyskano w oparciu o wektory nośne co potwierdzają wartości MAE i $MAPE$ (tabela 4, tabela 5).

W przypadku modeli otrzymanych metodami k-NN i BT wartości błędów w zbiorach testowych zmieniają się w zakresach $MAE = 1060-1151 \text{ m}^3$ i $MAPE = 16,02-16,25\%$ dla 6-dniowej prognozy. Natomiast, dla 7-dniowej symulacji wielkości błędów wynoszą odpowiednio $MAE = 1222-1284 \text{ m}^3$ i $MAPE = 16,13-16,63$ dla 7 dni. W celu wizualizacji wyników obliczeń na rysunku 1 przedstawiono przykładowo dobowe pomierzone i wyznaczone metodą wektorów nośnych ilości osadów nadmiernych w tygodniowych odstępach czasu.

Mając na uwadze niniejsze wyniki analiz można stwierdzić, że do prognozy od 1 do 7-dniowej ilości osadu nadmiernego można zastosować metody wektorów nośnych, drzew wzmacnianych jak i k – najbliższego sąsiada co potwierdzają otrzymane wartości dopasowania parametrów wyników obliczeń do danych pomiarowych.



Rys. 1. Porównanie wyników obliczeń metodą SVM i pomiarów dobowej ilości osadów nadmiernych

Fig. 1. Comparison of the results of the calculation method of support vectors machine and the measurement of the daily amount of excess sludge

5. Podsumowanie

Obecnie tematem licznych publikacji jest problem optymalizacji działania oczyszczalni ścieków w oparciu o ilość pobieranej energii elektrycznej, stężenie zanieczyszczeń na odpływie, ilość powstających osadów, ilość powietrza dostarczanego do komór nityfikacji itp. W związku z powyższym celowe jest tworzenie modeli matematycznych dających możliwość prognozowania parametrów stanowiących podstawę oceny działania OS. Ze względu na problemy z kalibracją modeli fizycznych jak i długo okresową prognozą ilości i jakości ścieków dopływających do oczyszczalni nie zawsze jest możliwość opracowania modeli hybrydowych do prognozy działania oczyszczalni. W związku z powyższym do modelowania procesów zachodzących na oczyszczalni stosuje się modele czarnej skrzynki.

Przeprowadzone analizy wykazały, że do krótko i długookresowej prognozy ilości powstających osadów nadmiernych można zastosować metodę wektorów nośnych, k – najbliższego sąsiada i drzew wzmacnianych. Na podstawie wykonanych obliczeń stwierdzono, że lepszymi

zdolnościami predykcyjnymi dobowej oraz 6 i 7 dniowej ilości osadów nadmiernych charakteryzuje się model opracowany w oparciu o metodę wektorów nośnych. Z kolei najmniejsze wartości błędów względnych i bezwzględnych prognozy ilości osadów powstających w przeciągu 3-5 dni uzyskano przy zastosowaniu metody drzew wzmacnianych, natomiast w metodzie k – najbliższego sąsiada wielkości tych błędów były największe.

Literatura

- ATV – DVWK – A 131 P (2001). *Wymiarowanie oczyszczalni ścieków z osadem czynnym*. Warszawa: Wydawnictwo Seidel-Przywecki.
- Bień, J., Neczaj, E., Worwąg, M., Grosser, A., Nowak, D., Milczarek, M., Janik, M. (2001). Kierunki zagospodarowania osadów w Polsce po roku 2013. *Inżynieria i Ochrona Środowiska*, 14(4), 375-384.
- Dellana, S., West, D. (2009). Predictive modeling for wastewater applications: Linear and nonlinear approaches. *Environmental Modelling and Software*, 24, 96-106.
- Friedman, J. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, 38(4), 367-378.
- Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Gawdzik, J., Długosz, J., Urbaniak, M. (2015). General Characteristics of The Quantity And Quality of Sewage Sludge From Selected Wastewater Treatment Plants in The Świętokrzyskie Province. *Environment Protection Engineering*, 41(2), 107-117.
- Heidrich, Z., Witkowski, A. (2005). *Urządzenia do oczyszczania ścieków – projektowanie, przykłady obliczeń*. Warszawa: Seidel-Przywecki.
- Henze, M., Gujer, W., Mino, T., Loosdrecht, M. (2000). *Activated Sludge Models*. London: IWA Publishing.
- Kalinowska-Przybyś, E. (2002). Uwaga na stopień recyrkulacji. *Forum Eksplo-ratora*, 3(8), 8-9.
- Kiczko, A., Romanowicz, R., Osuch, M., Karamuz, E. (2013). Maximising the usefulness of flood risk assessment for the River Vistula in Warsaw. *Natural Hazards and Earth System Sciences*, 13, 3443-3455.
- Licznar, P., Szeląg, B., (2014). Analiza zmienności czasowej opadów atmosferycznych w Warszawie. *Ochrona Środowiska*, 3, 23-28.
- Martin, C., Shaw, A.R., Phillips, H.M., Gilley, A. and Ayesa, E. (2010). *Comparison of methods for dealing with uncertainty in wastewater modelling and design*. In: Proc. WWTmod2010. Mont-Sainte-Anne, Québec, Canada, 63-169.

- Piotrowski, A., Napiorkowski, J., Rowiński, P. (2006). Flash-flood forecasting by means of neural networks and nearest neighbour approach – a comparative study. *Nonlinear Processes Geophysics*, 13, 443-448.
- Raha, D. (2007). Exploring Artificial Neural Networks (ANN) Modelling for a Biological Nutrient Removal (BNR) sewage treatment Plant (STP) to Forecast Effluent Suspended Solids. *Indian Chemical Engineering*, 49(3), 205-220.
- Rydziński, R. (2012). *Dobre praktyki związane z gospodarką osadami ściekowymi*. PURE. European Regional Development Fund and European Neighbourhood and Partnership Instrument.
- Serón, N., Puig, S., Meijer, S., Balaguer, M., Colprim, J. (2011). Sludge production based on organic matter and nitrogen removal performances. *Water Practice and Technology*, 6, 2.
- Studzinski, J., Bartkiewicz, L., Stachura, M. (2013). *Development of mathematical models for forecasting hydraulic loads of water and wastewater networks*. EnviroInfo 2013: Environmental Informatics and Renewable Energies Copyright 2013 Shaker Verlag.
- Tao, X., Chengwen, W. (2013). Estimating and modelling the sludge excess discharge in wastewater treatment plants in China. *Environmental Engineering and Management Journal*, 12(7), 1509-1514.
- Vapnik, V. (1998). *Statistical Learning Theory*. New York: John Wiley and Sons.
- Wei, X., Kusiak, A. (2015). Short-term prediction of influent flow in wastewater treatment plant. *Stochastic Environmental Research and Risk Assessment*, 29(1), 241-249.

Application of Selected Nonlinear Methods to Forecast the Amount of Excess Sludge

Abstract

Operation of a sewage treatment plant is a complex task because it requires maintaining the parameters of its activities at the appropriate level in order to achieve the desired effect of reducing pollution and reduce the flow of sediment discharged from the biological reactor. The basis for predicting the amount of excess sludge and operational parameters WWTP can provide physical models describing the biochemical changes occurring in the reactor, in which the input parameters, ie. Indicators of effluent quality and quantity of wastewater are modeled in advance.

However, due to numerous interactions and uncertainty of the data in the physical models and forecast errors parameters of the inlet to the treatment plant Simulation results may be affected by significant errors. Therefore, to

minimize the prediction error parameters of operation of the technological objects deliberate use of a black box model. In these models at the stage of learning is generated model structure underlying the projections analyzed the operating parameters of the plant.

This publication presents the possibility of the use of methods: support vector, k – nearest neighbour and trees reinforced to predict the amount of the resulting excess sludge during wastewater treatment in the WWTP located in Sitkówka – News with a capacity of 72,000 m³/d with a load of 275,000 PE . Due to the fact that did not have the quality parameters of wastewater at the inlet to the activated sludge chambers it was not possible to verify the empirical relationships commonly used in engineering practice to determine the size of the daily flow of excess sludge. Due to the significant differences in the amount of excess sludge generated in the period ($t = 1-7$ days) the simulation of the amount of sludge into the time were performed. To assessment the compatibility of measurement results and simulations quantities of sludge the mean absolute error and relative error of prediction for the considered parameter of technology was used.

The analyzes carried out revealed that the amount of generated excess sludge can be predicted on the basis of parameters describing the quantity and quality of influent waste water (slurry concentration of total nitrogen and total phosphorus, BOD₅) and the operating parameters of the biological reactor (recirculation rate, concentration and temperature of the sludge, the dosed amount of methanol and PIX). On the basis of computations, it can be concluded that the most accurate forecasting results amounts of sediment were obtained by using a reinforced trees

($t = 2$ to 5 days) and Support Vector Machines methods ($t = 1, 6, 7$ days). While the highest values of forecast errors sediments was obtained using a k – nearest neighbor ($t = 2$ to 5 days) and reinforced trees ($t = 1, 6, 7$ days).

Słowa kluczowe:

osady nadmierne, oczyszczanie ścieków, metoda wektorów nośnych, k – najbliższego sąsiada, drzewa wzmacniane

Keywords:

excess sludge, wastewater treatment, support vector machine, k – nearest neighbour, boosted tree