



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

BADANIA SYSTEMOWE

Tom 76

**MODELOWANIE MATEMATYCZNE,
OPTYMALIZACJA I STEROWANIE
PRACĄ PRZEPLYWOWYCH
OCZYSZCZALNI ŚCIEKÓW**

Bartosz Szelaĝ

Warszawa 2019

**MODELOWANIE MATEMATYCZNE, OPTYMALIZACJA I STEROWANIE PRACĄ
PRZEPLYWOWYCH OCZYSZCZALNI ŚCIEKÓW, Bartosz Szelaĝ**



**POLSKA AKADEMIA NAUK
INSTYTUT BADAŃ SYSTEMOWYCH**

**Seria: BADANIA SYSTEMOWE
Tom 76**

**Redaktor naukowy:
Prof. dr hab. inż. Olgierd Hryniewicz**

Warszawa 2019

Rada redakcyjna serii: BADANIA SYSTEMOWE

Prof. Janusz Kacprzyk – przewodniczący

Prof. Olgierd Hryniewicz

Prof. Tadeusz Kaczorek

Prof. Jerzy Klamka

Prof. Józef Korbicz

Prof. Zbigniew Nahorski

Prof. Ngoc Thanh Nguyen

Prof. Marek Niezgódka

Prof. Ewaryst Rafajłowicz

Prof. Leszek Rutkowski

Prof. Roman Słowiński

Prof. Jan Studziński

Prof. Tomasz Szapiro

Bartosz Szelaĝ

**MODELOWANIE MATEMATYCZNE,
OPTYMALIZACJA I STEROWANIE
PRACĄ PRZEPIYWOWYCH
OCZYSZCZALNI ŚCIEKÓW**

Warszawa 2019

Copyright © by Instytut Badań Systemowych PAN
Warszawa 2019

Autor:

Dr inż. Bartosz Szelaĝ
Wydział Inżynierii Środowiska, Geomatyki i Energetyki
Politechnika Świętokrzyska
bszelag@tu.kielce.pl

Recenzenci:

Dr hab. inż. Ewa Burszta-Adamiak, prof. nadzwyczajny
Uniwersytet Przyrodniczy we Wrocławiu

Dr hab. inż. Izabela Rojek, prof. nadzwyczajny
Uniwersytet Kazimierza Wielkiego w Bydgoszczy

Skład: Anna Gostyńska / Aneta M. Pielak

Wydawca:

Instytut Badań Systemowych
Polska Akademia Nauk
Newelska 6, 01-447 Warszawa
www.ibspan.waw.pl

*Publikacja sfinansowana ze środków projektu RID
Ministerstwa Nauki i Szkolnictwa Wyższego nr 025/RID/2018/19*

ISSN 0208-8029

ISBN 978-83-89475-62-6

1. Wprowadzenie

Poprawa jakości wód odbiorników i ograniczenie niekorzystnego wpływu na środowisko stanowi obecnie jeden z najaktualniejszych tematów zarówno w kraju, jak i zagranicą (Thunberg i in. 2009, Minsoo i in. 2016). W celu poprawy jakości wód powierzchniowych wprowadzono w ustawodawstwie unijnym i krajowym szereg wymagań formalno-prawnych, narzucających wielkości dopuszczalnych stężeń wskaźników jakości ścieków oczyszczonych, które nie mogą zostać przekroczone. Potwierdzeniem tego w skali europejskiej jest to dyrektywa unijna 91/271/EWG odnosząca się do kwestii oczyszczania ścieków komunalnych. Natomiast, w kraju jest to Rozporządzenie Ministra Środowiska z dnia 18 listopada 2014 r. w sprawie warunków, jakie należy spełnić przy wprowadzaniu ścieków do wód lub do ziemi, oraz w sprawie substancji szczególnie szkodliwych dla środowiska wodnego. Zgodnie z obowiązującymi aktami prawnymi podstawą projektowania oczyszczalni jest równoważna liczba mieszkańców (RLM), która stanowi podstawę doboru technologii oczyszczania ścieków. Zaprojektowany układ technologiczny ma zapewniać odpowiednie usunięcie związków organicznych (BZT_5 , $ChZT_{(Cn)}$), azotu ogólnego (N_{og}) i fosforu ogólnego (P_{og}) (tab.1.1).

W warunkach rzeczywistych wartości wskaźników jakości przyjęte do obliczeń oczyszczalni ścieków nie przyjmują wartości stałych, ale zmieniają się w szerokim zakresie. Równocześnie, zmienia się ilość dopływających ścieków i temperatura powietrza, co ma istotny wpływ na przebieg procesów oczyszczania przebiegających w reaktorze biologicznym. Podczas intensywnych opadów znacznie zwiększa się ilość dopływających ścieków, co prowadzi zwykle do hydraulicznego przeciążenia całego obiektu i wpływa na skrócenie czasu zatrzymania ścieków w reaktorze biologicznym. Równocześnie, zwiększony dopływ ścieków może prowadzić do przeciążenia hydraulicznego wydzielonych obiektów oczyszczalni, na przykład osadnika wtórnego, co niekorzystanie wpływa na proces klarowania oczyszczonych ścieków. Jednocześnie, z uwagi na rozcieńczenie dopływających ścieków, mogą wystąpić problemy związane z niedostateczną ilością związków biogennych niezbędnych do prawidłowego

przebiegu procesu oczyszczania (usuwania związków organicznych, azotu i fosforu).

Tab. 1.1. Wpływ wielkości oczyszczalni ścieków (RLM) na wymagany stopień redukcji zanieczyszczeń na odpływie z oczyszczalni ścieków; BZT₅ – pięciodniowe biochemiczne zapotrzebowanie na tlen, ChZT_(Cr) – chemiczne zapotrzebowanie na tlen wykonane metodą dwuchromianową, Z_{og} – zawiesiny ogólne, N_{og} – azot ogólny, P_{og} – fosfor ogólny, x^a – wartości wymagane wyłącznie w ściekach wprowadzanych do jezior i ich dopływów oraz bezpośrednio do sztucznych zbiorników wodnych usytuowanych na wodach płynących, x^b – minimalny procent redukcji nie ma zastosowania do ścieków wprowadzanych do jezior i ich dopływów, bezpośrednio do sztucznych zbiorników wodnych usytuowanych na wodach płynących oraz do ziemi (Dz. U. poz. 1800, 2014).

Wskaźniki	RLM				
	<2000	2000÷9999	10000÷14999	15000÷99999	>100000
BZT ₅ , mgO ₂ /l	40	25	25	15	15
BZT ₅ , min %		70-90	70-90	90	90
ChZT _(Cr) , mgO ₂ /l	150	125	125	125	125
ChZT _(Cr) , min %		75	75	75	75
Z _{og} , mg/l	50	35	35	35	35
Z _{og} , min %		90	90	90	90
N _{og} , mg/l	30 ^a	15 ^a	15	15	10
N _{og} , min %		-	70-80 ^b	70-80	70-80
P _{og}	5 ^a	2 ^a	2	2	1
P _{og} , min %		-	70-80 ^b	80	80

Eksploatując oczyszczalnię ścieków należy mieć na uwadze fakt, że w okresie zimowym na skutek obniżenia temperatury powietrza ma miejsce obniżenie kinetyki procesów zachodzących w reaktorach biologicznych, co także może prowadzić do pogorszenia jakości ścieków na odpływie. Należy także liczyć się z faktem, że mimo stabilnych warunków pracy reaktora (tj. brak opadów atmosferycznych, okres lata lub wiosny wpływające pozytywnie na stabilność procesu oczyszczania) mogą wystąpić niekontrolowane zrzuty ładunków zanieczyszczeń do systemu kanalizacyjnego, co może mieć niekorzystny wpływ na działanie obiektów technologicznych oczyszczalni i powodować konieczność zmiany nastaw w regulatorach reaktora. Wymienione czynniki mogą prowadzić do zakłóceń ciągłości przebiegu procesu oczyszczania ścieków. Dlatego też, w celu

uzyskania na odpływie wymaganych wartości wskaźników jakości ścieków, zachowania ciągłości procesów zachodzących w reaktorze biologicznym i utrzymania wysokiej niezawodności działania oczyszczalni, konieczne jest aktywne sterowanie i bieżąca kontrola nastaw reaktora biologicznego. W praktyce jest to zadanie bardzo skomplikowane, wymagające ciągłych zmian wielu nastaw regulatorów w taki sposób, aby zapewnić wymaganą jakość ścieków oczyszczonych i aby sterując jednym procesem, nie zakłócić innych procesów przebiegających w bioreaktorze.

Z uwagi na to, że eksploatacja oczyszczalni ścieków wymaga równoczesnej kontroli wielu procesów, stosuje się szereg urządzeń pomiarowych zamontowanych na jej obiektach technologicznych. Pożądaną sytuacją byłoby prowadzenie pomiarów ilości i jakości ścieków na dopływie, nastaw reaktora oraz jakości ścieków na odpływie z oczyszczalni. Obecnie większość obiektów krajowych posiada monitoring on-line nastaw reaktora biologicznego i jakości ścieków na odpływie. Natomiast, rzadko występują przypadki, gdy prowadzone są ciągłe pomiary jakości ścieków na dopływie do obiektu. Stan ten wynika z faktu, iż zgodnie z obowiązującymi aktami prawnymi nie ma takich wymagań i zachodzi jedynie potrzeba okresowej oceny efektywności działania oczyszczalni ścieków w zależności od wielkości obiektu wyrażonej za pomocą równoważnej liczby mieszkańców – RLM (tab. 1.2).

Tab. 1.2. Wpływ wielkości oczyszczalni ścieków (RLM) na roczną liczbę próbek poboru jakości ścieków (Dz. U. poz. 1800, 2014)

RLM	Liczba próbek	
	Pierwszy rok	Kolejny rok po spełnieniu wymagań
<2000	4	2
2000 ÷ 9999	12	4
10000 ÷ 49999	12	12
>50000	24	24

Standardowo, w przypadku obiektów RLM < 2000 pomiary efektywności pracy oczyszczalni należy kontrolować pobierając 4 próbki w ciągu roku; jeżeli pobrane próbki spełnią wymagane warunki, to w kolejnym roku pobierane mogą być 2 próbki (tab. 1.2). Dla obiektów, gdzie $9999 > \text{RLM} > 2000$, w przypadku uzyskiwania wysokiej efektywności możliwe jest także ograniczenie w kolejnym roku liczby próbek. Z kolei, dla oczyszczalni RLM > 10000 nie jest możliwa redukcja ilości analizowanych próbek na dopływie.

Pomiar jakości ścieków na dopływie nie stanowi obecnie dużego problemu technicznego z uwagi na to, że istnieje szereg analizatorów (np. firmy Endress–Hausser, HahLange itp.) umożliwiających pomiar odpowiednich wskaźników jakości, jednak koszt ich zakupu i eksploatacji jest wysoki. Równocześnie do stosowania ich zniechęca fakt, że urządzenia te należy okresowo kalibrować w celu uzyskania wiarygodnych wyników pomiarów, co generuje dodatkowe koszty i wymaga uciążliwego ciągłego nadzoru.

Należy przy tym zauważyć, że w oczyszczalniach ścieków zarówno o wysokim stopniu zautomatyzowania i rozbudowanym systemie monitoringu, jak i w obiektach, gdzie pomiary wykonywane są głównie na drodze oznaczeń laboratoryjnych, gromadzone są ogromne ilości danych mogących służyć do opracowania modeli procesowych. Fakt ten wzbudza szerokie zainteresowanie, na co wskazuje znaczna liczba publikacji zagranicznych. W publikacjach tych na podstawie zgromadzonych danych pomiarowych opracowuje się modele matematyczne do prognozowania jakości ścieków na odpływie oraz kontroli przebiegu procesów nityfikacji, denityfikacji i defosfatacji. Jednak z punktu widzenia eksploatatora oczyszczalni ścieków interesujące jest nie tylko prognozowanie przebiegu procesu i jego symulacja. Istotne znaczenie ma także sterowanie procesem, to znaczy wyznaczanie takich nastaw bioreaktora, aby zapewnić przebieg procesu warunkujący zadane wartości ścieków na odpływie i zminimalizowanie kosztów eksploatacyjnych obiektu. Takie postawienie zadania cieszy się ostatnio coraz większym zainteresowaniem, co potwierdza zwiększająca się liczba odnośnych prac w kraju i zagranicą, choć publikacji krajowych na ten temat jest zdecydowanie mniej.

Parametry działania reaktora biologicznego można ustalać w taki sposób, że bez względu na ilość i jakość ścieków dopływających oraz na porę roku, na odpływie uzyska się wartości wskaźników jakości ścieków poniżej wartości dopuszczalnych. Jednak problem polega na tym, że realizując takie nastawy zwykle nie uwzględnia się kosztów eksploatacyjnych obiektu, które mogą być wysokie. Wobec powyższego nasuwa się pytanie, czy nie można byłoby nastaw w bioreaktorze skorygować w taki sposób, aby uzyskać zbliżony efekt oczyszczania ścieków, ale niższym kosztem. Aby odpowiedzieć na to pytanie i uniknąć ewentualnych problemów eksploatacyjnych związanych z możliwym zakłóceniem pewnych procesów technologicznych przy sterowaniu innym procesem, jest konieczna optymalizacja pracy oczyszczalni i zdefiniowanie w tym celu odpowiednich kryteriów optymalizacyjnych. Podstawowym kryterium oceny efektywności pracy oczyszczalni jest kryterium ekonomiczne: proces oczyszczania powinien być realizowany w taki sposób, aby uzyskać najniższe koszty eksploatacji przy zapewnieniu wymaganej jakości ścieków oczyszczonych. Kolejne kryteria mogą mieć charakter techniczny i odnosić się do utrzymania na odpowiednim poziomie przebiegu poszczególnych procesów w reaktorze

biologicznym, do minimalizacji zużycia dawkowanych środków chemicznych lub minimalizacji zużycia tlenu na napowietrzanie ścieków. Kryteriów technicznych może być wiele, ponieważ jest to uzależnione od tego, co technolog chce przede wszystkim uzyskać. Kryteriami stanowiącymi kombinacje kryterium ekonomicznego i kryteriów technicznych są kryteria niezawodnościowe. W takim przypadku można skoncentrować się na analizie niezawodności działania obiektu w odniesieniu do jednego procesu najistotniejszego dla technologa lub do kilku procesów równocześnie. W ostatnich latach przedstawiono szereg interesujących prac z zakresu zastosowania modelowania matematycznego w optymalizacji i sterowaniu oczyszczalniami ścieków (np. Mulas i in. 2015, Corominas i in. 2013, Asadi i in. 2016), jednak mimo licznych analiz o charakterze teoretycznym, jedynie w jednym z 40 badanych obiektów wdrożono opracowany model matematyczny (Haimi i in. 2013). W obszarze sterowania do rozwiązania pozostaje cały szereg zagadnień dotychczas nie badanych, które dotyczą m.in. sterowania reaktorem w przypadku, gdy nie są prowadzone ciągle pomiary wskaźników jakości ścieków na dopływie a jedynie pomiary okresowe, co ma miejsce w większości krajowych oczyszczalni ścieków. Mimo potwierzonego wpływu procesu sedymentacji na jakość ścieków na odpływie w dalszym ciągu fakt ten jest uwzględniany w ograniczonym stopniu przy doborze strategii sterowania reaktorem biologicznym. Jednocześnie, mimo wielu analiz dotyczących procesu sedymentacji osadu czynnego, nie opracowano dotychczas uniwersalnej zależności pozwalającej ocenić podatność osadu na puchnięcie, która byłaby pomocna na etapie sterowania reaktorem biologicznym.

O skali poruszanego problemu i zapotrzebowania na odpowiednie algorytmy obliczeniowe może świadczyć fakt, że na terenie Polski wg danych GUS (BDL, 2017) znajduje się 3258 komunalnych oczyszczalni ścieków, z czego 11 obiektów oczyszcza ścieki mechanicznie, 2439 biologicznie i 808 obiektów realizuje podwyższony stopień redukcji zanieczyszczeń. Całkowita liczba oczyszczalni nie odzwierciedla bezpośrednio skali problemu, bowiem trzeba na to popatrzeć szerzej, biorąc pod uwagę ilość gmin w kraju tj. 2478 oraz liczbę miast na prawach powiatu tj. 66. Mianowicie dla „dużych jednostek osadniczych” zaprojektowany układ oczyszczania ścieków musi warunkować usuwanie związków organicznych, w tym azotu i fosforu. W tych przypadkach, ze względu na przepustowość obiektów przekroczenie dopuszczalnych wartości wskaźników prowadziłoby do wprowadzenia do odbiornika ścieków znacznych ilości zanieczyszczeń, które mogłyby prowadzić do naruszenia równowagi ekosystemu i mieć daleko idące następstwa ekologiczne. W związku z tym, na tych obiektach instaluje się systemy do kontroli i sterowania reaktorem, co istotnie wpływa na efektywność ich pracy. Rozwiązanie to nie jest tak powszechne w mniejszych oczyszczalniach ścieków, które są bardziej wrażliwe na zmiany ilości i jakości dopływających ścieków, co ma wpływ na ilość zanieczyszczeń trafiającą do odbiorników. Dlatego też, szczególnie w przypadku mniejszych oczyszczalni zachodzi potrzeba ciągłego

monitoringu i aktywnego sterowania bioreaktorem w celu uzyskania wymaganej jakości ścieków na odpływie. W przypadku dużych obiektów, z uwagi na wielkość zamontowanych energochłonnych urządzeń i ilość dopływających ścieków, bardziej wskazana jest optymalizacja procesu oczyszczania mająca na celu obniżenie kosztów eksploatacyjnych.

W pracy przedstawiono koncepcję systemu do kontroli, sterowania i optymalizacji działania reaktora biologicznego w przepływowej oczyszczalni ścieków. W systemie tym przewidziano zastąpienie pomiarów wskaźników jakości ścieków na dopływie wynikami obliczeń uzyskanymi w oparciu o pomiary natężenia przepływu. Do symulacji wartości wskaźników jakości ścieków zaproponowano zastosowanie metod eksploracji danych (ang. data mining). Po wykonaniu wstępnych obliczeń modelowania przedstawiono koncepcję modeli hybrydowych, w których wartości pomiarów wskaźników jakości ścieków dopływających do oczyszczalni zastąpiono wynikami ich symulacji. Modele te zastosowano do prognozy: sedymentacji osadu czynnego, zmieniających się wartości kluczowych parametrów technologicznych reaktora biologicznego wskaźników jakości ścieków na odpływie z oczyszczalni.

Opracowano także modele matematyczne do analizy niezawodności działania reaktora biologicznego w oparciu o tzw. współczynniki niezawodności w odniesieniu do procesu sedymentacji osadu czynnego i poprawności procesu usuwania związków biogennych.

Ponadto, zaproponowano nowatorski parametr uwzględniający interakcje między zmiennymi procesowymi o charakterze stochastycznym. Zmienne te dotyczą ilości i jakości ścieków na dopływie, warunków atmosferycznych i ich wpływu na podatność osadu czynnego na puchnięcie oraz doboru strategii sterowania reaktorem biologicznym.

Przedstawiono także model probabilistyczny niezawodności działania reaktora biologicznego w odniesieniu do sedymentacji osadu czynnego, w którym uwzględniono występowanie awarii systemu pomiarów wskaźników jakości ścieków na dopływie do oczyszczalni.

Przyjęto następujący układ pracy:

W rozdziale 2 omówiono metody biologicznego oczyszczania ścieków i scharakteryzowano rolę poszczególnych obiektów zlokalizowanych w ciągu technologicznym oczyszczalni. Omówiono w nim także rozwiązania technologiczne eksploatowanych systemów oczyszczania ścieków, przy czym skoncentrowano się na parametrach operacyjnych reaktorów i ich wpływie na efektywność działania obiektu.

W rozdziale 3 omówiono obecny stan wiedzy z zakresu modelowania pracy reaktorów biologicznych, ich sterowania i optymalizacji. W rozdziale opisano

także opracowane dotychczas modele do analizy i oceny niezawodności działania oczyszczalni ścieków.

W rozdziale 4 scharakteryzowano obiekty badań uwzględnione w pracy, to znaczy oczyszczalnie ścieków w Sitkówce–Nowiny i Trepczy (dalej również: oczyszczalnia Sitkówka–Nowiny i oczyszczalnia Trepcza).

W rozdziale 5 przedstawiono stosowane w badanych oczyszczalniach metody pomiarów ilości i jakości ścieków oraz parametrów działania reaktorów biologicznych. Przeanalizowano także wiarygodność i jakość danych pomiarowych pozyskiwanych w badanych oczyszczalniach ścieków. Na podstawie zgromadzonych danych wyznaczano podstawowe miary statystyczne opisujące zmienność ilości i jakości ścieków oraz parametrów działania reaktorów w tych oczyszczalniach.

W rozdziale 6 badano możliwość zastosowania wyznaczanych modeli matematycznych do symulacji reaktora biologicznego oczyszczalni i oceny niezawodności jego funkcjonowania. Metody obliczeniowe zastosowane do wyznaczania modeli porównano pod kątem ich złożoności i jej wpływu na szybkość i dokładność wykonywanych obliczeń.

W rozdziale 7 przedstawiono i oceniono wyniki obliczeń dotyczące ilości i jakości ścieków na dopływie i odpływie z oczyszczalni oraz parametrów technologicznych reaktora biologicznego. W rozdziale zaprezentowano także nowatorskie zastosowanie modelu regresji logistycznej do oceny wpływu wrażliwości poszczególnych zmiennych wejściowych modelu na prognozę dopływu ścieków do oczyszczalni. Podejście to jest innowacyjne z uwagi na to, że taki model nie był dotychczas stosowany w tego typu zadaniach, a jedynie wykorzystywany w typowych zagadnieniach klasyfikacji. Wykonane obliczenia pozwalają na uogólnienie uzyskanych wyników z uwagi na to, że symulacje komputerowe przeprowadzono dla dwóch istotnie różnych obiektów (w oczyszczalniach Sitkówka–Nowiny i w Rzeszowie), a otrzymane wyniki okazały się w obu przypadkach zbliżone.

Na przykładzie danych z zakresu natężenia przepływu ścieków przeanalizowano wpływ długości szeregów czasowych, sposobu gromadzenia danych (ciągły, nieciągły) w zależności od dobranej metody eksploracji danych na zdolności predykcyjne modelu. Wykonane analizy pozwoliły ustalić minimalną długość okresu prowadzenia pomiarów warunkującą otrzymanie zadowalających wyników symulacji w zależności od sposobu zapisywania danych i dobranej metody.

W rozdziale 7 przedstawiono także koncepcję modelowania wskaźników jakości ścieków na podstawie wartości parametrów mierzonych w każdej oczyszczalni, to znaczy natężenia przepływu dopływających ścieków. Podejście to

można potraktować jako innowacyjne z uwagi na to, że analizy wykonywane przez innych autorów wykazywały wprawdzie podobną zależność, jednak nie była ona szczegółowo badana. Należy przy tym podkreślić, że wykonane w pracy badania dotyczyły nie tylko jednego obiektu, ale dwóch różnych oczyszczalni ścieków (Sitkówka–Nowiny i w Rzeszowie), co pozwala na uogólnienie wyników. Fakt, że możliwe jest modelowanie jakości ścieków w oparciu o pomiary przepływu ścieków dopływających wykorzystano przy prognozowaniu działania reaktora biologicznego w odniesieniu do jakości ścieków na odpływie (stężenia $N - NH_4$, P_{og} i N_{og}) i sedymentacji osadu czynnego (objętościowego indeksu osadu). Przedstawione w pracy podejście do prognozy działania bioreaktora stanowi również pewną innowację w porównaniu z pracami innych autorów z uwagi na to, że w wyznaczanych obecnie modelach trudno uzyskiwane wartości pomiarów wskaźników jakości ścieków zostają wyeliminowane i zastąpione wartościami obliczeń. W praktyce pozwala to na symulację działania reaktora biologicznego również w przypadku braku ciągłych pomiarów wskaźników jakości ścieków, co do tej pory w pracach innych autorów było pomijane.

W rozdziale 7 omówiono także porównanie wyników symulacji wyznaczonych za pomocą różnych modeli klasyfikacyjnych, zaczynając od najprostszych modeli w postaci regresji logistycznej, modelu liniowego dyskryminacyjnego (LDA), a kończąc na bardziej skomplikowanych modelach uzyskanych metodami data mining. Otrzymane wyniki analiz wykazały, że możliwe jest modelowanie zjawiska sedymentacji osadu czynnego z zadowalającą dokładnością za pomocą jawnych zależności regresyjnych, co w odniesieniu do prac innych autorów (Han Qiao i in. 2012, 2013, 2016, Boztoprak in. 2016, Bagherii i in. 2016) stanowi znaczne uproszczenie symulacji rozpatrywanego zjawiska.

W rozdziale 8 zaproponowano algorytm obliczeniowy modelu do doboru sterowania oczyszczalnią ścieków, wykorzystując wprowadzone wcześniej współczynniki niezawodności w odniesieniu do sedymentacji osadu i poprawności przebiegu procesu usuwania związków biogenych w przypadku, gdy w ciągach pomiarów dotyczących wskaźników jakości ścieków występują nieciągłości. W rozdziale zaproponowano także nowy parametr uwzględniający interakcje między zmiennymi operacyjnymi reaktora o charakterze losowym (ilość i jakość ścieków oraz warunki atmosferyczne), na podstawie którego można ocenić, czy wystąpi puchnięcie osadu i będzie można ustalić odpowiednią strategię sterowania bioreaktorem. W porównaniu z pracami innych autorów zaproponowane podejście jest nowatorskie, bowiem daje możliwość kontroli i sterowania reaktorem biologicznym w przypadku ograniczonego dostępu do pomiarów jakości ścieków opisujących analizowane zjawisko. Nowy parametr, zdefiniowany dla przypadku oczyszczalni Sitkówka–Nowiny, przetestowano także na przykładzie oczyszczalni

Trepcza, badając jego przydatność do oceny podatności osadu czynnego na puchnięcie.

W rozdziale 8 opracowano także model probabilistyczny do oceny niezawodności działania osadnika wtórnego przy wykorzystaniu generatorów liczb losowych (Monte Carlo), w którym ujęto możliwość wystąpienia awarii. Należy podkreślić, że dotychczas potencjalne awarie nie były ujmowane w modelach probabilistycznych do symulacji pracy oczyszczalni.

W rozdziale 9 zaproponowano rozwiązanie kilku zadań optymalizacji pracy reaktora biologicznego na przykładzie oczyszczalni w Sitkówce–Nowiny. W tym przypadku dysponowano kompletem danych dotyczących ilości dopływających ścieków, ich jakości na dopływie i odpływie z oczyszczalni oraz parametrów operacyjnych bioreaktora.

6. Aparat matematyczny zastosowany do obliczeń

W praktyce inżynierskiej obecnie powszechnie stosuje się modelowanie matematyczne. Szczególnie w zadaniach inżynierii środowiska modele znajdują szerokie zastosowanie, symulacje procesów środowiskowych przy użyciu modeli dają możliwość poprawy efektywności działania i bezpieczeństwa obiektów infrastruktury podziemnej i inżynierii rzecznej, redukcji kosztów eksploatacyjnych obiektów, redukcji zanieczyszczeń wprowadzanych do atmosfery i do odbiorników itp. Powszechnie w zadaniach modelowania stosuje się metody statystyczne, za pomocą których wyznacza się modele: parametryczne tj. szarej skrzynki (regresja wieloraka, modele autoregresyjne, logika rozmyta itp.) i czarnej skrzynki (sieci neuronowe, wektory nośne itp.).

Oddzielną grupę metod statystycznych stanowią metody służące do grupowania (clusteringu), tj. wydzielenia w danych pomiarowych oddzielnych klas (np. wyznaczenie skupisk roślin lub zwierząt, składu wód podziemnych, obszarów zagrożonych występowaniem zjawisk w rodzaju huraganów, opadów nawalnych, osuwisk itp.). W metodach tych zastosowanie znajdują algorytmy z tzw. uczeniem nadzorowanym (gdy zgodnie z wiedzą ekspercką dokonuje się wstępnie podziału danych na klasy) i nienadzorowanym (wyznaczona liczba klas i przynależność do nich jest wynikiem obliczeń). Z dokonanego przeglądu literatury (Sanchez i in. 1997, Garcia i Gonzáles 2004, Rustum i in. 2007) wynika, że w zadaniach clusteringu są wykorzystywane zwykle metody drzew klasyfikacyjnych, drzew wzmacnianych, lasów losowych, k-średnich, metoda k-najbliższych sąsiadów, hierarchiczna analiza skupień, metoda logitowa, dyskryminacyjna oraz samoorganizujące się sieci neuronowe.

Na podstawie danych literaturowych (Belanche i in. 2000, Ribeiro i in. 2013, Han i in. 2014, Boztopak i in. 2015, Szelaąg i Gawdzik 2017a,b) można stwierdzić, że w większości przypadków opracowywane modele statystyczne charakteryzują się zadowalającymi zdolnościami predykcyjnymi. Jednocześnie, metody implementowane początkowo wyłącznie w zadaniach regresji znajdują z powodzeniem zastosowanie w zagadnieniach klasyfikacyjnych i odwrotnie. W obu rodzajach problemów obliczeniowych bardzo trudny jest dobór metody, której użycie warunkuje najlepsze dopasowanie wyników obliczeń do danych

pomiarowych. Często, stosując różne metody do modelowania określonego procesu, najlepsze dopasowanie wyników obliczeń do pomiarów otrzymuje się metodą nie najbardziej złożoną, a prostszą. Poza doбором metody obliczeniowej istotny wpływ na otrzymywane wyniki mają również inne czynniki, jak liczba użytych danych pomiarowych oraz ich zmienność (Caruana i in. 2008, Nitze i in. 2012, Kuswanto i in. 2017).

Omawiając problem związany z doбором odpowiedniej metody do prognozy badanego procesu należy zwrócić uwagę na wzajemną zależność złożoności stosowanych metod i wielkości zbioru danych pomiarowych, którymi się dysponuje. W przypadku modeli o złożonej strukturze, przy ograniczonych zasobach danych, mogą pojawić się problemy z estymacją parametrów modelu i ich wiarygodnością. Ponadto, może się zdarzyć sytuacja, że model matematyczny, dla którego uzyskano zadowalające dopasowanie wyników symulacji do pomiarów oceniane na podstawie różnych wskaźników statystycznych, nie będzie poprawnie odwzorowywać prognozowanego procesu. Ten aspekt modelowania był wielokrotnie pomijany w większości prac dotyczących wyznaczania modeli czarnej skrzynki.

Mając na względzie powyższe uwagi, powinno się na etapie symulacji rozpatrywać i porównywać wiele metod w celu uzyskania modelu o najlepszych zdolnościach predykcyjnych.

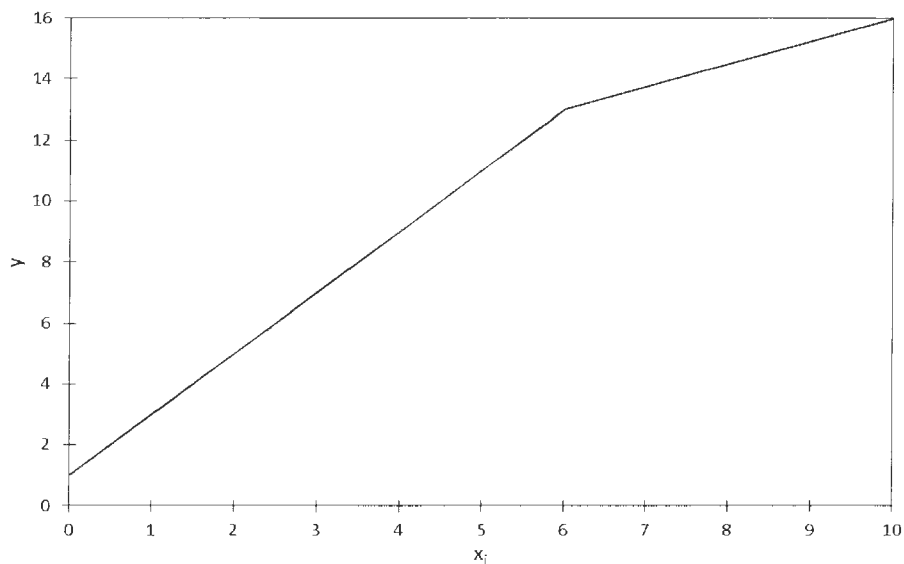
6.1. Wieloraka regresja liniowa

Jedną z najprostszych i równocześnie często stosowanych metod budowy modeli liniowych jest metoda regresji wielorakiej (ang. Multivariate Linear Regression – MLR). W metodzie tej prognozowana zależność ma postać (Box i Jenkins 1983):

$$y = \sum_{j=1}^M a_j \cdot x_j \quad (6.1)$$

gdzie: a_j – parametry modelu wyznaczone metodą najmniejszych kwadratów, x_j – zmienne wejściowe (predyktory), y – wyjście modelu, j – liczba zmiennych wejściowych $j = 1, 2, 3, \dots, M$.

W metodzie przyjmuje się założenia, że sygnały wejściowe modelu są zmiennymi losowymi o rozkładzie normalnym i że są one niezależne, co ocenia się na podstawie współczynnika korelacji liniowej między nimi. W celu doboru niezależnych zmiennych wejściowych i jednocześnie ograniczenia złożoności modelu, na etapie tworzenia modelu stosuje się specjalną metodę doboru zmiennych – regresję krokową z krokiem wprzód i wstecz.



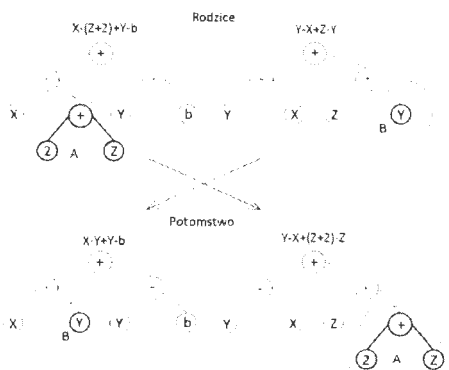
Rys. 6.1. Przykładowy przebieg zmienności funkcji $y = f(x_j)$.

Założenia metody ograniczają jednak jej zastosowanie, ponieważ na etapie doboru zmiennych eliminuje się zmienne, dla których zależność $y = f(x_j)$ jest nieliniowa, co jest częstym przypadkiem przy modelowaniu rzeczywistych procesów (rys. 6.1).

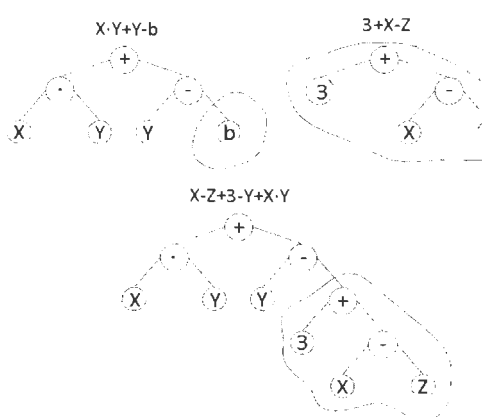
6.2. Programowanie genetyczne

Mając na uwadze ograniczenia metody MLR, opracowano szereg algorytmów obliczeniowych poprawiających jej zdolności predykcyjne. Jednym z nich, który został ujęty w pracy, jest algorytm ewolucyjny nazywany programowaniem genetycznym (GP – ang. Genetic Programming), będący rozwinięciem algorytmu genetycznego (Koza 1992). Przy pomocy opracowanego algorytmu tworzy się tzw. drzewa parsyjne a sam algorytm polega na procesie uczenia populacji składającej się z tzw. osobników. Osobnik w GP reprezentuje drzewo o węzłach losowo wybranych z dwóch zbiorów: zbioru liści będącego zbiorem argumentów (T) i zbioru funkcji (F). Zbiór T stanowią zmienne niezależne opisujące badany proces (wysokość opadu, temperatura, wilgotność powietrza itp.) i stałe, natomiast zbiór F zawiera podstawowe operatory matematyczne (+, -, /, sinus, cosinus, eksponent, funkcje warunkowe itp.), jakie mogą być wykorzystane na etapie tworzenia modelu. Proces ewolucji rozpoczyna się losowym wyborem n osobników populacji i następnie powtarzane są N -krotnie operacje: reprodukcji, genetyczna, oceny i sukcesji, dopóki nie zostanie spełnione kryterium zatrzymania algorytmu. Operator reprodukcji polega na wyborze z aktualnej populacji osobników

rodzicielskich, przy użyciu których generowane są osobniki potomne. Osobniki rodzicielskie poddane są operacji genetycznej mającej na celu wymieszanie informacji w nich zawartych poprzez krzyżowanie (rys. 6.2) i mutację (rys. 6.3).



Rys. 6.2. Schemat działania operatorów genetycznych: krzyżowanie.



Rys. 6.3. Schemat działania operatorów genetycznych: mutacja.

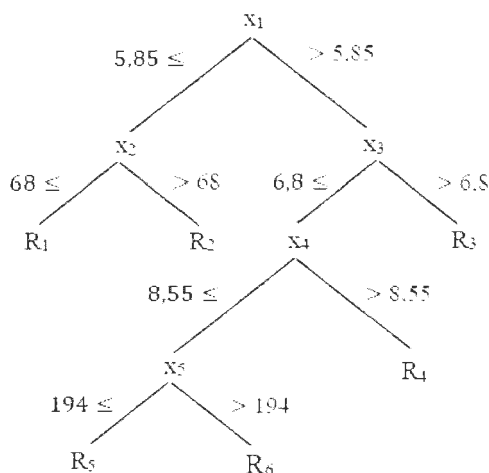
Tak uzyskane osobniki potomne poddawane są ocenie, tj. określa się wartości optymalizowanego kryterium jakości, stanowiącego funkcję dopasowania. Mutacja punktowa zmienia funkcję w wybranej części drzewa, tworząc tzw. poddrzewa. Krzyżowanie tworzy nowe drzewa poprzez zamianę losowo wybranych poddrzew. Zastosowane podejście, w odróżnieniu od regresji wielorakiej, pozwala na uwzględnienie w tworzonym modelu zależności o charakterze nieliniowym.

Modele opracowane za pomocą metody GP charakteryzują się lepszymi zdolnościami predykcyjnymi, niż modele regresji wielorakiej, co wynika z założeń teoretycznych przyjętych na etapie budowy modelu (Koza 1992). Dlatego też, metoda ta znajduje szerokie zastosowanie przy symulacji złożonych procesów w hydrologii (Fallah-Mehdipour i in. 2013), ekologii (Bozorg-Haddad i in. 2017), oczyszczaniu ścieków (Hong i Bhamidimarri 2003), budownictwie (Razenia i in. 2007), inżynierii chemicznej (Cheema i in. 2002), ekonomii (Salpasaranis i Stylianakis 2017), geotechnice (Shanin 2015), medycynie (Bannister i in. 2014) itp. Mimo wielu zalet przedstawiona metoda ma również wady: po pierwsze, brak jest analitycznej zależności opisującej globalne minimum funkcji; po drugie, czas tworzenia modelu może być długi i dochodzić do kilkunastu godzin, co w wielu przypadkach uniemożliwia praktyczne zastosowanie metody. Z uwagi na specyficzny przebieg tworzenia modelu w metodzie programowania genetycznego, w przypadku znacznej liczby zmiennych niezależnych (taki przykład stanowią oczyszczalnie ścieków), na etapie generowania kolejnych osobników (rys. 6.2,

rys. 6.3) mogą być wybierane takie zmienne, których wpływ nie jest decydujący na przebieg badanego procesu, co spowoduje, że uzyskany ostatecznie model może nie być dokładny.

6.3 Regresyjne drzewa decyzyjne

Kolejną badaną w pracy metodą modelowania jest metoda regresyjnych drzew decyzyjnych (rys. 6.4). (ang. classification and regression tree) W metodzie dokonuje się w sposób rekurencyjny podziału przestrzeni zmiennych wejściowych modelu (zmiennych niezależnych) X_j , tzn. w każdym kroku obliczeniowym wydzielony fragment przestrzeni X_j jest dzielony na dwie lub więcej części przy użyciu jednej zmiennej (x_j), co prowadzi do tworzenia decyzyjnego drzewa regresyjnego (Łapczyński 2010).



Rys. 6.4. Przykładowa struktura decyzyjnego drzewa regresyjnego.

Zmienne w strukturze drzewa regresyjnego, a więc również miejsca podziału przestrzeni, określane są w taki sposób, aby wyznaczone części przestrzeni były możliwie jednorodnie ze względu na zmienną zależną (y). Uzyskany model opisuje się równaniem:

$$y = \sum_{p=1}^P a_p \cdot I(\mathbf{x}_j \in R_p) = \sum_{p=1}^P a_p \cdot I(p) \quad (6.2)$$

obrazującym wielowymiarową przestrzeń X_j w której znajdują się poszczególne zmienne ($x_1, x_2, x_3, \dots, x_n$), i gdzie $I(q)$ – funkcja wskaźnikowa opisana zależnością:

$$I(q) = \begin{cases} 1 & \text{gdy } \mathbf{x}_j \in R_p \\ 0 & \text{gdy } \mathbf{x}_j \notin R_p \end{cases} \quad (6.3)$$

oraz α_p – parametry modelu wyznaczone za pomocą formuły:

$$\alpha_p = \frac{1}{N^{(p)}} \cdot \sum_{t=1}^{N^{(p)}} y_i \quad (6.4)$$

W modelach regresyjnych heterogeniczność podzbiorów uzyskanych na skutek rekurencyjnych podziałów mierzona jest za pomocą wariancji:

$$D^2(S_p) = \frac{1}{N^{(p)}} \cdot \sum_{t=1}^{N^{(p)}} (y_i - \bar{y}_k)^2 \quad (6.5)$$

Zaletą decyzyjnych drzew regresyjnych jest prosta interpretacja uzyskanych wyników symulacji, dzięki czemu można stosunkowo szybko ocenić wpływ poszczególnych zmiennych x_j na badane zjawisko. Zaproponowana w metodzie regresyjnych drzew decyzyjnych koncepcja grupowania danych daje możliwość symulacji nieliniowych procesów. Mimo wymienionych zalet metoda drzew regresyjnych ma również wady: wyniki obliczeń mają charakter skokowy, ponieważ niewielka zmiana wartości jednej zmiennej wejściowej może prowadzić do dużego skoku wartości wyjścia modelu, co w rezultacie osłabia jego zdolności predykcyjne; możliwe gwałtowne zmiany wartości zmiennej wyjściowej często powodują niestabilność metody.

6.4. Metoda lasów losowych i drzew wzmocnianych

Mając na uwadze zalety metody regresyjnych drzew decyzyjnych i jednocześnie potrzebę poprawy jej zdolności predykcyjnych, dokonano modyfikacji jej algorytmu obliczeniowego. W rezultacie opracowano metodę drzew wzmocnianych (ang. Boosting Tree, BT) i metodę lasów losowych (ang. Random Tree, RT). W metodzie lasów losowych zastąpiono pojedyncze drzewo regresyjne lasem drzew. Wybór zmiennych niezależnych przy tworzeniu drzew regresyjnych lasu losowego odbywa się w taki sposób, że na każdym etapie podziału przestrzeni \mathbf{X}_j losowanych jest S zmiennych, dopiero spośród których wybierana jest zmienna warunkująca uzyskanie np. jednorodnych danych w wydzielonym segmencie wielowymiarowej przestrzeni.

Zaletami metody RF jest jej odporność na nieprawidłowości występujące w danych pomiarowych (wartości bardzo odstające od średniej, braki w ciągach danych, częściowo skorelowane zmienne wejściowe), odporność na przeuczenie, poprawa zdolności predykcyjnych w porównaniu z metodą regresyjnego drzewa decyzyjnego, możliwość bardziej wiarygodnej oceny wpływu poszczególnych zmiennych (x_j) na wynik symulacji w oparciu o ranking predyktorów, w którym

dla każdej zmiennej określana jest jej tzw. wartość ważności (IMP) (ang. IMPortance).

Metoda lasów losowych może się nie sprawdzić w przypadku dużej liczby zmiennych wejściowych (objasniających), z których tylko część ma decydujący wpływ na modelowany proces. Sytuacja ta może wystąpić wówczas, gdy podczas losowania podzbioru zmiennych do tworzenia drzewa regresyjnego w lesie losowym nie zostaną wybrane zmienne szczególnie ważne, co zdarza się na przykład w przypadku modelowania procesów oczyszczania ścieków, w których występuje wiele zmiennych ważnych, ale jednak o istotnie zróżnicowanym wpływie na przebieg badanego zjawiska. Natomiast, metoda lasów losowych bardzo dobrze nadaje się do rozwiązywania problemów klasyfikacyjnych, gdy liczba zmiennych wejściowych jest duża i dochodzi nawet do kilku tysięcy. Dlatego też, metoda RF na szeroką skalę jest stosowana w badaniach genetycznych (szczególnie do analizy matrycy DNA), przewidywania aktywności biologicznej cząstek (QASR), w medycynie (Ogutu i in. 2011), do analiz przestrzennych w geodezji i kartografii (Nitze i in. 2012, Noi i Kappas 2017), do identyfikacji gatunków roślin i zwierząt oraz analizy jakości wody (Fernandes i in. 2008, Kocev i in. 2009). Metoda lasów losowych znajduje także zastosowanie w rozwiązywaniu problemów przestrzennych dotyczących modelowania zmienności jakości powietrza (Siwek i Osowski 2016), gleby (Nawar i Mouazen 2017), przepływów i jakości wód w zlewniach (Yang i in. 2017). Metoda RF znajduje coraz częściej zastosowanie także przy modelowaniu procesów oczyszczania ścieków z wykorzystaniem technologii sorpcji, utlenienia i reakcji Fentona (Dąbek i in. 2017). Wyniki wykonanych obliczeń w takich przypadkach wskazują, że modele uzyskane za pomocą lasów losowych charakteryzują się często lepszymi zdolnościami predykcyjnymi, niż w przypadku użycia metod, w których struktura modelu jest znacznie bardziej złożona. Jednak, mimo tych zalet metoda RF do tej pory była bardzo rzadko stosowana do symulacji procesów zachodzących w komunalnych oczyszczalniach ścieków (Kusiak i in. 2013), a można sądzić, że jest ona odpowiednią alternatywą dla metod obecnie powszechnie stosowanych.

Modyfikację metody drzew regresyjnych w metodzie lasów losowych stanowi metoda drzew wzmacnianych, w której zaimplementowano koncepcję stochastycznego gradientowego wzmacniania tworzonych drzew dla poprawy zdolności predykcyjnych modelu. W algorytmie wzmacniania gradientowego kolejne drzewa tworzone są na podstawie losowej próby z całego zbioru danych. To rozwiązanie ma na celu eliminację przeuczenia modelu (każde kolejne drzewo regresyjne w strukturze modelu jest budowane w oparciu o inne zbiory danych) i pozwala na uzyskanie modeli mających właściwości generalizacji, co poprawia ich zdolności predykcyjne. Zaletą metody BT, podobnie jak metody lasów losowych, jest możliwość oceny wpływu poszczególnych predyktorów na wynik modelowania na podstawie wyznaczonych wartości ich współczynników ważności

(IMP). Z uwagi na to, że metoda drzew wzmocnianych, podobnie jak lasów losowych, została pierwotnie opracowana dla zadań klasyfikacji danych, więc jej zakres zastosowań jest obecnie podobny, jak w przypadku metody RF (Kusiak i in. 2013, Asadi i in. 2016, Lee i in. 2017, Dąbek i in. 2017).

6.5. Metoda k-najbliższych sąsiadów

Jedną z najprostszych, najstarszych i w dalszym ciągu stosowanych metod eksploracji danych (ang. data mining) jest metoda k-najbliższych sąsiadów (k-NN) (ang. k-Nearest Neighbours). Należy ona do tzw. metod pamięciowych (ang. memory-based), w których uczenie polega na zapamiętywaniu danych treningowych i wyszukiwaniu podobnych obiektów. Podstawą metody k-NN jest założenie, że obiekty podobne, opisane za pomocą cech $[x_1, x_2, x_3, \dots, x_j]$, są przydzielone do tej samej klasy. W metodzie k-NN zmienną wyjściową (objaśnianą) modelu wyznacza się z zależności:

$$y = \frac{1}{K} \cdot \sum_{i=1}^T y_i \cdot J(x_i, x_1) \quad (6.6)$$

gdzie x_1 , jest jednym z K najbliższych sąsiadów x_j , gdy odległość $d(x_i, x_1)$ należy do najmniejszych odległości między obserwacjami ze zbioru $Z_N = \{(x_1, y_1), \dots, (x_n, y_n)\} \in R^{j+1}$, przy czym: $x_i = (x_{1,i}, \dots, x_{j,i})$ stanowi i-ty wektor zmiennych objaśniających o liczbie współrzędnych j, y_i , jest i-tą wartością zmiennej objaśnianej, T – ilość obserwacji, J(x_i, x_1) J – funkcja postaci:

$$J(x_i, x_j) = \begin{cases} 1, & \text{gdy } x_i \text{ jest jednym z } K \text{ najbliższych sąsiadów } x_j \\ 0, & \text{w przeciwnym przypadku} \end{cases} \quad (6.7)$$

Metoda k-NN, podobnie jak RF, BT, znalazła zastosowanie przede wszystkim przy klasyfikacji danych, jednak z czasem zaczęto ją stosować także do symulacji zagadnień regresyjnych. Obliczenia prowadzone za pomocą metody k-NN wykazały (Sarkar i Leong 2000, Hoef i in. 2013), że uzyskiwane wyniki obliczeń (szczególnie w analizie obrazów) charakteryzują się lepszym dopasowaniem do pomiarów, niż otrzymywane za pomocą bardziej złożonych metod. Regresyjne modele statystyczne wyznaczone metodą k-NN dla ograniczonej liczby zmiennych objaśniających często nie są gorsze, niż modele wyznaczone metodami sieci neuronowych lub wektorów nośnych. Znajduje to potwierdzenie na przykład w badaniach wykonanych dla oczyszczalni ścieków, gdy na podstawie pomiarów ilości i jakości ścieków na dopływie i nastaw reaktora opracowano modele do prognozy jakości ścieków na odpływie z oczyszczalni (Lee i Scholz 2006, Asadi i in. 2016). Gorsze rezultaty symulacji niż Asadi i in. (2016) otrzymali Kusiak

i Wei (2013), którzy zastosowali metody k-najbliższych sąsiadów i lasów losowych do prognozy ilości biogazu, uzyskując w wynikach symulacji skokowe przebiegi sygnału wyjściowego. Mogło to wynikać, z niewystarczającej dla tych metod zmienności wartości zmiennych wejściowych, co skutkowało niepoprawną strukturą wyznaczanych modeli. Oznacza to, że w metodach drzew regresyjnych zmienność danych stanowi istotny czynnik determinujący zdolności predykcyjne modeli i zmienność ta powinna być uwzględniana zarówno przy wybotrze metody modelownia, jak i później w procesie tworzeniu modelu.

Zaletą metody k-NN jest jej bardzo prosta implementacja oraz brak procesu uczenia, w którym dokonuje się estymacji parametrów modelu. Do wad metody należy zaliczyć długi czas przeszukiwania dużych zbiorów danych w celu wyznaczenia k-najbliższych sąsiadów, potrzebę zapamiętywania wszystkich danych, co wymaga znacznej pamięci komputera, brak możliwości zastosowania metody do sterowania oczyszczalnią w trybie on-line ze względu na ograniczenia objętościowe i czasowe.

6.6. Sztuczne sieci neuronowe

Bardziej złożoną metodą, niż opisane wcześniej metody k-NN, GP, RF, BT, jest metoda sztucznych sieci neuronowych (ANN), wykorzystywana do modelowania zjawisk, w których zależność między zmiennymi wejściowymi a zmienną wyjściową jest nieliniowa. Opracowano szereg modyfikacji tej metody (Rustum i in. 2007, Chen i in. 2010, Qiao i Yang 2011, Han i Qiao 2013, Hongbin i in. 2014, Heddam 2016, Mirbagherii 2015) jednak jednym z najczęściej stosowanych wariantów ANN jest perceptron wielowarstwowy (MLP) (Côté 1995, Hong i Bhamidimarri 2003, Güçlü i Dursun 2010, Jami i in. 2011, Bartkiewicz i in. 2016).

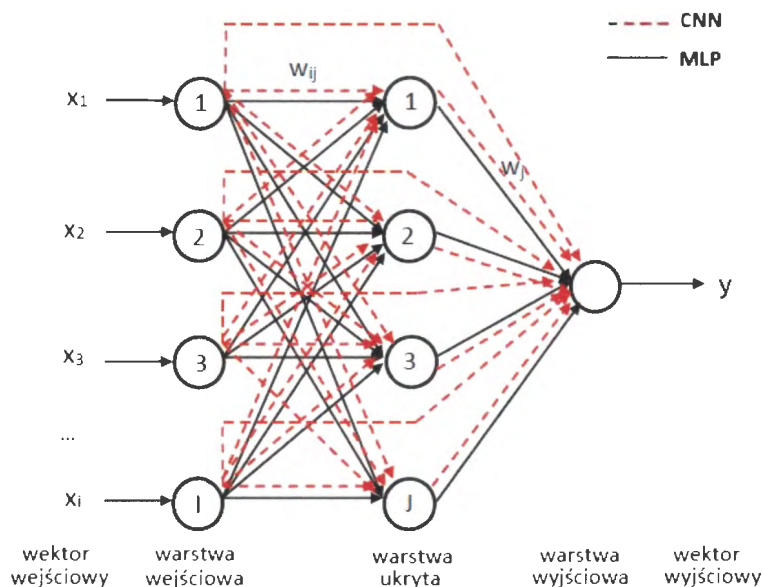
W modelu sieci sygnały wejściowe (x_i) dochodzące do warstwy wejściowej przemnażane są przez wartości wag (w_{ij}). Uzyskane sumy poddawane są transformacji za pomocą liniowej lub nieliniowej funkcji aktywacji f (funkcja eksponentialna, tangens hiperboliczny, sinus, funkcja logistyczna) i następnie przekazywane do neuronów (lub neuronu) warstwy wyjściowej. Estymacja wartości wag (w_{ij}) w modelu ANN odbywa się na etapie uczenia za pomocą odpowiednich algorytmów numerycznych (Rutkowski 2003), w celu wyznaczenia minimum funkcji o postaci:

$$E = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^n (\widehat{y}_n - y_n)^2 \quad (6.8)$$

Wartości wyjść (y) oblicza się z następującego wzoru:

$$y = \sum_{j=1}^J w_j f(\sum_{i=1}^I w_{ij} \cdot x_i + b_j) \quad (6.9)$$

gdzie: I – liczba wejść modelu, J – liczba neuronów w warstwie ukrytej, w_{ij} – wartości wag między wejściami a neuronami warstwy ukrytej, b_j – obciążenia neuronów warstwy ukrytej, w_j – wartości wag między neuronami warstwy ukrytej a neuronem warstwy wyjściowej, f – funkcja aktywacji.



Rys. 6.5. Porównanie struktury sieci neuronowej MLP i CNN.

Na etapie tworzenia modelu MLP kluczowe znaczenie odgrywa określenie liczby neuronów w warstwie ukrytej. Zgodnie z ogólnymi zaleceniami liczba neuronów (J) w warstwie ukrytej powinna być nie mniejsza, niż liczba zmiennych objaśniających (j), ale nie większa, niż $2j + 1$ (Hecht–Nielsen 1987). Aby uniknąć przeuczenia modelu, Rogers i Dowla (1994) sugerują, aby wartość J nie była mniejsza, niż $T/(j + 1)$ (gdzie T – liczba danych w zbiorze uczącym). Liczbę neuronów w warstwie ukrytej można określać także metodą prób i błędów minimalizując błąd prognozy, ale nie dopuszczając do przeuczenia modelu (gdy ze wzrostem J błąd prognozy zwiększa się, to następuje spadek zdolności uogólniających modelu). W przypadku, gdy wyniki symulacji uzyskane za pomocą modelu MLP nie spełniają wymagań, tj. błędy predykcji przekraczają dopuszczalne wartości błędów przyjęte przez użytkownika, to można przeprowadzić optymalizację wyznaczonych wag za pomocą bardziej złożonych metod numerycznych (algorytm genetyczny, mrówkowy, roju pszczół itp.) lub dokonać modyfikacji struktury sieci neuronowej (rys. 6.5). Jednym z możliwych rozwiązań jest wprowadzenie dodatkowych połączeń między neuronami warstwy wejściowej a neuronem (lub neuronami) warstwy wyjściowej i neuronami warstwy ukrytej (tzw. kaskadowa

sieć neuronowa – CNN). W podejściu tym tworzy się zwykle modele sieci (tzw. deep learning) z większą liczbą warstw ukrytych, w których wartość zmiennej zależnej opisuje się wzorem postaci:

$$y = \sum_{j=1}^J w_{j1} f(\sum_{i=1}^I w_{ij} \cdot x_i + b_j) + (\sum_{i=1}^I w_{i1} \cdot x_i + b_1) \quad (6.10)$$

Przeprowadzone badania wykazały, że dodatkowe połączenia prowadzą do przyspieszenia procesu uczenia i poprawy zdolności predykcyjnych modelu CNN w porównaniu z modelem MLP (Setti i Rao 2014, Capizzi i in. 2015).

Metoda MLP znalazła szerokie zastosowanie w zadaniach symulacji biologicznych oczyszczalni ścieków komunalnych (modelowanie ilości i jakości ścieków na dopływie i odpływie z oczyszczalni oraz procesów zachodzących w bioreaktorze, na przykład sedimentacja osadu czynnego) i oczyszczalni ścieków przemysłowych m.in. pracujących w oparciu o metody sorpcji i utleniania (Tomczak i Kamiński 2012, Dąbek i in. 2017). Z kolei, mimo swoich zalet metoda CNN dotychczas stosowana była wyłącznie do prognozy jakości ścieków (BZT₅, ChZT) na odpływie z reaktora biologicznego, do którego trafiały ścieki przemysłowe (Cinar i in. 2006). Metoda ta znajduje szerokie zastosowanie przy analizie zdjęć i obrazów (Shoumy i in. 2015, Zhao i in. 2017).

Podobnie jak w większości omówionych wyżej metod, istotny wpływ na zdolności predykcyjne modelu MLP ma wielkość zbioru danych. W przypadku ograniczonej liczby danych (dane uczące i testowe), z uwagi na złożoną strukturę modelu, mogą wystąpić problemy z wiarygodnością estymowanych parametrów, prowadząc w rezultacie do wyników gorszych, niż na przykład uzyskanych metodą GP (Hong i Bhamidimarri 2003). Metoda sztucznych sieci neuronowych zawdzięcza swą uniwersalność i na ogół zadowalające zdolności predykcyjne dzięki m.in. odporności na uszkodzenia i braki w ciągach danych pomiarowych. Jednocześnie, metoda ANN ma wady, do których należy m.in. zbieżność stosowanych algorytmów optymalizacji do minimów lokalnych oraz długi czas uczenia.

Poza sieciami neuronowymi typu CNN i MLP opracowano także szereg innych wariantów sieci, m.in. wykorzystano wnioskowanie rozmyte przy estymacji wag (sieci ANFIS), co poprawiło dokładność prognozowania w porównaniu z modelem MLP (Pai i in. 2011). Poprawa ta jest dostrzegalna nawet w przypadku gorszej jakości danych wejściowych do obliczeń. Kolejną modyfikacją są sieci neuronowe ze sprzężeniami zwrotnymi (RNN), co również poprawiło dokładność obliczeń w porównaniu z sieciami MLP (Thiery i in. 2005). Także zwiększenie liczby parametrów sieci poprawia na ogół dokładność modeli, chociaż wówczas zachodzi potrzeba posiadania odpowiednio wielkich zbiorów danych (uczących, testowych i walidacyjnych). W przypadku ograniczonej liczby danych rezultaty

obliczeń uzyskiwane zmodyfikowanymi sieciami ANFIS i RNN nie są lepsze, niż otrzymane sieciami o znacznie mniej złożonej strukturze modelu.

6.7. Metoda wektorów nośnych

Kolejną modyfikacją sieci neuronowej typu perceptron wielowarstwowy jest metoda wektorów nośnych (SVM) (ang. Support Vector Machines). Metoda ta pierwotnie została opracowana do rozwiązywania zagadnień klasyfikacyjnych przez Vapnika (1997). Po czym, ponieważ okazała się szybka i efektywna, została dostosowana do zagadnień regresyjnych. Z uwagi na to, że w problemach regresyjnych zależność między zmienną objaśnianą y a zmiennymi objaśniającymi x_i ma często charakter nieliniowy, w metodzie SVM stosuje się transformację nieliniową $\varphi: X^L \rightarrow Z$ obserwacji w zbiorze uczącym do przestrzeni o dużo większym wymiarze. W metodzie SVM stosowana jest funkcja błędu z progiem niezności ε zdefiniowana następująco:

$$L^\varepsilon(y, y(x)) = \begin{cases} 0 & \text{gd}y |y - f(x)| \leq \varepsilon \\ |y - y(x)| - \varepsilon & \text{gd}y |y - f(x)| > \varepsilon \end{cases} \quad (6.11)$$

gdzie: ε – założona dokładność modelu, y – wartość zadana, x – wektor wejść modelu, $y(x)$ – wartość sygnału wyjściowego wyrażona zależnością:

$$y(x) = \sum_{j=1}^K w_j \cdot \varphi_j(x) + b = w^T \cdot \varphi(x) + b \quad (6.12)$$

gdzie: $w = [w_1, \dots, w_k]^T$ – wektor transponowany wag, $\varphi(x) = [\varphi_1(x), \dots, \varphi_k(x)]$ – wektor funkcji bazowych (jądrowych) postaci $K(x, x_i)$.

W metodzie wektorów nośnych uczenie sieci polega na minimalizacji funkcji celu postaci:

$$R = \frac{1}{r} \cdot \sum_{i=1}^R L_s(y_i, y(x_i)) \quad (6.13)$$

w której p – liczba par uczących (x_i, y_i) .

Zakładając, że nie jest możliwa liniowa transformacja przestrzeni L -wymiarowej do Z -wymiarowej, wprowadza się zmienne dopełniające ξ_i i ξ'_i (określane dla każdego przykładu uczącego); problem uczenia sieci sprowadza się wtedy do obliczania zależności:

$$\varphi(w, \xi, \xi') = C \left[\sum_{i=1}^R (\xi_i + \xi'_i) \right] + \frac{1}{2} \cdot w^T \cdot w \quad (6.14)$$

pod warunkiem, że spełnione są ograniczenia:

$$\begin{cases} d_i - w^T \varphi(x) \leq \varepsilon + \xi_i \\ w^T \varphi(x) - y_i \leq \varepsilon + \xi_i \\ \xi_i \geq 0 \\ \xi'_i \geq 0 \end{cases} \quad (6.15)$$

Stała C (współczynnik regularyzacji) we wzorze (6.14) jest dobierana przez użytkownika w zależności od ε (Burges i in. 2000). Wykorzystując metodę mnożników Lagrange'a (α_i) układ równań zdefiniowany równaniami (6.14) i (6.15), przy wykorzystaniu tzw. funkcji jądra kernela $K(x, x_i)$ (funkcja liniowa, gaussowska, wielmianowa, sigmoidalna), można zapisać następująco:

$$y(x) = \sum_{i=1}^R (\alpha_i - \alpha'_i) K(x, x_i) + w \quad (6.16)$$

gdzie α_i, α'_i – wartości niezerowych mnożników Lagrange'a zależne od C i ε .

W metodzie SVM, niezależnie od rodzaju zastosowanego jądra i rodzaju rozważanego zadania (klasyfikacja, regresja), wyznaczenie parametrów w strukturze modelu $y = f(x_j)$ sprowadza się do wykorzystania programowania kwadratowego z ograniczeniami liniowymi. Stanowi to istotną przewagę metody SVM nad sieciami neuronowymi. Na etapie tworzenia modelu w metodzie SVM struktura modelu nie jest przyjmowana arbitralnie, co stanowi kolejną zaletę w odniesieniu do metody ANN. Ponadto fakt, że wartości sygnału wyjściowego uzyskiwane z modelu opisuje zależność analityczna (6.16), umożliwia wyznaczenie pochodnych $\frac{\partial y}{\partial x_i}$, co pozwala na bezpośrednią ocenę wpływu zmiennych niezależnych na wyniki symulacji i stanowi ważną zaletę w odniesieniu do wielu innych metod modelowania.

Poza istotnymi zaletami metody, ma ona również wady. Z uwagi na złożoność modelu mogą pojawić w celu poprawy zdolności predykcyjnych modelu można zastosować do estymacji jego parametrów algorytm roju pszczoł, genetyczny lub boostingu (wzmacniania) (Tian i in. 2012, Li i Kong 2014).

Z uwagi na swoje zalety metoda SVM w ostatnich latach rozwijała się bardzo dynamicznie i znajduje obecnie zastosowanie w wielu zagadnieniach regresyjnych i klasyfikacyjnych. W wielu przypadkach (prognoza przepływów i zmian klimatu, analiza danych przestrzennych i meteorologicznych, analiza obrazów i tekstów) uzyskane za jej pomocą wyniki charakteryzują się wysokim dopasowaniem do danych pomiarowych i są lepsze, niż otrzymane innymi metodami data mining.

6.8. Modele klasyfikacyjne (regresja logistyczna, liniowy model dyskryminacyjny)

Metody omówione w punktach 6.3 ÷ 6.7 można implementować w zadaniach klasyfikacji, co znajduje potwierdzenie w wielu pracach (Belanche i in. 2000, Comas i in. 2001, Szeląg i Siwicki 2017). Jednak metody te, co jest typowe dla modeli czarnej skrzynki, mają jedną kluczową wadę i jest nią brak fizycznej interpretacji parametrów w strukturze modelu. Dlatego też, w problemach z zakresu klasyfikacji, o ile to możliwe, stosuje się modele empiryczne, w których parametry mają interpretację fizyczną.

W zagadnieniach klasyfikacyjnych znajdują powszechnie zastosowanie model regresji logistycznej (RL) i liniowy model dyskryminacyjny (LDA). W modelu regresji logistycznej wyniki obliczeń mogą przyjmować wartości od 0 do 1, przez co możliwe jest także wyznaczanie prawdopodobieństwa przewyższenia zadanej granicznej wartości badanej zmiennej (zrzut ścieków przelewem, wystąpienie choroby, puchnięcie osadu czynnego w bioreaktorze, ryzyko bankructwa, wystąpienie awarii, ryzyko uszkodzenia obiektu (Bagley i in. 2001, Bayo i in. 2006, Tran i in. 2007, Szeląg i Siwicki 2017, Szeląg i in. 2018c). Regresja logistyczna stanowi szczególny przypadek uogólnionego modelu liniowego (GLM) (ang. Generalized Linear Model) postaci:

$$g(\mu) = \alpha_0 + \alpha_1 \cdot X_1 + \alpha_2 \cdot X_2 + \alpha_3 \cdot X_3 + \dots + \alpha_j \cdot X_j \quad (6.17)$$

gdzie: $g(\mu)$ – tzw. funkcja wiążąca opisująca zależność między średnią wartością zmiennej objaśnianej $\mu = E(Y|X_1 = x_1, X_2 = x_2, \dots, X_j = x_j)$ i liniową kombinacją predyktorów x_i , α_0 – wyraz wolny; $\alpha_1, \alpha_2, \dots, \alpha_j$ – współczynniki regresji wyznaczone metodą największej wiarygodności.

W modelu logitowym funkcja wiążąca nazywana logitem jest opisana równaniem:

$$g(p) = \text{logit}(p) = \ln\left(\frac{p}{1-p}\right) \quad (6.18)$$

gdzie $p = \mu = P(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots, X_j = x_j)$.

Na podstawie (6.17) i (6.18) otrzymuje się nową postać modelu logitowego:

$$p = P(Y = 1|X_1 = x_1, X_2 = x_2, \dots, X_j = x_j) = \frac{\exp(\alpha_0 + \sum_{i=1}^j \alpha_i x_i)}{1 + \exp(\alpha_0 + \sum_{i=1}^j \alpha_i x_i)} \quad (6.19)$$

gdzie p – prawdopodobieństwo przekroczenia przyjętej wartości granicznej badanej zmiennej.

Zaletą regresji logistycznej jest możliwość interpretacji parametrów funkcji $\exp(\alpha_j)$, stosując przy tym pojęcie szansy, wyznaczonej jako iloraz prawdopodobieństwa wystąpienia zdarzenia (p) do prawdopodobieństwa jego niewystąpienia ($1 - p$), co można wyrazić równaniem:

$$\frac{p}{1-p} = \theta(x_1, x_2, \dots, x_j) = \exp(\alpha_0 + \sum_{i=1}^j \alpha_i \cdot x_i) \quad (6.20)$$

Natomiast wpływ przyrostu o wartość Δx_i ($i = 1, 2, \dots, j$) wartości zmiennych niezależnych na szansę wystąpienia badanego zjawiska można wyznaczyć, określając tzw. iloraz szans opisany wzorem:

$$OR = \frac{\theta(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_j + \Delta x_j)}{\theta(x_1, x_2, \dots, x_j)} = \exp(\sum_{i=1}^j \alpha_i \cdot \Delta x_i) \quad (6.21)$$

Na podstawie ilorazu szans można stwierdzić, jak zmieni się szansa, że zmienna x_1 zwiększy się o Δx_1 przy ustalonych wartościach pozostałych zmiennych.

Poza tym, że zależność (6.19) ma charakter jawny, co stanowi niewątpliwą zaletę modelu RL, to rozkłady zmiennych (x_j) nie muszą być zgodne z normalnymi (jest to kluczowe założenie w wielu modelach). Kolejną zaletą modelu RL, to brak wymagań liniowości między zmienną zależną a zmienną niezależną. Istotną wadę metody RL stanowi jej ograniczone zastosowanie z uwagi na wyznaczaną liczbę separowanych klas. W klasycznym podejściu model logitowy pozwala identyfikować zależności opisujące przynależność zaledwie do dwóch a maksymalnie trzech klas, to jest znacznie mniej, niż w metodach RF, BT czy drzew klasyfikacyjnych. Wadę tę wyeliminowano w liniowym modelu dyskryminacyjnym (LDA), w którym można wyznaczyć modele do opisu wielu klas równocześnie.

6.9. Algorytm budowy modelu statystycznego do symulacji procesów biologicznych w oczyszczalniach ścieków

Ze względu na brak wytycznych dotyczących doboru odpowiedniej metody do modelowania określonych procesów środowiskowych wydaje się konieczne, aby na etapie konstruowania modelu rozważano zastosowanie wielu metod statystycznych, uwzględniając przy tym:

- wielkość zbioru danych,
- zależności (liniowa, nieliniowa lub mieszana) między zmiennymi objaśniającymi (x_i) a zmienną objaśnianą (y),
- liczbę zmiennych objaśniających (x_i , $i = 1, 2, 3, \dots, j$),

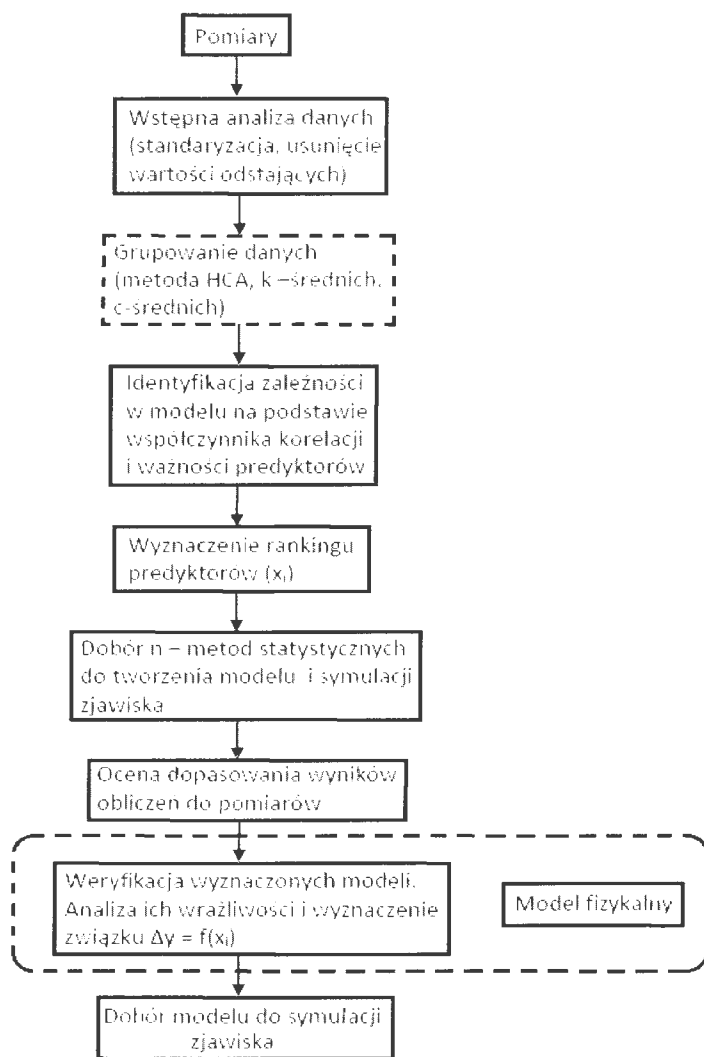
- znane ograniczenia, zalety i wady metod data mining i ich lokalne uwarunkowania.

Mimo tego, że obecnie metody eksploracji danych (ang. data mining) są powszechnie stosowane i są opracowywane ich liczne modyfikacje, to jednak wyraźnie odczuwalny jest brak sformalizowanych, jednoznacznych zasad tworzenia modeli statystycznych do symulacji oczyszczalni ścieków. Przedstawiane w literaturze zasady tworzenia takich modeli są bardzo ogólne i nie uwzględniają czynników, które mają istotny wpływ na metodykę obliczeń. W związku z powyższym w pracy zaproponowano ogólny algorytm doboru i budowy modelu statystycznego, przedstawiony na rys. 6.6.

Opracowany algorytm stanowi modyfikację klasycznej metody tworzenia modeli, w której jednym z kluczowych etapów jest wstępna analiza danych pomiarowych polegająca na usunięciu wartości odstających oraz standaryzacji i normalizacji danych. W przypadku modeli szeregów czasowych istotne znaczenie odgrywa analiza stacjonarności procesu, do której oceny można wykorzystać nieparametryczny test Manna–Kendalla na trend w szeregu (Rutkowska i Ptak 2012).

Kolejnym etapem tworzenia modelu, który powinien być obligatoryjnie wykonywany, jest analiza zbioru danych mająca na celu wydzielenie w nim danych o podobnych cechach, co zwykle ma istotny wpływ na dobór metody modelowania i konstrukcję modelu. Do tego celu można wykorzystać nienadzorowane metody grupowania danych. Do powszechnie stosowanych metod modelowania procesów hydrologicznych, meteorologicznych i ekologicznych należą metody k–średnich, c–średnich, analizy skupień i samoorganizujących się sieci neuronowych. Gdy w badanym obiekcie pomiary nie są wykonywane systematycznie (np. oczyszczalnia ścieków bez systemu on–line rejestracji jakości ścieków na dopływie) i dysponuje się jedynie ograniczonymi informacjami o dynamice modelowanego procesu, braki w danych mogą istotnie zniekształcić wyniki grupowania danych.

Następny etap tworzenia modelu dotyczy doboru zmiennych objaśniających (predyktorów). W przypadku procesów, w których wpływ poszczególnych czynników na analizowane zjawisko wykazano wcześniejszymi badaniami i wieloletnią eksploatacją obiektów, do modelu przyjmowane są predyktory podane przez technologa. W przeciwnym przypadku konieczne są analizy statystyczne dostępnych danych w celu odpowiedniego doboru predyktorów.



Rys. 6.6. Schemat doboru metody do tworzenia modelu statystycznego.

Pierwszą z zalecanych analiz jest wyznaczenie współczynnika korelacji (R) Pearsona między zmiennymi wejściowymi (niezależnymi lub objaśniającymi) a zmienną wyjściową (zależną) modelu. Gdy wartości R przyjmują odpowiednio duże wartości i wartość tzw. wskaźnika podbicia wariancji $VIF = 1: (1 - R^2)$ jest większa od 10, to można przyjąć, że występuje zależność liniowa między badanymi zmiennymi. Natomiast, gdy wartość VIF jest mniejsza, to liniowa zależność między wejściem a wyjściem modelu nie występuje i w celu doboru odpowiednich zmiennych objaśniających oblicza się wartość współczynnika

korelacji rangowej (Spearmana) lub wyznacza się tzw. ważność predyktorów IMP za pomocą drzew regresyjnych C&RT, drzew wzmacnianych lub lasów losowych. Wartości ważności IMP dla poszczególnych zmiennych mogą zmieniać się w zakresie 0÷1 i na podstawie minimalnej progowej wartości IMP_{\min} odnośna zmienna jest uwzględniana lub pomijana w modelu. Wyniki obliczeń ważności predyktorów są pomocne przy doborze zmiennych objaśniających do modelu, jednak często występują problemy z ustaleniem dolnej wartości IMP_{\min} i dlatego przy doborze zmiennych objaśniających stosuje się także metody wykorzystywane w ekonometrii.

Jedną z nich jest test F Snedecora, w którym formułuje się dwie hipotezy:

$H_0: (\alpha_1 = \alpha_2 = \dots \alpha_k = 0)$ – hipoteza zerowa mówiąca, iż parametry strukturalne nie różnią się istotnie od zera,

$H_A: (\alpha_1 \neq 0 \vee \alpha_2 \neq 0 \vee \dots \vee \alpha_k \neq 0)$ – hipoteza alternatywna mówiąca, iż istnieje chociaż jeden parametr istotnie różniący się od zera.

Następnie wyznacza się statystykę testową F w postaci:

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-(k+1)}} \quad (6.22)$$

gdzie: R – współczynnik korelacji, n – liczba pomiarów, k – liczba przyjmowanych zmiennych objaśniających.

Statystyka F ma rozkład F –Snedecora z k i $n - (k + 1)$ stopniami swobody. Dla przyjętego poziomu istotności α oraz przyjętych wartości dla k i $n - (k + 1)$ określa się wartość krytyczną $F_{(\alpha, k, n-(k+1))}^*$.

Gdy $F \leq F^*$, to brak przesłanek do odrzucenia H_0 , co oznacza, że przyjęte zmienne objaśniające nie objaśniają zmiennej objaśnianej, natomiast gdy $F > F^*$, to hipotezę H_0 należy odrzucić na rzecz H_A .

Kolejny etap tworzenia modelu statystycznego polega na doborze kilku metod obliczeniowych, co pozwoli wyznaczyć model najbardziej dokładny. W wielu przypadkach najlepsze dopasowanie wyników modelowania do danych pomiarowych nie stanowi jednego kryterium doboru metody, bo należy jeszcze uwzględniać czas symulacji i złożoność modelu. Równoczesne spełnienie wszystkich kryteriów jest bardzo trudne. Analizując szereg publikacji z zakresu modelowania procesów środowiskowych w tab. 3.1 ÷ 3.6 zaproponowano w pracy podział stosowanych metod ze względu na sposób badania zależności między wejściem a wyjściem modelu (obliczanie współczynnika korelacji, współczynnika podbicia wariancji i ważności predyktorów), na złożoność metody i możliwość interpretacji fizycznej zależności między zmiennymi.

W przypadku ustalenia na podstawie wartości współczynnika korelacji liniowej zależności między zmiennymi objaśniającymi a zmienną objaśnianą, do obliczeń modelowania zaleca się stosowanie regresji wielorakiej lub metod, w których istnieje możliwość uwzględnienia zależności liniowych (na przykład programowanie genetyczne). Otrzymane modele stanowią jawne zależności empiryczne, w których można bezpośrednio ocenić poprawność uzyskanych relacji między zmiennymi niezależnymi (x_i) a zmienną zależną (y).

Gdy rozważana zmienność $y = f(x_i)$ jest nieliniowa i dodatkowo dane wejściowe charakteryzują się stosunkowo niską jakością (wartości odstające, braki w danych, nieciągłości w ciągach pomiarowych), przy czym kluczowym kryterium stawianym modelowi jest możliwość identyfikacji wpływu zmiennych (x_i) na wynik symulacji, to wskazane jest zastosowanie metody RF lub BT. Przy tym należy mieć na uwadze, że do tej pory nie przeanalizowano dokładnie wpływu liczby zmiennych i ich rozkładów na wyniki symulacji, co może prowadzić do skokowych zmian wyjścia i spadku zdolności predykcyjnych modelu. Metody RF i BT mają ograniczone zastosowanie w przypadku, gdy zmienne x_i mają istotnie zróżnicowany wpływ na badane zjawisko. W zagadnieniach klasyfikacyjnych alternatywą dla metod RF i BT jest metoda regresji logistycznej, charakteryzującą się porównywalnymi zdolnościami prognozy, jak modele bardziej złożone. Ponadto analiza wpływu zmiennych objaśniających na wyniki symulacji jest także możliwa w metodzie SVM. W przypadku metod ANN nie jest możliwa ocena wpływu zmiennych x_i na wyniki obliczeń, co eliminuje zastosowanie tej grupy metod wówczas, gdy zależność $y = f(x_i)$ jest opisana funkcją analityczną.

W przypadku, gdy głównym kryterium uwzględnianym w procesie modelowania jest złożoność użytej metody, to można stosować metody MLR, RF lub BT charakteryzujące się stosunkowo nieskomplikowaną strukturą modelu. Największą złożonością charakteryzują się modele ANFIS i SVM. Gdy dane do obliczeń mają niską jakość, to lepsze dopasowanie wyników do pomiarów uzyskuje się metodą ANFIS, chociaż czas jej obliczeń jest dłuższy, niż metodą SVM. Mniejszą złożonością, niż wymienione wyżej metody, charakteryzują się metody CNN i MLP. W przypadku kaskadowej sieci neuronowej wprowadzenie kolejnych połączeń prowadzi do poprawy zdolności predykcyjnej modelu i dlatego tę metodę zaleca się stosować w pierwszej kolejności. Należy przy tym pamiętać, że dobierając strukturę modelu ANN należy poszukiwać kompromisu między jego złożonością i liczbą danych, którymi będą dysponowały zbiory uczący i testowy.

Powyższe rozważania odnośnie doboru metody modelowania mają charakter ogólnych wskazówek uzasadnionych obszerną analizą metod i licznymi pracami (tab. 3.1+tab.3.4).

W celu oceny dokładności wyznaczanych modeli statystycznych, czyli określenia ich zdolności predykcyjnych, stosuje się szereg miar dopasowania

modeli do pomiarów, przy czym do najczęściej stosowanych należą (Lee i in. 2002, Hong i Bhamidimarri 2003, Cinar i in. 2006, Lee i in. 2009, Chen i in. 2010, Han i Qiao 2012, Mirbagherii 2015, Boztoprak i in. 2015):

❖ w przypadku modeli regresyjnych:

– współczynnik korelacji (R):

$$R = \frac{\sum_{i=1}^N (y_{i,pom} - \bar{y}_{i,pom}) \cdot (y_{i,obl} - \bar{y}_{i,obl})}{\sqrt{\sum_{i=1}^N (y_{i,pom} - \bar{y}_{i,pom})^2} \cdot \sqrt{\sum_{i=1}^N (y_{i,obl} - \bar{y}_{i,obl})^2}} \quad (6.23)$$

– średni błąd kwadratowy (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (y_{i,pom} - y_{i,obl})^2} \quad (6.24)$$

– średni błąd bezwzględny (MAE):

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^N |y_{i,pom} - y_{i,obl}| \quad (6.25)$$

– średni błąd względny (MAPE):

$$MAPE = \frac{1}{N} \cdot \sum_{i=1}^N \left| \frac{y_{i,pom} - y_{i,obl}}{y_{i,pom}} \right| \quad (6.26)$$

❖ w przypadku modeli klasyfikacyjnych:

– czułość (SENS):

$$SENS = 100 \cdot \frac{TP}{TP+FN} \quad (6.27)$$

– swoistość (SPEC):

$$SPEC = 100 \cdot \frac{TN}{FP+FN} \quad (6.28)$$

– błąd zliczeniowy (R_z^2):

$$R_z^2 = 100 \cdot \frac{TN+TP}{TP+TN+FP+FN} \quad (6.29)$$

gdzie: TN, TP, FP, FN – liczba wskazań sklasyfikowanych zgodnie z tab. 6.1, odpowiednio: prawdziwie negatywna, prawdziwie pozytywna, fałszywie pozytywna, fałszywie negatywna.

Tab. 6.1. Przykładowa tabela klasyfikacyjna.

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

Stosunek wartości RMSE do MAE stanowi cenną informację o zmienności błędów w otrzymanych wynikach. Znacząca różnica między MAE i RMSE może wskazywać na występowanie znacznych błędów prognozy w zbiorze wyników. Współczynnik korelacji jest powszechnie stosowaną miarą do porównywania zdolności predykcyjnych modeli opracowanych przez różnych autorów.

Kolejny etap w zaproponowanym algorytmie doboru metody modelowania (rys. 6.6) może stanowić podstawę oceny poprawności utworzonej struktury modelu oraz jest istotny w odniesieniu do modeli, które mają być wdrożone w rzeczywistych obiektach. Etap weryfikacji modeli ma na celu określenie wpływu zmian w danych wejściowych ($\Delta x_i = x_{i+1} - x_i$) na wyniki symulacji ($\Delta y = y_{i+1} - y_i$). W przypadku układów eksploatowanych już wiele lat (oczyszczalnie ścieków, układy wentylacji i uzdatnianie wody itp.) stwierdzenie poprawności utworzonych zależności nie jest trudne i jest poparte odpowiednią wiedzą i doświadczeniem. W zaproponowanym algorytmie doboru metody zwrócono uwagę także na możliwość weryfikacji modelu statystycznego za pomocą modelu fizycznego. Podejście to jest bardzo rzadko stosowane, o czym świadczy mała liczba prac z tego zakresu (Sant'in 2015).

Ostatni etap algorytmu doboru metod polega na wyborze jednego modelu matematycznego do jego implementacji w warunkach rzeczywistych. Polega to na wyborze modelu charakteryzującego się najlepszym dopasowaniem wyników obliczeń do pomiarów i poprawnie prognozującego wpływ zmiennych wejściowych na badany proces.

W pracy wykorzystano zaproponowany algorytm przy doborze metod do modelowania procesów w oczyszczalniach ścieków. Zwracano przy tym uwagę na kilka własności charakteryzujących rozważaną metodę, mianowicie na: złożoność metody, szybkość obliczeń oraz dotychczasową częstość jej stosowania.

Należy przy tym zauważyć, że szereg metod było do tej pory rzadko wykorzystywanych w modelowaniu bioreaktorów, a wyniki uzyskiwane za pomocą jednej metody rzadko porównywano z wynikami innych metod. Dlatego

trudno jest ocenić zalety i wady poszczególnych metod i ich wartości aplikacyjne i takie zadanie było jednym z celów obecnej pracy.

9. Podsumowanie i wnioski końcowe

Oczyszczalnia ścieków stanowi złożony obiekt technologiczny, w którym zachodzi szereg procesów biochemicznych. Podstawą oceny działania oczyszczalni jest uzyskiwanie na wylocie założonej jakości ścieków. Z uwagi na losowy i zależny od pory roku charakter ilości i jakości ścieków na dopływie, zapewnienie wymaganej jakości ścieków na odpływie jest trudne. Wymaga to regulacji w trybie dynamicznym nastaw bioreaktora i wybranych obiektów technologicznych, co pozwala na spełnienie obowiązujących wymagań prawnych. W związku z powyższym podejmowane są próby wspomaganie procesu usuwania związków biogennych przy wykorzystaniu modeli matematycznych opracowanych na podstawie pomiarów gromadzonych na obiekcie. Do tego celu można stosować modele fizykalne lub parametryczne, tzw. modele czarnej skrzynki. Jednak z uwagi na znaczenie czasu przy wykonywaniu pojedynczej symulacji modelu oczyszczalni, obecnie w większości przypadków stosuje się modele parametryczne (szarej lub czarnej skrzynki).

W przypadku obiektów, na których gromadzone są pomiary w trybie ciągłym (on-line) (ilość i jakość ścieków na dopływie i odpływie z oczyszczalni, parametry operacyjne bioreaktora), takie modele pozwalają na uzyskanie zadowalającej jakości ścieków na wylocie i obniżenie kosztów eksploatacyjnych. W przypadku braku ciągłości w pomiarach jakości ścieków na dopływie do oczyszczalni, na przykład na skutek wadliwej pracy urządzeń pomiarowych lub gdy wartości wskaźników oznaczane są drogą analityczną, modele matematyczne zasilane danymi w trybie on-line mają ograniczone zastosowanie. Wynika to z braku wartości liczbowych zmiennych niezależnych uwzględnianych w modelach procesowych.

Poza problemami z dostępem do danych wejściowych w modelach matematycznych oczyszczalni pojawia się kolejna istotna kwestia dotycząca doboru metody do symulacji i interpretacji uzyskanych wyników. Należy rozważyć kwestię, czy tworzyć modele o prostszej strukturze wymagające mniejszych nakładów obliczeniowych i krótszego czasu obliczeń, czy może stosować modele bardziej złożone? Sformułowane zadanie jest trudne, bowiem stanowi kompromis między dokładnością prognoz i złożonością modelu i ma istotny wpływ na

podejmowanie decyzji przez technologa podczas eksploatacji oczyszczalni. Ponadto w przypadku modeli procesowych uzyskanych przy pomocy metod czarnej skrzynki (z uwagi na brak fizycznej interpretacji wyznaczonej struktury) mogą pojawić się problemy z jednoznacznością oceną ich przydatności. Warunkiem użyteczności tworzonych modeli do symulacji bioreaktorów jest poprawne odzwierciedlenie wpływu zmiennych niezależnych (ilość i jakość ścieków na dopływie do obiektu, parametry operacyjne reaktora) na zjawiska zachodzące w osadzie czynnym.

W pracy przedstawiono metodykę tworzenia modeli procesowych do symulacji ilości i jakości ścieków na wlocie i wylocie z oczyszczalni i osadu czynnego w bioreaktorze. Do tego celu zastosowano szereg metod data mining, zaczynając od najprostszych modeli liniowych, poprzez modele bazujące na teorii drzew regresyjnych i klasyfikacyjnych (drzewa wzmacniane, lasy losowe), a kończąc na złożonych metodach uczenia maszynowego, jak metoda wektorów nośnych, sztuczne sieci neuronowe typu perceptron wielowarstwowy (3 warstwy) oraz sieci neuronowe kaskadowe (tzw. deep learning) z co najmniej dwiema warstwami ukrytymi. W przypadku symulacji osadu czynnego obliczenia dotyczą jego sedymentacji z zastosowaniem modeli do prognozy wartości liczbowych i zmiennych lingwistycznych (osad puchnie lub nie). Wykorzystując opracowane modele statystyczne do symulacji procesów zachodzących w bioreaktorze przedstawiono metodykę optymalizacji działania oczyszczalni ścieków. Wykonane modele statystyczne wyznaczono na podstawie wieloletnich ciągów pomiarowych pochodzących z oczyszczalni ścieków Sitkówka–Nowiny, Trepcza i Rzeszów.

Ze względu na złożoną postać opracowanych modeli przedstawiono metodykę ich tworzenia pozwalającą na analizę niezawodności pracy oczyszczalni w ujęciu długo- i krótko-okresowym.

Na podstawie wykonanych w pracy obliczeń sformułowano następujące wnioski końcowe:

1. Do analizy wrażliwości struktury sieci neuronowej (funkcje aktywacji, liczby neuronów w warstwach ukrytych) i zmiennych niezależnych (wartości przepływów i opadów) w przypadku prognozy dopływu należy zastosować metodę regresji logistycznej. Opracowany model szarej skrzynki pozwala na ocenę wpływu uwzględnionych w nim zmiennych niezależnych na dokładność predykcji dopływu. W opracowanym modelu można analizować wpływ poszczególnych zmiennych i również grupy zmiennych w porównaniu z innymi rozwiązaniami, w których rozważa się jedynie wpływ pojedynczych zmiennych. Zaprezentowane w pracy podejście stanowi alternatywne rozwiązanie do obecnie stosowanych metod, gdzie zachodzi potrzeba wykonania obliczeń o dużym nakładzie czasowym lub wymaga się implementacji złożonych algorytmów numerycznych.

2. Wykazano na podstawie przeprowadzonych badań, że możliwe jest modelowanie jakości ścieków (BZT_5 , $ChZT$, Z_o , N_{og} , $N-NH_4$, P_{og}) na dopływie do oczyszczalni ścieków na podstawie pomiarów natężenia przepływu. Rezultaty te wskazują na to, że zróżnicowany stopień rozcieńczenia ścieków determinuje jakość ścieków, co wykazano na przykładzie dwóch obiektów (oczyszczalnie Sitkówka–Nowiny i Rzeszów). Spośród badanych w pracy metod najlepsze wyniki modelowania (najmniejsze wartości błędów prognozy wybranych wskaźników jakości ścieków) otrzymano stosując metodę kaskadowych sieci neuronowych (wielowarstwowych). Z kolei największe błędy prognozy otrzymano metodami lasów losowych bądź drzew wzmocnianych.

W przypadku niezadawalających zdolności predykcyjnych modeli do prognozy jakości ścieków wyznaczanych na podstawie wartości dopływów w oparciu o metody uczenia maszynowego, alternatywne rozwiązanie może stanowić opracowany klasyfikator wskaźników jakości i ładunków zanieczyszczeń. Zaproponowane narzędzie zbudowane przy użyciu metody wektorów nośnych pozwala na identyfikację stanu pracy oczyszczalni na dopływie do obiektu (przeciążenie hydrauliczne, niedobór biogenów) z dobowym wyprzedzeniem, na podstawie wartości natężenia przepływu zmierzonych w dobach poprzednich. Przedstawiony model pozwala na identyfikację typowych stanów na dopływie do oczyszczalni i umożliwia prognozowanie stanów incydentalnych niebezpiecznych dla bioreaktora. Wykonany w pracy model może stanowić użyteczne narzędzie dla technologa, który przy jego użyciu może podjąć odpowiednie decyzje eksploatacyjne dotyczące korekty parametrów operacyjnych bioreaktora (zmiana stopni recyrkulacji, stężenia tlenu rozpuszczonego w komorach osadu czynnego, dawkowania metanolu itp.).

3. Na podstawie wykonanych badań stwierdzono, że jest możliwe modelowanie sedymentacji osadu czynnego w osadniku wtórnym i jakości ścieków na wylocie z oczyszczalni ścieków w przypadku nieciągłości pomiarów jakości ścieków na dopływie do obiektu. Uzyskano to tworząc modele hybrydowe. Rozwiązanie to polega na zastąpieniu wartości zmiennych niezależnych (jakość ścieków) w modelach procesowych wynikami obliczeń. Najlepsze wyniki modelowania (najmniejsze błędy prognozy) sedymentacji osadu czynnego otrzymano za pomocą modeli hybrydowych bazujących na kombinacjach modeli kaskadowych sieci neuronowych jako modelu procesowego i modeli kaskadowych sieci neuronowych względnie wektorów nośnych lub perceptronu wielowarstwowego jako modeli wskaźników jakości ścieków. W przypadku modeli do prognozy jakości ścieków (azot amonowy, fosfor ogólny) najmniejsze wartości błędów uzyskano stosując metody wektorów nośnych i kaskadowych sieci neuronowych. Dla azotu ogólnego

zadowalające rezultaty obliczeń otrzymano modelami wykonanymi przy pomocy kaskadowych sieci neuronowych.

4. Do identyfikacji zdolności sedymentacyjnych osadu czynnego na istniejących obiektach można zastosować metodę regresji logistycznej (szara skrzynka), wykorzystując pomiary ilości i jakości ścieków na dopływie do oczyszczalni (BZT_5 , N_{og} , P_{og} , $N-NH_4$), pomiary temperatury w bioreaktorze i parametrów operacyjnych bioreaktora (stężenie osadu czynnego, stężenie tlenu rozpuszczonego, ilość dawkowanego PIX).

Należy podkreślić, że wyniki obliczeń otrzymane modelem logitowym nie odbiegają od rezultatów uzyskanych złożonymi modelami klasyfikacyjnymi opartymi na metodach sztucznych sieci neuronowych, wektorów nośnych, lasów losowych lub drzew wzmacnianych. Wykonane obliczenia (oczyszczalnie Sitkówka-Nowiny i Trepcza) wykazały, że wartości empirycznych współczynników w modelu regresji logistycznej zależą od dynamiki jakości ścieków na dopływie do oczyszczalni i rozwiązań technologicznych przyjętych na obiektach.

5. Kluczowe znaczenie na etapie tworzenia modeli procesowych (jakość ścieków na wylocie, sedymentacja osadu) przy pomocy metod czarnej skrzynki (sztuczne sieci neuronowe, metoda wektorów nośnych, lasy losowe, drzewa wzmacniane itp.) ma analiza wrażliwości modelu. Zaproponowana w pracy metoda analizy pozwala ocenić, na ile opracowany model poprawnie symuluje wpływ zmiennych niezależnych (w tym przede wszystkim zmiennych sterujących) na rozpatrywane zjawisko, co jest kluczowe z punktu widzenia wdrożenia opracowanych modeli statystycznych.
6. Opracowany model matematyczny oczyszczalni ścieków w oparciu o metody data mining (sieci neuronowe, regresja logistyczna) pozwala na badanie interakcji między procesami zachodzącymi w osadzie czynnym (sedymentacja) i jakością ścieków na wylocie (związki azotu i fosforu). Stanowi to cenne i oryginalne narzędzie, którego opracowania nie podjęto się do tej pory w kraju i zagranicą. Wykonane obliczenia wykazały, że na bazie opracowanego modelu możliwa jest optymalizacja pracy oczyszczalni ścieków (dobór odpowiednich nastaw, tj. korekta stężenia osadu czynnego, stężenia tlenu rozpuszczonego w komorach, ilości dawkowanego PIX) nawet w przypadku braku ciągłości danych pomiarowych dotyczących jakości ścieków na dopływie do obiektu.
7. Opracowany model probabilistyczny do analizy efektywności pracy bioreaktora (w odniesieniu do sedymentacji osadu czynnego) umożliwia ocenę niezawodności pracy oczyszczalni dla przyjętej strategii sterowania reaktorem w ujęciu długo-okresowym. Model ten pozwala na ocenę wpływu zmienności (prognozowanej za pomocą generatorów Monte Carlo bazujących

na wyznaczonych rozkładach teoretycznych na podstawie danych pomiarowych) ilości i jakości ścieków na dopływie i temperatury osadu czynnego (w ujęciu sezonowym wynikającym z pór roku) na dobór odpowiednich nastaw, co do tej pory w modelach o podobnym charakterze nie było uwzględniane. Jako nowe podejście w zaprezentowanym modelu należy uznać możliwość symulacji wpływu awarii analizatorów jakości ścieków na dopływie na dobór nastaw i niezawodność pracy oczyszczalni.

8. Wykazano, że długość szeregu pomiarowego, liczba danych i sposób ich gromadzenia (sposób ciągły - dane gromadzone w układzie on-line ze stałą rozdzielczością; sposób nieciągły, tj. okres między kolejnymi pomiarami (krok próbkowania) nie jest stały, są czynnikami determinującymi dobór metody do prognozowania jakości pracy oczyszczalni ścieków. Na podstawie opracowanych w pracy algorytmów można optymalizować długość okresu pomiarowego, przy uwzględnieniu którego ma być tworzony model matematyczny (o zadawalających zdolnościach predykcyjnych), przyjmując na początku eksperymentu określoną metodę data mining do symulacji zjawiska i sposób gromadzenia danych do jego opisu. W pracy wykazano, że stosując do prognozy działania oczyszczalni ścieków metody sieci neuronowych (MLP, CNN, SVM) wystarczą dane z rozdzielczością dobową obejmujące okres 6 miesięcy, tj. 180 wartości (dla zmiennej zależnej i zmiennych niezależnych) gromadzonych w układzie ciągłym lub nieciągłym, aby opracowany model charakteryzował się wysokimi zdolnościami predykcyjnymi. Przy czym modele opracowane na podstawie danych mierzonych w trybie on-line dla przyjętych długości szeregów czasowych charakteryzują się lepszymi zdolnościami predykcyjnymi, niż w przypadku danych mierzonych w trybie nieciągłym. Pośród badanych metod (MLR, GP, RF, BT, k-NN, MLP, CNN, SVM) stwierdzono, że metody MLP, CNN i SVM wykazują mniejszą wrażliwość na liczbę danych pomiarowych, niż modele opierające się na metodach lasów losowych, drzew wzmacnianych, k-najbliższych sąsiadów i programowania genetycznego. Tym samym przede wszystkim te metody powinny być dedykowane do modelowania oczyszczalni ścieków.
9. Możliwa jest ocena, symulacja i kontrola efektywności pracy oczyszczalni ścieków w trybie on-line na podstawie wartości współczynników niezawodności w odniesieniu do sedimentacji osadu czynnego (COR_{I_0} i jakości ścieków na wylocie ($COR_{F/M}$), wyrażonej w sposób uproszczony (obciążenie substratowe bioreaktora). Rozwiązanie to jest możliwe nawet w przypadku nieciągłości pomiarów wskaźników jakości ścieków na dopływie do oczyszczalni. Brakujące wartości wskaźników są modelowane na podstawie pomiarów natężenia dopływu do oczyszczalni przy zastosowaniu metod data mining. Podstawę doboru nastaw (stężenie osadu czynnego,

stężenie tlenu rozpuszczonego, ilość dawkowanego PIX) w bioreaktorze stanowi wartość współczynnika Φ będącego liniową kombinacją zmiennych dotyczących ilości i jakości ścieków (BZT_5/N_{og} , BZT_5/P_{og} , $N-NH_4$) oraz temperatury w komorach osadu czynnego. W pracy wykazano, że współczynnik Φ ma charakter sezonowy, zależy od pory roku i jest wrażliwy na przeciążenia hydrauliczne. Przyjęte rozwiązanie stanowi uproszczenie w odniesieniu do opracowywanych do tej pory modeli, gdy wartości nastaw ustalano drogą symulacji komputerowych metodami uczenia maszynowego.

Literatura

- Aarnio P., Minkkinen P. (1986): *Application of partial least – squares modelling in the optimization of a wastewater treatment plant*. Analytica Chimica Acta, 191, 457-460.
- Abba S.I., Elkiran G. (2017): *Effluent prediction of chemical oxygen demand from the wastewater treatment plant using artificial neural network application*. Procedia Computer Science, 120, 156–163.
- Abunama T., Othman F. (2017): *Time Series Analysis and Forecasting of Wastewater Inflow into Bandar Tun Razak Sewage Treatment Plant in Selangor, Malaysia*. IOP Conf. Series: Materials Science and Engineering, 210, 1-10.
- Abyaneh H.Z. (2014): *Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters*. Journal of Environmental Health Science and Engineering, 12, 1–8.
- Ahnert M., Marx C., Krebs P., Kuehn V. (2016): *A black-box model for generation of site-specific WWTP influent quality data based on plant routine data*. Water Science and Technology, 74, 2978–2986.
- Al-batah M.S., Alkhasawneh M.S., Tay L.T., Ngah U.K., Lateh H.H., Isa N.A.M. (2015): *Landslide occurrence prediction using trainable cascade forward network and multilayer perceptron*. Mathematical Problem in Engineering, 2015, 1–9.
- Alex J., Benedetti L., Copp J., Gernaey K.V., Jeppsson U., Nopens I., Pons M.N., Rieger L., Rosen Ch., Steyer J.P., Vanrolleghem P.A., Winkler S. (2008): *Benchmark Simulation Model no. 1 (BSM1)*. Technical Report, Department of Industrial Electrical Engineering and Automation, Lund University, LUTEDX/(TEIE7229)/1-62/(2008).
- Alsina X.F., Roda I.R., Sin G., Gernaey K.V. (2008): *Multi-criteria evaluation of wastewater treatment plant control strategies under uncertainty*. Water Research, 42, 4485–4497.
- Alsina X. F., Comas J., Roda I.R., Gernaey K.V., Rosen C. (2009): *Including the effects of filamentous bulking sludge during the simulation of wastewater treatment plants using a risk assessment model*. Water Research, 43, 4527–4538.

- Alsina X.F., Corominas L., Neumann M.B., Vanrolleghem P.A. (2012): *Assessing the use of activated sludge process design guidelines in wastewater treatment plant projects: A methodology based on global sensitivity analysis*. Environmental Modelling & Software, 38, 50-58.
- Alsina X.F., Arnell M., Amerlinck Y., Corominas L., Gernaey K.V., Guo L., Lindblom E., Nopen I., Porro J., Shaw A., Snip L., Vanrolleghem P.A., Jeppsson U. (2014): *Balancing effluent quality, economic cost and greenhouse gas emissions during the evaluation of (plant-wide) control/operational strategies in WWTPs*. Science of the Total Environment, 46, 616–624.
- Alsina X.F., Saagi R., Lindblom E., Thirsing C., Thornberg D., Gernaey K.V., Jeppsson U. (2014): *Calibration and validation of a phenomenological influent pollutant disturbance scenario generator using full-scale data*. *Water Research*, 51, 172-185.
- Andraka D., Dzienis L. (2013): *Modelowanie ryzyka w eksploatacji oczyszczalni ścieków*. Rocznik Ochrony Środowiska, 15, 1111-1125.
- Asadi A., Verma A., Yang K. (2016): *Wastewater treatment aeration process optimization: A data mining approach*. Journal of Environmental Management, 203, 1-10.
- Avella A.C., Görner T., Yvon J., Chappe P., Guinot-Thomas P., Donato P. (2011): *A combined approach for a better understanding of wastewater treatment plants operation: Statistical analysis of monitoring database and sludge physico-chemical characterization*. Water Research, 45, 981–992.
- Ayesa E., Sota A. De la, Grau P., Sagarna J.M., Salterain A., Suescun J. (2006): *Supervisory control strategies for the new WWTP of Galindo-Bilbao: the long run from the conceptual design to the full-scale experimental validation*. Water Science and Technology, 53, 193–201.
- Ayesha S., Balasubramanian S., Latha K.C. (2013): *A Time Series Analysis of Wastewater Inflow of Sewage Treatment Plant in Mysore, India*. International Journal of Current Research, 5, 248-253.
- Bagheri M., Mirbagheri S.A., Bagheri Z., Kamarkhani A.M. (2015): *Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach*. Process Safety and Environmental Protection, 95, 12-25.
- Bagley S.C., White H., Golomb B.A. (2001): *Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain*. Journal of Clinical Epidemiology, 54, 979–985.
- Balku S., Berber R. (2006): *Dynamics of an activated sludge process with nitrification and denitrification: Start-up simulation and optimization using evolutionary algorithm*. Computers and Chemical Engineering, 30, 490–499.
- Bannister C.A., Currie C.J., Preece A., Spasic I. (2014), *Automatic development of clinical prediction models with genetic programming: A case study in cardiovascular disease*. Health, 17, 200–201.

- Barbu M., Vilanova R., Meneses M., Santin I. (2017): *Global Evaluation of Wastewater Treatment Plants Control Strategies Including CO₂ Emissions*. IFAC PapersOnLine, 50-1, 12956–12961.
- Barbusiński K., Kościelniak H. (1995): *Influence of substrate loading intensity on floc size in activated sludge process*. Water Research, 29, 1703–1710.
- Bartkiewicz L., Studziński J. (2010): *Mathematical modeling of the hydraulic load of communal wastewater networks*. In: Modeling and Simulation 2010, G.K. Janssens, K. Ramakers, A. Caris. (eds), EUROSIS-ETI, Hasselt Belgium 2010, 156–160.
- Bartkiewicz L., Szelał B., Studziński J. (2016): *Ocena wpływu zmiennych wejściowych oraz struktury modelu sztucznej sieci neuronowej na prognozowanie dopływu ścieków komunalnych do oczyszczalni*. Ochrona Środowiska, 38, 29–36.
- Bayo J., Angosto J.M., Serrano-Aniorte J. (2006): *Evaluation of physicochemical parameters influencing bulking episodes in a municipal wastewater treatment plant*. Water Pollution VIII: Modelling, Monitoring and Management, Bologna, September 4-6, 531–541.
- Bayo J., López-Castellanos J. (2016): *Principal factor and hierarchical cluster analyses for the performance assessment of an urban wastewater treatment plant in the Southeast of Spain*. Chemosphere, 155, 152-162.
- Bechmann H., Nielsen M.K., Madsen H., Kjølstad-Poulsen N. (1999): *Grey-box modelling of pollutant loads from a sewer system*. Urban Water, 1, 71-78.
- Belanche L., Valde's J.J., Comas J., Roda I.R., Poch M. (2000): *Prediction of the bulking phenomenon in wastewater treatment plants*. Artificial Intelligence in Engineering, 14, 307-317.
- Berthouex P.M., Box G.E. (1996): *Time series models for forecasting wastewater treatment plant performance*. Water Research, 30, 1865 – 1875.
- Bezak-Mazur E., Stoińska R., Szelał B. (2016): *Ocena wpływu parametrów operacyjnych i występowania bakterii nitkowatych na objętościowy indeks osadu czynnego – studium przypadku*. Rocznik Ochrona Środowiska, 18, 487–498.
- Béraud B., Steyer J.P., Lemoine C., Latrille E., Manic G., Printemps-Vacquier C. (2007): *Towards a global multi objective optimization of wastewater treatment plant based on modeling and genetic algorithms*. Water Science and Technology, 56, 109-16.
- Bixio D., Parmentier G., Rousseau D., Verdonck F., Meirlaen J., Vanrolleghem P.A., Thoeye C. (2002): *A quantitative risk analysis tool for design/simulation of wastewater treatment plants*. Water Science and Technology, 46, 1-7.
- Bloom H.A. (1996): *Indirect measurement of key water quality parameters in sewage treatment plants*. Journal of Chemometrics, 10, 697-706.
- Bojanowska I., Downar D., Siedlecka E., Ropel E. (2000): *Ocena skuteczności miejskiej oczyszczalni ścieków w Kartuzach*. Ochrona Środowiska, 3, 21–24.
- Box G.E.P., Jenkins G.M. (1983): *Analiza szeregów czasowych. Prognozowanie i sterowanie*. PWN, Warszawa 1983.

- Bozorg-Haddad O, Soleimani S., Loáiciga H.A. (2017): *Modeling water-quality parameters using genetic algorithm–least squares support vector regression and genetic programming*. Journal of Environmental Engineering, 143, 145-153.
- Boztoprak H., Özbay Y., Güçlü D., Küçükhemek M. (2015): *Prediction of sludge volume index bulking using image analysis and neural network at a full-scale activated sludge plant*. Desalination and Water Treatment, 57, 1–11.
- Brdjanovic D. (1998): *Modeling biological phosphorus removal in activated sludge systems*. Ph.D, Thesis, Delft University of Technology, Delft 1998.
- Brdyś M.A., Grochowski M., Gminski T., Konarczak K., Drewa M. (2008): *Hierarchical predictive control of integrated wastewater treatment systems*. Control Engineering Practice, 16, 751–767.
- Brdyś M.A., Konarczak K. (2002): *Model predictive medium control layer for integrated sewer-wastewater treatment plant systems*. IIFAC International Conference on Technology, Automation and Control of Wastewater and Drinking Water Systems-Ti4SWiK'02, Gdansk–Sobieszewo, June 19-21, Poland.
- Brdyś M.A. (2014): *Integrated monitoring, control and security of Critical Infrastructure Systems*. Annual Reviews in Control, 38, 47–70.
- Breiman L. (2000): *Random forests*. Journal Machine Learning, 45, 5–32.
- Burges C. (1998): *A Tutorial on Support Vector Machines for Pattern Recognition*. In: U. Fayyad, Knowledge Discovery and Data Mining, Kluwer, 1–43.
- Butler D., Graham N.J.D. (1995): *Modeling Dry Weather Wastewater Flow in Sewer Networks*. Journal of Environmental Engineering, 121, 161–173.
- Canete J.F., Saz-Orozco P.D., Baratti R., Mulas M., Ruano A., Garcia-Cerezo A. (2016): *Soft-sensing estimation of plant effluent concentrations in a biological wastewater treatment plant using an optimal neural network*. Expert Systems With Applications, 63, 8–19.
- Capizzi G., Sciuto G.L., Monforte P., Napoli C. (2015): *Cascade feed forward neural network-based model for air pollutants evaluation of single monitoring stations in urban areas*. International Journal of Electronics and Telecommunications, 61, 327–332.
- Capodaglio A., Jones H., Novotny V., Feng X. (1991): *Sludge bulking analysis and forecasting: Application of system identification and artificial neural computing technologies*. Water Research, 25, 1217-1224.
- Carstensen J., Nielsen M.K., Strandbæk H. (1998): *Prediction of hydraulic load for urban storm control of a municipal WWT plant*. Water Science and Technology, 37, 363-370.
- Caruana R., Karampatziakis N., Yessenalina A. (2008): *An Empirical Evaluation of Supervised Learning in High Dimensions*. In: Proceedings of the 25. International Conference on Machine Learning, Helsinki, Finland, 2008.

- Chang T., Duzinkiewicz K., Brdys M.A. (2004): *Bounding approach to parameter estimation without priori knowledge on model structure error*. IFAC Large Scale Systems: Theory and Applications, Osaka, Japan, 2004, 221–226.
- Cheema J.J.S., Sankpal N.V., Tambe S.S., Kulkarni B.D. (2002): *Genetic programming assisted stochastic optimization strategies for optimization of glucose to gluconic acid fermentation*. Biotechnology Progress, 18, 1356–1365.
- Chen H.W., Ning S.K., Yu R.F., Hung M.S. (2006): *Optimizing the monitoring strategy of wastewater treatment plants by multiobjective neural networks approach*. Environmental Monitoring and Assessment, 125, 325–32.
- Chen H.W., Yu R.F., Ning S.K., Huang H.C. (2010): *Forecasting effluent quality of an industry wastewater treatment plant by evolutionary grey dynamic model*. Resources, Conservation and Recycling 54, 235–241.
- Chen Q., Qiao J., Zou Y.M. (2017): *A self-organizing recurrent neural network*, International Journal of Artificial Intelligence and Applications, 8, 11–23.
- Chuchro M. (2009): *Prediction of the sewage treatment plant inflow parameters*. Akademia Górniczo-Hutnicza, Wydział Geologii, Geofizyki i Ochrony Środowiska, Kraków 2009.
- Çinar Ö., Hasar H., Kinaci C. (2006): *Modeling of submerged membrane bioreactor treating cheese whey wastewater by artificial neural network*. Journal of Biotechnology, 123, 204–209.
- Clara N. (2008): *Neural networks complemented with genetic algorithms and fuzzy systems for predicting nitrogenous effluent variables in wastewater treatment plants*. WSEAS Transactions on Systems, 6, 695–705.
- Comas J., Dzeroski S., Gibert K., Roda I.R., Sánchez-Marré M. (2001), *Knowledge discovery by means of inductive methods in wastewater treatment plant data*. AI Communications, 14, 45–62.
- Comas J., Roda I.R., Gernaey K.V., Rosen C., Jeppsson U., Poch M. (2008): *Risk assessment modelling of microbiology-related solids separation problems in activated sludge systems*. Environmental Modelling & Software, 23, 1250–1261.
- Corominas L., Larsen H.F., Alsina X.F., Vanrolleghem P.A. (2013): *Including Life Cycle Assessment for decision-making in controlling wastewater nutrient removal systems*. Journal of Environmental Management, 128, 759–767.
- Corominas L., Foley J., Guest J.S., Hospido A., Larsen H.F., Morera S., Shaw A. (2013): *Life cycle assessment applied to wastewater treatment: State of the art*. Water Research, 47, 5480–5492.
- Cortés U., Martínez M., Comas J., Sánchez-Marré M., Poch M., Rodríguez-Roda I. (2003): *A conceptual model to facilitate knowledge sharing for bulking solving in wastewater treatment plants*. AI Communications, 16, 279–289.
- Côté M., Jean B.P.A.G., Lessard P., Thibault J. (1995): *Dynamic modelling of the activated sludge process: improving prediction using neural networks*. Water Research, 29, 995–1004.

- Cristea V.M., Pop C., Agachi P.S. (2009): *Artificial neural networks modelling of PID and model predictive controlled waste water treatment plant based on the Benchmark Simulation Model No.1*. 19th European Symposium on Computer Aided Process Engineering–ESCAPE19J, Jeżowski and J, Thullie (eds.), 1183–1188.
- Czekański A. (2017): *Efekty stosowania PIX-u 113 do wspomagania procesu technologicznego w rozbudowanej oczyszczalni ścieków w Trepczy*. Seminarium naukowo–techniczne Opalenica–Poznań, 20–22 wrzesień 2017, 175–211.
- Dąbek L., Szelaż B., Picheta – Oleś A. (2017): *Assessment of the possibility of using data mining methods to predict sorption isotherms of selected organic compounds on activated carbon*. E3S Web of Conferences 22, 00174 (2017).
- Dellana S.A., West D. (2009): *Predictive modeling for wastewater applications: Linear and nonlinear approaches*. Environmental Modelling & Software, 24, 96–106.
- Devisscher M., Ciacci G., Fe L., Benedetti L., Bixio D., Thoeye C., Gueldre G., Marsili-Libelli S., Vanrolleghem P.A. (2006): *Estimating costs and benefits of advanced control for wastewater treatment plants – the MAgIC methodology*. Water Science and Technology, 53, 215–223.
- Dogan E., Ates A., Yilmaz E.C., Eren B. (2008): *Application of Artificial Neural Networks to Estimate Wastewater Treatment Plant Inlet Biochemical Oxygen Demand*. Environmental Progress, 27, 439–446.
- Djeddou M., Achour B. (2015): *The use of a neural network technique for the prediction of sludge volume index in municipal wastewater treatment plant*. Larhyss Journal, 24, 351-370.
- Drewnowski J., Zmarzły M. (2017): *The use of mathematical models for diagnosis of activated sludge systems in WWTP*. E3S Web of Conferences 22, 00037 (2017).
- Ebrahimi M., Gerber E.L., Rockaway T.D. (2017): *Temporal performance assessment of wastewater treatment plants by using multivariate statistical analysis*. Journal of Environmental Management, 193, 234-246.
- Eisenberge D., Soller J., Sakaji R., Olivier A. (2001): *Methodology to evaluate water and wastewater treatment plant reliability*. Water Science and Technology, 43, 91-99.
- El-Din A.G., Smith D.W. (2002): *Modelling approach for high flow rate in wastewater treatment operation*. Journal of Environmental Engineering Science, 1, 275–291.
- Ellis G.W., Grasso D., Ge X. (1993): *ARMA Processes and Reliability-Based Design of Wastewater-Treatment Facilities*. Journal of Environmental Engineering, 119, 463-477.
- Fallah-Mehdipour E., Bozorg Haddad O., Mariño M.A. (2013): *Prediction and simulation of monthly groundwater level by Genetic Programming*. Journal of Hydro-environment Research, 7, 253–260.
- Fernandes J.A., Irigoien X., Boyra G., Lozano J.A., Inza I. (2008): *Optimizing the number of classes in automated zooplankton classification*. Journal of Plankton Research, 31, 19–29.

- Fernandez F.J., Seco A., Ferrer J., Rodrigo M.A. (2009): *Use of neurofuzzy networks to improve wastewater flow-rate forecasting*. Environmental Modelling & Software, 24, 686–693.
- Fiałkowska E., Fyda J., Pajdak-Stós A., Wiąckowski K. (2010): *Osad czynny. Biologia i analiza mikroskopowa*. Wyd. Seidel-Przywecki, Warszawa 2010.
- Friedman J.H. (2001): *Greedę function approximation: A gradient boosting machine*. The Annals of Statistics, 29, 1189–1232.
- Friedman J.H. (2002): *Stochastic gradient boosting*. Computational Statistics and Data Analysis, 38, 367–378.
- Garcia H.L., Gonzales I.M. (2004), *Self-organizing map and clustering for wastewater treatment monitoring*. Engineering Applications of Artificial Intelligence, 17, 215–225.
- Gernaey K.V., Mussati M., Yuan Z., Nielsen M.K., Jørgensen S.B. (2002): *Control strategy evaluation for combined N and P removal using a benchmark wastewater treatment plant*. 15th Triennial World Congress, Barcelona, Spain, 1–6.
- Gernaey K.V., Alsina X.F., Rosen Ch., Benedetti L., Jeppsson U. (2011): *Dynamic influent pollutant disturbance scenario generation using a phenomenological modelling approach*. Environmental Modelling & Software, 26, 1255–1267.
- Grochowski M., Brdýs M.A., Gminski T. (2004): *Intelligent control structure for control integrated wastewater systems*. IFAC Large Scale Systems: Theory and Applications, Osaka, Japan, 2004, 245 – 250.
- Gyeongdong B., Seong-Pyo C., Sudaе K., Yejin K., Hyosoo K., Changwon K., Sungshin K. (2012): *Modular Neural Networks Prediction Model Based A2/O Process Control System*. International Journal of Precision Engineering and Manufacturing, 13, 905–913.
- Güçlü D., Dursun Ş. (2010): *Artificial neural network modelling of a large-scale wastewater treatment plant operation*. Bioprocess and Biosystems Engineering, 33, 1051–1058.
- Haimi H., Mulas M., Corona F., Marsili-Libelli S., Lindell P., Heinonen M., Vahala R. (2016): *Adaptive data-derived anomaly detection in the activated sludge process of a large-scale wastewater treatment plant*. Engineering Applications of Artificial Intelligence, 52, 65–80.
- Han H.G., Qiao J.F. (2012): *Prediction of activated sludge bulking based on a self-organizing RBF neural network*. Journal of Process Control, 22, 1103–1112.
- Han H.G., Qiao J.F. (2013): *Hierarchical Neural Network Modeling Approach to Predict Sludge Volume Index of Wastewater Treatment Process*. IEEE Transactions on Control Systems Technology, 21, 2423–2431.
- Han H., Li Y., Qiao J. (2014): *A fuzzy neural network approach for online fault detection in wastewater treatment process*. Computers & Electrical Engineering, 40, 2216–2226.

- Han H.G., Li Y., Guo Y.N., Qiao J.F. (2016): *A soft computing method to predict sludge volume index based on a recurrent self-organizing neural network*. Applied Soft Computing, 38(C), 477-486.
- Han H., Chen Z.Y., Qiao J., Zhang H. (2016): *Soft-sensor method for effluent ammonia nitrogen based on interval type-2 fuzzy neural networks*. 36th Chinese Control Conference (CCC), 26–28 lipiec 2017, 120–126.
- Häck M., Köhne M. (1996): *Estimation of wastewater process parameters using neural networks*. Water Science and Technology, 33, 101-115.
- Hecht-Nielsen R. (1987): *Kolmogorov's mapping neural network existence theorem*. In Proceedings of the 1st IEEE International Joint Conference of Neural Networks, 3, 11–14, New York, NY, USA, 1987.
- Heddad S., Lamda H., Filali S. (2016): *Predicting effluent Biochemical Oxygen Demand in a wastewater treatment plant using Generalized Regression Neural Network based approach: a comparative study*. Environmental Processes, 3, 153–165.
- Helling C., Schellen A.A.J.C., Mulder J.W., Loosdrecht M.C.M., Heijnen J.J. (1998): *The sharon process: An innovative method for nitrogen removal from ammonium-rich waste water*. Water Science and Technology, 37, 135-142.
- Henze M., Gujer W., Mino T., Matsuo T., Wentzel M.C., Marais G.V.R., Loosdrecht M.C.M. (1999): *Activated Sludge Model No 2d*. Water Science and Technology, 39, 165–182.
- Henze M., Harremoes P., Jansen J., Arvin E. (2002): *Wastewater Treatment, Biological and Chemical Processes*. Springer-Verlag Berlin Heidelberg.
- Hoef J.M., Temesgen H., Gómez S. (2013): *A Comparison of the Spatial Linear Model to Nearest Neighbor (k-NN) Methods for Forestry Applications*. PLoS ONE, 8, 1–13.
- Hong Y.S., Bhamidimarri R. (2003): *Evolutionary self-organising modelling of a municipal wastewater treatment plant*. Water Research, 37, 1199–1212.
- Hong Y.S.T., Paik B.C. (2007): *Evolutionary Multivariate Dynamic Process Model Induction for a Biological Nutrient Removal Process*. Journal of Environmental Engineering, 133, 1–10.
- Hongbin L., Mingzhi H., ChangKyoo Y. (2014): *A fuzzy neural network-based soft sensor for modeling nutrient removal mechanism in a full-scale wastewater treatment system*. Desalination and Water Treatment, 51, 6184–6193.
- Huo J.S., Seaver W.L., Robinson R.B., Cox Ch.D. (2005): *Application of Time Series Models to Analyze and Forecast the Influent Components of Wastewater Treatment Plants (WWTPs)*. World Water and Environmental Resources Congress 2005, 1–5.
- Iacopozzi I., Innocenti V., Marsili-Libelli S., Giusti E. (2007): *A modified Activated Sludge Model No. 3 (ASM3) with two-step nitrification – denitrification*. Environmental Modelling & Software, 22, 847-861.
- Ingildsen P., Rosen C., Gernaey K.V., Nielsen M.K., Guildal T., Jacobsen B.N. (2006): *Modelling and control strategy testing of biological and chemical phosphorus removal at Avedøre WWTP*. Water Science and Technology, 53, 105–113.

- Jami M.S., Mujeli M., Kabbashi N.A. (2011): *Simulation of ammoniacal nitrogen effluent using feedforward multilayer neural networks*. African Journal of Biotechnology, 10, 18755-18762.
- Janus T. (2014): *Integrated mathematical model of a MBR reactor including biopolymer kinetics and membrane fouling*. Procedia Engineering, 70, 882-891.
- Jenkins D., Richard M.G., Digger G.T. (2003): *Manual on the Caused and Control of Activated Sludge Bulking, Foaming and other Solids Separation Problems*. Lewis Publishers, New York, NY, USA.
- Jurasz J., Piasecki A. (2015): *Application of artificial neural networks in discharged wastewater volume forecasting case study Toruń*. Logistyka, 4, 9061–9066.
- Kaczor G. (2008): *Influence of air temperature on sewage temperature in sewerage system and in biological reactor*. Infrastructure and Ecology of Rural Areas, 3, 129–137.
- Khademikia S., Haghizadeh A., Godini H., Khorramabadi G.S. (2016): *Artificial Neural Network-Cuckoo Optimization Algorithm (ANN-COA) for Optimal Control of Khorramabad Wastewater Treatment Plant*. Iranian Civil Engineering Journal, 2, 555-567.
- Kiczko A., Szeląg B., Koziół A.P., Krukowski M., Kubrak E., Kubrak J., Romanowicz R.J. (2018): *Optimal Capacity of a Stormwater Reservoir for Flood Peak Reduction*. Journal of Hydrologic Engineering, 23, 1–9.
- Kim D., Bowen J.D., Ozelkan E.C. (2015): *Optimization of wastewater treatment plant operation for greenhouse gas mitigation*. Journal of Environmental Management, 163, 39-48.
- Kocev D., Dzeroski S., White M.D., Newell G.R., Griffioen P. (2009): *Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition*. Ecological Modelling, 220,1159–1168.
- Koza J.R. (1992): *Genetic programming: on the programming of computers by natural selection*. MIT Press, Cambridge, MA, 1992.
- Kudłak B., Wiczerzak M., Yotova G., Tsakovski S., Simeonov V., Namieśnik J. (2016): *Environmental risk assessment of Polish wastewater treatment plant activity*. Chemosphere, 160, 181-188.
- Kundu P., Debsarkar A., Mukherjee S., Kumar ., (2014): *Artificial neural network modelling in biological removal of organic carbon and nitrogen for the treatment of slaughterhouse wastewater in a batch reactor*. Environmental Technology, 35, 1296-1306.
- Kusiak A., Zhang Z. (2010): *Short-horizon prediction of wind power: A Data-Driven Approach*. IEEE Transactions on Energy Conversion, 25, 1112–1122.
- Kusiak A., Verma A., Wei X. (2013): *A data-mining approach to predict influent quality*. Environmental Monitoring and Assessment, 185, 2197–2210.

- Kusiak A., Zeng Y., Zhang Z. (2013): *Modeling and analysis of pumps in a wastewater treatment plant: A data-mining approach*. Engineering Applications of Artificial Intelligence, 26, 1643–161.
- Kusiak A., Wei X. (2013): *Optimization of the Activated Sludge Process*. Journal of Energy Engineering, 139, 12–17.
- Kusiak A., Wei X. (2014): *Prediction of methane production in wastewater treatment facility: a data-mining approach*. Annals of Operations Research, 216, 71–81.
- Kuswanto H., Salamah M., Fachruddin M.I. (2017): *Random Forest Classification and Support Vector Machine for Detecting Epilepsy using Electroencephalograph Records*. American Journal of Applied Science, 14, 533–539.
- Lackner S., Gilbert E.M., Vlaeminck S.E., Joss A., Horn H., Loosdrecht M.C.M. (2014): *Full-scale partial nitrification/anammox experiences - An application survey*. Water Research, 55, 292–303.
- Langergraber G., Alex J., Weissenbacher N., Woener D., Ahnert M., Frehman T., Halft N., Hobus I., Plattes M., Spering V., Winkler S. (2008): *Generation of diurnal variation for influent data for dynamic simulation*. Water Science and Technology, 50, 131-138.
- Lawrence A.W., McCarty P.L., Fachruddin M.I. (1970): *Unified basis for biological treatment design and operation*. Journal of the Sanitary Engineering Division 96, 757–778.
- Lee D.S., Jeon C.O., Park J.M., Chang K.S. (2002): *Hybrid neural network modelling of a full - scale Industrial Wastewater Treatment Process*. Biotechnology and Bioengineering, 78, 670–682.
- Lee D.S., Vanrolleghem P.A., Park J.M. (2005): *Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant*. Journal of Biotechnology, 115, 317–328.
- Lee D.S., Lee M.W., Woo S.H., Kim Y.J., Park J.M. (2006): *Nonlinear dynamic partial least squares modeling of a full-scale biological wastewater treatment plant*. Process Biochemistry, 41, 2050–2057.
- Lee Y., Cho J., Seo Y., Lee J.W., Ahn K. (2002), *Modeling of submerged membrane bioreactor process for wastewater treatment*. Desalination, 146, 451–457.
- Lee H.W., Lee M.W., Park J.M. (2009): *Multi-scale extension of PLS algorithm for advanced on-line process monitoring*. Chemometrics and Intelligent Laboratory Systems, 98, 201-212.
- Lee S., Kim J.C., Jung H.S., Lee M.J., Lee S. (2017): *Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea*. Journal Geomatics, Natural Hazards and Risk, 8, 1185–1203.
- Li X.Z., Kong J.M. (2014): *Application of GA-SVM method with parameter optimization for landslide development prediction*. Natural Hazards and Earth System Sciences, 14, 525–533.

- Liu Y., Guo J., Wang Q., Huang D. (2016): *Prediction of Filamentous Sludge Bulking using a State-based Gaussian Processes Regression Model*. Scientific Reports, 6, 31303, 1-11.
- Lubos J., Kaletova T., Sedmakova M., Balazova P., Cervenanska A. (2017): *Comparison of service characteristics of two town's WWTP*. Journal of Ecological Engineering, 18, 61-67.
- Luccarini L., Porrá E., Spagni A., Ratini P., Grilli S., Longhi S., Bortone G. (2002): *Soft sensors for control of nitrogen and phosphorus removal from wastewaters by neural networks*. Water Science and Technology, 45, 101-107.
- Luo F., Yu R., Xu Y., Li Y. (2009): *Effluent quality prediction of wastewater treatment plant based on fuzzy-rough sets and artificial neural networks*. 6th International Conference on Fuzzy Systems and Knowledge Discovery, 47-51.
- Lou I., Zhao Y. (2012): *Sludge Bulking Prediction Using Principle Component Regression and Artificial Neural Network*. Mathematical Problem in Engineering, 2012, 1-17.
- Łapczyński M. (2010): *Drzewa klasyfikacyjne i regresyjne w badaniach marketingowych*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków 2010.
- Łomotowski J., Szpindor A. (1999): *Nowoczesne systemy oczyszczania ścieków*. Wydawnictwo Arkady, Warszawa 1999.
- Ma S., Zeng S., Dong X., Chen J., Olsson G. (2013): *Short-term prediction of influent flow rate and ammonia concentration in municipal wastewater treatment plants*. Frontiers of Environmental Science & Engineering, 8, 128-136.
- Maachou R., Lefkir A., Merabtene T., Hamriche A., Bermad A. (2017): *Contribution to optimize decision parameters in activated-sludge process using ANFIS model*. MATEC Web of Conferences 120, 05001, ASCMCES-17, 1-8.
- Manu D.S., Thalla A.K. (2017): *Artificial intelligence models for predicting the performance of biological wastewater treatment plant in the removal of Kjeldahl Nitrogen from wastewater*. Applied Water Science, 7, 3783-3791.
- Martin C., Vanrolleghem P.A. (2014): *Analysing, completing, and generating influent data for WWTP modelling: A critical review*. Environmental Modelling & Software, 60, 188-201.
- Martins A.M.P., Heijnen J.J., Loosdrecht M.C.M. (2004): *Bulking sludge in biological nutrient removal systems*, Biotechnology and Bioengineering, 86, 25-135.
- Martins A.M.P., Pagilla K.R., Heijnen J.J., Loosdrecht M.C.M. (2004): *Bulking filamentous sludge - a critical review*. Water Research, 38, 793-817.
- Martins A.M.P., Picioreanu C., Heijnen J.J., Loosdrecht M.C.M. (2004): *Three-Dimensional Dual-Morphotype Species Modeling of Activated Sludge Flocs*. Environmental Science & Technology, 38, 5632-5641.
- Mąkinia J. (2010): *Mathematical Modelling and Computer Simulation of Activated Sludge Systems*. IWA Publishing, London.

- Messaoud D., Bachir A., Maurice M. (2013): *Determination and analysis of daily reliability level of municipal wastewater treatment plant*. Courier du Savoir, 17, 39-46.
- Minsoo K., Yejin K., Hyosoo K., Wenhua P., Changwon K. (2016): *Operator decision support system for integrated wastewater management including wastewater treatment plants and receiving water bodies*. Environmental Science and Pollution Research, 23, 10785–10798.
- Minsoo K., Yejin K., Hyosoo K., Wenhua P., Changwon K. (2016): *Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant*. Frontiers of Environmental Science & Engineering, 10, 299–310.
- Mirbagheri S.A., Bagheri M., Ehteshami M., Bagheri Z., Poursaghar M. (2015): *Modeling of mixed liquor volatile suspended solids and performance evaluation for sequencing batch reactor*. Journal of Urban and Environmental Engineering, 9, 54-65.
- Mirbagheri S.A., Bagheri M., Boudaghpour S., Ehteshami M., Bagheri Z. (2015): *Performance evaluation and modeling of a submerged membrane bioreactor treating combined municipal and industrial wastewater using radial basis function artificial neural networks*. Journal of Environmental Health Science & Engineering, 13, 13–17.
- Mulas M., Tronci S., Corona F., Haimi H., Lindell P., Heinonen M., Vahala R., Baratti R. (2015): *Predictive control of an activated sludge process: An application to the Viikinmäki wastewater treatment plant*. Journal of Process Control, 35(C), 89–100.
- Mulas M., Corona F., Sirviö J., Hyvönen S., Vahala R. (2016): *Full-scale implementation of an advanced control system on a biological wastewater treatment plant*. IFAC-PapersOnLine, 49-7, 1163–1168.
- Mulder A. (2003): *The quest for sustainable nitrogen removal technologies*. Water Science and Technology, 48, 67-75.
- Nawar S., Mouazen A.M. (2017): *Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon*. Sensors, 17, 2–22.
- Nguyen D.H., Latifi M.A., Lesage F., Mulholland M. (2013): *Dynamic simulation and optimization of wastewater treatment plants*. 2013 International Conference on Process Control (PC) June 18–21, 2013, Štrbské Pleso, Slovakia, 407–414.
- Nielsen P.H., Kragelund C., Seviour R.J., Nielsen J.L., *Identity and ecophysiology of filamentous bacteria in activated sludge*. FEMS Microbiology Reviews, 33, 969–998.
- Niku S., Schroeder E.D., Tchobanous G., Samaniego F.J. (1981): *Performance of activated sludge process: reliability, stability and variability*. Environmental Protection Agency, EPA Grant No, R805097-01, 124p.

- Nitze U., Schulthess U., Asche H. (2012): *Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised cro type classification*. Proceedings of the 4th GEOBIA, May 7-9, 2012, Rio de Janeiro, Brazil, 35–40.
- Noi P.T., Kappas M. (2017): *Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery*. Sensors (Basel) 2018, 18, 2–20.
- Ogutlu J.O., Piepho H.P., Streeck T.S. (2011): *A comparison of random forests, boosting and support vector machines for genomic selection*. BMC Proceedings 2011, 5(Suppl 3):S11.
- Olivera S.C., Sperling M. (2008): *Reliability analysis of wastewater treatment plants*. Water Research, 42, 1182-1194.
- Ossowski S. (2013): *Neural Networks for information processing*. Publishing House of the Warsaw University of Technology, Warszawa 2013.
- Ozkan O., Ozdemir O., Azin S.T. (2009): *Prediction of Biochemical Oxygen Demand in a wastewater treatment plant by Artificial Neural Networks*. Asian Journal of Chemistry, 21, 4821-4830.
- Pai T.Y., Tsai Y.P., Lo H.M., Tsai C.H., Lin C.Y. (2005): *Grey and neural network prediction of suspended solids and chemical oxygen demand in hospital wastewater treatment plant effluent*. Computers and Chemical Engineering, 31, 1272–1281.
- Pai T.Y., Yang P.Y., Wang S.C., Lo M.H., Chiang C.F., Kuo J.L., Chu H.H., Su H.C., Yu L.F., Hu H.C., Chang Y.H. (2011), *Predicting effluent from the wastewater treatment plant of industrial park based on fuzzy network and influent quality*. Applied Mathematical Modelling, 35, 3674–3684.
- Pawlowski, C.W., Rhea, L., Shuster, W.D., Barden, G. (2014). *Some factors affecting inflow and infiltration from residential sources in a core urban area: Case study in Columbus, Ohio, neighborhood*. Journal of Hydraulic Engineering, 140, 105-114.
- Pijáková I., Derco J. (2013): *The assessment of potential operation of wastewater treatment plant by dynamic simulations*. Acta Chimica Slovaca, 6, 20—24.
- Ping Y.U. (2015): *Multi-model modeling and its application of urban sewage treatment based on clustering analysis*. 2015 Global Conference on Polymer and Composite Materials (PCM 2015), IOP Conf, Series: Materials Science and Engineering 87, 012-029.
- Poutiainen H., Niska H., Heinonen H.T., Kolehmainen M. (2010): *Use of sewer on-line total solids data in wastewater treatment plant modelling*. Water Science and Technology, 62, 743-750.
- Qiao J., Chen Q., Han H. (2011): *The Chemical Oxygen Demand Modelling Based on a Dynamic Structure Neural Network*. In: F, Sebastián, G, Einschlag (Editors), Waste Water- Evaluation and Management, 93–114.
- Qiao J., Yang W. (2011): *Recurrent High Order Neural Network Modeling for Wastewater Treatment Process*. Journal of Computers, 8, 1570–1577.

- Qiao J., Hu Z., Li W. (2016): *Soft Measurement Modeling Based on Chaos Theory for Biochemical Oxygen Demand (BOD)*. Water, 8, 581, 2–21.
- Rak J. (2004): *Istota ryzyka w funkcjonowaniu systemu zaopatrzenia w wodę*. Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów 2004.
- Rak J., Tchórzewska-Cieślak B. (2009): *O bezpieczeństwie zaopatrzenia w wodę*. Monografie Komitetu Inżynierii Środowiska PAN, vol. 59, tom 2, s.171-178.
- Rak J., Tchórzewska-Cieślak B. (2006): *Five – parametric matrix to estimate risk connected with water supply system operating*. Environment Protection Engineering, 2, 37-47.
- Rak J. (2008). *Methods of reliability index determination concerning municipal water quality*. Journal of KONBiN, 5(2), 157 – 173.
- Rak J., Tchórzewska-Cieślak B. (2013). *Ryzyko w eksploatacji systemów zbiorowego zaopatrzenia w wodę*. Wydawnictwo Scidel Przywecki, Warszawa 2013.
- Raduly B., Gemaey K.V., Capodaglio A.G., Mikkelsen P.S., Henze M. (2007): *Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study*. Environmental Modelling & Software, 22, 1208-1216.
- Rene E.R., Saidutta M.B. (2008): *Prediction of Water Quality Indices by Regression Analysis and Artificial Neural Networks*. International Journal of Environmental Research, 2, 183-188.
- Rezania M., Javadi A. (2007): *A new genetic programming model for predicting settlement of shallow foundations*. Canadian Geotechnical Journal, 44, 1462-1473.
- Ribeiro D., Sanfins A., Belo O. (2013): *Wastewater treatment plant performance prediction with support vector machines*. In: *Industrial Conference on Data Mining*. In: P. Perner (Editor), *Proceedings of the 13th international conference on Advance in Data mining: applications and theoretical aspects*, New York 2013, 99–111.
- Roche N., Vaxelaire J., Prost C. (1995): *A simple empirical model for hindered settling in activated sludge clarifier*. Water Environment Research, 67, 775-780.
- Rodríguez J.P., McIntyre N., Díaz-Granados M., Achleitner S., Hochedlinger M., Maksimović C. (2013): *Generating time-series of dry weather loads to sewers*, Environmental Modelling & Software, 43, 133-143.
- Rogers L.L., Dowla F.U. (1994): *Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling*. Water Resources Research, 30, 457–481.
- Romanowicz R.J., Kiczko A., Napiórkowski J.J. (2010): *Stochastic transfer function model applied to combined reservoir management and flow routing*. Hydrological Sciences Journal, 55, 27–40.
- Romanowicz R.J., Osuch M. (2015): *Stochastic Flood Forecasting System: The Middle River Vistula Case Study*. Springer Verlag, New York, 2015.
- Rosen Ch., Yuan Z. (2001): *Supervisory control of wastewater treatment plants by combining principal component analysis and fuzzy c-means clustering*. Water Science and Technology, 43, 147-156.

- Rosenblatt F. (1958): *The perceptron: A probabilistic model for information storage and optimization in the brain*. Psychological Review, 65, 386 - 408.
- Rousseau D., Verdancq F., Moerman D., Carrette R., Thoeye C., Meirlaen J., Venrolleghem P.A. (2001): *Development of a risk assessment based technique for design/retrofitting WWTP*. Water Science & Technology, 43, 287–294.
- Rössle W.H., Pretorius W.A. (2008): *Batch and automated SVI measurements based on short-term temperature variations*. Water SA, 34, 237–243.
- Rustum R., Adeloye A., Simala A. (2007): *Kohonen self-organising map (KSOM) extracted features for enhancing MLP-ANN prediction models of BOD₅*. Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century (Proceedings of Symposium HS2005 at IUGG2007, Perugia, July 2007), IAHS Publ, 314.
- Rustum R. (2009): *Modelling Activated Sludge Wastewater Treatment Plants Using Artificial Intelligence Techniques (Fuzzy Logic and Neural Networks)*. PhD Thesis, Heriot, 2009.
- Salpasaranis K., Stylianakis V. (2017): *Forecasting the OECD fixed broadband penetration with genetic programming method, diffusion models and macro-economic indicators*. Image Processing & Communications, 21, 47-56.
- Sant´ın I. (2015): *Effluent predictions in wastewater treatment plants for the control strategies selection*. Actas de las XXXVI Jornadas de Automática, 2 - 4 de septiembre de 2015, Bilbao, 1009–1016.
- Sarkar M., Leong T.Y. (2000): *Application of K-nearest neighbors algorithm on breast cancer diagnosis problem*. AMIA Annual Symposium Proceedings of Archive, 759–763.
- Sathyamoorthy S., Chandran K., Ramsburg C.A. (2013): *Biodegradation and cometabolic modeling of selected beta blockers during ammonia oxidation*. Environmental Science & Technology, 47, 12835–12843.
- Sánchez M., Cortés U., Béjar J., Gracia J.De., Lafuente J., Poch M. (1997): *Concept Formation in WWTP by Means of Classification Techniques: A Compared Study*. Applied Intelligence, 7, 147–165.
- Schmidt L., Sliemers O., Schmid M., Bock E., Gijs Kuenen J., Jetten M.S.M., Strous M. (2003): *New concepts of microbiological treatment processes for nitrogen removal in wastewater*. Microbiology Review, 772, 1 – 12.
- Seggelke K., Rosenwinkel K.H., Vanrolleghem P.A., Krebs P. (2005): *Integrated operation of sewer system and WWTP by simulation-based control of the WWTP inflow*. Water Science and Technology, 5, 195–203.
- Setti S.G., Rao R.N. (2014): *Artificial neural network approach for prediction of stress-strain curve of near b titanium alloy*. Rare Metals, 33, 249–257.
- Sezgin M., Jenkins D., Parker D.S. (1978): *A unified theory of filamentous activated sludge bulking*. Journal (Water Pollution Control Federation), 50, 362–381.

- Shahin M. (2015): *Genetic programming for modelling of geotechnical engineering systems*. In: Handbook of Genetic Programming Applications, A, H, Gandomi, A, H, Alavi and C, Ryan (eds.), 37-57, Springer Verlag Heidelberg.
- Shoumy N.J., Yaakob S.N., Ehkan P., Ali M.S., Khatun S. (2016): *Cascade-forward neural network performance study for bloodstain image analysis*. 3rd International Conference on Electronic Design (ICED).
- Singh N.K., Kazmi A.A. (2018): *Performance and Cost Analysis of Decentralized Wastewater Treatment Plants in Northern India: Case Study*. Journal of Water Resources Planning and Management, 144, 1–12.
- Siwek K., Osowski S. (2016): *Data mining methods for prediction of air pollution*. International Journal of Applied Mathematics and Computer Science, 26, 467–478.
- Snip L.J.P., Alsina X.F., Gy Plósz B., Jeppsson U., Gernaey K.V. (2014): *Modelling the occurrence, transport and fate of pharmaceuticals in wastewater systems*. Environmental Modelling & Software, 62, 1-16.
- Sobota J., Szetela R. (2005): *Jednowymiarowe modele natężenia i jakości ścieków dopływających do oczyszczalni komunalnych*. Ochrona Środowiska, 27, 15-22.
- Sobota J., Szetela R. (2006): *Tworzenie szeregów czasowych natężenia dopływu i parametrów jakości ścieków na potrzeby symulacji pracy oczyszczalni komunalnych*. Ochrona Środowiska, 28, 39–46.
- Studziński J. (2004): *Identyfikacja, symulacja i sterowanie oczyszczalniami ścieków*. Badania Systemowe, Warszawa 2004, Monografia Instytutu Badań Systemowych PAN.
- Sweetapple C.G. (2014): *Developing Strategies for the Reduction of Greenhouse Gas Emissions from Wastewater Treatment*. Ph, D, Thesis, Exeter University.
- Szeląg B., Barbusiński K., Studziński J. (2018a): *Activated sludge process modelling using selected machine learning techniques*. Desalination and Water Treatment, 1–10.
- Szeląg B., Barbusiński K., Studziński J., Bartkiewicz L. (2017c): *Prediction of wastewater quality indicators at the inflow to the wastewater treatment plant using data mining methods*. E3S Web of Conferences 22, 00174 (2017).
- Szeląg B., Bartkiewicz L., Studziński J., Barbusiński K. (2017): *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, 43, 74–81.
- Szeląg B., Gawdzik J. (2017a): *Assessment of the effect of wastewater quantity and quality, and sludge parameters on predictive abilities of non-linear models for activated sludge settleability predictions*. Polish Journal of Environmental Studies, 26, 315–322.
- Szeląg B., Gawdzik J. (2017b): *Application of Selected Methods of Artificial Intelligence to Activated Sludge Settleability Predictions*. Polish Journal of Environmental Studies, 25, 1709-1714.

- Szeląg B., Gawdzik J., Studziński J. (2018b). *Sludge Volume Index (SVI) Modelling: Data Mining Approach*. Z. Wilimowska et al. (eds.), Information Systems Architecture and Technology: Proceedings of 38th International Conference on Information Systems Architecture and Technology – ISAT 2017, Advances in Intelligent Systems and Computing 657, 324 – 335.
- Szeląg B., Kiczko A., Studziński J., Dąbek L. (2018c): *Hydrodynamic and probabilistic modelling of storm overflow discharges*. Journal of Hydroinformatics, Vol. 20, No. 5, 1100 – 1110.
- Szeląg B., Kiczko A., Dąbek L. (2016): *Analiza wrażliwości i niepewności modelu hydrodynamicznego (SWMM) do prognozowania odpływu wód opadowych ze zlewni zurbanizowanej studium przypadku*. Ochrona Środowiska, 38, 15–22.
- Szeląg B., Siwicki P. (2017): *Application of the selected classification models to the analysis of the settling capacity of the activated sludge – case study*. In: B. Kaźmierczak, M. Kutylowska, K. Piekarska, A. Trusz-Zdybek, E3S Web of Conferences 17, Boduszów-Gorce 2017, 1–7.
- Szeląg B., Studziński J. (2017): *A data mining approach to the prediction of food-to-mass ratio and mixed liquor suspended solids*. Polish Journal of Environmental Studies, 26, 2231-2238.
- Szeląg B., Studziński J. (2018). *Modelling and Forecasting the Sludge Bulking in Biological Reactors of Wastewater Treatment Plants by Means of Data Mining Methods*. 296 – 305, Proceedings of the First International Conference on Intelligent Systems in Production Engineering and Maintenance ISPEM 2017, 296 – 305.
- Szeląg B., Studziński J., Barbusiński K. (2018d). *Application of the model of sludge volume index forecasting to assess reliability and improvement of wastewater treatment plant operating conditions*. Desalination and Water Treatment (w druku).
- Szetela R., Dymaczeński Z. (2002): *Modyfikacja obecnej postaci modelu osadu czynnego ASM2d*. Ochrona Środowiska, 1, 3–8.
- Taheriyoun M., Moradinejad S. (2015): *Reliability analysis of a wastewater treatment plant using fault tree analysis and Monte Carlo simulation*. Environmental Monitoring and Assessment, 187, 1-13.
- Thiery F., Grieu S., Traore A., Esteben M., Polit M. (2005): *Neural networks for estimating the efficiency of a WWTP biologic treatment*. In: Software Engineering: Evolution and Emerging Technologies.(red. K. Zieliński, T. Szmuc), IOS Press, Amsterdam, 25–33.
- Thunberg A., Sundin A.M., Carlsson B. (2009): *Energy optimization of the aeration process at Käppala wastewater treatment plant*. 10th IWA Conference on Instrumentation, Control & Automation.
- Tian Z.X., Jiang J.P., Guo L., Wang P. (2012): *Anomaly detection of Municipal Wastewater Treatment Plant operation using Support Vector Machine*. International Conference on Automatic Control and Artificial Intelligence (ACAI 2012).

- Tomczak E., Kaminski W. (2012): *Application of ANN to the Sorption Equilibrium Modelling of Heavy Metal Ions on Clinoptilolite*. Ecological Chemistry and Engineering S, 19, 227-237.
- Tran D.H., Ng A.W.M., Perera B.J.C., Burn S., Davis P. (2007): *Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes*. Urban Water Journal 3, 175-184.
- Tomita R.K., Sotomayor O.A.Z., Park S.W., Tisza-Contreras J.F. (2007): *Multiobjective optimization of an upflow anaerobic sludge blanket reactor*. Revista Peruana de Química e Ingeniería Química, 10, 39-49.
- Traczewska T. (1997): *Biotyczne i abiotyczne uwarunkowania pęcznienia osadu czynnego*. Ochrona środowiska, 2, 29-32.
- Tümer A.E., Edebalı S. (2015): *An Artificial Neural Network Model for Wastewater Treatment Plant of Konya*. Intelligent Systems and Applications in Engineering, 3, 131-135.
- Van Hulle S.W.H. (2005): *Modelling, simulation and optimization of autotrophic nitrogen removal processes*. PhD Thesis, University of Gent.
- Vapnik V. (1998): *Statistical Learning Theory*. John Wiley and Sons, New York, 1998.
- Venkadesh S., Hoogenboom G, Potter W., McClendon R. (2013): *A genetic algorithm to refine input data selection for air temperature prediction using artificial neural networks*. Applied Soft Computing, 13, 2253-2260.
- Verma A., Wei X., Kusiak A. (2012): *Predicting the total suspended solids in wastewater: A data-mining approach*. Engineering Applications of Artificial Intelligence, 26, 1366-1372.
- Vojinovic Z., Kecman V., Babovic V. (2003): *Hybrid approach for modeling wet weather response in wastewater systems*. Journal of Water Resources Planning and Management, 129, 511-521.
- Vrecko D., Gernaey K.V., Rosen C., Jeppsson U. (2006): *Benchmark Simulation Model No 2 in Matlab-Simulink: towards plant-wide WWTP control strategy evaluation*. Water Science and Technology, 54, 65-72.
- Wan T.J., Chen Ch.H., Chang Y.S., Chen J.C. (2000): *An application of artificial neuromolecular system for effluent quality prediction of wastewater treatment plant*. Journal of Chinese Institute of Environmental Engineering, 10, 155-162.
- Wei X., Kusiak A. (2012): *Optimization of Biogas Production Process in a Wastewater Treatment Plant*. Proceedings of the 2012 Industrial and Systems Engineering Research Conference, 1-9.
- Wei X., Kusiak A., Sadat H.R. (2013): *Prediction of Influent Flow Rate: Data-Mining Approach*. Journal of Energy Engineering, 139, 118-123
- Weiß, G., Brombach, H., Haller, B. (2002). *Infiltration and inflow in combined sewer systems: longterm analysis*. Water Science & Technology, Vol. 45, p. 227-230.

- Wett B., Podmirseg S.M., Gomez-Brand M., Hell M., Nyhuis G., Bott C., Murthy S. (2015): *Expanding DEMON Sidestream Deammonification Technology Towards Mainstream Application*. Water Environment Research, 87, 2084–2089.
- Wintgens T., Rosen J., Melin T., Brepols C., Drensla K., Engelhardt N. (2003): *Modeling of a membrane bioreactor system for municipal wastewater treatment*. Journal of Membrane Science, 216, 55-6.
- Woo S.H., Jeon C.O., Yun Y.S., Choi H., Lee C.S., Lee D.S. (2009): *On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant*. Journal of Hazardous Materials, 161, 538–544.
- Xie W.M., Zhang R., Li W.W., Ni B.J., Fang F., Sheng G.P., Yu H.Q., Song J., Le D.Z., Bi X.J., Liu Ch.Q., Yang M. (2011): *Simulation and optimization of a full-scale Carrousel oxidation ditch plant for municipal wastewater treatment*. Biochemical Engineering Journal, 56, 9-16
- Yang J.H., Cheng Ch.H., Chan Ch.P. (2017): *A Time-Series Water Level Forecasting Model Based on Imputation and Variable Selection Method*. Computational Intelligence and Neuroscience, 2017, Article ID 8734214, 1-11.
- Yasmin N.S.A., Gaya M.S., Wahab N.A., Sam Y.M. (2017): *Estimation of pH and MLSS using neural network*. Telkomnika, 15, 912–918.
- Yetilmmezsoy K., Ozkaya B., Cakmakci M. (2011): *Artificial intelligence-based prediction models for environmental engineering*. Neural Network World, 3, 193-218.
- Yoo C.K., Bang Y.H., Lee B., Vanrolleghem P.A., Rosén Ch. (2004): *Application of Fuzzy Partial Least Squares (FPLS) Modeling Nonlinear Biological Processes*. Korean Journal of Chemical Engineering, 21, 1087-1097.
- Zhang R., Xie W.M., Yu H.Q., Li W.W. (2014): *Optimizing municipal wastewater treatment plants using an improved multi-objective optimization method*. Bioresource Technology, 157, 161–165.
- Zhang M., Jing H., Liu Y., Shi H. (2017): *Estimation and optimization operation in dealing with inflow and infiltration of a hybrid sewerage system in limited infrastructure facility data*. Frontiers of Environmental Science & Engineering, 11, 1–11.
- Zhao Y., Guo L., Liang J., Zhang M. (2014): *Seasonal artificial neural network model for water quality prediction via a clustering analysis method in a wastewater treatment plant of China*. Desalination and Water Treatment, 57, 3452–3465.
- Zhao G., Liu J., Jiang J., Wang W. (2017): *A deep cascade of neural networks for image inpainting, deblurring and denoising*. Multimedia Tools and Application, 1–16.
- Zhu J.J., Kang L., Anderson P.R. (2018): *Predicting influent biochemical oxygen demand: Balancing energy demand and risk management*. Water Research, 128, 304-313.

ISSN 0208-8029
ISBN 978-83-89475-62-6

INSTYTUT BADAŃ SYSTEMOWYCH
POLSKIEJ AKADEMII NAUK
tel.: (+48) 22 3810246 / 22 3810277 / 22 3810241 / 22 3810273
e-mail: biblioteka@ibspan.waw.pl