



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

BADANIA SYSTEMOWE

Tom 76

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

Bartosz Szeląg

Warszawa 2019

MODELOWANIE MATEMATYCZNE, OPTYMALIZACJA I STEROWANIE PRACĄ
PRZEPLYWOWYCH OCZYSZCZALNI ŚCIEKÓW, Bartosz Szeląg



**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 prowadziłyby 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.

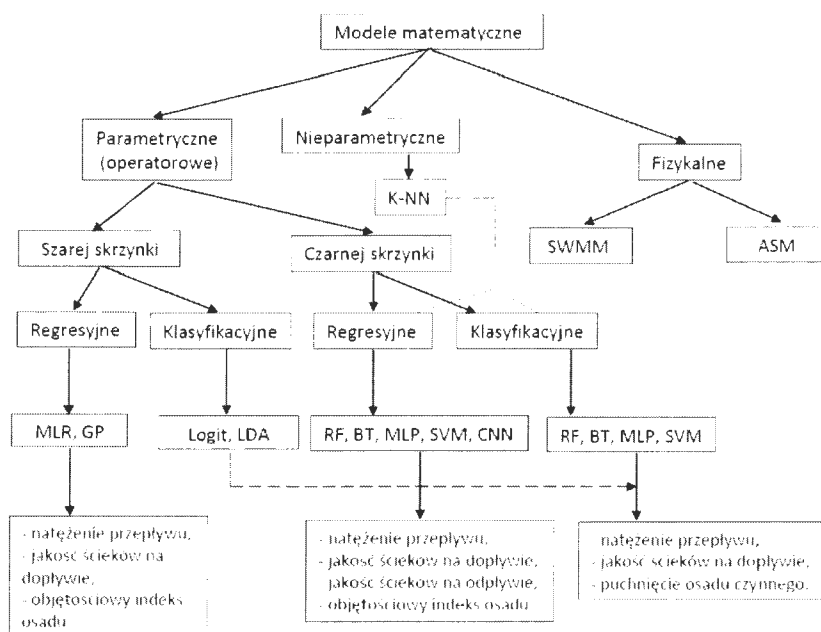
W rozdziale 8 opracowano także model probabilistyczny do oceny niezawodności działania osadnika wtórnego przy wykorzystaniu generatorów liczb losowych (Monte Carlo), w którym ujęto możliwość wystąpienia awarii. Należy podkreślić, że dotychczas potencjalne awarie nie były ujmowane w modelach probabilistycznych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

7. Dobór metod data mining do symulacji oczyszczalni ścieków

W ramach prowadzonych badań wykonano symulacje ilości i jakości ścieków na dopływie i odpływie z oczyszczalni oraz sedimentacji osadu czynnego w oparciu o zaproponowany algorytm doboru metody (rys. 6.6). Do tego celu zastosowano modele parametryczne i nieparametryczne (rys. 7.1). W obrębie modeli parametrycznych (operatorowych) rozważono wykorzystanie modeli szarej i czarnej skrzynki, które zaimplementowano w zagadnieniach do prognozy wartości liczbowych (modele regresyjne) i lingwistycznych (klasyfikacyjnych). Poza modelami parametrycznymi rozważono także implementację modeli *nieparametrycznych*, do których można zaliczyć metodę *k-najbliższych sąsiadów*. *W pracy pominięto* modele fizykalne do prognozy ilości, jakości ścieków na dopływie tworzone na bazie programu SWMM (Stormwater Management Model) i odpływie do oczyszczalni w oparciu o modele ASM.

Z uwagi na złożoność procesów, jakie zachodzą na oczyszczalniach ścieków, przed przystąpieniem do symulacji rozważono w pierwszej kolejności badanie zależności (liniowa, nieliniowa) między zmienną zależną i zmiennymi objaśniającymi. Gdy analizowane zależności $y = f(x_i)$ miały charakter liniowy lub mieszany, przewidziano zastosowanie metody wielorakiej regresji liniowej (MLR). W przeciwnym przypadku (nieliniowe zależności i złożone procesy w bioreaktorze) wykorzystywano metodę programowania genetycznego (GP), bowiem uzyskane w niej równanie modelu, podobnie jak w metodzie MLR, ma charakter empirycznej zależności i możliwa jest fizykalna interpretacja estymowanych parametrów. Gdy wyniki obliczeń nie spełniały zakładanych wymagań dotyczących dokładności modelowania, zaplanowano wykorzystanie metod *k-najbliższych sąsiadów* (*k-NN*) oraz lasów losowych (RF), drzew wzmacnianych (BT), w których otrzymana struktura modeli ma interpretację fizyczną i jest możliwa ocena wpływu wybranych zmiennych objaśniających na wyniki obliczeń.



Rys. 7.1. Podział metod zastosowanych do modelowania poszczególnych procesów na oczyszczalni ścieków.

W kolejnym etapie badań stosowano modele o bardziej złożonej strukturze, w których wartości parametrów w uzyskiwanej strukturze modelu nie mają już interpretacji fizycznej. Są to modele ANN typu MLP, CNN oraz SVM. W modelach sieci neuronowych badano wpływ złożoności struktury modelu na wyniki uzyskiwanych symulacji, dlatego też w pierwszej kolejności ujęto metodę MLP, a następnie wykorzystano metodę CNN (w celu oceny wpływu dodatkowych połączeń w sieci na wyniki symulacji) i metodę SVM.

W pracy przeanalizowano najpierw szczegółowo zastosowanie wymienionych wyżej metod do prognozy ilości i jakości ścieków na dopływie do oczyszczalni ścieków. Wykonane analizy miały na celu ocenę wpływu złożoności opracowanego modelu na wyniki symulacji. Aspekt ten z praktycznego punktu widzenia jest bardzo istotny. Mianowicie należy dążyć do tego, aby implementowane modele były z jednej strony jak najmniej skomplikowane, a z drugiej strony aby otrzymane przy ich pomocy wyniki obliczeń były bliskie wartościom pomierzonym. Wyznaczone w wyniku modelowania wskaźniki jakości ścieków na dopływie były zmiennymi niezależnymi w wyznaczanych następnie modelach procesowych (jakości ścieków na wylocie i sedimentacji osadu czynnego). W związku z tym zachodziła potrzeba poszukiwania metod, dla których błędy prognozy były możliwie najmniejsze, ponieważ zmienność jakości ścieków

na wlocie do obiektu stanowi podstawę doboru nastaw w reaktorach i optymalizacji ich pracy.

Po wyznaczeniu wymienionych modeli do prognozy jakości ścieków na dopływie i odpływie z oczyszczalni oraz sedymentacji osadu czynnego, przystąpiono następnie do wyznaczania modeli hybrydowych, w których wartości pewnych trudno mierzalnych zmiennych niezależnych zastąpiono wynikami obliczeń tych zmiennych uzyskanymi za pomocą modeli statystycznych. Do dalszych rozważań wybrano modele hybrydowe do prognozy jakości ścieków na wylocie i sedymentacji osadu, dla których otrzymano najmniejsze wartości błędów. Modele te stanowiły następnie podstawę stworzenia modelu matematycznego bioreaktora, służącego do optymalizacji pracy oczyszczalni ścieków.

7.1. Ogólne zasady tworzenia modeli do prognozy ilości, jakości ścieków na dopływie, odpływie i sedymentacji osadu czynnego

W celu uogólnienia zasad tworzenia modeli do prognozy ilości i jakości ścieków na dopływie i odpływie z oczyszczalni oraz sedymentacji osadu czynnego zdefiniowano następujące ogólne założenia:

- w metodzie MLR stosowano regresję krokową pozwalającą na eliminowanie zmiennych nieistotnych statystycznie na przyjętym poziomie ufności ($\alpha = 0,05$),
- w metodzie programowania genetycznego (GP) przyjmowano zmienne niezależne ustalone na podstawie przeprowadzonych analiz statystycznych (test Fischera–Snedecora, chi–kwadrat), a następnie uwzględniano podstawowe operatory genetyczne typu +, -, /, *, ^ . Dla założonej początkowej liczby osobników i pokoleń nie większej, niż 500, dla przyjętego prawdopodobieństwa mutacji 0,25 i krzyżowania 0,65 poszukiwano zależności empirycznych,
- w metodach RF i BT, aby otrzymany model nie był przeuczony, przyjęto maksymalną liczbę drzew w modelach nie większą, niż 300 (Kusiak i in. 2013),
- w metodzie k-NN liczbę sąsiadów w każdym modelu dobierano metodą kolejnych podstawień wartości k, dopóki wartości miar dopasowania modeli nie były najmniejsze
- w metodzie MLP analizowano liczbę neuronów w warstwie ukrytej w zakresie $j \div (2 * j + 1)$ dla różnych funkcji aktywacji (liniowa, logistyczna, eksponentialna, tangens hiperboliczny), uznając za optymalny model, dla którego uzyskane miary dopasowania były najmniejsze,

- w metodzie CNN, bazując na danych literaturowych (Setti i Rao 2014, Capizzi i in. 2015, Szelaąg i in. 2018a), przyjęto dwa dodatkowe połączenia między sąsiednimi warstwami i identyczną liczbę neuronów w kolejnych warstwach ukrytych. Ponadto, z uwagi na istotny wpływ wartości startowych wag w modelach ANN na wyniki symulacji i problemy z ich optymalizacją, każdy model sieci neuronowej (ustalona liczba neuronów i funkcja aktywacji) był wielokrotnie generowany (1000–krotnie) dla różnych warunków brzegowych,
- w metodzie SVM, bazując na danych literaturowych (Borges i in. 2000, Ribiero i in. 2013, Szelaąg i Gawdzik 2017a, b), przyjęto do badań gaussowską funkcję jądrową a wartości C i ε poszukiwano metodą kolejnych podstawień, aż do momentu uzyskania minimalnych wartości współczynników opisanych równaniami (6.13 ÷ 6.15).

W prowadzonych badaniach dane do tworzenia modeli statystycznych podzielono na trzy zbiory: uczący (70%), testowy (15%) i walidacyjny (15%). Liczebność poszczególnych zbiorów danych wykorzystanych do wyznaczania modeli zamieszczono w tab. 7.1.1. Ponadto, wykonane analizy statystyczne przy pomocy testu Mann – Kendalla na poziomie istotności statystycznej $p = 0,05$ wykazały brak trendu, co wskazuje, że rozpatrywane szeregi są stacjonarne (Szelaąg i in. 2018a).

Tab. 7.1.1. Zestawienie wielkości poszczególnych zbiorów danych do tworzenia poszczególnych modeli statystycznych.

Zmienna zależna	Liczebność zbiorów danych
Natężenie przepływu (Q)	968
$(BZT_5, ChZT, Z_{og}, N_{og}, P_{og}, N-NH_4)_{dop}$	300
$(BZT_5, ChZT, Z_{og}, N_{og}, P_{og}, N-NH_4)_{odp}$	150
Objętościowy indeks osadu (I_o)	250

Modele matematyczne opracowane w oparciu o metody k-NN, RF, BT, SVM wyznaczano przy wykorzystaniu programu STATISTICA, z kolei modele opracowane w oparciu o metody CNN, MLP wyznaczono stosując program MATLAB.

7.2. Prognoza natężenia dopływu do oczyszczalni ścieków

Eksploatacja oczyszczalni ścieków jest złożonym zadaniem sterowania. Wartości nastaw w bioreaktorze zależą od warunków występujących na dopływie do oczyszczalni, charakteryzowanych przez jakość ścieków i ilość ścieków surowych. Drugi z tych czynników jest szczególnie istotny, bowiem determinuje wielkość ładunku dopływających zanieczyszczeń i obciążenie komór osadu czynnego, co

wpływa na działanie reaktora biologicznego (wiek osadu, obciążenie substratowe komór osadu czynnego, czas zatrzymania ścieków) i jakość ścieków na odpływie z obiektu. Gwałtowny wzrost ilości ścieków trafiających do oczyszczalni często prowadzi do zaburzenia procesu usuwania zanieczyszczeń stąd przewidywanie wielkości strumienia dopływających ścieków wpływa pozytywnie na eksploatację obiektu.

Obecnie wielkość dopływu do oczyszczalni ścieków prognozuje się przy pomocy modeli fizykalnych. Jednak modele te mają wady, bowiem zebranie odpowiednich informacji o zlewni jest trudne, a koszty prowadzenia ciągłych pomiarów ilości i jakości ścieków oraz opadów są wysokie. Ponadto, mimo zgromadzenia odpowiedniego zbioru danych uzyskiwane wyniki prognozowania dopływu nie zawsze są zadowalające (Szelaąg i in. 2016, Kiczko i in. 2018). Dlatego też, do symulacji ilości ścieków dopływających do oczyszczalni coraz częściej stosuje się modele parametryczne (czarnej skrzynki), przy czym często wykorzystuje się w tym celu metodę MLP (El-Din i Smith 2002, Fernandez i in. 2009, Jurasz i Pilecki 2013, Wei i in. 2013, Szelaąg i in. 2017). Metoda ta jest obecnie powszechnie stosowana, co wynika między innymi z faktu, iż jest ona zaimplementowana w większości dostępnych pakietów statystycznych. Mimo licznych zastosowań metody MLP i również pozostałych metod data mining (rozdział 6), autorzy wielu prac (El-Din i Smith 2002, Sobota i Szetela 2006, Fernandez i in. 2009, Ma i in. 2013) z reguły koncentrują się na używaniu jednej metody. Otrzymywane wyniki mają więc charakter jednostkowy, ponieważ nie pozwalają stwierdzić, czy dla analizowanego zbioru danych nie można było zaimplementować mniej złożonych modeli o prostszej strukturze, wymagających mniejszych mocy obliczeniowych, krótszego czasu obliczeń itp.

W związku z powyższym, w ramach prezentowanych badań do symulacji dopływu ścieków surowych rozważono szereg modeli statystycznych (MLR, GP, RF, BT, k-NN, MLP, SVM, CNN), próbując równocześnie ustalić optymalny zbiór zmiennych niezależnych, stanowiących kompromis między ich liczbą a dokładnością modelu. Prezentowane podejście jest istotne z uwagi na to, że pozwala precyzyjnie ocenić wpływ struktury badanego modelu statystycznego na wyniki obliczeń. W prowadzonych badaniach do oceny wpływu poszczególnych zmiennych niezależnych na wyniki modelowania dopływu skoncentrowano się na metodzie sztucznych sieci neuronowych, wykorzystując w obliczeniach dane pomiarowe z oczyszczalni ścieków w Sitkówce-Nowiny z okresu 2013–2015. Pomiary dotyczące opadów deszczów pozyskano z posterunku opadowego Politechniki Świętokrzyskiej zlokalizowanego na terenie miasta Kielce.

W badaniach podjęto próbę opracowania uniwersalnego modelu statystycznego prognozującego dopływ ścieków surowych. W tym celu rozważono zastosowanie metody regresji logistycznej, w której można ustalić w sposób jawny wpływ poszczególnych zmiennych niezależnych na dokładność predykcji dopływu

do oczyszczalni, oceniając tą dokładność za pomocą współczynnika korelacji. Aby otrzymać szeroką zmienność parametrów struktury modelu MLP zastosowano funkcję „Automatycznego kreatora sieci neuronowych” w programie STATISTICA. Na etapie tworzenia wspomnianego modelu uniwersalnego korzystano z pomiarów z oczyszczalni w Sitkówce–Nowiny, a do weryfikacji otrzymanych zależności wykorzystano dane z oczyszczalni ścieków w Rzeszowie.

7.2.1. Modelowanie dopływu ścieków do oczyszczalni ścieków w Sitkówce–Nowiny

Na podstawie zestawionych w tab. 3.2 prac z zakresu prognozy dopływu ścieków do oczyszczalni zebrano dane dotyczące dopływu mierzonego w różnych chwilach czasowych w sposób ciągły, całkowitej wysokości opadu deszczu, okresu bezdeszczowego (t_{bd}) oraz sezonowości odniesionej do pór roku. W oparciu o zgromadzone dane wyznaczono wartości współczynników korelacji między wartością przepływu Q a rozważanymi zmiennymi niezależnymi; w tab. 7.2.1. zestawiono wyniki wykonanych obliczeń.

Tab. 7.2.1. Wartości współczynników korelacji (R) między przepływem (Q) a zmiennymi niezależnymi; gdzie: 4 – jesień, 1 – zima, 2 – wiosna, 3 – lato, P – całkowita wysokość opadu deszczu, P , $Q(t - 1)$ – wartości pomierzonych wejść do modelu odpowiednio w i – dobach poprzedzających modelowaną wielkość w chwili t)

Zmienne	$Q(t - 1)$	$Q(t - 2)$	$Q(t - 3)$	$P(t - 1)$	$P(t - 2)$	t_{bd}	Pora roku			
							4	1	2	3
R	0,77	0,69	0,64	0,23	0,17	-0,26	-0,2	-0,07	0,34	-0,07
$p(\text{test F-S})$	0,00001	0,00003	0,00060	0,00100	0,01200	0,00240	0,0050	0,0040	0,0017	0,0050

Na podstawie otrzymanych wyników stwierdzono, że zmienne niezależne, które przyjęto do analiz statystycznych, określono poprawnie, na co wskazują otrzymane wartości $p < 0,05$. W przypadku $Q(t - 1)$ otrzymano istotną wartość $R = 0,77$. Dla zmiennych takich, jak P , t_{bd} i pory roku, stwierdzono zależności silnie nieliniowe z przepływem $Q(t)$, co również potwierdzają obliczone wartości współczynnika korelacji.

Wartość R wyznaczona dla pary zmiennych Q i (t_{bd}) wskazuje, że objętość wody zgromadzona w zlewni (mierzona od poprzedniego zdarzenia opadowego) ma wpływ na ilość ścieków dopływających do oczyszczalni w chwili t . Oznacza to, że im dłuższy jest okres od ostatniego opadu, tym ilość wody w zlewni zmniejsza się na skutek parowania, a co za tym idzie dopływ do oczyszczalni ścieków maleje. Równocześnie, otrzymane wartości współczynnika korelacji między Q

a poszczególnymi porami roku potwierdzają rolę procesu parowania w bilansie wodnym zlewni i wpływu dynamiki zmian położenia zwierciadła wód gruntowych w okresie roku na zróżnicowany dopływ ścieków do oczyszczalni (Weiss i in. 2002, Pawlowski i in. 2014).

W oparciu o zgromadzone dane i powyższe wyniki analizy, stosując metodę wielorakiej regresji liniowej, wyznaczono model postaci:

$$Q(t) = 0,955(\pm 0,003) \cdot Q(t-1) + 845,29(\pm 31,11) \cdot P(t-1) \quad (7.2.1)$$

dla, którego, przy pomocy programu STATISTICA, ustalono wartości $R=0,74$, $MAE=3073 \text{ m}^3/\text{d}$, $RMSE=5545 \text{ m}^3/\text{d}$ i $MAPE=8,15 \%$. Analizując zależność (7.2.1) można stwierdzić, że opisuje ona z zadowalającą dokładnością wartości dobowego dopływu ścieków do oczyszczalni. Jednak różnica między wartościami MAE i RMSE może wskazywać, że wielkości maksymalne są zaniżone (rys. 7.2.1).

Z punktu widzenia eksploatacji obiektu i doboru nastaw w oparciu o wyniki symulacji, niedoszacowanie dopływu może prowadzić do problemów z optymalnym doбором obciążenia osadu czynnego ładunkiem zanieczyszczeń, utrzymaniem jakości ścieków na wylocie z oczyszczalni itp. W związku z powyższym, zachodzi potrzeba poprawy zdolności predykcyjnych podanego wyżej modelu statystycznego.

W kolejnym etapie badań, stosując metodę programowania genetycznego i bazując na założeniach przyjętych w rozdziale 7.1, opracowano model postaci:

$$Q(t) = \frac{0,99 \cdot Q(t-1) + 385,15 \cdot P(t-1)}{1 - 0,00141 \cdot t_{bd}} \quad (7.2.2)$$

dla którego wyznaczono przy pomocy autorskiego programu w języku MATLAB wartości $R=0,73$, $MAE=3.050 \text{ m}^3/\text{d}$, $RMSE=5.490 \text{ m}^3/\text{d}$ i $MAPE=7,03\%$.

Mimo tego, że wyznaczona zależność charakteryzuje się zbliżonymi zdolnościami prognozy dopływu, co poprzednia, to jednak otrzymany z niej związek $Q(t) = f(t_{bd})$ nie ma merytorycznego uzasadnienia, stąd też w dalszych analizach została ona pominięta.

W następnej kolejności wykonano obliczenia dopływu do oczyszczalni ścieków w Sitkówce-Nowiny metodami k-NN, RF, BT, MLP, SVM i CNN. Zestawienie przykładowych kombinacji zmiennych niezależnych i uzyskanych wartości współczynników dopasowania wyników symulacji do pomiarów w metodzie MLP zamieszczono w tab. 7.2.3. dla zbioru uczącego i testowego, a w tab. 7.2.2. podano optymalne wartości parametrów struktury modeli MLP ustalone zgodnie z zasadami podanymi w rozdziale 7. Z kolei w tab. 7.2.4. zamieszczono wartości miar dopasowania pomiarów do obliczeń dla zbioru testowego i walidacyjnego, uzyskane metodami RF, BT, CNN, SVM.

Tab. 7.2.2. Optymalne parametry struktury (MAE, R, RMSE, MAPE) modeli MLP do symulacji dopływu do oczyszczalni w Sitkówce–Nowiny i Rzeszowie.

Zmienne	Oczyszczalnia w Sitkówce - Nowiny			Oczyszczalnia w Rzeszowie		
	liczba neuronów	funkcja aktywacji		liczba neuronów	funkcja aktywacji	
	ukrytych	wejście	wyjście	ukrytych	wejście	wyjście
Q(t-1)	3	lin	lin	5	log	tanh
Q(t-1), pora roku	3	tanh	lin	3	sin	lin
Q(t-1), t_{bd}	4	lin	lin	3	log	lin
Q(t-1), Q(t-2)	5	tanh	sin	6	wykl	lin
Q(t-1), Q(t-2), Q(t-3)	8	wykl	lin	4	tanh	wykl
P(t-1)	6	tanh	log	5	tanh	lin
P(t-1), P(t-2)	5	wykl	sin	5	tanh	sin
P(t-1), Q(t-1)	4	tanh	lin	4	log	tanh
P(t-1), Q(t-1), Q(t-2)	7	tanh	log		sin	log
P(t-1), Q(t-1), Q(t-2), P(t-2)	9	log	sin	8	tanh	lin
P(t-1), Q(t-1), Q(t-2)	11	tanh	lin	9	tanh	

Tab. 7.2.3. Wartości współczynników dopasowania wyników obliczeń (MAE, R, RMSE, MAPE) do pomiarów uzyskane metodą MLP dla oczyszczalni w Sitkówce–Nowiny.

Zmienna	Test				Walidacja			
	R	MAE, m ³ /d	RMS E, m ³ /d	MAP E, %	R	MAE, m ³ /d	RMS E, m ³ /d	MAP E, %
Q(t-1)	0,72	3550	5490	7,92	0,71	3470	5450	7,90
Q(t-1), pora roku	0,74	3250	5300	7,28	0,77	3257	5300	7,28
Q(t-1), t_{bd}	0,74	3270	5315	7,35	0,72	3290	5316	7,35
Q(t-1), Q(t-2)	0,72	3209	5260	7,34	0,71	3254	5259	7,31
Q(t-1), Q(t-2), Q(t-3)	0,70	3460	5440	7,86	0,70	3500	5490	7,92
P(t-1)	0,30	5389	7100	12,69	0,29	5421	7100	12,65
P(t-1), P(t-2)	0,35	5149	6800	12,33	0,37	5100	6796	12,3

Tab. 7.2.3. cd.

Zmienna	Test				Walidacja			
	R	MAE, m ³ /d	RMS E, m ³ /d	MAP E, %	R	MAE, m ³ /d	RMS E, m ³ /d	MAP E, %
P(t-1), Q(t-1)	0,75	3000	5020	6,91	0,76	2980	5000	6,91
P(t-1), Q(t-1), Q(t-2)	0,77	2800	4820	6,51	0,77	2780	4800	6,25
P(t-1), Q(t-1), Q(t-2), P(t-2)	0,78	2650	4650	6,00	0,78	2620	4620	6,03
P(t-1), Q(t-1), Q(t-2) Q(t-3)	0,75	3010	5030	6,91	0,75	3040	5020	6,90
P(t-1), Q(t-1), Q(t-2) Q(t-3), P(t-2)	0,76	2750	4810	6,51	0,77	2720	4857	6,50

Tab. 7.2.4. Zestawienie wartości wskaźników (R, RMSE, MAPE, MAE) dopasowania wyników obliczeń do pomiarów uzyskanych metodami RF, BT, SVM, CNN dla oczyszczalni w Sitkówce–Nowiny.

Metoda	Test				Walidacja			
	R	MAE, m ³ /d	RMSE, m ³ /d	MAPE, %	R	MAE, m ³ /d	RMSE, m ³ /d	MAPE, %
RF	0,76	3066	5301	7,17	0,74	3160	5390	7,41
BT	0,76	2962	5433	6,81	0,76	3020	5500	6,96
SVM	0,80	2726	4941	6,15	0,78	2780	5000	6,35
CNN	0,82	2650	4850	6,11	0,79	2730	4920	6,25
k-NN	0,63	3545	6509	7,95	0,57	3600	6700	8,25

Na podstawie danych w tab. 7.2.2. można stwierdzić, że w opracowanych modelach liczba neuronów w warstwie ukrytej nie przekraczała liczby $2 \cdot j + 1$, co wskazuje, że modele nie są przeuczone. Równocześnie stwierdzono, że w modelach tych w większości przypadków dla warstwy wejściowej sieci najlepsze okazały się funkcje tangens hiperboliczny, a dla warstwy wyjściowej funkcje liniowe. Fakt, iż funkcja tangens hiperboliczny jest dobrym rozwiązaniem w modelach MLP do prognozy procesów na oczyszczalni ścieków znajduje potwierdzenie w pracach Bartkiewicz i Studziński (2010), Lou i Zhao (2010), Szeląg i in. (2018d).

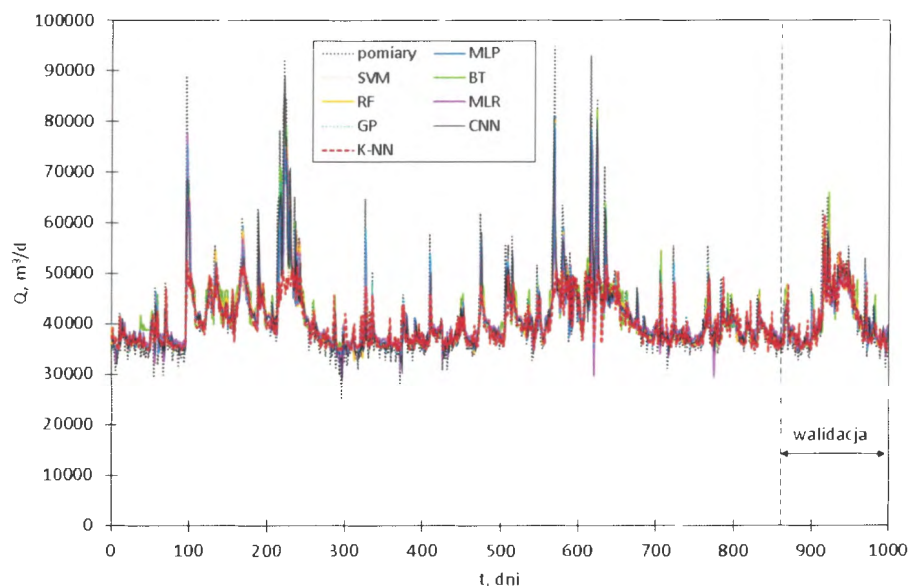
W przypadku metody CNN, bazując na wynikach otrzymanych metodą MLP ustalono, że zakres poszukiwanej liczby neuronów w modelu zmienia się od 4 do 10 (mniejsza liczba wynika z tego, że najmniejsze wartości błędów dopasowania uzyskano dla $j = 4$ zmiennych, a większa liczba neuronów wynika z zależności $2 \cdot j + 1$). W metodzie k-NN stwierdzono, że w przypadku liczby sąsiadów $k = 1 \div 15$ najmniejsze wartości MAE, RMSE, MAPE wyznaczono dla $k = 6$. Z kolei w metodzie SVM, podstawiając różne wartości C, ε i γ ustalono, że wartości błędów osiągają wartość minimalną dla $C = 25$ oraz $\varepsilon = 0,1$ i $\gamma = 0,35$.

Analizując dane w tab. 7.2.3 stwierdzono, że najmniejsze wartości błędów prognozy dopływu do oczyszczalni ścieków otrzymano na podstawie modelu, w którym uwzględniono wysokości opadów deszczu i dopływu z dwóch poprzednich dób, tj. $Q(t-1), Q(t-2), P(t-1), P(t-2)$. Zauważono przy tym, że nie jest możliwe prognozowanie dopływu w chwili t na podstawie całkowitej wysokości opadu deszczu $P(t-1)$ zmierzonej poprzedniego dnia. Mniejsze wartości błędów prognozy dopływu (o około 25%) otrzymano dla modelu MLP uwzględniającego wielkości dopływu $Q(t-1)$ zmierzone w poprzedniej dobie. Jednocześnie, stwierdzono wpływ wartości $Q(t-2)$ i $P(t-2)$ na poprawę średnio o 4,5% zdolności predykcyjnych modelu liczonego na podstawie $Q(t-1)$. W przypadku modelu MLP wykazano także nieznaczny wpływ pór roku (2,5%) i okresu bezdeszczowego (2%) na poprawę dokładności prognoz. W kolejnych etapach wyznaczania modeli MLP zmienne te, w kombinacji z innymi zmiennymi, okazały się pomijalne. Może to wynikać z faktu, że do tworzenia modelu wykorzystano zaledwie 3-letnie szeregi obserwacyjne. Wydłużenie długości szeregów pomiarowych (np. do 5 lat) może istotnie wpłynąć na poprawę uzyskanych wyników, pokazanych na rys. 7.2.1.

Istotnym aspektem, którego nie można pominąć w analizie danych (tab. 7.2.3, 7.2.4), jest zależność między wartościami MAE i RMSE. Znaczne różnice (dochodzące do 30÷60%) w ich wartościach mogą wskazywać, że w wyznaczonych modelach występują problemy z niedoszacowaniem wartości ekstremalnych; wniosek ten potwierdza zmienność krzywych uzyskanych z pomiarów i obliczeń (rys. 7.2.1).

Omawiając otrzymane wyniki obliczeń dopływu ścieków do oczyszczalni (rys. 7.2.1 i tab. 7.2.1÷7.2.4) nie można pozostawić bez komentarza faktu, że wartości błędów prognozy Q otrzymane przy pomocy metod RF, BT, CNN i SVM tylko nieznacznie różnią się od rezultatów symulacji uzyskanych metodą MLP. Wskazują na to również krzywe zobrazowane na rys. 7.2.1. Jedynie w przypadku metody k-NN wartości błędów prognozy dopływu (MAE) są większe o 24%, niż w metodzie MLP, o 15% większe, niż w metodzie RF, o 20% większe, niż w metodzie BT i o 35% większe, niż w metodzie CNN. W metodzie k-NN średnie wartości dopływu są modelowane z zadowalającą dokładnością, z kolei wartości maksymalne są niedoszacowane. Mając na uwadze sposób generowania prognoz

w metodzie k-NN można sądzić, iż otrzymany wynik symulacji może być wywołany niedostateczną zmiennością w danych pomiarowych lub zbyt małą wielkością zbioru danych.



Rys. 7.2.1. Porównanie wyników pomiarów i obliczeń dopływu ścieków różnymi metodami (MLR, GP, K-NN, RF, BT, MLP, SVM, CNN) dla oczyszczalni w Sitkówce – Nowiny.

Podsumowując uzyskane wyniki modelowania można stwierdzić, że chociaż przedstawiona metodyka nie jest nowatorska, to jednak podejście do prognozy natężenia przepływu jakie zaprezentowano w pracy pozwala na ocenę wpływu złożoności przyjętego modelu na wyniki symulacji, co do tej pory w tak szerokim zakresie w odniesieniu do metod uczenia maszynowego nie było rozpatrywane.

Przedstawione wyniki symulacji dopływu do oczyszczalni ścieków (tab. 7.2.3, tab. 7.2.4) są gorsze, niż uzyskane przez innych autorów. Znacznie lepsze dopasowanie wyników obliczeń do pomiarów ($R=0,98$) uzyskali Wei i in. (2013), którzy dla dużej zlewni w oparciu o ciągłe wyniki pomiarów opadów deszczu z sieci deszczomierzy opracowali model do prognozy $Q(t)$ z 15-minutowym krokiem czasowym. W obliczeniach tych, stosując różne metody czarnej skrzynki (RF, BT, SVM, MLP) stwierdzili, że wyniki modelowania wybranymi metodami tylko nieznacznie się różnią, a wyznaczone wartości współczynników korelacji wynoszą od 0,945 (RF) do 0,985÷0,988 (MLP, SVM). Większe wartości R wyznaczone przez Wei i in. (2013) mogą wynikać z tego, iż w tworzonym modelu uwzględnili oni poza wysokością opadu również kierunki przemieszczania się frontów opadowych deszczów, co wpływa na zmienność dopływu do oczyszczalni

ścieków. Mniejsze wartości błędów dopasowania (MAPE na poziomie 5%) uzyskali także Fernandez i in. (2009), którzy stosowali bardziej złożone modele do symulacji dopływu bazujące na sieciach neuronowych, stanowiących połączenie logiki rozmytej z klasycznym modelem MLP. Problem prognozy dopływu ścieków do oczyszczalni ścieków w Polsce był poruszany przez wielu autorów (tab. 3.2), których modele charakteryzowały się lepszymi zdolnościami prognozy, niż uzyskane w prezentowanych obliczeniach, co zostało szczegółowo omówione punkcie 3.1. Wyznaczone tutaj wartości błędów dopasowania wyników obliczeń do pomiarów były jednak nieznacznie lepsze, niż wyznaczone przez Sobotę i Szetelę (2006), którzy zmodyfikowali model autoregresyjny uwzględniając w nim wrażliwość modelu na zdarzenia nietypowe.

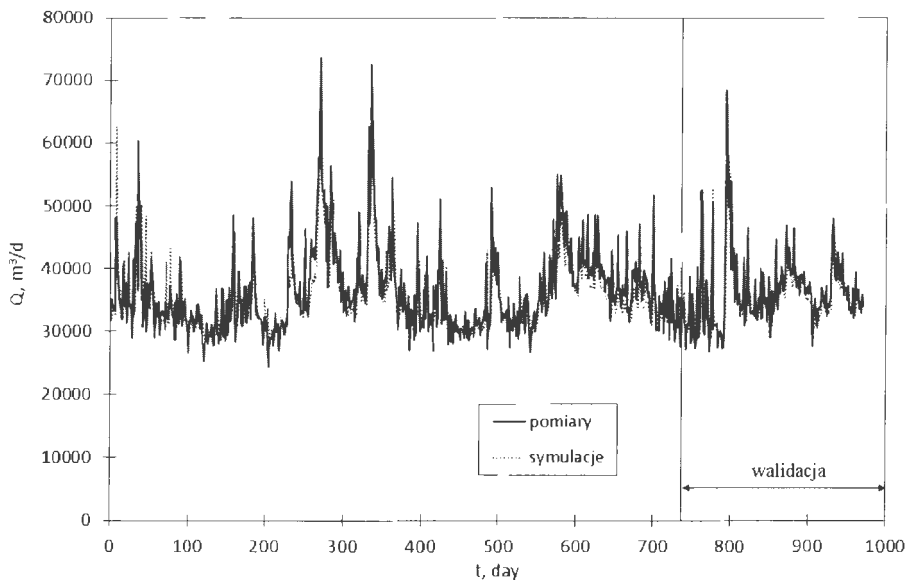
Omówione wyżej wyniki modelowania dopływu do oczyszczalni ścieków (tab. 7.2.2 i 7.2.3) stanowiły początkowy etap analiz do opracowania narzędzia pozwalającego ocenić wpływ różnych zmiennych niezależnych i struktury modelu sieci neuronowej typu perceptron wielowarstwowy na wyniki prognoz (Bartkiewicz i in. 2016). Fakt, że do dalszych rozważań przyjęto metodę MLP, znajduje uzasadnienie w otrzymanych wynikach obliczeń. Mianowicie, rezultaty symulacji dopływu ścieków do oczyszczalni uzyskane za pomocą modeli o bardziej złożonej strukturze (CNN, SVM) różnią się nie więcej, niż 5% od rezultatów obliczeń na bazie metody MLP.

W dalszym ciągu badań obliczenia wykonane dla oczyszczalni Sitkówka-Nowiny powtórzono dla oczyszczalni ścieków w Rzeszowie. Wartości parametrów opisujących strukturę otrzymanych modeli zamieszczono w tab. 7.2.2, a w tab. 7.2.5 podano wyznaczone wskaźniki dopasowania wyników obliczeń do pomiarów. Na rys. 7.2.2 pokazano wyniki modelowania ilości dopływających ścieków surowych modelem MLP. Na podstawie tab. 7.2.2 stwierdzono, że najmniejsze wartości błędów prognozy dopływu (w Sitkówce-Nowiny, Rzeszowie) uzyskano dla funkcji aktywacji tangens hiperboliczny w warstwie ukrytej oraz dla funkcji liniowej w warstwie wyjściowej.

Na podstawie danych w tab. 7.2.5, w odniesieniu do wpływu poszczególnych zmiennych na wyniki modelowania, można dostrzec, iż w większości badanych wariantów modelowania dla obu obiektów występowały podobne związki. Wyznaczone modele charakteryzowały się zadowalającymi zdolnościami predykcyjnymi, co wynika z ustalonych wartości miar dopasowania wyników obliczeń do pomiarów oraz rys. 7.2.2. Spośród rozważanych miar dopasowania dużą zmienność (w zależności od kombinacji zmiennych niezależnych i struktury modelu MLP) w przypadku symulacji wykonanych dla rozpatrywanych oczyszczalni (w Sitkówce-Nowiny i w Rzeszowie) wykazywał współczynnik korelacji. Wyznaczona zmienność jego wartości pozwala na wiarygodną ocenę wpływu poszczególnych czynników (zmiennych niezależnych, struktury modelu) na wyniki obliczeń.

Tab. 7.2.5. Wartości współczynników dopasowania wyników obliczeń (MAE, R, RMSE, MAPE) do pomiarów uzyskane w oparciu o metodę MLP dla oczyszczalni ścieków w Rzeszowie.

Zmienne	Test				Walidacja			
	R	MAE, m ³ /d	RMSE, m ³ /d	MAPE, %	R	MAE, m ³ /d	RMSE, m ³ /d	MAPE, %
Q(t-1)	0,73	3107	4709	8,31	0,74	2786	4288	7,20
Q(t-1), Q(t-2)	0,74	3067	5102	8,05	0,74	2720	4515	7,04
Q(t-1), Q(t-2), Q(t-3)	0,72	3283	5229	8,52	0,75	2950	4671	6,90
P(t-1)	0,50	4339	6970	11,76	0,58	4165	6597	10,89
P(t-1), P(t-2)	0,53	4229	6583	11,45	0,56	4312	6696	11,33
P(t-1), Q(t-1)	0,77	3028	4029	8,14	0,79	2705	3578	7,06
P(t-1), Q(t-1), Q(t-2)	0,75	3009	3934	8,03	0,81	2686	3543	6,15
Q(t-1), Q(t-2), Q(t-3), P(t-1)	0,81	2814	4615	7,62	0,82	2545	4185	6,65
Q(t-1), Q(t-2), Q(t-3), P(t-1), P(t-2)	0,81	2779	4736	7,50	0,83	2482	4243	6,47



Rys. 7.2.2. Porównanie wyników obliczeń z pomiarami doływu ścieków do oczyszczalni w Rzeszowie w analizowanym okresie objętym badaniami.

W dalszych badaniach przy tworzeniu modelu klasyfikacyjnego przyjęto wartość współczynnika korelacji jako miarę wrażliwości modelu na zmiany w doborze zmiennych niezależnych i struktury sieci neuronowej MLP. Ponadto, w dalszych badaniach dotyczących modelowania dopływu ścieków do oczyszczalni przyjęto podział modeli na dwie klasy. W pierwszej klasie otrzymane modele charakteryzują się dobrymi zdolnościami predykcyjnymi, a wartości współczynnika korelacji są nie mniejsze, niż $R = R_{gr} = 0,70$ (korelacja bardzo wysoka wg skali Guilforda). Kolejna klasa obejmuje modele, dla których wyznaczone wartości współczynnika korelacji R nie przekraczają wartości 0,7.

Z uwagi na szereg zalet modelu regresji logistycznej (możliwość jawnej interpretacji uzyskanych wyników, możliwa bezpośrednia ocena wpływu np. liczby neuronów czy funkcji aktywacji na wyniki) i szeroką jego dostępność w pakietach statystycznych został on zastosowany w dalszych rozważaniach.

Ogólną postać modelu logitowego do oceny wpływu zmiennych niezależnych na zdolności predykcyjne modelu MLP można zapisać równaniem:

$$p = \frac{\exp(F(Q,P)+F(N,\theta_1,\theta_2))}{1 + \exp(F(Q,h,P)+F(N,\theta_1,\theta_2))} \quad (7.2.3)$$

gdzie: p – prawdopodobieństwo przekroczenia granicznej wartości współczynnika korelacji (R_{gr}), $F(Q,P)$ – funkcja opisująca liniową kombinację między binarnymi wartościami danych wejściowych $Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $P(t-1)$, $P(t-2)$, $F(N, \theta_1, \theta_2)$ – funkcja definiująca liniową kombinację między binarnymi wartościami zmiennych opisujących strukturę modelu, przy czym: N – liczba sygnałów wejściowych do modelu sieci, θ_1, θ_2 – funkcje aktywacji na wejściach odpowiednio do warstwy ukrytej i warstwy wyjściowej sieci (1 – wykładnicza, 2 – logistyczna, 3 – tangens hiperboliczny, 4 – liniowa, 5 – sinus).

W oparciu o powyższe założenia, stosując program STATISTICA, wyznaczono wartości liczbowe parametrów (α_i) uwzględnionych w równaniu regresji logistycznej oraz odpowiadające im wartości prawdopodobieństw testowych (p). Równocześnie, obliczono parametry dopasowania danych pomiarowych do wyników prognoz (SPEC, SENS, R_2^2) dla badanych oczyszczalni (tab. 7.2.6). Wyniki modelowania wskazują, że istotny wpływ na dokładność prognozy modelu ma dobór jego struktury i uwzględnione w nim zmienne niezależne (por. wartości współczynników α_i). Ponadto, otrzymane wartości wskaźników dopasowania wyników obliczeń do pomiarów pozwalają ocenić, że wyznaczone modele charakteryzują się zadowalającymi zdolnościami predykcyjnymi. W celu oceny wpływu poszczególnych zmiennych na prawdopodobieństwo przekroczenia granicznej wartości współczynnika korelacji określono wartości ilorazów szans (OR) dla opracowanych modeli; wyniki obliczeń przedstawiono w tab. 7.2.7.

Tab.7.2.6. Zestawienie wartości liczbowych parametrów β_i i prawdopodobieństw testowych (p) dla zmiennych wejściowych oraz parametrów opisujących strukturę sieci neuronowej w wyznaczonych modelach regresji logistycznej.

Zmienne	Rzeszów		Sitkówka-Nowiny	
	α_i	p	α_i	p
β_0	-5,256	< 0,0001	-8,052	< 0,0001
Q(t-3)	-0,021	< 0,0001	-0,390	< 0,0001
Q(t-2)	1,121	< 0,0001	0,548	< 0,0001
Q(t-1)	6,125	< 0,0001	8,540	< 0,0001
P(t-1)	0,427	0,00200	0,215	0,00300
P(t-2)	0,114	0,00300	0,183	< 0,0001
act _{wej} (1)	0,756	0,01200	0,021	0,00300
act _{wej} (2)	0,811	0,02300	0,351	0,00100
act _{wej} (3)	0,921	0,03400	0,368	0,00200
act _{wej} (4)	0,201	0,02500	0,466	0,00100
act _{wyj} (1)	0,712	0,02200	0,800	< 0,0001
act _{wyj} (2)	0,821	< 0,0001	0,655	0,00000
act _{wyj} (3)	0,756	< 0,0001	0,596	0,00100
act _{wyj} (4)	0,812	< 0,0001	0,644	0,00200
l. neuronów	0,100	0,03000	0,070	0,00300
SENS		88,15%		85,21%
SPEC		82,23%		91,94%
R_c^2		84,655		84,925

Tab. 7.2.7. Zestawienie wartości ilorazów szans OR dla modeli regresji logistycznej otrzymanych dla oczyszczalni w Sitkówce–Nowiny i Rzeszowie.

Zmienne	OR	
	Rzeszów	Sitkówka-Nowiny
Q(t-3)	0,98	0,68
Q(t-2)	3,07	1,73
Q(t-1)	457,14	5116,02
P(t-1)	1,53	1,24

Tab. 7.2.7. cd.

Zmienne	OR	
	Rzeszów	Sitkówka-Nowiny
P(t-2)	1,12	1,20
act _{wej} (1)	2,13	1,02
act _{wej} (2)	2,25	1,42
act _{wej} (3)	2,51	1,44
act _{wej} (4)	1,22	1,23
act _{wyj} (1)	2,04	2,22
act _{wyj} (2)	2,27	1,92
act _{wyj} (3)	2,13	1,81
act _{wyj} (4)	2,25	1,90
l. neuronów	1,10	1,07

W oparciu o uzyskane ilorazy szans (OR) można stwierdzić, że spośród rozpatrywanych zmiennych wejściowych największy wpływ na prawdopodobieństwo przewyższenia przyjętej wartości $R_{gr} = 0,70$ i na zdolności prognostyczne modelu mają dobowy dopływ ścieków do oczyszczalni i wysokość opadu deszczu, zmierzone w poprzedniej dobie. Ilorazy szans dla oczyszczalni w Sitkówce–Nowiny i Rzeszowie wynoszą odpowiednio $OR=5116,02$ i $OR=457,14$ dla dopływu ścieków oraz $OR=2,342$ i $OR=1,690$ dla opadów deszczu. Otrzymane wyniki znajdują potwierdzenie w rezultatach obliczeń wykonanych przez Bartkiewicz i Studziński (2010), w pracy doktorskiej Chuchro (2009) oraz w symulacjach wykonanych przez Wei in. (2013) i Fernandez i in. (2009). Wielokrotnie większa wartość OR dla oczyszczalni ścieków w Sitkówce-Nowiny niż obiektu w Rzeszowie może wskazywać na krótszy czas przepływu ścieków siecią kanalizacyjną i mniejszą retencję ścieków. Bardzo ciekawą informacją, którą uzyskano za pomocą modelu logitowego, jest możliwość oceny wpływu na wyniki funkcji aktywacji w warstwie ukrytej i wyjściowej. W analizowanych przypadkach dobór tych funkcji miał istotny wpływ na wartość prawdopodobieństwa przekroczenia granicznej wartości R_{gr} .

Jest to cenna informacja, bowiem analizując wartości OR wyznaczone dla różnych funkcji aktywacji można dobrać taką ich kombinację, która zapewnia najlepsze dopasowanie wyników obliczeń do pomiarów. W przypadku doboru funkcji aktywacji dla warstwy wyjściowej stwierdzono, że skrajne różnice między wartościami OR uzyskanymi dla obu oczyszczalni nie są większe niż 20%.

W przypadku doboru funkcji aktywacji dla warstwy ukrytej stwierdzono, że dla obu oczyszczalni największy wpływ naprawdopodobieństwo przekroczenia wartości R_{gr} ma wybór funkcji tangens hiperboliczny. Zależność tą potwierdzają eksperymenty numeryczne wykonane przez Luo i Zhao (2011) zajmujących się modelowaniem oczyszczalni ścieków.

W zakończeniu rozważań można stwierdzić, że model logitowy może być stosowany do oceny wpływu zmiennych objaśniających i wybranych parametrów struktury modelu sieci neuronowej na dokładność modelu do prognozy dopływu. Jego zaletą jest możliwość jawnej oceny wpływu poszczególnych zmiennych niezależnych na prawdopodobieństwo przekroczenia granicznej wartości współczynnika korelacji, czego nie umożliwiają modele czarnej skrzynki typu ANN, SVM itp. Kolejną zaletą modelu logitowego opisanego równaniem (7.2.3) jest możliwość równoczesnej oceny wpływu kilku czynników na wyniki obliczeń. Otrzymane wartości współczynników α_i wykazały zróżnicowanie w odniesieniu do wartości dopływu z poprzedniej doby, jednak dla pozostałych zmiennych (wysokość opadu deszczu i natężenie przepływu zmierzone w dwóch ostatnich dobach) ich wpływ na prawdopodobieństwo przekroczenia (p) wartości granicznej współczynnika korelacji był zbliżony.

7.2.2. Wpływ długości czasu prowadzenia pomiarów na dobór metody symulacji na przykładzie dopływu do oczyszczalni ścieków

Jednym z kluczowych czynników mających wpływ na dokładność modelowania dopływu ścieków do oczyszczalni, w tym również procesów zachodzących w oczyszczalni, jest długość okresu prowadzenia pomiarów. Jest to istotny czynnik determinujący zmienność danych w zbiorze i ich rozkłady empiryczne, a co za tym idzie ma także duży wpływ na dobór właściwej metody do prognozy badanego procesu w oczyszczalni ścieków. Poza długością okresu prowadzenia pomiarów istotne znaczenie może mieć również sposób gromadzenia danych na obiektach, tj. czy dane zbierane są w sposób ciągły ze stałym krokiem czasowym, czy może w różnych odstępach czasu. Mimo tego, że jest to istotny aspekt modelowania, to nie był on do tej pory badany. Nie rozpatrywano także problemu minimalnej wymaganej długości szeregu czasowego w odniesieniu do tworzenia modeli statystycznych. Im dłuższe są zbierane ciągi danych, tym model powinien lepiej odwzorowywać rzeczywistość. Jednak, im dłuższe są te ciągi, tym, w przypadku układów z zamontowanymi na dopływie ścieków analizatorami, koszty ich działania i eksploatacji są wyższe. Identyczny problem dotyczy również pomiarów wykonywanych drogą oznaczeń analitycznych, które są czasochłonne i pracochłonne.

Mając na względzie powyższe uwagi podjęto w pracy próbę ustalenia minimalnego wymaganego okresu prowadzenia pomiarów na przykładzie modelu do prognozy dopływu ścieków do oczyszczalni ścieków wyznaczanego za pomocą

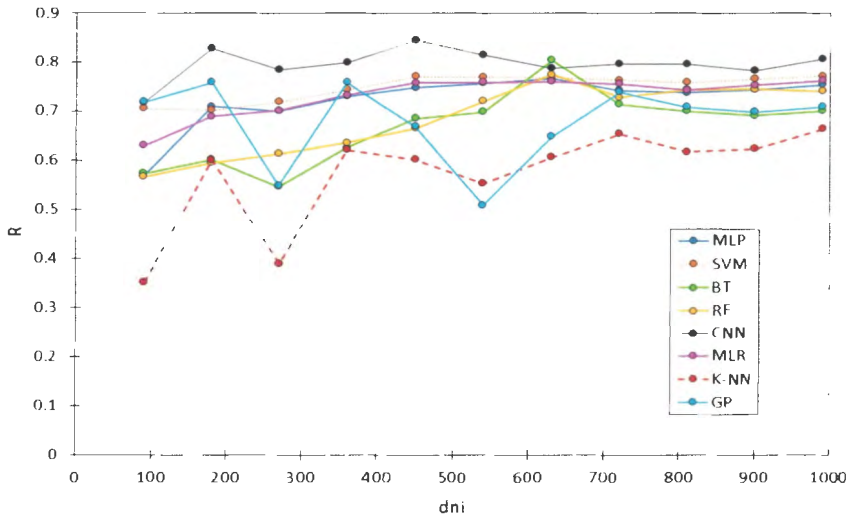
metod MLR, BT, RF, MLP, SVM, CNN i GP. Podjęto również próbę ustalenia referencyjnej metody spośród rozpatrywanych w pracy metod dla przypadków, gdy występują braki w ciągłości danych. W prowadzonych analizach rozpatrzono dwa warianty. W pierwszym z nich przyjęto, że dane gromadzone są w układzie on-line ze stałym krokiem próbkowania odpowiednio przez okres 6, 9, 12, 15, 18, 21, 24, 27, 30 miesięcy. W drugim wariantcie przyjęto, że dane są gromadzone w sposób nieciągły. W oparciu o ustalone zmienne do prognozy dopływu, tj. wartości $Q(t-1)$, $Q(t-2)$, $P(t-1)$, $P(t-2)$ dla różnych długości ciągów pomiarowych opracowano modele za pomocą metod MLR, k-NN, RF, BT, MLP, SVM oraz CNN. Dla wyznaczonych modeli obliczono wartości R dla przyjmowanych okresów pomiarowych, rys. 7.2.3 i 7.2.4.

W celu odpowiedniej interpretacji otrzymanych wyników w pierwszej kolejności zajęto się pomiarami wyznaczonymi w sposób nieciągły. Wykonując pomiary na oczyszczalni ścieków przykładowo 1 lub 3 razy w tygodniu roczna liczba danych zgromadzonych w ciągu roku wynosi odpowiednio 48 i 144 wartości. Tym samym, aby uzyskać na przykład zbiór o wielkości 280 danych, w pierwszym przypadku należy prowadzić pomiary przez okres blisko 6 lat, podczas gdy w drugim przypadku już zaledwie przez 2 lata. Dlatego też, należy mieć na uwadze, że prowadzenie pomiarów w układzie nieciągłym może wiązać się z koniecznością prowadzenia pomiarów wieloletnich w celu uzyskania zbiorów danych odpowiedniej wielkości.

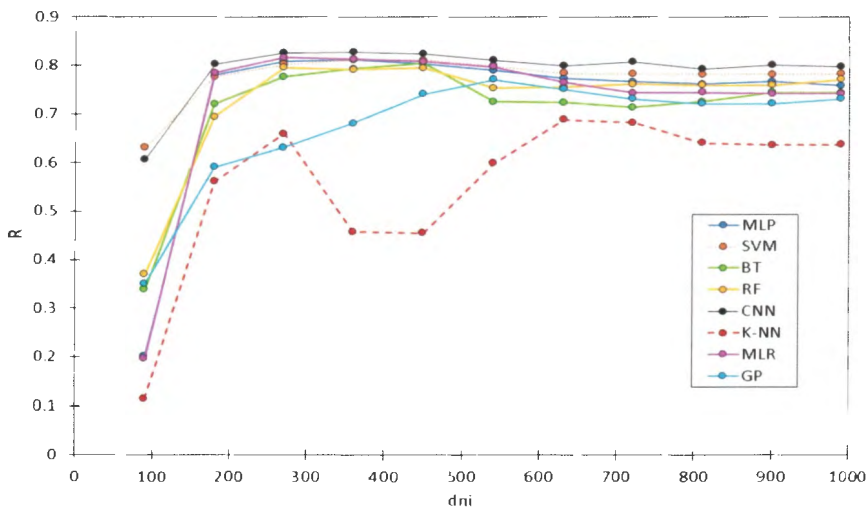
Analizując zmienność wyznaczonych krzywych (rys. 7.2.3 i 7.2.4) wykazano, że liczba danych i sposób ich gromadzenia istotnie wpływa na wyniki modelowania dopływu ścieków do oczyszczalni. Najgorsze wyniki prognoz rozpatrywanych metodami w obu rozważanych wariantach uzyskano dla danych gromadzonych przez okres 3 miesięcy (na rysunkach 90 dni). W przypadku danych zbieranych w sposób ciągły i nieciągły największe błędy prognozy dopływu stwierdzono w metodach k-NN i GP, co potwierdzają obliczone wartości R . Na podstawie otrzymanych wyników można też stwierdzić, że po okresie nie krótszym niż 24 miesiące rezultaty modelu wyznaczonego za pomocą metody GP są zbliżone do rezultatów ustalonych pozostałymi metodami. Z kolei, w przypadku metody k-NN, mimo stosunkowo długiego okresu prowadzenia pomiarów wyniki modelowania są gorsze od pozostałych metod. Tak duża zmienność wartości R w zależności od liczby danych uczących w modelu k-NN wskazuje na dużą wrażliwość metody.

Na podstawie otrzymanych wyników można stwierdzić, że spośród analizowanych metod najmniejszą wrażliwością na liczbę danych pomiarowych i ich zmienność charakteryzują się modele sieci neuronowych (MLP, SVM, CNN). Krótszy okres stabilizacji wyników obliczeń (9 miesięcy), niż w metodzie GP, stwierdzono w modelach otrzymanych metodami SVM, MLR, BT i RF (wartość R nie zmieniała się o więcej niż 10%, mimo wydłużania okresu pomiarów).

W przypadku modelu opracowanego za pomocą metody CNN uzyskano szybszą stabilizację wyników (po okresie 6 miesięcy), niż w przypadku pozostałych metod. Dalsze wydłużanie okresu pomiarów prowadziło do zmian wartości R zaledwie na poziomie 3,5%, co może wskazywać na wysoką stabilność metody.



Rys. 7.2.3. Wpływ długości okresu pomiarów natężenia przepływu na wyniki modelowania dla nieciągłego pomiaru danych na przykładzie oczyszczalni w Sitkówce–Nowiny.



Rys. 7.2.4. Wpływ długości okresu prowadzenia pomiarów natężenia przepływu na wyniki modelowania dla ciągłego pomiaru danych na przykładzie oczyszczalni w Sitkówce–Nowiny.

Bazując na wynikach modelowania można stwierdzić, że różnice w otrzymanych wartościach R dla poszczególnych metod korzystających z danych gromadzonych w sposób ciągły są mniejsze, niż w przypadku modeli tworzonych przy użyciu danych gromadzonych w sposób nieciągły.

W przypadku modeli opracowanych o dane gromadzone w sposób nieciągły (rys. 7.2.3) stwierdzono, że podobnie jak w modelach tworzonych na podstawie danych ciągłych najlepsze wyniki symulacji i najszybszą stabilność rozwiązania uzyskano metodą CNN (już w przypadku 180 danych). Wydłużanie okresu pomiarowego w przedziale 270÷990 dni (w przypadku pomiarów realizowanych w sposób nieciągły, np. 3 razy w tygodniu, odpowiada to dla dolnej granicy przedziału liczbie 1,875 lat, tj. 1 rok i 10 miesięcy, lub 6,875 lat, tj. 6 lat i 10 miesięcy, dla górnej granicy przedziału) prowadziło do zmian współczynnika korelacji na poziomie 5%, przy czym minimalna wartość R nie była mniejsza, niż 0,79. Mniejsze dopasowanie wyników w rozpatrywanych okresach pomiarowych ($R = 0,70 \div 0,76$) uzyskano stosując metody MLR, MLP i SVM, przy czym stabilność uzyskiwanych wyników na stałym poziomie otrzymywano po okresie 9 miesięcy pomiarowych.

W przypadku modeli wyznaczonych metodami RF i BT stwierdzono, że wydłużanie okresu pomiarowego do 21 miesięcy ma istotny wpływ na poprawę dokładności modelowania, a rezultaty symulacji są lepsze, niż wyznaczone bardziej złożonymi modelami (SVM i MLP). Dalsze wydłużanie okresu pomiarowego prowadzi w tym przypadku do zmniejszenia się zdolności predykcyjnych modeli i jednocześnie stabilizacji wyników na poziomie $R=0,70 \div 0,72$.

Otrzymane wyniki znajdują potwierdzenie w obliczeniach wykonanych przez innych autorów. Kusiak i in. (2013), zajmując się prognozą jakości ścieków na dopływie do oczyszczalni (CBZT₃) i korzystając ze zbiorów danych o liczebności 855, wykazali nieznacznie lepsze dopasowanie wyników obliczeń do pomiarów metodą MLP niż RF, co obrazują również krzywe na rys. 7.2.3 i 7.2.4. Kusiak i Wei (2014) w oparciu o zbiór 577 danych uczących opracowali modele (ANFIS, SVM, MLP, k-NN i RF) do prognozy dobowej ilości produkowanego w oczyszczalni biogazu. Wyznaczone przez nich wyniki wykazały, że spośród rozpatrywanych metod najgorsze wyniki otrzymano metodą k-NN. Mniejsze wartości błędów uzyskano metodami RF i MLP, z kolei model opracowany przy pomocy metody SVM charakteryzował się lepszymi zdolnościami predykcji biogazu, niż model MLP. Wyniki te również są potwierdzone na rysunkach 7.2.3 i 7.2.4.

Podsumowując powyższe rozważania można stwierdzić, że na bazie wyznaczonych krzywych można ustalić wpływ długości okresu pomiarowego na wyniki modelowania dopływu ścieków do oczyszczalni i mają one również aspekt

praktyczny. Dysponując wielkością zbioru danych i informacją o sposobie ich gromadzenia można dobrać odpowiednią metodę do prognozowania dopływu do oczyszczalni ścieków. Wyznaczone na zamieszczonych rysunkach krzywe pozwalają na ustalenie minimalnej długości okresu pomiarów, dla którego opracowany model (np. MLP, RF, BT, MLP itp.) charakteryzuje się satysfakcjonującymi zdolnościami predykcyjnymi.

7.3. Prognoza jakości ścieków na dopływie do obiektu

Jednym z istotnych czynników determinujących przebieg procesów biochemicznych zachodzących w bioreaktorze jest jakość ścieków dopływających do obiektu. Znajduje to potwierdzenie w licznych modelach procesowych (tab. 3.2÷3.4), w których szereg zmiennych niezależnych (x_i) stanowią wskaźniki jakości ścieków. Równocześnie, informacja o jakości ścieków z odpowiednim wyprzedzeniem czasowym stwarza możliwość odpowiedniego doboru nastaw reaktora w celu uzyskania optymalnych warunków prowadzenia procesu przez technologa. Dlatego też, podjęto badania (Cartensen i Nielsen 1998, Bechman i Nielsen 1999, Sobota i Szetela 2005, 2006, Martin i in. 2014) zmierzające do opracowania modeli fizykalnych i statystycznych do symulacji jakości ścieków. Z uwagi na ograniczone zastosowanie modeli fizykalnych do prognozy jakości ścieków i potrzebę identyfikacji zdarzeń prowadzących do zaburzeń w działaniu oczyszczalni ścieków wykorzystanie znalazły modele statystyczne (szarej i czarnej skrzynki). Modele te zwykle mają charakter autoregresyjny, czyli wskaźniki jakości ścieków w chwili t oraz $(t + \tau)$ prognozowane są w oparciu o pomiary tych zmiennych w chwilach poprzedzających. Na skutek awarii sond pomiarowych lub inne problemy techniczne mogą wystąpić nieciągłości w szeregach pomiarowych. Powoduje to braki w pomiarach zmiennych ujętych w modelach procesowych. Dlatego też, w niniejszych analizach w celu zapewnienia ciągłości w ciągach pomiarowych tych zmiennych (ChZT, BZT_s, N_{og}, I_o itp.) proponuje się ich zastąpienie wynikami obliczeń uzyskanymi przy pomocy innych modeli matematycznych. W pracy przyjęto, że wartości wybranych wskaźników jakości ścieków na wlocie do oczyszczalni będą prognozowane na podstawie równania (Rosseau i in. 2008, Ahnet i in. 2016, Lubos i in. 2016, Szlag i Studziński 2017):

$$C(t + \tau)_{dop,f} = f(Q(t - 1), \dots, Q(t - z)) \quad (7.3.1)$$

w którym: f – operator wyznaczanych wskaźników jakości ścieków, τ – krok czasowy prognozy, z – opóźnienia czasowe wyznaczone indywidualnie dla każdego obiektu, Q – wartość przepływu mierzona na dopływie do oczyszczalni ścieków, C – modelowane wartości wskaźników jakości ścieków.

Tab. 7.3.1. Wartości testu F i prawdopodobieństw testowych p dla wybranych wskaźników jakości ścieków (Z_{og} , BZT_5 , N_{og}) dla oczyszczalni w Sitkówe–Nowiny.

Z_{og}			BZT_5			N_{og}		
Zmienna	F	p	Zmienna	F	p	Zmienna	F	p
Q(t-3)	5,188955	0,000019	Q(t-3)	5,606284	0,000006	Q(t-5)	8,594092	0,000000
Q(t-2)	4,348797	0,000074	Q(t-4)	5,883967	0,000011	Q(t-4)	9,375573	0,000000
Q(t-5)	4,367232	0,000157	Q(t-1)	5,323337	0,000013	Q(t-1)	7,463390	0,000000
Q(t-4)	4,722765	0,000159	Q(t-6)	3,875105	0,000287	Q(t-2)	6,574459	0,000000
Q(t-1)	3,495234	0,001468	Q(t-5)	3,837461	0,000613	Q(t-3)	5,951143	0,000003
Q(t-6)	2,807908	0,005705	Q(t-2)	3,500874	0,000829	Q(t-6)	5,333198	0,000004
Q(t-7)	2,331556	0,020506	Q(t-8)	2,984251	0,005342	Q(t-7)	4,629288	0,000033
Q(t-8)	2,200964	0,028850	Q(t-7)	2,686018	0,007951	Q(t-8)	2,545482	0,011645
Q(t-9)	2,168029	0,038599	Q(t-9)	2,272668	0,030195	Q(t-9)	2,138568	0,041315

Tab. 7.3.2. Wartości testu F i prawdopodobieństw testowych p dla wybranych wskaźników jakości ścieków ($ChZT$, P_{og} , $N-NH_4$) dla oczyszczalni w Sitkówe–Nowiny.

$ChZT$			P_{og}			$N-NH_4$		
Zmienna	F	p	Zmienna	F	p	Zmienna	F	p
Q(t-3)	4,805356	0,000051	Q(t-4)	5,831312	0,000013	Q(t-1)	9,435602	0,000000
Q(t-4)	4,512563	0,000127	Q(t-5)	5,148807	0,000021	Q(t-5)	6,938248	0,000000
Q(t-5)	3,985647	0,0004	Q(t-3)	4,333345	0,000172	Q(t-2)	5,907549	0,000001
Q(t-1)	3,305466	0,002375	Q(t-2)	4,035140	0,000183	Q(t-3)	6,214499	0,000001
Q(t-7)	2,697349	0,007710	Q(t-6)	3,278977	0,001554	Q(t-4)	6,019130	0,000008
Q(t-8)	2,310079	0,021715	Q(t-7)	2,591124	0,010302	Q(t-7)	4,903307	0,000015
Q(t-2)	1,965463	0,032496	Q(t-9)	2,685765	0,011201	Q(t-6)	4,211407	0,000110
Q(t-6)	1,627349	0,042496	Q(t-1)	2,643186	0,012374	Q(t-8)	2,132220	0,041980

Zależność opisana równaniem (7.3.1) bazuje na założeniu, że jakość ścieków na dopływie do oczyszczalni kształtowana jest przez stopień rozcieńczenia ścieków wywołany wielkością dopływu (Kaczor i in. 2008, Ahnet i in. 2016). Postępując zgodnie z algorytmem na rys. 6.6 wykonano obliczenia na podstawie pomiarów

wskaźników fizyko–chemicznych ścieków na dopływie do oczyszczalni w Sitkówce-Nowiny. Do analizy wyników został wykorzystany test F–Snedecora. Na podstawie pomiarów natężenia dopływu wyznaczono wartości testu na poziomie istotności $\alpha < 0,05$ dla wybranych wskaźników jakości ścieków (BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄); w tab. 7.3.1 i 7.3.2 podano wartości testu F i prawdopodobieństw testowych p .

Na podstawie otrzymanych wyników można stwierdzić, że wartości wybranych wskaźników jakości są poprawnie opisywane przez wartości przepływów $Q(t - 1)$ do $Q(t - 9)$, co oznacza, że dynamika procesu dopływu ścieków w czasie kształtuje jakość ścieków na wlocie do obiektu. W rezultacie przystąpiono do wyznaczania odpowiednich modeli dla wskaźników jakości ścieków na dopływie do oczyszczalni.

Stosując metody MLR uzyskano zależności:

$$BZT(t) = -0,0031(\pm 0,0012) \cdot Q(t - 1) + 544,98(\pm 36,82) \quad (R=0,45) \quad (7.3.2)$$

$$ChZT(t) = -0,0031(\pm 0,0012) \cdot Q(t - 1) + 544,98(\pm 36,82) \quad (R=0,42) \quad (7.3.3)$$

$$Z_{og}(t) = -0,0050(\pm 0,0011) \cdot Q(t - 1) + 1153,58(\pm 75,91) \quad (R=0,36) \quad (7.3.4)$$

$$N_{og}(t) = -0,00031(\pm 0,00015) \cdot Q(t - 1) + 116,92(\pm 4,34) \quad (R=0,44) \quad (7.3.5)$$

$$N - NH_4(t) = -0,00033(\pm 0,0001) \cdot Q(t - 1) - 0,00032(\pm 0,00001) \cdot Q(t - 3) + 75,03(\pm 2,85) \quad (R=0,53) \quad (7.3.6)$$

$$TP(t) = -0,0005(\pm 0,0012) \cdot Q(t - 1) + 11,56(\pm 0,64) \quad (R=0,44) \quad (7.3.7)$$

Następnie przystąpiono do budowy modeli nieliniowych. Wykorzystując metodę programowania genetycznego otrzymano modele postaci:

$$BZT_5 = \frac{0,263}{0,001 \cdot Q(t-1)^{0,17} - 0,006 \cdot Q(t-2)^{1,74}} - 161,03 \quad (R=0,45) \quad (7.3.8)$$

$$ChZT = \frac{0,523}{0,0002 \cdot Q(t-1)^{0,25} - 0,0003 \cdot Q(t-2)^{-0,21}} + 35,85 \quad (R=0,42) \quad (7.3.9)$$

$$Z_0 = \frac{0,306}{0,0002 \cdot Q(t-1)^{0,25} - 0,0003 \cdot Q(t-2)^{-0,21}} + 41,12 \quad (R=0,36) \quad (7.3.10)$$

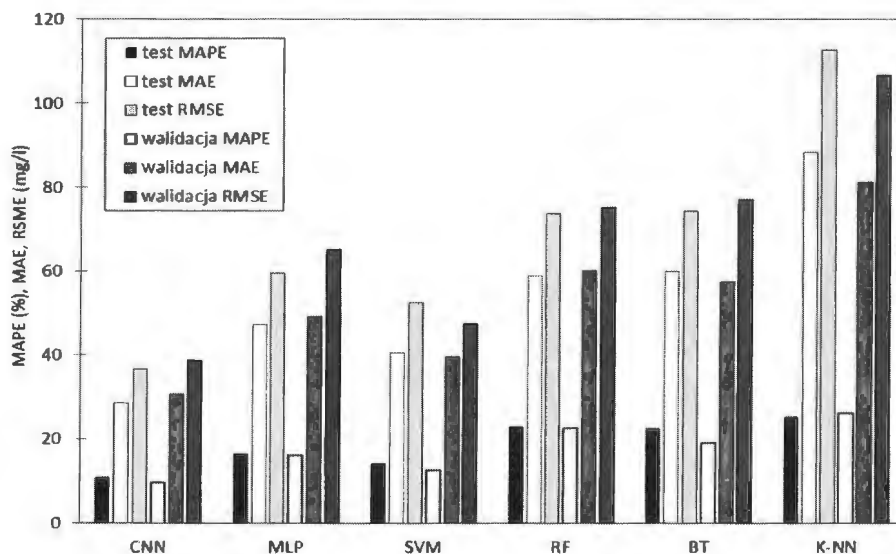
$$N_{og} = \frac{0,343}{0,001 \cdot Q(t-1)^{0,09} - 0,008 \cdot Q(t-2)^{-0,21}} - 105,69 \quad (R=0,50) \quad (7.3.11)$$

$$N - NH_4 = \frac{0,279}{0,001 \cdot Q(t-1)^{0,063} - 0,006 \cdot Q(t-2)^{-0,21}} - 110,19 \quad (R=0,60) \quad (7.3.12)$$

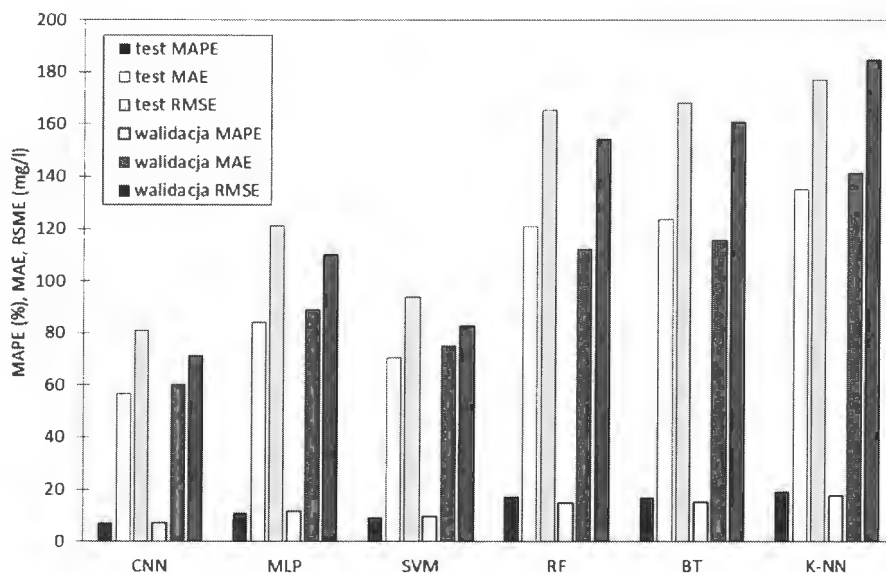
$$P_{og} = \frac{0,203}{-0,0117 \cdot Q(t-1)^{-0,0935} + 0,0081 \cdot Q(t-2)^{0,046}} - 18,02 \quad (R=0,44) \quad (7.3.13)$$

Powyższe wyniki obliczeń wskazują na niezadowalające dopasowanie wyników obliczeń do pomiarów wyznaczonych metodą programowania genetycznego (por. wartości R i RMSE). Podobne wyniki uzyskali Kusiaka i Wei (2013), którzy wykorzystując metodę GP próbowali opracować modele do prognozowania wskaźników jakości ścieków na dopływie do oczyszczalni i otrzymali również wysokie wartości błędów prognozy.

W związku z powyższym, w następnym etapie badań do modelowania wskaźników jakości ścieków na dopływie przyjęto metody RF, BT, k-NN, MLP, SVM i CNN. Wartości miar dopasowania wyników obliczeń do pomiarów (RMSE, MAE, MAPE) wyznaczone dla prognozowanych wskaźników jakości ścieków dla zbioru testowego i walidacyjnego pokazano na rys. 7.3.1÷7.3.6, z kolei na rys. 7.3.7÷7.3.12 przedstawiono wyniki symulacji tych wskaźników. Wartości współczynników R i MAPE stanowiły podstawę porównania uzyskanych wyników z pracami innych autorów (tab. 3.3), którzy zajmowali się symulacją jakości ścieków na dopływie do oczyszczalni ścieków. W tab. 7.3.3 zamieszczono parametry opisujące optymalne struktury modeli wyznaczonych metodami MLP, CNN, SVM.



Rys. 7.3.1. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości BZT₅ dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.

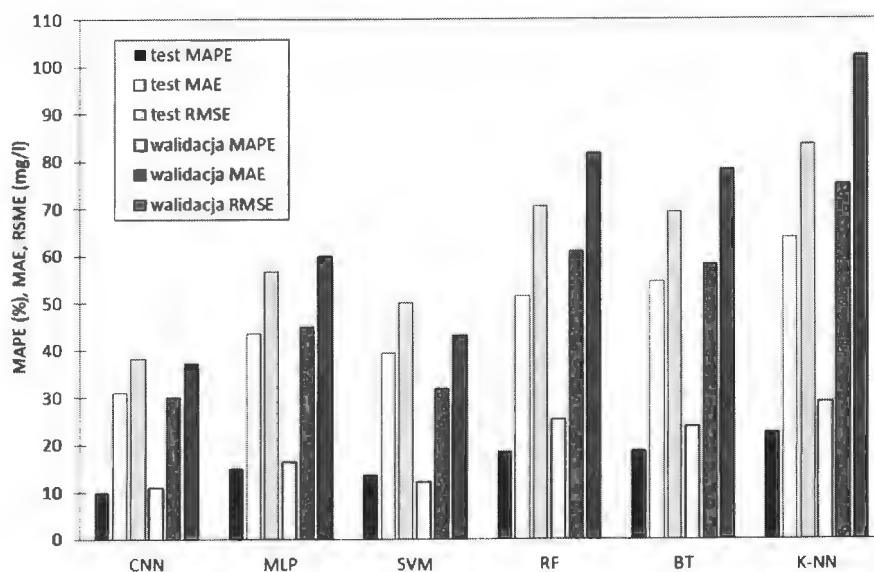


Rys. 7.3.2. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości ChZT dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.

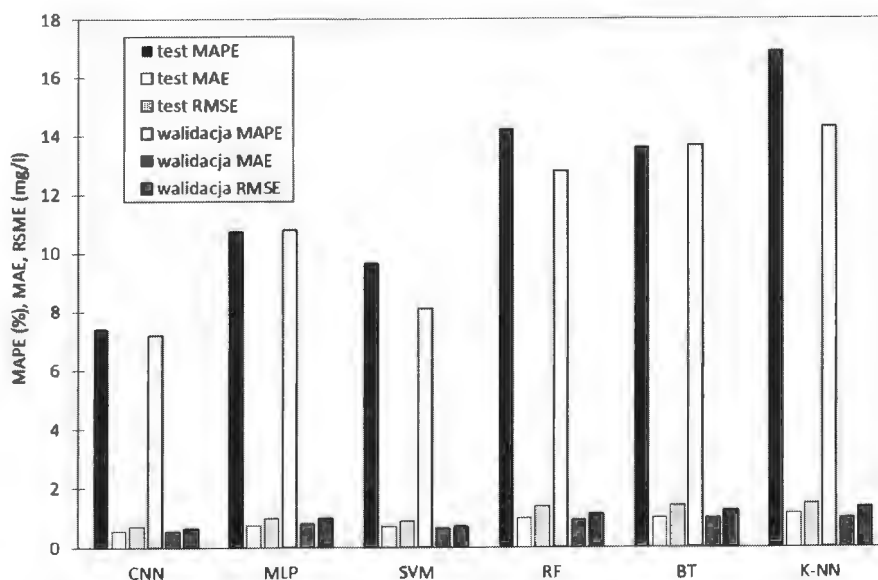
Na podstawie tab. 7.3.3 można stwierdzić, że liczba neuronów w warstwie ukrytej dla modeli MLP wynosiła $13 \div 16$ i nie przekroczyła wartości $(2 \cdot j + 1)$, co wskazuje, że otrzymane modele nie są przeuczone. W przypadku modeli określonych przy pomocy metody CNN liczba neuronów w poszczególnych warstwach ukrytych nie jest większa, niż 8. W modelach MLP i CNN funkcjami aktywacji najczęściej był tangens hiperboliczny. W modelach do prognozy wartości BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄ uzyskanych metodą SVM stwierdzono znaczną zmienność wartości C, γ i ϵ .

Tab. 7.3.3. Zestawienie parametrów opisujących strukturę modeli (MLP, CNN, SVM) do prognozy wybranych wskaźników jakości ścieków na dopływie do oczyszczalni ścieków w Sitkówce–Nowiny.

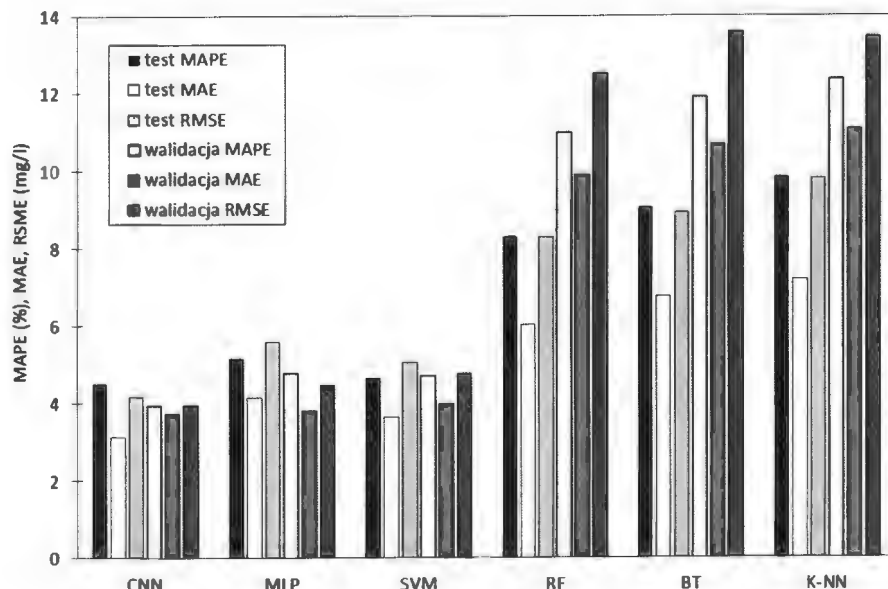
Wskaźniki	MLP		CNN		SVM		
	l. neuron	f. aktyw.	l. neuron	f. aktyw.	C	ϵ	γ
BZT ₅	16	exp	2x7	tanh	40	0,001	0,25
ChZT	14	tanh	2x8	exp	200	0,004	0,50
Z _{og}	15	tanh	2x8	tanh	120	0,007	0,61
N _{og}	12	tanh	2x7	lin	80	0,010	0,45
N-NH ₄	13	log	2x7	log	110	0,023	0,32
P _{og}	13	tanh	2x7	tanh	90	0,004	0,44



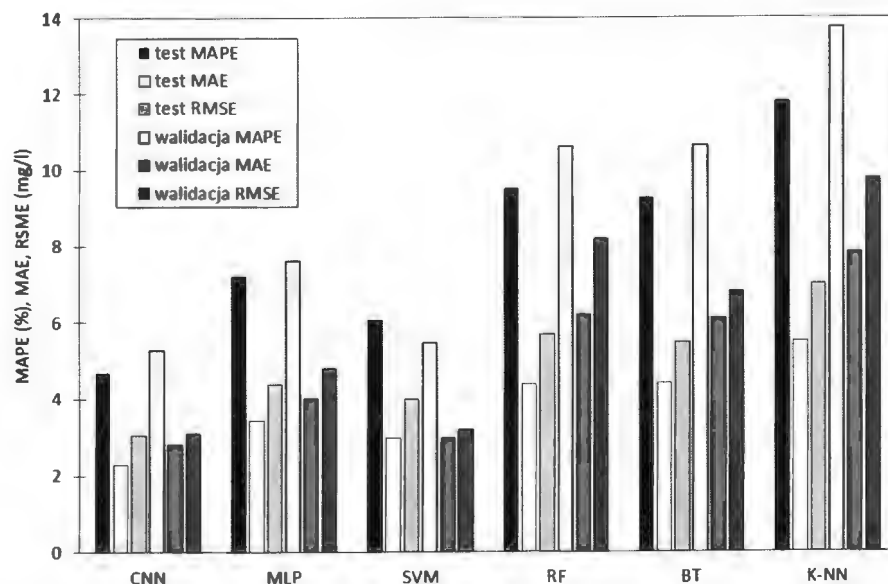
Rys. 7.3.3. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości Z_{og} dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.



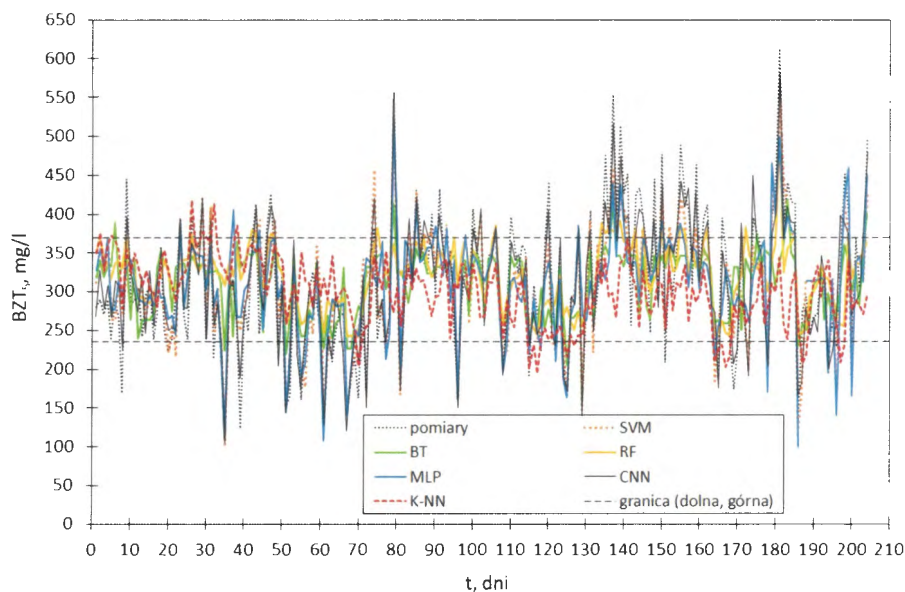
Rys. 7.3.4. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości P_{og} dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.



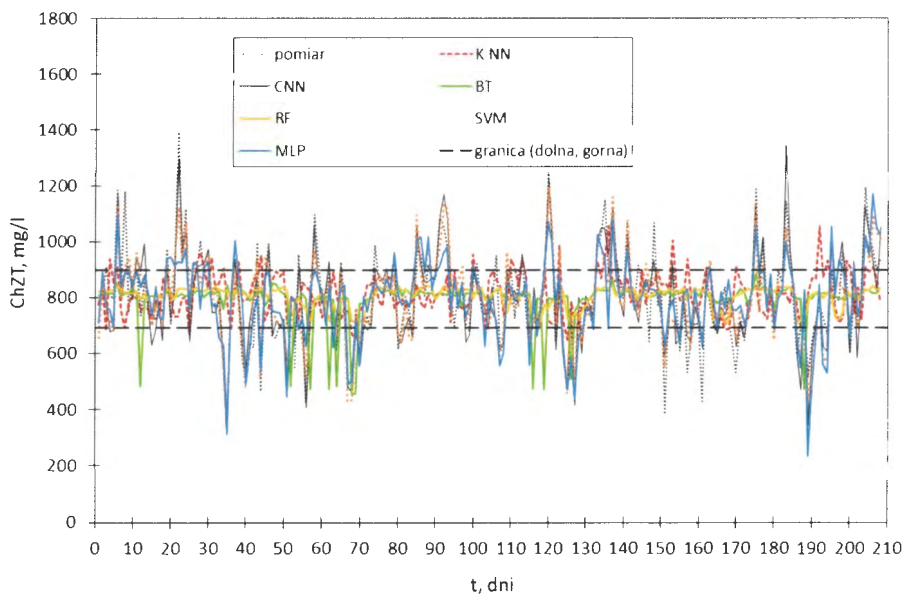
Rys. 7.3.5. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości N_{Og} dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.



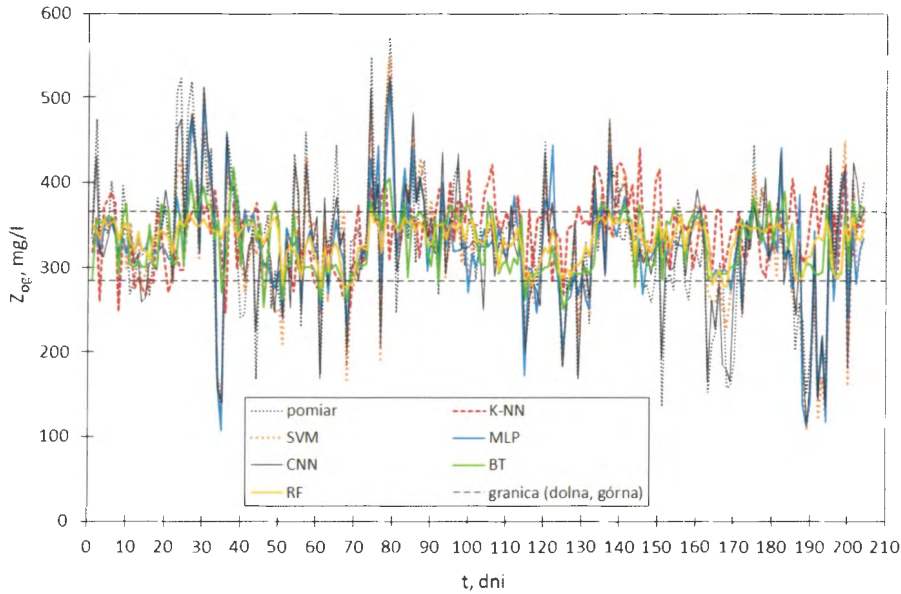
Rys. 7.3.6. Zestawienie wartości współczynników dopasowania wyników obliczeń (MAE, RMSE, MAPE) wartości $N-NH_4$ dla zbioru testowego i walidacyjnego dla oczyszczalni w Sitkówce–Nowiny.



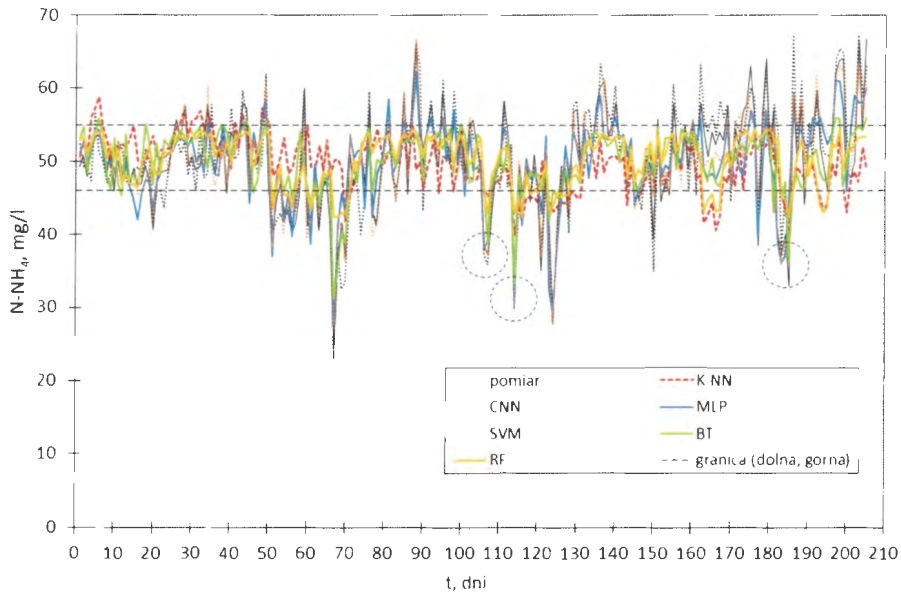
Rys. 7.3.7. Porównanie pomiarów wartości BZT₅ z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce–Nowiny.



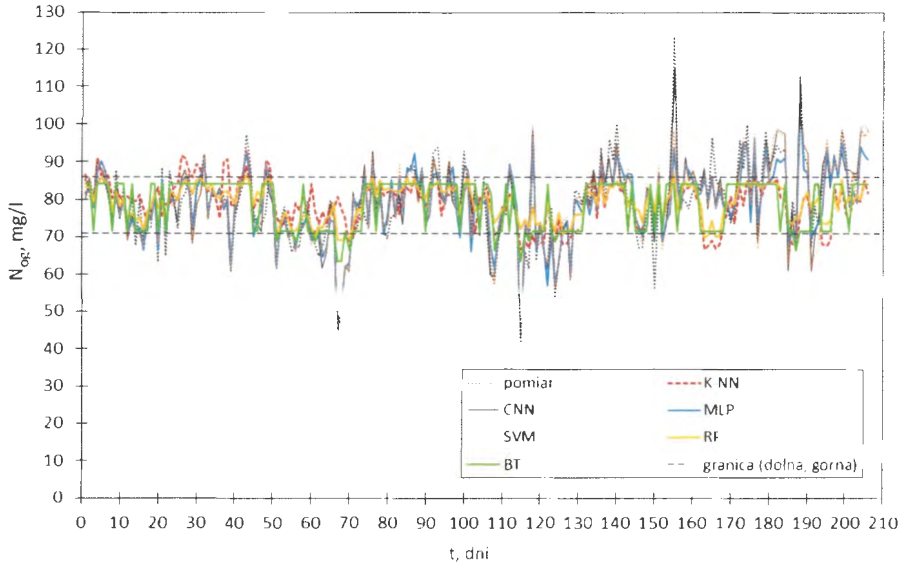
Rys. 7.3.8. Porównanie pomiarów wartości ChZT z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce–Nowiny.



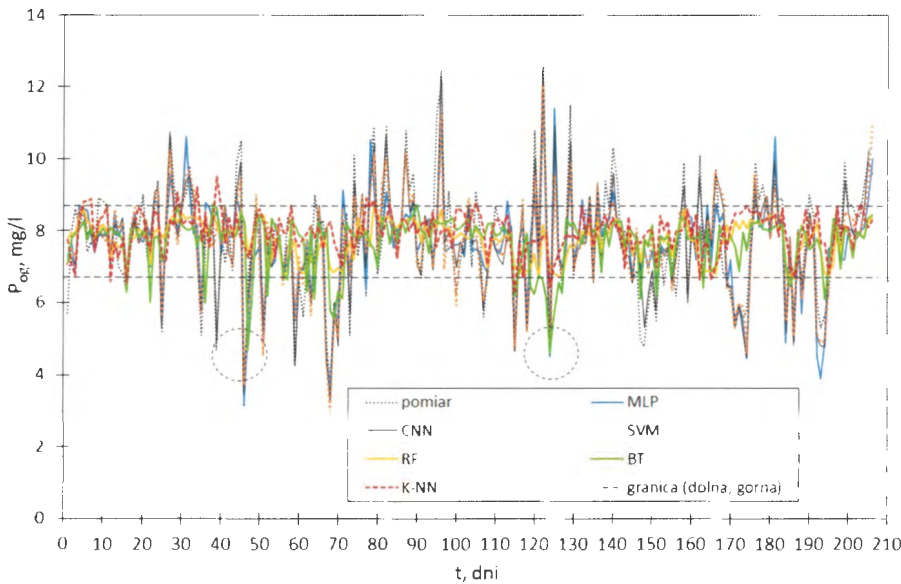
Rys. 7.3.9. Porównanie pomiarów wartości Z_{og} z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce-Nowiny.



Rys. 7.3.10. Porównanie pomiarów wartości $N-NH_4$ z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce-Nowiny.



Rys. 7.3.11. Porównanie pomiarów wartości N_{og} z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce–Nowiny.



Rys. 7.3.12. Porównanie pomiarów wartości P_{og} z obliczeniami metodami RF, BT, k-NN, MLP, SVM i CNN dla oczyszczalni w Sitkówce–Nowiny.

Analiza zmienności krzywych na rys. 7.3.1÷7.3.12 wykazała, że gorsze dopasowanie wyników obliczeń do pomiarów wybranych wskaźników jakości

ścieków otrzymano przy pomocy metod RF, BT, k-NN. Wskazują na to wartości błędów prognozy wskaźników jakości ścieków dla zbioru testowego i walidacyjnego, które, przykładowo, w metodzie RF dla N_{og} wynoszą $MAE=10,76$ mg/l, $RMSE=12,16$ mg/l i $MAPE=10,78\%$. W modelu wyznaczonym metodą CNN wynoszą one $MAE=3,75$ mg/l, $RMSE=5,03$ mg/l i $MAPE=5,18\%$. Wartości błędów są znacząco mniejsze w modelach wyznaczonych przy pomocy metod CNN, MLP i SVM niż w pozostałych metodach. Przykładowo, w modelu do prognozy P_{og} uzyskany w oparciu o metodę SVM wynoszą one $MAE=0,64$ mg/l, $RMSE=0,70$ mg/l i $MAPE=8,1\%$. Z kolei, w modelu otrzymanym metodą CNN błędy te są równe $MAE=0,56$ mg/l, $RMSE=0,62$ mg/l i $MAPE=7,20\%$. Wynik ten wskazuje na lepsze dopasowanie wartości odstających niż w modelach wyznaczonych za pomocą metod RF, BT, k-NN. W modelu opracowanym metodą CNN wartości maksymalne i minimalne są modelowane z większą dokładnością, niż w modelu wyznaczonym metodą wektorów nośnych (rys. 7.3.7÷7.3.12). Z wykreślonych krzywych wynika także, że w większości przypadków modele otrzymane za pomocą metod k-NN, RF i BT zaniżają wartości minimalne i maksymalne. Z kolei, wartości średnie są za pomocą tych metod identyfikowane z zadowalającą dokładnością.

Fakt, iż otrzymane wyniki obliczeń wskaźników jakości ścieków za pomocą metod RF, BT i k-NN są gorsze, niż z użyciem metod SVM, CNN, MLP, może wynikać z ograniczonej wielkości zbioru danych uczących, co determinuje ich rozkład statystyczny i zakres zmienności. Ma to istotny wpływ na poszukiwanie podobieństw między wektorami danych $[x_1, x_2, x_3 \dots, x_i]$ w metodzie k-NN, a w metodach RF i BT wpływa na tworzenie reguł generujących strukturę drzew w układzie równoległym lub szeregowym. Otrzymane wyniki znajdują potwierdzenie w analizach Kusiaka i Wei (2014), którzy podjęli próbę implementacji metod k-NN i RF do symulacji ilości biogazu uzyskując także niezadowalające wyniki prognoz.

W oparciu o krzywe na rys. 7.3.1÷7.3.12 można również stwierdzić, że modele opracowane metodami sztucznych sieci neuronowych (MLP, CNN, SVM) charakteryzują się lepszymi zdolnościami predykcyjnymi wybranych wskaźników jakości ścieków (BZT_5 , ChZT, N_{og} , Z_{og} , P_{og} , $N-NH_4$), niż modele wyznaczone metodami k-NN, RF i BT. W modelach sieci neuronowych możliwości doboru i kształtowania struktury modelu poprzez zmianę odpowiednich współczynników (liczba neuronów, funkcja aktywacji i ilość dodatkowych połączeń między sąsiednimi warstwami w modelach CNN i MLP; stała regularyzacji C , próg nieczułości ϵ i funkcja Kernela w metodzie SVM) są większe, niż w pozostałych metodach, mimo ograniczonej wielkości zbioru danych (Bagherii i in. 2015). W efekcie pozwala to na ustalenie struktury modeli o zadowalających zdolnościach predykcyjnych.

Analiza zmienności uzyskanych krzywych pozwala także stwierdzić, że w przypadku metod CNN, MLP, SVM największe wartości błędów prognozy wskaźników jakości ścieków otrzymano metodą MLP. Przykładowo, w modelu do prognozy ChZT wyznaczonym metodą MLP wartości błędów wynoszą MAPE=11,82%, MAE=89,1 mg/l i RMSE=110,31 mg/l. Natomiast, w modelu CNN są one równe MAPE=7,13%, MAE=60,15 mg/l i RMSE=71,33 mg/l. Różnice w wartościach MAE i RMSE wskazują, że w metodzie MLP występuje szereg wartości wskaźników, które są niedoszacowane.

Lepsze zdolności predycyjne wybranych wskaźników jakości ścieków otrzymane metodą SVM w porównaniu z MLP znajdują teoretyczne uzasadnienie. Z uwagi na to, że w metodzie SVM problem uczenia sprowadzono do programowania kwadratowego, nie ma się w nim do czynienia z funkcją celu z wieloma minimami, co ma miejsce w przypadku uczenia sieci MLP. Jednocześnie, wprowadzenie dodatkowych połączeń w modelach MLP prowadzi do poprawy zdolności predycyjnych modeli, a uzyskane rezultaty obliczeń dla modelu CNN są lepsze, niż dla modelu SVM (Szelaąg i in. 2018a). Modyfikacja modeli MLP w przypadku symulacji zmiennych BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄ prowadzi do zmniejszenia wielkości RMSE średnio o 66%. Z kolei, w przypadku modeli wyznaczonych metodą CNN wartości błędów RMSE i MAE są mniejsze odpowiednio o 11÷18% i 5÷12% w porównaniu z metodą SVM.

W celu potwierdzenia poprawności zależności opisanej równaniem (7.3.1) wykonano obliczenia dla tych samych wskaźników jakości ścieków, lecz dla innego obiektu - oczyszczalni ścieków w Rzeszowie (Szelaąg i in. 2017c). W badaniach ograniczono się do stosowania jedynie metod kaskadowych sieci neuronowych CNN i lasów losowych RF; wyniki obliczeń zamieszczono w tab. 7.3.4 i 7.3.5.

Tab. 7.3.4. Zestawienie współczynników dopasowania wyników obliczeń metodą CNN do pomiarów wybranych wskaźników jakości ścieków w Rzeszowie.

Parametry	Liczba	Test			Walidacja		
	neuronów	MAE, mg/l	MAPE, %	RMSE, mg/l	MAE, mg/l	MAPE, %	RMSE, mg/l
BZT ₅	14	37,29	11,29	47,25	32,36	10,09	43,71
ChZT	16	96,70	12,59	117,56	112,5	13,14	120,21
Z _{og}	16	60,9	16,61	80,15	65,23	15,2	86,21
N _{og}	12	5,76	8,96	7,25	5,13	8,36	6,95
N-NH ₄	12	4,83	10,34	6,26	5,12	11,02	7,31
P _{og}	16	1,76	15,54	2,56	1,41	12,48	2,12

Tab. 7.3.5. Zestawienie współczynników dopasowania wyników obliczeń metodą RF do pomiarów wybranych wskaźników jakości ścieków w Rzeszowie.

Parametry	Test			Walidacja		
	MAE, mg/l	MAPE, %	RMSE, mg/l	MAE, mg/l	MAPE, %	RMSE, mg/l
BZT ₅	42,71	13,49	53,12	46,78	14,32	54,15
ChZT	102,56	15,38	119,87	107,12	17,57	120,12
Z _{og}	71,32	19,27	85,21	77,62	21,27	87,27
N _{og}	6,25	10,03	7,91	6,87	10,78	8,04
N-NH ₄	5,00	10,75	7,12	5,84	11,96	7,99
P _{og}	1,89	16,54	2,88	2,11	18,86	3,45

Wykonane badania potwierdziły przypuszczenie, że pomiary dopływu pozwalają na prognozę jakości ścieków surowych. Jest to bardzo istotna konkluzja, bowiem wykorzystując łatwe do uzyskania pomiary natężenia dopływu można poprawnie prognozować zawartość w ściekach wybranych zanieczyszczeń (BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄). Stanowi to istotne uproszczenie w odniesieniu do prac innych autorów (Sobota i Szetela 2005, 2006, Abyaneh 2013, Verma i Kusiak 2013, Minsoo i in. 2016), którzy do symulacji wskaźników jakości ścieków na dopływie wykorzystywali ich wartości mierzone z odpowiednim opóźnieniem czasowym. Zaprezentowane w pracy rozwiązanie daje możliwość eliminacji prowadzenia kosztownych oznaczeń wskaźników jakości ścieków. Ponadto, pozwala na redukcję liczby automatycznych analizatorów jakości ścieków montowanych na dopływie i kosztów ich eksploatacji, co jest istotne z ekonomicznego punktu widzenia. Spośród rozpatrywanych wyżej metod najlepsze wyniki symulacji wybranych wskaźników jakości ścieków otrzymano kaskadowymi sieciami neuronowymi i metodą wektorów nośnych. Rezultaty te potwierdzają, że modele te nadają się do symulacji silnie nieliniowych zależności. Wyniki obliczeń zmodyfikowanymi metodami drzew regresyjnych (RF) wykonane przez Verma i Kusiaka (2012), Minsoo i in. (2016) oraz Kusiaka i Verma (2013) wskazują, że metody te mogą być porównywalne z metodami ANN.

Zagadnieniem prognozy wskaźników jakości ścieków na dopływie do oczyszczalni ścieków za pomocą modeli statystycznych zajmował się szereg autorów, dlatego wskazane jest porównanie otrzymanych rezultatów modelowania z ich wynikami. W tab. 7.3.6 zestawiono wartości miar dopasowania wyników obliczeń do pomiarów (MAPE, R) wyznaczone w pracy i przez innych autorów. Można stwierdzić, że lepsze dopasowanie wyników obliczeń (metodami MLR, MLP) do pomiarów BZT₅, niż otrzymane w pracy (R = 0,93), uzyskali Dogan i in.

(2008). Wynika to z faktu, iż do opisu BZT₅ wykorzystali oni większą liczbę zmiennych (ChZT, N, P, Z_{og} i Q), niż jest to w obecnej pracy. Mniejsze wartości R (o ok. 1%) w odniesieniu do modelowania BZT₅ podał także Sobota i Szetela (2006). Podobne wartości R otrzymali również Ebrahimi i in. (2017), którzy stosując nieliniowy model regresji z interakcjami prognozowali BZT₅ na podstawie pomiarów P, N, Z_{og}. Z kolei większe wartości błędów dopasowania wyników obliczeń do pomiarów BZT₅ (o ok. 40%) uzyskali Hou i in. (2005), którzy zastosowali w badaniach metodę wygładzania wykładniczego.

Tab. 7.3.6. Porównanie wyników obliczeń wybranych wskaźników jakości ścieków (BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄); gdzie: x/x – dopasowanie modelu dla okresu bezdeszczowego/opadowego, x – autorzy w pracy nie podali wartości miar dopasowania, a jedynie zobrazowali wyniki symulacji na rysunkach

Wskaźniki jakości	Źródło	Wartości miar dopasowania
BZT ₅	Dogan i in. (2008)	R=0,96/0,95
	Ebrahimi i in. (2017)	R=0,91
	Huo (2008)	R=0,63
	Sobota i Szetela (2005)	R=0,92
	b. własne	R=0,93
ChZT	Ahnet i in. (2017)	R=0,82
	Minsoo i in. (2016)	MAPE=3,63/7,35 %
	Sobota i Szetela (2005)	R = 0,92
	b. własne	R=0,89; MAPE=9,68%
Z _{og}	Sobota i Szetela (2005)	R=0,77
	Verma i in. (2012)	MAPE=13,34%
	b. własne	R=0,88; MAPE=11,02%
N-NH ₄	Hacke i Kohne (1996)	R=0,92
	Sobota i Szetela (2006)	x*
	Huo (2008)	R=0,79
	b. własne	R=0,90
N _{og}	Minsoo i in. (2016)	MAPE=4,54/5,79 %
	b. własne	MAPE=4,01%
P _{og}	Ebrahimi i in. (2017)	R=0,92
	Minsoo i in. (2016)	MAPE=3,81/8,87 %
	b. własne	R=0,87; MAPE=7,21%

W przypadku modelowania ChZT nieznacznie większą wartość R (o ok. 3%), niż otrzymaną w pracy, uzyskali Sobota i Szetela (2006). Z kolei Minsoo i in. (2016), stosując zmodyfikowaną metodę k-NN, wyznaczyli mniejsze wartości błędów MAPE. Natomiast, większe wartości błędów prognozy ChZT podali Ahnet i in. (2016), którzy opracowali model bazujący na funkcji wykładniczej.

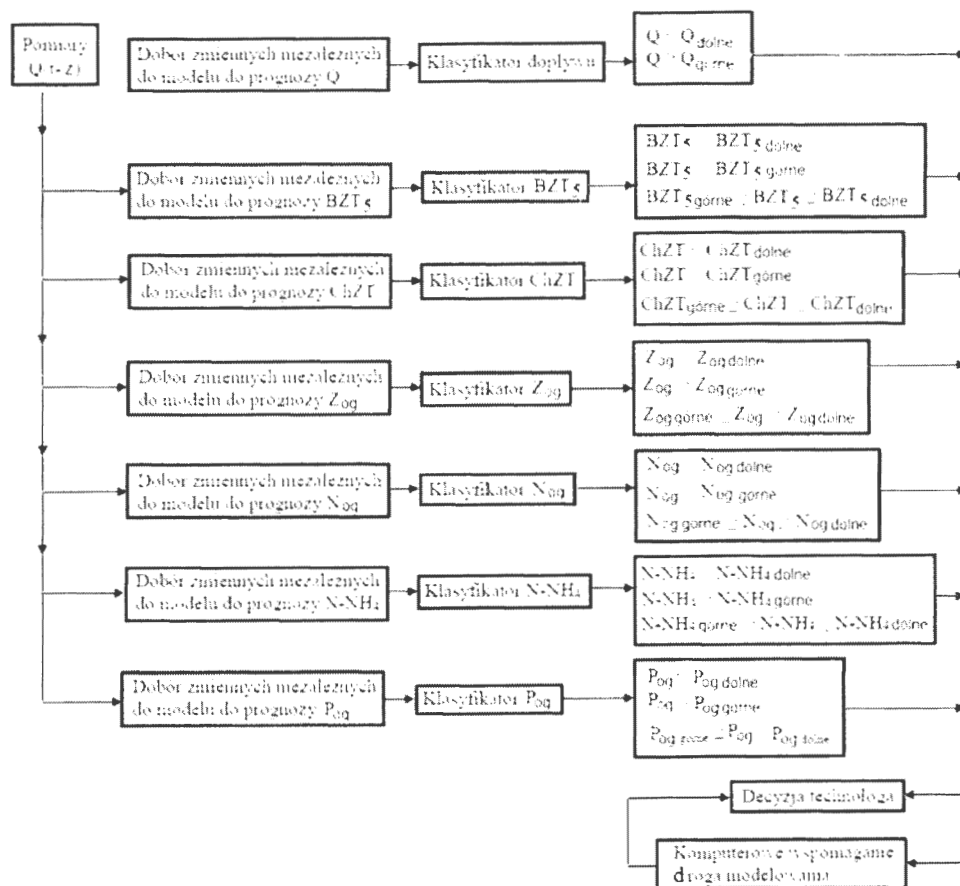
Sobota i Szetela (2006) do prognozy Z_{og} zastosowali identyczną metodykę obliczeń, jak dla BZT₅ i ChZT, uzyskując tym razem mniejszą wartość dopasowania wyników obliczeń do pomiarów (o ok. 14%), niż w obecnej pracy. Interesujące rozwiązanie przedstawili Verma i in. (2012), którzy opracowali model MLP do symulacji Z_{og} w oparciu o pomiary przepływu ścieków z 10-minutową rozdzielczością, otrzymując mniejsze wartości błędów, niż podane w pracy.

Analiza danych w tab. 7.3.6 wskazuje, że wyznaczony w niniejszej pracy model do prognozy N-NH₄ charakteryzuje się porównywalną dokładnością, co model zaproponowany przez Häck i Köhne (1996), którzy wykorzystali w swoich obliczeniach wyniki pomiarów ChZT i mętności.

Wyniki modelowania N_{og} i P_{og} metodą CNN wykazały rezultaty zbliżone do uzyskanych przez Minsoo i in. (2016), którzy w obliczeniach zastosowali zmodyfikowaną metodę k-NN. Wyniki obliczeń P_{og} przedstawione przez Ebrahimi i in. (2017) przy pomocy nieliniowej regresji z interakcją okazały się lepsze, niż w pracy. Można to przypuszczalnie wyjaśnić faktem, że Ebrahimi i in. (2017) do budowy modelu wykorzystali więcej zmiennych (BZT₅, Z_{og} , N_{og}) silnie skorelowanych z wartością rozpatrywanego wskaźnika.

7.3.1. Identyfikacja stanów incydentalnych na dopływie do oczyszczalni ścieków

Poza symulacją wskaźników jakości ścieków istotne znaczenie może odgrywać identyfikacja stanów nietypowych prowadzących do zaburzeń w działaniu reaktora biologicznego. W wielu przypadkach identyfikacja stanu na dopływie poszczególnych wartości wskaźników jakości ścieków (BZT₅, ChZT, N_{og} , Z_{og} , P_{og} , N-NH₄) może być bardzo cenna dla technologa obiektu. Doświadczonemu technologowi daje to możliwość na podjęcie odpowiedniej decyzji, która pozwoli na właściwy dobór nastaw w komorach osadu czynnego (korektę stopnia recyrkulacji, ilości dawkowanego koagulanta, zewnętrznego źródła węgla itp.). Obecnie problem identyfikacji i ograniczania nietypowych stanów działania bioreaktora jest tematem wielu opracowań (Avella i in. 2011, Haimi i in. 2016), a stosowane w oczyszczalniach układy sterowania pozwalają uzyskać ciągłość procesów zachodzących w komorach osadu czynnego. Dla zapewnienia wysokiej niezawodności działania tych układów wymagana jest znajomość ciągłych szeregów czasowych badanych wskaźników jakości, co nie zawsze jest wykonalne.



Rys. 7.3.13. Schemat ideowy klasyfikatora ilości i jakości ścieków.

Dlatego też, biorąc pod uwagę problemy związane z prognozowaniem ciągłych wartości wskaźników jakości ścieków, zaproponowano w pracy koncepcję systemu (rys. 7.3.13) pozwalającego identyfikować stany nietypowe na dopływie w odniesieniu do wybranych wskaźników (BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄). System ten bazuje na modelach do identyfikacji stanów wybranych wskaźników jakości ścieków (wartości obniżone, typowe i podwyższone) i modelach symulujących zwiększone dopływy do oczyszczalni ścieków.

W obliczeniach przyjęto, że 50% wszystkich wartości pomiarów zlokalizowanych w obrębie mediany stanowią wartości typowe dla dopływu do oczyszczalni dla analizowanego okresu obserwacji. Pozostałe wartości wskaźników jakości ścieków zlokalizowane poniżej mediany przyjęto za obniżone, a powyżej jako podwyższone. Z punktu widzenia eksploatacji oczyszczalni

ścieków szczególnie niebezpieczne są zdarzenia, którym towarzyszy zwiększony dopływ, w czasie którego następuje spadek wartości wskaźników, co może potencjalnie prowadzić do zaburzeń w działaniu reaktora.

W przedstawionym schemacie identyfikacja zmienności wybranych wskaźników jakości ścieków na dopływie w chwili t bazuje na wartościach natężenia przepływu $Q(t-z)$. W obliczeniach przewiduje się wykorzystanie modeli klasyfikacyjnych (dla pojedynczego wskaźnika jakości), opisanych równaniami postaci:

- identyfikator obniżonych wartości wskaźników jakości ścieków:

$$K(C(t))_{f,dolne} = f^K(Q(t-1), \dots, Q(t-z)) = \begin{cases} 1 & \text{gdy } C(t)_f < C(t)_{f,dolne} \\ 0 & \text{gdy } C(t)_f > C(t)_{f,dolne} \end{cases} \quad (7.3.13)$$

- identyfikator podwyższonych wartości wskaźników jakości ścieków:

$$K(C(t))_{f,gorne} = f^K(Q(t-1), \dots, Q(t-z)) = \begin{cases} 1 & \text{gdy } C(t)_f > C(t)_{f,gorne} \\ 0 & \text{gdy } C(t)_f < C(t)_{f,gorne} \end{cases} \quad (7.3.14)$$

Doboru zmiennych niezależnych w modelach klasyfikatorów dokonano za pomocą obliczeń testu chi-kwadrat. Pierwszy z klasyfikatorów systemu z rys. 7.3.13 pozwala na prognozę zdarzeń, gdy $C(t)_f < C(t)_{f,dolne}$ (identyfikacja minimalnych wartości) i $C(t)_{f,gorne} > C(t)_f > C(t)_{f,dolne}$ (identyfikowanie typowych wartości). Zaletą tego rozwiązania jest możliwość rozpoznawania przypadków, gdy $C(t)_f > C(t)_{f,gorne}$, co pozwala na wyznaczanie trzech zakresów zmienności wybranych wskaźników jakości ścieków. W przypadku modeli do klasyfikacji przepływu, dla każdego wskaźnika jakości ustalono mediany dopływu dla zakresów $C(t)_f < C(t)_{f,dolne}$ oraz $C(t)_{f,gorne} > C(t)_f > C(t)_{f,dolne}$ i $C(t)_f > C(t)_{f,gorne}$.

Gdy analiza danych wskazuje, że wartości Q nie odpowiadają wielkościom wywołanym intensywnymi opadami, można przyjąć kolejne proggi ustalone za pomocą analizy statystycznej szeregów czasowych. Rozwiązanie to jest kluczowe dla wydzielenia grupy zdarzeń kształtujących warunki przeciążenia hydraulicznego i niedociążenia ładunkiem zanieczyszczeń oczyszczalni ścieków. Na podstawie wartości $Q_{gr,f}$ (dopływu granicznego) wykonuje się modele klasyfikacyjne do identyfikacji dopływu w odniesieniu do każdego ujętego w analizach wskaźnika jakości ścieków. Połączenie dwóch modeli (ilość i jakość ścieków) pozwala na

klasyfikowanie ładunku poszczególnych zanieczyszczeń trafiających na obiekt, co jest istotne z punktu widzenia stabilności pracy bioreaktora. Bazując na powyższych założeniach na rys. 7.3.7÷7.3.12 naniesiono dolne i górne ograniczenie wyznaczające typową jakość ścieków na dopływie.

W przypadku poszczególnych wskaźników jakości zakresy typowe (między dolnym i górnym ograniczeniem) dla rozpatrywanych wskaźników jakości wynoszą (przypadek oczyszczalni w Sitkówce-Nowiny): $BZT_5=230\div375$ mg/l, $ChZT=691\div898$ mg/l, $N_{og}=71\div 86$ mg/l, $Z_{og}=284\div366$ mg/l, $P_{og}=6,7\div8,7$ mg/l i $N-NH_4=44,0\div65$ mg/l. Wartości median natężenia dopływu do oczyszczalni odpowiadające zmienności wybranych wskaźników jakości ścieków wynosiły: dla $C(t)_f < C(t)_{f,dolne} - 40224\div42782$ m³/d (co daje medianę 41778 m³/d); dla $C(t)_{f,górne} > C(t)_f > C(t)_{f,dolne} - 37604\div39464$ m³/d (co daje medianę 37698 m³/d); dla $C(t)_f > C(t)_{f,górne} - 36072 \div 36800$ m³/d (co daje medianę 36463 m³/d).

Dla wyznaczonych zakresów zmienności wskaźników jakości i dopływu ścieków zidentyfikowano następnie zmienne zależne i stosując test chi-kwadrat dobrano zmienne niezależne do modeli klasyfikacyjnych (tab. 7.3.7÷7.3.10). W tab. 7.3.11 zamieszczono wartości parametrów struktury modeli klasyfikacyjnych otrzymanych za pomocą metody SVM, charakteryzującej się wysokimi zdolnościami predykcyjnymi (Ossowski 2013). Z kolei, w tab. 7.3.12 zestawiono dopasowanie wyznaczonych wyników obliczeń klasyfikacji dla wybranych wskaźników jakości ścieków i natężenia dopływu do oczyszczalni ścieków.

Na podstawie wyników w tab. 7.3.7÷7.3.10 można stwierdzić, że typową jakość ścieków w odniesieniu do wybranych wskaźników opisuje mniej zmiennych niezależnych, niż w modelach do identyfikacji wartości poniżej $C(t)_{f,dolne}$. W przypadku modeli do prognozy dopływu wykonane analizy wykazały, że na poziomie istotności $\alpha = 0,05$ należy w nich uwzględnić pomiary $Q(t-1) \div Q(t-7)$. Z tab. 7.3.11 wynika, że modele do identyfikacji wartości wskaźników jakości BZT_5 , $ChZT$, N_{og} , Z_{og} , P_{og} , $N-NH_4$, uzyskane dla $C(t)_f > C(t)_{f,dolne}$, charakteryzują się zadowalającymi zdolnościami klasyfikacji, co potwierdzają obliczone wartości współczynników $SPEC=87,11\div93,03\%$, $SENS=91,50\div97,62\%$ i $R^2 = 90,47\div97,02$. Modele do identyfikacji wartości $C(t)_{f,górne} > C(t)_f > C(t)_{f,dolne}$ również dają zadowalające wyniki, bowiem $SPEC=89,30\div97,29\%$, $SENS=87,70\div95,38\%$ i $R^2=88,10\div95,83$. Analizując wartości współczynników dopasowania można stwierdzić, iż średnio statystycznie na 10 przypadków jeden z nich może zostać błędnie sklasyfikowany.

Tab. 7.3.7. Zestawienie wartości testu chi–kwadrat i prawdopodobieństw testowych p dla zmiennych opisujących wskaźniki jakości ścieków (Z_{og} , BZT_5 , N_{og}) w modelach do identyfikacji minimalnych wartości $C(t)_f$ dla oczyszczalni w Sitkówce–Nowiny.

Z_{og}			BZT_5			N_{og}		
Zmienna	chi-kw	p	Zmienna	chi-kw	p	Zmienna	chi-kw	p
Q(t-4)	32,42068	0,000014	Q(t-4)	25,34981	0,000294	Q(t-4)	46,08528	0,000000
Q(t-3)	32,25462	0,000036	Q(t-3)	26,79023	0,000364	Q(t-1)	42,66964	0,000000
Q(t-1)	23,53151	0,001377	Q(t-1)	24,51877	0,000923	Q(t-2)	42,21697	0,000001
Q(t-2)	24,84461	0,001652	Q(t-5)	18,64596	0,009372	Q(t-5)	39,74766	0,000001
Q(t-5)	21,53823	0,003050	Q(t-6)	17,47480	0,025528	Q(t-6)	40,15858	0,000003
Q(t-7)	17,81771	0,022635	Q(t-2)	16,46696	0,036164	Q(t-3)	38,00547	0,000003
Q(t-6)	16,96332	0,030493	Q(t-7)	14,78519	0,038854	Q(t-7)	33,01311	0,000061

Tab. 7.3.8. Zestawienie wartości testu F i prawdopodobieństw testowych p dla zmiennych opisujących wskaźniki jakości ścieków (Z_{og} , BZT_5 , N_{og}) w modelach do identyfikacji minimalnych wartości $C(t)_f$ dla oczyszczalni w Sitkówce–Nowiny.

ChZT			N-NH ₄			P _{og}		
Zmienna	chi-kw	p	Zmienna	chi-kw	p	Zmienna	chi-kw	p
Q(t-1)	25,25100	0,000685	Q(t-1)	43,04561	0,000000	Q(t-4)	22,95434	0,000812
Q(t-3)	23,32212	0,001498	Q(t-2)	38,10843	0,000007	Q(t-2)	26,47954	0,000869
Q(t-2)	18,83438	0,008722	Q(t-4)	31,22635	0,000023	Q(t-3)	20,32829	0,004903
Q(t-5)	19,48597	0,012466	Q(t-3)	31,97009	0,000041	Q(t-1)	19,97723	0,005619
Q(t-6)	17,45207	0,014703	Q(t-5)	27,69588	0,000250	Q(t-5)	14,77164	0,039041
Q(t-7)	18,97392	0,015000	Q(t-6)	26,47088	0,000872			
Q(t-4)	15,06211	0,019779	Q(t-7)	23,16626	0,003157			

Tab. 7.3.9. Zestawienie wartości testu chi–kwadrat i prawdopodobieństw testowych p dla zmiennych opisujących wskaźniki jakości ścieków (Z_{og} , BZT_5 , N_{og}) w modelach do identyfikacji typowych wartości $C(t)_f$ dla oczyszczalni w Sitkówce–Nowiny.

Z_{og}			BZT_5			N_{og}		
Zmienna	chi-kw	p	Zmienna	chi-kw	p	Zmienna	chi-kw	p
Q(t-1)	29,46273	0,000119	Q(t-4)	14,47017	0,024803	Q(t-1)	20,51941	0,004551
Q(t-2)	24,43654	0,000434	Q(t-1)	14,56696	0,036164	Q(t-4)	17,05255	0,009092
Q(t-3)	26,25911	0,000453	Q(t-3)	14,63351	0,040214	Q(t-2)	16,67740	0,033649
Q(t-4)	25,74203	0,001163	Q(t-2)	14,23351	0,045995	Q(t-5)	12,74766	0,037649

Tab. 7.3.10. Zestawienie wartości testu chi–kwadrat i prawdopodobieństw testowych p dla zmiennych opisujących wskaźniki jakości ścieków (ChZT, $N-NH_4$, P_{og}) w modelach do identyfikacji typowych wartości $C(t)_f$ dla oczyszczalni w Sitkówce–Nowiny.

ChZT			$N-NH_4$			P_{og}		
Zmienna	chi-kw	p	Zmienna	chi-kw	p	Zmienna	chi-kw	p
Q(t-1)	20,37353	0,005275	Q(t-2)	23,73972	0,002534	Q(t-4)	13,59467	0,002067
Q(t-2)	18,25676	0,011278	Q(t-3)	17,73359	0,013232	Q(t-2)	5,148807	0,010057
Q(t-3)	16,47903	0,021351	Q(t-1)	16,39001	0,021783	Q(t-3)	4,333345	0,020569
Q(t-6)	14,69606	0,040495	Q(t-4)	13,17944	0,040273	Q(t-1)	13,59467	0,034507

Tab. 7.3.11. Zestawienie wartości C , γ i ε dla otrzymanych modeli klasyfikacyjnych dla oczyszczalni w Sitkówce–Nowiny.

Wskaźnik	C	γ	ε	Wskaźnik	C	γ	ε
BZT_5^{sr}	30	0,32	0,1	$N - NH_4^{sr}$	25	0,75	0,1
BZT_5^{min}	25	0,36	0,1	$N - NH_4^{min}$	39	0,50	0,1
$ChZT^{sr}$	50	0,15	0,1	P_{og}^{sr}	40	0,25	0,005
$ChZT^{min}$	46	0,41	0,1	P_{og}^{min}	50	0,35	0,005
Z_{og}^{sr}	32	0,52	0,1	Q dolne	20	0,20	0,01
Z_{og}^{min}	30	0,15	0,1	Q górne	30	0,15	0,01
N_{og}^{sr}	55	0,27	0,1	Q d-g	18	0,35	0,02
N_{og}^{min}	50	0,32	0,1	Q ₁	15	0,15	0,01
				Q ₂	20	0,20	0,01

Tab. 7.3.12. Zestawienie współczynników dopasowania wyników obliczeń do pomiarów (SENS, SPEC, R_z^2) dla opracowanych modeli klasyfikacyjnych do identyfikacji zmienności wybranych wskaźników jakości ścieków dla oczyszczalni w Sitkówce–Nowiny.

Wskaźnik	Test			Walidacja		
	SENS, %	SPEC, %	R_z^2	SENS, %	SPEC, %	R_z^2
BZT ₅ ^{sr}	88,37	97,20	91,67	96,15	90,22	93,51
BZT ₅ ^{min}	91,86	91,46	96,43	93,03	98,40	97,02
ChZT ^{sr}	91,25	92,04	91,67	95,00	90,91	92,86
ChZT ^{min}	85,94	96,52	95,24	87,11	97,52	95,05
Z _{og} ^{sr}	99,18	86,67	95,83	96,75	93,33	95,83
Z _{og} ^{min}	86,67	99,18	95,83	88,89	96,75	94,65
N _{og} ^{sr}	88,06	88,79	88,11	89,30	87,70	88,10
N _{og} ^{min}	84,08	90,50	88,56	88,70	91,50	90,47
N – NH ₄ ^{sr}	88,72	91,75	90,47	92,95	90,73	91,67
N – NH ₄ ^{min}	85,62	94,59	90,47	89,47	91,89	91,07
P _{og} ^{sr}	91,89	96,69	95,80	97,29	94,65	95,24
P _{og} ^{min}	89,47	99,23	97,03	89,47	95,38	94,05
Q ^{dołne}	93,25	92,81	93,02	90,78	90,26	90,50
Q ^{górne}	95,21	94,25	94,83	94,25	95,55	95,03
Q ^{d-g}	90,23	92,25	91,56	92,51	93,14	92,82
Q ₁	95,21	97,26	96,01	96,38	97,25	97,00
Q ₂	95,32	94,78	95,12	96,54	95,52	96,01

Dla modeli do klasyfikacji przepływów uzyskano zadowalającą zgodność danych pomiarowych z obliczeniami, na co wskazują wyznaczone SPEC, SENS i R_z^2 . Analiza danych w tab. 7.3.12 wskazuje, że w oparciu o otrzymane modele, mając na uwadze ich wysokie zdolności predykcyjne, możliwa będzie identyfikacja przypadków, gdy $C(t)_f > C(t)_{f,górne}$ tj. identyfikacja wartości wskaźników jakości o podwyższonych wartościach.

Na podstawie rezultatów symulacji można stwierdzić, że w przypadku niezadowalających rezultatów modelowania wskaźników jakości ścieków, przedstawiony wyżej model klasyfikacyjny daje możliwość identyfikacji zakresów zmienności BZT₅, ChZT, N_{og}, Z_{og}, P_{og}, N-NH₄ (wartości obniżone, typowe i podwyższone) w dopływie. Jest to istotna zaleta modelu. Wyznaczone przy jego

pomocy wyniki dają możliwość równoczesnej oceny wpływu zwiększonego dopływu ścieków i zmian w wartościach wybranych wskaźników jakości ścieków na działanie oczyszczalni. Otrzymane rezultaty symulacji mogą okazać się pomocne przy podejmowaniu odpowiednich decyzji przez technologa oczyszczalni. Ponadto, model ten pozwala na ocenę skorelowania wielkości dopływu do oczyszczalni z wartościami wskaźników jakości ścieków. Przy pomocy opracowanego modelu możliwa jest także identyfikacja zdarzeń, w czasie których, mimo braku istotnej zmiany dopływu, stężenie zanieczyszczeń ulegnie zmianie, co ma istotny wpływ na przebieg procesów w bioreaktorze.

7.4. Prognoza funkcjonowania oczyszczalni ścieków

Z wykonanego przeglądu literatury (Hong i Paik 2007, Hongbin i in. 2014, Mirbagherii i in. 2015, Heddam i in. 2016) wynika, że w większości modeli procesowych do symulacji oczyszczalni ścieków zmiennymi objaśniającymi są wskaźniki jakości ścieków. Modele te charakteryzują się zadowalającą dokładnością i pozwalają na ciągłą prognozę działania bioreaktora pod warunkiem, że pomiary wszystkich zmiennych niezależnych są dostępne. W przypadku ich braku (awaria sond pomiarowych, problemy techniczne z wykonaniem oznaczeń laboratoryjnych itp.) pojawiają się ograniczenia w wykorzystaniu takich modeli. Wobec powyższego, w pracy zaproponowano algorytm tworzenia modelu (rys. 7.4.1), w którym uwzględniono dwa typowe przypadki.



Rys. 7.4.1. Schemat tworzenia modelu procesowego w przypadku nieciągłości w szeregach czasowych.

Pierwszy przypadek dotyczy rozwiązania, gdy dostępne są ciągłe wyniki pomiarów ilości i jakości ścieków, temperatury, parametrów operacyjnych determinujących działanie reaktora. Drugi przypadek dotyczy sytuacji, gdy występują nieciągłości w pomiarach pewnych zmiennych, i wówczas przyjmuje się, że zmienne te będą obliczane za pomocą modeli statystycznych, wyznaczanych zgodnie z algorytmem zamieszczonym na rys. 6.6. Takie podejście pozwala na uzyskanie wysokiej niezawodności działania obiektu w przypadku braku pomiarów zmiennych mających kluczowe znaczenie dla realizacji procesu oczyszczania ścieków i jego optymalizacji.

W warunkach eksploatacyjnych daje to możliwość ograniczenia kosztów prowadzenia ciągłych pomiarów wskaźników jakości ścieków przy pomocy analizatorów, co jest kosztowne ze względu na wysokie koszty odczynników i często uciążliwe ze względu na konieczność częstej kalibracji urządzeń.

7.4.1. Prognoza sedymentacji osadu czynnego

Z przeprowadzonych rozważań w rozdziale 2 wynika, że sedymentacja osadu czynnego stanowi istotny wskaźnik oceny działania oczyszczalni ścieków, decydujący o efektywności jej funkcjonowania. Dlatego też dotychczas, podjęto szereg prac (tab. 3.5) mających na celu opracowanie modelu do prognozy tego wskaźnika. Modele te (czarna skrzynka) w większości przypadków stosowane są do symulacji wartości liczbowych (Rossle, Pretorius 2008, Qiao i in. 2012, 2014, Bagheri i in. 2015, Szelağ i Studziński 2018, Szelağ i in. 2018b) i interpretacja uzyskanych zależności między zmienną zależną a zmiennymi niezależnymi jest ograniczona. Capodaglio i in. (1991), Côté (1995), Aarnio i Minkkinen (1986), Lou i Zhao (2012), skupiali się na prezentacji przedstawianego algorytmu obliczeniowego, nie badając przy tym, czy rozpatrywane zjawisko można modelować prostszymi metodami. Jednocześnie, podjęto badania (Belanche i in. 2000, Bayo i in. 2006), mające na celu opracowanie modeli klasyfikacyjnych do identyfikacji zdolności sedymentacyjnych osadu czynnego. Mimo licznych zalet modeli klasyfikacyjnych w odniesieniu do modeli regresyjnych, liczba prac z tego zakresu jest znacznie mniejsza niż modeli do symulacji wartości ciągłych. Jednocześnie wyznaczone modele miały ograniczone zastosowanie, co wynikało z liczby zmiennych niezależnych uwzględnianych w modelu.

W związku z powyższym w pracy rozważono możliwość implementacji licznych metod data mining do identyfikacji zdolności sedymentacyjnych osadu czynnego (klasyfikacji zdarzenia puchnięcia lub nie osadu). W wykonanych badaniach wykorzystano:

- model logitowy (pozwala na symulację zależności nieliniowych a otrzymane równanie ma charakter jawny),
- model LDA (w celu porównania z modelem logitowym),

- model k–najbliższych sąsiadów (jako najprostszy model nieparametryczny),
- modele drzew wzmacnianych BT i lasów losowych RF (modele pierwotnie stosowane wyłącznie w klasyfikacji, ponieważ w metodach RF i BT można ocenić wpływ poszczególnych zmiennych na wyniki modelowania),
- modele sieci neuronowych MLP i SVM (w celu oceny wpływu złożoności modelu na dokładność uzyskiwanych wyników w porównaniu z prostszymi metodami).

7.4.2.1. Prognoza sedymentacji osadu czynnego (modele klasyfikacyjne)

W oparciu o przesłanki literaturowe (Martins i in. 2004, Nielsen i in. 2009), jako zmienne objaśniające sedymentację osadu czynnego przyjęto wskaźniki jakości ścieków (BZT_5 , $ChZT$, N_{og} , P_{og} , Z_{og} , $N-NH_4$) i nastawy bioreaktora (X_{OC} , T_{KOC} , m_{PIX} , pH). Przy opracowywaniu modelu postępowano zgodnie z algorytmem obliczeniowym na rys. 6.6. Tym samym, w pierwszej kolejności wyznaczono współczynniki korelacji między analizowanymi zmiennymi a objętościowym indeksem osadu; wyniki obliczeń przedstawiono w tab. 7.4.1. Na ich podstawie stwierdzono, że największy wpływ w rozważanym obiekcie na sedymentację osadu ma stężenie osadu ($R = -0,61$) i temperatura w komorach osadu czynnego ($R = -0,46$), znacznie mniejszy wpływ na zdolności sedymentacyjne ma jakość ścieków (BZT_5 , $ChZT$, Z_{og} , $N-NH_4$, pH) oraz ilość dawki PIX i stężenie tlenu rozpuszczonego w komorach osadu czynnego.

Tab. 7.4.1. Wartości współczynnika korelacji między analizowanymi zmiennymi niezależnymi a objętościowym indeksem osadu dla oczyszczalni w Sitkówce–Nowiny.

Zmienne	BZT_5	$ChZT$	Z_{og}	$N-NH_4$	N_{og}	pH	T_{KOC}	X_{oc}	DO	m_{PIX}	Q
R	0,21	0,15	0,17	0,25	0,19	-0,11	-0,46	-0,61	-0,21	-0,12	-0,10
VIF	1,05	1,02	1,03	1,07	1,04	1,01	1,27	1,59	1,05	1,01	1,01

Otrzymane wartości R i VIF wskazują, że rozpatrywane zależności mają charakter nieliniowy, co determinuje dobór modelu – regresji logistycznej. Na podstawie wykonanych obliczeń otrzymano zależność postaci (Szelaąg i in. 2018d):

$$p = \frac{\exp\left(\alpha_0 + \alpha_1 \cdot X_{OC} + \alpha_2 \cdot T_{KOC} + \alpha_3 \cdot m_{PIX} + \alpha_4 \cdot DO + \alpha_5 \cdot \frac{BZT_5}{P_{og}} + \alpha_6 \cdot \frac{BZT_5}{N_{og}} + \alpha_7 \cdot L_{N-NH_4}\right)}{1 + \exp\left(\alpha_0 + \alpha_1 \cdot X_{OC} + \alpha_2 \cdot T_{KOC} + \alpha_3 \cdot m_{PIX} + \alpha_4 \cdot DO + \alpha_5 \cdot \frac{BZT_5}{P_{og}} + \alpha_6 \cdot \frac{BZT_5}{N_{og}} + \alpha_7 \cdot L_{N-NH_4}\right)} \quad (7.4.1)$$

gdzie: X_{OC} – stężenie osadu czynnego, T_{KOC} – temperatura w komorach osadu czynnego, m_{PIX} – ilość dawki PIX w ciągu doby, DO – stężenie tlenu

w komorach osadu czynnego, BZT_5 – biochemiczne zapotrzebowanie na tlen, P_{og} – fosfor ogólny, N_{og} – azot ogólny, L_{N-NH_4} – ładunek azotu amonowego.

Zestawienie uzyskanych współczynników α_i w równaniu 7.4.1 zamieszczono w tab. 7.4.2, a na rys. 7.4.2 i 7.4.3 przedstawiono graficzną interpretację uzyskanej zależności i wpływu zmienności poszczególnych zmiennych ($\Delta x_i/x_i$) na wartość p. W oparciu o ustalone zmienne niezależne opracowano na podstawie równania (7.4.1) modele do identyfikacji puchnięcia osadu; wyniki obliczeń przedstawiono w tab. 7.4.3.

Tab. 7.4.2. Zestawienie wartości wyznaczonych współczynników α_i i odchyłeń standardowych (σ_i) dla oczyszczalni w Sitkówce–Nowiny.

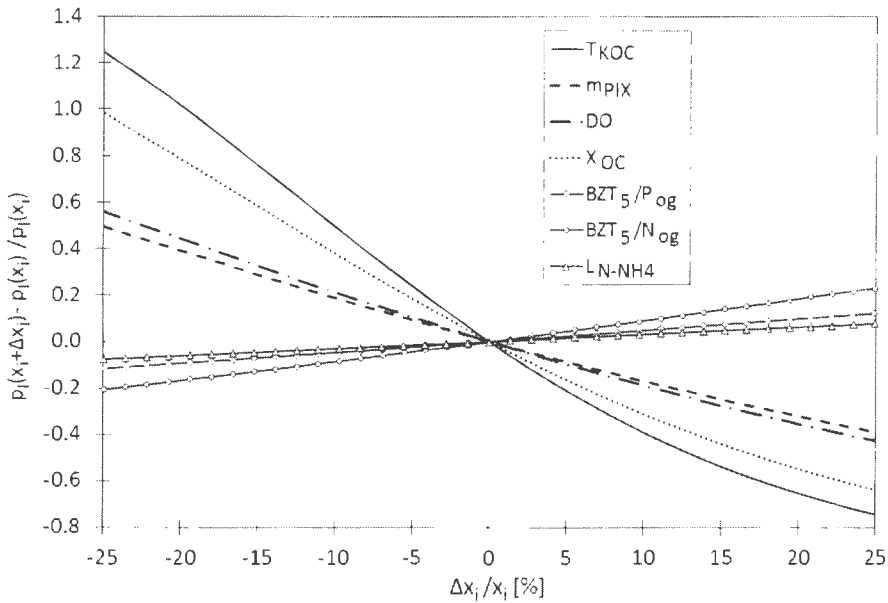
Zmienna	logit		LDA			
	α_i	odchylenie standardowe (σ_i)	$\alpha_{i,1}$	odchylenie standardowe ($\sigma_{i,1}$)	$\alpha_{i,2}$	odchylenie standardowe ($\sigma_{i,2}$)
BZT_5/N_{og}	0,019	0,003	3,951	0,568	5,431	0,592
BZT_5/P_{og}	0,32	0,032	0,225	0,001	0,173	0,011
L_{N-NH_4}	0,0009	0,00005	0,009	0,0012	0,011	0,0002
T_{KOC}	-0,368	0,233	3,394	0,161	2,732	0,138
X_{OC}	-1,365	0,39	8,964	1,046	6,921	0,703
DO	-1,183	0,272	12,463	0,404	11,316	0,324
m_{PIX}	-1,756	0,239	6,514	0,004	5,586	0,386
α_0	-14,385	1,252	-91,923	-3,478	-74,596	-2,679

Z wykonanych analiz wynika, że model logitowy charakteryzuje się zadowalającymi zdolnościami predykcyjnymi, co potwierdzają obliczone wartości parametrów dopasowania wyników do pomiarów. Wartość SPEC=90% oznacza, że na 80 przypadków, gdy $I_o > I_{o,lim}$, poprawnie zidentyfikowano 72 zdarzenia, a wartość SENS=86,7% wskazuje na to, że na 100 przypadków, gdy $I_o > 150 \text{ cm}^3/\text{g}$, poprawnie zaklasyfikowano 87 zdarzeń. Na 180 analizowanych zdarzeń przy pomocy wyznaczonego modelu poprawnie zidentyfikowano 169 przypadków. Walidacja modelu wykazała, że na 10 zdarzeń model poprawnie zaklasyfikował 8 przypadków.

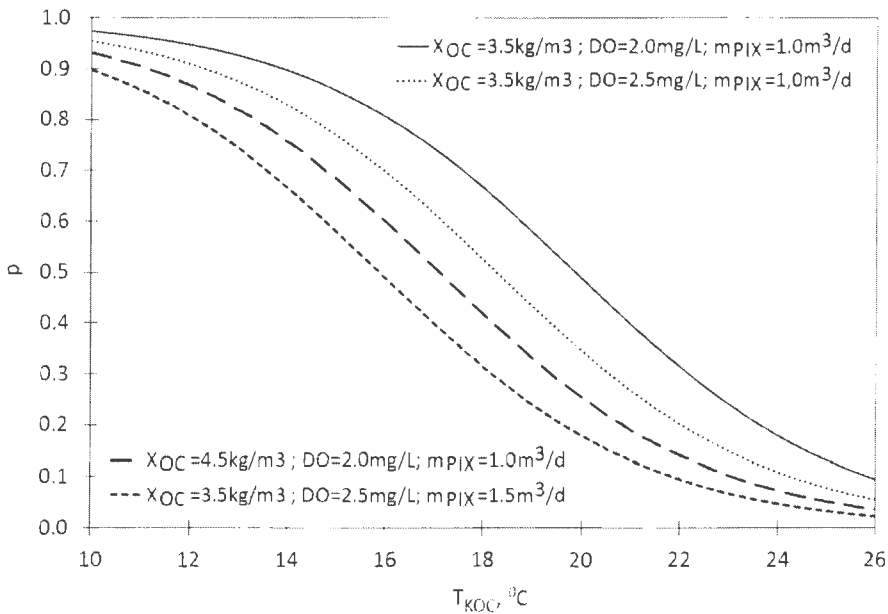
Tab. 7.4.3. Porównanie dopasowania wyników obliczeń do pomiarów wartości I_0 metodami regresji logistycznej, LDA, RF, BT, k-NN, MLP i SVM dla oczyszczalni w Sitkówce–Nowiny.

Metoda	Test			Walidacja		
	SENS	SPEC	R_z^2	SENS	SPEC	R_z^2
logit	86,7	98,0	91,2	85,23	98,10	90,02
LDA	80,3	87,3	83,3	79,86	86,12	82,10
RF	96,5	76,5	81,5	95,80	77,50	81,00
BT	96,5	97,2	96,9	96,50	97,00	96,80
k-NN	96,5	95,3	95,8	96,30	95,10	95,60
MLP	94,0	95,0	94,5	95,12	95,00	94,61
SVM	100,0	95,0	96,5	100,00	97,50	98,50

Na podstawie rys. 7.4.2 można stwierdzić, że największy wpływ na zmianę zdolności sedymentacyjnych osadu czynnego mają nastawy, tj. temperatura i stężenie osadu w komorach osadu czynnego, w następnej kolejności stężenie tlenu rozpuszczonego i ilość dawkowanego PIX (przykładowo, wzrost DO i m_{PIX} o 5% prowadzi do zmniejszenia wartości p o ok. 10%). Uzyskane wyniki znajdują potwierdzenie w badaniach wykonanych przez Bayo i in. (2006), którzy stosując również metodę regresji logistycznej wykazali wpływ temperatury w komorach osadu czynnego na jego puchnięcie. Wyznaczony przez nich model charakteryzował się zadowalającymi zdolnościami predykcyjnymi (SPEC=98%), jednak nie uwzględniono w nim nastaw reaktora, co by dawało możliwość korekty i kontroli przebiegu procesu sedymentacji. Znacznie mniejszy wpływ na sedymentację osadu czynnego ma jakość ścieków dopływających do oczyszczalni, na co wskazuje wyznaczona zmienność $\Delta p = f(BZT_5/N_{Og}, BZT_5/P_{Og}, L_N-NH_4)$. Przykładowo, wzrost wartości BZT_5/P_{Og} o 10% prowadzi do zwiększenia wartości p o ok. 7%. Podsumowując zmienności zobrazowane na rys. 7.4.2 można stwierdzić, iż znajdują one potwierdzenie w analizach wykonanych przez innych autorów (Comas i in. 2008, Rossle i Pretorius 2008, Szelağ i Siwicki 2017).



Rys. 7.4.2. Wpływ względnej zmiany wartości x_i na prawdopodobieństwo przewyższenia objętościowego indeksu osadu dla oczyszczalni w Sitkowie-Nowiny.



Rys. 7.4.3. Przykładowe krzywe $p = f(T_{KOC}, X_{OC}, DO, m_{PIX})$ dla oczyszczalni w Sitkowie-Nowiny.

Z kolei krzywe wyznaczone na rys. 7.4.3 w oparciu o równanie (7.4.1) pozwalają w prosty sposób ocenić, czy osad czynny puchnie, czy nie. Na tej podstawie można określać takie warunki pracy oczyszczalni, przy których nie będzie problemów z sedymentacją osadu czynnego.

Mimo tego, że otrzymane wyniki analiz z zakresu sedymentacji znajdują potwierdzenie w badaniach innych badaczy, w niniejszej pracy przeprowadzono weryfikację wyznaczonej zależności (7.4.1) na bazie danych dla oczyszczalni ścieków w Trepczy. Stosując program STATISTICA w oparciu o dane zgromadzone na obiekcie ustalono wartości współczynników α_i w modelu regresji logistycznej, które zamieszczono w tab. 7.4.4. Na ich podstawie można stwierdzić, że dla oczyszczalni ścieków w Trepczy otrzymano inne wartości współczynników α_i niż w modelu dla oczyszczalni w Sitkówce – Nowiny. Uzyskane wyniki obliczeń dla oczyszczalni w Trepczy wskazują, że wpływ wzrostu stężenia tlenu rozpuszczonego prowadzi do pogorszenia zdolności sedymentacyjnych osadu czynnego. Fakt, ten można wyjaśnić zastosowanym systemem napowietrzania (szczotki rotorowe), który wymaga wtłaczania znacznych ilości powietrza, aby otrzymać odpowiedni rozkład stężenia tlenu na wysokości komory. Uzyskane w pracy wyniki wskazują, że osad ulega przetlenieniu, co może prowadzić do wzrostu wartości objętościowego indeksu osadu.

Tab. 7.4.4. Zestawienie wyznaczonych wartości współczynników α_i i odchyłeń standardowych (σ_i) dla oczyszczalni ścieków w Trepczy.

Zmienna	Logit	
	α_i	odchylenie standardowe (σ_i)
BZT ₅ /N _{og}	0,221	0,005
BZT ₅ /P _{og}	0,562	0,029
L _{N-NH4}	0,0007	0,00002
T _{KOC}	-0,135	0,012
X _{OC}	-0,845	0,210
DO	0,198	0,012
m _{PIX}	0,000	0,000
α_0	-22,12	2,341

W rozpatrywanym przypadku, wartości współczynników α_i dla BZT₅/N_{og} i BZT₅/P_{og} są znacznie większe niż dla oczyszczalni w Sitkówce–Nowiny. Wyniki te mogą dowodzić, że dynamika i zmienność jakości ścieków na dopływie istotnie wpływają na zjawisko sedymentacji. W przypadku oczyszczalni w Trepczy ma to

miejsce, bowiem występują zrzuty ścieków z zakładów przemysłowych, co prowadzi do zaburzeń w działaniu osadnika wtórnego i wzrostu opadalności. Dla oczyszczalni ścieków w Trepczy, w okresie objętym obliczeniami nie dawkowano PIX, stąd też zmienna ta nie była ujmowana w modelu.

Uzyskane wyniki obliczeń dla oczyszczalni w Sitkówce–Nowiny i Trepczy wykazały, że regresja logistyczna stanowi cenne narzędzie, które na oczyszczalniach ścieków można stosować do prognozy sedymentacji osadu czynnego. Równocześnie wykazano, że zarówno jakość ścieków na dopływie i przyjęte rozwiązania technologiczne w obrębie bioreaktora mają wpływ na proces sedymentacji, a w konsekwencji na estymowane współczynniki empiryczne. Mając na uwadze otrzymane wyżej wyniki obliczeń, celowe są dalsze analizy również na innych obiektach, w celu opracowania uniwersalnego modelu matematycznego.

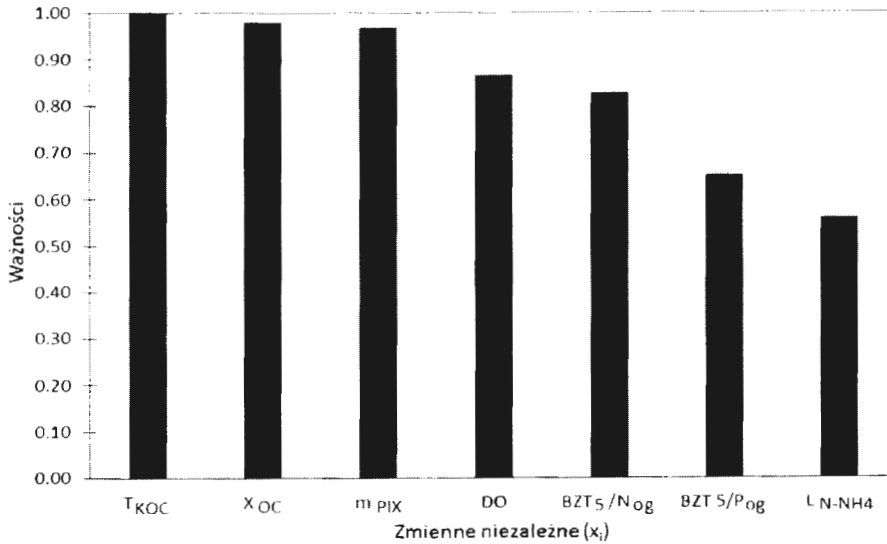
Mimo licznych zalet i zadowalającej zdolności predykcyjnej modelu logitowego zdecydowano się na opracowanie kolejnych modeli w celu ustalenia najlepszej metody do prognozy puchnięcia osadu. W następnej kolejności do identyfikacji zdolności sedymentacyjnych osadu czynnego zastosowano liniowy model dyskryminacyjny. Model otrzymany metodą LDA przy użyciu programu STATISTICA można opisać zależnością:

$$Z_L = \alpha_{0,L} + \alpha_{1,L} \cdot X_{OC} + \alpha_{2,L} \cdot T_{KOC} + \alpha_{3,L} \cdot m_{PIX} + \alpha_{4,L} \cdot DO + \\ + \alpha_{5,L} \cdot \frac{BZT_5}{P_{og}} + \alpha_{6,L} \cdot \frac{BZT_5}{N_{og}} + \alpha_{7,L} \cdot L_{N-NH_4} \quad (7.4.2)$$

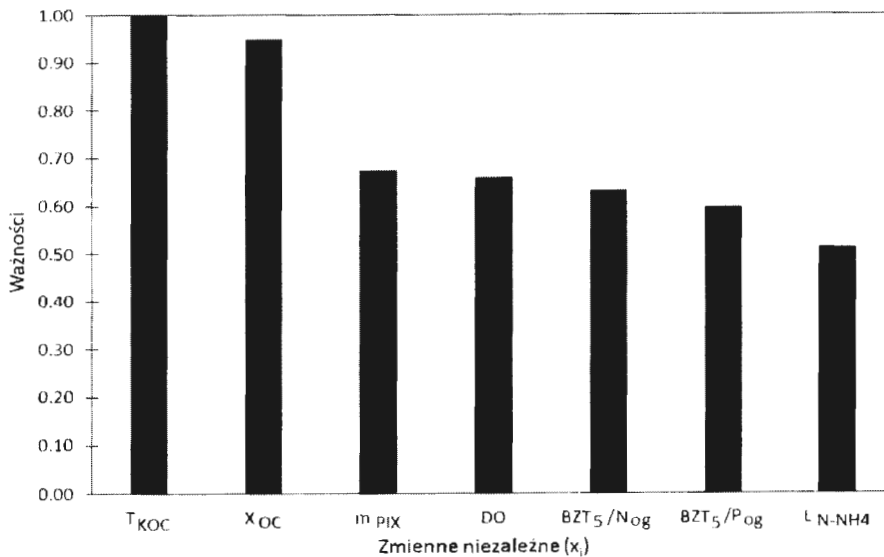
gdzie L – liczba klas przyjętych w obliczeniach, $L = 1, 2$.

Gdy z obliczeń wynika, że $Z_1 > Z_2$ dla przyjętych zmiennych x_i , to osad puchnie, w przeciwnym przypadku osad nie puchnie; wartości wyznaczonych współczynników w równaniu (7.4.2) dla $L = 1$ i $L = 2$ zamieszczono w tab. 7.4.2. Na podstawie otrzymanych wyników stwierdzono, że model statystyczny otrzymany przy pomocy LDA charakteryzuje się gorszymi zdolnościami predykcyjnymi, niż model logitowy.

Mając na uwadze powyższe wyniki w kolejnym etapie badań do prognozy puchnięcia osadu czynnego zastosowano metody pierwotnie stosowane wyłącznie w zagadnieniach klasyfikacyjnych, w których uzyskana struktura modelu ma ograniczone możliwości fizycznej interpretacji. W oparciu o zmienne niezależne wyznaczone dla modelu logitowego opracowano modele klasyfikacyjne przy pomocy metod drzew wzmocnianych i lasów losowych, na bazie założeń z rozdziału 7.1. Modele na bazie metod BT i RF zbudowane są odpowiednio ze 181 i 103 drzew, co oznacza, że nie są one przeuczone. Na rys. 7.4.4 i 7.4.5 przedstawiono wartości ważności określone dla badanych zmiennych niezależnych.



Rys. 7.4.4. Wpływ poszczególnych zmiennych niezależnych na wartość I_0 w modelu klasyfikacyjnym otrzymanym metodą RF dla oczyszczalni w Sitkówce–Nowiny.



Rys. 7.4.5. Wpływ poszczególnych zmiennych niezależnych na wartość I_0 w modelu klasyfikacyjnym otrzymanym metodą BT dla oczyszczalni w Sitkówce–Nowiny.

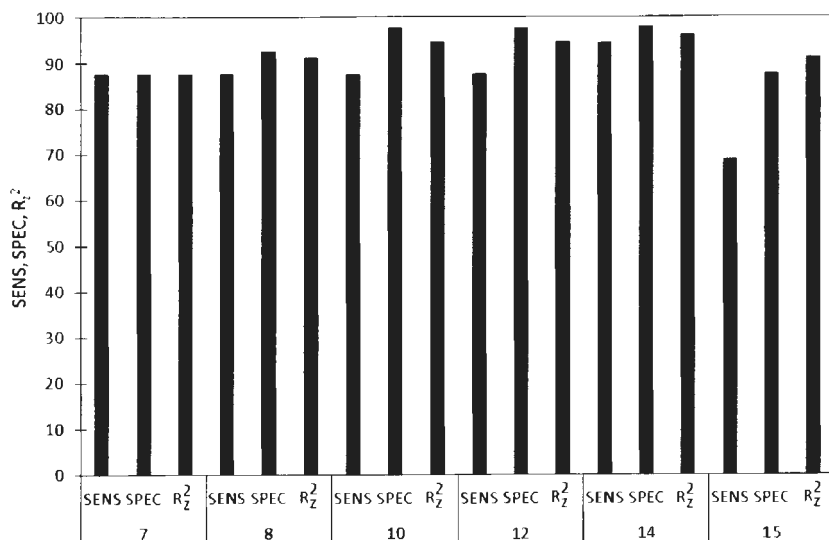
Na ich podstawie można stwierdzić, że decydujący wpływ na wartości I_0 w analizowanych modelach ma temperatura i stężenie w komorach osadu czynnego. Podobne rezultaty otrzymano dla modelu logitowego, co wynika z wykresów pokazanych na rys. 7.4.2. Przy tym model wyznaczony metodą RF charakteryzuje się gorszymi zdolnościami predykcyjnymi, niż określony na bazie metody BT i w porównaniu z innymi metodami jego wyniki obliczeń są najgorsze. Może to wskazywać na to, że model RF jest bardzo wrażliwy na zróżnicowanie wartości ważności poszczególnych zmiennych niezależnych. Dlatego też, będzie on miał ograniczone zastosowanie przy symulacji procesów opisanych przez wiele zmiennych niezależnych, gdy wartości współczynnika ważności (IMP) są zróżnicowane w szerokim zakresie (rys. 7.4.5).

Metodą z grupy metod pamięciowych implementowaną do prognozy puchnięcia osadu czynnego jest metoda k – najbliższych sąsiadów (k -NN), w której przyjmując kolejne wartości k poszukiwano optymalnej struktury modelu. Wykonane obliczenia wykazały, że najmniejsze wartości błędów (SENS, SPEC, R_z^2) uzyskano dla $k = 3$ (tab. 7.4.3). Otrzymane wyniki wskazują na zadowalające dopasowanie wyników symulacji do pomiarów i rezultaty te są porównywalne z wynikami na bazie prognoz metodą BT i z modelem logitowym. Jednak, z uwagi na to, że w wyznaczonym modelu nie ma możliwości fizikalnej interpretacji uzyskanej w nim struktury, wykonano jeszcze dodatkowe obliczenia.

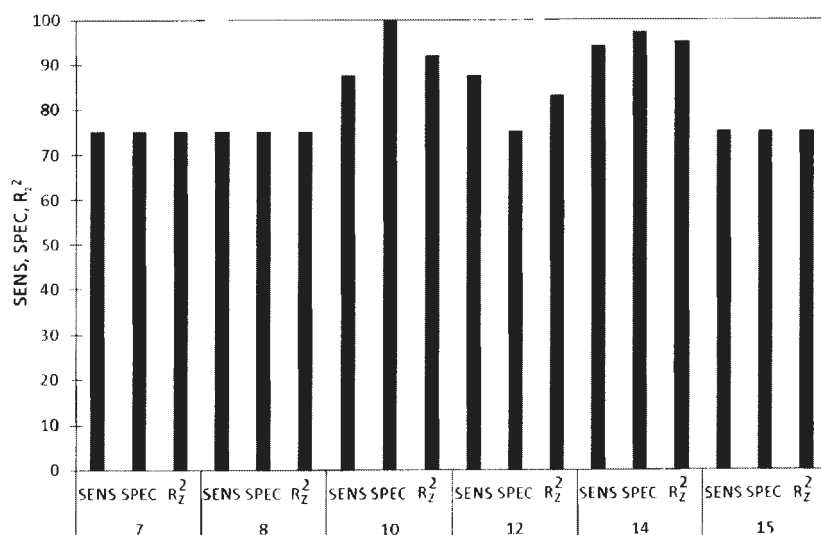
W ramach dodatkowych badań dla wartości otrzymanych z pomiarów ($BZT_5/N_{og} = 1,9 \div 6,5$, $BZT_5/P_{og} = 19,8 \div 95$, $T_{KOC} = 11 \div 22^{\circ}C$) zmieniano wartości nastaw (DO, X_{OC} , m_{PIX}) i prognozowano sedymentację osadu czynnego modelem opracowanym przy pomocy metody k -NN. Wykonane analizy wykazały, że wyznaczony model jest mało wrażliwy na zmiany wartości DO dla analizowanego zakresu zmian temperatury i jakości ścieków (wzrost stężenia tlenu rozpuszczonego powyżej 3,5mg/l prowadził do poprawy sedymentacji osadu, co nie jest prawdą, jak pokazuje eksploatacja obiektów komunalnych, a może prowadzić do przetlenienia osadu). Powyższy przykład potwierdza wadę metody k -NN (brak jawnej zależności opisującej wpływ poszczególnych zmiennych na badane zjawisko), która ogranicza stosowanie opracowanego modelu do prognozy sedymentacji osadu, mimo uzyskanych przez niego zadowalających zdolności predykcyjnych.

Kolejną z metod, którą rozważono przy symulacji objętościowego indeksu osadu za pomocą modeli klasyfikacyjnych, jest metoda sieci neuronowych typu perceptron wielowarstwowy. Bazując na ogólnych założeniach dotyczących budowy modeli parametrycznych, zdefiniowanych w rozdziale 7.1, dla rozpatrywanej liczby zmiennych niezależnych ($j = 7$) rozważano w obliczeniach dla warstwy ukrytej sieci 7÷15 neuronów i przyjęcie licznych funkcji aktywacji (funkcja liniowa, tangens hiperboliczny, wykładnicza, sinus). Wyniki modelowania

metodą MLP dla zbioru uczącego i testowego przedstawiono na rys. 7.4.6 a dla zbioru walidacyjnego na rys. 7.4.7.



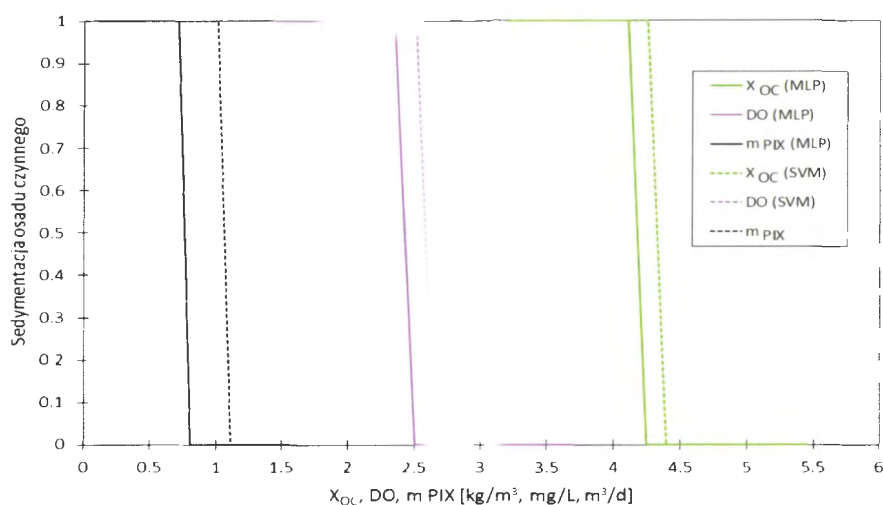
Rys. 7.4.6. Wpływ liczby neuronów w warstwie ukrytej sieci na dopasowanie wyników obliczeń do pomiarów (SENS, SPEC, R^2) dla zbioru uczącego (oczyszczalnia w Sitkówce–Nowiny).



Rys. 7.4.7. Wpływ liczby neuronów w warstwie ukrytej sieci na dopasowanie wyników obliczeń do pomiarów (SENS, SPEC, R^2) dla zbioru walidacyjnego (oczyszczalnia w Sitkówce–Nowiny).

Na podstawie otrzymanych wyników można stwierdzić, że model MLP charakteryzuje się zadowalającymi zdolnościami predykcyjnymi, a najlepsze dopasowanie otrzymano dla 10 neuronów w warstwie ukrytej. Analizując obliczone wartości współczynników (SENS, SPEC, R^2) opisujących zgodność wyników symulacji z pomiarami można zauważyć, że stosując metodę MLP uzyskuje się rezultaty minimalnie gorsze, niż uzyskane metodami BT i k-NN. Wynik ten może wskazywać na to, że struktura modelu generowanego przy pomocy metody MLP jest na tyle złożona, iż mogą pojawić się problemy z wiarygodnością wag wyznaczanych w sieci, co prowadzi do pogorszenia dokładności modelu. Jedno z możliwych rozwiązań zapobiegawczych polega na zwiększeniu wielkości zbioru danych, lub zastosowaniu odpowiednich algorytmów obliczeniowych (genetyczne, mrówkowe, roju pszczół itp.) do optymalizacji wag.

W celu oceny wpływu wyznaczonej struktury modelu na wyniki obliczeń wykonano analizę wrażliwości uwzględnianych w modelu zmiennych (rys. 7.4.8). Na podstawie otrzymanych wyników można stwierdzić, że model MLP poprawnie prognozuje wpływ wartości X_{OC} , DO i m_{PIX} na objętościowy indeks osadu czynnego, co znajduje potwierdzenie w wynikach badań laboratoryjnych (Barbusiński i Kościelniak 1995).



Rys. 7.4.8. Wyniki analizy wrażliwości w modelach MLP i SVM pokazujące wpływ zmiany wartości X_{OC} , DO i m_{PIX} na przekroczenie $I_{o,lim}$ dla oczyszczalni Sitkówka–Nowiny.

Najbardziej złożoną metodą spośród metod rozpatrywanych, którą zastosowano do symulacji objętościowego indeksu osadu (do identyfikacji zdarzeń, gdy $I_o > 150 \text{ cm}^3/\text{g}$), jest metoda wektorów nośnych. Otrzymany tą metodą model

jest bardziej złożony, niż w przypadku innych metod z uwagi na to, że trzeba w nim zdefiniować równocześnie funkcję Kernela, opisującą jej kształt współczynniki empiryczne oraz wartości parametrów ε i C . Opierając się na wynikach eksperymentów wykonanych przez licznych autorów (Ribeiro i in. 2013, Ossowski 2013, Kusiak i Wei 2014, Szelaąg i in. 2018b) w prowadzonych badaniach przyjęto funkcję gaussowską Kernela w postaci zależności:

$$K(x, x_i) = \exp(-\gamma \cdot \|x - x_i\|^2) \quad (7.4.3)$$

w której γ – współczynnik empiryczny poszukiwany metodą kolejnych podstawień dla kolejno przyjmowanych wartości C i ε (Burgess i in. 2000).

Na podstawie wykonanych obliczeń stwierdzono, że najlepsze dopasowanie wyników obliczeń do pomiarów uzyskano dla $\varepsilon = 0,01$, $C = 1$ i $\gamma = 3,0$. Dla tych wartości współczynników otrzymano zgodność klasyfikacji zdarzeń, gdy $150 \text{ cm}^3/\text{g} < I_o < 150 \text{ cm}^3/\text{g}$, nie mniejszą, niż 95%, nie mniejszą, niż 95% w odniesieniu do wartości SPEC (na 80 przypadków gdy $I_o > 150 \text{ cm}^3/\text{g}$ właściwie zidentyfikowano 80 epizodów) i SENS (na 100 przypadków gdy $I_o < 150 \text{ cm}^3/\text{g}$ właściwie zidentyfikowano 95 zdarzeń). Wskazuje to, że model ten charakteryzuje się najlepszymi zdolnościami predykcyjnymi spośród wszystkich rozpatrywanych metod. Jednak, należy zwrócić uwagę również na to, że stosując prostsze metody (RF, BT, k-NN) uzyskane wyniki modelowania nie są znacznie gorsze, aniżeli otrzymane przy pomocy metody wektorów nośnych. Poprawa dokładności prognoz indeksu objętościowego modelem SVM w porównaniu z modelem MLP może być spowodowana tym, że:

- uczenie w metodzie SVM sprowadza się do rozwiązywania zadania programowania kwadratowego, gdzie występuje minimum funkcji celu,
- model SVM jest mniej wrażliwy na ograniczoną liczbę danych uczących, niż model MLP.

Dla modelu SVM wykonano również analizę wrażliwości, podobnie jak dla modelu MLP (por. rys. 7.4.8). W rezultacie stwierdzono, że wyznaczony model klasyfikacyjny z punktu fizyki zjawiska poprawnie prognozuje wpływ zmian stężenia osadu, stężenia tlenu rozpuszczonego w komorach osadu czynnego i ilości dawkowanego PIX na przekroczenie wartości $I_o = 150 \text{ cm}^3/\text{g}$.

Podsumowując powyższe rozważania dotyczące modeli klasyfikacyjnych można stwierdzić, że najlepsze dopasowanie wyników obliczeń do pomiarów otrzymano modelem SVM, a najgorsze modelem LDA. Wyniki te potwierdziły ograniczone możliwości implementacji liniowego modelu dyskryminacyjnego do symulacji przebiegu nieliniowych zjawisk determinujących przebieg procesu sedimentacji osadu czynnego w oczyszczalni ścieków. Równocześnie wykazano, że przy identyfikacji złożonych, nieliniowych zależności procesowych oprócz

metod sztucznych sieci neuronowych (MLP, SVM) można z powodzeniem zastosować również metody, w których uzyskiwana struktura modelu jest znacznie prostsza i mniejsze są nakłady obliczeniowe. Wykonane badania potwierdziły, że poza oceną dopasowania wyników obliczeń do pomiarów opartą na współczynnikach SENS, SPEC i R_z^2 istotne znacznie odgrywa również analiza wrażliwości modeli, szczególnie wówczas, gdy badana zależność $y = f(x_i)$ ma charakter niejawny (MLP, SVM, RF, BT). Alternatywne rozwiązanie dla modeli czarnej skrzynki wyznaczanych w pracy stanowi model obliczany metodą regresji logistycznej. Charakteryzuje się on nieznacznie gorszymi zdolnościami predykcyjnymi, niż większość modeli czarnej skrzynki, ale otrzymana przy jego pomocy zależność opisująca model ma charakter jawnej zależności empirycznej (nie ma wówczas potrzeby wykonywania analizy wrażliwości), na podstawie której można badać interakcje między poszczególnymi zmiennymi niezależnymi.

7.4.2.2. Prognoza sedymentacji osadu czynnego (symulacja wartości liczbowych)

Modele klasyfikacyjne są szczególnie użyteczne przy ustalaniu, czy zachodzi zjawisko puchnięcia osadu czynnego. Jednak gdy chcemy ustalić, o ile należy zmienić wybraną nastawę w reaktorze, aby uzyskać poprawę jakości ścieków na wylocie z reaktora i zmniejszyć o zadaną wartość objętościowy indeks osadu, to wtedy zachodzi potrzeba opracowania modeli regresyjnych do symulacji wartości I_o . Z uwagi na złożony przebieg procesu sedymentacji osadu czynnego (Martins i in. 2004, Henze i in. 2002) i duże problemy z analitycznym opisem tego zjawiska, w większości dotychczasowych prac wykorzystano modele czarnej skrzynki. Koncentrowano się przy tym zwykle na metodzie sztucznych sieci neuronowych, w których otrzymana struktura modelu nie ma fizycznej interpretacji. Przy symulacji objętościowego indeksu osadu nie rozpatrywano dotychczas modeli bazujących na drzewach regresyjnych i klasyfikacyjnych, mimo tego, że były one już stosowane do symulacji pracy oczyszczalni, a uzyskane wyniki potwierdzały dobre dopasowanie modeli do pomiarów.

Równocześnie, w większości prac autorzy, może poza Comas i in. (2008) I Flores–Alsina i in. (2009), koncentrowali się tylko na zdolnościach predykcyjnych modeli i nie zajmowali się ich wykorzystaniem do doboru nastaw reaktora w celu poprawy efektywności działania oczyszczalni ścieków. Dlatego też, tylko w ograniczonym stopniu badali oni zależności między zmiennymi niezależnymi modelu a zmienną zależną. Jest to bardzo istotne zagadnienie, bowiem aby odpowiednio dobrać wartości nastawy bioreaktora wyznaczony model powinien poprawnie prognozować wpływ poszczególnych zmiennych na wartość objętościowego indeksu osadu czynnego. Aby rozwiązać ten problem, opierając się na wynikach obliczeń przedstawionych w tab. 7.4.1, zbadano w dalszym ciągu

możliwości implementacji wybranych metod data mining do symulacji objętościowego indeksu osadu, uwzględniając przy tym następujące metody:

- liniowej regresji wielorakiej (w aspekcie stopnia przydatności modelu do prognozy nieliniowych procesów),
- programowania genetycznego (w aspekcie zdefiniowania zależności jawnej),
- k- najbliższych sąsiadów,
- drzew wzmocnianych i lasów losowych (metody te są rzadko stosowane w modelowaniu oczyszczalni ścieków, a zgodnie z obecną wiedzą są to użyteczne narzędzia do symulacji wielu procesów),
- MLP, SVM (przykłady metod o bardziej złożonej strukturze).
- kaskadowych sieci neuronowych CNN; na podstawie poprzednich badań przyjęto w modelu sieci dwa dodatkowe połączenia i jako funkcje aktywacji w warstwie ukrytej i wyjściowej, odpowiednio, tangens–hiperboliczny i liniową; w warstwie ukrytej przyjęto minimalną liczbę neuronów równą 11 a liczbę maksymalną równą $(11 \cdot 2 + 1) = 23$; przyjęto przy tym identyczną liczbę neuronów w poszczególnych warstwach sieci.

Kluczowe znaczenie przy modelowaniu procesu sedymentacji osadu czynnego ma dobór zmiennych niezależnych (x_i). Bazując na wyznaczonych wartościach współczynników korelacji (R) i danych literaturowych (tab. 3.5) uwzględniono w obliczeniach zmienne opisujące zmienność związków węgla, azotu i fosforu w dopływających ściekach (BZT_5/N_{og} , BZT_5/P_{og} , L_{N-NH_4}), nastawy w bioreaktorze (DO , X_{OC}), ilość dawkowanego koagulantu (m_{PIX}) oraz temperaturę w komorach osadu czynnego (T_{KOC}), która determinuje kinetykę procesów. Dotychczasowe eksperymenty laboratoryjne i eksploatacja oczyszczalni ścieków (Henze i in. 2002, Fialkowska i in. 2010) wykazały, że dodatek koagulantów (m_{PIX}) ma wpływ na obniżenie objętościowego indeksu osadu, jednak to zjawisko nie było do tej pory badane za pomocą modelowania matematycznego.

Dlatego też, zajęto się tym zagadnieniem w obecnej pracy mając na uwadze, że korekta parametrów X_{OC} , DO i m_{PIX} w bioreaktorze będzie prowadzić do poprawy sedymentacji osadu. Przy tworzeniu modelu do prognozy wartości I_o skoncentrowano się na zmiennych, dla których wyznaczone wartości R są: największe i istotne statystycznie (por. dane w tab. 7.4.1). Stosując metodę MLR i krokowy algorytm doboru zmiennych zdefiniowano następującą zależność opisującą model:

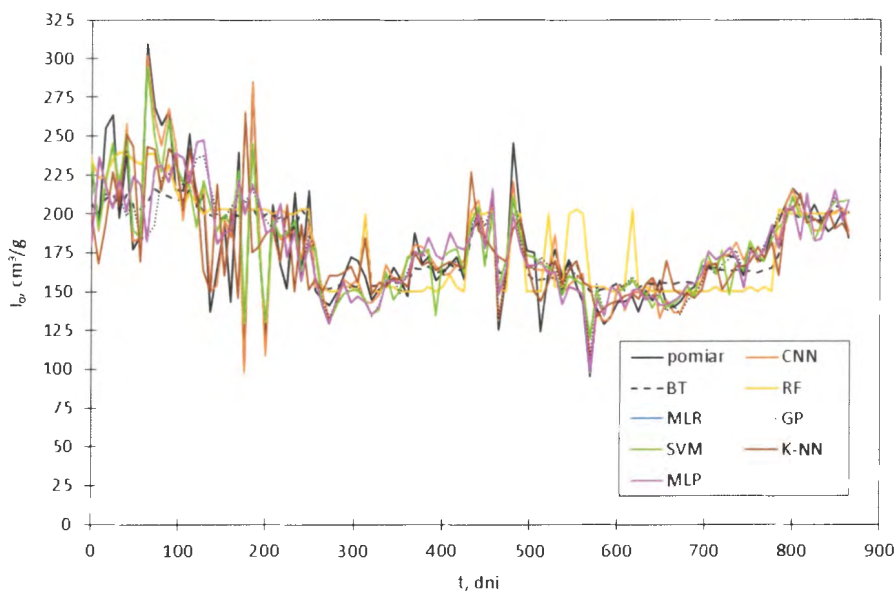
$$I_o = 338,12(\pm 16,19) - 20,37(\pm 2,33) \cdot X_{OC} - 4,24(\pm 0,77) \cdot T_{KOC} \quad (7.4.4)$$

dla której uzyskano wartość $R = 0,67$.

Mimo ograniczonych zdolności predykcyjnych modelu opisanego równaniem (7.4.4), jego opis postaci $I_o = f(X_{OC}, T_{KOC})$ jest zasadniczo prawidłowy, co potwierdzają prace Sezgin i in. (1978), Comas i in. (2001), Martins i in. (2004), Nielsen i in. (2009), Lou i Zhao (2012) oraz Szelaąg i Siwicki (2017). Jednak, uzyskana dla modelu niska wartość R wskazuje, że badany związek jest bardziej złożony i należy w nim uwzględnić dodatkowe zmienne niezależne (tab. 7.4.4). Dlatego, stosując tym razem metodę programowania genetycznego, zdefiniowano model postaci:

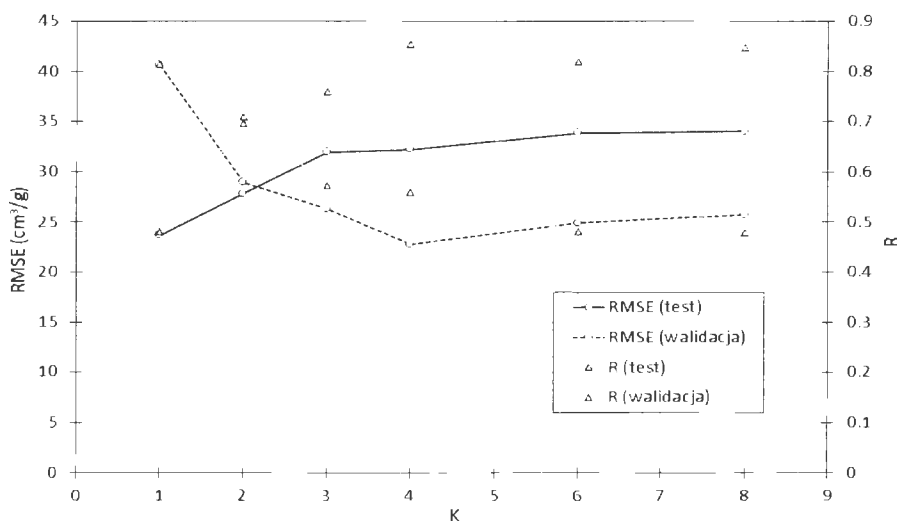
$$I_o = \frac{314,19 \cdot T_{KOC}^{-0,007}}{X_{OC}^{0,008} - 0,97} - 3,16 \cdot m_{PIX}^{0,94} \quad (R=0,70) \quad (7.4.5)$$

Zaletą wyznaczonej zależności (7.4.5) jest jej jawny charakter oraz możliwość uwzględnienia nieliniowości występujących między zmienną zależną a zmiennymi niezależnymi (Rössle i Pretorius 2008). Mimo zalet modeli opisanych równaniami (7.4.4) i (7.4.5) należy mieć na uwadze, iż otrzymane wartości R wskazują na ich ograniczone zastosowanie w zadaniach prognozy i kontroli nastaw na oczyszczalni ścieków. Potwierdzają to także wyznaczone wartości indeksu osadu dla rozpatrywanego okresu (rys. 7.4.9); maksymalne wartości I_o obliczone za pomocą metod MLR, GP są zaniżone w stosunku do pomiarów.



Rys. 7.4.9. Wyniki symulacji objętościowego indeksu osadu metodami data mining dla oczyszczalni w Sitkówce-Nowiny.

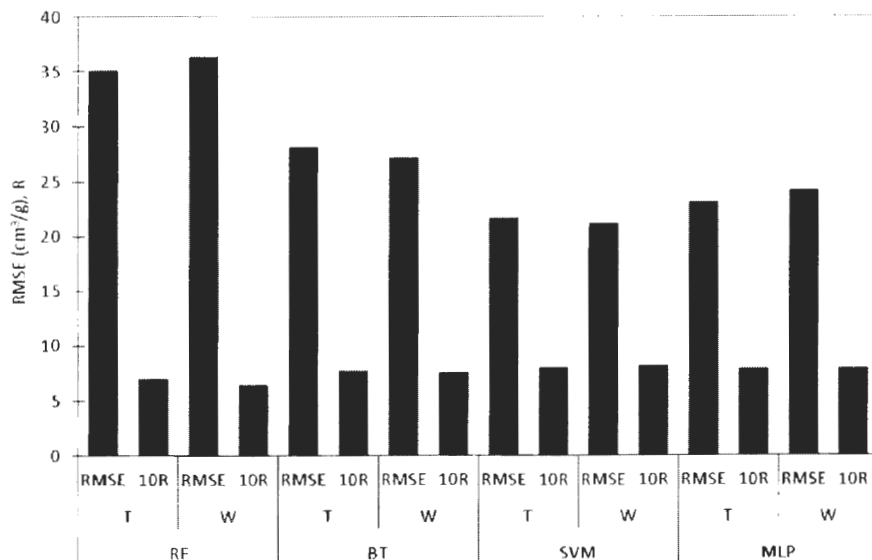
Wobec powyższego, w kolejnym etapie obliczeń badano możliwość użycia metod k-NN, RF, BT, MLP, SVM, CNN, w których zależność $y = f(x_i)$ ma charakter niejawni. Wpływ liczby sąsiadów w metodzie k-najbliższych sąsiadów na zgodność wyników obliczeń z pomiarami zamieszczono na rys. 7.4.10, natomiast na rys. 7.4.11 i 7.4.12 przedstawiono wyniki otrzymane dla metod RF, BT, MLP, SVM i CNN.



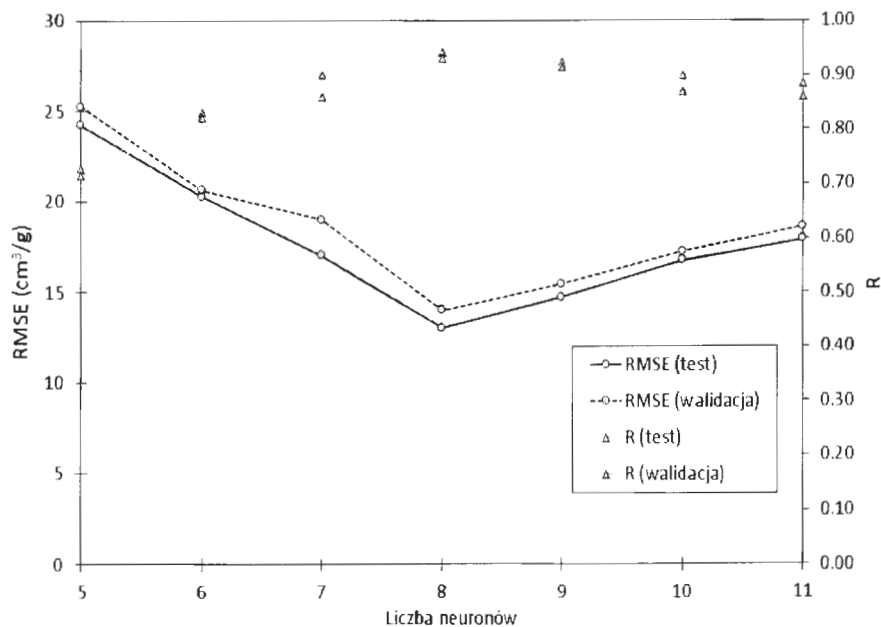
Rys. 7.4.10. Wpływ liczby sąsiadów (k) na błędy (RMSE, R) prognozy objętościowego indeksu osadu dla zbioru testowego i walidacyjnego.

Analiza rezultatów symulacji wskazuje na gorsze dopasowanie modeli k-NN, RF i BT (rys. 7.4.10, rys. 7.4.11) do pomiarów w porównaniu z modelami MLR i GP (równanie 7.4.1, 7.4.5). Na podstawie wykreślonych krzywych (rys. 7.4.9) można stwierdzić, że wartości I_0 otrzymane metodami RF i BT są niedoszacowane, na co wskazują również obliczone wartości R i RMSE (rys. 7.4.11). Oznacza to, iż przejścia między poszczególnymi „skupieniami” w strukturze modeli nie są ciągłe i mają charakter skokowy. Wyniki obliczeń otrzymane dla metody RF znajdują potwierdzenie w symulacjach prowadzonych przez Kusiaka i Wei (2014) dla oczyszczalni ścieków, którzy modelując ilość biogazu także wykazali ograniczone zastosowanie tej metody.

Model wyznaczony metodą k-NN prowadzi do niedoszacowania wartości maksymalnych i minimalnych I_0 , natomiast wartości typowe prognozowane są z zadowalającą dokładnością. W praktyce daje to możliwość zastosowania modelu do symulacji wartości średnich.



Rys. 7.4.11. Współczynniki dopasowania (R, RMSE) wyników obliczeń do pomiarów I_0 metodami RF, BT, SVM i MLP dla zbioru testowego (T) i walidacyjnego (W) dla oczyszczalni w Sitkówce–Nowiny.



Rys. 7.4.12. Wpływ liczby neuronów w warstwie ukrytej na błędy (R, RMSE) prognozy objętościowego indeksu osadu metodą CNN dla zbioru testowego i walidacyjnego.

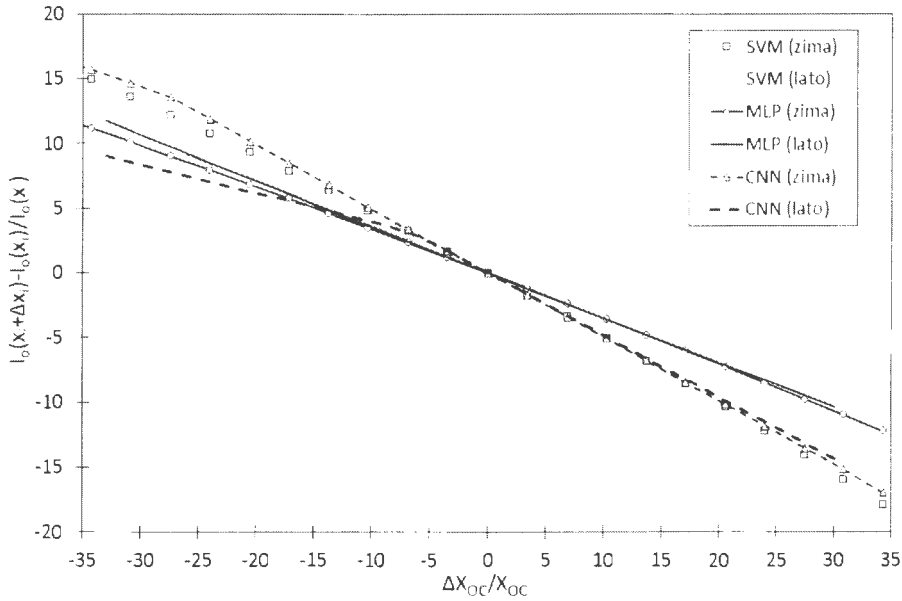
Znacznie lepszymi zdolnościami predykcyjnymi objętościowego indeksu osadu, niż modele k-NN, RF i BT, charakteryzują się modele otrzymane metodami MLP, SVM i CNN (rys. 7.4.11). W modelu MLP nie stwierdzono, dla przyjętych funkcji aktywacji, wpływu liczby neuronów w warstwie ukrytej na dokładność prognoz. W metodzie SVM, poszukując optymalnych wartości $C = 0,01 \div 10^4$, γ i $\varepsilon = 0,001 \div 0,100$, wyznaczono minimalną wartość błędu prognozy I_0 dla $C=6,00$ oraz $\gamma = 0,01$ i $\varepsilon = 0,01$. Na podstawie rys. 7.4.12 można zauważyć, że najmniejszy błąd prognozy wartości I_0 uzyskano dla 5 neuronów w pojedynczej warstwie ukrytej.

Wykonane obliczenia wykazały także, że najmniejsze wartości błędów prognozy wielkości I_0 otrzymano metodą kaskadowych sieci neuronowych CNN. Wyniki te potwierdzają wpływ dodatkowych połączeń w modelu MLP na poprawę jego zdolności predykcyjnych. Dalsza analiza rezultatów symulacji (rys. 7.4.9) wskazuje, że różnice wartości maksymalnych I_0 obliczonych metodą CNN i określone z pomiarów są mniejsze, niż w przypadku pozostałych metod. Jest to istotne z punktu widzenia doboru nastaw w reaktorze i warunkuje wysoką niezawodność działania obiektu. Biorąc pod uwagę otrzymane zmienności I_0 pokazane na rys. 7.4.11 można stwierdzić także, że w modelu wyznaczonym metodą SVM otrzymano lepsze dopasowanie obliczeń do danych pomiarowych (w odniesieniu do wartości średnich, minimalnych i maksymalnych), niż w metodzie MLP. Z uwagi na to, że w modelach opracowanych metodami MLP, SVM i CNN nie jest możliwa fizyczna interpretacja wyznaczanych parametrów modeli, to zgodnie z zaproponowanym algorytmem doboru metody (rys. 6.6) wykonano analizę wrażliwości modeli, której wyniki przedstawiono na rys. 7.4.13.

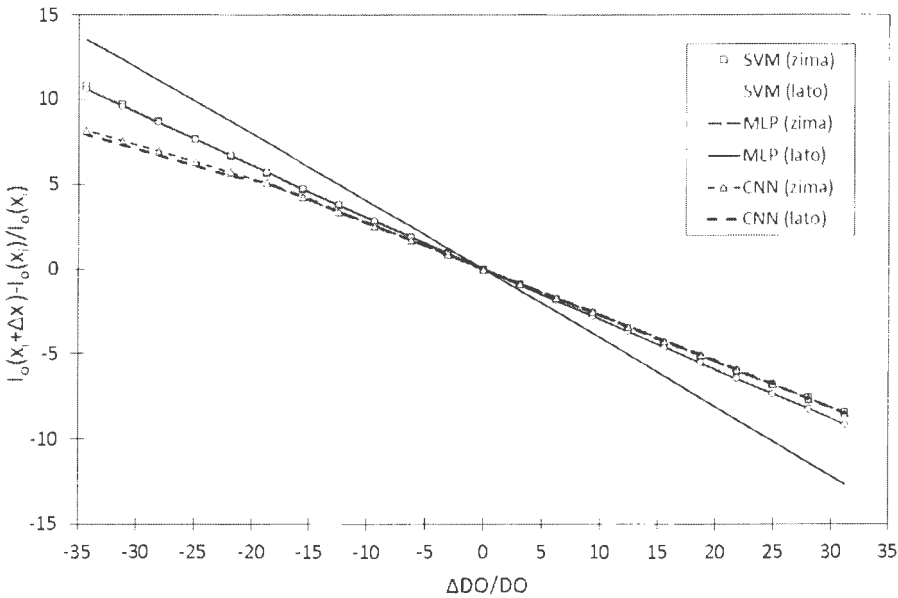
Wyniki te wskazują, że modele SVM i MLP prawidłowo odzwierciedlają wpływ wartości X_{OC} , DO i m_{PIX} na objętościowy indeks osadu. Jednocześnie w modelu MLP obserwuje się mniejszy wpływ zmiennej m_{PIX} na wyniki, niż w modelach CNN, SVM czy również w modelu regresji logistycznej (rys. 7.4.3). Wyniki modelowania wskazują także na zróżnicowany wpływ poszczególnych nastaw reaktora na względną zmianę wartość I_0 w zależności od pory roku w modelach SVM, MLP i CNN. Związek ten wydaje się logiczny z uwagi na to, że w okresie letnim w komorach osadu czynnego są niższe wartości X_{OC} , niż w okresie zimy, i w celu uzyskania zadanej wartości I_0 względna korekta X_{OC} będzie mniejsza, niż w okresie obniżonych temperatur powietrza. Z powyższych rozważań wynika, że modele statystyczne do prognozy objętościowego indeksu osadu wyznaczone metodami CNN, SVM i MLP charakteryzują się zadowalającymi zdolnościami prognozy i potencjalnie mogą być zaimplementowane w oczyszczalni ścieków. Jednak, podobnie jak większość opracowanych do tej pory modeli, mają one podstawową wadę, mianowicie w przypadku braku pomiarów wskaźników jakości ścieków w nich uwzględnionych ich zastosowanie do symulacji badanego procesu jest

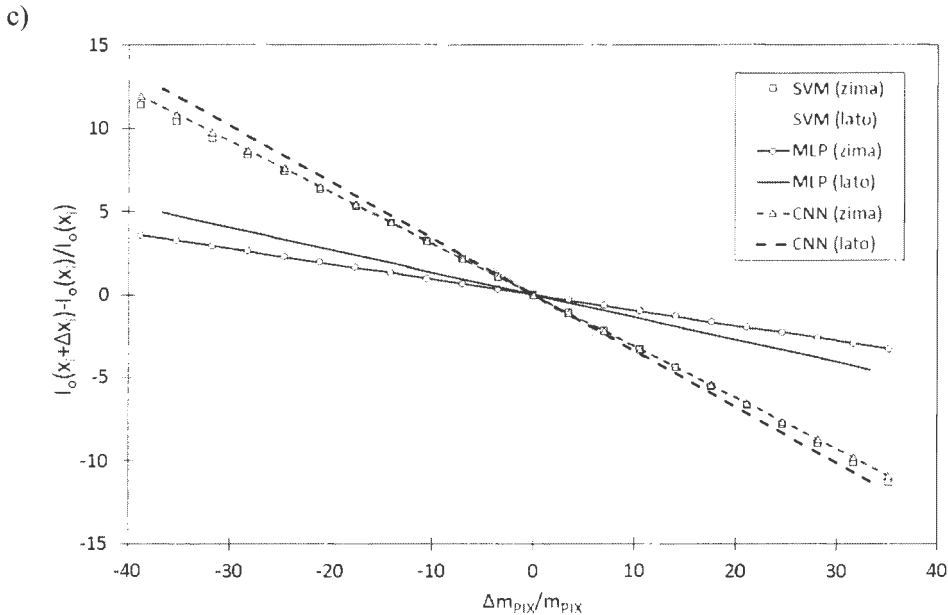
ograniczone. Dlatego też, w dalszych badaniach zastosowano koncepcję modelowania przedstawioną w rozdz. 7.4, polegającą na zastąpieniu mierzonych wartości wskaźników jakości ścieków wartościami obliczonymi (Szeląg i in. 2017, 2018).

a)



b)





Rys. 7.4.13. Wrażliwość modeli do prognozy I_0 , wyznaczonych metodami CNN, SVM i MLP, na zmiany wartości nastaw (a) X_{OC} , (b) DO, (c) m_{PIX} .

W kolejnych analizach badano implementację metod CNN, SVM, MLP wykorzystanych do obliczania wybranych wskaźników jakości ścieków (BZT_5 , N_{og} , P_{og} , $N-NH_4$; rozdział 7.3), uwzględnionych w równaniu (7.3.1), i następnie zastosowanie modelu CNN na bazie wartości ujętych w zależności (7.4.1) do obliczania wartości I_0 (Szelağ i in. 2018b).

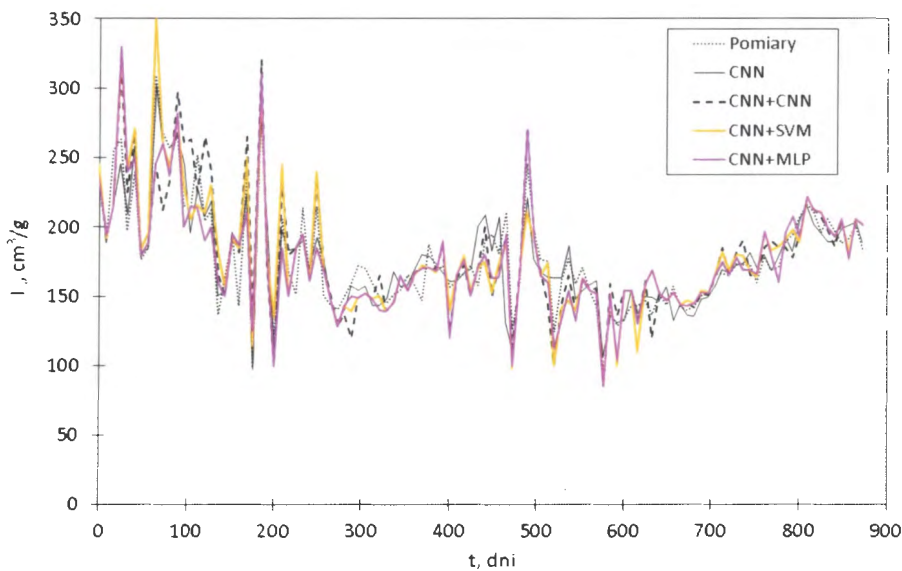
Uzyskane wyniki przedstawiono w tab. 7.4.5 i na rys. 7.4.14. W tab. 7.4.5 znajdują się wartości obliczonych współczynników dopasowania wyników obliczeń do pomiarów, przy czym: CNN+CNN oznacza, że do prognozy $I_0 = f(BZT_5, N_{og}, P_{og}, N-NH_4)$ i wskaźników jakości ścieków zastosowano metodę CNN; CNN+SVM oznacza, że do prognozy I_0 przyjęto metodę CNN, a do obliczeń wskaźników jakości ścieków założono metodę SVM; CNN+MLP oznacza, że do prognozy I_0 zastosowano metodę CNN, natomiast wskaźniki jakości wyznaczono metodą MLP). Na rys. 7.4.14 przedstawiono porównanie wyników obliczeń wartości I_0 otrzymanych bezpośrednio metodą CNN (wartości wskaźników jakości ścieków pochodziły z pomiarów) i ustalonych za pomocą modelowania złożonego (wartości wskaźników jakości ścieków pochodziły z obliczeń).

Na podstawie wyznaczonych krzywych i danych w tab. 7.4.5 można stwierdzić, że modele do prognozy I_0 wyznaczone w sposób kombinowany, gdy

zmienna I_0 była określona za pomocą metody CNN, a wskaźniki jakości ścieków były obliczane za pomocą metod CNN, SVM i MLP, charakteryzując się zadowalającymi zdolnościami predykcyjnymi. W praktyce oznacza to możliwość ich zastosowania do modelowania pracy oczyszczalni ścieków w celu poprawy efektywności jej pracy.

Tab. 7.4.5. Porównanie wartości współczynników (MAE, RMSE, MAPE, R) dopasowania wyników obliczeń I_0 do pomiarów wyznaczonych modelami złożonymi dla oczyszczalni w Sitkówce–Nowiny.

Metoda	MAE, cm ³ /g	RMSE, cm ³ /g	MAPE, %	R -
CNN+CNN	15,72	21,06	9,02	0,874
CNN+SVM	14,70	19,03	8,41	0,900
CNN+MLP	14,43	18,96	8,18	0,881

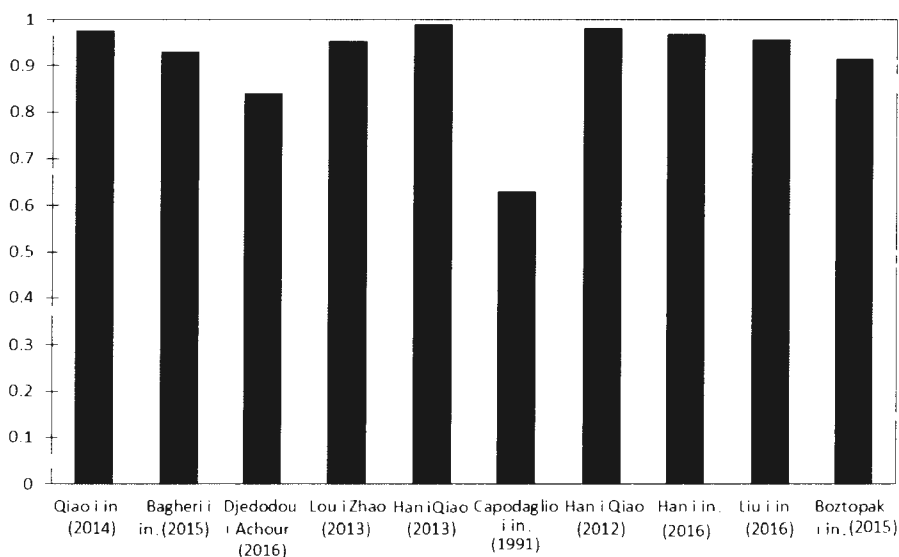


Rys. 7.4.14. Porównanie wyników obliczeń objętościowego indeksu osadu modelami złożonymi dla oczyszczalni w Sitkówce–Nowiny.

Analiza danych w tab. 7.4.5 wykazała, że otrzymane wartości miar dopasowania wyników obliczeń do pomiarów wykazują nieznaczne zróżnicowanie, dlatego nie można jednoznacznie stwierdzić, który model jest najlepszy. Mając to na uwadze wykonano test ANOVA, który wykazał, że średnie wartości błędów prognozy wartości I_0 nie wykazują istotnych różnic statystycznych. Ostateczna

konkluzja jest taka, że zastępując w modelu procesowym wartości zmiennych niezależnych (BZT_5/N_{og} , BZT_5/P_{og} i $N-NH_4$) wynikami obliczeń, następuje spadek wartości współczynników dopasowania wyników symulacji do pomiarów, w porównaniu z modelowaniem bezpośrednim na podstawie wyłącznie pomiarów.

Kolejny wniosek wynikający z wykonanych obliczeń jest taki, że najlepszymi zdolnościami predykcyjnymi indeksu osadu czynnego charakteryzuje się model bazujący na metodzie CNN (wartości wskaźników jakości ścieków uzyskane z pomiarów), a najgorszymi model wyznaczony metodą lasów losowych. Zaletą modelu CNN w porównaniu z modelami opracowanymi przez innych autorów (tab. 3.5) jest jego względnie prosta struktura, szczególnie w porównaniu z modelami zaproponowanymi przez Han i Qiao (2012, 2014, 2016). Przykładowe wyniki dotyczące modelowania procesu sedymentacji osadu czynnego, referowane w literaturze, są pokazane na rys. 7.4.15.



Rys. 7.4.15. Porównanie zdolności predykcyjnej modeli regresyjnych do symulacji wartości I_0 na bazie modelu CNN+CNN z wynikami uzyskanymi przez innych badaczy.

Należy tutaj zauważyć, że aparat obliczeniowy, który stoi za modelami badanymi przez Qiao i Hana (2012, 2013, 2014, 2016) (hierarchiczne sieci neuronowe, hybrydowe modele na bazie logiki rozmytej, metody grupowania danych), jest bardzo wyrafinowany i w rezultacie modele o badanej przez nich strukturze prognozują zależności silnie nieliniowe. Model o zdolnościach predykcyjnych porównywalnych z modelem wyznaczonym w tej pracy zaproponowali Boztoprak et al. (2015), którzy wykorzystali w obliczeniach wyniki z obserwacji kłaczek

osadu czynnego wykonane za pomocą kamer o wysokiej rozdzielczości. Zaproponowane rozwiązanie wydaje się użyteczne z uwagi na to, że nie zachodzi wówczas potrzeba wykonywania kosztownych oznaczeń jakości ścieków i pomiarów w reaktorze (DO, X_{OC} , pH, T_{KOC} , itp.). Na podstawie pomiarów jakości ścieków ($ChZT$, N_{og} , Z_{og}) i nastaw reaktora (pH, DO, T_{KOC} , X_{OCV}) Bagheri i in. (2015), stosując metodę sztucznych sieci neuronowych MLP, opracowali model o gorszych zdolnościach predykcyjnych, niż przedstawiony w pracy. Model MLP do symulacji I_o w oparciu na wyniki pomiarów zastosowali także Zhao i Luo (2012), którzy na bazie ciągłych wartości ilości i jakości ścieków oraz nastaw reaktora z okresu 9 miesięcy uzyskali model o zadowalających zdolnościach predykcyjnych. Ciekawe podejście do rozwiązania omawianego problemu przedstawili z kolei Liu i in. (2016), którzy w oparciu o ciągłe wyniki pomiarów jakości ścieków i nastaw reaktora pracującego w systemie Caroussel, za pomocą metody GPR (gaussowskie procesy regresyjne) opracowali model do prognozy I_o z wyprzedzeniem czasowym 7 dni, uzyskując wartości $R > 0,957$. Model MLP do symulacji I_o w oparciu o wyniki pomiarów jakości ścieków i nastaw reaktora (F/M) wyznaczyli również Capodaglio in. (1991), przy czym otrzymane przez nich wyniki nie były zadowalające, ponieważ nie uwzględnili oni w modelu zmiennej T_{KOC} determinującej kinetykę sedimentacji kłaczków w bioreaktorze. Djedoduo i Archour (2016), podobnie jak Capodaglio i in. (1991) oraz Zhao i Luo (2012), zastosowali metodę MLP do prognozy wartości I_o , przy czym zamiast nastaw reaktora uwzględniali oni w modelu efektywność redukcji zanieczyszczeń. Otrzymany w ten sposób model charakteryzował się gorszymi zdolnościami obliczeniowymi, niż modele, w których uwzględniano X_{OC} , DO, pH i inne zmienne procesowe. Ciekawą obserwacją z podanego wyżej przeglądu wyników modelowania jest to, że stosując podobne metody i uwzględniając w modelach podobne zmienne procesowe liczni autorzy otrzymują zwykle różne wyniki symulacji. Wydaje się, że decyduje o tym wielkość stosowanego w obliczeniach zbioru danych i zakres zmienności predyktorów modelu.

Omawiając uzyskane w pracy rezultaty modelowania objętościowego indeksu osadu przy użyciu autorskiej koncepcji wyznaczania modeli, polegającej na zastąpieniu w modelu wartości wskaźników jakości ścieków wynikami ich obliczeń, należy zaznaczyć, że otrzymane tym sposobem modele charakteryzują się na ogół gorszym dopasowaniem do pomiarów, niż w przypadku modeli wyznaczanych przez innych autorów (tab. 3.5, rys. 7.4.15). Mimo tego, mogą one być wykorzystane w sytuacjach, gdy występują nieciągłości w pomiarach wskaźników determinujących modelowany proces, co często się zdarza w praktyce eksploatacyjnej oczyszczalni ścieków.

Analizując uzyskane wyniki modelowania objętościowego indeksu osadu czynnego można stwierdzić, że proces ten można z zadowalającą dokładnością zidentyfikować w oparciu o wyniki pomiarów ilości i jakości ścieków (BZT_5/N_{og} ,

BZT₅/P_{og}, N-NH₄), parametry działania reaktora (temperatura osadu czynnego, stężenie tlenu rozpuszczonego w komorze nityfikacji, stężenie osadu czynnego, ilość dawkowanego PIX), stosując do tego celu modele klasyfikacyjne i regresyjne. W przypadku modeli klasyfikacyjnych najlepsze wyniki otrzymano metodą SVM. Zadowolające wyniki uzyskano również modelem regresji logistycznej. Jego zaletą jest to, iż ma on postać jawnej zależności, na podstawie której można ocenić wpływ poszczególnych zmiennych niezależnych na wynik modelowania bez konieczności wykonywania dodatkowych analiz. Jednocześnie w przypadku modeli czarnej skrzynki, klasyfikacyjnych i regresyjnych, istotne znaczenie na etapie weryfikacji poprawności modelu ma analiza wrażliwości, gdyż pozwala ona ocenić charakter zależności w modelu między zmienną zależną a zmiennymi niezależnymi. Analiza wrażliwości powinna stanowić jedno z podstawowych kryteriów doboru właściwej metody do modelowania procesów w oczyszczalniach ścieków.

7.4.3. Prognoza jakości ścieków na odpływie

Podstawowym kryterium oceny efektywności działania oczyszczalni ścieków jest uzyskanie wymaganej jakości ścieków na wylocie z obiektu. Wskaźniki jakości ścieków, których wartości dopuszczalne są regulowane aktami prawnymi, omówiono w rozdz. 1, przy czym odnosząc się do ich wpływu na środowisko wodne odbiorników należy zwrócić szczególnie uwagę na związki azotu, w tym również azot amonowy, oraz związki fosforu. Wzrost zawartości N i P w ciekach wodnych prowadzi do ich eutrofizacji. Z drugiej strony, zwiększona zawartość azotu amonowego jest szkodliwa i niebezpieczna dla żyjących w nich organizmów.

Aby nie zostały przekroczone dopuszczalne wartości wybranych wskaźników jakości ścieków, opracowano specjalne układy sterowania nastawami reaktora biologicznego. Prawidłowe prowadzenie procesu oczyszczania ścieków wymaga ciągłej kontroli i regulacji nastaw reaktora (stopień recyrkulacji, stężenie tlenu rozpuszczonego, ilość dawkowanych reagentów chemicznych, stężenie osadu czynnego itp.). Aby eksploatacja oczyszczalni ścieków była racjonalna, wartości parametrów charakteryzujących pracę reaktora powinny być ustalane w taki sposób, aby ograniczyć koszty eksploatacji, zachować odpowiednią jakość ścieków oczyszczonych i nie pogorszyć stanu wód odbiornika. Trzeba też zwrócić uwagę na potrzebę uzyskania wysokiej niezawodności działania obiektu w przypadku awarii analizatorów jakości ścieków na dopływie do oczyszczalni.

Wartości wskaźników jakości ścieków (Qiao i in. 2011, Han i in. 2016) stanowią zmienne objaśniające procesy przemian biochemicznych zachodzących na oczyszczalni. Ich nieznanie w znacznym stopniu ogranicza kontrolę działania reaktora (Chen 2003, Clara 2008, Luo 2009, Mirbagherii i in. 2015). Aby temu zapobiec, zachodzi potrzeba tworzenia odpowiednich modeli matematycznych. Modele pozwalają na ilościową ocenę wpływu zmiennych

niezależnych (ilość i jakość ścieków, parametry operacyjne reaktora) na zmienną zależną (wartość biochemicznego zapotrzebowania na tlen, stężenie azotu ogólnego i fosforu ogólnego na wylocie itp.), nawet w przypadku braku pomiarów w ciągach pomiarowych dotyczących wskaźników jakości ścieków.

Obecnie do modelowania oczyszczalni stosuje się modele fizykalne (bazujące na układach równań różniczkowych) i statystyczne (szarej i czarnej skrzynki). Jednak z uwagi na to, że kalibracja modeli fizykalnych jest złożona, a uzyskany wynik nie zawsze jest zadowalający, szerokim zainteresowaniem cieszą się modele statystyczne. O ile w modelu fizykalnym i szarej skrzynki można ustalić wpływ poszczególnych zmiennych niezależnych na wyniki symulacji, to w przypadku modeli czarnej skrzynki jest to niemożliwe i zachodzi potrzeba wykonywania dodatkowych analiz.

W związku z powyższym badano w pracy modele statystyczne do symulacji wybranych wskaźników jakości ścieków na wylocie z oczyszczalni ścieków dla przypadków, gdy występują nieciągłości w szeregach pomiarowych tych zmiennych. Przy tym wykonywano analizę wrażliwości opracowanych modeli. Na podstawie wniosków wynikających obliczeń modelowania wykonanych i omówionych powyżej (dotyczących prognozy ilości i jakości ścieków na dopływie oraz sedimentacji osadu), do prognozy wskaźników jakości ścieków na odpływie z oczyszczalni ($N-NH_4$, N_{og} , P_{og}) wybrano metody:

- kaskadowych sieci neuronowych (CNN) ,
- wektorów nośnych (SVM),
- drzew wzmocnianych (BT) (w celu oceny wpływu złożoności zastosowanego modelu na wyniki symulacji).

Za koniecznością tworzenia modeli do prognozy N_{og} i P_{og} przemawia fakt, że na przykład na oczyszczalni ścieków Sitkówka–Nowiny występują przekroczenia dopuszczalnych wartości N_{og} i P_{og} na odpływie w okresie obniżonych temperatur (tab. 5.1).

Stosując algorytm przedstawiony na rys. 6.6, na podstawie pomiarów zgromadzonych na oczyszczalni Sitkówka–Nowiny, danych literaturowych (Chen 2003, Raduly i in. 2007, Clara 2008, Luo i in. 2009, Mirbagherii i in. 2015, Han i in. 2016) i wyników wykonanych analiz statystycznych (test Fischera–Snedecora), ustalono zmienne niezależne opisujące wartości wskaźników jakości ścieków na odpływie. Wyniki obliczeń testu Fischera–Snedecora dla analizowanych w pracy wskaźników jakości ścieków zamieszczono w tab. 7.4.6 i tab. 7.4.7.

Na podstawie tych wyników można stwierdzić, że na usuwanie związków azotu (N_{og} , $N-NH_4$) i fosforu (P_{og}) ma wpływ ładunek dopływających do oczyszczalni związków organicznych, niezbędnych jako pożywienie dla bakterii,

i czas zatrzymania mikroorganizmów w reaktorze biologicznym, które determinują nastawy bioreaktora. Powyższe spostrzeżenia znajdują potwierdzenie w praktyce eksploatacyjnej obiektów i w literaturze z zakresu modelowania bioreaktorów przy pomocy modeli fizykalnych oraz statystycznych (Henze i in. 2002, Wan i in. 2000, Manu i Thalla 2003, Mąkinia 2010, Poutiainen i in. 2010, Gernaey i in. 2014).

Tab. 7.4.6. Zestawienie wyników obliczeń testu Fischera–Snedecora dla wskaźników jakości ścieków (N_{og} , $N-NH_4$) dla oczyszczalni w Sitkówce–Nowiny.

N-NH ₄			N _{og}		
Zmienna	F	p	Zmienna	F	p
T _{KOC}	10,017	0,000000	T _{KOC}	7,217	0,000000
DO	9,149	0,000000	DO	7,949	0,000000
N _{og}	7,863	0,000060	N _{og}	4,668	0,000020
N-NH ₄	6,448	0,000520	N-NH ₄	3,668	0,000320
BZT ₅	5,326	0,001237	BZT ₅	3,226	0,013308
X _{OC}	5,008	0,003414	X _{OC}	2,508	0,014173
REC	4,507	0,004083	REC	2,307	0,013308
M _{nadm}	4,234	0,008742	M _{nadm}	2,234	0,002742

Tab. 7.4.7. Zestawienie wyników obliczeń testu Fischera–Snedecora dla wskaźników jakości ścieków (P_{og}) dla oczyszczalni w Sitkówce–Nowiny.

P _{og}		
Zmienna	F	p
T _{KOC}	4,575	0,000039
DO	3,342	0,000630
N _{og}	2,542	0,015462
N-NH ₄	2,423	0,016253
BZT ₅	2,866	0,011803
X _{OC}	5,387	0,000004
REC	1,986	0,022453
M _{nadm}	2,332	0,020491
ChZT	2,631	0,013222
P _{og}	2,945	0,000926
m _{PIX}	2,345	0,018370

W oparciu o uzyskane wyniki analiz (tab. 7.4.6 ÷ 7.4.7) opracowano modele do prognozy wybranych wskaźników jakości ścieków na odpływie z oczyszczalni, stosując metody CNN, SVM i BT oraz opis modelu postaci:

$$C(t)_{odp} = f(C(t)_{dop}, x_1, x_2, x_3, \dots, x_i) \quad (7.4.6)$$

gdzie: $C(t)_{dop}$ – wartości wskaźników jakości ścieków na dopływie do oczyszczalni mierzone lub wyznaczone z równania (7.3.1); $C(t)_{odp}$ – wartości wskaźników jakości ścieków na odpływie z oczyszczalni ścieków; $x_1, x_2, x_3, \dots, x_i$ – wartości nastaw bioreaktora: stężenie osadu, stopień recyrkulacji, ilość dawkowanego koagulanta, stężenie tlenu w komorze nityfikacji itp.

Optymalizację parametrów w modelach statystycznych wykonano zgodnie z zasadami opisanymi w rozdziale 7.1. Wyznaczone wartości współczynników opisujących strukturę modeli opracowanych metodami CNN, SVM i BT dla wybranych wskaźników jakości ścieków przedstawiono w tab. 7.4.8, a w tab. 7.4.9 podano wartości współczynników dopasowania wyników obliczeń do pomiarów (MAE, RMSE, MAPE, R).

Tab. 7.4.8. Zestawienie wartości współczynników opisujących strukturę modeli do prognozy wybranych wskaźników jakości ścieków przy pomocy metod SVM, CNN i BT dla oczyszczalni w Sitkówce–Nowiny.

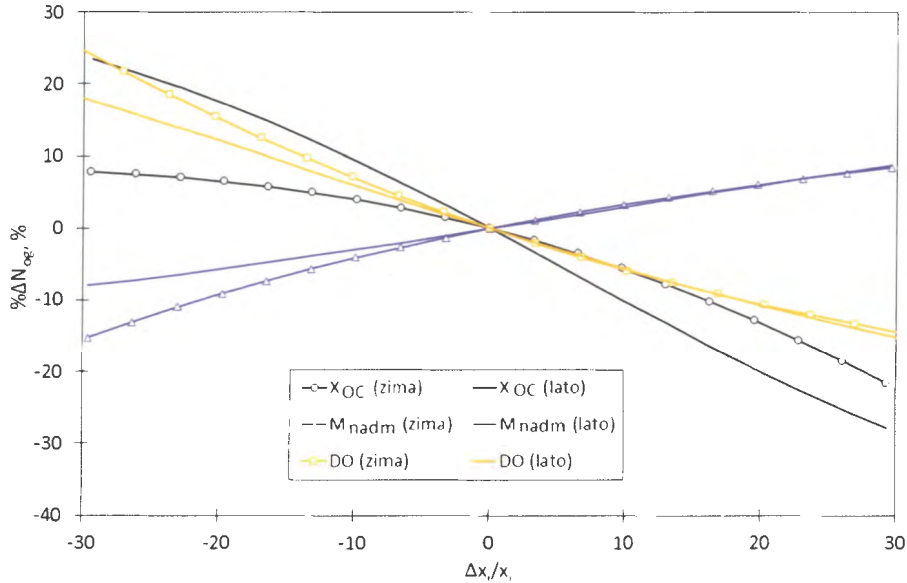
Wskaźnik	SVM			CNN	BT
	C	ϵ	γ	Liczba neuronów	Liczba drzew
N_{og}	100	0,01	0,35	14	55
N-NH ₄	80	0,10	0,67	16	60
P_{og}	300	0,03	0,15	18	72

Na podstawie wyników zamieszczonych w tab. 7.4.9 można stwierdzić, że mniejsze wartości błędów prognozy wybranych wskaźników jakości ścieków uzyskano metodami SVM i CNN, niż metodą BT. Otrzymane rezultaty analiz potwierdzają, że w przypadku ograniczonej ilości danych uczących metody o bardziej złożonej strukturze z większą ilością współczynników do kalibracji (CNN lub SVM) dają możliwość lepszego dopasowania wyników obliczeń do pomiarów, niż modele o prostszej strukturze, jak na przykład metoda drzew wzmocnianych.

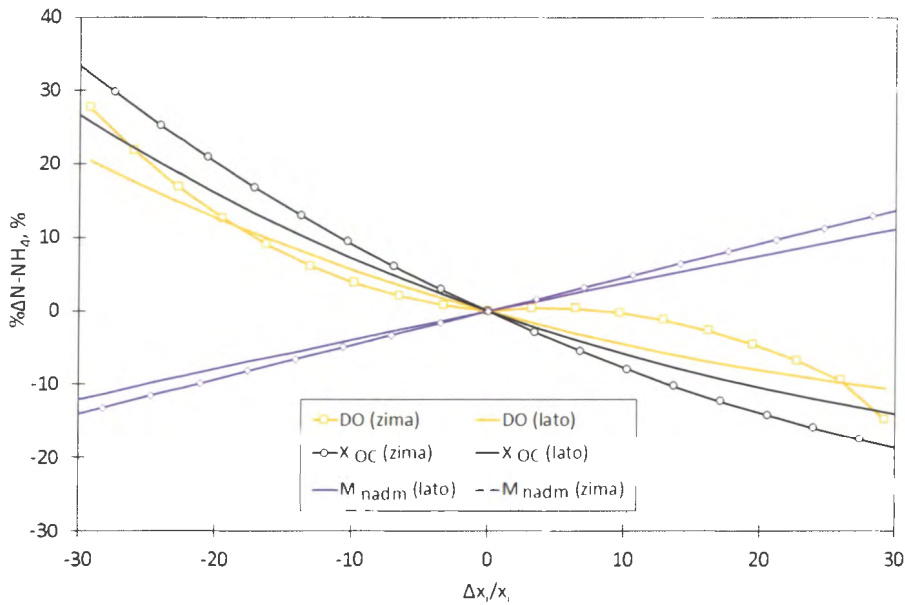
Tab. 7.4.9. Zestawienie wartości współczynników dopasowania wyników obliczeń do pomiarów (MAE, RMSE, MAPE, R) wybranych wskaźników jakości ścieków uzyskanych metodami SVM, CNN i BT dla oczyszczalni w Sitkówce–Nowiny.

Wskaźnik	Metoda	Test				Walidacja			
		MAE mg/l	RMSE mg/l	MAPE %	R -	MAE mg/l	RMSE mg/l	MAPE %	R -
N _{og}	CNN	0,61	0,84	8,12	0,94	0,59	0,82	7,58	0,95
	SVM	0,68	0,93	9,37	0,93	0,71	0,95	9,52	0,91
	BT	0,80	1,43	10,11	0,86	0,84	1,52	10,96	0,82
N-NH ₄	CNN	0,27	0,37	25,21	0,95	0,26	0,37	23,20	0,96
	SVM	0,18	0,32	20,10	0,97	0,19	0,31	19,95	0,97
	BT	0,29	0,37	26,03	0,87	0,31	0,41	28,90	0,86
P _{og}	CNN	0,041	0,052	13,50	0,93	0,044	0,054	13,8	0,93
	SVM	0,021	0,027	6,96	0,98	0,024	0,027	7,50	0,96
	BT	0,044	0,056	13,70	0,93	0,047	0,059	15,0	0,91

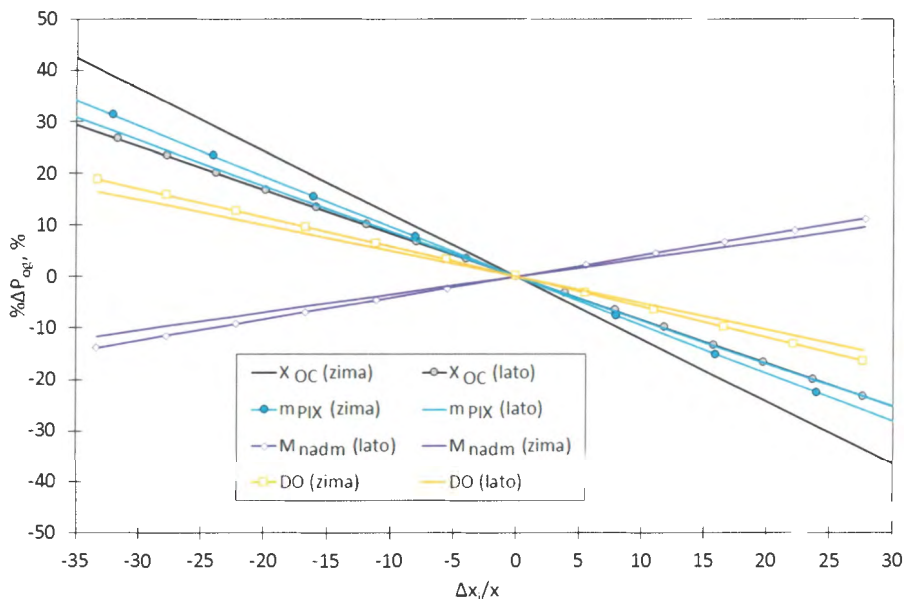
W kolejnym etapie analiz, bazując na algorytmie obliczeniowym z rys. 6.6, wykonano analizę wrażliwości modelu, ograniczając się do metody CNN ze względu na najmniejsze uzyskane przez nią wartości błędów. Wyniki analizy wrażliwości dla wybranych wskaźników jakości ścieków (N_{og}, N-NH₄, P_{og}) zilustrowano na rys. 7.4.16 ÷ 7.4.18 dla pory letniej i zimowej.



Rys. 7.4.16. Wpływ zmiennych sterujących (X_{OC} , M_{nadm} , DO) w różnych porach roku na zmianę zawartości azotu ogólnego na odpływie z oczyszczalni ścieków w Sitkówce-Nowiny.



Rys. 7.4.17. Wpływ zmiennych sterujących (X_{OC} , M_{nadm} , DO) w różnych porach roku na zmianę zawartości azotu amonowego na odpływie z oczyszczalni ścieków w Sitkówce-Nowiny.



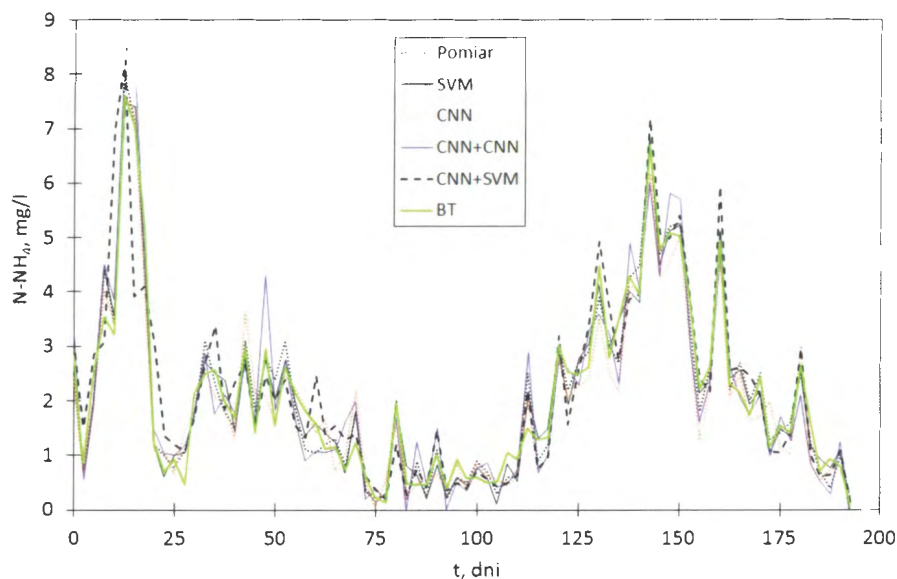
Rys. 7.4.18. Wpływ zmiennych sterujących (X_{OC} , M_{nadm} , DO , m_{pIX}) w różnych porach roku na zmianę zawartości fosforu ogólnego na odpływie z oczyszczalni ścieków w Sitkówce-Nowiny.

W oparciu o wyznaczone krzywe stwierdzono, że opracowane modele do prognozy wybranych wskaźników jakości ścieków (N_{og} , $N-NH_4$, P_{og}) poprawnie prognozują wpływ zmiennych sterujących, tj. X_{OC} , M_{nadm} , REC , DO , na jakość ścieków na wylocie z oczyszczalni. Stanowi to cenną informację z punktu możliwości implementacji modelu na działającym obiekcie.

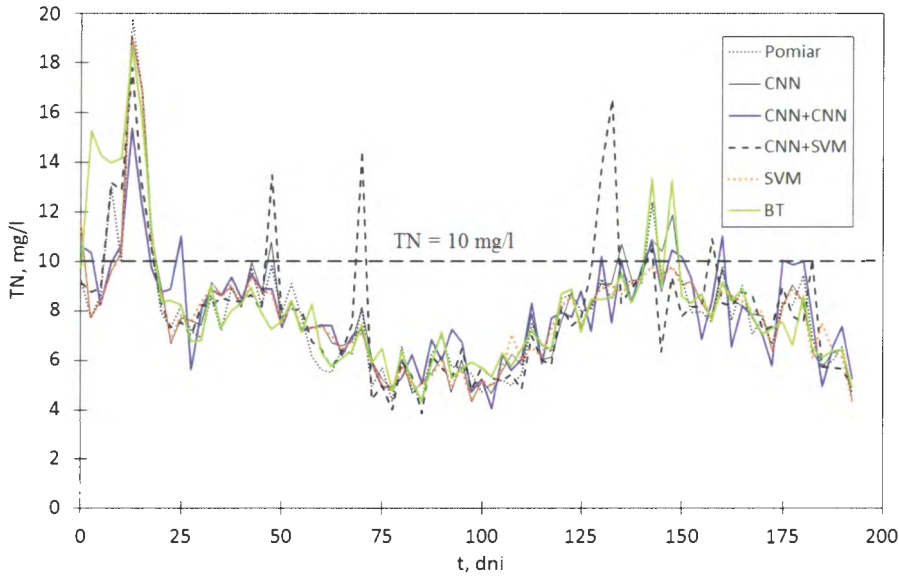
Wyznaczone modele opisane równaniem (7.4.6) mogą być stosowane, gdy zmienne niezależne (tab. 7.4.10 i 7.4.11) opisujące koncentrację zanieczyszczeń w ściekach dopływających są mierzone w sposób ciągły. Gdy nie jest to możliwe, konieczne jest zastąpienie pomiarów wartości wskaźników jakości ścieków, wymaganych do prognozy N_{og} i $N-NH_4$, wynikami ich obliczeń na podstawie równania (7.3.1). W modelu do symulacji P_{og} zachodzi potrzeba wykonania symulacji wartości BZT_5 , $ChZT$, N_{og} i $N-NH_4$. Wykorzystując opracowane wcześniej modele do prognozy wskaźników jakości ścieków wyznaczone metodą CNN (rozdz. 7.3.1) i równanie (7.3.1), otrzymane za ich pomocą wyniki obliczeń $C(t)_{in}$ podstawiono do równania (7.4.6). W ten sposób określono wartości wskaźników na wylocie z oczyszczalni dla zadanych wartości nastaw bioreaktora; wyniki obliczeń przedstawiono na rys. 7.4.19 ÷ 7.4.21 i w tab. 7.4.10.

Tab. 7.4.10. Zestawienie wartości współczynników dopasowania wyników obliczeń do pomiarów (MAE, RMSE, MAPE, R) dla wybranych wskaźników jakości ścieków uzyskanych dla oczyszczalni w Sitkówce-Nowiny.

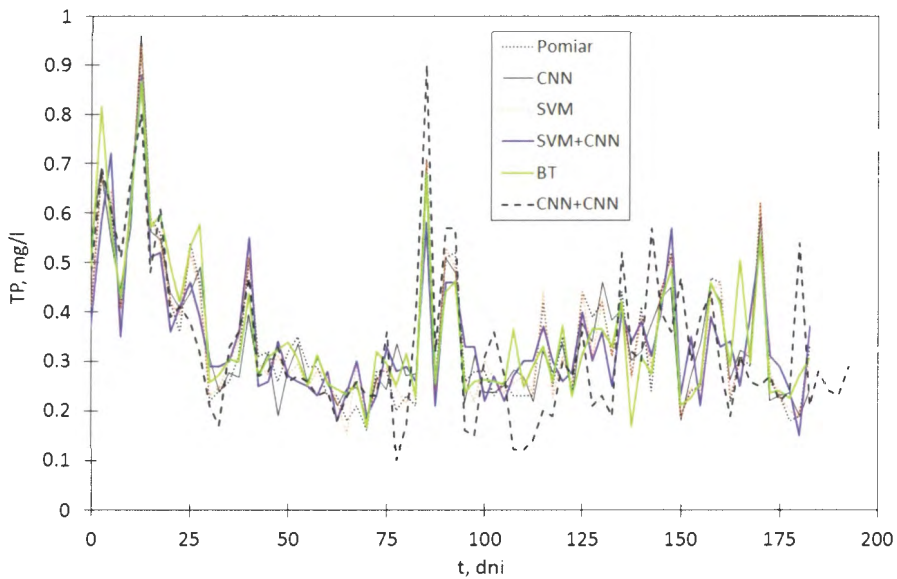
Wskaźnik	Metody	MAE mg/l	RMSE mg/l	MAPE %	R -
N _{og}	CNN+CNN	1,07	1,32	13,79	0,86
	CNN+SVM	1,07	1,75	13,12	0,78
N-NH ₄	SVM+CNN	0,32	0,49	25,29	0,92
	SVM+SVM	0,46	0,74	29,38	0,90
P _{og}	CNN+CNN	0,078	0,092	25,12	0,81
	SVM+CNN	0,061	0,065	19,12	0,90



Rys. 7.4.19. Porównanie wyników obliczeń i pomiarów wartości N-NH₄ na wylocie z obiektu w Sitkówce-Nowiny.



Rys. 7.4.20. Porównanie wyników obliczeń i pomiarów wartości N_{Og} na wylocie z obiektu w Sitkówce- Nowiny.



Rys. 7.4.21. Porównanie wyników obliczeń i pomiarów wartości P_{Og} na wylocie z obiektu w Sitkówce- Nowiny.

W oparciu o wyznaczone krzywe i stabilizowane dane można stwierdzić, że opracowane modele do prognozy jakości ścieków, będące kombinacją równań (7.3.1) i (7.4.9), charakteryzują się zadowalającymi zdolnościami predykcyjnymi. Wyniki te wskazują na to, że jest możliwa symulacja, a co za tym idzie również kontrola nastaw bioreaktora w celu zapewnienia na wylocie wymaganej jakości ścieków (N_{og} , $N-NH_4$ i P_{og}) w przypadku nieciągłości w pomiarach zmiennych niezależnych w równaniu (7.4.6). W oparciu o wyznaczone krzywe (rys. 7.4.19 ÷ 7.4.21) stwierdzono, że pojedyncze maksymalne wartości wskaźników (N_{og}) otrzymane z pomiarów mogą być niedoszacowane przy użyciu modeli (CNN+CNN) bazujących na równaniach (7.4.6) i (7.3.1). Tym samym w przypadkach, które można uznać za skrajne (jednak ważne dla działania obiektu), może dochodzić do spadku niezawodności działania obiektu, jednak skutki tego będą znacznie mniejsze niż w sytuacji, gdy nie ma możliwości doboru nastaw reaktora.

W przypadku modeli do prognozy P_{og} (rys. 7.4.21) szereg wartości maksymalnych jest zawyżony. Pozwala to na stwierdzenie, że wartości wyznaczanych wskaźników jakości ścieków mogą być niższe, niż w rzeczywistości, ale mimo tego dobrane nastawy reaktora zapewnią większy zapas bezpieczeństwa w działaniu obiektu, niż brak ich regulacji.

Tab. 7.4.11. Porównanie wartości dopasowania wyników obliczeń do pomiarów wybranych wskaźników jakości ścieków na wylocie z oczyszczalni Sitkówka-Nowiny wyznaczonych w pracy i przez innych autorów; b. własne* - modele opracowane na podstawie kombinacji równań (7.3.1) i (7.4.6).

Wskaźniki jakości	Źródło	Wartości miar dopasowania
N-NH ₄	Chen i in. (2003)	R=0,980
	Raduly i in. (2007)	R=0,971
	Jami i in. (2011)	R=0,921
	Qiao i in. (2011)	MSE = 0,018 mg/l
	Gaya i in. (2014)	R=0,948
	Mirbagherii i in. (2014)	R=0,902
	Mirbagherii i in. (2015)	R=0,992
	Han i in. (2016)	R=0,990
	b. własne	R=0,950
	b. własne*	R=0,920

Tab. 7.4.11. cd.

Wskaźniki jakości	Źródło	Wartości miar dopasowania
N _{og}	Clara (2008)	R=0,955
	Woo i in. (2009)	RMSE = 16,81 mg/l
	Luo i in. (2009)	R=0,938
	Lee i in. (2009)	R=0,920
	Mirbagherii i in. (2014)	R=0,882
	Hongbin i in. (2014)	R=0,748
	b. własne	R=0,940
	b. własne*	R=0,873
P _{og}	Chen i in. (2003)	R=0,970
	Lee i in. (2009)	R=0,936
	Hongbin i in. (2014)	R=0,800
	Mirbagherii i in. (2015)	R=0,994
	b. własne	R=0,970
	b. własne*	R=0,900

Podsumowując obliczenia wybranych wskaźników jakości ścieków na odpływie z oczyszczalni ścieków w oparciu o klasyczne podejście, tj. równanie (7.4.6), oraz kombinację równań (7.3.1) i (7.4.6), celowe jest porównanie otrzymanych wyników z wynikami modelowania uzyskanymi przez innych autorów (tab. 7.4.11).

Analizując dane w tab. 7.4.11 w odniesieniu do wartości N-NH₄ można stwierdzić, że nieznacznie lepsze dopasowanie (nie więcej niż o 10%) wyników obliczeń do pomiarów otrzymali Han i in. (2016), Raduly i in. (2007), Qiao i in. (2011) oraz Mirbagherii (2015). W pierwszym przypadku autorzy do tworzenia modelu wykorzystali pomiary P_{og}, Z_{og}, DO, T_{KOC}, ORP i jednocześnie zastosowana przez nich metoda (model hybrydowy bazujący na metodzie zbiorów rozmytych i c-średnich) była bardziej złożona obliczeniowo, niż wykorzystana w pracy metoda CNN. Implementacja metody c-średnich jako klasyfikatora pozwoliła wydzielić grupy podobnych danych i dla każdej grupy wyznaczyć osobny model statystyczny na bazie teorii zbiorów rozmytych. Podejście to, w przypadku zbioru danych odpowiedniej wielkości, jest stosowane na etapie tworzenia modeli statystycznych i istotnie wpływa na poprawę dokładności prognozy (Lee i in. 2009). W pracy Qiao i in. (2011) wykorzystano do obliczeń model hierarchiczny, a zmiennymi niezależnymi w modelu były udziały poszczególnych frakcji

zanieczyszczeń w ściekach, stosowane standardowo jako wejścia do modeli ASM. Ustalenie wartości poszczególnych frakcji w ściekach wymaga wykonania czasochłonnych analiz, które nie są rutynowo wykonywane w oczyszczalniach, dlatego ich wykorzystanie jako zmiennych niezależnych w modelach statystycznych jest ograniczone. Lepsze dopasowanie wyników obliczeń do pomiarów, niż wyznaczone w pracy, uzyskał również Chen (2003), który zastosował bardziej złożony model (rekurencyjne sieci neuronowe) i uwzględnił także w obliczeniach dynamiczne zmiany parametrów bioreaktora (DO, pH, ORP) w czasie. W modelu MLP podanym przez Mirbagherii i in. (2015), podobnie jak w pracy, ujęto wskaźniki jakości ścieków dotyczące zawartości związków organicznych, fosforu i azotu oraz wskaźniki działania reaktora biologicznego. Jednak, dane dla modelu pochodziły z instalacji laboratoryjnej, w której wykonywanie pomiarów znacznie odbiega od warunków panujących na obiekcie, co może wyjaśniać uzyskaną przez autorów wysoką dokładność modelu.

Zadawalające dopasowanie wyników obliczeń do pomiarów wartości N_{og} w porównaniu z wynikami podanymi w Clara (2008), Woo i in. (2009), Luo i in. (2009), Lee i in. (2009), Mirbagherii i in. (2015) stwierdzono w przypadku modeli, w których mierzone wartości zmiennych niezależnych dotyczących wskaźników jakości ścieków zastąpiono wynikami obliczeń równania (7.3.1). Wyniki modelowania N_{og} otrzymane w pracy są jednak gorsze, niż uzyskane przez Luo i in. (2009) (o 8,5%), Clara (2008) (o 10,5%), Mirbagherii i in. (2014) (o 2,1%) i Lee i in. (2009) (o 6,5%). W pracy Luo i in. (2009) zmienne niezależne uwzględniane w modelu obejmowały tylko wskaźniki działania bioreaktora (pH, ORP, T, DO, X_{OC} , M_{nadm} , REC), natomiast w pracach Clara (2008), Mirbagherii i in. (2014) i Lee i in. (2009) korzystano w symulacji zarówno ze wskaźników jakości ścieków na dopływie do oczyszczalni, jak i z wielkości nastaw działania bioreaktora (DO, X_{OC} , M_{nadm} , REC).

Z kolei Hongbin i in. (2014), bazując na wynikach pomiarów jakości ścieków na dopływie do oczyszczalni (BZT₅, ChZT, Z_{og} , N_{og} , P_{og}) i przy zastosowaniu metody ANFIS, podjęli próbę opracowania modeli do prognozy N_{og} i P_{og} . Uzyskane przez nich wyniki okazały się jednak niezadawalające. Wskazują one na duży wpływ na wartości N_{og} na wylocie z oczyszczalni wskaźników opisujących działanie bioreaktora. Mniejsze błędy prognozy P_{og} , niż Hongbin i in. (2014) uzyskali Chen i in. (2008), Lee i in. (2009) oraz Mirbagherii (2014).

Na uwagę zasługuje fakt, że różnice w wynikach obliczeń dotyczących P_{og} , uzyskane za pomocą kombinacji równań (7.3.1) i (7.4.6) i jedynie równania (7.4.6), są większe (wynoszą 7,8% na korzyść modelu (7.4.6)), niż w przypadku obliczania pozostałych wskaźników (dla N_{og} wynoszą 7,7% i dla $N-NH_4$ wynoszą 4%). Może to wynikać z tego, że zawartość fosforu na wylocie z oczyszczalni ścieków zależy

od większej liczby wskaźników na dopływie (BZT_5 , $ChZT$, N_{og} , $N-NH_4$, P_{og}), niż w przypadku N_{og} i $N-NH_4$ (BZT_5 , N_{og} , $N-NH_4$).

Na podstawie przeprowadzonych badań można stwierdzić, że modele do prognozy jakości ścieków na wylocie (N_{og} , $N-NH_4$ i P_{og}) z oczyszczalni, wyznaczone przy pomocy metod sztucznych sieci neuronowych (CNN, SVM), charakteryzują się lepszymi zdolnościami predykcyjnymi, niż modele wyznaczone metodą drzew wzmocnianych. Ponadto, podobnie jak poprzednio, przeprowadzone obliczenia wykazały, że cennym narzędziem pozwalającym na weryfikację opracowanych modeli statystycznych typu czarnej skrzynki jest analiza wrażliwości. Uzyskane przy jej pomocy wyniki pozwalają ustalić charakter zależności między poszczególnymi zmiennymi niezależnymi a wyjściem z modelu. Zaproponowane podejście stanowi dodatkowy element oceny zdolności predykcyjnej modeli i do tej pory nie było tematem analiz prowadzonych dla oczyszczalni ścieków przez innych autorów.

Analizując wyniki obliczeń można również stwierdzić, że możliwe jest poprawne modelowanie jakości ścieków na wylocie z oczyszczalni w przypadku nieciągłości w szeregach pomiarowych dotyczących zmiennych N_{og} , $N-NH_4$ i P_{og} . Z punktu widzenia eksploatacji istniejących i projektowanych obiektów jest to bardzo ważne, bowiem pozwala na uzyskanie wysokiej niezawodności działania obiektu mimo braku pomiarów dla kluczowych zmiennych procesowych.

9. Podsumowanie i wnioski końcowe

Oczyszczalnia ściek6w stanowi złoŝony obiekt technologiczny, w którym zachodzi szereg proces6w biochemicznych. Podstawą oceny działania oczyszczalni jest uzyskiwanie na wylocie załoŝonej jakości ściek6w. Z uwagi na losowy i zaleŝny od pory roku charakter ilości i jakości ściek6w na dopływie, zapewnienie wymaganej jakości ściek6w na odpływie jest trudne. Wymaga to regulacji w trybie dynamicznym nastaw bioreaktora i wybranych obiekt6w technologicznych, co pozwala na spełnienie obowiązujących wymagań prawnych. W związku z powyŝszym podejmowane sã próby wspomaganienia procesu usuwania związkw6w biogennych przy wykorzystaniu modeli matematycznych opracowanych na podstawie pomiar6w 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 przypadkw6w stosuje się modele parametryczne (szarej lub czarnej skrzynki).

W przypadku obiekt6w, na których gromadzone sã pomiary w trybie ciãgłym (on-line) (ilość i jakość ściek6w na dopływie i odpływie z oczyszczalni, parametry operacyjne bioreaktora), takie modele pozwalajã na uzyskanie zadowalajãcej jakości ściek6w na wylocie i obniŝenie koszt6w eksploatacyjnych. W przypadku braku ciãgłośc6i w pomiarach jakości ściek6w na dopływie do oczyszczalni, na przykład na skutek wadliwej pracy urzãdzeń pomiarowych lub gdy wartośc6i wskaźnik6w oznaczane sã drogã analitycznã, modele matematyczne zasilane danymi w trybie on-line majã ograniczone zastosowanie. Wynika to z braku wartośc6i liczbowych zmiennych niezaleŝnych uwzględnianych w modelach procesowych.

Poza problemami z dostępem do danych wejściowych w modelach matematycznych oczyszczalni pojawia się kolejna istotna kwestia dotyczãca doboru metody do symulacji i interpretacji uzyskanych wynik6w. Naleŝy rozwaŝyć kwestię, czy tworzyć modele o prostszej strukturze wymagajãce mniejszych nakładow6w obliczeniowych i kr6tszego czasu obliczeń, czy może stosować modele bardziej złoŝone? Sformułowane zadanie jest trudne, bowiem stanowi kompromis międy dokładnoścã prognoz i złoŝonoścã modelu i ma istotny wplyw 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 wzmacnianych.

W przypadku niezadowalających 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., Gonzáles 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., Tchobanoglous 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., Adeloje 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., Edebali 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