

Zbigniew Kowalewski, Elena Neverova-Dziopak, Michał Preisner

## Próba opracowania modelu regresyjnego do szacowania wartości BZT<sub>5</sub> ścieków komunalnych

Poprawna eksploatacja nowoczesnych urządzeń do biologicznego oczyszczania ścieków – oprócz znajomości aspektów technicznych i rozwiązań konstrukcyjnych poszczególnych urządzeń – wymaga także wiedzy na temat podstawowych procesów biochemicznych. W rzeczywistych warunkach niestabilności charakterystyki jakościowej i ilościowej ścieków kierowanych do części biologicznej oczyszczalni, podejmowanie właściwych decyzji jest warunkiem zapewnienia wymaganej jakości ścieków oczyszczonych oraz minimalizacji skutków sytuacji awaryjnych. Podstawowym składnikiem ścieków komunalnych są zanieczyszczenia organiczne, których zawartość charakteryzują dwa podstawowe wskaźniki – biochemiczne (BZT) oraz chemiczne zapotrzebowanie na tlen (ChZT). Wartości tych wskaźników są wykorzystywane do ustalania podatności ścieków na biodegradację oraz obliczania obciążenia urządzeń do oczyszczania biologicznego ładunkiem zanieczyszczeń organicznych, a także w doborze i obliczaniu bioreaktorów oraz kontroli sprawności działania oczyszczalni.

Biochemiczne zapotrzebowanie ścieków na tlen zasługuje na szczególną uwagę nie tylko dlatego, że stanowi jeden z głównych wskaźników projektowych i eksploatacyjnych w procesie ich oczyszczania, ale także w związku z problemami dotyczącymi prawidłowego oznaczania jego wartości, które mogą być obarczone dość dużymi błędami wynikającymi z możliwości występowania w ściekach zanieczyszczeń specyficznych, które mogą zawyżać lub zaniżać wynik analizy. Uwzględniając dość długi czas potrzebny do oznaczenia wartości BZT (5 d, 7 d, 20 d), zaleca się opracowanie szybkich metod oznaczania wartości tego wskaźnika, co pozwoli na bardziej sprawne sterowanie warunkami technologicznymi panującymi w bioreaktorach i ich korektę w celu zachowania aktywności biocenozy osadu czynnego oraz wymaganej sprawności usuwania zanieczyszczeń. Do oceny sprawności procesów biologicznych w oczyszczalniach ścieków w Europie wykorzystuje się najczęściej BZT<sub>5</sub>, w niektórych krajach (Szwecja, Finlandia, Litwa, Łotwa i Estonia) – BZT<sub>7</sub> [1, 2], natomiast w Federacji Rosyjskiej – w zależności od celu oznaczania tego wskaźnika – używa się BZT<sub>5</sub> oraz BZT<sub>20</sub> (BZT<sub>c</sub> – zapotrzebowanie całkowite) [3, 4]. W ostatnich latach nastąpił znaczący rozwój nowoczesnych metod wykorzystujących

czujniki biologiczne (biosensory), dzięki którym możliwe jest szybsze uzyskanie wartości biochemicznego zapotrzebowania ścieków na tlen do bieżącego monitoringu procesów oczyszczania biologicznego [5]. Do szybszych sposobów określania wartości tego wskaźnika należą metody szacowania oparte na modelach opisujących jego zależność od takich parametrów, jak potencjał elektryczny czy aktywność bakterii (bioluminescencja) [6].

Nowoczesne urządzenia pomiarowe pracujące w sposób ciągły oraz technologie informatyczne umożliwiają prowadzenie stałego monitoringu parametrów technologicznych i eksploatacyjnych procesu oczyszczania ścieków i gromadzenie odpowiedniej bazy danych do opracowania modeli, które pozwalają na optymalizację bieżącej kontroli warunków eksploatacji oczyszczalni i podejmowanie właściwych decyzji. Opracowanie i wykorzystanie takich modeli pozwoli na skrócenie czasu uzyskania wartości biochemicznego zapotrzebowania ścieków na tlen oraz bieżącą kontrolę warunków prowadzenia procesu biologicznego oczyszczania i jego sprawności.

Celem badań opisanych w artykule było opracowanie modelu matematycznego do szacowania wartości BZT<sub>5</sub> ścieków na podstawie danych z dziesięcioletniego monitoringu prowadzonego w oczyszczalni komunalnej (Terrence J. O'Brien Water Reclamation Plant – WRP) w Chicago (Stany Zjednoczone). Dane wykorzystane w badaniach zostały udostępnione przez eksploatatora oczyszczalni (Metropolitan Water Reclamation District of Greater Chicago).

### Charakterystyka danych wykorzystanych w badaniach

Oczyszczalnia ścieków o nazwie Terrence J. O'Brien WRP jest jedną z siedmiu oczyszczalni komunalnych funkcjonujących obecnie w Chicago. Została uruchomiona w 1928 r. pod nazwą North Side WRP, pochodzącej od nazwy kanału stanowiącego odbiornik ścieków North Shore oraz faktu, że obsługuje głównie północną część ścisłego centrum Chicago. Oczyszczalnia ta pierwotnie oczyszczała ścieki pochodzące od około 800 tysięcy mieszkańców, natomiast obecnie odbiera ścieki od ponad 1,3 miliona mieszkańców, przy czym średnia dobowo ilość oczyszczanych ścieków wynosi około 870 tys. m<sup>3</sup>/d [7]. Technologia stosowana w T. J. O'Brien WRP wykorzystuje konwencjonalny układ z tlenowym reaktorem biologicznym. W celu ochrony zasobów wodnych Jeziora Michigan, od 2016 r. prowadzona jest dezynfekcja ścieków oczyszczonych promieniami nadfioletowymi [8].

Dr inż. Z. Kowalewski, prof. dr hab. inż. E. Neverova-Dziopak, mgr inż. M. Preisner: AGH Akademia Górniczo-Hutnicza w Krakowie, Wydział Geodezji Górniczej i Inżynierii Środowiska, Katedra Kształtowania i Ochrony Środowiska, al. Adama Mickiewicza 30, 30-059 Kraków  
kowalew@agh.edu.pl, elenad@agh.edu.pl, preisner@agh.edu.pl

Bazę danych wykorzystaną do sformułowania modelu stanowiły wyniki monitoringu jakości ścieków dopływających do oczyszczalni T.J. O'Brien WRP w latach 1990–2000. W tym czasie wykonano regularne (około 400 rocznie) pomiary wartości takich wskaźników, jak BZT<sub>5</sub>, pH, zawiesiny ogólne (Z), azot amonowy (NH<sub>4</sub><sup>+</sup>), azot Kjeldahla (N<sub>Kj</sub>) oraz azotany (NO<sub>3</sub><sup>-</sup>). Podstawowe statystyki opisowe ogółu danych przedstawiono w tabeli 1.

Ścieki charakteryzowały się znacznym udziałem azotu Kjeldahla i niewielką ilością azotanów. Wartości BZT<sub>5</sub> i pH, a także zawartość zawiesin (oprócz wartości maksymalnych) były typowe w przypadku tego rodzaju ścieków. Średnioroczne wartości poszczególnych danych przedstawiono w tabeli 2.

Przeprowadzona analiza wykazała, że na przestrzeni lat 1990–2000 wartości azotu Kjeldahla i azotu amonowego miały nieznaczną tendencję rosnącą, przy czym zawartość azotanów malała wraz ze zmniejszaniem się ich zakresu. Podobna sytuacja wystąpiła w przypadku pH, jednak pomimo zmniejszenia zakresu, wartości tego wskaźnika w każdym roku były na podobnym poziomie. Wartości BZT<sub>5</sub>

oraz zawiesin ogólnych utrzymywały się w każdym roku na jednakowym poziomie, jednak w latach 1997 i 1999 wystąpiły pojedyncze wartości wyraźnie odstające od pozostałych.

### Budowa modeli regresyjnych

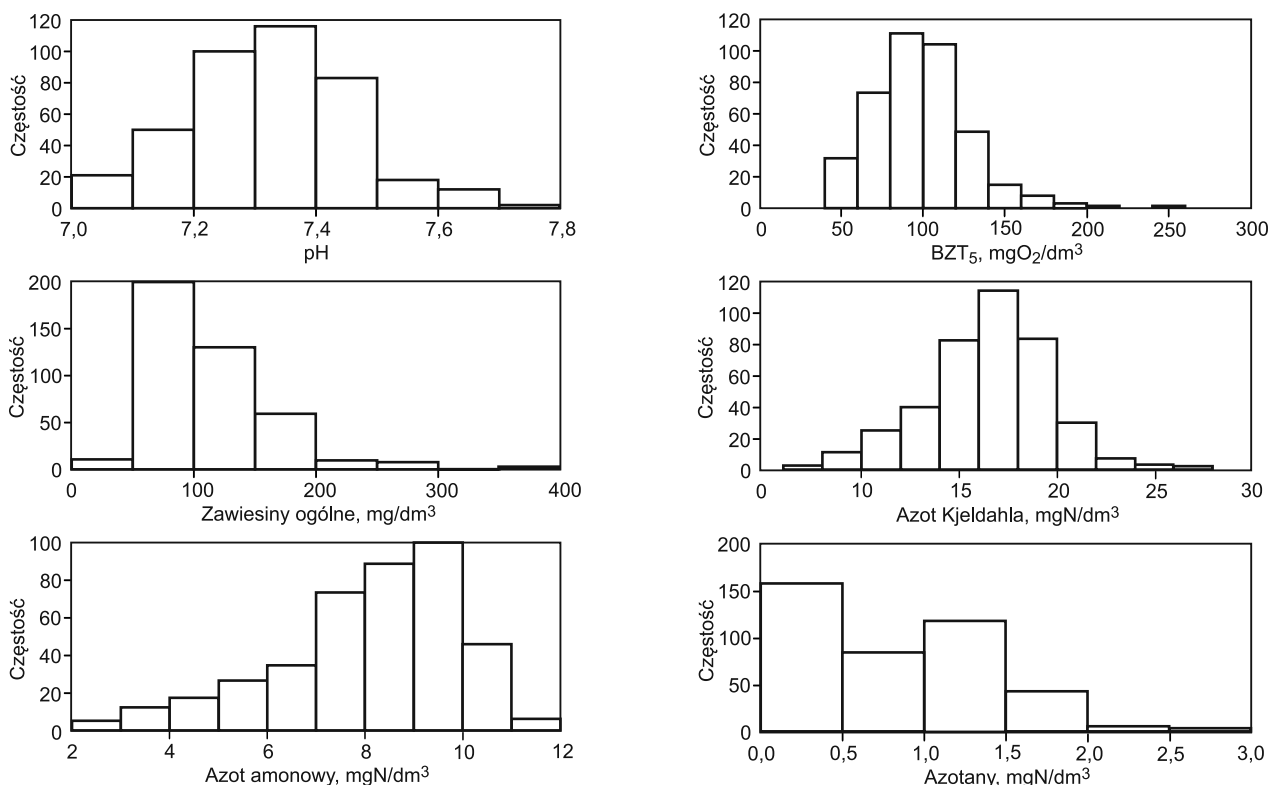
Obliczenia zostały wykonane w programie R z nakładką R Studio w wersji 1.0.143 [9], a wszystkie testy zostały przeprowadzone przy współczynniku istotności  $\alpha=0,05$ . Jako dane wyjściowe do budowy modeli regresyjnych posłużyły wyniki pomiarów z 1992 r., natomiast dane z pozostałych lat wykorzystano do testowania modeli. Wybór danych z 1992 r. uznano za korzystny z punktu widzenia tworzenia modeli regresyjnych, ponieważ wyniki monitoringu ścieków w tym roku dobrze odzwierciedlały całość danych. Jednakże należy mieć na uwadze specyfikę funkcjonowania oczyszczalni ścieków, a zwłaszcza procesów oczyszczania biologicznego, a także brak możliwości gromadzenia i swobodnego wyboru odpowiedniego zestawu danych w warunkach eksploatacyjnych.

Tabela 1. Statystyki opisowe zbioru danych z lat 1990–2000  
Table 1. Descriptive statistics for the data set 1990-2000

Statystyka	Wskaźnik jakości ścieków, jednostka					
	pH	BZT <sub>5</sub> , mgO <sub>2</sub> /dm <sup>3</sup>	Zawiesiny ogólne, mg/dm <sup>3</sup>	Azot amonowy, mgN/dm <sup>3</sup>	Azot Kjeldahla, mgN/dm <sup>3</sup>	Azotany, mgN/dm <sup>3</sup>
Minimum	6,50	24,00	16,00	0,60	3,10	0,00
1. Kwartyl	7,20	80,00	84,00	7,04	14,90	0,03
Mediana	7,30	97,00	102,10	8,79	17,60	0,24
Średnia	7,31	99,92	114,70	8,52	17,58	0,59
3. Kwartyl	7,40	114,00	128,00	10,26	20,30	1,00
Maksimum	8,00	475,00	1720,00	10,23	40,90	4,79

Tabela 2. Średnioroczne wartości analizowanych wskaźników jakości ścieków  
Table 2. Average annual values of analyzed sewage quality parameters

Rok	Liczba pomiarów	Wskaźnik jakości ścieków, jednostka					
		pH	BZT <sub>5</sub> , mgO <sub>2</sub> /dm <sup>3</sup>	Zawiesiny ogólne, mg/dm <sup>3</sup>	Azot amonowy, mgN/dm <sup>3</sup>	Azot Kjeldahla, mgN/dm <sup>3</sup>	Azotany, mgN/dm <sup>3</sup>
1990	400	7,12	86,43	101,97	7,89	14,40	1,15
1991	401	7,23	92,44	110,39	7,32	14,99	0,68
1992	402	7,37	100,40	109,16	8,11	16,52	0,82
1993	400	7,42	91,13	104,32	7,79	15,65	0,69
1994	397	7,34	97,48	113,64	9,10	17,48	0,38
1995	400	7,34	99,75	121,51	18,04	8,77	0,56
1996	400	7,34	114,81	125,08	18,78	9,37	0,51
1997	397	7,31	110,88	126,31	18,35	8,01	0,46
1998	393	7,31	93,94	108,68	18,43	8,39	0,46
1999	394	7,33	101,32	114,22	19,02	9,04	0,38
2000	363	7,31	111,61	128,18	22,18	10,09	0,34

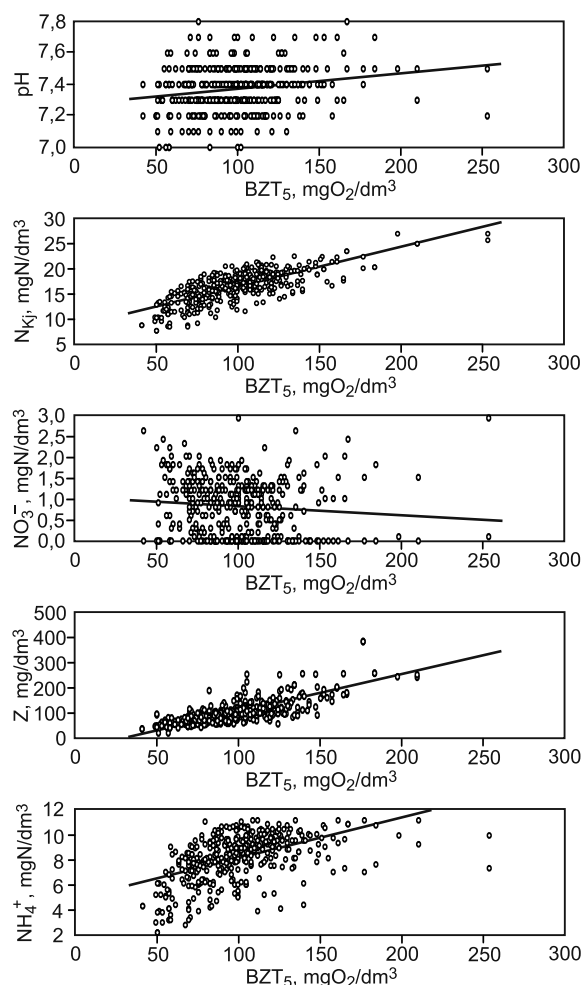


Rys. 1. Histogramy rozkładu zmiennych  
Fig. 1. Histograms of data distribution

Pierwszym etapem modelowania było sprawdzenie charakteru rozkładów wartości badanych parametrów, ich wzajemnych korelacji i liniowości oraz stworzenie prostego modelu regresyjnego ujmującego zależność BZT<sub>5</sub> od wybranych wskaźników jakości ścieków. Histogramy z rozkładami zmiennych przedstawiono na rysunku 1, przy czym zbieżność z rozkładem normalnym badano testem Shapiro-Wilka. Histogramy pH, BZT<sub>5</sub> oraz azotu Kjeldahla charakteryzowały się regularnym kształtem zbliżonym do rozkładu normalnego, histogramy zawiesin ogólnych oraz azotanów były prawoskośne, natomiast histogram azotu amonowego lewoskośny. Test Shapiro-Wilka nie wykazał zgodności z rozkładem normalnym żadnego z analizowanych wskaźników.

Za pomocą współczynnika korelacji Pearsona sprawdzono zależności korelacyjne między danymi. Wysokimi wartościami współczynnika korelacji charakteryzowały się zależności modelowanego BZT<sub>5</sub> z zawiesinami ogólnymi (0,78), azotem Kjeldahla (0,75) i azotem amonowym (0,51). Oprócz nich wysoki współczynnik korelacji miały azot Kjeldahla i azot amonowy (0,83) oraz zawiesiny ogólne i azot Kjeldahla (0,55). Pozostałe wskaźniki nie wykazywały wzajemnych silnych korelacji. Wartości współczynnika zmienności poszczególnych wskaźników wynosiły odpowiednio: pH – 2%, BZT<sub>5</sub> – 30%, azotu Kjeldahla – 19%, zawiesiny ogólne – 51%, azot amonowy – 23% oraz azotan – 81%. Oprócz wartości pH, pozostałe wskaźniki jakości ścieków wykazywały zmienność w zakresie od przeciętnej do dużej.

W celu określenia charakteru zależności pomiędzy wartościami BZT<sub>5</sub> i pozostałych wskaźników wykonano wykresy rozrzutu wraz z liniami regresji. Analizując wykresy zamieszczone na rysunku 2 uznano za liniowe zależności między wartościami BZT<sub>5</sub> i pH, zawiesinami ogólnymi, azotem Kjeldahla oraz azotem amonowym. Zależność liniowa między BZT<sub>5</sub> i azotanami była mało widoczna.



Rys. 2. Diagramy korelacyjne z liniami regresji  
Fig. 2. Correlation diagrams with regression lines

Do podstawowych założeń regresji liniowej należą liniowość modelu, rozkład normalny zmiennych i brak korelacji między zmiennymi objaśniającymi [12]. W wypadku danych pomiarowych z 1992 r. część tych założeń nie była spełniona, jednak zdecydowano się na zastosowanie modelu liniowego do ich opisu. Wynikało to z łatwości użycia prostych narzędzi statystycznych, w postaci modeli liniowych, do podejmowania decyzji w warunkach eksploatacji oczyszczalni ścieków.

W pierwszej kolejności podjęto próbę oszacowania wartości BZT<sub>5</sub> ścieków za pomocą pojedynczych danych pochodzących z 1992 r., przy czym modele powstałe w ten sposób nie zostały poddane pełnej diagnostyce modeli regresyjnych. Do oceny jakości oszacowanych wartości stosuje się najczęściej średni błąd bezwzględny (mean absolute error – MAE) [10, 11] w postaci:

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

w której:

$n$  – liczebność analizowanego zbioru danych

$y_{i,obs/i,osz}$  – wartość obserwowana/oszacowana

Parametry funkcji regresji oraz wartości błędów przedstawiono w tabeli 3.

Tabela 3. Współczynniki funkcji regresji oraz wartości błędów  
Table 3. Regression function coefficients and error values

Parametr	Współczynniki funkcji regresji i błędy				
	pH	Z	NH <sub>4</sub> <sup>+</sup>	N <sub>Kj</sub>	NO <sub>3</sub> <sup>-</sup>
Wyraz wolny	-222,28	54,23	35,21	-17,12	103,96
Współczynnik kierunkowy	43,76	0,42	8,04	7,11	-4,35
Średni błąd bezwzględny (MAE)	21,89	14,66	18,22	15,14	22,01
Test Fishera-Snedecora	17,9	663,3	140,9	529,6	3,7
Skorygowany współczynnik determinacji	0,04	0,62	0,26	0,57	0,01

Z – zawiesiny ogólne, NH<sub>4</sub><sup>+</sup> – azot amonowy  
N<sub>Kj</sub> – azot Kjeldahla, NO<sub>3</sub><sup>-</sup> – azotany

W grupie wskaźników jakości ścieków, które posłużyły do budowania modeli, wartości MAE, wyniki testu Fishera-Snedecora oraz wartości współczynnika determinacji były najbardziej korzystne w przypadku równania regresji opartego jedynie na zawartości zawiesin ogólnych. Model ten oznaczono symbolem I. W następnym etapie badań sformułowano trzy kolejne modele w postaci regresji liniowej, oznaczone symbolami II, III i IV. Do budowy modelu II wykorzystano wszystkie wskaźniki jakości ścieków. W wyborze wskaźników istotnych do budowy modelu zastosowano kryterium informacyjne Akaikego (AIC) (parametry wyboru: direction=both, start=2065,06) oraz algorytm Leaps (parametry wyboru: nbest=10, selection algorithm=exhaustive, r2) [13], w wyniku czego w modelu III nie uwzględniono zawartości azotanów, a w skład modelu IV nie weszły wartość pH i zawartość azotanów. Wartości współczynników regresji czterech sformułowanych modeli przedstawiono w tabeli 4.

Tabela 4. Współczynniki regresji sformułowanych modeli  
Table 4. Regression coefficients in formulated models

Model	Wyraz wolny	Współczynnik kierunkowy				
		pH	Z	NH <sub>4</sub> <sup>+</sup>	N <sub>Kj</sub>	NO <sub>3</sub> <sup>-</sup>
I	54,23	–	0,42	–	–	–
II	-129,58	17,89	0,28	-0,25	4,31	-0,77
III	-125,28	16,91		-0,31	4,40	–
IV	-2,28	–		0,03	4,30	

Z – zawiesiny ogólne, NH<sub>4</sub><sup>+</sup> – azot amonowy, N<sub>Kj</sub> – azot Kjeldahla, NO<sub>3</sub><sup>-</sup> – azotany

W tabeli 5 zestawiono podstawowe statystyki służące do diagnostyki regresyjnych modeli liniowych. Założenie o zerowych średnich i sumach reszt zostało spełnione we wszystkich proponowanych modelach. Normalność rozkładu reszt była testowana testami Shapiro-Wilka i Jarque-Bera. Założenie o normalności ich rozkładu zostało spełnione jedynie w przypadku modelu I. Autokorelacja predyktorów modelu była sprawdzana za pomocą testu Durбина-Watsona oraz czynników inflacji wariancji VIF (variance inflation factor) (tab. 6) i autokorelacji ACF (autocorrelation function). Założenie dotyczące autokorelacji nie zostało spełnione, a wartości czynników VIF oraz ACF nie były zadowalające. Założenie o jednorodności wariancji sprawdzono testem Breuscha-Pagana i nie została ona osiągnięta w żadnym z modeli.

Tabela 5. Statystyki sformułowanych modeli regresyjnych  
Table 5. Statistics of formulated regression models

Statystyka	Model			
	I	II	III	IV
Minimum	-56,5	-57,07	-57,45	-55,93
Maksimum	54,2	42,73	43,38	44,08
Średnia	0	~0	~0	~0
Suma	~0	~0	~0	~0
S-W (p)	0,4851	0,00758	0,007724	0,0152
J-B (p)	0,701	0,004015	0,00398	0,006761
D-W (p)	3,97e-14	1,215e-09	7,388e-10	2,336e-11
B-P (p)	1,733e-09	0,001091	0,0004057	0,0001212

S-W – test Shapiro-Wilka, J-B – test Jarque-Bera  
D-W – test Durбина-Watsona, B-P – test Breuscha-Pagana

Tabela 6. Czynniki inflacji wariancji (VIF) sformułowanych modeli regresyjnych  
Table 6. Variance inflation factor (VIF) in formulated regression models

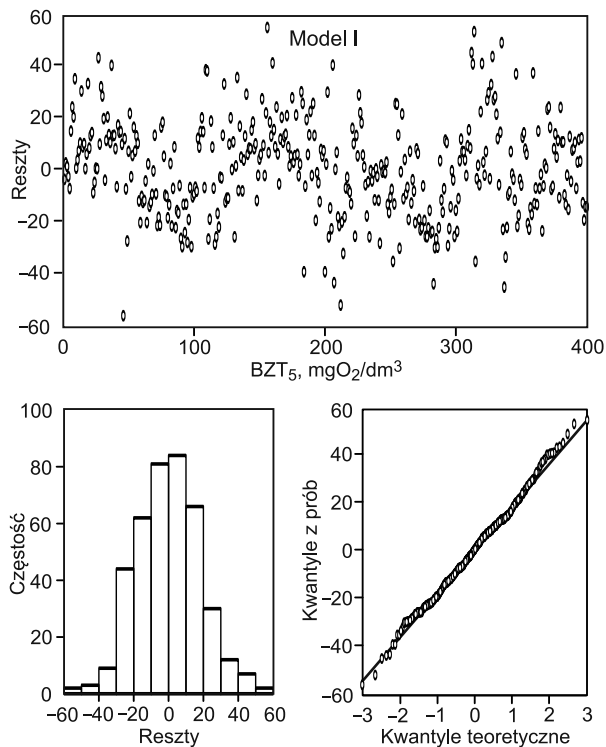
Model	Czynnik inflacji wariancji (VIF)				
	pH	Z	NH <sub>4</sub> <sup>+</sup>	N <sub>Kj</sub>	NO <sub>3</sub> <sup>-</sup>
I	–	–	–	–	–
II	1,09	1,95	4,39	6,19	1,15
III	1,04	1,88	4,32	5,78	–
IV	–	1,87	4,24	5,76	

Z – zawiesiny ogólne, NH<sub>4</sub><sup>+</sup> – azot amonowy  
N<sub>Kj</sub> – azot Kjeldahla, NO<sub>3</sub><sup>-</sup> – azotany



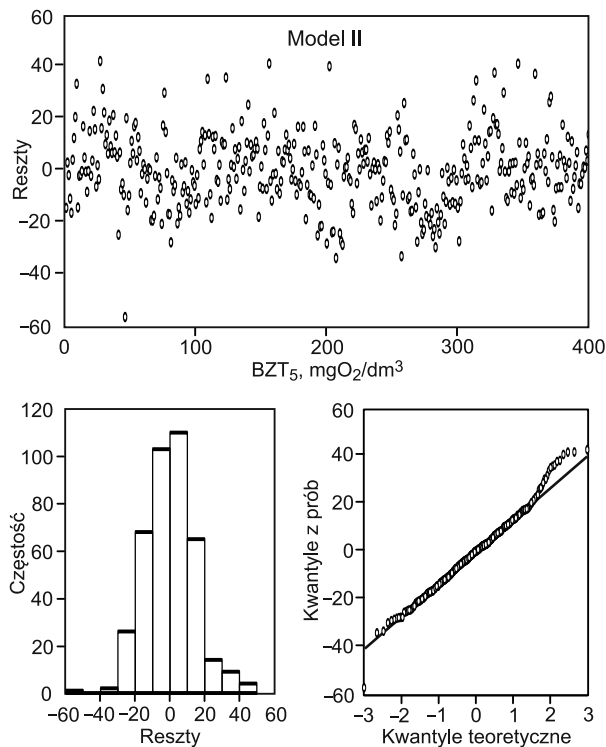
Na rysunkach 3–6 przedstawiono charakterystykę reszt – wykresy rozrzutu, histogramy oraz wykresy kwantyl-kwantyl. Rozrzut wartości reszt był podobny we wszystkich czterech modelach, punkty rozkładały się równomiernie, nie osiągając wartości większych niż ±58. Histogramy

były regularne, bez tendencji do skośności, jednak – jak wykazały testy – reszty (oprócz modelu I) nie miały rozkładu normalnego. Najlepsze dopasowanie na wykresie kwantyl-kwantyl wykazywał również model I, natomiast pozostałe modele miały tendencję do asymetrii prawostronnej.



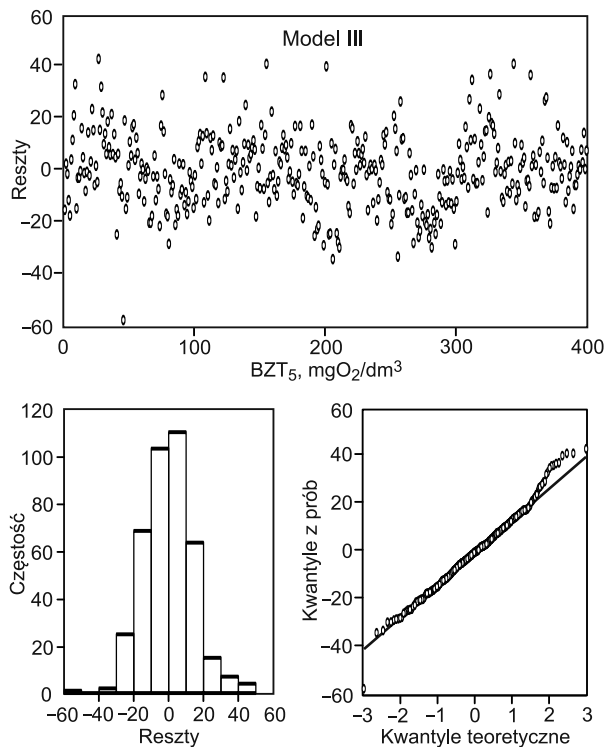
Rys. 3. Rozrzut, histogram i wykres kwantyl-kwantyl reszt modelu I

Fig. 3. Dispersion, histogram and quantile-quantile (Q-Q) plot of the model I residues



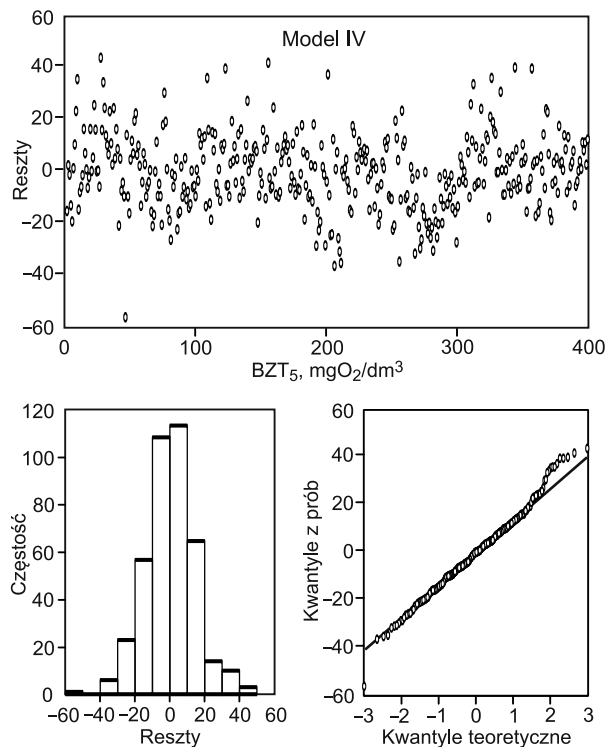
Rys. 4. Rozrzut, histogram i wykres kwantyl-kwantyl reszt modelu II

Fig. 4. Dispersion, histogram and quantile-quantile (Q-Q) plot of the model II residues



Rys. 5. Rozrzut, histogram i wykres kwantyl-kwantyl reszt modelu III

Fig 5. Dispersion, histogram and quantile-quantile (Q-Q) plot of the model III residues



Rys. 6. Rozrzut, histogram i wykres kwantyl-kwantyl reszt modelu IV

Fig. 6. Dispersion, histogram and quantile-quantile (Q-Q) plot of the model IV residues

## Weryfikacja modeli regresyjnych

Weryfikację czterech modeli regresyjnych stworzonych z użyciem danych z 1992 r. przeprowadzono wykorzystując wyniki pomiarów ze wszystkich lat (1990–2000). Wykresy rozrzutu i histogramy reszt czterech modeli w przykładowym 1995 r. przedstawiono na rysunku 7.

Wartości średniego błędu bezwzględnego (MAE) oszacowania BZT<sub>5</sub> ścieków według przyjętych modeli przedstawiono w tabeli 7. W przypadku danych pomiarowych pochodzących z lat 1990–1995 poszczególne modele charakteryzowały się najmniejszymi wartościami tego błędu (9,90÷16,94), natomiast pomiary wykonane w późniejszych latach dały już jego większe wartości (maksymalnie do 24,30). Najmniejszą wartością średniego błędu bezwzględnego w szacowaniu BZT<sub>5</sub> ścieków charakteryzował się model III oparty na danych z lat 1990–1995, natomiast dane z lat 1996–2000 umożliwiły najlepsze oszacowanie BZT<sub>5</sub> według modelu I.

Podczas doboru wskaźników jakości ścieków do budowy modelu stwierdzono, że trzy (spośród badanego zakresu jakości ścieków) standardowo oznaczane wskaźniki – zawiesiny ogólne, azot Kjeldahla i azot amonowy były wystarczające do sformułowania modelu matematycznego do szacowania wartości BZT<sub>5</sub>. Średni błąd bezwzględny (MAE) modelu IV wyniósł w przypadku danych z lat 1990–1995 około 13, a z lat 1996–2000 około 17. Zdecydowana większość oszacowanych wartości BZT<sub>5</sub> nie różniła się więcej niż o 25% od wartości rzeczywistych. Stwierdzono również, że model I, oparty wyłącznie na zawartości zawiesin ogólnych, również pozwolił na otrzymanie zadowalających wyników szacowania wartości BZT<sub>5</sub>.

Na podstawie wyników badań ukierunkowanych na sformułowanie modelu matematycznego do szacowania

Tabela 7. Średni błąd bezwzględny (MAE) szacowania wartości BZT<sub>5</sub> ścieków według sformułowanych modeli regresyjnych  
Table 7. Mean absolute error (MAE) of BOD<sub>5</sub> value estimation in wastewater according to formulated regression models

Rok	Średni błąd bezwzględny (MAE) modelu			
	I	II	III	IV
1990	16,20	9,94	9,90	10,15
1991	17,05	12,93	12,81	13,06
1992	14,66	10,94	10,94	11,13
1993	16,94	12,47	12,29	12,68
1994	14,82	13,01	13,84	14,25
1995	15,14	15,44	15,44	15,61
1996	20,15	16,73	16,69	16,63
1997	20,96	17,80	17,75	18,02
1998	18,06	20,47	20,42	21,18
1999	17,88	19,98	19,91	20,13
2000	16,23	23,36	23,41	24,30

wartości BZT<sub>5</sub> ścieków w warunkach eksploatacji oczyszczalni Terrence J. O'Brien oraz wykorzystując wyniki testów statystycznych stwierdzono, że bieżąca kontrola zmian wartości tego wskaźnika może być realizowana na podstawie modelu IV sformułowanego w postaci regresji liniowej:

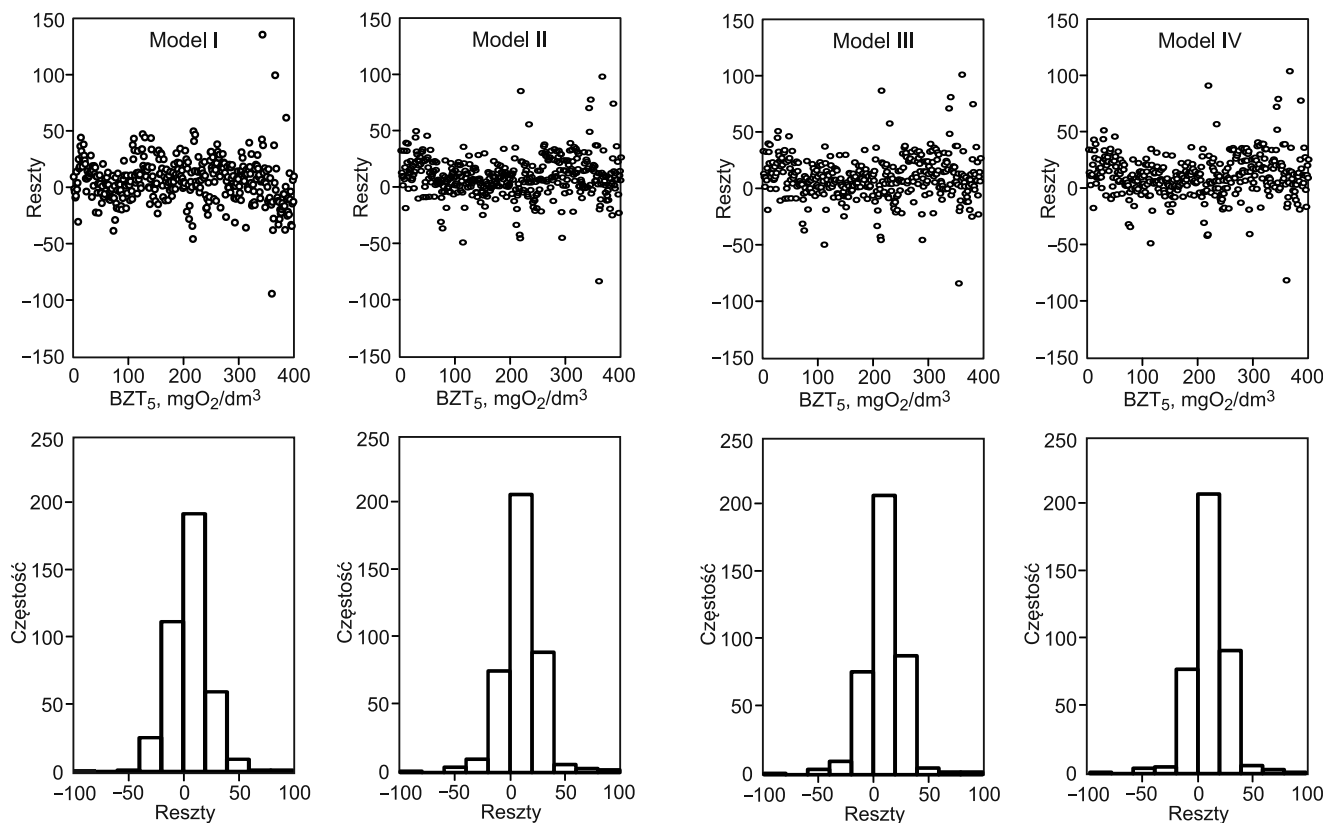
$$[BZT_5] = -2,28 + 0,28[Z] + 4,30[N_{Kj}] + 0,03[NH_4^+] \quad (2)$$

w którym:

[Z] – zawartość zawiesin ogólnych, mg/dm<sup>3</sup>

[N<sub>Kj</sub>] – zawartość azotu Kjeldahla, mgN/dm<sup>3</sup>

[NH<sub>4</sub><sup>+</sup>] – zawartość azotu amonowego, mgN/dm<sup>3</sup>



Rys. 7. Przykładowy rozrzut i histogram reszt modeli (dane z 1995r.)  
Fig. 7. The exemplary dispersion and histogram of model residues (data from 1995)

## Podsumowanie

W warunkach niestabilnej charakterystyki jakościowej i ilościowej ścieków dopływających do oczyszczalni, podjęcie właściwych decyzji eksploatacyjnych jest warunkiem zapewnienia wymaganych standardów jakości ścieków oczyszczonych i minimalizacji skutków zaburzenia warunków pracy oczyszczalni. Jednym z podstawowych składników ścieków komunalnych są zanieczyszczenia organiczne, których zawartość jest określana najczęściej na podstawie ich biochemicznego zapotrzebowania na tlen (BZT). Uwzględniając długi czas oczekiwania na wynik analizy (5 d, 7 d lub 20 d), zaleca się opracowanie szybkich metod oznaczania wartości tego wskaźnika, co pozwoliłoby na bardziej sprawne sterowanie warunkami technologicznymi w bioreaktorach i ich korektę w celu zachowania aktywności biocenozy osadu czynnego i wymaganej sprawności usuwania zanieczyszczeń.

Wykorzystując obszerną bazę danych pomiarowych jakości ścieków dopływających do oczyszczalni Terrence J. O'Brien WRP w Chicago, zbudowano prosty model regresyjny do szacowania wartości BZT<sub>5</sub> oczyszczanych ścieków. Model ten został opracowany na podstawie obliczeń wykonanych w programie statystycznym R, przy czym algorytm modelowania polegał na analizie korelacyjno-regresyjnej i sformułowaniu modelu w postaci wieloparametrowej regresji liniowej. Opracowany w warunkach tej oczyszczalni algorytm formułowania modelu regresyjnego przydatnego do szybkiego uzyskiwania przybliżonych wartości BZT może posłużyć do budowy podobnych modeli na potrzeby innych oczyszczalni ścieków, bez konieczności częstego wykonywania oznaczeń laboratoryjnych i długiego oczekiwania na ich wyniki.

## LITERATURA

1. Present concentration of BOD<sub>5</sub>, BOD<sub>7</sub> in rivers. European Environment Agency, 2009 (<https://www.eea.europa.eu/data-and-maps/figures/present-concentration-of-bod5-bod7-mg-o2-l-in-rivers>).
2. Biological Oxygen Demand, 7 days, BOD<sub>7</sub>, Swedish Pollutant Release and Transfer Register, Swedish Environmental Protection Agency, 2010 (<https://utslappisiffror.naturvardsverket.se/en/Substances/Inorganic-substances/Bod/>).
3. E. NEVEROVA-DZIOPAK, M. PREISNER: Analiza metod ustalania warunków wprowadzania ścieków komunalnych do odbiorników w wybranych państwach (Analysis of methods for determination of conditions of municipal wastewater discharge into recipients in selected countries). *Ochrona Środowiska* 2015, vol. 37, nr 1, ss. 3–9.
4. Teoreticheskie osnovy manometricheskogo metoda opredelenija BPK ([http://www.ecoinstrument.ru/service/public/chto\\_nuzhno\\_znat\\_o\\_bpk/](http://www.ecoinstrument.ru/service/public/chto_nuzhno_znat_o_bpk/)).
5. S. HEDDAM, H. LAMDA, S. FILALI: Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: A comparative study. *Environmental Processes* 2016, Vol. 3, No. 1, pp. 153–165.
6. S. JOUANNEAU, L. RECOULES, M.J. DURAND, A. BOUKABACHE, V. PICOT, Y. PRIMAULT, A. LAKEL, M. SENGELIN, B. BARILLON, G. THOUAND: Methods for assessing biochemical oxygen demand (BOD): A review. *Water Research* 2014, Vol. 49, pp. 62–82.
7. TERRENCE J. O'BRIEN WRP – Metropolitan Water Reclamation District of Greater Chicago (<https://www.mwrdd.org/irj/portal/anonymous?NavigationTarget=navurl://373e8fd63e3e6ce3c9fbec450abb3cfb>).
8. A. FORE: Innovation of traditional technology sparks MWRD to channel cleaner Chicago waterways. Metropolitan Water Reclamation District of Greater Chicago, Press Release 2016 ([https://www.mwrdd.org/irj/go/km/docs/documents/MWRD/internet/News&Media/Newsroom/Media/Press\\_Releases/2016/16\\_0324\\_0%27Brien\\_ribbon\\_cutting\\_PR.pdf](https://www.mwrdd.org/irj/go/km/docs/documents/MWRD/internet/News&Media/Newsroom/Media/Press_Releases/2016/16_0324_0%27Brien_ribbon_cutting_PR.pdf)).
9. R Development Core Team: R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna (Austria) 2011 (<http://www.gbif.org/resource/81287>).
10. C.J. WILLMOTT, K. MATSUURA: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 2005, Vol. 30, pp. 79–82.
11. T. CHAI, R. R. DRAXLER: Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development* 2014, Vol. 7, No. 3, pp. 1247–1250.
12. A. MUCIEK: Wyznaczanie modeli matematycznych z danych eksperymentalnych. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2012.
13. H. BOZDOGAN: Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 1987, Vol. 52, No. 3, pp. 345–370.

**Kowalewski, Z., Neverova-Dziopak, E., Preisner, M. An Attempt at Development of a Regression Model for Estimating BOD<sub>5</sub> Values of Municipal Wastewater. *Ochrona Środowiska* 2018, Vol. 40, No. 1, pp. 21–27.**

**Abstract:** One of the primary components of municipal sewage is organic pollution, the content of which is determined most often on the basis of its biochemical oxygen demand (BOD<sub>5</sub>). In the paper, a simple regression model was developed to estimate BOD<sub>5</sub> values of municipal sewage in order to ensure ongoing efficiency control of the biological treatment process and possibilities for optimizing the operating conditions of biological reactors. The model was developed from the long-term monitoring data (1990–2000) on wastewater quality in the Terrence J. O'Brien Water Reclamation Plant in Chicago (USA). Statistical analysis covered the following quality indicators for the sewage material flowing into the treatment plant: pH, BOD<sub>5</sub>, total suspended solids,

ammonium nitrogen, Kjeldahl nitrogen and nitrates. All calculations were made using a statistical program R with R Studio patch, version 1.0.143. Akaike Information Criterion (AIC) and the Leaps algorithm were employed to select parameters relevant to constructing the model. Based on the results of model testing, the regression model for BOD<sub>5</sub> values estimation in municipal wastewaters was formulated. The model employed parameters such as total suspended solids, Kjeldahl nitrogen and nitrates wastewater content. The algorithm of formulating the regression model that allows for quick generation of approximate BOD<sub>5</sub> values in wastewater can be applied to development of similar models for other treatment plants, with no need for frequent laboratory testing and long wait for the results.

**Keywords:** Municipal sewage, water reclamation plant, BOD<sub>5</sub>, total suspended solids, nitrogen compounds, modeling, estimation, regression model, operation conditions.