

Lidia BARTKIEWICZ, Jan STUDZIŃSKI

*Instytut Badań Systemowych PAN
Warszawa*

*Politechnika Świętokrzyska
Kielce*

PROGNOZOWANIE DOPŁYWU ŚCIEKÓW SUROWYCH DO OCZYSZCZALNI ZA POMOCĄ MODELOWANIA MATEMATYCZNEGO[†]

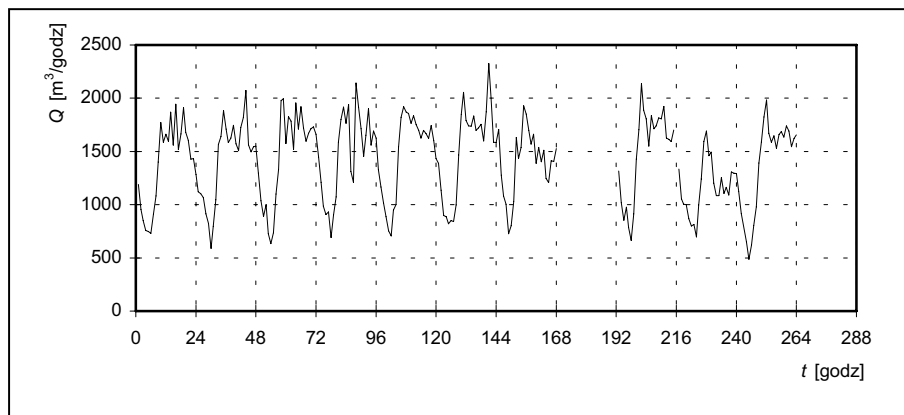
FORECASTING OF THE RAW SEWAGE INFLOWS WITH MATHEMATICAL MODELS

In the paper some results of mathematical modeling of the raw sewage are presented. The models created are used for forecasting the sewage inflow into the wastewater plants for better management of the sewage clarification process. The models are made with time series methods and for the comparison with neuronal nets. The own developed program library idol is used for creating the models.

1. Wprowadzenie

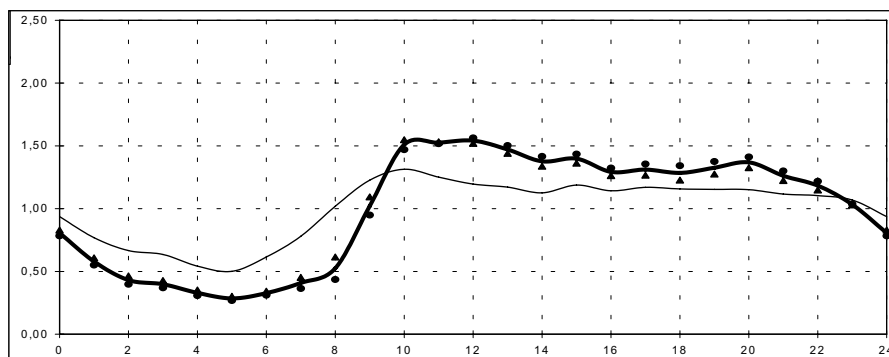
Oczyszczalnie ścieków często mają problemy z właściwym prowadzeniem procesu technologicznego z powodu dużych nierównomierności w dopływie ścieków surowych i związanych z tym gwałtownych i dużych zmian w dopływającym do oczyszczalni ładunku zanieczyszczeń. Przykładowe zmiany dopływu ścieków są pokazane na rysunku 1, natomiast zmieniający się wraz ze zmianami dopływu ładunek zanieczyszczeń jest pokazany na rysunku 2 [4]. Z wykresów widać, że dopływ ścieków ma w przeważającym stopniu charakter okresowy, co wiąże się z okresową pracą pompowni ścieków surowych, natomiast ładunek zanieczyszczeń zmienia się na ogół zgodnie ze zmianami dopływu ścieków. Ta obserwacja prowadzi do wniosku, że stosunkowo łatwe powinno być modelowanie dopływu ścieków i jednocześnie modele dopływu mogłyby być przydatne do prognozowania dopływającego ładunku zanieczyszczeń. Prognozowanie z odpowiednim wyprzedzeniem dopływu ścieków i ładunku zanieczyszczeń powinno z kolei ułatwić sterowanie oczyszczalnią, stwarzając operatorowi procesu możliwość odpowiednio wczesnego przygotowania obiektu na spodziewane zmiany parametrów wejściowych.

[†] Praca powstała w ramach projektu badawczego MNiSzW nr N514 2977 33.



Rys. 1. Przykładowy dopływ ścieków surowych do oczyszczalni

Fig. 1. Exemplary raw sewage inflow



Rys. 2. Przykładowe wykresy zmienności dopływu ścieków surowych (cienka linia) i zawartego w nich ładunku związków organicznych (gruba linia)

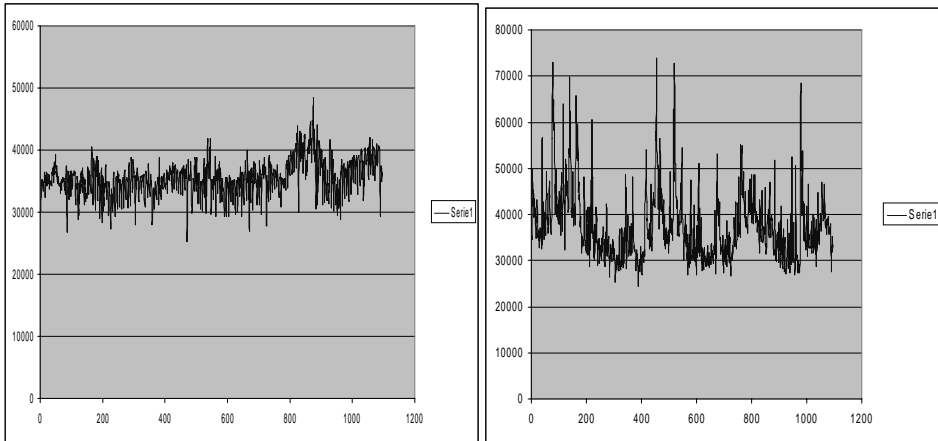
Fig. 2. Exemplary raw sewage inflow (thin line) and the included load of organic components (fat line)

Stąd powstała koncepcja wyznaczenia modeli prognostycznych dopływu ścieków surowych, które następnie byłyby wykorzystane do wspomagania sterowania oczyszczalnią na podstawie jej modelu fizycznego. Modele fizyczne oczyszczalni ścieków są opracowywane w Polsce już od kilkunastu lat [3, 4, 5], jednak służą obecnie jedynie do obliczeń symulacyjnych, głównie z powodu trudności z ich kalibracją i dopasowaniem do rzeczywistego obiektu oraz braku odpowiednich algorytmów sterowania. Wyznaczenie poprawnych modeli prognostycznych powinno ułatwić zmianę tej sytuacji.

W dalszym ciągu pokażemy kilka przykładów takich modeli wyznaczonych w postaci szeregów czasowych i sieci neuronowych na podstawie rzeczywistych danych pomiarowych, które dotyczą produkcji wody i produkcji ścieków w przykładowym przedsiębiorstwie wodociągowym [1].

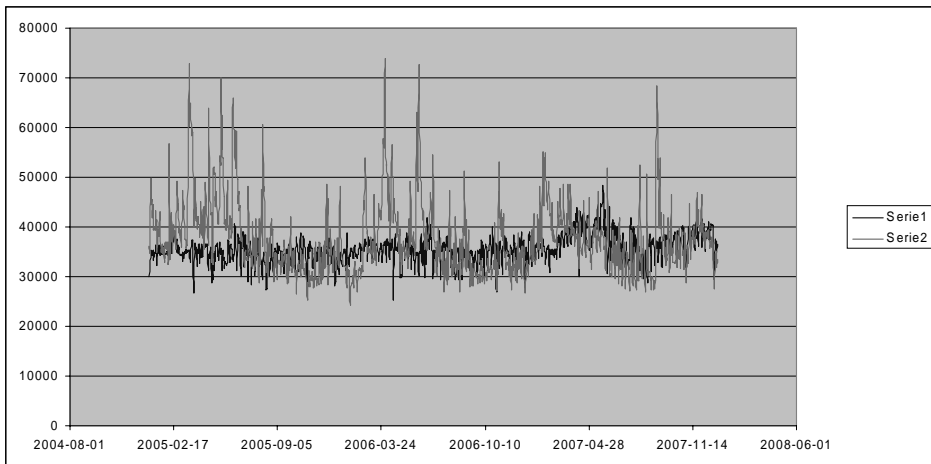
2. Dane wejściowe do modelowania

Danymi do modelowania są dobowe pomiary produkcji wody wprowadzanej do sieci wodociągowej oraz dobowe pomiary ścieków wprowadzanych do oczyszczalni, zbierane przez okres trzech lat. W rezultacie mamy do dyspozycji dwa ciągi danych o liczności 1095 elementów każdy, pokazane na rysunkach 3.



Rys. 3. *Produkcja wody (lewy rysunek) i ścieków w badanym przedsiębiorstwie wodociągowym*

Fig. 3. *Water (left) and wastewater (right) production in the waterworks investigated*



Rys. 4. *Produkcja wody (mniejsze wartości) i ścieków w badanym przedsiębiorstwie*

Fig. 4. *Water (smaller values) and wastewater production in the waterworks*

Istotnymi cechami danych są, ich okresowość oraz widoczna przewaga ilości ścieków nad ilością produkowanej wody. Widać to wyraźnie na rysunku 4, gdzie oba ciągi danych są na siebie nałożone. Okresowość w produkcji wody jest spowodowana zmiennością jej zużycia w cyklu tygodniowym, co w konsekwencji wpływa na okresowość produkcji ścieków. Ponieważ dane dotyczą pomiarów dobowych a nie godzinowych, więc nie występują w nich szybsze okresowości związane ze zmiennym rozbiorem wody, a więc i produkcją ścieków, w cyklu dobowym. Ilości ścieków większe od ilości produkowanej wody są wynikiem przede wszystkim opadów deszczowych, które nie były monitorowane. Brak tych danych powoduje, że w ciągu pomiarowym dotyczącym ścieków należy wyeliminować wpływ opadów względnie uwzględnić go w ciągu pomiarowym dotyczącym produkcji wody. Nie jest możliwe zrobienie tego w sposób dokładny bez odpowiednich informacji o opadach, dlatego postanowiono wyeliminować wpływ opadów w pomiarach ścieków w sposób przybliżony poprzez wygładzenie danych.

3. Metody modelowania

Modele prognostyczne dopływu ścieków wyznaczano głównie z zastosowaniem postaci opisu modeli stochastycznych klasy ARMA (*autoregression moving average*). Do ich identyfikacji stosuje się zwykle tak zwane metody *szeregów czasowych*, oparte na metodzie najmniejszej sumy kwadratów. W Instytucie Badań Systemowych PAN opracowano bibliotekę programów komputerowych IDOL (*Identyfikacja Dynamicznych Obiektów Liniowych*), służących do wyznaczania modeli złożonych procesów dynamicznych [2]. Programy zawierają podstawowe metody identyfikacji liniowych i nieliniowych modeli dyskretnych, w tym: metodę najmniejszej sumy kwadratów Kalmana, metodę uogólnionej sumy kwadratów Clarke'a, metodę największej wiarygodności, metodę regresji liniowej i metodę regresji nieliniowej. Ponadto biblioteka zawiera programy do wygładzania danych pomiarowych algorytmami średniej ruchomej i wielomianów aproksymacyjnych, a także zawiera algorytmy oceny statystycznej wyznaczanych modeli.

Badane modele prognostyczne dopływu ścieków są dynamicznymi modelami liniowymi. Przyjęta w programach biblioteki postać opisu takich modeli jest w ogólności następująca

$$y_n = -A(z^{-1})y_n + \sum_{i=1}^M B_i(z^{-1})u_{in} + v_n$$

gdzie: M – liczba wejść modelu, $A(z^{-1}), B_i(z^{-1})$ – operatory różnicowe wyjścia i wejść modelu: $A(z^{-1}) = a_1 z^{-1} + \dots + a_R z^{-R}$ i $B_i(z^{-1}) = b_{i0} + b_{i1} z^{-1} + \dots + b_{iP_i} z^{-P_i}$; R, P_i – rzędy operatorów $A(z^{-1})$ i $B_i(z^{-1})$, v_n – zakłócenie skorelowane.

Z kolei w przypadku modelu różnicowego typu ARMA (*autoregressive moving average*) każdy bieżący pomiar chwilowy jest traktowany jak funkcja pewnej liczby poprzednich pomiarów z dodanym zakłóceniem oznaczającym błąd pomiaru. Liczba uwzględnionych poprzednich pomiarów oznacza rząd modelu. Opis modelu jest postaci

$$y_n = a_1 y_{n-1} + a_2 y_{n-2} + \dots + a_R y_{n-R} + v_n$$

gdzie R jest rzędem modelu, a_i jest parametrem modelu a v_n oznacza błąd pomiaru. Wyznaczenie modelu różnicowego polega na wyznaczeniu jego rzędu oraz obliczeniu wartości parametrów.

Doświadczenia ze stosowania biblioteki IDOL wykazują, że w przypadku modeli dynamicznych metodą najbardziej niezawodną jest stosunkowo wolna metoda największej wiarygodności. Z kolei metoda Kalmana jest najszybsza, dlatego jej właśnie użyto do modelowania dopływu ścieków. Metoda Clarke'a jest dokładniejsza od metody Kalmana i szybsza od metody największej wiarygodności, jednak bardziej od niej zawodna, to znaczy może zbiegać do złych rozwiązań, tak zwanych *optimów lokalnych*, w przypadku bardziej złożonych modeli. W metodzie Kalmana nie stosuje się dodatkowego opisu dla zakłócenia skorelowanego v_n . Wyznaczony estymator parametrów w modelu procesu jest asymptotycznie nieobciążony dla metody Clarke'a i największej wiarygodności i jest on asymptotycznie obciążony dla metody Kalmana.

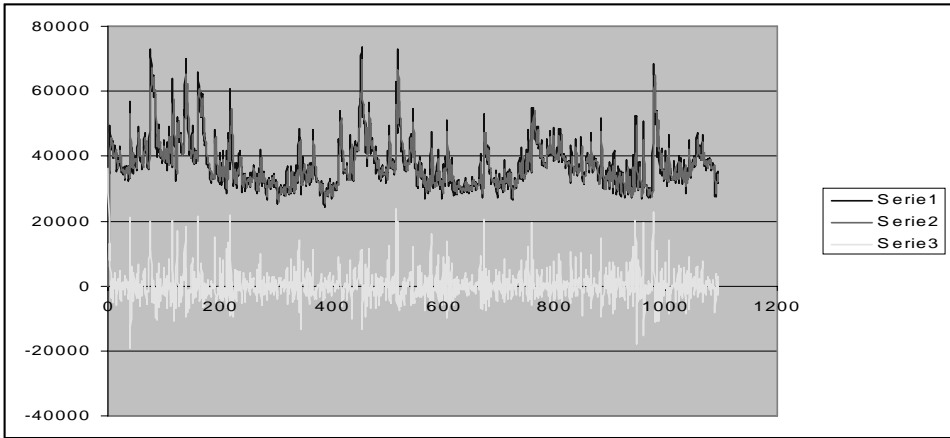
Drugi sposób modelowania dopływu ścieków polegał na zastosowaniu sieci neuronowych. Zastosowano sieć typu BP (*back propagation feedforward network*), składającą się z 3 warstw: wejściowej, ukrytej i wyjściowej, oraz mającej liniowe funkcje przejścia [4]. Wyznaczano modele przy użyciu tych samych danych pomiarowych, na podstawie których wyznaczono modele szeregów czasowych za pomocą metody Kalmana. Obliczenia modelowania za pomocą sieci neuronowych wykonano przy użyciu pakietu obliczeniowego MATLAB.

4. Wyniki modelowania

Pierwszy etap modelowania polegał na wyznaczeniu modelu autoregresji jedynie na podstawie danych dotyczących ścieków surowych, bez ich wygładzenia (Model 1 w tabeli 1). Wyznaczony model jest trzeciego rzędu z trzema parametrami a_i , przy czym wszystkie parametry są istotne (ich wariancje d_i są mniejsze od odpowiednich wartości parametrów). Jednak błąd modelowania jest stosunkowo duży, co widać na rysunku 5 (dolny wykres) a także po dużej wartości sumy resztowej SS i stosunkowo małej wartości współczynnika korelacji RR między wyjściem modelu i pomiarami.

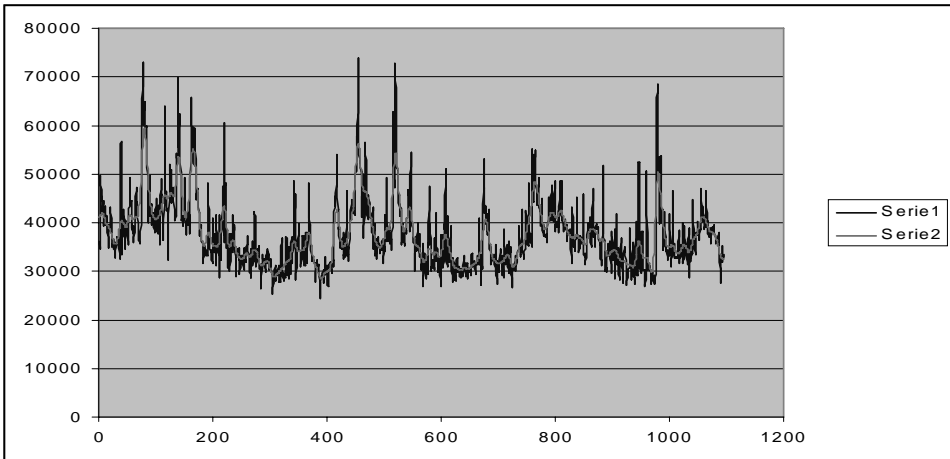
Dlatego w kolejnym kroku wygładzono pomiary ścieków za pomocą algorytmu średniej ruchomej, eliminując w ten sposób z danych piki odpowiedzialne za opady deszczowe. Przyjęto szerokość okna równą siedmiu pomiarom i proces wygładzania wykonano trzy razy. Wyniki wygładzania są widoczne na rysunku 6.

Kolejny model był również modelem autoregresji wyznaczonym dla wygładzonych pomiarów ścieków (Model 2). Wyniki modelowania zamieszczono na rysunku 7 i również w tabeli 1. Widać obecnie, że wygładzanie znacznie poprawiło wyniki modelowania, ponieważ suma resztowa modelu SS jest znacznie mniejsza, a wyznaczony współczynnik korelacji RR jest znacznie większy, niż poprzednio. Wyznaczony model jest także trzeciego rzędu, przy czym drugi parametr w modelu a_2 jest praktycznie nieistotny ze względu na dużą wariancję w stosunku do jego wartości (wartości wariancji i parametru są tego samego rzędu). Szczególnie na rysunku 7 widać wyraźnie, jak dobrze model dopasowuje się do pomiarów i jednocześnie jak niewielki jest błąd modelowania.



Rys. 5. Model autoregresji dopływu ścieków bez wygładzenia danych (Model 1; seria 3 – błąd modelowania)

Fig. 5. Autoregression model of the sewage inflow without smoothing the data (Model 1; serie 3 – model error)

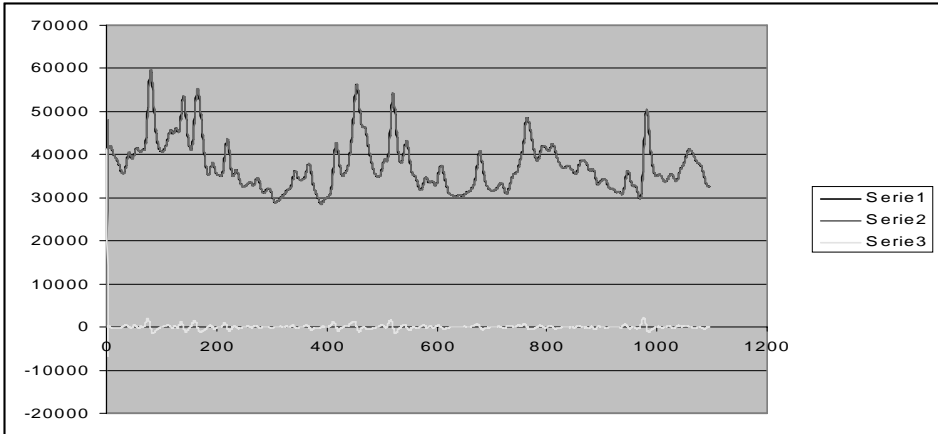


Rys. 6. Ścieki surowe przed (seria 1) i po wygładzeniu (seria 2)

Fig. 6. Raw sewage inflow before (serie 1) and after smoothing (serie 2) the data

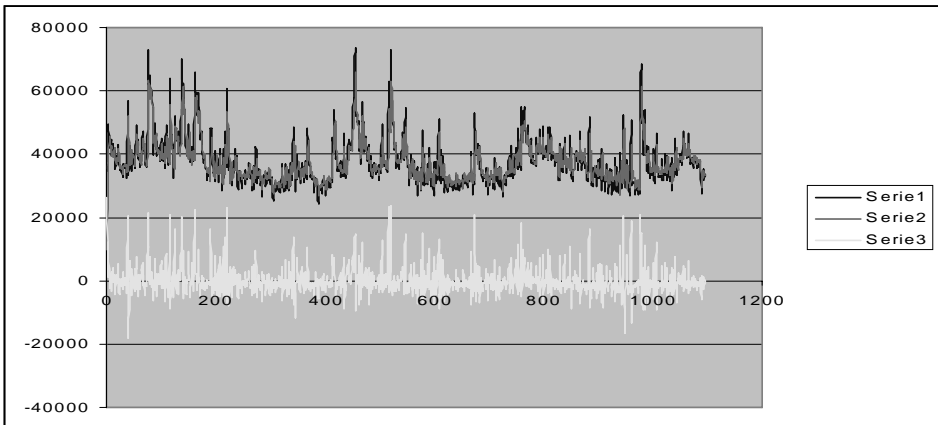
W kolejnych krokach modelowania wyznaczono modele dopływu ścieków w funkcji produkcji wody, chcąc sprawdzić, czy istnieje funkcyjna zależność między tymi zmiennymi. Podobnie jak poprzednio, korzystano z metody Kalmana użytej tym razem do modelu z jednym wejściem i jednym wyjściem. Podobnie też jak poprzednio, w przypadku pomiarów ścieków surowych stosowano dane niewygładzone i dane wygładzone za pomocą algorytmu średniej ruchomej.

Wyniki obliczeń zamieszczono w tabeli 1 i na rysunkach 8 i 9. Wyniki te w pewnym sensie są podobne do otrzymanych wcześniej. Model dla danych niewygładzonych (Model 3) jest gorszy od modelu dla danych wygładzonych (Model 4). Jednocześnie jednak modele z uwzględnieniem produkcji wody są znacznie lepsze od modeli autoregresji, korzystających jedynie z pomiarów ścieków.



Rys. 7. Model autoregresji dopływu ścieków po wygładzeniu danych (Model 2; seria 3 – błąd modelowania)

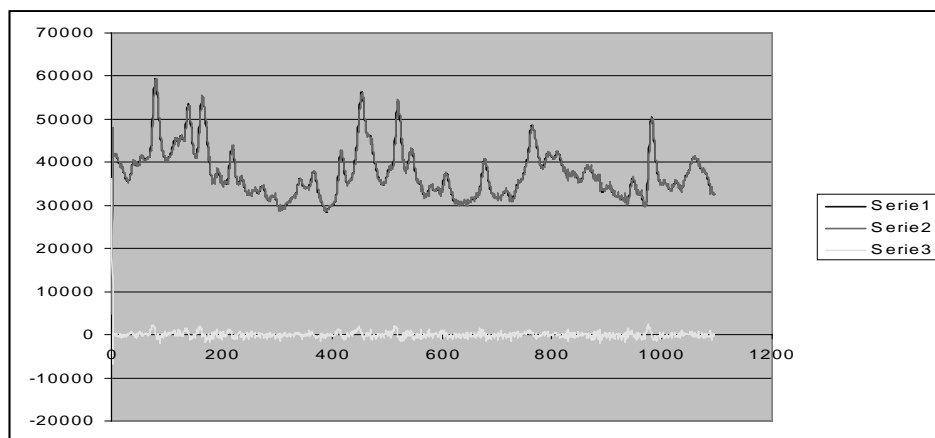
Fig. 7. Autoregression model of the sewage inflow after smoothing the data (Model 1; serie 3 – model error)



Rys. 8. Model dopływu ścieków w funkcji produkcji wody bez wygładzenia danych (Model 3; seria 3 – błąd modelowania)

Fig. 8. Model of the sewage inflow depending on the water production without smoothing the data (Model 3; serie 3 – model error)

Model 3 jest modelem trzeciego rzędu, w którym w operatorze wyjścia wszystkie parametry a_i są istotne, natomiast w operatorze wejścia naprawdę istotny jest jedynie parametr wolny b_0 , natomiast pozostałe parametry b_i są na granicy istotności. W przypadku Modelu 4, również trzeciego rzędu, w operatorze wyjścia drugi parametr a_2 jest nieistotny, podobnie jak w Modelu 2, natomiast w operatorze wejścia na granicy istotności są parametry b_2 i b_3 , podobnie jak w Modelu 3. Niemniej Model 4 wydaje się bardzo dobry, ponieważ suma resztowa SS jest już bardzo mała a współczynnik korelacji RR jest praktycznie równy 1, czy osiąga wartość maksymalną. Można więc powiedzieć, że stosunkowo prosty model dobrze opisuje zależność między ilością produkowanej wody a ilością produkowanych ścieków, bez uwzględnienia jednak dodatkowych ścieków tworzonych przez opady.



Rys. 9. Model dopływu ścieków w funkcji produkcji wody z wygładzeniem danych (Model 4; seria 3 – błąd modelowania)

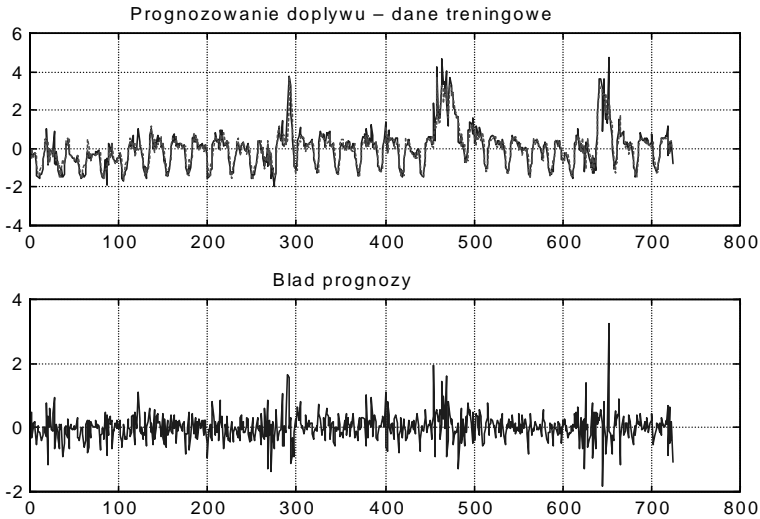
Fig. 9. Model of the sewage inflow depending on the water production after smoothing the data (Model 4; serie 3 – model error)

Tab. 1. Wyznaczone modele dopływu ścieków surowych; SS – suma resztowa; RR – współczynnik korelacji pomiarów z wyjściem modelu

Tab. 1. Calculated models of the raw sewage inflow; SS – residuum value, RR - correlation coefficient

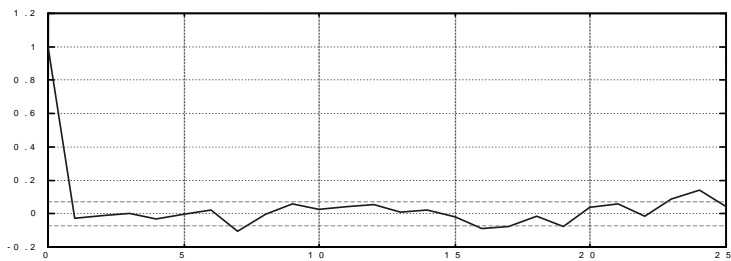
	Model 1	Model 2	Model 3	Model 4
a_1/d_1	-0.72 / 0.03	-1.17 / 0.03	-0.65 / 0.03	-1.14 / 0.03
a_2/d_2	-0.12 / 0.04	0.03 / 0.05	-0.10 / 0.04	0.00 / 0.04
a_3/d_3	-0.14 / 0.03	0.14 / 0.03	-0.09 / 0.03	0.16 / 0.03
b_0/d_0			0.33 / 0.05	0.17 / 0.02
b_1/d_1			-0.03 / 0.07	-0.14 / 0.02
b_2/d_2			-0.06 / 0.06	-0.02 / 0.02
b_3/d_3			-0.06 / 0.05	0.02 / 0.02
SS	4820	1350	4616	1275
RR	0.897	0.974	0.658	0.999

Oprócz wyznaczonych wyżej modeli różnicowych wykonano również obliczenia dodatkowe dla modeli neuronowych, chcąc sprawdzić przydatność również takiego sposobu modelowania do prognozowania dopływu ścieków surowych. Zastosowano model typu NNAR (*neural net autoregressive model*), posiadający 5 neuronów na warstwie wejściowej, 7 neuronów na warstwie ukrytej i 1 neuron na warstwie wyjściowej [4]. Odpowiada on modelowi różnicowemu autoregresji piątego rzędu. Przed przystąpieniem do obliczeń dokonano skalowania danych pomiarowych, sprowadzając je do przedziału $<-1, 1>$ z wartością średnią 0 i wariancją 1. Takiego skalowania nie robiono w przypadku modeli różnicowych ARMA. Przy wyznaczaniu modelu neuronowego zastosowano algorytm badania istotności połączeń między neuronami i usuwania połączeń nieistotnych (*pruning*). Dane pomiarowych przy tworzeniu modeli i prognozowaniu, czyli w fazie uczenia i fazie testowania sieci neuronowej, używano w identyczny sposób, jak w modelu różnicowym, to znaczy w przesuwającym się oknie pomiarowym o zadanej długości pierwszy pomiar oznaczał wyjście sieci a pozostałe pomiary oznaczały wejścia sieci.



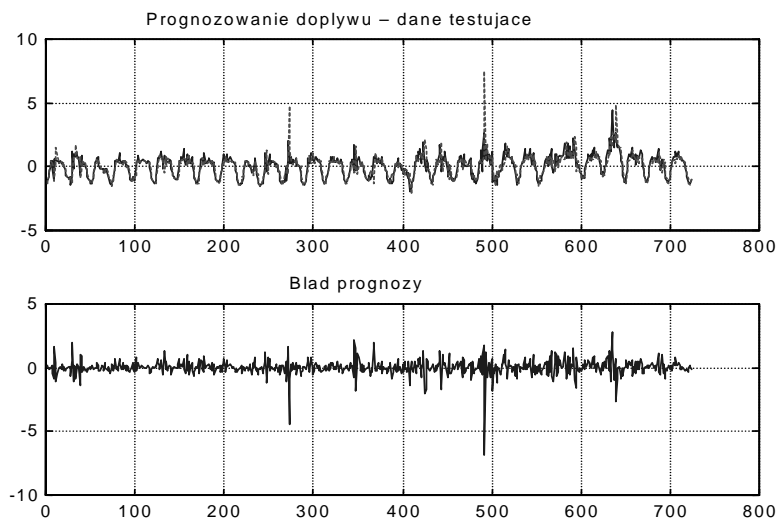
Rys. 10. Model neuronowy dopływu ścieków (rysunek górny) i błąd prognozowania (rysunek dolny); dane treningowe

Fig. 10. Neuronal model of the sewage inflow (on top) and the forecasting error (below); training data



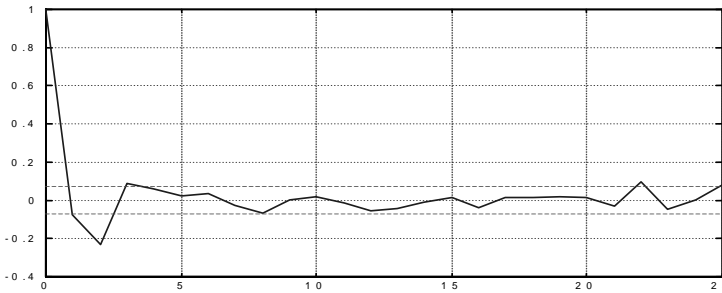
Rys. 11. Funkcja autokorelacji dla błędu prognozowania modelu neuronowego; dane treningowe

Fig. 11. Autocorrelation function for the forecasting error of the neuronal model; training data



Rys. 12. Model neuronowy dopływu ścieków (rysunek górny) i błąd prognozowania (rysunek dolny); dane testujące

Fig. 12. Neuronal model of the sewage inflow (on top) and the forecasting error (below); testing data



Rys. 13. Funkcja autokorelacji dla błędu prognozowania modelu neuronowego; dane testujące

Fig. 13. Autocorrelation function for the forecasting error of the neuronal model; testing data

Na rysunkach 10 – 13 pokazano przykładowe wyniki modelowania z użyciem sieci neuronowych. Ocena otrzymanych wyników jest pozytywna, ponieważ widać na rysunkach, że model neuronowy wyznacza poprawne prognozy dopływu ścieków do oczyszczalni. Na rysunkach 11 i 13 pokazano, że błąd prognozowania z użyciem modelu neuronowego ma charakter białego szumu. Obliczenia wskazują, że sieci neuronowe jako modele prognostyczne są konkurencyjne w stosunku do klasycznych modeli różnicowych, chociaż trzeba jednocześnie zaznaczyć, że wyznaczanie i stosowanie modeli różnicowych jest prostsze niż w przypadku modeli neuronowych.

5. Uwagi końcowe

W pracy przedstawiono pewne wyniki wyznaczania modeli matematycznych przeznaczonych do prognozowania dopływu ścieków surowych do oczyszczalni. Wyznaczone modele są opisane równaniami różnicowymi względnie sieciami neuronowymi. Wyznaczano modele autoregresji, w których prognoza dopływu odbywa się jedynie na podstawie wcześniejszych jego wartości, oraz modele, w których prognozowany dopływ jest funkcją produkcji wody. Otrzymane wyniki wskazują, że zarówno równania różnicowe, jak i sieci neuronowe są w stanie dobrze opisać modelowane procesy a wyznaczone modele poprawnie prognozują dopływ ścieków, chociaż modele autoregresji są mniej dokładne, niż modele uwzględniające zależność ilości ścieków od ilości produkowanej wody. Jednocześnie modele neuronowe są bardziej złożone i mniej wygodne w eksploatacji, niż klasyczne modele różnicowe.

Problemem w obliczeniach był fakt, że nie posiadano danych pomiarowych dotyczących opadów deszczowych w okresie zbierania pomiarów. Stąd istniała niezgodność między danymi dotyczącymi ilości produkowanej wody i danymi dotyczącymi ilości ścieków dopływających do oczyszczalni. W rezultacie modele tworzone dla takich ‘niezgodnych’ danych były mniej dokładne, niż modele tworzone dla danych, w których poprzez wygładzanie wyeliminowano wpływ opadów. Wydaje się, że wyniki modelowania

wania będzie można znacznie poprawić, jeżeli będzie się dysponować pomiarami opadów deszczowych i uwzględni się je w danych dotyczących produkcji wody.

W wyznaczonych modelach różnicowych część parametrów była nieistotna względnie mało istotna, co świadczy o okresowości badanych procesów. Nieuwzględnienie tych parametrów w modelach powinno również poprawić wyniki modelowania.

Wykonane obliczenia miały charakter rozpoznawczy i są wstępem do dalszych badań związanych z opracowaniem algorytmów do optymalnego sterowania oczyszczalnią ścieków. Prognozowanie dopływu ścieków i konsekwencji również ładunku zanieczyszczeń dopływających do oczyszczalni powinno w koncepcji autorów znacznie usprawnić sterowanie złożonym procesem oczyszczania ścieków.

Bibliografia

- [1] Bartkiewicz L. *Modelowanie ilości ścieków dopływających do oczyszczalni*. INSTAL, 2008 (w druku)
- [2] Licznar P., Łomotowski J., Rojek M. i M., Studziński J. *Wybrane zagadnienia z zakresu pomiaru i metod opracowywania danych z automatycznych stacji meteorologicznych*. Monografie AR we Wrocławiu, XXV, Wrocław 2001
- [3] Stachura M. J. *Komputerowa symulacja i optymalizacja oczyszczalni ścieków*. Warszawa: PAN IBS, Seria Badania Systemowe, 2008 (w druku)
- [4] Studziński J. *Identyfikacja, symulacja i sterowanie oczyszczalniami ścieków*. Warszawa: PAN IBS, Seria Badania Systemowe, 2004
- [5] Szetela W. *Model dynamiczny oczyszczalni ścieków z osadem czynnym*. Wydawnictwo Politechniki Wrocławskiej, Wrocław 1990