

Marcin Stachura, Jan Studziński

## Prognozowanie obciążenia hydraulicznego miejskiego systemu wodociągowego z wykorzystaniem modeli rozmytych typu TSK

Racjonalna eksploatacja systemu zaopatrzenia w wodę powinna być oparta na prognozowaniu obciążenia hydraulicznego sieci, przy uwzględnieniu wpływu zmiennych czynników losowych. Prognozowanie polega na wykrzystaniu znanych w chwili obecnej i chwilach przeszłych wartości opisywanych zmiennych w celu przewidywania wartości tych lub innych zmiennych w przyszłości. Do analizy i prognozowania szeregów czasowych poboru wody w miejskich systemach zaopatrzenia w wodę mogą być wykorzystane modele różnego typu. Opisują one proces na podstawie wartości pomiarów wejścia i wyjścia z obiektu (obciążonych szumem) z wykorzystaniem metod interpolacyjnych, aproksymacyjnych, analizy szeregów czasowych, modeli neuronowych itp. [1, 2]. Mogą być to modele tzw. czarnej skrzynki, które wyznacza się przy wykorzystaniu metod parametrycznych, np. ARIMA lub ARIMAX, metod wygładzania szeregów czasowych, sztucznych sieci neuronowych (SSN) oraz modeli rozmytych. Model poboru wody, który z zadowalającą dokładnością odwzorowuje zarejestrowane pomiary może być wykorzystany między innymi do kalibracji modeli przepływu w systemach dystrybucji wody oraz komputerowej symulacji działania tych systemów [3], optymalnego sterowania dystrybucją wody i procesami jej oczyszczania [4, 5], określania ilości ścieków powstających w aglomeracji, a co za tym idzie – poprawy pracy miejskiej oczyszczalni [5, 6] oraz do wykrywania wycieków wody [7].

### Sformułowanie problemu

Celem badań było przedstawienie możliwości prognozowania dobowego poboru wody na podstawie jego wartości archiwalnych, czyli wyznaczenie modelu w postaci:

$$p_w(k+1) = f(p_w(k), p_w(k-1), \dots, p_w(k-r)) \quad (1)$$

w którym:

$p_w$  – dobowy pobór wody,  $m^3/d$

$k$  – dyskretna chwila czasowa

$r$  – rząd modelu

W ostatnich latach popularne stało się wykorzystywanie w tym celu sztucznych sieci neuronowych (SSN). Prognozowanie z użyciem SSN daje dobre rezultaty, charakteryzuje się jednak pewnymi wadami, do których można zaliczyć [8]:

– brak możliwości analizy struktury wyznaczonego modelu; zapis sztucznej sieci neuronowej zakodowany jest w niezrozumiałej formie,

– sztuczne sieci neuronowe mają ograniczone zdolności ekstrapolacji poza zakres danych, na podstawie których zostały wyznaczone.

Inną grupą modeli, intensywnie badaną i rozwijaną w ostatnich latach, są modele oparte na logice rozmytej. Ma ona zdolność opisu obiektu w bardzo ogólnej i zrozumiałej formie, co w połączeniu ze swobodą doboru nieliniowości oraz możliwościami opracowania dla niej algorytmów identyfikacyjnych daje interesujący typ modelu. Model rozmyty jest zapisywany w postaci ogólnej reguły: jeżeli <stwierdzenie rozmyte> to <stwierdzenie rozmyte>, przy czym stwierdzenie rozmyte określone jest na zbiorze funkcji przynależności zdefiniowanych na obszarze zmiennych fizycznych połączonych operatorami rozmytymi. Stwierdzenie występujące jako pierwsze jest przesłanką bądź poprzednikiem, drugie – konkluzją bądź następnikiem.

W niniejszej pracy sprawdzono możliwość wykorzystania modeli o strukturze Takagi-Sugeno-Kanga (TSK), w których następnikiem jest klasyczny model liniowy uwzględniający dynamikę, co pozwala na określenie struktury jako „modele liniowe strefami rozmywane” [9]. Tego typu model, w porównaniu do wcześniej wymienionych, charakteryzuje się następującymi interesującymi cechami:

– dzięki uwzględnieniu dynamiki obiektu w funkcji następnika, modele TSK mogą dobrze naśladować dany proces, nawet przy niewielkiej liczbie przedziałów rozmywania,

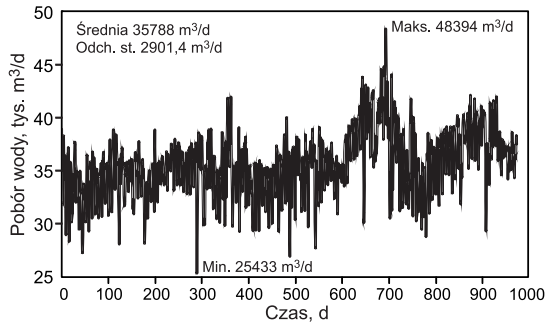
– następnikiem jest klasyczny model liniowy, co pozwala traktować nieliniowy model rozmyty jako zbiór lokalnych modeli liniowych; daje to możliwość stosowania bezpośrednio teorii sterowania opracowanej dla obiektów liniowych przy syntezie regulatora,

– bardzo prosta jest linearyzacja tego modelu w dowolnym punkcie pracy,

– modele liniowe są dobrze przebadane i szeroko znane, co umożliwia porównanie ich podstawowych właściwości (opóźnienia, stałe czasowe, rząd) z przesłankami wynikającymi z analizy teoretycznej zidentyfikowanego procesu [10].

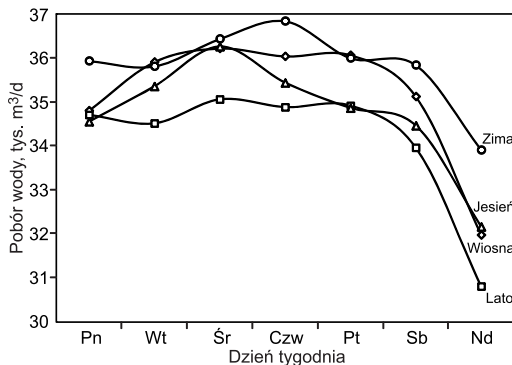
### Badane szeregi czasowe

Do przeprowadzenia badań związanych z modelowaniem wykorzystano szeregi czasowe zarejestrowanych wartości dobowego poboru wody (rys. 1). Dane pozyskano z przedsiębiorstwa wodociągowego w Rzeszowie, pomiary wykonano w czasie od 01-07-2005 do 29-02-2008, a liczebność zbioru danych wynosiła 974.



Rys. 1. Zarejestrowane wartości dobowego poboru wody  
Fig. 1. Recorded values of daily water demand

W analizowanym systemie wodociągowym zdecydowanie największa ilość wody pobierana jest przez gospodarstwa domowe (73,1%), a tylko 11,5% przez przedsiębiorstwa przemysłowe. Pobór wody przez mieszkańców ma zazwyczaj charakter cykliczny, więc uznano, że możliwe jest wyznaczenie przyszłych jej wartości na podstawie zarejestrowanych sygnałów z chwil przeszłych. Na podstawie zarejestrowanych danych wyznaczono dobowe wartości poboru wody w poszczególnych dniach tygodnia w różnych porach roku (rys. 2). Z danych tych wynika, że w ciągu całego okresu badawczego najmniejszy pobór wody występował w niedziele oraz że największy sumaryczny pobór wody miał miejsce w zimie, a najmniejszy w lecie (pobór wody jesienią i wiosną był porównywalny).



Rys. 2. Tygodniowy pobór wody w różnych porach roku  
Fig. 2. Weekly water demand in different seasons

## Model rozmyty Takagi-Sugeno-Kanga

Model TSK zalicza się do ogólnej grupy modeli rozmytych. Schemat przetwarzania takich modeli można w uproszczeniu opisać następująco: zmierzone wartości wejściowe są poddawane procesowi rozmywania, a następnie przeprowadzane jest wnioskowanie rozmyte w oparciu o zestaw reguł wchodzących w skład modelu. Na koniec następuje proces wyostrzenia i wyliczenie konkretnej – ostrej – wartości wyjścia modelu. Charakterystyką modelu TSK jest to, że w konkluzji każdej reguły występuje w miejsce zbioru rozmytego analityczna funkcja zmiennych wejściowych. Tak więc każda reguła modelu TSK przyjmuje następującą postać:

$$R_i: \text{jeżeli } [x_1 = A_1^i \wedge \dots \wedge x_n = A_n^i] \text{ to } [y = f_i(u)] \quad (2)$$

w której:  $R_i$  –  $i$ -ta reguła wiedzy,  $x_j$  –  $j$ -te wejście rozmywane,  $A_j^i$  –  $j$ -ta partycja zapisana w  $i$ -tej regule modelu,  $n$  – liczba wejść rozmywanych,  $y$  – wyjście modelu,  $f_i$  – funkcja następnika (konkluzja)  $i$ -tej reguły,  $u$  – wektor wejść modelu liniowego (funkcji następnika).

Model składa się z  $M$  reguł. Rozmyte przesłanki tych reguł dzielą przestrzeń wejściową na określoną liczbę rozmytych regionów, natomiast funkcje następnika opisują zachowanie się modelu w tych regionach. Przesłanka jest zbudowana z jednego lub więcej stwierdzeń rozmytych połączonych rozmytym operatorem koniunkcji. Zmienne  $x_1, x_2, \dots, x_n$  tworzą  $n$ -wymiarowy wektor wejść rozmywanych. Wypadkowa wartość wyjścia jest obliczana jako ważona suma wyjść poszczególnych reguł, gdzie wagą jest wypadkowy stopień przynależności danej reguły  $\mu_A^i(x)$ :

$$y = \sum_{i=1}^M \mu_A^i(x) f_i(u) \quad (3)$$

w której  $\mu_A^i(x)$  to spełnienie  $i$ -tej reguły, określone jako iloczyn spełnienia poszczególnych jej przesłanek:

$$\mu_A^i(x) = \prod_{j=1}^n \mu_{A_j^i}(x_j) \quad (4)$$

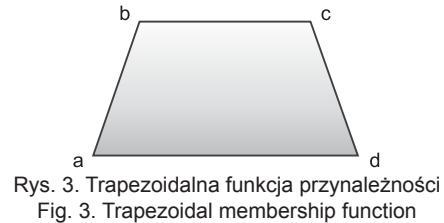
Takie podejście upraszcza sposób obliczania wyjścia z modelu, dzięki wyeliminowaniu procesu wyostrzenia. Najczęściej wykorzystywaną postacią funkcji  $f_i$  jest wielomian I-stopnia składowych wektora  $u$ :

$$f_i(u) = a_0^i + a_1^i u_1 + \dots + a_L^i u_L \quad (5)$$

Wektor  $x$  definiuje punkt pracy modelu, natomiast wektor  $u$  zawiera informacje potrzebne do obliczenia jego dynamicznej odpowiedzi. Jako funkcje przynależności wykorzystane zostały funkcje trapezoidalne:

$$\mu_A(x) = \max \left( 0, \min \left( 1, \frac{x-a}{b-a}, \frac{d-x}{d-c} \right) \right) \quad (6)$$

w których:  $a, b, c, d$  – punkty charakterystyczne funkcji trapezoidalnej ( $d > c, b > a$ ) (rys. 3):



Rys. 3. Trapezoidalna funkcja przynależności  
Fig. 3. Trapezoidal membership function

## Wyznaczanie modeli TSK

Identyfikacja modelu TSK jest procesem iteracyjnym, który można zawrzeć w dwóch fazach [11]:

1. Określenie wszystkich parametrów rozmywania, tj. zbioru sygnałów rozmywanych, parametrów poszczególnych funkcji przynależności itd. Zadanie to wykonane zostało przy wykorzystaniu algorytmów ewolucyjnych [12].

2. Estymacja współczynników  $a_0^i, a_1^i, \dots, a_L^i$  wielomianu (5) każdej ( $i$ -tej) konkluzji z wykorzystaniem klasycznych metod obliczeniowych – w niniejszej pracy wykorzystano metodę najmniejszej sumy kwadratów (LS). Rozwiązanie zadania tego typu można przedstawić w postaci następujących równań:

$$\tilde{\Theta} = [V^T V]^{-1} V^T y \quad (7)$$

$$\tilde{\Theta} = [\Theta_1, \Theta_2, \dots, \Theta_M]^T \quad (8)$$

$$\Theta_i = [a_0^i, a_1^i, \dots, a_L^i] \quad (9)$$

$$V = \begin{bmatrix} \mu_A^1(x(1))u(1) & \dots & \mu_A^M(x(1))u(1) \\ \vdots & \ddots & \vdots \\ \mu_A^1(x(N))u(N) & \dots & \mu_A^M(x(N))u(N) \end{bmatrix} \in \mathbb{R}^{N \times (M \times L)} \quad (10)$$

$$y = \begin{bmatrix} \underbrace{y(1), \dots, y(N)}_I & \dots & \underbrace{y(1), \dots, y(N)}_M \end{bmatrix}^T \in \mathbb{R}^{N \times M} \quad (11)$$

w których:  $u(k)$  – wektor wartości wejść modelu w dyskretnej chwili czasowej  $k$ , zaś  $y(k)$  – wartość wyjścia modelu w dyskretnej chwili czasowej  $k$ .

Przy wyznaczaniu modelu rozmytego typu TSK należy zwrócić szczególną uwagę na poziom weryfikacji poszczególnych modeli cząstkowych, tzn. liczebność zbioru danych [12], na podstawie których wyznaczona zostanie każda z konkluzji (8) modelu:

$$\kappa_i = \sum_{k=1}^N \mu_A^i(k) \quad (12)$$

Wartość ta nie może być zbyt mała, tzn.  $\kappa_i \gg L$  ( $L$  – liczba współczynników równania modelu (5)), ponieważ w przeciwnym wypadku może zostać wyznaczony model o błędnie określonej dynamice. Zwykle przyjmuje się relację  $\kappa_i \gg (10 \div 20)L$ .

Zarejestrowane dane podzielono na dwa zbiory – uczący (50%) i weryfikujący (walidacyjny) (50%). Współczynniki modelu wyznaczano na 60% losowo wybranych próbek ze zbioru uczącego, natomiast ocena modelu odbywała się z wykorzystaniem średniego błędu modelu (MAE – mean absolute error) na wszystkich próbkach zbioru uczącego.

### Struktury wyznaczanych modeli

Poprzez strukturę modelu TSK rozumie się sygnały rozmywające oraz ich podział na partycję oraz postać konkluzji modelu, czyli równania różnicowego. W przeprowadzonych rozważaniach dostępne były trzy sygnały:

- pobór wody ( $p_w$ ),
- dzień tygodnia ( $d_t$ ), przy czym kolejnym dniom przypisano kolejne (całkowite) wartości liczbowe, począwszy od poniedziałku (wartość 1) i skończywszy na niedzieli (wartość 7),
- pora roku ( $p_r$ ), przy czym kolejnym porom roku przypisano kolejne (również całkowite) wartości liczbowe, począwszy od wiosny (wartość 1), a skończywszy na zimie (wartość 4).

Wpływ poszczególnych sygnałów na wartość poboru wody oszacowano na podstawie wartości współczynnika korelacji r-Pearsona poszczególnych sygnałów ( $p_w$ ,  $d_t$ ,  $p_r$ ) z sygnałem poboru wody ( $p_w$ ). Analizując wyniki przedstawione w tabeli 1 można zauważyć, że wartości auto-korelacji sygnału produkcji wody bardzo szybko maleją. W dyskretnej chwili czasowej  $k-1$  wynosi on 0,416, co jest wartością porównywalną z wartością współczynnika r-Pearsona w przypadku korelacji dni tygodnia z sygnałem poboru wody w chwili  $k$ . Korelacja poboru wody z porą roku oscyluje w praktyce w okolicach zera. Biorąc pod uwagę

Tabela 1. Współczynnik korelacji r-Pearsona dostępnych sygnałów z sygnałem poboru wody  
Table 1. Pearson's r correlation coefficient of available signals with water demand signal

Sygnał	$p_w$	$d_t$	$p_r$
$x[k]$	1	-0,381	0,04
$x[k-1]$	0,416	-0,318	0,05
$x[k-2]$	0,171	0,017	0,06
$x[k-3]$	0,058	0,09	0,07
$x[k-4]$	0,013	0,261	0,08
$x[k-5]$	0,052	0,286	0,08
$x[k-6]$	0,013	0,259	0,09

wyznaczone wartości współczynników korelacji przyjęto następującą strukturę wyznaczanych modeli:

- sygnałami rozmywającymi były: wartość poboru wody w chwili  $k$  (ostatnia dostępna pomiarowo wartość), i/lub indeks dnia tygodnia w chwili  $k+1$  (indeks dnia tygodnia, dla którego wyznaczana jest prognoza),
- równanie różnicowe konkluzji zawierało archiwalne dane poboru wody i/lub indeks dnia tygodnia.

### Dyskusja wyników

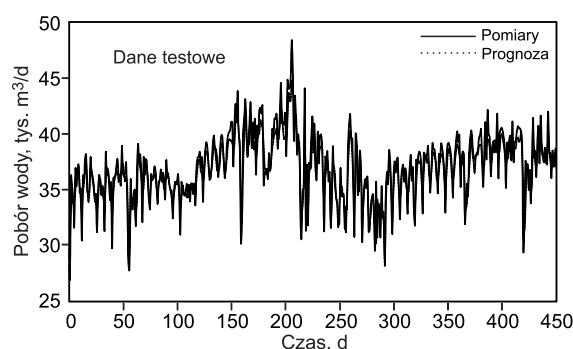
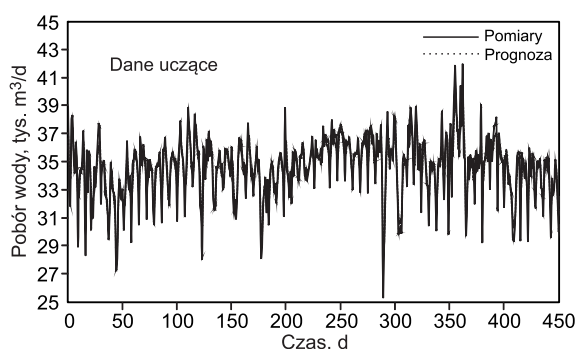
Zbiór danych uczących zawierał 487 próbek. Aby zależność (12) została spełniona, należało wyznaczyć odpowiednią liczbę zbiorów rozmytych tak, aby łączna liczba współczynników modeli cząstkowych nie przekraczała 48. Najlepsze wyniki modelowania uzyskano w przypadku konkluzji zawierających wyraz wolny, wartość indeksu dnia tygodnia, w chwili  $k+1$  oraz 6 archiwalnych wartości poboru wody, co daje (maksymalnie) liczbę 8 współczynników w równaniu. Na tej podstawie określono, że maksymalna liczba zbiorów rozmytych może wynosić 6. W badaniach wyznaczono modele, zakładając różne kombinacje rozmywania dostępnych sygnałów (tab. 2), przy czym należy zaznaczyć, że wyniki uzyskane według różnych modeli były porównywalne.

Analizując uzyskane wyniki modelowania można uznać, że wyznaczone modele obniżają poziom reakcji na pobudzenia, co jest częstym efektem przy wykorzystaniu estymatora LS. Powodowane jest to zakłóceniami pomiarowymi, bądź wpływem innych zakłóceń o charakterze losowym. Wysoki poziom korelacji sygnału modelowanego i zarejestrowanego (0,87 – dane uczące i 0,92 – dane testowe; współczynnik korelacji był wyższy w przypadku danych testowych ze względu na większą zmienność sygnału

Tabela 2. Parametry jakości i opis struktury wyznaczonych modeli  
Table 2. Quality parameters and structural description of selected models

Struktura modelu		postać konkluzji	Parametr jakości					
liczba zbiorów rozmytych (przesłanki)	pobór wody		MAE, m <sup>3</sup> /d		$e_w$ , %		r	
dni tygodnia			$d_u$	$d_t$	$d_u$	$d_t$	$d_u$	$d_t$
4	0	$y(k+1)=a_0^0+a_0^1d_t(k+1)+\sum_{i=1}^6 a_i^2 p_w(k-i)$	852	997	2,461574	2,69773	0,876	0,924
3	2	$y(k+1)=a_0^0+\sum_{i=1}^6 a_i^2 p_w(k-i)$	850	1026	2,455796	2,776199	0,871	0,916
6	0	$y(k+1)=a_0^0+\sum_{i=1}^6 a_i^2 p_w(k-i)$	855	1004	2,470242	2,716671	0,877	0,922
0	6	$y(k+1)=a_0^0+a_0^1d_t(k+1)+\sum_{i=1}^6 a_i^2 p_w(k-i)$	832	989	2,403791	2,676083	0,895	0,897
6	0	$y(k+1)=a_0^0+a_0^1d_t(k+1)+\sum_{i=1}^6 a_i^2 p_w(k-i)$	873	986	2,522247	2,667965	0,872	0,92

$d_u$  – dane uczące,  $d_t$  – dane testujące, MAE – średni błąd absolutny,  $e_w$  – błąd względny modelu,  $r$  – współczynnik korelacji Pearsona między prognozą a pomiarami



Rys. 4. Szeregi czasowe zarejestrowanych i prognozowanych wartości poboru wody (z lewej – dane uczące, z prawej – dane testowe)  
Fig. 4. Recorded and predicted time series of water demand (left – training data, right – testing data)

poboru wody w tym zbiorze) pozwala określić jakość modelowania jako zadowalającą. Średni błąd absolutny w przypadku danych uczących wynosił ok.  $850 \text{ m}^3/\text{d}$ , zaś testowych –  $990 \text{ m}^3/\text{d}$ , co pozwala określić wartość względnego błędu absolutnego na 2,4% (dane uczące) i 2,7% (dane testowe).

Wartości korelacji sygnału prognozowanego i modelowanego (przykładowe przebiegi w przypadku drugiego modelu w tab. 2 przedstawiono na rys. 4) dotyczące danych testowych były niewiele mniejsze od wartości uzyskanych w pracy [1], przy prognozowaniu z wykorzystaniem sieci neuronowych i użyciem radialnych funkcji bazowych ( $r=0,93\pm 0,94$ ). Należy pamiętać, że autorzy w badaniach z zastosowaniem sieci neuronowych wykorzystywali zbiór 7-letniego szeregu czasowego, a nie jak w niniejszym artykule – zbiór szeregu 2-letniego.

## Podsumowanie

Dokładność identyfikacji i prognozowania, a w konsekwencji właściwe wykorzystanie systemu monitoringu w dużym stopniu zależy od poprawnego wyboru rodzaju i struktury modelu. Przedstawiony sposób prognozowania dobowych wartości poboru wody wydaje się być dobrą alternatywą dla sztucznych sieci neuronowych. Uzyskane wyniki wykazały, że modelowanie może odbyć się w oparciu o znajomość tygodniowych zmian poboru wody, a wynikowy model pozwala prognozować wartości poboru wody w szerokim zakresie zmian punktu pracy procesu. Uzyskane wyniki, przy różnych kombinacjach rozmywania i doboru struktury konkluzji modelu, były porównywalne. Stąd można wnioskować, że metoda wyznaczania modeli rozmytych może być stosowana do określania właściwości procesów dynamicznych, w przypadku których nie jest znany dokładny opis modelowanych zjawisk. Otrzymane wyniki są zadowalające, gdyż opracowane modele pozwoliły na dobre określenie dynamiki procesu.

**Stachura, M., Studziński, J. Forecasting Hydraulic Load of Urban Water Supply System Using TSK Fuzzy Models. *Ochrona Środowiska* 2014, Vol. 36, No. 1, pp. 57–60.**

**Abstract:** The paper presents possibilities of daily water demand forecasting for municipal water supply system. For this purpose, Takagi-Sugeno-Kang's (TSK) models were applied. In this type of models the conclusion is in the form of a classical linear function, which allows describing their structure as 'fuzzified linear models'. For the purpose of this study data from the water supply network for the city of Rzeszów was used (974 samples). It covered the period from 1 July 2005 to 29 February 2008. Based on the collected data weakly water demand values were

## LITERATURA

1. P. LICZNAR, J. ŁOMOTOWSKI: Zastosowanie sztucznych sieci neuronowych Kohonena do prognozowania dobowego poboru wody. *Ochrona środowiska* 2006, vol. 28, nr 1, ss. 45–48.
2. J. ŁOMOTOWSKI, Z. SIWON: Metodyka analizy danych pochodzących z monitoringu systemów wodociągowo-kanalizacyjnych. *Gaz, Woda i Technika Sanitarna* 2010, nr 3, ss. 16–20.
3. M. STACHURA, B. FAJDEK, J. STUDZIŃSKI: Model based decision support system for communal water networks. Proc. of 'Industrial Simulation Conference ISC', Brno 2012.
4. U. SHAMIR, E. SALOMONS: Optimal real-time operation of urban water distribution systems using reduced models. *Journal of Water Resources Planning and Management* 2008, Vol. 134, No. 2, pp. 181–185.
5. S.T.M. BEZERRA, S.A. da SILVA, H.P. GOMES: Operational optimisation of water supply networks using fuzzy system. *Water SA* 2012, Vol. 38, No. 4, pp. 565–572.
6. M. GHIASSI, D. K. ZIMBRA, H. SAIDANE: Urban water demand forecasting with a dynamic artificial neural network model. *Journal of Water Resources Planning and Management* 2008, Vol. 134, No. 2, pp. 138–146.
7. C. ARSENE, B. GABRYŚ, D. AL-DABBASS: Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection. *Expert Systems with Applications* 2012, Vol. 39, pp. 13214–13224.
8. S. OSOWSKI: Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnictwa Politechniki Warszawskiej, Warszawa 2000.
9. T. TAKAGI, M. SUGENO: Fuzzy identifications of systems and its applications to modeling and control. *IEEE Transactions on System, Man, and Cybernetic* 1985, Vol. SMC-15, No. 1, pp. 116–132.
10. L. LJUNG: System Identification: Theory for the User. Prentice Hall, Englewood Cliffs 1987.
11. H. HELLENDORRN, D. DRIANKOV: Fuzzy Model Identification: Selected Approaches. Springer-Verlag, Berlin 1997.
12. P. WNUK: Evolutionary Optimization for Fuzzy TSK Model Identification. AI-METH 2005 Recent Developments in Artificial Intelligence Methods.

compared for different seasons. Subsequent TSK models were described together with the way they were developed. It was shown that modeling could be based on weakly water demand data and that resulting model allowed predicting water demand values over a wide range of variability. The results received for different combinations of fuzzification and model conclusion structure selection were comparable. Therefore, it could be concluded that the method used for fuzzy model development might be used to determine dynamic properties of the processes for which the exact description of modeled phenomena was unknown.

**Keywords:** Water supply system, water demand, modeling, forecasting, fuzzy TSK models.