

Andrzej Wałęga, Tomasz Bergel

**ZASTOSOWANIE METOD DATA MINING
W PROGNOZOWANIU ZUŻYCIA WODY
NA CELE BYTOWE W GOSPODARSTWACH WIEJSKICH**

***DATA MINING IMPLEMENTATION IN HOUSEHOLD
WATER USAGE FORECASTING IN THE FARMHOUSES***

Streszczenie

Krótkoterminowe prognozy wielkości zużycia wody, mają podstawowe znaczenie przy optymalizacji pracy sieci wodociągowych i kanalizacyjnych, a także oczyszczalni ścieków. W pracy dokonano oceny możliwości prognozowania szeregów czasowych dobowego zużycia wody na cele bytowe w gospodarstwie wiejskim z wykorzystaniem metod Data Mining. Do opracowania 10-dniowej prognozy zużycia wody wykorzystano wyrównanie wykładnicze i metodę ARIMA. Materiałem źródłowym do analizy były dobowe ilości wody zużywanej na cele bytowe w wytypowanym gospodarstwie wiejskim w okresie 22 miesięcy. Szczególnie przydatne do prognozowania zużycia wody okazało się wyrównanie wykładnicze, które uwzględnia nie tylko wartości, ale także zróżnicowanie ważności przeszłych obserwacji. Znaczna nierównomierność dobowego zużycia wody powoduje wzrost błędów prognoz. Metody prognozowania oparte na algorytmach wyrównania wykładniczego są łatwe do zastosowania i nie wymagają założenia o stacjonarności szeregu czasowego. W analizowanym przypadku stosunkowo dobrą prognozę dobowego zużycia wody na cele bytowe uzyskano stosując addytywny model Wintersa. Z kolei modele klasy ARIMA pozwalają na dokładną prognozę zużycia wody pod warunkiem właściwej identyfikacji parametrów modelu i spełnienia warunku stacjonarności szeregu. W przypadku braku stacjonarności, przed analizą szeregu, należy przekształcić go wykorzystując np. metodę różnicowania. W celu prognozowania wielkości dobowego zużycia w gospodarstwie wiejskim zastosowano model ARIMA (0,1,2) przy dwóch parametrach średniej ruchomej. Zarówno wyrównanie wykładnicze jak i model typu ARIMA pozwoliły na uzyskanie podobnych wyników prognozowania, przy czym średnia wartość 10-dniowej prognozy zużycia wody na cele bytowe w modelu wykładniczym była o 4,5% wyższa od uzyskanej z modelu ARIMA.

Słowa kluczowe: woda wodociągowa, prognozowanie zużycia wody, model ARIMA

Summary

Short-term water usage forecasts are fundamental for the waterworks and sewerage systems as well as for the sewage treatment plants' optimization. In this research the capability of forecasting the time series of daily household water usage in the farms with implementation of Data Mining methods was evaluated. To prepare the 10-days water usage forecast, exponential smoothing and ARIMA method was used. The source material were the daily amounts of water used for household purposes in the selected farmhouse during 22 months. Exponential smoothing turned out to be the most useful in water usage forecasting, because it includes not only the values, but also the diversification of the future forecasts' importance. Significant inequality of the daily water usage causes the increase of the forecasts' errors. The forecasting methods which base on the exponential smoothing algorithms are easy to apply and do not require the assumption of the stationarity of the time series. In the analyzed case relatively good forecast of the daily household water usage was obtained after applying the additive Winters model. On the other hand, ARIMA models allow for the precise forecast of the water usage, providing that the model parameters will be correctly identified and the condition of the time series stationarity will be met. In the case of non-stationary time series, before the analysis the series has to be transformed with e.g. differentiation method. In order to forecast the daily water usage in the farmhouse, ARIMA model was used (0,1,2) with two parameters of the moving average. The exponential smoothing as well as the ARIMA model allowed to obtain the similar forecast results, whereas the average value of the 10-days household water usage forecast in the exponential model was 4,5% higher than the forecast obtained with the ARIMA model.

Key words: tap water, water usage forecasting, ARIMA model

WPROWADZENIE

Bieżące pomiary oraz krótkoterminowe prognozy dotyczące ilości zużywanego wody, mają podstawowe znaczenie przy optymalizacji pracy sieci wodociągowych i kanalizacyjnych, a także oczyszczalni ścieków. W stanach zakłóceń znaczenie tych informacji jeszcze wzrasta. Zjawisko poboru wody i zapotrzebowania na nią przez konsumentów ma charakter złożony deterministyczno-losowy. Szczególnie jest to widoczne w gospodarstwach wiejskich, które w odróżnieniu od gospodarstw na terenach zurbanizowanych, charakteryzują się zdecydowanie odmienną strukturą zużycia wody [Bergel, Kaczor 2007]. Zużycie wody podlega pewnym prawidłowościom, na które nakładają się czynniki losowe. Z tego powodu do krótkoterminowego prognozowania zużycia wody są wykorzystywane najczęściej modele stochastyczne, wprowadzone w postaci zalgorytmizowanej do struktury zarządzania procesem sterowania. Przykładem takiego podejścia jest praca Siwonia i in. [2005], w której do krótkoterminowego prognozowania godzinowych poborów wody dla miast Brzeg i w wydzielonej części sieci wodociągowej we Wrocławiu, zastosowano wy-

równanie wykładnicze i modele klasy ARIMA. Modele te wykorzystywane są także do prognozowania dopływu ścieków do oczyszczalni, czego przykładem są badania Chotkowskiego i Lisa [2006]. Wymienieni autorzy prognozowali godzinowy dopływ ścieków do oczyszczalni w Kartuzach na podstawie prognozy poboru wody opartej na wyrównaniu wykładniczym modelem Wintersa.

Celem pracy jest ocena możliwości prognozowania szeregów czasowych dobowego zużycia wody na cele bytowe w gospodarstwie jednorodzinnym z wykorzystaniem metod Data Mining.

OPIS OBIEKTU I METODYKA BADAŃ

Materiał źródłowy do analizy stanowiły dobowe ilości wody zużywanej na cele bytowe w gospodarstwie wiejskim we Włostowicach (pow. proszowicki) zamieszkiwanym na stałe przez 5 osób, z czego 1 osoba przebywała w domu tylko od soboty wieczór do poniedziałku rano. Dane obejmowały okres od 7.07.2003 r. do 30.04.2005 r. i pochodziły z dokonywanych codziennie o tej samej porze odczytów wskazań wodomierza skrzydełkowego typu JS 1,5 17 firmy Metron wyposażonego w nadajnik impulsów, który współpracował z rejestratorem Mini-Log B produkcji Endress+Hauser. Podstawą do przeprowadzenia analizy szeregów czasowych zużycia wody było zastosowanie technik Data Mining. Jedną z definicji Data Mining zawarta jest w podręczniku Berry'ego i Linnoffa [1997], w którym techniki te zdefiniowane są jako proces badania i analizy dużych ilości danych metodami automatycznymi i półautomatycznymi w celu odkrycia znaczących wzorców i reguł. W podejściu Data Mining kluczowe jest uzyskanie odpowiedzi na pytanie nurtujące badacza, rozwiązanie konkretnego problemu, przewidzenie pewnej ważnej z praktycznego punktu widzenia wartości. Zazwyczaj mniej ważne jest sformułowanie ogólnego wniosku czy reguły [Demski 2007].

W pracy skupiono się przede wszystkim na analizie struktury szeregu czasowego i prognozowaniu dobowego zużycia wody na cele bytowe. W pierwszej kolejności dokonano weryfikacji szeregu czasowego zużycia wody pod kątem występowania wartości odstających. W dalszej kolejności zbadano czy w analizowanym ciągu danych występują określone prawidłowości. W tym celu posłużono się funkcją autokorelacji i autokorelacji cząstkowej. Z wykonanych korelogramów można odczytać siłę wzajemnych zależności między kolejnymi wartościami w ciągu. Opierając się na wynikach uzyskanych z analizy funkcji autokorelacji, dokonano dekompozycji szeregu czasowego zużycia wody, przyjmując na podstawie obserwacji trendu model addytywny. Celem metody dekompozycji sezonowej było wyodrębnienie składników trendu i sezonowości ciągu danych, to znaczy dekompozycja szeregu na efekt trendu, efekty sezonowe i pozostałą zmienność. Klasyczna technika dekompozycji jest znana jako metoda Census I opisana m.in. w pracy Makridakis i Wheelwright [1989] i tak została przyjęta w analizowanym przypadku.

Powszechnie w opisie zjawisk zachodzących w czasie stosuje się wyrównanie wykładnicze. W celu wyrównania szeregu i dokonania prognozy 10 kolejnych wartości zużycia wody wykorzystano wyrównanie wykładnicze z trendem gasnącym, gdzie parametry modelu α (parametr wygładzania), δ (parametr wyrównania sezonowego) i ϕ (parametr wyrównania trendu) zostały określone poprzez optymalizację ich wartości w przestrzeni parametrów tak, aby otrzymać najmniejszą sumę kwadratów (lub najmniejszy średni kwadrat) różnicy pomiędzy wartościami empirycznymi a wartościami prognozowanymi na jeden okres naprzód. Jako metodę wygładzania zastosowano model Wintersa [Siwoń i in. 2005]. Wersja modelu przyjęta w niniejszej pracy ma postać:

$$q_{t+m} = F_t + m \cdot S_t + C_{t+m-s} \quad (1)$$

$$F_{t-1} = \alpha \cdot (q_{t-1} - C_{t-1-s}) + (1 - \alpha) \cdot (F_{t-2} + S_{t-2}) \quad (2)$$

$$S_{t-1} = \delta \cdot (F_{t-1} - F_{t-2}) + (1 - \delta) \cdot S_{t-2} \quad (3)$$

$$C_{t-1} = \phi(q_{t-1} - F_{t-1}) + (1 - \phi) \cdot C_{t-1-s} \quad (4)$$

gdzie:

- F_{t-1} – ocena wartości średniej w okresie t-1,
- S_{t-1} – ocena przyrostu trendu w czasie t-1,
- C_{t-1} – ocena wskaźnika sezonowości w okresie t-1,
- α, δ, ϕ – parametry gładzenia,
- m – ilość kroków prognozy.

Ostatnim elementem analizy było zastosowanie metody ARIMA (*Auto Regressive Integrated Moving Average*) do prognozowania zużycia wody na cele bytowe. Prawidłowe przeprowadzenie analizy jest możliwe w przypadku spełnienia warunku o stacjonarności szeregu. W analizowanym przypadku na podstawie oceny przebiegu szeregu stwierdzono, że jest on niejednorodny, a przeprowadzona analiza funkcji autokorelacji szeregu danych wyjściowych wykazała, że szereg jest niestacjonarny. Aby usunąć niejednorodność oraz uzyskać stacjonarność szereg został przekształcony przez zastosowanie różnicowania (odejmowanie od kolejnej wartości szeregu czynnika związanego z opóźnieniem sezonowym). Stacjonarność przekształconego szeregu sprowadzono metodą autokorelacji i autokorelacji cząstkowej. W przypadku stwierdzenia stacjonarności przeprowadzono procedurę doboru modelu ARIMA poprzez ustalenie wartości liczbowych parametrów związanych z funkcją autokorelacji p i średnią ruchomą q . Model ARIMA ma ogólną postać:

$$\phi(B)X_t = \theta(B)A_t \quad (5)$$

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \sum \theta_j A_{t-j} \quad (6)$$

$$\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p) \quad (7)$$

$$\theta(B) = 1 - \theta_1(B) - \theta_2(B^2) - \dots - \theta_q(B^q) \quad (8)$$

gdzie:

$\phi(B)$ – operator autoregresji rzędu p ,

$\theta(B)$ – operator średniej ruchomej rzędu q ,

X_t – wartość procesy w czasie t ,

A_t – szereg czasowy reszt modelu, biały szum,

B – operator przesunięcia wstecz.

Parametry te zostały określone na podstawie analizy autokorelogramów szeregu przekształconego. Po ustaleniu parametrów modelu dokonano obliczeń prognozy za pomocą modelu ARIMA. Weryfikacja poprawności obliczeń została przeprowadzona na podstawie analizy istotności parametrów modelu przy pomocy statystyki t oraz sprawdzenia normalności i braku autokorelacji reszt. Wszystkie obliczenia zostały przeprowadzone w programie Statistica 8.

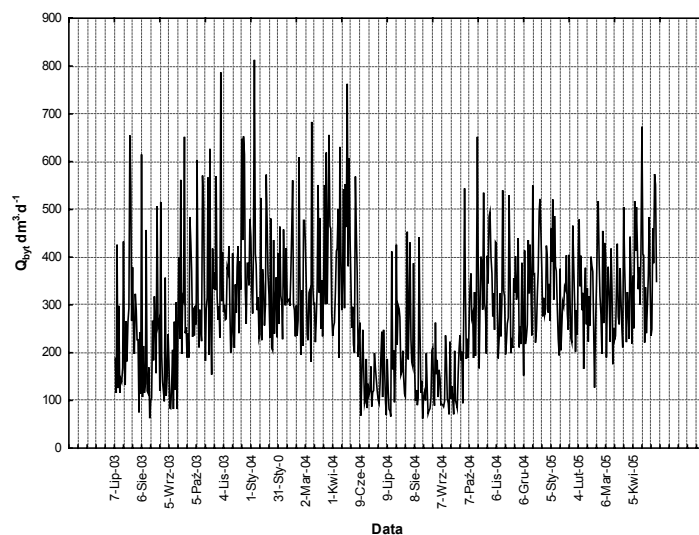
WYNIKI I DYSKUSJA

Na rysunku 1 przedstawiono szereg czasowy dobowego zużycia wody na cele bytowe w gospodarstwie wiejskim w analizowanym okresie. Przedstawiony szereg charakteryzuje się znacznymi wahaniami zużycia wody z brakiem wyraźnej zarysowującej się sezonowości.

Średnie dobowe zużycie wody wyniosło $295,4 \text{ dm}^3 \cdot \text{d}^{-1}$, przy odchyleniu standardowym $\sigma = 136,5 \text{ dm}^3 \cdot \text{d}^{-1}$. Wyliczone współczynniki nierównomierności wynoszą: $N_{\text{dmax}} = 2,75$ i $N_{\text{dmin}} = 0,21$. Maksymalny pobór wody wynoszący $813,0 \text{ dm}^3 \cdot \text{d}^{-1}$ miał miejsce 4.01.2004 r. (niedziela), a minimalny równy $62,0 \text{ dm}^3 \cdot \text{d}^{-1}$ 16.08.2004 r. (środa). Duże wartości odchylenia standardowego i współczynnika N_{dmax} potwierdzają dużą zmienność szeregu. W analizowanym szeregu można wyróżnić trzy okresy: pierwszy – trwający od lipca 2003 do końca maja 2004 r., gdzie występuje wyraźny trend rosnący dobowego zużycia wody; drugi – od czerwca do końca września 2004, gdzie występuje względnie stabilne zużycie wody na najniższym poziomie w całym badanym okresie i trzeci okres – od października 2004 do końca kwietnia 2005 r. z wyższymi niż poprzednio wartościami poboru wody i o nieznacznie rosnącej tendencji wzrostowej. Taki charakter zużycia wody wynikał zarówno z charakteru gospodarstwa jak i z pory roku. Wątpliwość mogłoby budzić jedynie zmniejszenie wielkości zużycia wody w okresie od czerwca do września 2004 r. Okazało się jednak, że wynikało ono z faktu przebywania w tym czasie w gospodarstwie jedynie 3 spośród 5 domowników (2 osoby wyjechały za granicę).

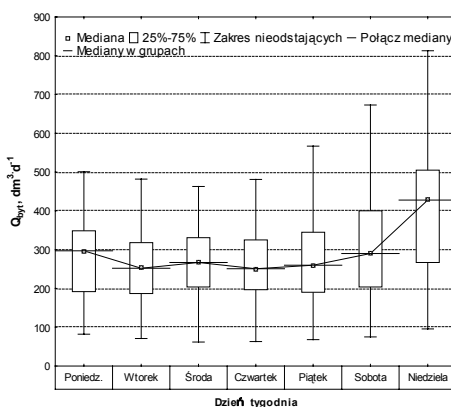
Na rysunku 2 przedstawiono mediany zużycia wody na cele bytowe wraz z zakresem zmienności dla poszczególnych dni tygodnia. Zauważyć można, że

wraz ze wzrostem wartości mediany rośnie zmienność poboru wody. Widać również, że od poniedziałku do piątku obserwuje się małe wahania wielkości zużycia wody, z ogólną tendencją do jego zmniejszania. Wzrost zużycia wody obserwuje się z kolei w weekend.



Rysunek 1. Szereg czasowy dobowego zużycia wody na cele bytowe w analizowanym gospodarstwie

Figure 1. Time series of the daily household water usage in the analyzed farmhouse



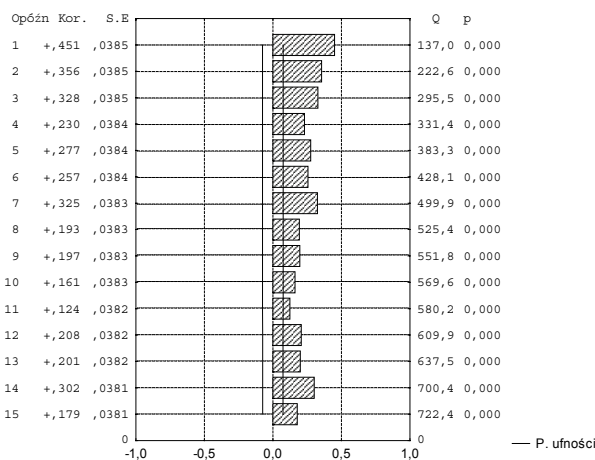
Rysunek 2. Charakterystyki statystyczne zużycia wody na cele bytowe w okresie uśrednionego tygodnia

Figure 2. Statistical characteristics of the household water usage during the averaged week

Wyniki przedstawione na rysunku 2 potwierdzają występowanie tzw. tygodniowej cykliczności zużycia wody, która wynika ze stylu życia mieszkańców i związanych z nim prac domowych. Zdziwienie może budzić jedynie fakt, że największe zużycie wody obserwuje się w niedzielę, podczas gdy w przeważającej większości gospodarstw największe zużycie wody na cele bytowe występuje w dzień powszedni – zazwyczaj w sobotę. Jednak jak wynika z wywiadu przeprowadzonego z domownikami analizowanego gospodarstwa, występowanie największego zużycia w niedzielę jest bardzo prawdopodobne. Wynika to z faktu, że gospodyni przeprowadza wiele prac domowych (kilkakrotne pranie bielizny, gotowanie potraw, pieczenie ciast) w związku z niedzielnym pobytem w domu jednego z domowników.

Celem sprawdzenia czy w analizowanym szeregu występuje czynnik sezonowy dokonano analizy autokorelacji oraz dekompozycji metodą Census I. Wyznaczony korelogram przedstawiono na rysunku 3. Widać na nim, iż poszczególne elementy szeregu są ze sobą istotnie skorelowane (na poziomie $\alpha = 0,05$). Jedną z przyczyn tego stanu może być niestacjonarność analizowanego szeregu.

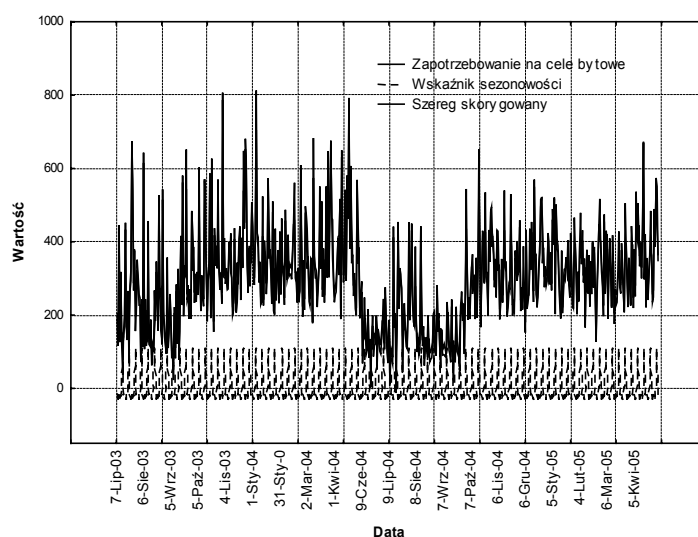
W dalszej analizie, przy stosowaniu metody ARIMA szereg będzie musiał zostać przekształcony tak, aby był stacjonarny (posiada stałą wartość przeciętną i wariancję).



Rysunek 3. Wykres funkcji autokorelacji badanego szeregu; oznaczenia na korelogramie: Opóźn. – wartości opóźnień, Kor. – wartość oceny funkcji autokorelacji, S.E. – błąd standardowy, Q – wartość statystyki do weryfikacji hipotezy, iż proces jest białym szumem, p- prawdopodobieństwo testowe

Figure 3. Graph of the auto-correlation function of the tested series; the correlogram symbols: Opóźn. – the delay values, Kor. – value of the auto-correlation function evaluation, S.E. – standard error, Q – statistic value for the hypothesis verification that the process is white noise, p- testing probability

Nieusunięta niestacjonarność jest w stanie „zagłuszyć” ważne prawidłowości procesu, np. sezonowość, możliwe do odgadnięcia na podstawie funkcji autokorelacji [Kot i in. 2007]. Z wykresu funkcji autokorelacji widać także, że dana wartość szeregu jest silnie skorelowana z obserwacją opróżnioną o 7 dni. To sugeruje, iż w badanym szeregu występują pewne prawidłowości sezonowe o długości 1 tygodnia. Wykorzystując to spostrzeżenie dokonano dekompozycji sezonowej badanego szeregu przyjmując opóźnienie sezonowe równe 7 dni. Z analizy przebiegu „surowego” szeregu przyjęto jego addytywność, czyli wahania procesu „dodają” się do tendencji rozwojowej i ich wielkość nie zależy od aktualnego poziomu trendu. Wyniki dekompozycji sezonowej przedstawiono na rysunku 4.

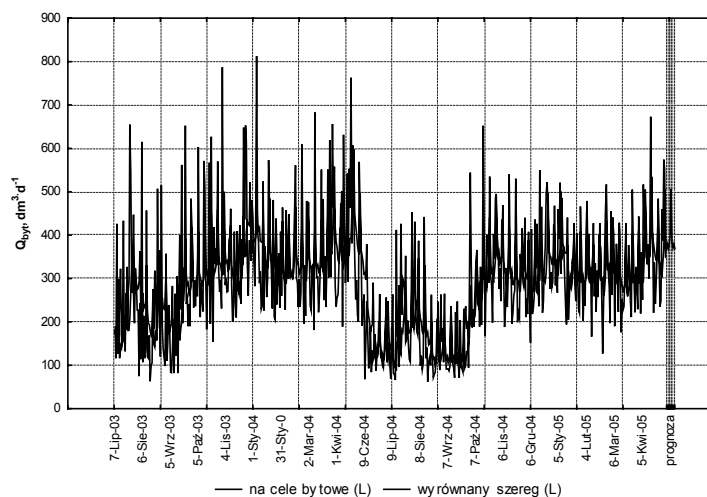


Rysunek 4. Wyniki analizy dekompozycji sezonowej dla analizowanego szeregu czasowego

Figure 4. Results of seasonal decomposition analysis for the analyzed time series

Składnik sezonowy został wyodrębniony z pierwotnego szeregu poprzez określenie różnic między szeregiem empirycznym (pierwotnym) i wygładzonym średnią ruchomą o okresie uśrednienia 7 dni. Z analizy wyodrębnionego składnika sezonowości wynika, iż jest on stały w czasie oraz w odstępach tygodniowych zużycie wody jest o $108,0 \text{ dm}^3 \cdot \text{d}^{-1}$ większe niż wynikałoby to z trendu oraz w takich samych odstępach mniejsze o $29,0 \text{ dm}^3 \cdot \text{d}^{-1}$ od określonych przez trend. Ze względu na fakt, iż wahania sezonowe są stałe w czasie, szereg skorygowany (nie zawierający wahań sezonowych) i pierwotny mają podobny przebieg.

Wyrównanie wykładnicze stało się bardzo popularne jako metoda prognozowania dla wielu typów szeregów czasowych. Umożliwiają one wyznaczenie przyszłych wartości szeregu (prognoz) poprzez obliczenie pewnego rodzaju średniej ruchomej, gdzie bieżącym i bezpośrednio poprzedzającym obserwacjom przypisuje się większą wagę niż obserwacjom starszym.



Rysunek 5. Wyrównanie wykładnicze dobowego zużycia wody wraz z 10-dniową prognozą (okres prognozy został wyróżniony siatką)
Figure 5. Exponential smoothing of the daily water usage with the 10-days forecast (the forecast period has been marked with the net)

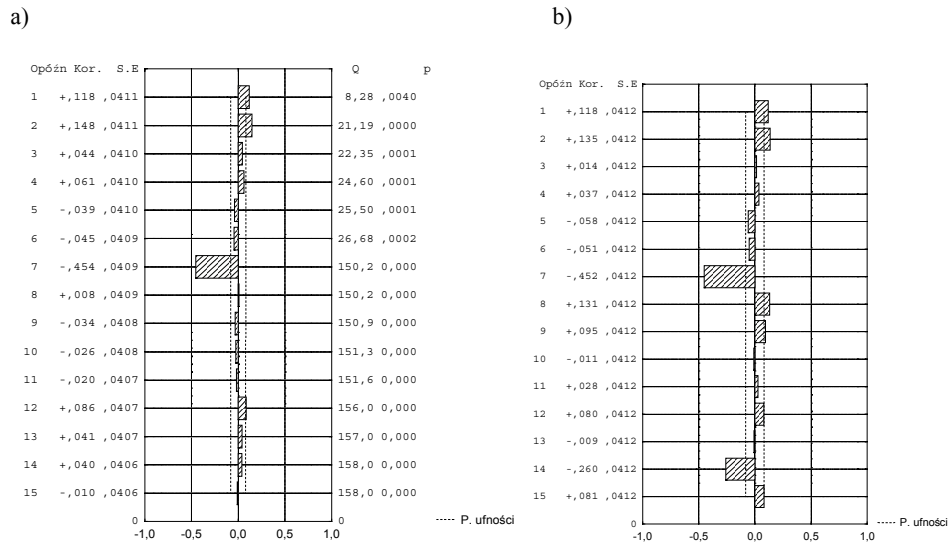
Na rysunku 5 przedstawiono rezultaty wyrównania wykładniczego wraz z 10-dniową prognozą zużycia wody w analizowanym gospodarstwie. W wyniku optymalizacji ustalone zostały następujące wartości parametrów modelu: $\alpha = 0,076$, $\delta = 0,000$ i $\varnothing = 0,137$. Wartość α bliska 0 oznacza, iż bieżąca przy obliczaniu kolejnej wartości szeregu bieżąca obserwacja zostaje praktycznie całkowicie zignorowana, z kolei $\delta = 0,0$ świadczy o dużej stabilności wskaźnika sezonowego. Według prognozy zużycie wody będzie oscylować od wartości 368,9 do 506,7 $\text{dm}^3 \cdot \text{d}^{-1}$ i przyjmie nieznaczną tendencję spadkową. W 10. dniu prognozy zużycie wody osiągnie wartość 368,9 $\text{dm}^3 \cdot \text{d}^{-1}$. Średnie zużycie wody w czasie 10-dniowej prognozy wyniesie 392,3 $\text{dm}^3 \cdot \text{d}^{-1}$. Weryfikację poprawności dopasowania przyjętego modelu do rzeczywistego szeregu określono przy pomocy błędu średniego, średniego błędu bezwzględnego, średniego kwadratu reszt, średniego błędu procentowego i średniego bezwzględnego błędu procentowego [Makridakis i in. 1983]. Wyniki przedstawiono w tabeli 1.

Tabela 1. Wyniki poprawności dopasowania modelu do rzeczywistego szeregu
Table 1. The correctness' results of the model adjustment to the actual series

Wskaźnik dopasowania	Wartość wskaźnika
Błąd średni, [$\text{dm}^3 \cdot \text{d}^{-1}$]	2,17
Średni błąd bezwzględny, [$\text{dm}^3 \cdot \text{d}^{-1}$]	80,41
Średni kwadrat reszt, [-]	11554,43
Średni błąd procentowy, [%]	-12,87
Średni bezwzględny błąd procentowy, [%]	32,43

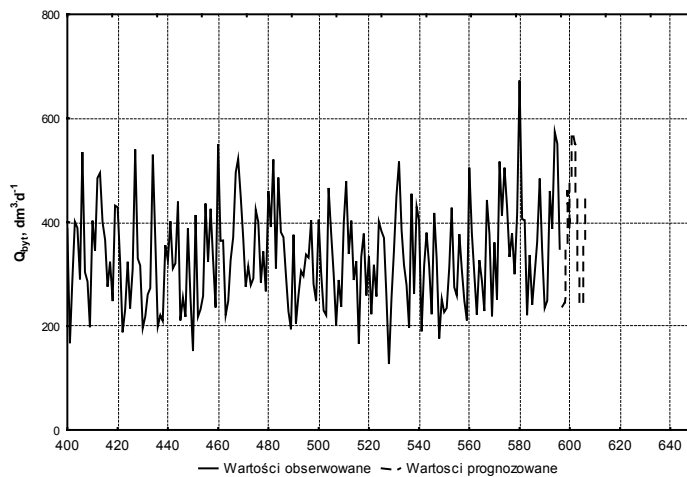
Częste wahania zużycia wody wpływają na dokładność wyników uzyskanych z modelu. O ile wartość błędu średniego jest niewielka, o tyle średni błąd procentowy wynoszący prawie – 13 % świadczy o tym, iż z takim błędem możemy przewidywać rzeczywiste wartości szeregu. Jakość modelu znacznie obniża wysoki średni bezwzględny błąd procentowy, który oznacza, że przeciętna prognoza odchyła się o $\pm 32,4$ % od wartości rzeczywistych.

Scałkowane modele autoregresji i średniej ruchomej (ARIMA) odwzorowują właściwości statyczne i dynamiczne szeregów stacjonarnych. Nadają się szczególnie do bieżącego i krótkoterminowego prognozowania szeregów czasowych poboru wody [Siwoń i in. 2005]. Warunkiem prawidłowego prognozowania za pomocą modelu ARIMA jest, by szereg był stacjonarny, a więc, że jego rozkład nie zależy od wyboru początku układu, tzn. nie zmienia się wraz z „przesunięciem” całego układu o τ jednostek w prawo lub lewo. Oznacza to również, iż wszystkie zmienne losowe X_t mają jednakowy rozkład [Box, Jenkins 1983]. Z uwagi na to, iż poszczególne wartości poborów wody są ze sobą istotnie skorelowane (rys. 3), przypuszczalnie analizowany szereg jest niestacjonarny. W związku z tym dokonano jego różnicowania, przyjmując czas opóźnienia 7 dni. Po dokonanych przekształceniach uzyskano nowy szereg stacjonarny. Wykres funkcji autokorelacji i autokorelacji cząstkowej szeregu przekształconego przedstawiono na rysunku 6. Z uwagi na fakt, iż funkcja autokorelacji wykazuje duże wartości przy opóźnieniach 1 i 2 oraz brak korelacji przy pozostałych opóźnieniach oraz, że funkcja autokorelacji cząstkowej ma kształt sinusoidy w modelu ARIMA przyjęto dwa parametry średniej ruchomej (q), a cały model charakteryzuje się strukturą (0,1,2). Obliczenia przeprowadzono dla opóźnienia 7 dni. Na rysunku 7 przedstawiono fragment surowego szeregu wraz z prognozą 10 obserwacji. Wydaje się, że model poprawnie prognozuje szereg czasowy dobowego zużycia wody. Według prognozy zużycie wody będzie rosło od wartości 237,7 do 460,0 $\text{dm}^3 \cdot \text{d}^{-1}$, osiągając maksimum 574,0 $\text{dm}^3 \cdot \text{d}^{-1}$ w piątym dniu prognozy. Średnie zużycie wody w prognozowanym okresie wyniesie 374,8 $\text{dm}^3 \cdot \text{d}^{-1}$. Średnie zużycie wody w okresie 10-dniowej prognozy jest wyższe o 4,5% w modelu wykładniczym w stosunku do modelu ARIMA. Model ARIMA można uznać za poprawny, ponieważ reszty można opisać rozkładem normalnym (rys. 8).

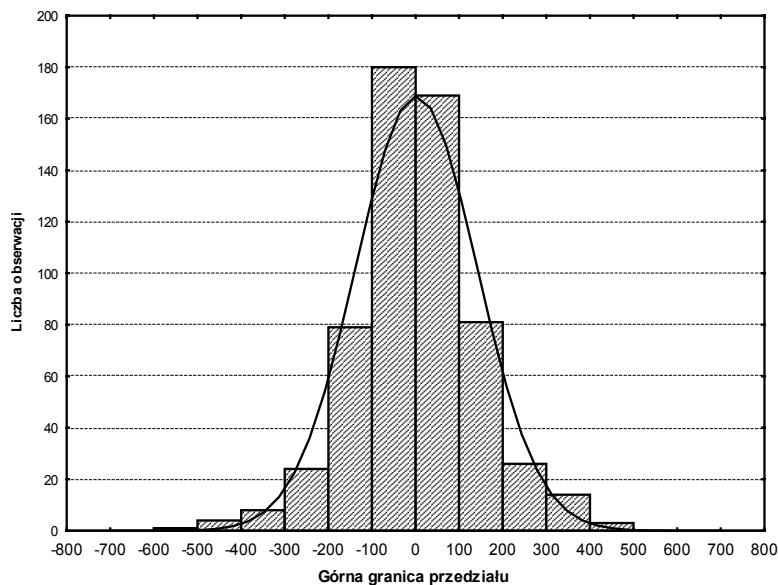


Rysunek 6. Wykresy funkcji: a) autokorelacji, b) autokorelacji cząstkowej szeregu przekształconego

Figure 6. Function graphs: a) auto-correlation, b) partial auto-correlation function of the transformed series



Rysunek 7. Wyniki prognozowania dobowego zużycia wody modelem ARIMA (0,1,2)
 Figure 7. The results of the daily water usage forecasting with the ARIMA model (0,1,2)



Rysunek 8. Histogram rozkładu reszt z modelu ARIMA (0,1,2) wraz z funkcją gęstości
Figure 8. Histogram of the rest distribution from the ARIMA model (0,1,2) with the density function

WNIOSKI

1. Dobowe zużycie wody na cele bytowe w gospodarstwach jednorodzinnych charakteryzuje się dużą nierównomiernością i cyklicznością tygodniową.
2. Techniki Data Mining pozwalają na wszechstronne modelowanie i prognozowanie szeregów czasowych dobowego zużycia wody z wykorzystaniem zarówno metod dekompozycji sezonowej, jak i wyrównania wykładniczego oraz modeli ARIMA
3. Szczególnie przydatne do prognozowania zużycia wody jest wyrównanie wykładnicze, bowiem uwzględnia ono nie tylko wartości, ale także zróżnicowanie ważności przeszłych obserwacji. Znaczna nierównomierność dobowego zużycia wody powoduje wzrost błędów prognoz. Metody prognozowania oparte na algorytmach wyrównania wykładniczego są łatwe do zastosowania i nie wymagają założenia o stacjonarności szeregu czasowego. W analizowanym przypadku stosunkowo dobrą prognozę dobowego zużycia wody na cele bytowe uzyskano stosując addytywny model Wintera.
4. Modele klasy ARIMA pozwalają na dokładną prognozę (obarczoną małym błędem) zużycia wody pod warunkiem poprawnej identyfikacji paramet-

trów modelu i spełnienia warunku stacjonarności szeregu. W przypadku braku stacjonarności, przed analizą szereg należy przekształcić go wykorzystując np. metodę różnicowania. W celu prognozowania wielkości dobowego zużycia w gospodarstwie jednorodzinnym zastosowano model ARIMA (0,1,2) przy dwóch parametrach średniej ruchomej.

5. Zarówno wyrównanie wykładnicze jak i model typu ARIMA pozwoliły na uzyskanie podobnych wyników prognozowania, przy czym średnia wartość 10-dniowej prognozy zużycia wody na cele bytowe jest o 4,5% wyższa od uzyskanej z modelu ARIMA (0,1,2).

BIBLIOGRAFIA

- Bergel T., Kaczor G. *The volume of wastewater discharged from rural households to the sewer system in the light of tap water consumption structure*. Polish Journal of Environmental Studies, Hard Olsztyn, Vol. 16, No. 2A, Part II, 109–112, 2007.
- Berry M. J. A., Linoff G. *Data mining techniques: for marketing, sales and customing support*. John Willey and Sons, 1997.
- Box G., Jenkins G. *Analiza szeregów czasowych*. PWN, Warszawa 1983.
- Chotkowski W., Lis G. *Krótkoterminowe prognozowanie dopływu ścieków do oczyszczalni na podstawie ilości zużywanej wody*. Gaz, Woda i Technika Sanitarna 5/2006, 13–18.
- Demski T. *Przykład prognozowania z zastosowaniem technik data mining*. StatSoft, 2007.
- Kot S. M., Jakubowski J., Sokołowski A. *Statystyka*. Difin, Warszawa 2007.
- Makridakis S., Wheelwright S. C., McGee V. E. *Forecasting. Methods and Applications*. John Wiley, New York 1983.
- Makridakis S., Wheelwright S. C. *Forecasting methods for management (Fifth Edition)*. Chichester: Wiley, New York 1989.
- Siwoń Z., Cieżak W., Cieżak J. *Bieżące prognozowanie godzinowego poboru wody z miejskich sieci wodociągowych*. Mat. XII Ogólnopolskiej Konferencji Naukowo-Technicznej z cyklu „Problemy gospodarki wodno-ściekowej w regionach rolniczo-przemysłowych”. Monografie Komitetu Inżynierii Środowiska PAN, nr 30, 15–33, 2005.

Dr inż. Andrzej Wałęga
Dr inż. Tomasz Bergel
Katedra Inżynierii Sanitarnej i Gospodarki Wodnej
Uniwersytet Rolniczy w Krakowie
al. Mickiewicza 24/28, 30-059 Kraków
tel. (012) 632 57 88, 662 41 08
e-mail: awalega@ar.krakow.pl tbergel@ar.krakow.pl

Recenzent: *Dr hab. Stanisław Węglarczyk, prof. PK*