

RYSZARDA IWANEJKO, JAROSŁAW BAJER\*

TEORETYCZNE PODSTAWY PROGNOZOWANIA  
USZKADZALNOŚCI SIECI WODOCIĄGOWEJTHEORETICAL BASIS OF PROGNOZING THE FAILURES  
IN WATER SUPPLY SYSTEMS

## Streszczenie

Sieci wodociągowe wody stanowią jeden z elementów strategicznej infrastruktury miejskiej. Są systemami złożonymi i dynamicznymi. Optymalna strategia eksploatacji sieci powinna uwzględniać stan aktualny i prognozy. Aktualny stan sieci może być charakteryzowany przez tzw. uszkodzalność. Parametr ten może być wyznaczony przez eksploatatora sieci jako średnia liczba uszkodzeń przypadająca na jednostkę długości i na jednostkę czasu. W artykule przedstawiono podstawy teoretyczne wybranych modeli matematycznych, które mogą być zastosowane do prognoz krótko- i średnioterminowych.

*Słowa kluczowe: sieć wodociągowa, sieć dystrybucji wody, uszkodzalność, prognoza*

## Abstract

Water supply systems are an important element of a strategic urban infrastructure. Their configurations are usually complex and dynamic. The optimal operation strategy of the water supply system should consider both current conditions as well as the future prognosis. Current conditions of the water supply system may be characterized by so called: failure rate. The parameter may be determined directly by the water system operator and defined as the average number of failures per the unit length of pipe and the time unit. The work presents a theoretical background for the selected mathematical models, that may be applied for the both short-term and mid-term predictions.

*Keywords: water supply system, water distribution system, failures, predictions*

\* Dr Ryszarda Iwanejko, dr inż. Jarosław Bajer, Instytut Zaopatrzenia w Wodę i Ochrony Środowiska, Wydział Inżynierii Środowiska, Politechnika Krakowska.

## 1. Wstęp

Zadaniem sieci wodociągowej (sieci dystrybucji wody) jest zapewnienie ciągłej dostawy do odbiorców wody w potrzebnej ilości, o odpowiedniej jakości i pod wymaganym ciśnieniem. Skutki nieprawidłowej pracy sieci są odczuwane przez odbiorców jako uciążliwości, a przez eksploatatora jako straty finansowe (m.in. usuwanie awarii, straty wody w sieci). Optymalna strategia eksploatacji sieci wodociągowej, minimalizująca straty i uciążliwości, powinna uwzględniać zarówno sytuację aktualną, jak i odpowiednie prognozy.

Sytuacja aktualna, charakteryzowana m.in. przez tzw. uszkodzalność sieci  $\lambda_0$  (roczna liczba uszkodzeń odniesiona do długości sieci) i wielkość strat wody, jest wynikiem stanu technicznego sieci, struktury materiałowej i wiekowej, bieżących prac prowadzonych na sieci i sposobu eksploatacji. Sieci wodociągowe to złożone systemy niejednorodnych przewodów, uzbrojenia i innych obiektów towarzyszących. Były rozbudowywane wraz z rozwojem miast, a w ciągu lat zmianie mogły ulegać struktura sieci, funkcje poszczególnych przewodów, ich obciążenie i warunki pracy.

Gromadzone przez każdego eksploatatora sieci wodociągowej informacje dotyczące uszkodzeń pozwalają wyznaczyć uszkodzalność sieci  $\lambda_0$ . Te wartości uzyskane dla poszczególnych lat tworzą szeregi dynamiczne  $\{\lambda_{0,i}\}_{i=1,\dots,n}$ , gdzie  $n$  – liczba danych (w konkretnych przypadkach jest to liczba lat, miesięcy, kwartałów, dób). Na zmienność w czasie każdego zjawiska, a więc i zjawiska uszkodzalności sieci, mają wpływ: tendencja rozwojowa, wahania okresowe i wahania przypadkowe [6]. Tendencja rozwojowa (trend) to powolne, regularne i systematyczne zmiany określonego zjawiska obserwowane w dostatecznie długim okresie (najlepiej co najmniej 10 lat) i będące rezultatem działania tzw. przyczyn głównych. W przypadku uszkodzalności przyczynami głównymi może być starzenie się sieci (trend rosnący) lub konsekwentna polityka remontowa (trend malejący). W przypadku niektórych sieci można zauważyć istotne, powtarzające się regularnie zróżnicowanie uszkodzalności w zależności od pory roku. Takie wahania sezonowe mogą wynikać np. z uszkodzeń spowodowanych przemarzaniem przewodów w miesiącach zimowych. Natomiast wahania przypadkowe są powodowane występowaniem czynników losowych.

Obecnie w wielu miastach polskich obserwuje się uszkodzalność sieci wodociągowych znacznie wyższą (3–10 razy) niż w wielu innych krajach świata, co wynika głównie z wieloletnich zaniedbań w zakresie konserwacji i modernizacji sieci oraz z niedoinwestowania w wymianę przewodów [4]. Dla wielu miast obserwuje się wzrost uszkodzalności sieci wodociągowych, co świadczy o starzeniu się sieci, przy czym część tych miast już wchodzi w okres konieczności renowacji przewodów sieci [5]. Niektórzy eksploatatorzy podjęli już decyzję o konieczności przystąpienia do remontów generalnych sieci wodociągowych. Rozpoczęcie tych prac zaczęło skutkować zmniejszeniem uszkodzalności sieci. Równocześnie ze względu na rozległość sieci wodociągowych, wysokie koszty odnowy przewodów, długi czas odnow czy prowadzenie innych prac przez przedsiębiorstwa wodociągowe, poprawa stanu sieci będzie procesem powolnym. Dlatego często powstaje pytanie: „jak może się zmienić uszkodzalność sieci, jeśli strategia eksploatacyjna nie ulegnie istotnej zmianie?”. Uzyskanie odpowiedzi pozwoli eksploatatorowi na ocenę sytuacji i podjęcie strategicznych decyzji co do konieczności zintensyfikowania remontów kapitalnych fragmentów sieci, możliwości odsunięcia w czasie tych remontów czy zmiany dotychczasowego sposobu eksploatacji.

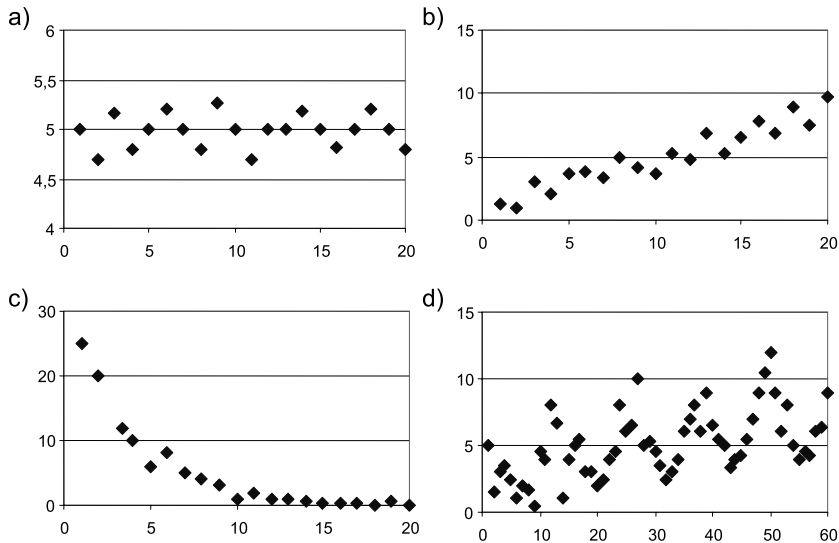
## 2. Ogólne uwagi o programowaniu

Każda prognoza jest wynikiem wnioskowania na przyszłość na podstawie znajomości danych opisujących zachodzenie badanego zjawiska w przeszłości (w tzw. okresie bazowym). Prognoza przyszłości jest tym bardziej wiarygodna, im krótszy jest horyzont prognozy, dłuższy jest okres bazowy i gdy mechanizm rozwoju badanego zjawiska nie ulega istotnym zmianom w czasie. Dla analizowanego problemu ostatnie założenie oznacza, że w najbliższej przyszłości nie będą podejmowane i wdrażane żadne nowe decyzje strategiczne i równocześnie nie ulegnie istotnej zmianie oddziaływanie środowiska. Drugi warunek odnosi się w szczególności do wystąpienia mroźnych, bezśnieżnych zim, gdy bardzo wzrasta głębokość przemarzania gruntu, co może być dodatkowym istotnym powodem wzrostu liczby awarii. Ze względu na charakter badanego zjawiska można stosować różne modele. W praktyce wyróżnia się 4 grupy modeli [8]:

- klasyczne modele trendu (modele tendencji rozwojowej), w których jedyną zmienną objaśniającą jest czas, a prognoza polega na ekstrapolacji wyznaczonej linii trendu; zastosowanie modeli z tej grupy wymaga spełnienia podstawowych założeń (np. że w okresie prognozowanym nie następuje zmiana mechanizmu rozwojowego zjawiska); niespełnienie założeń prowadzi do uzyskania niewiarygodnych prognoz,
- modele adaptacyjne, dla których klasyczne założenia nie muszą być spełnione; modele są elastyczne, mają zdolności dostosowawcze w przypadku nieregularnych zmian trendu i znacznych wahań przypadkowych; są szczególnie polecane do sporządzania prognoz krótkoterminowych i średnioterminowych,
- modele przyczynowo-skutkowe, które umożliwiają uwzględnienie wpływu różnych czynników (zmiennych objaśniających) na wyniki (zmienną objaśnianą); w zastosowaniach praktycznych trudnością może być dobór zmiennych objaśniających słabo skorelowanych ze sobą a silnie skorelowanych ze zmienną objaśnianą,
- modele autoregresyjne, w których zmiennymi objaśniającymi są opóźnione w czasie wartości zmiennej objaśnianej; do utworzenia dobrego modelu, tj. do określenia jak daleko należy sięgnąć wstecz, konieczne jest posiadanie dodatkowych, niestatystycznych informacji.

W praktyce najczęściej stosuje się klasyczne modele trendu i modele adaptacyjne, lecz wybór modelu jest uzależniony od posiadanych danych (rys. 1), od ich rozrzutu, trendu itp.

Ocena dokładności prognoz uzyskanych za pomocą dowolnego modelu może być typu *ex post* (na podstawie dopasowania wyników uzyskanych z zastosowania wybranego modelu do danych empirycznych z okresu bazowego) lub *ex ante* (dla przyszłych prognoz, przez podanie wartości spodziewanego odchylenia od prognozy). Przyjmuje się, że jeśli średni błąd względny prognozy *ex post* nie przekracza 5%, to prognozę uznaje się za dopuszczalną. Błędy typu *ex ante* dla pewnych modeli nie można wyznaczyć. Podstawowym sposobem zwiększenia pewności prognozy jest zastosowanie kilku metod (modeli) prognozowania i porównanie ich wyników ze sobą. Poniżej dokładnie omówiono wybrane modele prognozowania.



Rys. 1. Zbiory wyników obserwacji  $X$  i  $Y$  o tej samej średniej wykazujące: a) brak trendu, b) wyraźny rosnący trend liniowy, c) krzywoliniowy trend malejący, d) trend rosnący i cykliczność

Fig. 1. Pools of  $X$  and  $Y$  observations with same average, showing a) no trend, b) strong rising linear trend, c) declining non-linear trend, d) rising trend and periodicity

### 3. Modele trendu

Dotychczas do analiz uszkodzalności sieci wodociągowej, oprócz metod zwykłych opisowych, wykorzystywano modele statystyczne. Modele z tej grupy mogą zostać zastosowane, jeśli zostaną spełnione pewne podstawowe założenia (np. że w okresie prognozowanym nie następuje zmiana mechanizmu rozwojowego zjawiska, składniki losowe mają rozkład normalny o zerowej średniej, stałej wariancji i są od siebie niezależne).

W sytuacji, gdy nie zaobserwowano wyraźnego trendu i wartości  $\{\lambda_{0,i}\}$  dla kolejnych lat ( $i = 1, 2, \dots, n$ ) były zbliżone, uszkodzalność sieci w okresie bazowym charakteryzowano przez wartość średnią  $\bar{\lambda}_0$ , czasem wyznaczano przedział ufności dla tej średniej ( $\lambda_{0d}, \lambda_{0g}$ ). Można było sądzić, że w kolejnym okresie (np. roku) wartość uszkodzalności nie będzie się zbytnio różnić od średniej  $\bar{\lambda}_0$  i z żądanym prawdopodobieństwem znajdzie się w przedziale wyznaczonym na odpowiednim poziomie ufności. Estymatory (punktowy i przedziałowy) można było wyznaczać za pomocą metod statystycznych:

- klasycznych, gdy rozkład zmiennej losowej opisującej uszkodzalność przewodów był znany; w praktyce przyjmowano założenie o tzw. normalnym okresie eksploatacji, w którym występują jedynie uszkodzenia losowe, a zmienna losowa opisująca liczbę uszkodzeń przypadających na jednostkę czasu ma rozkład Poissona z parametrem  $\lambda_0$ ; wówczas

średnią uszkodzalność przewodów szacowano przez  $\bar{\lambda}_0 = \frac{\sum_{i=1}^n \lambda_{0,i}}{n}$ , gdzie  $n$  – liczba lat,

– nieklasycznych, np. metody bootstrapowej, gdy rozkład zmiennej losowej opisującej uszkodzalność przewodów nie był znany [3].

Natomiast w sytuacji, gdy można zaobserwować trend rozwojowy, to funkcję trendu można traktować jako szczególny przypadek funkcji regresji, w której zmienną niezależną jest czas. Wówczas prognozowanie wymaga kolejno:

- 1) przyjęcia modelu funkcji regresji,
- 2) oszacowania parametrów modelu,
- 3) statystycznej weryfikacji modelu,
- 4) wyznaczenia prognozy wartości zmiennych zależnych w oparciu o przyjęty model.

Postać funkcji regresji typuje się na podstawie wykresu rozproszenia sporządzonego dla odpowiednich par obserwacji  $(x_p, y_p)_{p=1 \dots n}$  – tutaj par  $(i, \lambda_{0,i})_{i=1 \dots n}$ . Najprostsza liniowa zależność między zmiennymi  $X$  (objaśniającą – dla szeregów czasowych: czas) i  $Y$  (objaśnianą, w tym przypadku wartości uszkodzalności  $\lambda_{0,i}$ ) ma postać:

$$\hat{y} = a \cdot x + b \quad (1)$$

gdzie parametr  $a$  interpretuje się jako przeciętną zmianę  $Y$  w przyjętej jednostce czasu (np. 1 rok), parametr  $b$  określa teoretyczny poziom badanej zmiennej na początku analizowanego okresu. Parametry modelu  $(a, b)$  szacuje się metodą najmniejszych kwadratów, minimalizując

błąd (*sum square error*)  $SSE = \sum_{i=1}^n e_i^2 = \sum (y_i - \hat{y}_i)^2$  [1]. Różnice  $e_i = y_i - \hat{y}_i$  są tzw.

błędami (resztami). Założeniem koniecznym do stosowania modelu jest, że błędy  $e_i$  mają zerową średnią, stałą wariancję i są od siebie niezależne. Ważne też jest, by dysponować wystarczającą liczbą danych ( $n$ ). W wyniku minimalizacji  $SSE$  uzyskuje się wartości

parametrów równe  $a = \frac{SS_{xy}}{SS_x}$  oraz  $b = \bar{y} - a \cdot \bar{x}$ , gdzie suma kwadratów oraz iloczyn

skalarny odchylen są wyznaczane jako  $SS_x = \sum_i (x_i - \bar{x})^2$ ,  $SS_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}_i)$ ,

a  $\bar{x}$ ,  $\bar{y}$  to odpowiednie średnie. W celu oceny przyjętego modelu liniowego stosuje się takie same miary i przeprowadza takie same testy jak dla regresji liniowej.

Miarę rozproszenia danych historycznych wokół linii trendu szacuje się przez wariancję resztową lub średni błąd kwadratowy (*mean square error*)  $MSE = \frac{SSE}{n-2}$ . Odchylenie stan-

dardowe składnika resztowego wynosi  $s = s_y = \sqrt{MSE}$ .

Współczynnik zmienności resztowej wyznaczany jako  $V = \frac{s}{\bar{y}}$  informuje, jaki procent zaobserwowanej zmienności  $Y$  stanowią odchylenia przypadkowe [6].

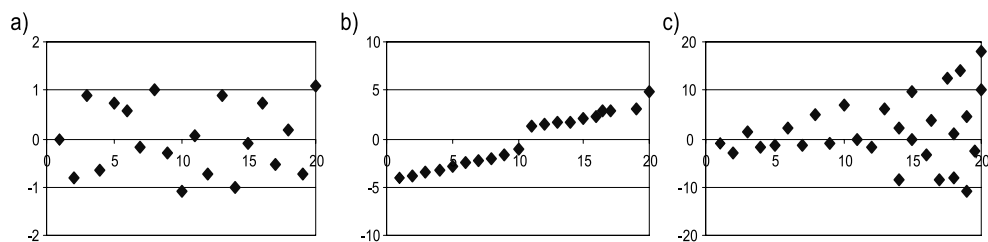
Współczynnik korelacji  $r$  będący miarą dopasowania linii trendu do danych wyznacza się jako  $r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}}$ , gdzie  $SS_y = \sum_i (y_i - \bar{y})^2$ . Wartość  $r^2$  jest interpretowana jako część zmienności  $Y$ , która jest wyjaśniona przez przyjęty liniowy związek. W prognozach zamiast współczynnika  $r^2$  stosuje się raczej współczynnik zbieżności  $\varphi^2 = 1 - r^2$ .

Najważniejszym testem statystycznym jest test weryfikujący istotność parametru  $a$  (równoważnie test badający czy między  $X$  i  $Y$  zachodzi związek liniowy). Sprawdzana jest hipoteza  $H_0: a = 0$  wobec  $H_1: a \neq 0$ . Podstawowym sprawdzianem w dwustronnym teście jest

$$t = \frac{a}{S(a)}, \quad S(a) - \text{standardowy błąd oceny parametru } a \text{ równy } s(a) = \frac{s}{\sqrt{SS_x}}.$$

rozkład  $t$  o  $(n - 2)$  stopniach swobody. W przypadku regresji liniowej testy  $t$  i  $F$  są równoważne [1], natomiast w przypadku regresji wielorakiej test  $F$  służy do ogólnego sprawdzenia hipotezy o zachodzeniu liniowego związku, zaś pojedyncze testy  $t$  są wykorzystywane do oceny istotności poszczególnych zmiennych. Możliwość zastosowania w praktyce testu  $F$  wynika z faktu, że wartość odpowiedniej statystyki może być wyznaczana za pomocą dostępnych programów komputerowych (np. przez formułę REGLINP w Excelu).

Reszty  $e_i$  nie powinny wykazywać trendu względem czasu, powinny być losowe. Oceny tej cechy można dokonać wizualnie na podstawie wykresu (rys. 2) lub analitycznie, np. za pomocą testu serii.



Rys. 2. Wizualna analiza rozrzutu reszt: a) reszty losowe, b) reszty wykazujące wzrastający trend czasowy, c) reszty o wzrastającej wariancji

Fig. 2. Visual analysis of the rest dispersion a) random rests, b) rests increasing in time c) rests with an increasing variance

Prognozowanie na przyszłość jest ekstrapolacją. Ten etap wymaga przyjęcia założenia, że czynniki kształtujące badane zjawisko nie ulegną zasadniczej zmianie w okresie objętym prognozą [8], co oznacza, że w najbliższej przyszłości nie będą podejmowane i wdrażane żadne nowe decyzje strategiczne. Przy tym założeniu prognozę dla chwili  $x_T$ , wykraczającej poza okres bazowy, wyznacza się jako  $\hat{y}_T = a \cdot x_T + b$ . Prognozę przedziałową typu *ex ante* dla wartości  $y_T$  wyznacza się na podstawie wzoru [2]:

$$P\{\hat{y}_T - t_\alpha \cdot D(\hat{y}_T) < y_T < \hat{y}_T + t_\alpha \cdot D(\hat{y}_T)\} = 1 - \alpha \quad (2)$$

gdzie:

$\hat{y}_T$  – prognoza punktowa,

$D(\hat{y}_T)$  – błąd standardowy prognozy,

$t_\alpha$  – wartość zmiennej  $t$ -Studenta dla  $n - 2$  stopni swobody i współczynnika ufności  $1 - \alpha$ .

$$\text{Błąd prognozy wyznacza się z wzoru } D(\hat{y}_T) = s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_T - \bar{x})^2}{\sum_{i=1..n} (x_i - \bar{x})^2}}.$$

Choć przedstawione powyżej wzory są skomplikowane, to w praktyce większość wielkości może być wyznaczona automatycznie przez zastosowanie odpowiedniego narzędzia (np. odpowiednie formuły w Excelu).

Prognozując zmienność  $Y$  należy uwzględnić, że np. mogą nie być spełnione podstawowe założenia modelu regresji, że kolejne obserwacje, tj. dane z szeregów czasowych, mają tendencje do wzajemnej korelacji, a błędy  $e_i$  mogą nie być wzajemnie niezależne lub że liczba obserwacji może nie być wystarczająco duża. To oznacza, że nie można mieć bezgranicznego zaufania do prognoz opartych na modelu regresji.

#### 4. Modele adaptacyjne

Metody z tej grupy są przydatne do prognozowania szeregów czasowych, dla których nie są spełnione klasyczne założenia. Praktycznie jedynym założeniem koniecznym do zastosowania modeli adaptacyjnych do prognozowania w przyszłości jest założenie stacjonarności (w czasie) błędów prognozy. W odróżnieniu od klasycznych metod analizy trendu i metod przyczynowo-skutkowych nie wyznacza się również parametrów modelu, których interpretacja wiązałaby się z charakterem badanego zjawiska. Ze względu na stosunkowo dużą dokładność i prostotę obliczeń, metody z tej grupy są często stosowane do sporządzania prognoz krótko- i średnioterminowych.

W części dotyczącej modeli adaptacyjnych w miejsce ogólnych oznaczeń danych  $(x_i, y_i)_{i=1..n}$  stosowane będą oznaczenia  $(t, y_t)_{t=1..n}$ , gdzie  $t$  – oznacza czas.

Błąd prognozy typu *ex post* (dla okresu bazowego) wyznacza się na podstawie błędów prognozy w kolejnych krokach  $e_t = y_t - \hat{y}_t$ ,  $t = 1..n$ , gdzie  $\hat{y}_t$  – tzw. prognoza wygasła, czyli prognoza dla czasu  $t$  z okresu bazowego wyznaczona za pomocą modelu. Najczęściej stosowaną miarą stopnia dokładności prognozy, podobnie jak w analizie trendu, jest błąd

$$\text{średniokwadratowy (mean square error) [1] } MSE = \frac{\sum_{i=1..n} (y_i - \hat{y}_i)^2}{n}. \text{ Drugą miarą jest średni}$$

błąd absolutny (*mean absolute error*)  $MAE = \frac{\sum_{i=1..n} |y_i - \hat{y}_i|}{n}$ . Pierwsza z tych miar jest bar-

dzo wrażliwa na wartości nietypowe, czyli duże rzadkie błędy prognozy. Niskie wartości błędów  $MSE$  i  $MAE$  pozwalają przypuszczać, że przy spełnionym założeniu o stałości charakteru badanego zjawiska, metoda z wysokim stopniem pewności może być zastosowana do prognozy wybiegającej poza okres bazowy. Spośród modeli adaptacyjnych w praktyce najczęściej wykorzystuje się trzy, które przedstawiono poniżej.

#### 4.1. Model wygładzania wykładniczego Browna

Metody wygładzania wykładniczego obejmują grupę metod o zróżnicowanym stopniu złożoności [1]. Technika wygładzania wykładniczego może być stosowana, gdy zmienna  $Y$  wykazuje trend i wahania przypadkowe, lecz nie wykazuje wahań okresowych. Prognozowane wartości zmiennej  $Y$  dla kolejnej chwili ( $t+1$ ) w tzw. prostym modelu Browna wyznacza się jako [1]:

$$\hat{y}_{t+1} = w \cdot y_t + (1-w) \cdot \hat{y}_t \quad (3)$$

gdzie:

$y_t$  – znana, zaobserwowana dla chwili  $t$  wartość zmiennej  $Y$ ,

$\hat{y}_t$  – wartość prognozy dla chwili  $t$ ,

$w$  – parametr wygładzania wykładniczego  $w \in (0, 1)$ .

Inaczej mówiąc, wartość prognozy jest średnią ważoną rzeczywistej, znanej wartości i prognozy dla chwili poprzedniej. Zwyczajowo przyjmuje się  $\hat{y}_1 = y_1$ . Parametr wygładzania  $w$  dobiera analityk. Jeśli zmiany  $y_t$  w czasie są częste i nieregularne i prognoza powinna reagować szybko na zmiany  $\{y_i\}_t$ , czyli gdy większy wpływ na prognozę powinna mieć najnowsza obserwacja, to należy przyjąć  $w$  bliskie 1. Jeśli natomiast prognoza powinna reagować z opóźnieniem to należy przyjmować niższe wartości  $w$ . Rozwinięcie rekurencyjnego wzoru (3) do postaci:

$$\hat{y}_{t+1} = wy_t + w(1-w)y_{t-1} + w(1-w)^2 y_{t-2} + w(1-w)^3 y_{t-3} + \dots + w(1-w)^{t-1} y_1$$

wskazuje, że prognoza opiera się na wszystkich poprzedzających rzeczywistych wartościach szeregu, lecz wpływ początkowych wyrazów szeregu maleje wykładniczo. Ta właściwość znalazła swoje odbicie w nazwie metody.

Czasem stosuje się zmodyfikowany model Browna [8]. Ocenę trendu  $m_t$  oraz prognozy  $\hat{y}_T$  wyznacza się na podstawie wzorów

$$m_t = \alpha \cdot y_t + (1-\alpha) \cdot m_{t-1} \quad \hat{y}_T = m_t + (m_t - m_{t-1}) \cdot h \quad (4)$$

gdzie  $T = t + h$ . Parametr wygładzania  $\alpha$  dobiera się na podstawie danych empirycznych tak, aby uzyskać jak najmniejsze rozbieżności między rzeczywistymi realizacjami  $\{y_i\}_{i=1..n}$  a wyznaczonymi prognozami  $\{\hat{y}_i\}_{i=1..n}$ . Modele wygładzania wykładniczego,



mimo łatwości stosowania, pozwalają zazwyczaj na dobre dopasowanie prognoz  $\{\hat{y}_t\}$  do danych empirycznych  $\{y_t\}$ .

#### 4.2. Model trendu pelzającego

Metoda powinna być stosowana, gdy szereg  $\{y_t\}_t$  charakteryzuje się dużą nieregularnością i załamaniem trendu. Metodę najlepiej przedstawić opisowo.

Prognoza dla chwili  $T$  wybiegającej do przodu poza okres bazowy składa się z kilku etapów. Najpierw na podstawie wizualnej oceny przebiegu badanego zjawiska, najlepiej na podstawie wykresu wyznaczonych indeksów łańcuchowych  $i_{t-1/t} = y_{t-1}/y_t$ , wyznacza się tzw. stałą wygładzania  $k < n$ . Przyjęta wartość odpowiada średniej długości fragmentów szeregu czasowego o podobnej zmienności trendu. Następnie dla  $(n - k + 1)$  fragmentów szeregu, których początkiem są kolejno  $y_1, y_2 \dots y_{n-k}$ , a każdy ma długość  $k$  („pelzające” segmenty) zakłada się liniową funkcję trendu i wyznacza teoretyczne prognozy. W ten sposób dla każdej chwili  $t$  z okresu bazowego ( $t = 1, \dots, n$ ) należącej równocześnie do kilku fragmentów szeregu uzyskuje się kilka prognoz, z których każda była generowana dla innego, kolejnego zakresu szeregu bazowego. Jako prognozę  $\hat{y}_t$  dla każdej chwili okresu bazowego przyjmuje się więc wartość średnią prognoz generowanych dla danej chwili  $t$  objętej kolejnymi zakresami prognoz. Na podstawie uśrednionych prognoz  $\{\hat{y}_t\}_{t=1..n}$  wyznacza się kolejne przyrosty uzyskanego trendu równe  $\Delta y_t = \hat{y}_{t+1} - \hat{y}_t$ . Średni przyrost trendu w całym okresie bazowym wyznacza

się jako  $\Delta y = \sum_{t=1}^{n-1} w_t \cdot \Delta y_t$ , gdzie  $w_t$  – wagi harmoniczne realizujące „postarzenie informacji”

(maleją monotonicznie z upływem czasu  $t$ ). Wagi są wyznaczone jako  $w_t = \frac{1}{n-1} \sum_{i=1}^t \frac{1}{n-i}$ .

Jako wartość prognozy punktowej dla chwili  $T$  przyjmuje się  $\hat{y}_T = y_n + (T - n) \cdot \Delta y$ . Metoda trendu pelzającego, choć jest złożona w zastosowaniu, pozwala na uzyskanie bardzo dobrej zgodności dopasowania prognoz  $\{\hat{y}_t\}$  do danych empirycznych  $\{y_t\}$ .

#### 4.3. Model Holta

Metoda może być szeroko stosowana. Model opiera się na dwóch równaniach, z których pierwsze wygładza wartości zmiennej prognozowanej  $\{y_t\}$ , a drugie wygładza przyrosty trendu [7]:

$$F_t = \alpha \cdot y_t + (1 - \alpha) \cdot (F_{t-1} + S_{t-1}) \quad (5)$$

$$S_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot S_{t-1} \quad (6)$$

gdzie:

$\alpha, \beta$  – parametry wygładzania z przedziału  $[0, 1]$ .

Wartości te dobiera się, minimalizując wartość średniego błędu kwadratowego prognoz wygasłych. Zazwyczaj przyjmuje się  $F_1 = y_1$  oraz  $S_1 = y_2 - y_1$ . Prognozy wygasłe dla  $t < n$  ( $n$  – liczba wyrazów szeregu czasowego) wyznacza się jako  $\hat{y}_{t+1} = F_t + S_1$ . Prognozy dla czasów  $T > n$  wyznacza się jako:

$$\hat{y}_T = F_t + (T - n)S_t \quad (7)$$

Prognozy  $\{\hat{y}_t\}$  uzyskane za pomocą modelu Holta zazwyczaj wykazują dobre dopasowanie do danych empirycznych  $\{y_t\}$ .

## 5. Wnioski

- Znajomość prognozy zmiany uszkodzalności sieci wodociągowej może być przydatna przy podejmowaniu decyzji co do zmiany lub kontynuowania sposobu eksploatacji sieci.
- Gdy prognozy wskazują na wzrost uszkodzalności  $\lambda_0$ , to wyniki te mogą i powinny pełnić rolę ostrzegawczą.
- Modele oparte na regresji liniowej, ze względu na konieczność spełnienia założeń, mogą mieć ograniczone zastosowanie. Modele trendu dla zjawisk i procesów dynamicznych o dużym stopniu losowości zazwyczaj generują duże błędy *ex ante*.
- Modele adaptacyjne nie wymagają spełnienia klasycznych założeń.
- Grupa modeli adaptacyjnych obejmuje modele o różnym stopniu złożoności, które pozwalają na uzyskanie dobrego lub bardzo dobrego dopasowania prognoz  $\{\hat{y}_t\}$  do danych empirycznych  $\{y_t\}$ . Prognozy, których błędy *ex post* nie przekraczają 5% uznaje się za przydatne [8].
- Modele adaptacyjne są szeroko stosowane do opisu zachodzenia zjawisk i prognozowania.
- Podstawowym sposobem zwiększenia pewności prognozy jest zastosowanie kilku metod (modeli) prognozowania i porównanie ich wyników ze sobą.
- W praktyce istnieje możliwość zwiększenia wiarygodności prognozy przez złożenie prognoz. W tym celu każdemu z wykorzystywanych modeli prognostycznych przypisuje się unormowane wagi (odwrotnie proporcjonalne do wariancji błędów prognoz modeli [1]).
- W kolejnym artykule, na przykładzie danych o uszkodzalności sieci wodociągowej miasta Krakowa, zostaną przedstawione wybrane metody służące do badania i sporządzania prognoz krótko- i średnioterminowych.

*Praca naukowa finansowana ze środków na naukę w ramach projektu rozwojowego Nr N R14 0006 10 nt. „Opracowanie kompleksowej metody oceny niezawodności i bezpieczeństwa dostawy wody do odbiorców”.*

## Literatura

- [1] A c z e l A., *Statystyka w zarządzaniu*, Wydawnictwo Naukowe PWN, Warszawa 2000.
- [2] G r e ń J., *Statystyka matematyczna. Modele i zadania*, Wydawnictwo Naukowe PWN, Warszawa 1982.

- [3] I w a n e j k o R., *Klasyczne i nieklasyczne metody szacowania uszkodzalności sieci dystrybucji wody*, INSTAL 6/2009, 58-60.
- [4] K w i e t n i e w s k i M., R a k J., *Niezawodność infrastruktury wodociągowej i kanalizacyjnej w Polsce*, Wydawnictwo Komitetu Inżynierii Lądowej i Wodnej PAN, Warszawa 2010, 41-59.
- [5] R a k J., *Podstawy bezpieczeństwa systemów zaopatrzenia w wodę*, Monografie Komitetu Inżynierii Środowiska PAN, vol. 28, Lublin 2005, 113-119.
- [6] S o b c z y k M., *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa 1996.
- [7] [http://www.megaedukacja.pl/model\\_holta.php](http://www.megaedukacja.pl/model_holta.php)
- [8] Z e l i a ś A., *Teoria prognozy*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1979.