

IWONA SULIKOWSKA

Gdańsk

## Estymacja parametrów modeli autoregresji z uwzględnieniem zakłóceń pomiarowych

### 1. Wprowadzenie

Nawiązując do wcześniej publikowanych prac [7, 3] dotyczących problematyki uwzględnienia błędu pomiaru w statystycznej analizie danych pomiarowych w artykule pokazano dalsze zastosowanie metody filtracji. Celem pracy było określenie zmian charakterystyk statystycznych ciągów przefiltrowanych i analiza wpływu zakłóceń pomiarowych na estymaty parametrów modelu autoregresji drugiego rzędu.

### 2. Podstawowy model rozkładu wilgotności

W świetle prowadzonej identyfikacji modelu [4, 5] można przyjąć, że dla ciągów reprezentujących rozkłady wilgotności naturalnej w ilach i pyłach piaszczystych struktura probabilistyczna ciągów obserwowanych  $z(k)$  odpowiada strukturze probabilistycznej liniowych modeli autoregresji ARIMA [2.0.0]. Zatem  $z(k)$  można wyrazić jako liniową kombinację dwóch wartości poprzednich obserwacji  $z(k-1)$  i  $z(k-2)$  oraz białego szumu  $a(k)$ . W postaci operatorowej, model rozkładu wilgotności ma postać:

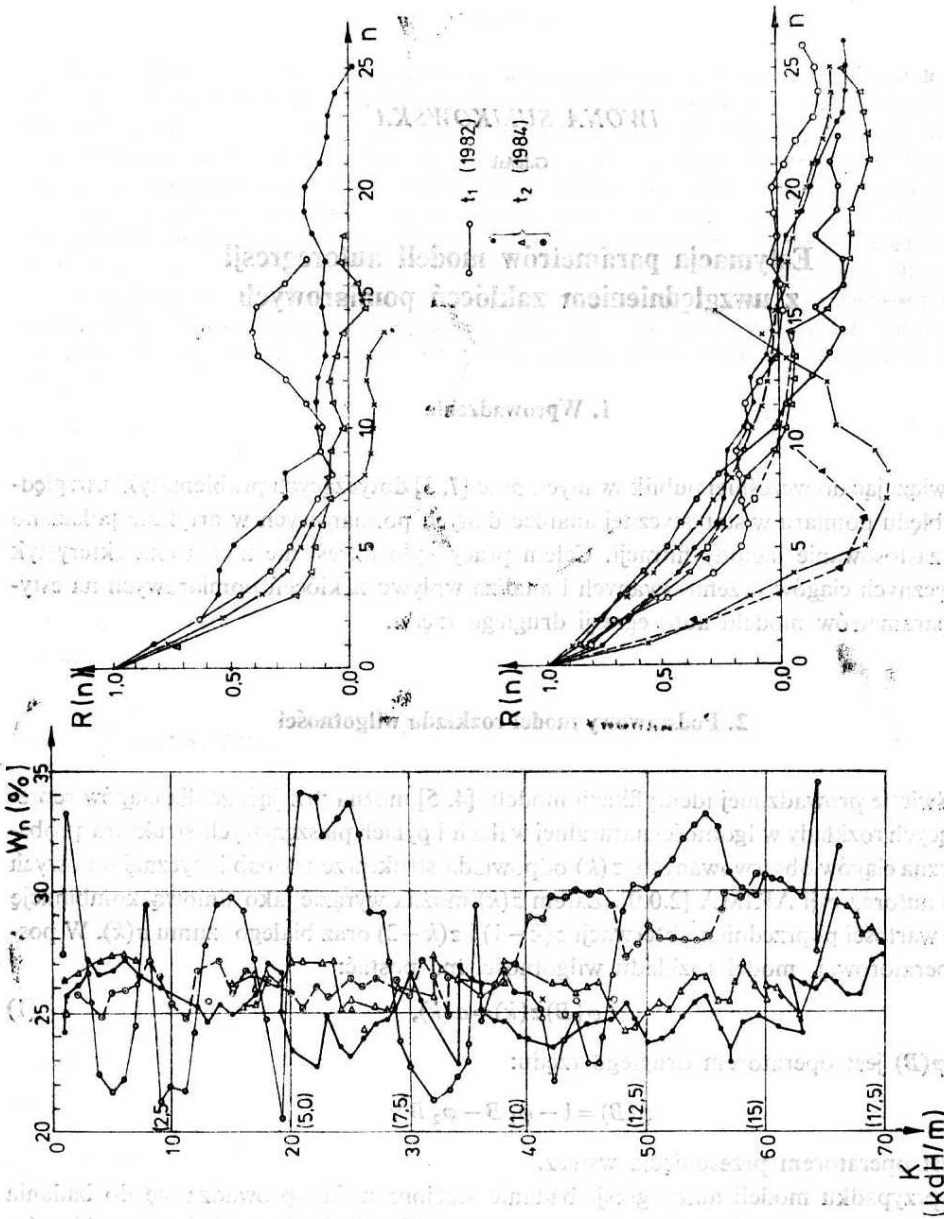
$$\varphi(B)z(k) = a(k), \quad (1)$$

gdzie  $\varphi(B)$  jest operatorem drugiego rzędu:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2,$$

a  $B$  jest operatorem przesunięcia wstecz.

W przypadku modeli autoregresji badanie stacjonarności sprowadza się do badania miejsc zerowych wielomianu charakterystycznego  $\varphi(\beta) = 0$ . Stąd wynikają warunki nałożone na parametry modelu:  $\varphi_1 + \varphi_2 < 1$ ,  $\varphi_2 - \varphi_1 < 1$  oraz  $-1 < \varphi_2 < 1$  (rys. 3). Analiza estymat funkcji autokorelacji  $r(k)$  ciągów rozkładu wilgotności naturalnej prowadzi do ograni-



Rys. 1. Przykłady funkcji autokorelacji  $r(k)$  dla ciągów pomiarowych  $z(k)$  rozkładu wilgotności naturalnej  
Uwaga: na rysunku zamiast  $n$  ma być względnie  $k$

Dr inż. J. Górniewicz, Instytut Badań i Wytworów Żywności, Państwowy Instytut Techniczny Żywności i Żywnościowych Surowców, ul. Górniewicza 11.

czenia parametrów modelu  $\varphi_1$  i  $\varphi_2$  do wielkości z obszaru zakreskowanego odpowiadającego zespolonym (obszar  $A$ ) lub rzeczywistym (obszar  $B$ ) pierwiastkom równania charakterystycznego. Przyjęty model autoregresji drugiego rzędu w zależności od wartości parametrów  $\varphi_1$  i  $\varphi_2$  opisuje funkcje autokorelacji w postaci sumy zanikających funkcji wykładniczych lub tłumionej sinusoidy [2]. Wybrane przykłady funkcji autokorelacji ciągów  $z(k)$  reprezentujących rozkłady wilgotności naturalnej w ilach i pyłach piaszczystych przedstawiono na rys. 1 i 2.

### 3. Estymacja parametrów modelu

Parametry modelu AR (2) określone jako estymatory największej wiarygodności, maksymalizują funkcję wiarygodności lub jej logarytm [2]. Określając dla funkcji gęstości prawdopodobieństwa logarytm funkcji wiarygodności i różniczkując względem parametrów modelu  $\varphi$ ,  $\sigma_a$  otrzymujemy układ równań, którego rozwiązanie określa parametry modelu AR(2). Z uwagi na trudności znalezienia dokładnego rozwiązania stosuje się do określenia estymatorów  $\hat{\varphi}_1$ ,  $\hat{\varphi}_2$ ,  $\hat{\sigma}_a^2$  metody przybliżone [2]. Pomijając niektóre z czynników otrzymuje się estymatory określone metodą najmniejszych kwadratów lub przybliżone estymatory największej wiarygodności, które są określane nie na podstawie sumy kwadratów, lecz wartości autokowariancji, przy czym dla dużej liczby danych obie metody określenia estymat prowadzą do zbieżnych wyników. Parametry modelu AR (2) można również wyrazić przez estymaty funkcji autokorelacji  $r(k)$  zastępując odpowiednie sumy iloczynów estymatorami autokowariancji  $c(k)$  i przyjmując, że  $r(k) = c(k)/c(0)$ . Dla AR(2) estymatory Yule'a-Walkera najczęściej stosowane w zagadnieniach identyfikacji modeli są określone następująco

$$\hat{\varphi}_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \quad \hat{\varphi}_2 = \frac{r_2-r_1^2}{1-r_1^2}. \quad (2)$$

Podobnie określa się estymatory  $\hat{\sigma}_a^2$  wariancji białego szumu [2]. Jako najczęściej stosowane podano estymator Yule'a-Walkera:

$$\hat{\sigma}_a^2 = c(0)(1-r_1\hat{\varphi}_1-r_2\hat{\varphi}_2). \quad (3)$$

Wartości parametrów modelu określonych jako estymatory Yule'a-Walkera dla rozpatrywanych ciągów pomiarowych rozkładu wilgotności naturalnej zestawiono w tablicy 1. Do analizy wpływu zakłóceń pomiarowych na estymaty parametrów modelu autoregresji wybrano z szeregu wyników (rys. 1) ciąg pomiarowy {ndH, 7W, 1982}. W tablicy 2 podano estymaty określone jako estymatory Yule'a-Walkera wzory (2), (3), największej wiarygodności (NW) oraz sumy kwadratów (SQ). Estymatory charakterystyk procesu reszt określono dla estymat parametrów modelu  $\hat{\varphi}$ ,  $\hat{\sigma}_a^2$  obliczonych według wzorów (2) i (3). Wartości funkcji autokorelacyjnej ciągu reszt  $r(a)$  w porównaniu z błędem standardowym pojedynczego współczynnika autokorelacji, który według [2] na rozkład normalny  $N(0; 0, 125)$  są małe.

Adekwatność modelu potwierdza również test zgodności. Zakładając, że wartość  $Q$  określona z wzoru

$$Q = n \sum_{k=1}^n r_k^2(a) \quad (4)$$

Tablica 1  
Wstępne identyfikacje parametrów modeli

Dane	Model ARIMA	$w \pm \sigma_w$	$\sigma_w^2 = C_0$	Zidentyfikowany model	$\sigma_a^2$
pyły piaszczyste	1.0.0	$20,81 \pm 1,8370$	$33,0136$ (5,7457)	$z(k) - 0,7570 z(k-1) = a(k)$	14,0551
	2.0.0	$20,81 \pm 1,8727$	$33,0136$ (5,7457)	$z(k) - 0,7431 z(k-1) - 0,01949 z(k-2) = a(k)$	14,0489
	0.1.0	$-0,033 \pm 0,436$	$15,7215$ (3,9650)	$Vz(k) = a(k)$	15,7215
realizacja ndZ, 4W, 1984	0.1.1	$-0,003 \pm 0,3883$	$15,7215$ (3,9650)	$Vz(k) = (1 - 0,159B) a(k)$	15,41
	2.0.0	-	-	$(1 - 0,8627B + 0,2325B^2) z(k) = a(k)$	-
realizacje ndZ, 4W, 1982	2.0.0	-	-	$(1 - 0,7500B + 0,5000B^2) z(k) = a(k)$	-

ma rozkład  $\chi^2$  o  $(k-p-q)$  stopniach swobody. Kwantyle 90% i 95% rozkładu  $\chi^2$  wynoszą  $Q_{\chi^2, 90} = 19,80$  i  $Q_{\chi^2, 95} = 22,40$ , podczas gdy wartość  $Q$  określona dla szeregu reszt ma wartość  $Q = 12,2892$ . Zatem pomimo bliskości niektórych współczynników autokorelacji procesu reszt  $r(a)$  błędowi standardowemu test zgodności wykazuje, że zidentyfikowany model i jego parametry są adekwatne.

Tablica 2

Estymaty parametrów modeli autoregresji wybranego ciągu pomiarowego  $z(k)$ 

Estymacja	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\sigma}_a^2$	Uwagi
Estymatory Y-W wz. (2), (3)	0,8433	-0,1151	7,3172	Observacje $z(k)$ $\sigma_a^2 = 0$ $\hat{\varphi}_z = \hat{\varphi}_z$
Estymatory (SQ)	0,8247	-0,0903	7,1559	
Estymatory (NW)	0,8263	-0,1057	7,4942	
Estymatory Y-W wz. (2), (3)	0,9680	-0,1933	4,3735	Estymowany ciąg $\hat{x}(k/k)$ $\sigma_a^2 = 3$
Estymatory (SQ)	0,9424	-0,1591	3,6768	
Estymatory (NW)	0,9679	-0,2044	2,8829	
Estymatory Y-W wz. (2), (3)	1,0604	-0,2795	3,2142	Estymowany ciąg $\hat{x}(k/k)$ $\sigma_a^2 = 6$
Estymatory (SQ)	1,0383	-0,2506	0,9625	
Estymatory (NW)	1,0535	-0,2432	0,7563	

$\hat{\varphi}_z, \hat{\sigma}_{az}$  - parametry modelu autoregresji dla ciągu estymowanego  $\hat{x}(k/k)$ ,  $\hat{\varphi}_z, \hat{\sigma}_{az}$  - parametry modelu autoregresji dla ciągu pomiarowego  $z(k)$

#### 4. Estymacja parametrów modelu z uwzględnieniem zakłóceń pomiarowych o charakterze białego szumu

Do rozdzielania niejednorodności statystycznej rozkładu wilgotności i zakłóceń metody pomiaru wykorzystano filtr KALMANA-BUCY [1]. Uzyskany estymator ma charakter równania różnicowego liniowego względem obserwacji  $z(k)$ , a zatem łatwej do realizacji komputerowej. Określona estymata  $\hat{x}(k/k)$  jest nieobciążona o minimalnej macierzy kowariancji błędu w zakresie estymatorów liniowych. W przypadku rozkładów łącznie gausowskich, przy czym założenie takie nie jest konieczne, określa warunkowe wartości oczekiwane.

##### 4.1. Ogólne sformułowanie problemu estymacji wielkości nie mierzonych w przypadku modelu ARIMA [2.0.0.]

Liniowy model (1) w przestrzeni stanów można zapisać w postaci [6]:

$$\begin{aligned} x(k+1) &= \begin{bmatrix} 0 & 1 \\ -\varphi_2 & -\varphi_1 \end{bmatrix} x(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(k) \\ z(k) &= H^T x(k) + v(k) \end{aligned} \quad (5)$$

gdzie:  $x(k)$  - jest zdefiniowane jako wektor stanu,  $z(k)$  - jest wektorem obserwacji,  $w(k)$  i  $v(k)$  - oznaczają zakłócenia na wejściu układu liniowego i zakłócenia pomiarowe.

Niepewności modelu  $x(0)$   $v(k)$ ,  $w(k)$  przy założeniu modelu Bayesa są modelowane jako wektory losowe nieskorelowane o zerowej wartości oczekiwanej dla których:

$$E\{x(0)x^T(0)\} = \psi$$

$$E\{v(k_1)r^T(k_2)\} = \begin{cases} R & \text{dla } k_1 = k_2 \\ 0 & \text{dla } k_1 \neq k_2 \end{cases} \quad (6)$$

$$E\{w(k_1)w^T(k_2)\} = \begin{cases} Q & \text{dla } k_1 = k_2 \\ 0 & \text{dla } k_1 \neq k_2 \end{cases}$$

Z uwagi na sterowalność i obserwowalność modelu (5) i dodatnią określoność  $R$  i  $Q$  błąd estymacji  $P(k/k)$  jest zbieżny do macierzy dodatnio określonej. Macierz błędu  $P_\infty$  dla układu w stanie ustalonym można otrzymać przez rozwiązanie macierzowego równania różnicowego nie zależnego od obserwacji aż do osiągnięcia stanu ustalonego. Można też znaleźć dodatnio określone rozwiązanie równania macierzowego w postaci:

$$P_\infty = \{H^T R^{-1} H + \{\phi P_\infty \phi^T + G Q G^T\}^{-1}\}^{-1}.$$

Dla AR (2) mamy

$$P_\infty = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\sigma_v^2} \end{bmatrix} + \left\{ \begin{bmatrix} 0 & 1 \\ -\varphi_2 & -\varphi_1 \end{bmatrix} P_\infty \begin{bmatrix} 0 & -\varphi_2 \\ 1 & -\varphi_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_a^2 \end{bmatrix} \right\}^{-1} \right\}^{-1}. \quad (7)$$

Dla stanu ustalonego ogólne równania rekurencyjne opisujące estymaty i ich błędy przyjmują postać:

$$\hat{x}(k+1/k+1) = \begin{bmatrix} 0 & 1 \\ -\varphi_2 & -\varphi_1 \end{bmatrix} \hat{x}(k/k) + K_\infty \{z(k+1) - [-\varphi_2, -\varphi_1] \hat{x}(k/k)\}$$

$$K_\infty = \frac{1}{\sigma_v^2} P_\infty \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$P_\infty(1/0) = \begin{bmatrix} 0 & 1 \\ -\varphi_2 & -\varphi_1 \end{bmatrix} P_\infty \begin{bmatrix} 0 & -\varphi_2 \\ 1 & -\varphi_1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_a^2 \end{bmatrix} \quad (8)$$

gdzie  $P_\infty(1/0)$  stanowi macierz kowariancji błędu dla stanu ustalonego związaną z predykcją o jeden krok.

Podaną metodą określono na podstawie różnicowych równań (8) estymaty wielkości  $\hat{x}(k/k)$ . Dla analizowanych modeli rozkładu wilgotności zachodzi przypadek filtracji ustalonej. Błąd estymacji dąży do wartości ustalonej i zależy od parametrów podstawowego modelu i wariancji zakłóceń pomiarowych. Charakterystyki statystyczne ciągów estymowanych  $\hat{x}(k/k)$  dla przyjęcia zakłóceń pomiarowych radiometrycznej metody pomiaru wilgotności o charakterze białego szumu i wariancji  $\sigma_v^2$  zestawiono w tablicy 3. Statystyczne charakterystyki ciągów wartości estymowanych:  $E\{\hat{x}(k/k)\hat{x}(k/k^T)\}$  oraz  $r(n)$  wykazują różnice w stosunku do statystycznych charakterystyk ciągu obserwacji  $z(k)$  (tabl. 3). Wylimitowanie z mierzonych wielkości zakłóceń pomiarowych prowadzi do zwiększenia wartości estymat funkcji autokorelacji i zmniejszenia wariancji ciągu.

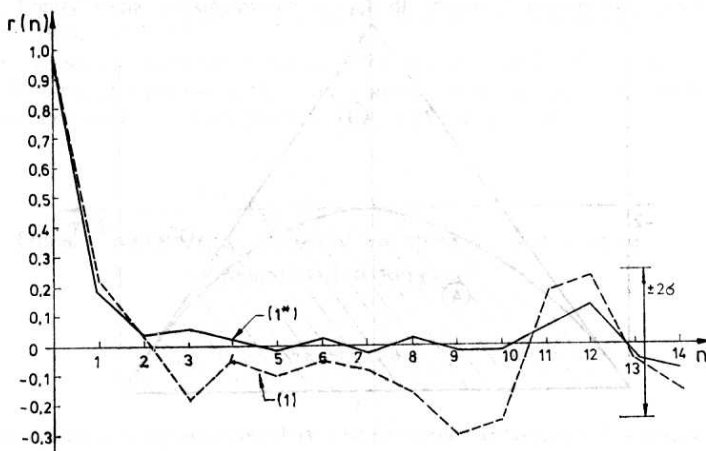
Podobne wyniki w estymatach funkcji autokorelacji z uwzględnieniem błędów metody oznaczeń wilgotności naturalnej metodą suszarkowo-wagową (rys. 2) uzyskano dla ciągów opisujących zmiany wilgotności w łąkach [4].

Tablica 3

Charakterystyki statystyczne ciągów estymowanych  $\hat{x}(k/k)$  dla modelu AR (2)

Ciąg $z(k)$									
$E\{z(k)\} = 27,6439$									
$E\{z(k)z^T(k)\} = 17,4130$									
$r(1)$	$r(2)$	$r(3)$	$r(4)$	$r(5)$	$r(6)$	$r(7)$	$r(8)$	$r(9)$	
0,7561	0,5223	0,3598	0,2659	0,2109	0,1668	0,0669	0,0	-0,095	
Ciąg $\hat{x}(k/k)$									
$\sigma_x^2 = 3,000$									
$E\{\hat{x}(k/k)\} = 0,0378$									
$E\{x(k)x^T(k)\} = 13,2814$									
$r(1)$	$r(2)$	$r(3)$	$r(4)$	$r(5)$	$r(6)$	$r(7)$	$r(8)$	$r(9)$	
0,8109	0,5912	0,4334	0,3180	0,2325	0,1536	0,0572	-0,0013	-0,0489	
Ciąg $\hat{x}(k/k)$									
$\sigma_x^2 = 6,000$									
$E\{\hat{x}(k)\} = 0,0535$									
$E\{\hat{x}(k)\hat{x}^T(k)\} = 11,1270$									
$r(1)$	$r(2)$	$r(3)$	$r(4)$	$r(5)$	$r(6)$	$r(7)$	$r(8)$	$r(9)$	
0,8286	0,5989	0,4359	0,3206	0,2489	0,1891	0,1109	0,0246	-0,0938	

Miarą dokładności estymacji jest zmienna resztowa, którą można podobnie jak w przypadku ciągu pomiarowego przyjąć jako biały szum o określonym estymatorze wariancji. Estymaty parametrów modelu określone według metod: SQ, NW i Y-W dla ciągów przefiltrowanych zestawiono w tabl. 2. Proces  $z(k)$  potraktowany jako suma dwóch niezależnych procesów, a mianowicie procesu generowanego przez podstawowy model rozkładu wilgotności ARIMA [p. 0, 0,] i procesu zakłóceń pomiarowych o charakterze białego szumu nie wykazuje zmiany rzędu  $p$  procesu autoregresji [2]. Jego teoretyczna funkcja autokorelacji przy założeniu braku korelacji szumu pomiarowego również nie ulega zmianie. Ponieważ jednak parametry funkcji autokorelacji są jedynie estymatorami stąd estymaty uzyskane dla ciągów przefiltrowanych wyrażają różnice w stosunku do ciągów obserwacji

Rys. 2. Przykład funkcji autokorelacji  $r(k)$ (1) dla ciągu pomiarowego  $z(k)$ , (1\*) dla ciągu estymowanego (bez zakłóceń pomiarowych)Uwaga: na rysunku zamiast  $n$  ma być  $k$

$z(k)$  (tabl. 3). Te zaś prowadzą do zmiany estymat parametrów modelu  $\hat{\varphi}_1, \hat{\varphi}_2, \hat{\sigma}_a$  (tabl. 2). Określmy wariancję i kowariancję parametrów modelu autoregresji AR (2) 1 jako:

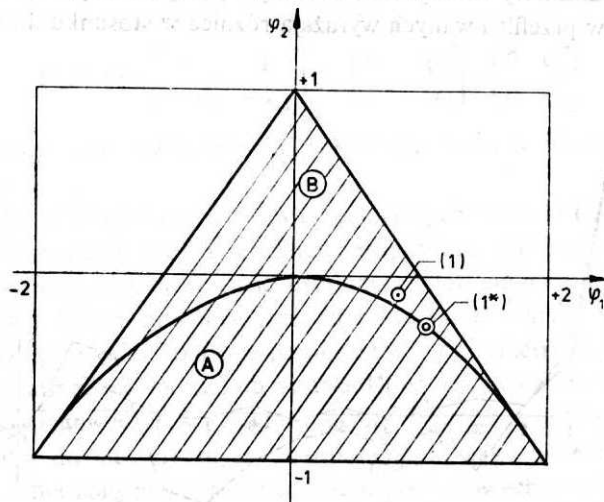
$$V(\hat{\varphi}_1, \hat{\varphi}_2) = \frac{1}{n} \left( 1 - [r_1, r_2] \begin{bmatrix} \hat{\varphi}_1 \\ \hat{\varphi}_2 \end{bmatrix} \right) \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}. \quad (9)$$

Uwzględniając błędy standardowe estymat i ich kowariancję (tabl. 4) należy stwierdzić, że parametry ciągów estymowanych  $\hat{\varphi}_{x1}, \hat{\varphi}_{x2}$  leżą w granicach przedziału  $\pm 2\sigma$  dla mniejszej wariancji zakłóceń pomiarowych i nieznacznie poza granicami przedziału dla założonej większej wariancji zakłóceń pomiarowych. Analiza błędu pomiarowego jest szczególnie

Tablica 4  
Wariancje i kowariancje estymatorów Yule'a-Walkera dla ciągu  $z(k)$  i ciągów estymowanych

Ciąg	Błąd standardowy	Współczynnik korelacji
$z(k)$	$\pm 0,0812$	0,7561
$\hat{x}(k/k)$ $\sigma_v^2 = 3,0$	$\pm 0,0717$	0,8109
$\hat{x}(k/k)$ $\sigma_v^2 = 6,0$	$\pm 0,0671$	0,8286

przydatna w przypadku parametrów modeli bliskich niestacjonarnym lub w przypadku parametrów o wartościach bliskim granicznemu rozdzielających poszczególne typy funkcji autokorelacyjnych. Na przykład w rozpatrywanym przypadku ciągu pomiarowego założenie wariancji błędu metody oznaczania wilgotności naturalnej  $\sigma_v^2 = 6\%$  prowadzi do zmiany parametrów modelu z obszaru A do B (rys. 3). Zatem ustalenie charakterystyki zakłóceń pomiarowych i ich eliminacja może ułatwić interpretację mierzonych wielkości.



Rys. 3. Obszar dopuszczalnych estymat parametrów modelu  $\hat{\varphi}_1, \hat{\varphi}_2$  z uwagi na stacjonarność modelu dla ciągów pomiarowych

(1) Estymaty parametrów  $\hat{\varphi}_1, \hat{\varphi}_2$  dla ciągu pomiarowego  $z(k)$ , (1\*) Estymaty parametrów  $\hat{\varphi}_1, \hat{\varphi}_2$  dla ciągu estymowanego (bez zakłóceń pomiarowych)



## 5. Podsumowanie

Dla ciągów reprezentujących rozkłady wilgotności naturalnej przyjęto, że struktura probabilistyczna ciągu odpowiada liniowemu modelowi autoregresji drugiego rzędu. Estymatory parametrów modelu ARIMA [2.0.0] określone różnymi metodami: SQ, NW, Y-W jak wykazano na przykładzie analizowanej realizacji nie wykazują dużych różnic co świadczy o poprawnie przyjętym modelu.

Celem uwzględnienia zakłóceń pomiarowych o charakterze białego szumu zastosowano metodę filtracji Kalmana-Bucy. Uzyskane estymaty charakterystyk statystycznych procesów przefiltrowanych (tab. 3) wykazują różnice w stosunku do statystycznych charakterystyk procesu obserwowanego.

Podana metoda filtracji może być stosowana do identyfikacji parametrów modelu autoregresji w przypadku ciągów pomiarowych  $z(k)$  zawierających zakłócenia pomiarowe o ustalonej wariancji. Reasumując można przyjąć, że w przypadku pomiarów podstawowych własności ośrodka gruntowego obciążonych błędami metody pomiarowej zastosowanie filtru Kalmana-Bucy umożliwia interpretację wpływu tych zakłóceń na statystyczne charakterystyki.

## LITERATURA

1. B. D. ANDERSEN, J. B. MOORE, Filtracja optymalna, WNT 1984.
2. G. E. P. BOX, G. M. JENKINA, Analiza szeregów czasowych, PWN, Warszawa 1983.
3. I. SULIKOWSKA, Metoda filtracji w estymacji funkcji prawdopodobieństwa charakterystyk gruntowych, Arch. Hydrot. Tom XXXI z. 4, 1984.
4. I. SULIKOWSKA, Statystyczne charakterystyki procesów losowych na przykładzie pomiarów wilgotności naturalnej, Arch. Hydrot. Tom XXXIII, z. 1 - 2, 1986.
5. I. SULIKOWSKA, Identyfikacja podstawowych modeli dla prostych parametrów gruntowych. Opr. wew. IBW PAN, 1986.
6. F. G. SCHWEPPE, Układy dynamiczne w warunkach losowych, WNT, Warszawa 1978.
7. P. WILDE, I. SULIKOWSKA, Estimation of probability function of soil properties on the basis of measurements with experimental error, Arch. Hydrot. Tom XXXI, z. 3, 1984.

## Оценка параметров моделей авторегрессии с учетом погрешностей измерений

### Содержание

Для последовательностей, представляющих распределения естественной влажности, принято, что вероятностная структура последовательности соответствует линейной модели авторегрессии второго порядка. Оценки параметров модели [ARIMA 2.0.0], полученные разными методами (табл. 2), для анализируемой реализации не проявляют больших различий, что свидетельствует о правильности принятой модели.

Разделение неоднородности анализируемого свойства грунта и погрешностей измерительного метода получено путем фильтрации Кальмана-Буци.

Полученные оценки статистических характеристик последовательности (табл. 3) указывают на целесообразность анализа погрешностей измерений при определении оценок параметров модели в практическом применении.

### Estimation of parameters of the autoregression models considering the measurement error

#### Summary

The filtering method was applied for identification of the parameters of autoregression models. The measurement error was assumed as a process of the white noise. The differences in the statistics characteristics of both filtered and measured processes are shown for the natural water content measurements.