

*Zdzisław Iwulski**

PROGNOZA WYSTĄPIENIA WSTRZĄSU ZA POMOCĄ SZEREGÓW CZASOWYCH

1. Wprowadzenie

Z szeregami czasowymi spotykamy się w inżynierii, geologii, ekonomii oraz w wielu innych naukach. Aby przedstawić wybrane techniki analizy tych szeregów, niezbędny jest znalezienie teoretycznego modelu, zwykle probabilistycznego, który możliwie najlepiej odzwierciedli dane. Po wybraniu modelu zajmujemy się estymacją parametrów, sprawdzaniem jakości jego dopasowania do danych oraz prognozowaniem. Model taki służy nam do opisu zjawiska oraz danych lub do wyodrębnienia części deterministycznej, czyli trendu bądź sezonowości.

Możliwość optymalnego prognozowania, zrozumienia związków dynamicznych pomiędzy zmiennymi i optymalnego sterowania mają wielkie znaczenie w praktyce. Jeżeli celem analizy jest prognoza, to dopasowany model statystyczny wykorzystuje się bezpośrednio do opracowania optymalnej prognozy poprzez proste obliczenia rekurencyjne. W szczególności model ten określa w pełni czy prognozy będą liniowe, wykładnicze i tak dalej. Ponadto, dopasowany model pozwala dokładnie zobaczyć, w jaki sposób prognozy wykorzystują dane z przeszłości, określić wariancje błędów prognozy oraz obliczyć przedziały, w których z zadaniem prawdopodobieństwem będą znajdowały się przyszłe wartości szeregów. Jeżeli model przedstawia również związki dynamiczne, to odpowiednie cykle identyfikacji, dopasowania i sprawdzenia diagnozy pozwalają na otrzymanie stochastycznego modelu funkcji przetwarzającej.

W większości metod statystycznych mamy do czynienia z modelami, w których zakłada się, że obserwacje są niezależne. W wielu zastosowaniach zależność pomiędzy obserwacjami traktowana jest jako przeszkoda, w doświadczeniach planowanych wprowadza się randomizację doświadczeń, co pozwala na przeprowadzenie takiej analizy, jak w przypadku obserwacji niezależnych. Niemniej różnorodne dane w handlu, ekonomii, technice i nau-

* Wydział Górnictwa i Geoinżynierii, Akademia Górniczo-Hutnicza, Kraków

kach przyrodniczych występują w postaci szeregów czasowych, w których obserwacje są zależne i zależność ta jest interesująca sama w sobie. Metody analizy takich ciągów obserwacji zależnych nazywane są analizą szeregów czasowych.

Celem jest otrzymanie modelu maksymalnie prostego oraz mającego minimalną liczbę parametrów i przy tym adekwatnie opisującego otrzymane obserwacje.

Otrzymanie takiego modelu jest ważne z następujących powodów:

- takie modele mogą dać informacje o charakterze układu generującego szereg czasowy;
- można z nich skorzystać dla otrzymania optymalnych prognoz przyszłych wartości szeregu.

Proponowane w tym artykule statystyczne podejście do analizy wstrząsów wynika z losowego charakteru tego zjawiska. Istnieje cały zespół wzajemnie nakładających się czynników, powodujących zaistnienie wstrząsu. Istnieją więc przesłanki do tego, aby zastosować statystyczną analizę do opisu tego zjawiska.

2. Próba analizy danych

W artykule tym przedstawiam przykładową analizę szeregu czasowego dla wybranego oddziału w KGHM Polska Miedź SA. Naturalnym narzędziem służącym do przeprowadzenia tejsze analizy są rekursje Kalmana oraz reprezentacja w przestrzeni stanów, ze względu na sekwencyjny napływ danych, z jakim mamy tutaj do czynienia.

Zgromadzone są obserwacje dotyczące wstrząsów górotworu między innymi:

- czasu wystąpienia wstrząsu,
- wielkości wydzielonej energii.

Za pomocą procedury agregacji (własny program AG) wyznaczono sumaryczne wartości energii wyzwolonej w poszczególnych dniach, a następnie przy pomocy procedury TS utworzono szereg czasowy wyzwolonej energii uzupełniając zagregowane dane zerami w dniach, w których nie obserwowano wstrząsów. Tak uzyskany szereg czasowy poddano stabilizującej transformacji:

$$z_t = \log_{10}(1 + \text{energ_agr}_t),$$

uzyskując podstawowy szereg czasowy z_t , $t = 1, \dots, N$. Szereg czasowy z_t traktujemy jako obserwację ciągu zmiennych losowych $Z(t, \omega)$, $t = 1, \dots, N$ zwykle stochastycznie zależnych. Wykrycie tej zależności stanowi podstawowy problem w analizie szeregów czasowych, którą można przeprowadzić w dziedzinie czasu i w dziedzinie częstotliwości. Aby taka analiza była możliwa, szereg czasowy musi być przynajmniej stacjonarny w szerszym sensie.

Jeśli założenie stacjonarności nie jest możliwe do przyjęcia, to należy szereg czasowy poddać transformacji tak, aby przekształcony szereg można było uznać za szereg stacjonarny w szerszym sensie. Stacjonarny szereg czasowy można analizować w dziedzinie czasu i w dziedzinie częstotliwości.

Analiza w dziedzinie czasu polegać będzie na wyestymowaniu funkcji:

- autokorelacji ACF,
- autokorelacji cząstkowej PACF

i ich jakościowej analizie.

Większość empirycznych szeregów czasowych można z powodzeniem aproksymować przy użyciu jednego z pięciu podstawowych modeli, który można zidentyfikować w oparciu o kształt autokorelogramu (ACF) i autokorelogramu cząstkowego (PACF).

Jeżeli struktura szeregu jest niewyraźna, wariancja składnika losowego jest znaczna, a my mimo to potrzebujemy nie tylko odkryć ukryty wzorzec, ale także generować prognozy, to możemy próbować użyć metodologii procesów ARIMA (*Autoregressive Integrated Moving Average*) rozwiniętej przez Boxa i Jenkinsa (1976). Zyskała sobie ona niezwykłą popularność w wielu dziedzinach, a praktyka badawcza potwierdza jej moc i elastyczność. Mimo swych zalet i elastyczności, ARIMA jest techniką złożoną; nie jest łatwa w użyciu, wymaga dużego doświadczenia i chociaż często daje zadowalające wyniki, to zależą one w dużym stopniu od badacza.

Skałkowany proces autoregresji i średniej ruchomej ARIMA (p, d, q) (*Autoregressive Integrated Moving Average*) z_t jest zdefiniowany równością:

$$\Phi_p(B)(1-B)^d z_t = const + \Theta_q(B)\varepsilon_t,$$

gdzie:

- B — operator przesunięcia wstecz tzn. $Bz_t = z_{t-1}$,
- D — rząd operatora różnicowego $(1-B)$,
- $\Phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ — operator autoregresji rzędu p (w skrócie $AR(p)$),
- $\Theta_q(B) = 1 - \theta_1 B + \dots - \theta_q B^q$ — operator średniej ruchomej rzędu q (w skrócie $MA(q)$),
- ε_t — ciąg niezależnych zmiennych losowych z zerową wartością oczekiwaną i wariancją σ^2 .

Zakładamy, że są spełnione następujące warunki stacjonarności i odwracalności (tzn. stabilność wielomianów w sensie Schura):

- pierwiastki (w ciele liczb zespolonych) wielomianu $\Phi(B)$ leżą na zewnątrz koła jednostkowego, czyli proces z_t jest stacjonarny (warunek stacjonarności);
- pierwiastki (w ciele liczb zespolonych) wielomianu $\Theta(B)$ leżą na zewnątrz koła jednostkowego (warunek odwracalności).

Dla wielomianów niskich stopni warunki stacjonarności i odwracalności można wypisać w sposób jawny:

— warunki stacjonarności

$$p = 1 \quad |\varphi_1| < 1,$$

$$p = 2 \quad |\varphi_2| < 1, |\varphi_1| < 1 - \varphi_2,$$

$$p = 3 \quad |\varphi_3| < 1, 1 - \varphi_1 - \varphi_2 - \varphi_3 > 0, 1 + \varphi_1 - \varphi_2 + \varphi_3 > 0, 1 - \varphi_3^2 + \varphi_2 + \varphi_1\varphi_3 > 0;$$

— warunki odwracalności

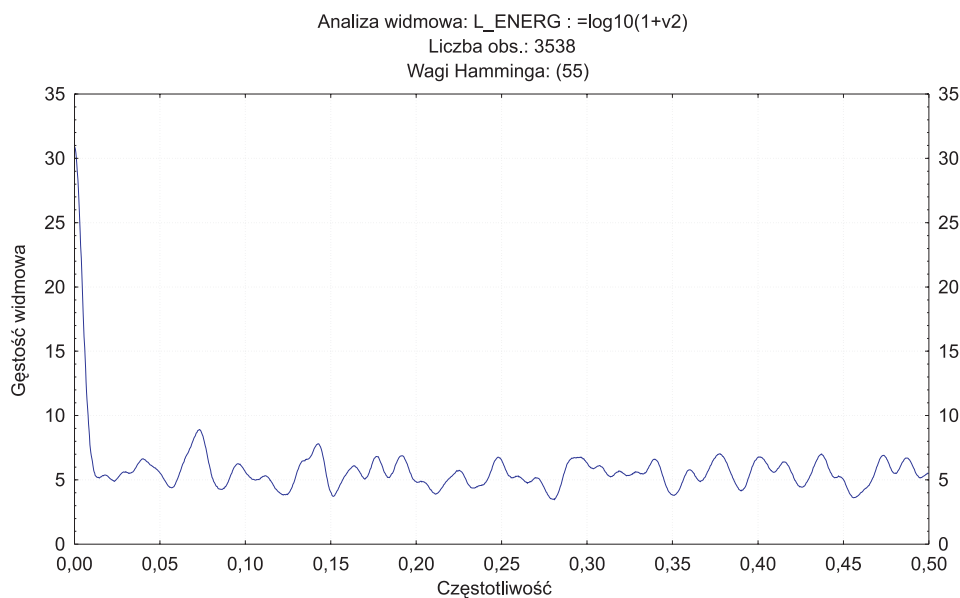
$$q = 1 \quad |\theta_1| < 1,$$

$$q = 2 \quad |\theta_2| < 1, |\theta_1| < 1 - \theta_2,$$

$$q = 3 \quad |\theta_3| < 1, 1 - \theta_1 - \theta_2 - \theta_3 > 0, 1 + \theta_1 - \theta_2 + \theta_3 > 0, 1 - \theta_3^2 + \theta_2 + \theta_1\theta_3 > 0.$$

2.1. Analiza w dziedzinie częstotliwości

Dla danych, które traktujemy jako obserwacje szeregu czasowego stacjonarnego (przynajmniej w szerszym sensie), wyestymowano gęstość widmową na rysunku 1.



Rys. 1. Analiza widmowa

Na przedstawionym rysunku widać, że gęstość widmowa ma:

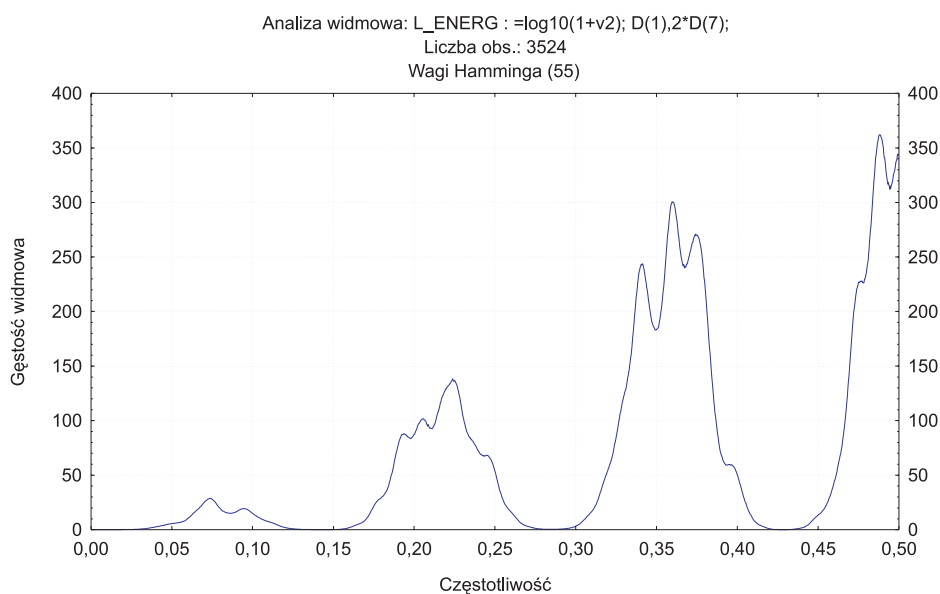
- wyraźne maksimum w pobliżu 0, co wskazuje na dominujący charakter niskich częstotliwości;
- kilka mniejszych pików okresowo powtarzających się, co sugeruje istnienie okresowych składowych.

Struktura częstotliwościowa jest więc niewyraźna.

W celu wykrycia wyraźnej struktury zastosujemy do badanego szeregu operatory różnicowania:

- niesezonowego $D(1) = (1 - B)$,
- sezonowego $D(7) = (1 - B^7)$,

co zostało przedstawione na rysunku 2.

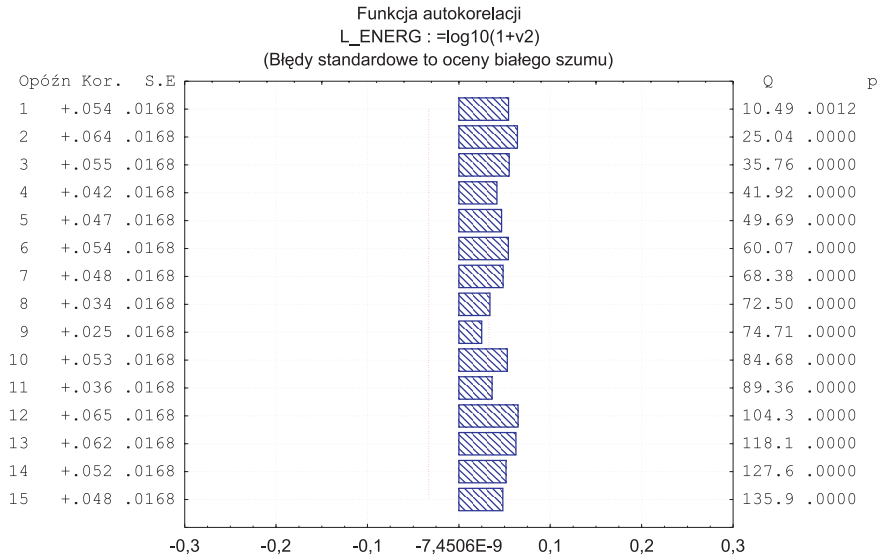


Rys. 2. Analiza widmowa dla różnicowania niesezonowego i sezonowego

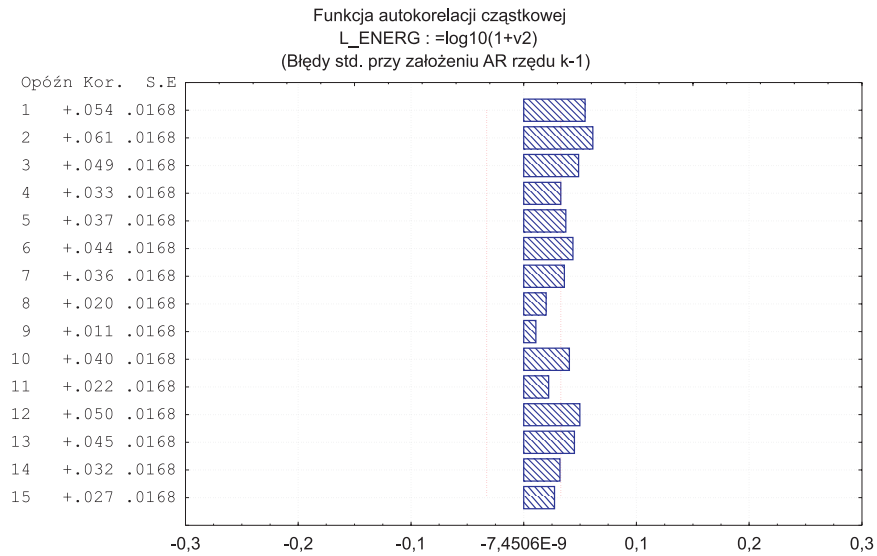
2.2. Analiza w dziedzinie czasu

W celu sprawdzenia poprawności przyjętego modelu matematycznego została wykonana funkcja autokorelacji (rys. 3) i funkcja autokorelacji cząstkowej (rys. 4).

Struktura korelacyjna badanego szeregu jest niewyraźna, co było do przewidzenia po przeprowadzeniu analizy widmowej. Następnie przekształcono szereg wyjściowy poprzez operację różnicowania i przeanalizowano tak przekształcony szereg.



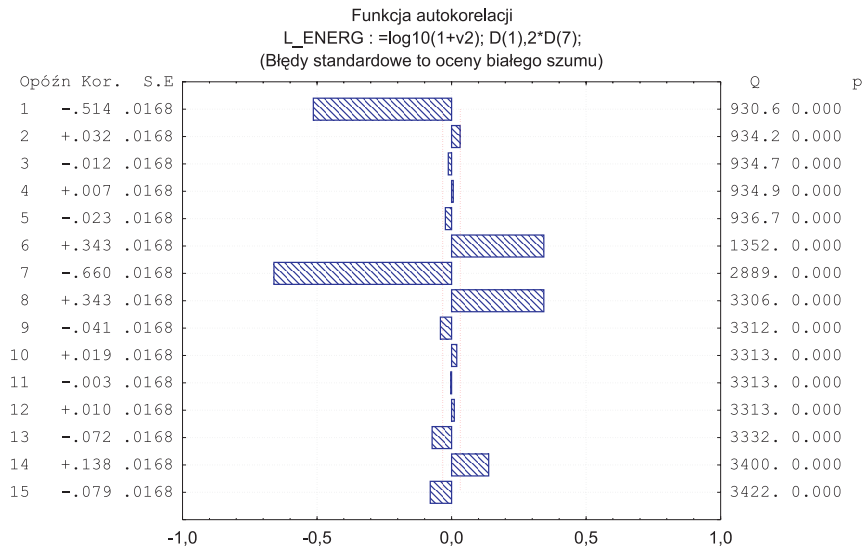
Rys. 3. Funkcja autokorelacji



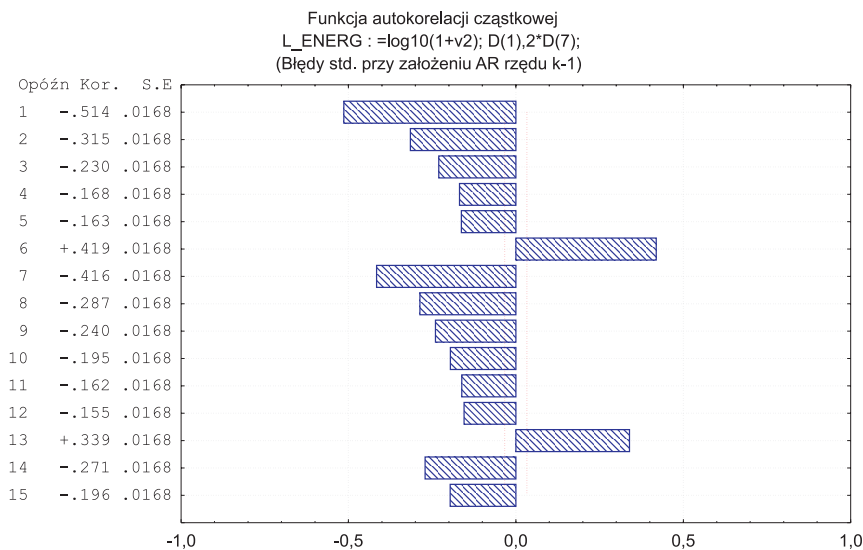
Rys. 4. Funkcja autokorelacji cząstkowej

2.3. Analiza szeregu po zastosowaniu niesezonowego i sezonowego różnicowania

Na rysunku 5 zostały przedstawione błędy standardowe do białego szumu dla funkcji autokorelacji oraz na rysunku 6 pokazano wielkość błędów standardowych przy założeniu AR rzędu $k - 1$ dla autokorelacji cząstkowej.



Rys. 5. Błędy standardowe do białego szumu



Rys. 6. Błędy standardowe przy założeniu AR rzędu $k - 1$

Kształt ACF i PACF dla szeregu różnic sugeruje możliwość użycia modelu SARIMA $(p,d,q) \times (P,D,Q)_7$ z parametrami:

$d = 1$ — co odpowiada niesezonowemu różnicowaniu szeregu wyjściowego;

$D = 2$ — co odpowiada dwukrotnemu sezonowemu różnicowaniu szeregu wyjściowego.

Model estymowany SARIMA (0,1,1)×(2,2,1) z opóźnieniem sezonowym 7 można zadać według następujących danych wejściowych:

$$\text{Dane: } L_ENERG = \log_{10}(1 + v_2),$$

$$\text{Przekształcenia: } D(1), 2 * D(7),$$

$$\text{Model: SARIMA}(0,1,1)(2,2,1) \text{ Opóź. sezon.: } 7 \text{ Resztowy } MS = 4,855.$$

Dla takich założeń otrzymano wyniki dla modelu SARIMA, które przedstawiono w tabeli 1, a rezultaty prognoz występowania wstrząsów w tabeli 2.

TABELA 1

Uzyskane wyniki dla modelu SARIMA

	Parametr	Asympt. std.	Asympt. t Błąd	p	Dolna gr. 95% p. ufn	Górna gr. 95% p. ufn
$q(1)$	0,804385	0,009693	82,9847	0,00	0,785381	0,823390
$Ps(1)$	-0,641031	0,016722	-38,3335	0,00	-0,673818	-0,608245
$Ps(2)$	-0,301941	0,016724	-18,0544	0,00	-0,334731	-0,269152
$Qs(1)$	0,900426	0,006799	132,4350	0,00	0,887095	0,913756

TABELA 2

Rezultaty obliczeń

Data	Prognoza energii log E	Dolne 90%	Górne 90%	Obserwowana energia log E
1.12.	0,4818	-1,7957	2,7593	0,0
2.12.	4,4017	2,0810	4,7225	0,0
3.12.	-0,0630	-2,3403	2,2142	0,0
4.12.	0,8343	-1,4423	3,1110	0,0
5.12.	0,9039	-1,3729	3,1809	0,0
6.12.	2,0474	-0,2295	4,3244	4,9
7.12.	1,9554	-0,3219	4,2329	0,0
8.12.	3,4031	0,9529	5,8533	7,1
9.12.	-0,2249	-2,5023	2,0524	0,0
10.12.	0,6605	-1,6203	2,9415	2,4
11.12.	4,2626	1,8127	6,7125	5,2
12.12.	2,5961	0,3151	4,8770	0,0
13.12.	3,6586	1,3775	5,9396	0,0

TABELA 2 cd.

Data	Prognoza energii log E	Dolne 90%	Górne 90%	Obserwowana energia log E
14.12.	1,8805	-0,4008	4,1619	0,0
15.12.	4,2565	1,9752	6,5379	0,0
16.12.	0,3030	-1,9783	2,5844	0,0
17.12.	1,4819	-0,7991	3,7630	0,0
18.12.	0,7143	-1,6102	3,0388	0,0
19.12.	-1,1929	-3,4738	1,0879	0,0
20.12.	0,9497	-1,3309	3,2304	0,0
21.12.	0,1598	-2,1205	2,4402	0,0
22.12.	1,6593	-0,62078	3,9393	5,2
23.12.	-0,5545	-2,8357	1,7260	0,0
24.12.	4,3220	1,8315	6,8124	6,4
25.12.	0,7766	-1,5492	3,1025	0,0
26.12.	0,2222	-2,0600	2,5048	0,0
27.12.	1,7373	-0,5447	4,0195	0,0
28.12.	-0,1891	-2,4711	2,0928	0,0
29.12.	4,5616	2,2800	6,8433	0,0
30.12.	-1,0778	-3,3605	1,2049	0,0
31.12.	5,4623	3,0101	7,9146	8,3

Rezultaty dokonanych obliczeń występowania wstrząsów oprócz przedstawienia ich powyżej (tab. 2) zostały przedstawione w postaci graficznej na rysunku 7.

Wynika z nich, że następujące parametry wynoszą:

— czułość prognozy:

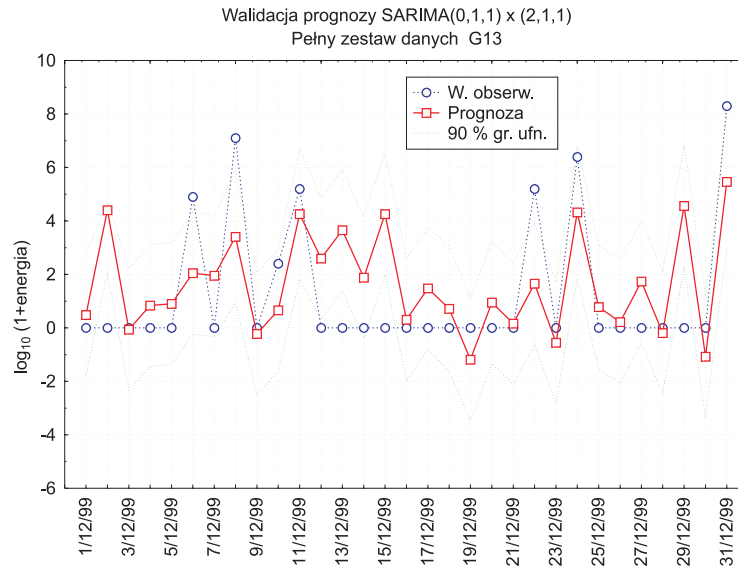
$$\frac{4}{4+2} 100\% = 67\%,$$

— specyficzność prognozy:

$$\frac{19}{19+6} 100\% = 76\%,$$

— całkowita dokładność:

$$\frac{4+19}{4+2+19+6} 100\% = 74\%.$$



Rys. 7. Walidacja prognozy

3. Podsumowanie

W prawdzie widzimy, że prognozy modelu nieco odbiegają od rzeczywistego procesu, jednak różnice te zachowane są w przedziale ufności. Analizując powyższe prognozy dochodzimy do wniosku, że ten model zachowuje się bardzo „ostrożnie” — nie prognozuje on bardzo dużych wartości wytworzonej energii, a przecież wiemy, że z takimi również mamy do czynienia. Być może, w celu uzyskania jeszcze lepszych prognoz, należałoby w dalszej analizie wykorzystać inne zmienne i zbudować model uwzględniający nie tylko przeszłe dane, lecz również inne czynniki towarzyszące wstrząsom, które mierzy się w kopalniach.

Przydatność tego typu modeli w prognozowaniu wystąpienia wstrząsu na dzień dzisiejszy nie gwarantuje całkowitego sukcesu. Wydaje się, że jeżeli szereg energii jest w ogóle sensownie prognozowalny, to należy oprócz tego szeregu uwzględnić szeregi towarzyszące — takie jak konwergencja, sejsmoakustyka itd., i próbować budować model funkcji przenoszenia.

LITERATURA

- [1] Hannan.E.J, Deistler M.: The statistical theory of linear systems. Wiley 1988
- [2] Wei W.W.S.: Time series analysis. Addison Wesley 1990
- [3] Box G.E.P, Jenkins G.M.: Analiza szeregów czasowych. Warszawa, PWN 1983
- [4] Iwulski Z.: Sposób budowania modelu matematycznego dla obliczeń przy wykorzystaniu szeregów czasowych
- [5] Iwulski Z.: Próba wykorzystania szeregów czasowych do prognozy występowania wstrząsów