

MARIOLA KĘDRA\*

ZASTOSOWANIE STATYSTYCZNEGO TESTU SSS  
DO ZIDENTYFIKOWANIA ISTNIENIA  
KRÓTKOTERMINOWEJ DYNAMIKI  
W PRZEPLÝWACH DOBOWYCH ZLEWNI RZEKI RABY

APPLICATION OF SSS TEST IN IDENTIFICATION  
OF SHORT-TERM DYNAMICS IN DAILY STREAM  
FLOW IN THE RABA RIVER BASIN

Streszczenie

W artykule badano dobową zmienność procesu przepływu rzeką w zlewni Raby w celu określenia jego deterministycznej bądź losowej natury. Na podstawie szeregów czasowych przepływów dobowych dla posterunków na cieku głównym i dopływach przeprowadzono test statystyczny SSS. Jego wyniki uzyskane dla wszystkich rozważanych przekrojów wodowskazowych wskazują na istnienie krótkoterminowej dynamiki potwierdzającej nielosową, deterministyczną naturę badanego procesu przepływu.

*Słowa kluczowe: przepływ dobowy, determinizm, losowość, test statystyczny SSS*

Abstract

In the paper deterministic or stochastic nature of daily flow in the Raba river catchment was investigated. The statistical test SSS was based on the series of daily flows from gauging stations in Raba basin. The results of the test obtained for all cross-sections show that there is short-term dynamics in irregular fluctuations of flow, so the determinism was detected by this test.

*Keywords: daily flow, determinism, randomness, statistical test SSS*

---

\* Dr Mariola Kędra, Instytut Inżynierii i Gospodarki Wodnej, Wydział Inżynierii Środowiska, Politechnika Krakowska.

### Oznaczenia

$AMI(\tau)$	– funkcja uśrednionej informacji wzajemnej
$N_{DZ}$	– liczba zbiorów danych zastępczych
$n$	– długość szeregu czasowego
$p$	– prawdopodobieństwo
$Q$	– natężenie przepływu
$\overline{SQ}$	– przepływ średni z wielolecia
$t$	– czas
$\Delta t$	– czas próbkowania
$\alpha$	– poziom istotności testu
$\tau$	– czas opóźnienia

### 1. Wstęp

Procesy i zjawiska zachodzące w przyrodzie, a w szczególności w obszarze zlewni rzecznej, mogą być postrzegane jako deterministyczne (nieprzypadkowe) bądź losowe. Jeśli układ czy proces jest deterministyczny, to na podstawie jego przeszłych zachowań można jednoznacznie określić przyszłe, zaś jego losowość pozwala to uczynić jedynie z pewnym prawdopodobieństwem. Rozpatrywanie zagadnienia determinizmu bądź losowości procesu jest badaniem jego istoty (natury), która może ukrywać rzeczywisty mechanizm generujący ten proces czy zjawisko<sup>1</sup>.

Dotychczas stosowane metody m.in. analizy korelacyjnej, spektralnej i falkowej pozwalają na identyfikację determinizmu w postaci różnej długości cykli, a więc służą do badania okresowości procesu czy zjawiska. Nowe możliwości w zakresie określania natury deterministycznej bądź losowej procesów i zjawisk stwarzają metody statystyczne oparte na tzw. danych zastępczych (ang. *surrogate data*) [6]. Są to dane generowane z danych źródłowych (na podstawie stosownych algorytmów), mające właściwości związane ze stawianą hipotezą zerową. Dane zastępcze są zatem danymi, które mogłyby się pojawić w wyniku niezależnie powtarzanych obserwacji procesu, który generował dane źródłowe, jeśli proces ten byłby zgodny z hipotezą zerową. Jednym z najnowszych takich testów statystycznych stosujących dane zastępcze jest test SSS [5]. Test ten pozwala zidentyfikować krótkoterminową dynamikę (determinizm) w szeregach czasowych o nieregularnym przebiegu, również tych zawierających długoterminowe trendy oraz cykle, stąd też jego zaletą jest uniwersalność w odniesieniu do danych<sup>2</sup>.

W niniejszym artykule zaprezentowano wyniki zastosowania testu SSS do zidentyfikowania istnienia krótkoterminowej dynamiki w przepływach dobowych zlewni rzeki Raby; rozpoznanie tej dynamiki jest zasadniczym celem artykułu. Posiadane zbiory danych były uprzednio poddane analizie spektralnej Fouriera [3], która ujawniła długoterminowy determinizm w postaci okresowości rocznej i 3-miesięcznej przepływu w zlewni rzeki Raby. Test SSS służy zaś zbadaniu, czy poza wykrytymi cyklami (długoterminowymi) istnieje

<sup>1</sup> Wykrycie determinizmu może m.in. otworzyć drogę do badań pod kątem istnienia chaosu deterministycznego.

<sup>2</sup> Dane mogą być regularne bądź nie, zawierać trendy i cykle albo być ich pozbawione.

również dynamika **krótkoterminowa** (w zakresie od jednego do kilkunastu dni; w artykule przyjęto maks. 20 dni).

## 2. Obszar i metodyka badań

Rzeka Raba jest jednym z głównych karpackich, prawobrzeżnych dopływów Wisły. Przekrój ujściowy Raby zlokalizowany w 134,7 km biegu Wisły zamyka zlewnię o powierzchni 1537 km<sup>2</sup>. Całkowita długość rzeki wynosi 132 km (rys. 1) [4].

Dane na temat przepływów w ciekach zlewni Raby pochodzą z obserwacji przeprowadzonych w konkretnych przekrojach wodowskazowych. Nazwy tych przekrojów i ich lokalizacja, a także okresy obserwacji, jakie wzięto pod uwagę, podano w tab. 1. Nie analizowano ciągów obserwacyjnych z lat 1986–1989 dla przekroju Gdów, ponieważ w okresie tym trwała budowa i napełnianie zbiornika retencyjnego w Dobczycach na rzece Rabie. Natomiast dane (rys. 1) z wodowskazu w Proszówkach, który leży znacznie niżej zbiornika, brane były pod uwagę w całości. Wartości przepływów z przekrojów w Dobczycach (1991–2002) i Gdowie (1991–1996) obarczone są wpływem pracy zbiornika, toteż są rozpatrywane z uwzględnieniem tego faktu.

W sumie przeanalizowano 21 ciągów wartości natężenia przepływu w wymienionych w tablicy ciekach. Jak wynika z tabeli, długość ciągów jest zróżnicowana. Zawierają one od 2191 do 15 340 codziennych pomiarów przepływów z okresów będących wielokrotnością roku (od 6 do 42 lat). Najkrótszy (2191-elementowy) ciąg obserwacyjny dotyczy przekroju wodowskazowego w Krzczonowie na Krzczonówce, zaś najdłuższy (15 340 elementy) odnosi się do przekrojów w Mszanie Dolnej, Stróży i Proszówkach na Rabie.

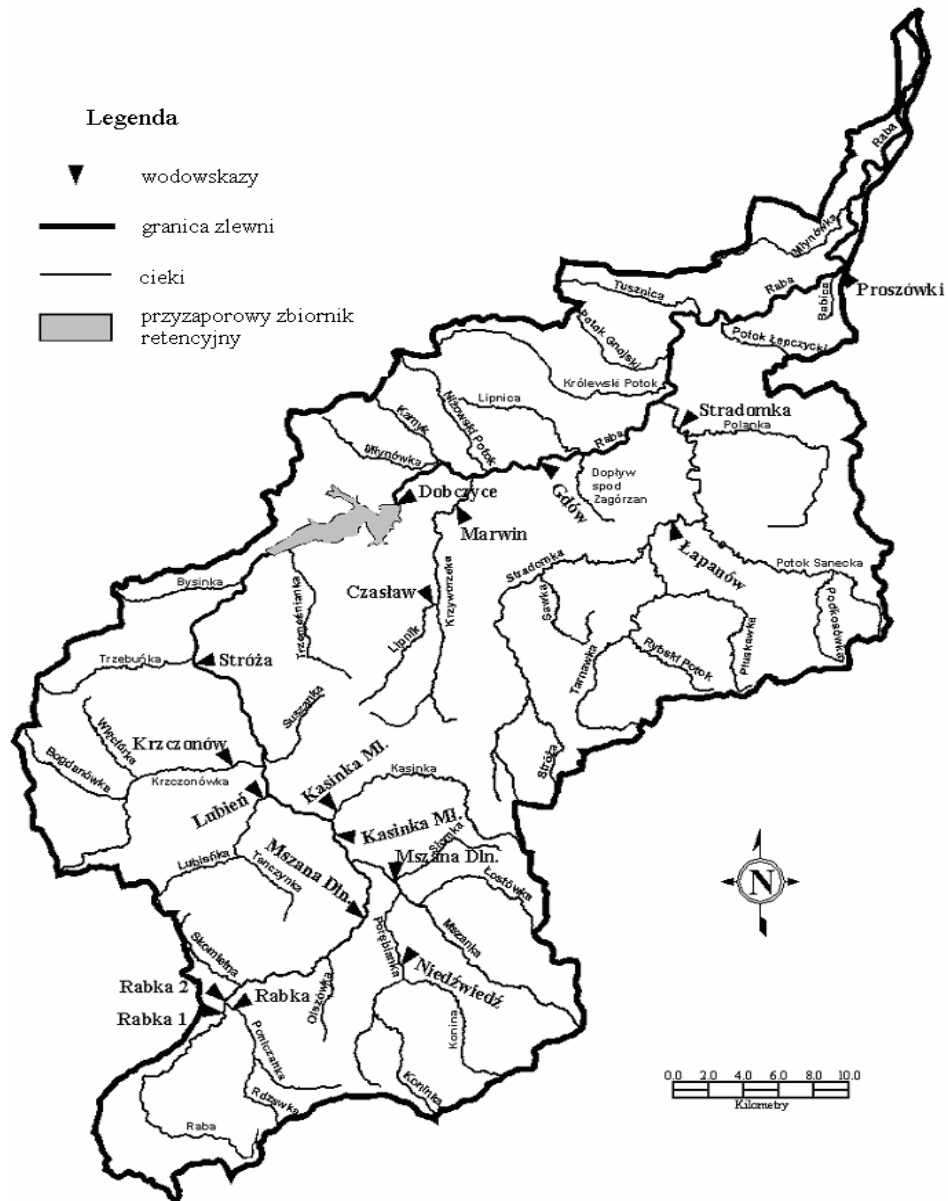
Materiałem empirycznym jest  $n$ -dniowy szereg czasowy natężenia przepływu  $Q_i$ ,  $i = 1, 2, \dots, n$  ( $n$  jest wielokrotnością dni w roku). Funkcja  $Q(t)$  dyskretyzowana jest przedziałem czasu  $\Delta t = 1$  dzień.

Aby uzyskać wstępny ogłąd danych, przebieg zmienności funkcji  $Q(t)$  zilustrowano serią wykresów (na rys. 2 zaprezentowano wybrane cztery wykresy, na rys. 3 – ich szczegóły). Następnie dla każdego szeregu czasowego  $Q_i$  z osobna zastosowano test SSS<sup>3</sup>. Test SSS polega na sprawdzeniu, czy w danych nieregularnych wykazujących (bądź nie) trend długoterminowy (lub cykle) istnieje dynamiczna struktura o krótkoterminowej zmienności. Stawia się hipotezę zerową: nieregularne fluktuacje krótkoterminowo zmienne są niezależnymi zmiennymi losowymi; oznacza to, że nie występują ani krótkoterminowa dynamika, ani też determinizm. Poprzez przetasowanie danych w niewielkiej skali czasowej uzyskuje się dane zastępcze, w których zostaje zmieniony przepływ informacji kreowany danymi. Przetasowanie takie niszczy bowiem potencjalną lokalną strukturę/korelację w nieregularnych fluktuacjach (o krótkoterminowej zmienności), a zachowuje struktury globalne (długoterminowe trendy, cykle). Istotne jest, że uzyskane w ten sposób dane zastępcze zachowują **ten sam rozkład prawdopodobieństwa** co dane źródłowe. Jako statystykę dyskryminacyjną<sup>4</sup> przyjęto uśrednioną informację wzajemną AMI<sup>5</sup>.

<sup>3</sup> Z ang. *small shuffle surrogate* – małe przetasowanie danych zastępczych.

<sup>4</sup> Jest stosowana do rozstrzygnięcia o przyjęciu lub odrzuceniu hipotezy zerowej.

<sup>5</sup> Z ang. *Average Mutual Information*.



Rys. 1. Zlewnia rzeki Raby z zaznaczonymi przekrojami wodowskazowymi – miejscami pomiaru badanych przepływów

Fig. 1. Raba river basin with marked gauging cross-sections – places of measurement of the investigated flows

Tabela 1

## Wzięte pod uwagę przekroje wodowskazowe i okresy obserwacji przepływów

Lp.	Rzeka	Przekrój wodowskazowy	Kilometr rzeki	Powierzchnia zlewni [km <sup>2</sup> ]	Okres obserwacji [lata]	Liczba lat obserwacji
1	2	3	4	5	6	7
1	Raba	Rabka 1	113,9	58	1961–1983	23
2	Raba	Rabka 2	113,1	92,8	1985–2002	18
3	Raba	Mszana Dolna	102,1	158,1	1961–2002	42
4	Raba	Kasinka Mała	95,8	353,1	1971–2002	32
5	Raba	Stróża	80,6	644,1	1961–2002	42
6	Raba	Dobczyce	61,9	768,4	1991–2002	12
7	Raba	Gdów	50,2	929,3	1961–1985	25
8					1991–1996	6
9	Raba	Proszówki	21,5	1470,4	1961–2002	42
10	Poniczanka	Rabka	0,5	33,1	1972–1983	12
11	Mszanka	Mszana Dolna	3,1	166,3	1961–1980	20
12					1985–2002	18
13	Lubieńka	Lubień	0,4	46,9	1971–1997	27
14	Krzczonówka	Krzczonów	2	87,9	1971–1986	16
15					1987–1995	9
16					1997–2002	6
17	Krzyworzeka	Marwin	2,5	77,7	1971–1986	16
18	Stradomka	Łapanów	14,8	213,1	1961–1994	34
19	Stradomka	Stradomka	2,6	361,8	1961–1985	25
20					1987–2002	16
21	Porębianka	Niedźwiedz	5,2	63,5	1982–1991	10

$$AMI(\tau) = \sum_{i=1}^n p(Q_i, Q_{i+\tau}) \log_2 \frac{p(Q_i, Q_{i+\tau})}{p(Q_i) \cdot p(Q_{i+\tau})}, \quad \tau = 1, 2, \dots, n-i+1 \quad (1)$$

gdzie  $p(Q_i, Q_{i+\tau})$  oznacza łączne prawdopodobieństwo zaobserwowania  $Q_i$  i  $Q_{i+\tau}$ , a  $p(Q_i)$  jest prawdopodobieństwem zaobserwowania  $Q_i$  w szeregu czasowym  $\{Q_1, Q_2, \dots, Q_n\}$ . Badając dynamikę krótkoterminową, obliczono AMI dla  $\tau = 1, 2, \dots, 20$ .

Funkcja ta dostarcza informacji o istnieniu i sile nieliniowych korelacji w danych. Jest uogólnieniem funkcji korelacji liniowej stosowanej w świecie liniowym – na świat nieliniowy [1]. Po obliczeniu wartości tej statystyki dla danych źródłowych i zastępczych następuje sprawdzenie, czy postawiona hipoteza zerowa może zostać odrzucona. Wykorzystując metodę Monte Carlo do testowania hipotez [7], należy zatem sprawdzić, czy obliczona wartość statystyki testowej dla danych źródłowych jest wewnątrz czy poza rozkładem wartości tej statystyki dla danych zastępczych. W przypadku gdy wyestymowana wartość statystyki dyskryminacyjnej pokrywa się z rozkładem tej statystyki dla danych zastępczych, hipoteza zerowa nie może zostać odrzucona. W przeciwnym przypadku (gdy wartość statystyki testowej jest poza rozkładem tej statystyki dla danych zastępczych) hipotezę zerową można odrzucić na przyjętym poziomie  $\alpha$  istotności testu.

Jak podaje [2], dla  $N_{DZ}$  zbiorów danych zastępczych  $\alpha = \frac{1}{N_{DZ} + 1}$  w przypadku testu jednostronnego, zaś  $\alpha = \frac{2}{N_{DZ} + 1}$  dla dwustronnego.

Dane zastępcze  $s(t)$  tworzone są z danych źródłowych  $x(t)$  na podstawie następującego algorytmu [5]. Niech  $i(t)$  oznacza indeks (numer) elementu w ciągu  $x(t)$ . Ponieważ  $i(t) = t$ , to  $x(i(t)) = x(t)$ .

Należy:

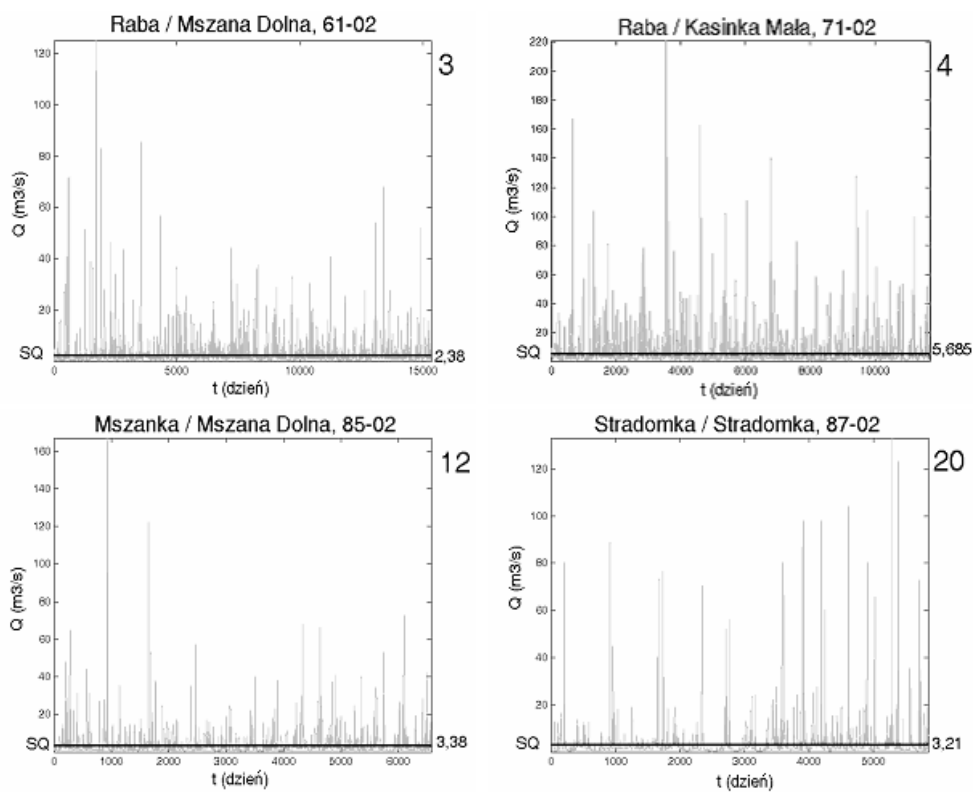
- utworzyć ciąg  $i'(t) = i(t) + A \cdot g(t)$ , gdzie  $A$  jest amplitudą (najczęściej<sup>6</sup>  $A = 1$ ), a  $g(t)$  ciągiem liczb losowych o rozkładzie Gaussa (choć  $i(t)$  jest ciągiem liczb całkowitych, to  $i'(t)$  nie jest),
- posortować ciąg  $i'(t)$  rosnąco, tworząc ciąg  $i''(t)$ ,
- utworzyć ciąg  $\hat{i}(t)$  indeksów, jakie miały kolejne elementy ciągu  $i''(t)$  w ciągu  $i'(t)$ ,
- utworzyć dane zastępcze poprzez przestawienie elementów ciągu  $x(t)$  zgodnie z indeksami wskazanymi przez elementy ciągu  $\hat{i}(t)$ , tj.  $s(t) = x(\hat{i}(t))$ .

### 3. Wyniki i dyskusja

Wstępna obserwacja zmian czasowych funkcji  $Q(t)$  (na rys. 2 zaprezentowano wybrane cztery wykresy, na rys. 3 – ich szczegóły) pozwala dostrzec ważną prawidłowość. Chodzi o rozkład wartości natężenia przepływu w danym ciągu czasowym: wartości wysokich jest niewiele, więcej pojawiło się wartości średnich, bardzo liczne są wartości małe. Trudno dostrzegalna jest regularność zmian wartości natężenia  $Q$ , choć zauważyć można pewne powtarzalne fluktuacje (nagłe wzrosty i spadki) wartości natężenia przepływu w odniesieniu do określonego niskiego ich poziomu. Ogólnie, taki stan rzeczy nie pozwala jeszcze na kategoryczne stwierdzenie, czy badany system (zlewnia) generuje wartości (przepływy) o charakterze losowym czy deterministycznym; nie uwidacznia się też żadna wyraźna okresowość tych wartości.

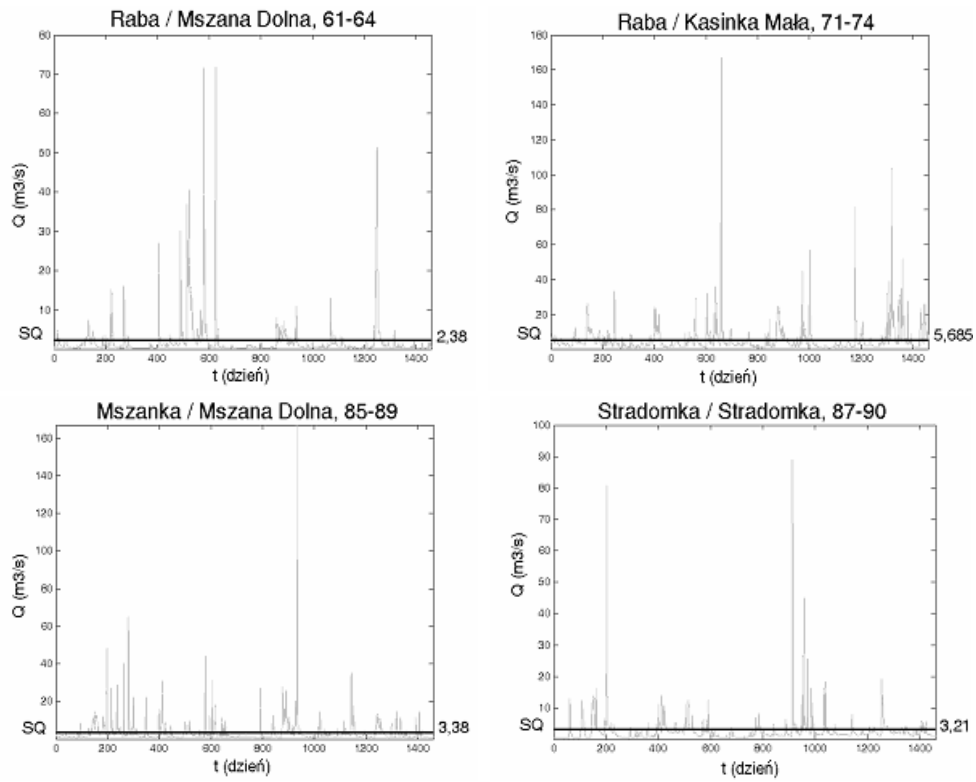
Przeprowadzenie testu SSS objęło wygenerowanie  $N_{DZ} = 40$  zestawów danych zastępczych dla każdego z 21 badanych ciągów  $Q_i$  wartości natężenia przepływu. Na rysunkach 4 i 5 zaprezentowano po pięć takich zestawów dla wybranych dwóch posterunków wodowskazowych: Mszana Dolna na Rabie i Mszana Dolna na Mszance. Porównując przebieg funkcji  $Q(t)$  dla danych źródłowych z jej przebiegiem dla danych zastępczych, widać duże podobieństwo kształtów i jedynie nieznaczne, mało zauważalne różnice. Stanowi to potwierdzenie poprawności działania zastosowanego algorytmu w stosunku do przyjętej strategii.

<sup>6</sup> Autorzy testu [5] proponują tę wartość jako poprawną w większości przypadków, przy czym wskazują, że jej dobór zależy od właściwości konkretnego zbioru danych źródłowych; stąd większe ( $A \in (1; 5)$ ) i mniejsze ( $A \in (0; 1)$ ) wartości amplitudy  $A$  mogą być uzasadnione.



Rys. 2. Natężenie przepływu  $Q$  (linia '-') w określonym przekroju ciekę rozważanej zlewni, w zależności od czasu  $t$ .  $SQ$  (linia '-') oznacza przepływ średni z wielolecia. Objasnienia: w nagłówku wykresu podano kolejno: nazwę rzeki, nazwę przekroju wodowskazowego, lata obserwacji. Numeracja wykresów podana w prawym górnym rogu zgodna z tab. 1

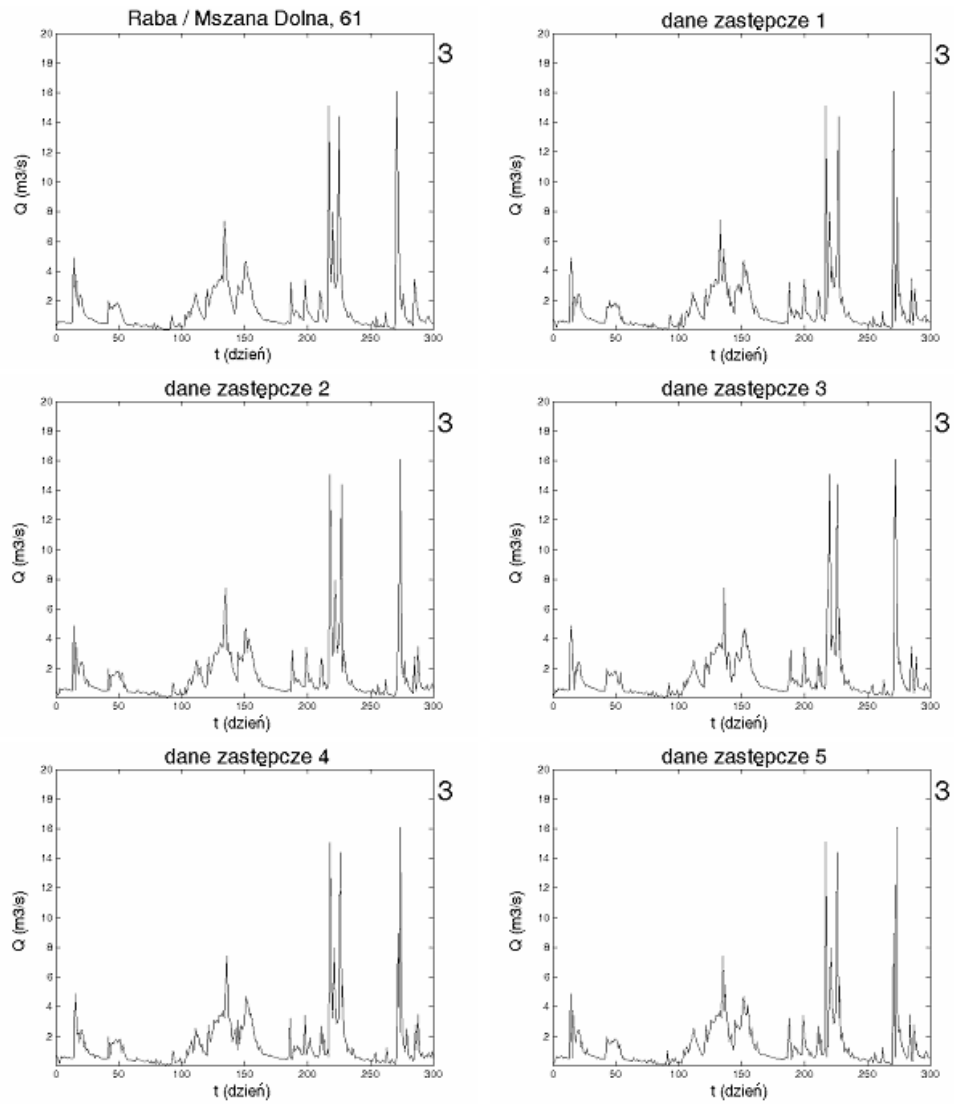
Fig. 2. The discharge  $Q$  ('-' line) in given cross-sections of the investigated basin versus time  $t$ .  $SQ$  ('-' line) means average flow of several years. Explanation: in the headline of the plot the following information is provided. That is: river name, cross-section name, observation time. The numbering of plots placed in the top right-hand corner is consistent with the Table 1



Rys. 3. Natężenie przepływu  $Q$  w zależności od czasu  $t$ : uszczegółowienie wykresów nr 3, 4, 12, 20

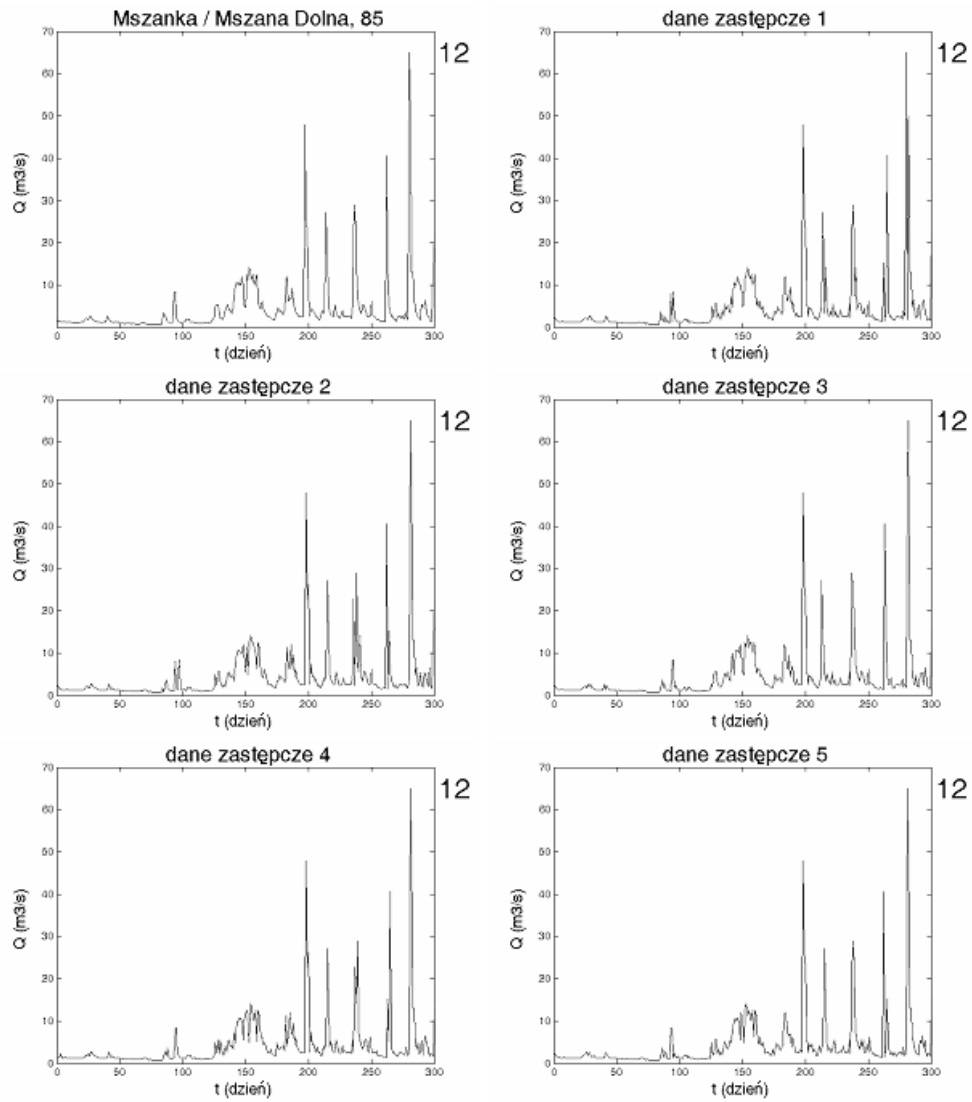
Fig. 3. The discharge  $Q$  versus time  $t$ : details of plots No. 3, 4, 12, 20





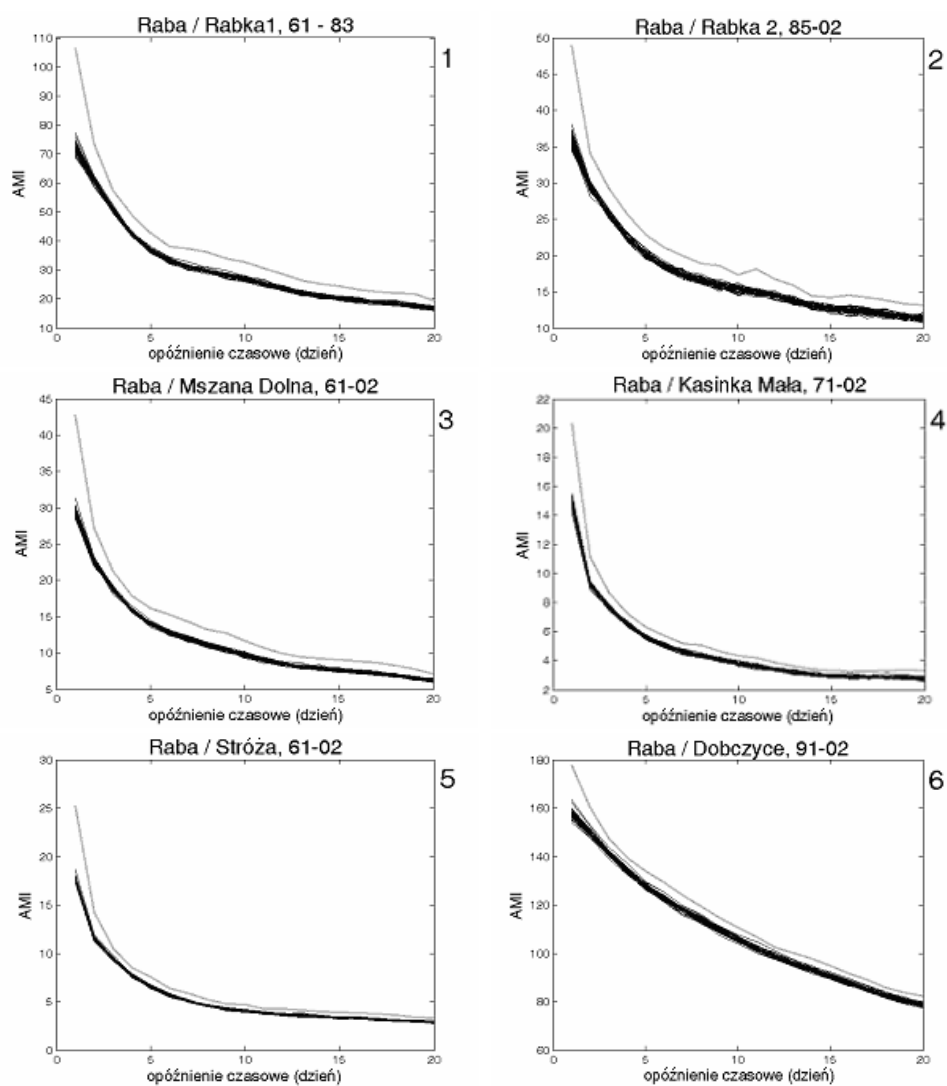
Rys. 4. Dla porównania: fragmenty przykładowych pięciu zestawów danych zastępczych dla posterunku Mszana Dolna na Rabcie. Objasnienia jak dla rys. 2

Fig. 4. For comparison: fragments of five exemplary surrogate data for Mszana Dolna cross-section on Raba river. Explanation is the same as on Fig. 2



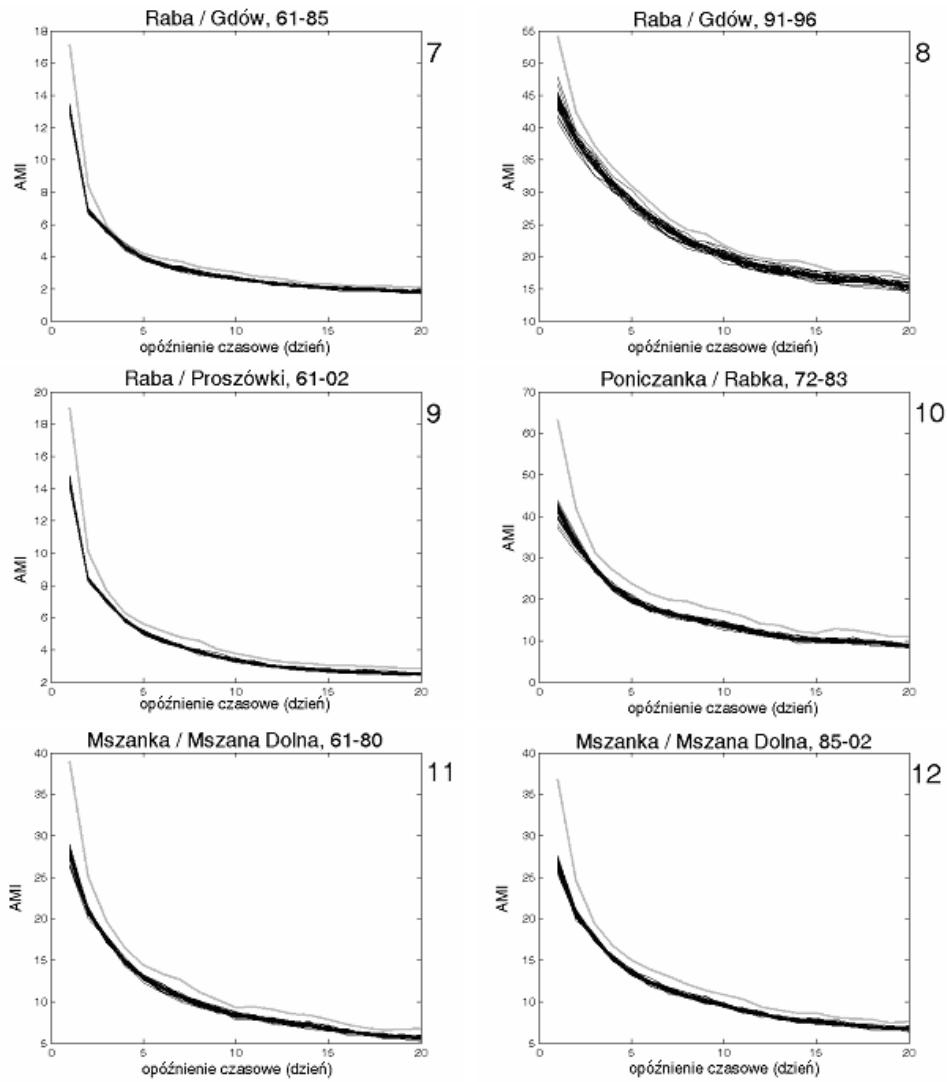
Rys. 5. Dla porównania: fragmenty przykładowych pięciu zestawów danych zastępczych dla posterunku Mszana Dolna na Mszance. Objasnienia jak dla rys. 2

Fig. 5. For comparison: fragments of five exemplary surrogate data for Mszana Dolna cross-section on Mszanka river. Explanation is the same as on Fig. 2



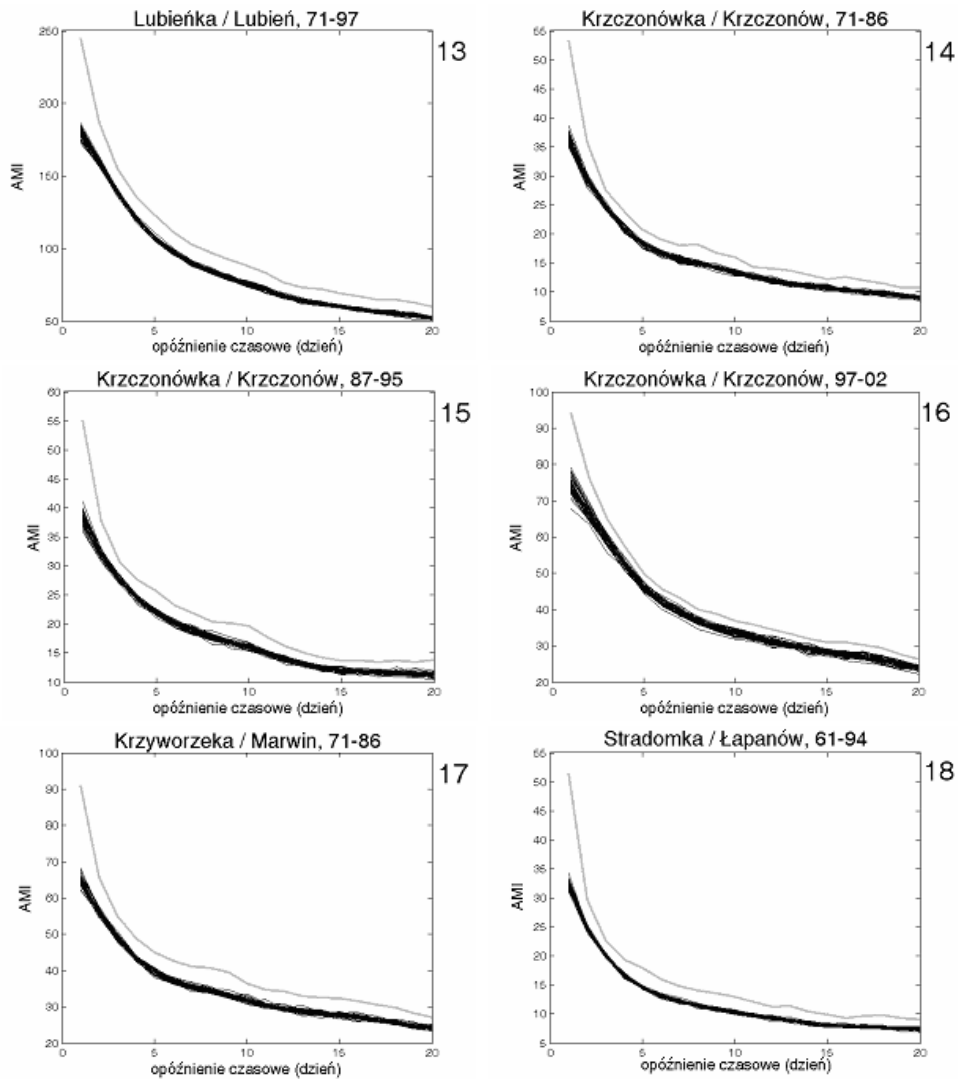
Rys. 6. Wyniki testu SSS dla poszczególnych posterunków wodowskazowych zlewni rzeki Raby. Pionowa oś wykresów podaje wartości statystyk testowych dla danych źródłowych (linia '-') i danych zastępczych (linie '-'). Objasnienia jak dla rys. 2

Fig. 6. The results of the test SSS for given cross-sections of Raba river basin. On vertical axes of the plots estimated test statistics are presented. The '-' line is the original data and '-' lines are the SSS data. Explanation is the same as on Fig. 2



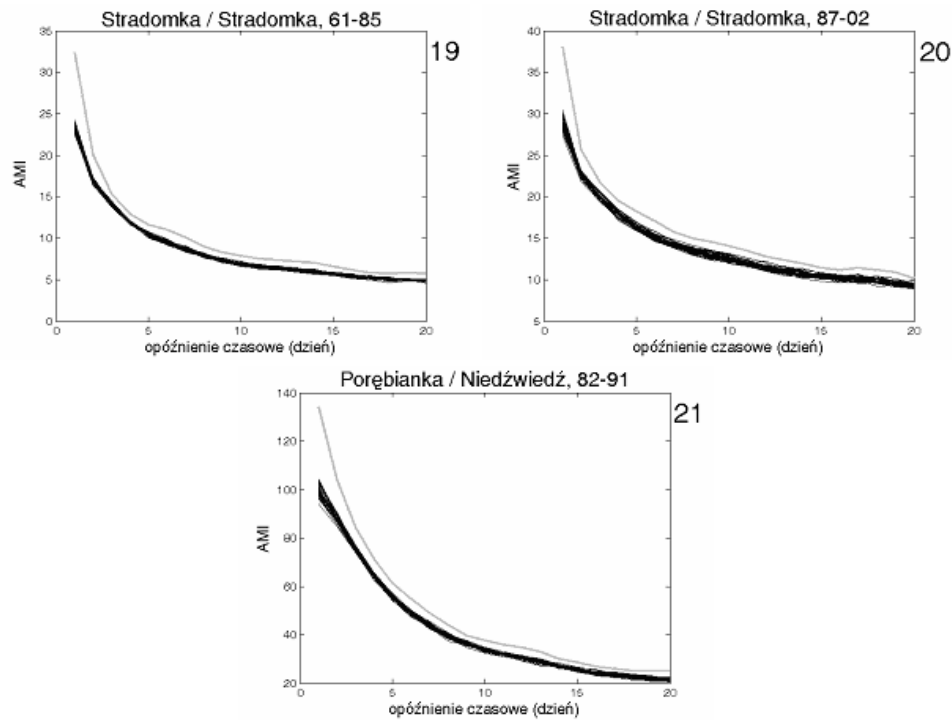
Rys. 6. Cd.

Fig. 6. Continued



Rys. 6. Cd.

Fig. 6. Continued



Rys. 6. Cd.

Fig. 6. Continued

Poprzez zastosowanie testu jednostronnego uzyskano poziom istotności  $\alpha = 0,0244$  (czyli poniżej 2,5%). Obliczone wartości statystyki dyskryminacyjnej (1) dla danych źródłowych i zastępczych przedstawiono na serii wykresów (rys. 6). Można na nich zaobserwować, że wartości funkcji uśrednionej informacji wzajemnej dla danych źródłowych leżą poza rozkładem estymat tej statystyki dla danych zastępczych. Oznacza to możliwość odrzucenia postawionej hipotezy zerowej na przyjętym poziomie istotności testu. Interesujące jest także spostrzeżenie dotyczące każdego z wykresów, że przebieg funkcji (1) dla danych źródłowych jest podobny do przebiegu tej funkcji dla danych zastępczych. Stanowi to potwierdzenie, że w danych SSS struktury lokalne zostały zniszczone, ale zachowane te globalne [5].

#### 4. Podsumowanie

Analizę natury (deterministycznej/losowej) przepływów dobowych z lat 1961–2002 w zlewni rzeki Raby przeprowadzono z zastosowaniem testu statystycznego SSS. Aby uzyskać poziom istotności testu  $\alpha \leq 2,5\%$  dla każdego badanego szeregu czasowego, wygenerowano 40 zestawów danych zastępczych. Przeprowadzone testy hipotezy postulującej, że nieregularne fluktuacje w danych źródłowych krótkoterminowo zmienne są niezależnymi zmiennymi losowymi, odrzuciły ją dla wszystkich 21 ciągów natężenia przepływu. Choć wstępny ogląd przebiegu funkcji  $Q(t)$  nie dawał wskazówek co do natury badanego przepływu, otrzymane wyniki obliczeń pozwalają z dużym prawdopodobieństwem (wyższym niż 97,5%) przyjąć istnienie krótkoterminowej dynamiki (determinizmu) w przepływach dobowych zlewni rzeki Raby.

#### Literatura

- [1] Abarbanel H.D.I., *Analysis of observed chaotic data*, Springer-Verlag, New York 1996.
- [2] Luo X.D., Nakamura T., Small M., *Surrogate test to distinguish between chaotic and pseudoperiodic time series*, Physical Review E, Vol. 71, 026230, 2005.
- [3] Kędra M., *Cykliczność przepływu w zlewni rzeki Raby, Część II: Przepływy dobowe*, Czasopismo Techniczne, z. 2-Ś/2006, Wyd. PK, Kraków 2006, 203-216.
- [4] Nachlik E. (red.), *Identyfikacja i ocena antropogenicznych oddziaływań na zasoby wodne zlewni Raby wraz z oszacowaniem ryzyka nieosiągnięcia celów środowiskowych*, Monografia Politechniki Krakowskiej, Kraków 2007.
- [5] Nakamura T., Small M., *Small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends*, Physical Review E, Vol. 72, 056216, 2005.
- [6] Theiler J., Eubank S., Longtin A., Galdrikian B., Farmer J.D., *Testing for nonlinearity in the time series: The method of surrogate data*, Physica D, Vol. 58, 1992, 77-94.
- [7] Theiler J., Prichard D., *Constrained-realization Monte Carlo method for hypothesis testing*, Physica D, Vol. 94, 1996, 221-239.