

# **Analiza i modelowanie przepływów w sieci Internet**

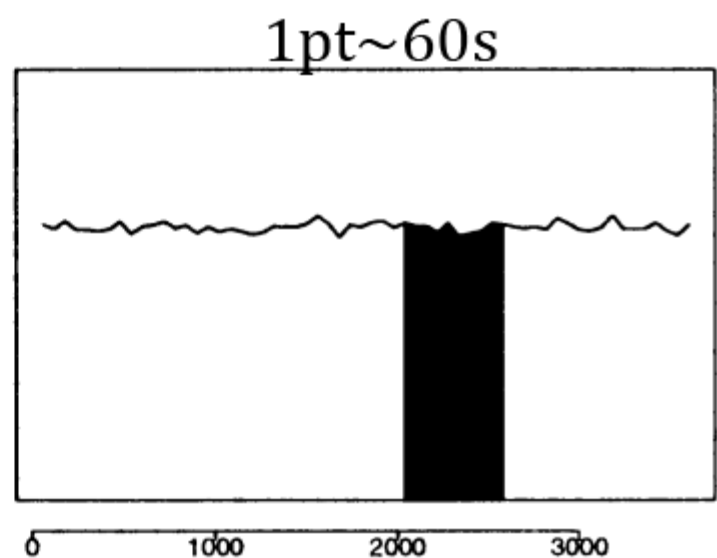
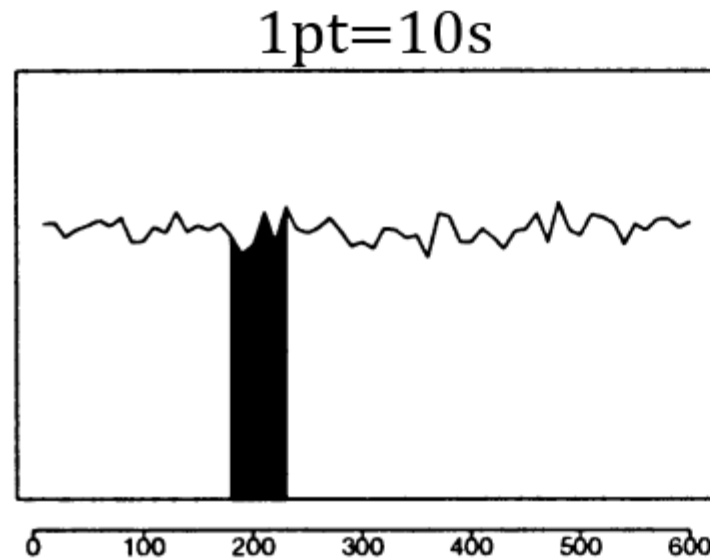
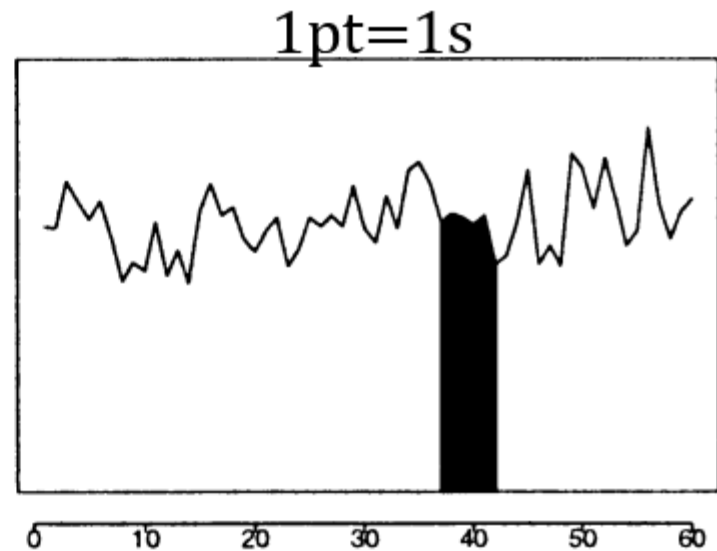
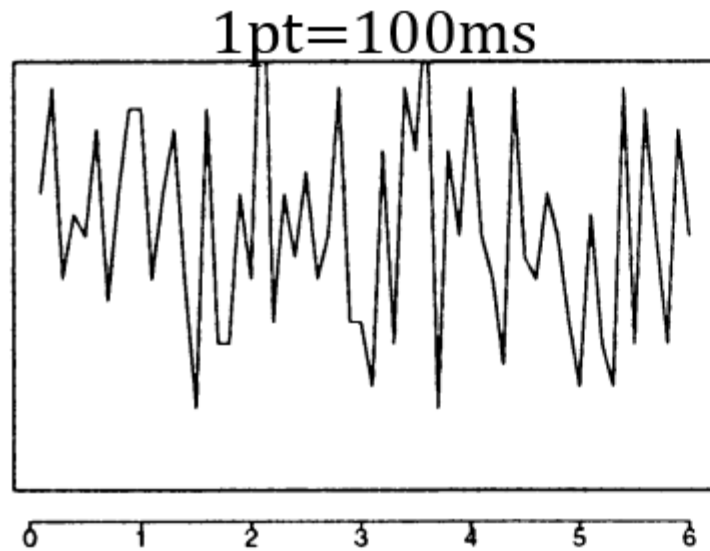
Andrzej Andrijew

# Plan referatu

- Samopodobieństwo w sieci Internet
- Samopodobne procesy stochastyczne
- Metody sprawdzania samopodobieństwa
- Modelowanie przepływów w sieciach komputerowych

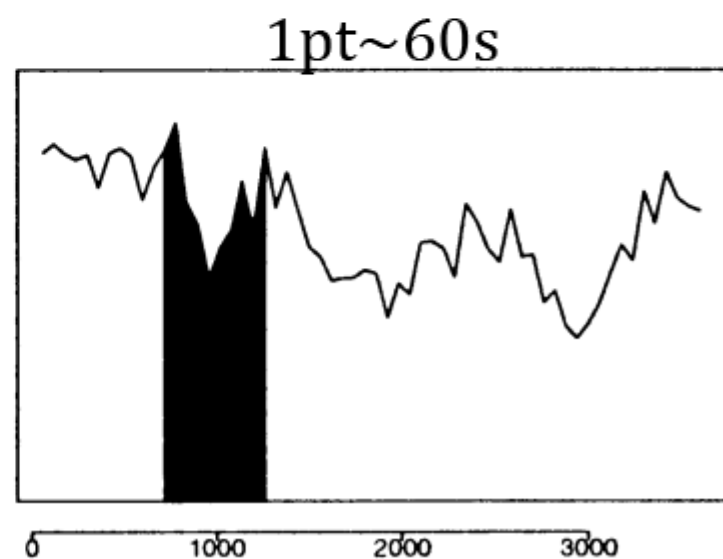
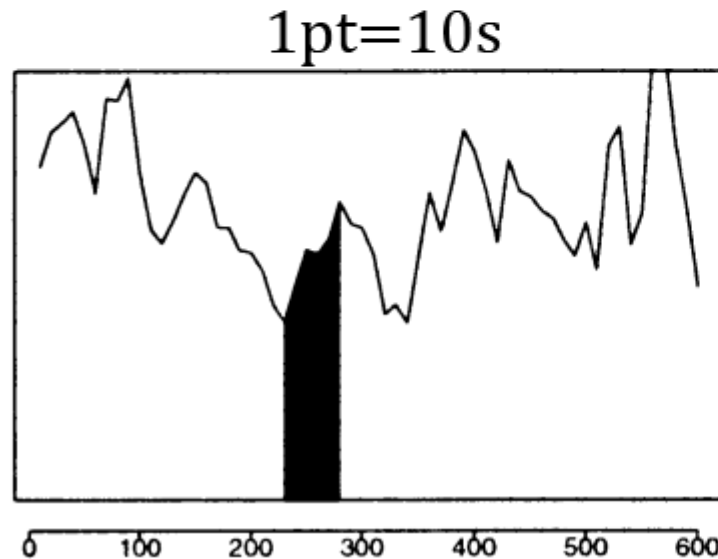
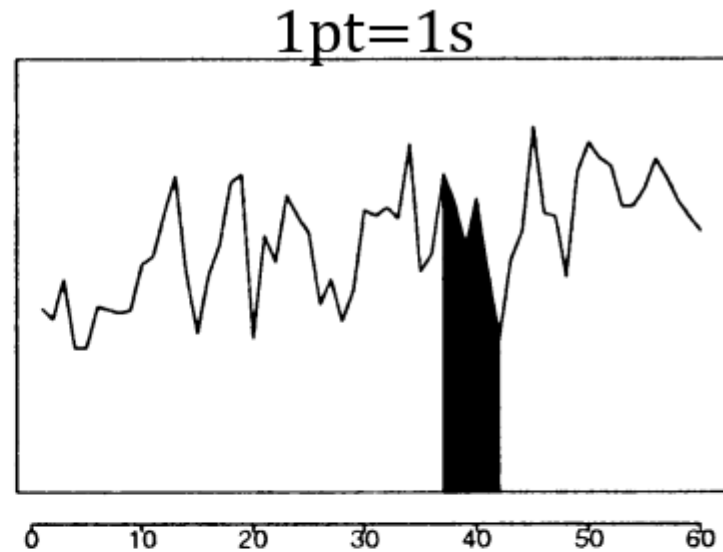
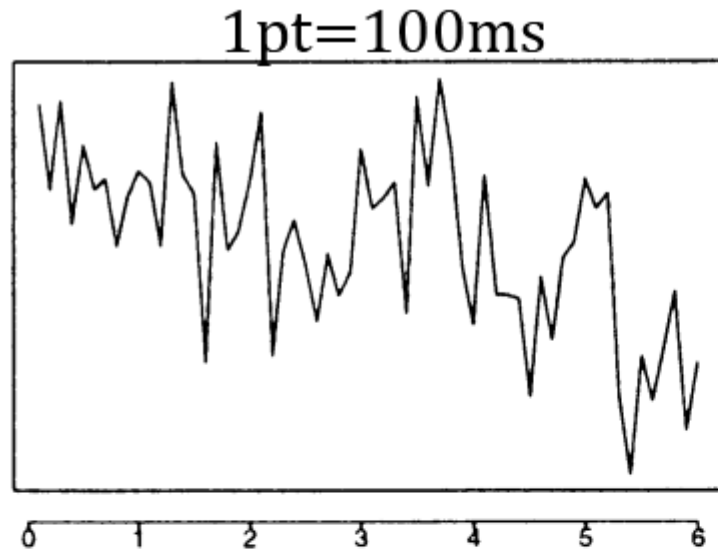
# Samopodobieństwo w sieci Internet

Rys [1]



# Samopodobieństwo w sieci Internet

Rys [1]



# Samopodobne procesy stochastyczne

Rodzina zmiennych losowych  $\{X(t), t \in \mathbb{R}\}$  jest samopodobnym procesem stochastycznym z parametrem samopodobieństwa  $H > 0$

nazywanym wykładnikiem Hursta, jeśli zachodzi

$$\{X(at), t \in \mathbb{R}\} \stackrel{d}{=} \{a^H X(t), t \in \mathbb{R}\}$$

dla każdego  $a > 0$ .

# Proces łącznych uśrednionych przyrostów

Dla danego procesu stacjonarnego  $\{X_k, k \in \mathbb{Z}\}$  w szerszym sensie (wartość oczekiwana jest stała, autokowariancja zależy tylko i wyłącznie od różnicy w czasie) można zdefiniować proces  $\{X_k^{(m)}, k \in \mathbb{Z}\}$  łącznych uśrednionych przyrostów jako

$$X_k^{(m)} = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i$$

# Samopodobieństwo drugiego rzędu

Proces  $\{X_k, k \in \mathbb{Z}\}$  jest asymptotycznie samopodobny w szerszym sensie z parametrem samopodobieństwa  $0.5 < H < 1$  jeśli

$$\lim_{m \rightarrow \infty} C^{(m)}(k) = \frac{\sigma^2}{2} [(k+1)^{2H} - 2(k)^{2H} + (k-1)^{2H}]$$

dla każdego  $k \geq 1$  oraz pewnego  $\sigma > 0$ , gdzie  $C^{(m)}$  jest autokowariancją procesu łącznych przyrostów określonego na rozważanym procesie.

# Samopodobieństwo drugiego rzędu i procesy H-sssi

- Procesy H-sssi

Proces ma stacjonarne przyrosty, jeśli

$$\begin{aligned} \{X(t + \Delta t) - X(\Delta t), t \in \mathbb{R}\} \\ = \{X(t) - X(0), t \in \mathbb{R}\} \end{aligned}$$

dla każdego  $\Delta t \in \mathbb{R}$ .

- Ułamkowe ruchy Browna *fBm* ( $0 < H < 1$ )
- Ułamkowy szum Gaussa *fGn*



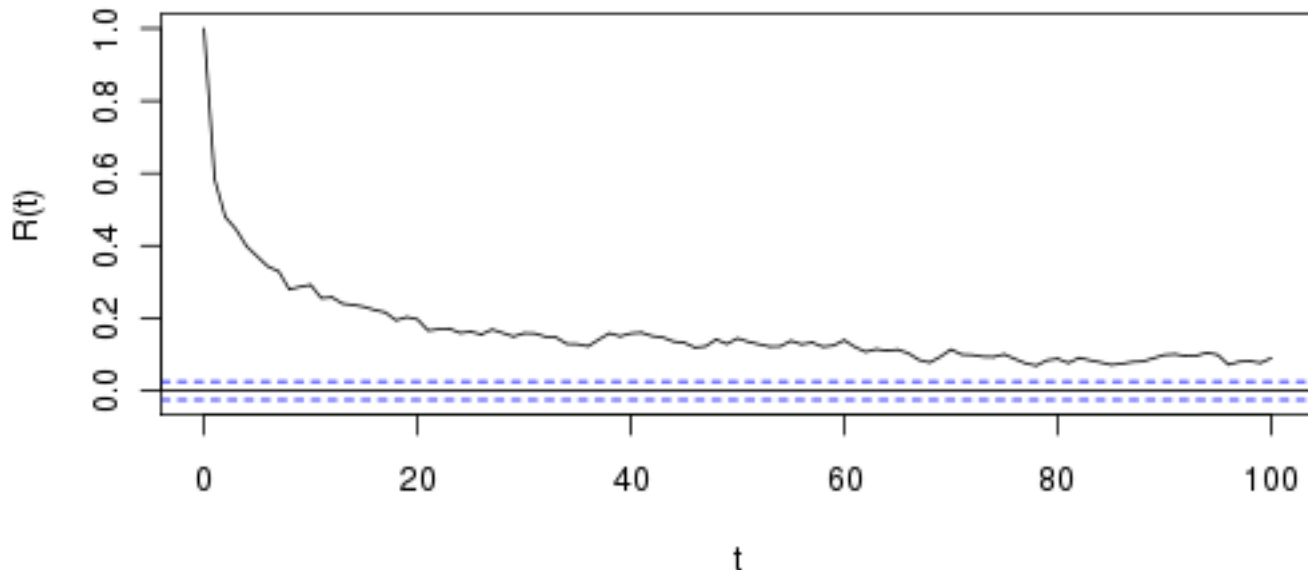
# Własności związane z samopodobieństwem

- Długoterminowe zależności (tzw. procesy *LRD*)
- Wolno opadająca wariancja
- Dystrybuanta o własności ciężkich ogonów

# Długoterminowe zależności

autokorelacje są dodatnie nawet dla dużych przesunięć, w szczególności istnieją takie  $c_r > 0$  oraz  $\alpha = 2 - 2H$ , że zachodzi

$$\lim_{k \rightarrow \infty} \frac{R(k)}{c_r k^{-\alpha}} = 1$$



# Długoterminowe zależności cd..

Funkcja gęstości widma mocy dla procesów dyskretnych

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{ik\omega},$$

$$\omega \in [-\pi, \pi], i = \sqrt{-1}$$

dla procesów LRD zachodzi przy  $\omega \rightarrow 0$

$$S(\omega) \sim \frac{1}{|\omega|^{2H-1}}$$

# Wolno opadająca wariancja

Dla procesu  $\{X_k, k \in \mathbb{Z}\}$  oraz odpowiedniego  $\{X_k^{(m)}, k \in \mathbb{Z}\}$  zachodzi

$$\sigma_{X^{(m)}}^2 \sim am^{2H-2}$$

przy odpowiednio dużych  $m \rightarrow \infty$ ,  
co pozwala rozróżnić procesy samopodobne od  
procesów z krótkoterminowymi zależnościami  
gdyż dla procesów *SRD* zachodzi

$$\sigma_{X^{(m)}}^2 \sim m^{-1}$$

# Dystrybuanta o własności ciężkich ogonów

W przepływach w sieciach komputerowych charakterystyki związane z długością czasu trwania połączeń, czasów pomiędzy przybyciami pakietów, wielkości przesyłanych plików są związane z dystrybuantami o własności ciężkich ogonów.

Wykazano, że właśnie one odpowiadają między innymi za długoterminowe zależności przepływów w sieciach komputerowych.

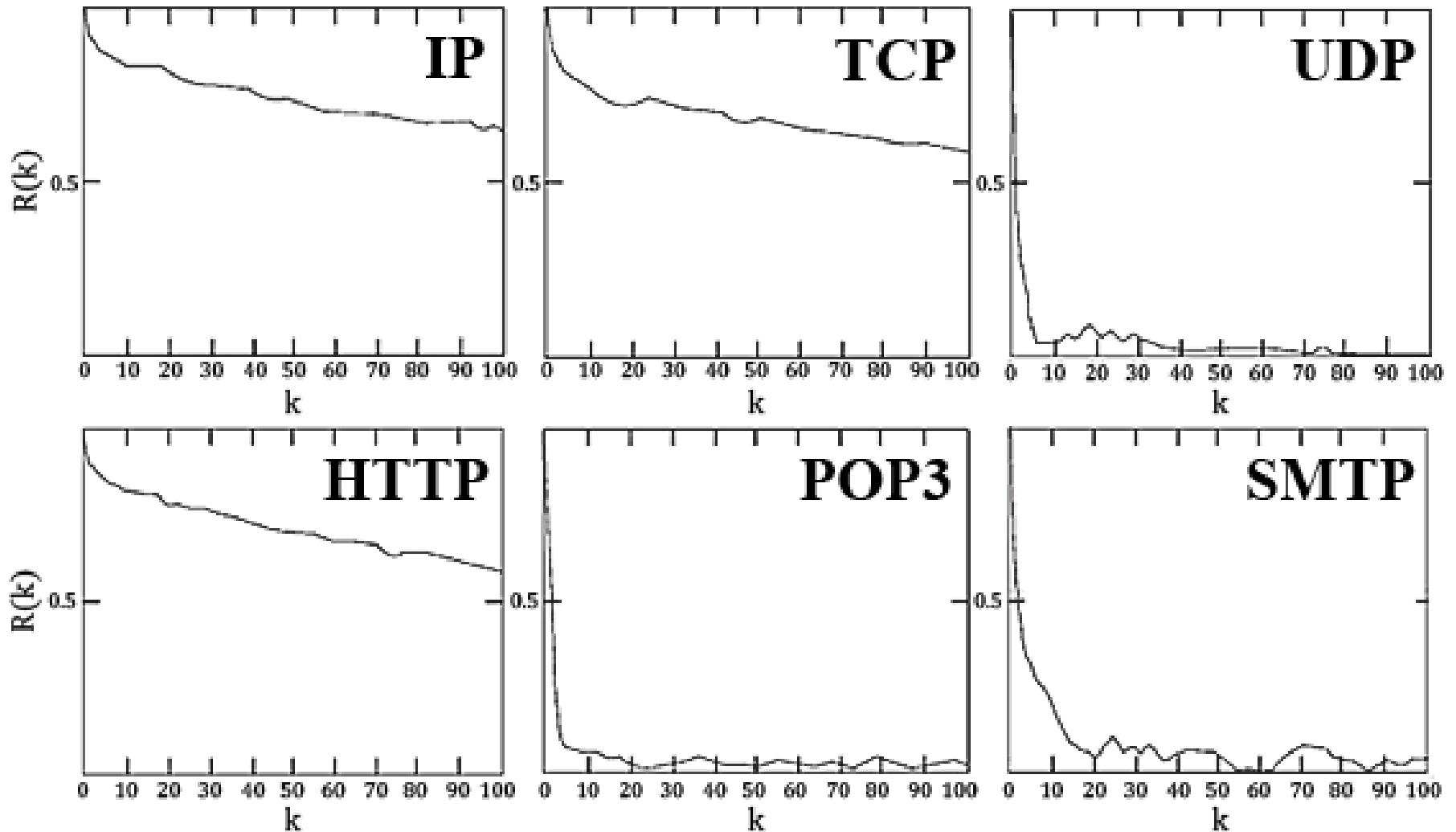
# Samopodobieństwo przepływów cd ...

- Autokorelacja agregowanego strumienia

$$R_A(k) = \frac{1}{\sum_{i=1}^N \sigma_{A_i}^2} \sum_{i=1}^N \sigma_{A_i}^2 R_{A_i}(k)$$

# Samopodobieństwo przepływów cd ...

Rys [2]

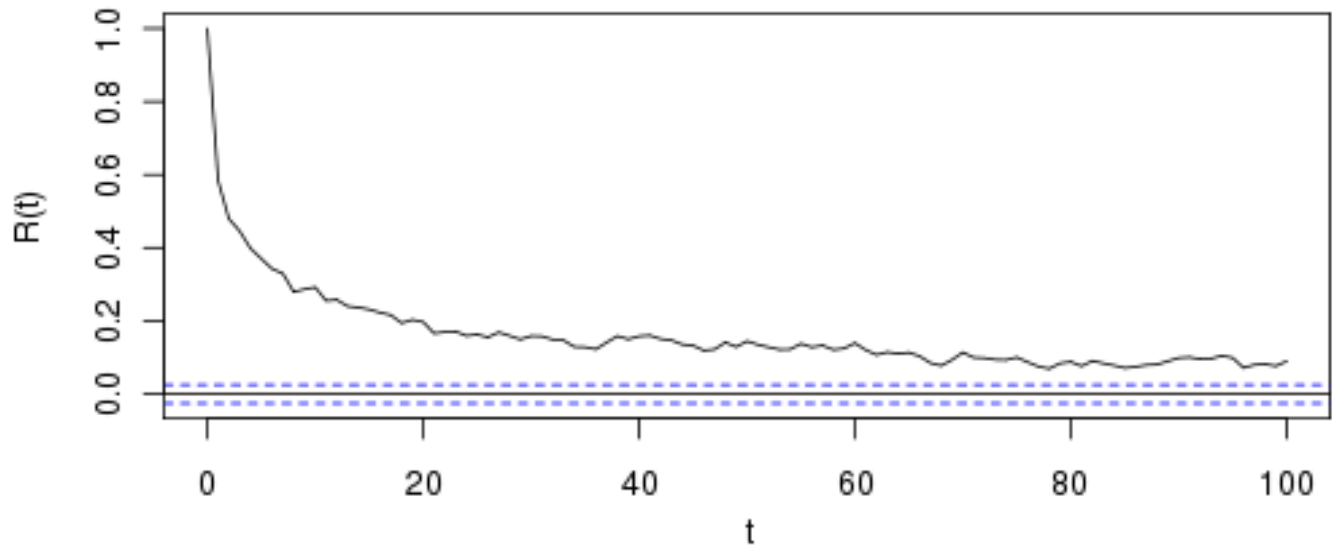
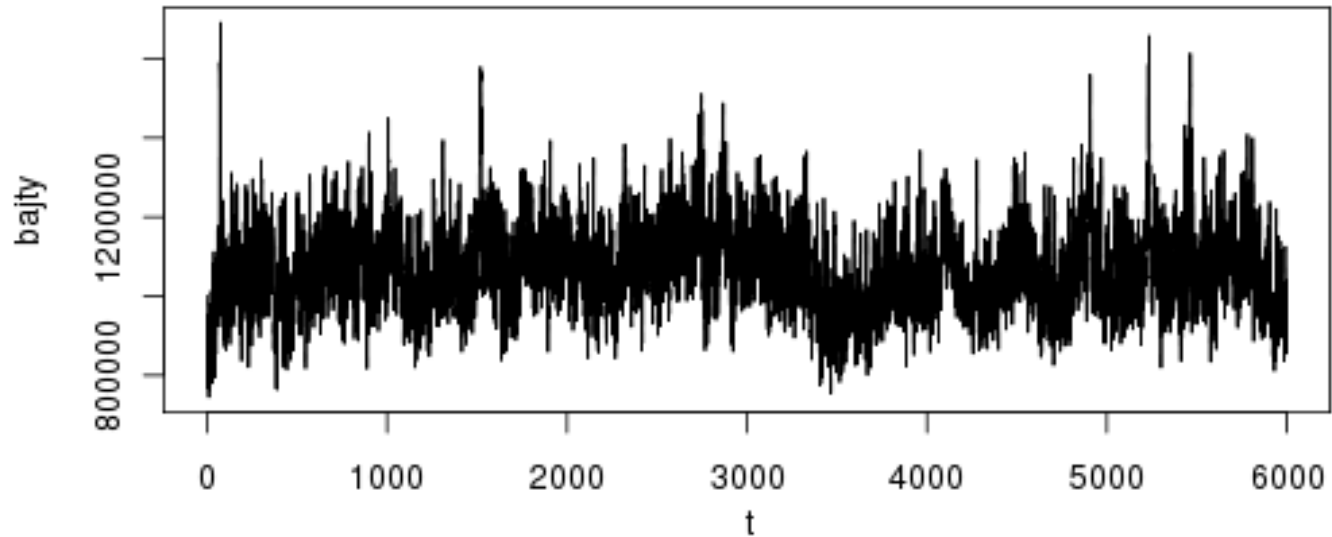


# Badanie rzeczywistych przepływów

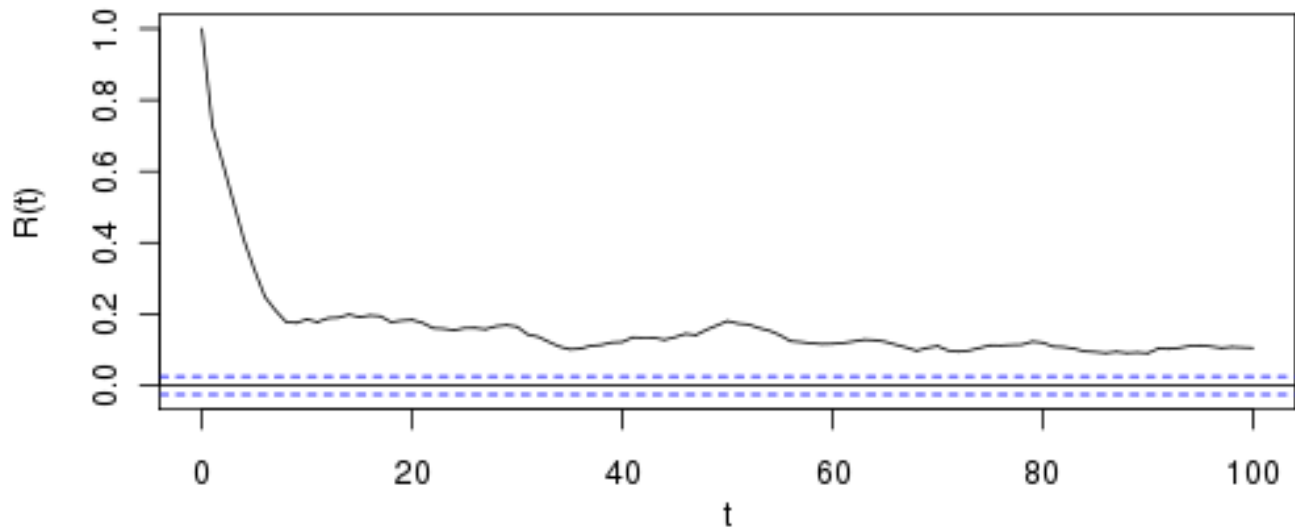
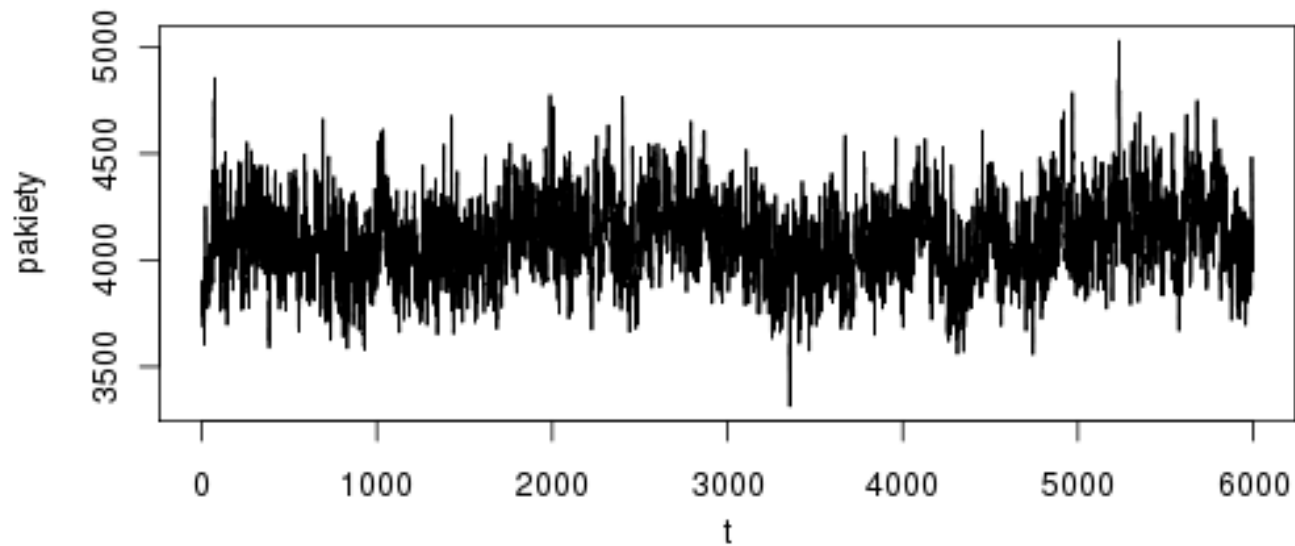
- Próbka z 20 kwietnia 2003 z jednego z serwerów na Uniwersytecie Carolina w Stanach Zjednoczonych
- Badanie ograniczone do zakresu pierwszych 10 minut
- Najmniejszy przedział czasowy to 100ms, próbka zawierała 24,5 mln pakietów



# Badane przepływy



# Badane przepływy



# Badane przepływy

| metoda      | H - bajty | H - pakiety |
|-------------|-----------|-------------|
| wariancji   | 0,7999328 | 0,793038    |
| R/S         | 0,7987437 | 0,799822    |
| periodogram | 0,9035105 | 0,7288805   |
| falkowa     | 0,8297308 | 0,8300158   |

# Źródła samopodobieństwa

- Zachowanie użytkowników sieci (np. preferencje wyboru dokumentów w sieci)
- Struktura przesyłanych danych (np. rozkład wielkości przesyłanych plików)
- Ograniczona przepustowość łącz
- Mechanizmy bezpośrednio powiązane z protokołami transmisji

# Metody sprawdzania samopodobieństwa

- metody operujące w dziedzinie czasu
  - wykres wariancji
  - statystyka R/S
- metody operujące w dziedzinie częstotliwości
  - periodogram
  - analiza falkowa

# Metoda wykresu wariacji

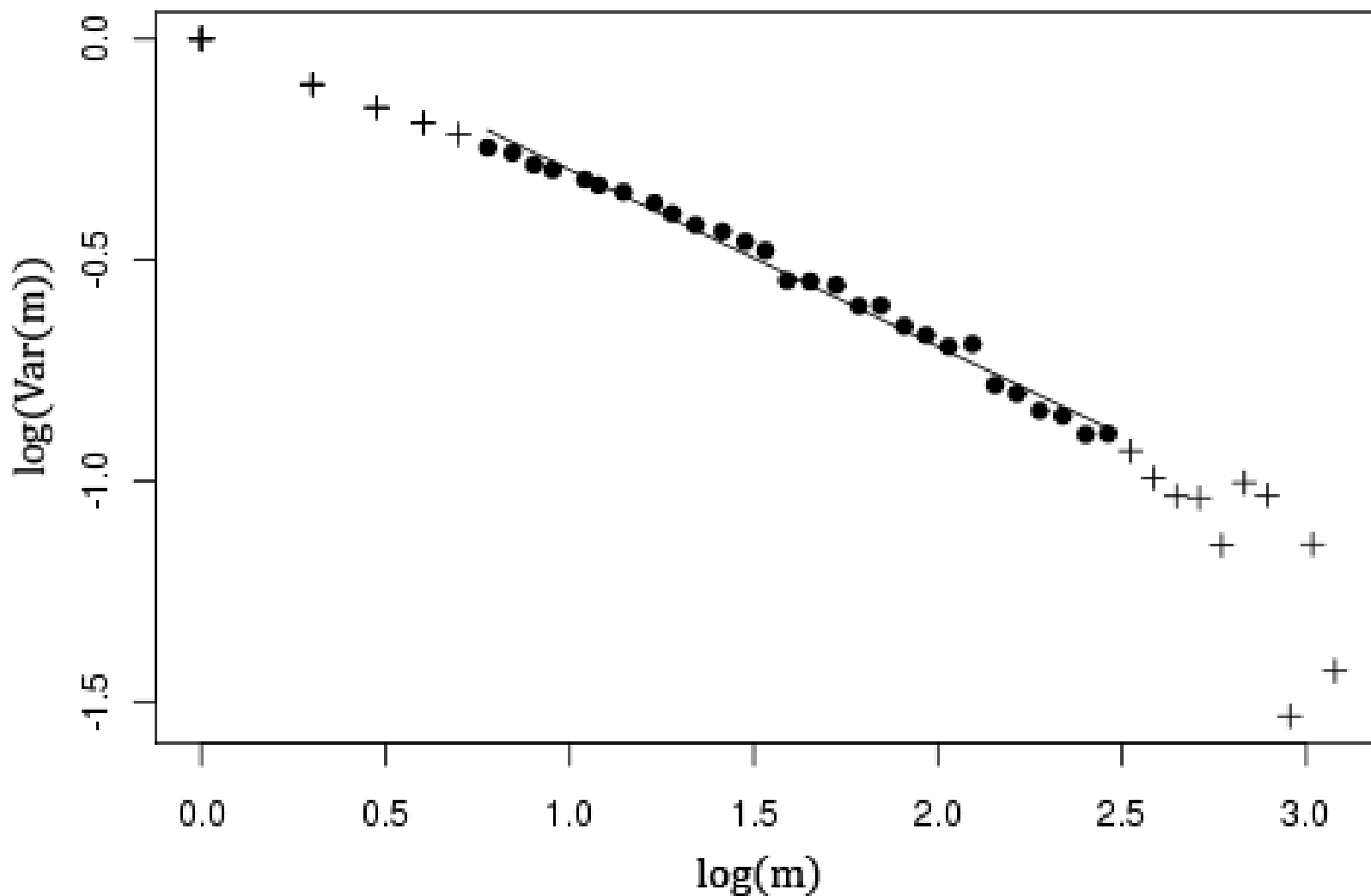
dla procesu samopodobnego zachodzi

$$\sigma_{X(m)}^2 \sim am^{2H-2}$$

skąd dalej

$$\log(\sigma_{X(m)}^2) \sim (2H - 2) \log(m) + \log(a)$$

# Metoda wykresu wariancji cd...



# Metoda statystyki R/S

dla zbioru  $n$  obserwacji  $\{X_k, k \in \mathbb{Z}_+\}$  można zdefiniować

$$\frac{R(n)}{S(n)} = \frac{\max_{1 \leq i \leq n} \Delta_i - \min_{1 \leq i \leq n} \Delta_i}{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

gdzie

$$\Delta_k = \sum_{i=1}^k X_i - k\bar{X}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$



# Metoda statystyki R/S cd...

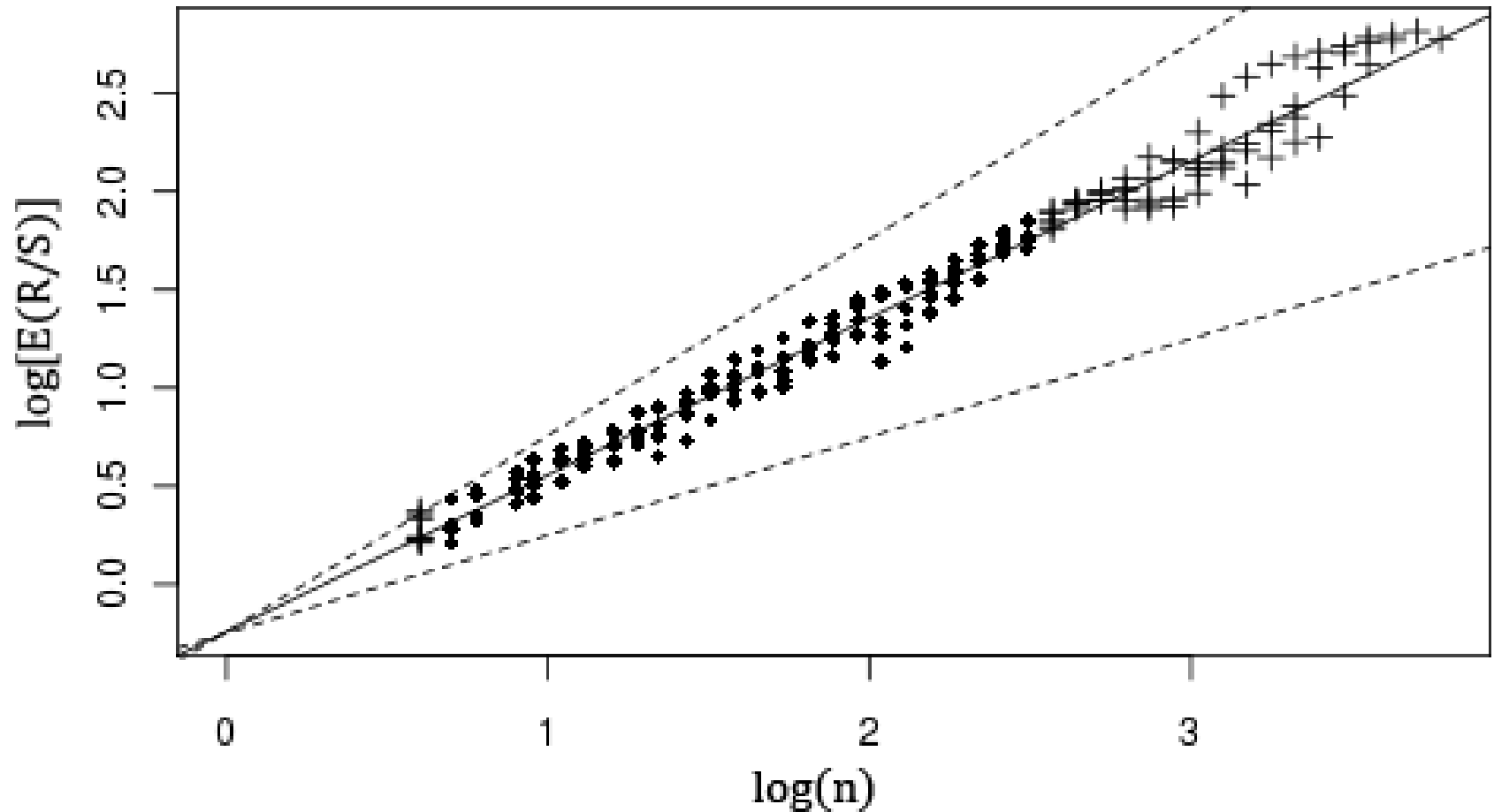
zachodzi

$$E \left[ \frac{R(n)}{S(n)} \right] \sim cn^H$$

skąd dalej

$$\log \left( E \left[ \frac{R(n)}{S(n)} \right] \right) \sim H \log(n) + \log(c)$$

# Metoda statystyki R/S cd...



# Metoda periodogramu

Periodogram ciągu obserwacji  $\{X_k, k = 1, 2 \dots N\}$

$$I_N(\omega) = \frac{1}{2\pi N} \left| \sum_{k=1}^N X_k e^{jk\omega} \right|^2$$

przybliża funkcję gęstości widma mocy

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R(k) e^{ik\omega},$$

$$\omega \in [-\pi, \pi], i = \sqrt{-1}$$

# Metoda periodogramu cd...

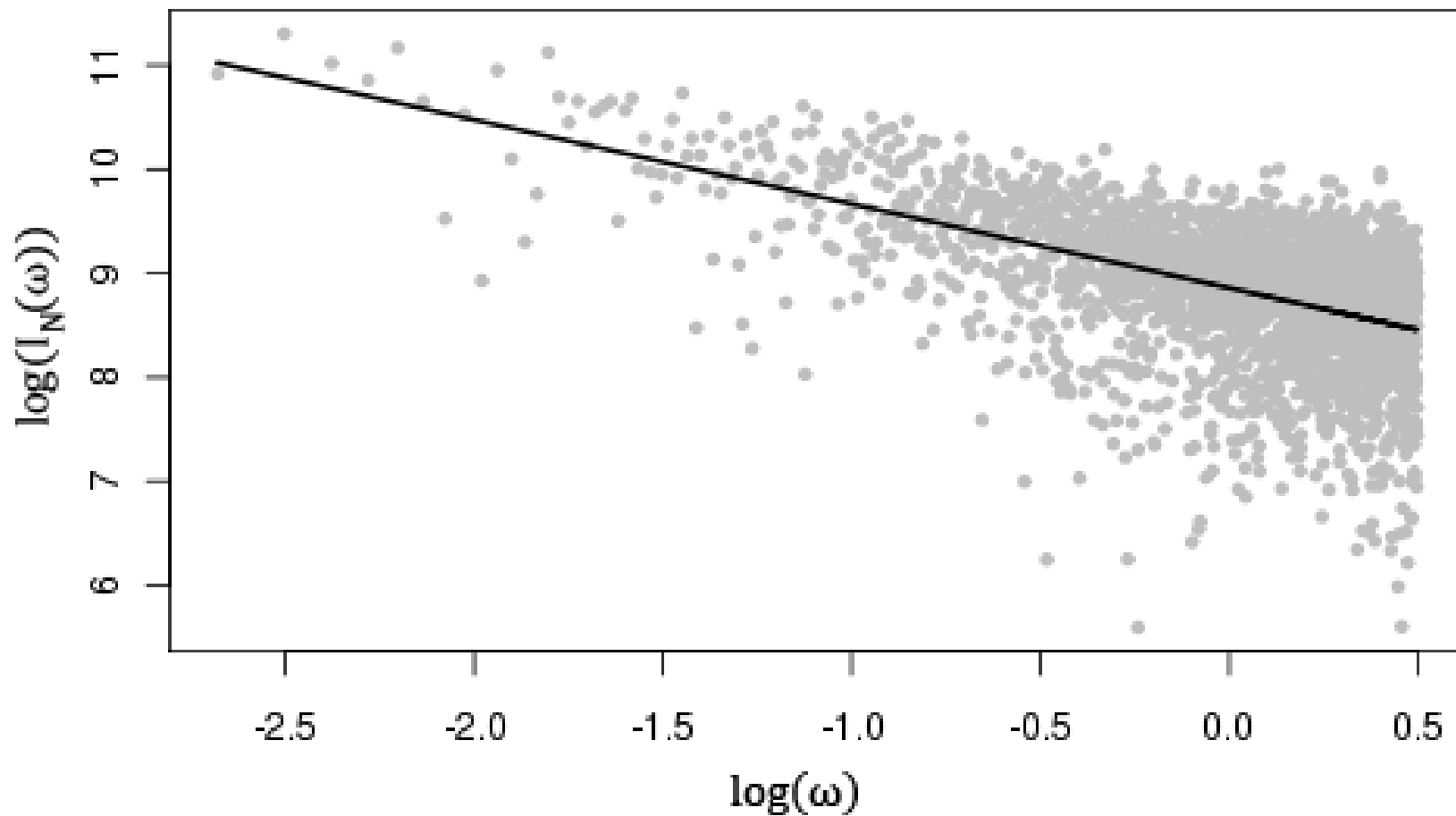
dla której przy  $\omega \rightarrow 0$  zachodzi

$$S(\omega) \sim \omega^{1-2H}$$

skąd dalej

$$\log(I_N(\omega)) \sim (1 - 2H) \log \omega$$

# Metoda periodogramu cd...



# Zarys analizy falkowej

- Ciągła transformata falkowa

$$\{ T_X(a, t) = \langle X, \psi_{a,t} \rangle, a \in \mathbb{R}^+, t \in \mathbb{R} \}$$

$$\left\{ \psi_{a,t}(u) \equiv \frac{1}{\sqrt{a}} \psi_0 \left( \frac{u-t}{a} \right), a \right.$$

$$\left. \in \mathbb{R}^+, t \in \mathbb{R} \right\}$$

gdzie  $\psi_0$  spełnia warunek dopuszczalności

$$\int_{-\infty}^{\infty} \psi_0(u) du = 0$$

# Zarys analizy falkowej

- CWT jest przekształceniem odwracalnym

$$\begin{aligned} X(t) \\ = C_{\Psi} \iint T_X(a, \tau) \psi_{a,\tau}(t) \frac{da d\tau}{a^2} \end{aligned}$$

Gdyby rozważać  $T_X(a, \tau)$  cała informacja zawarta jest w punktach  $(a, \tau)$  półpłaszczyzny.

Można zatem przypuszczać, że można wybrać tylko pewien podzbiór tych punktów.

# Zarys analizy falkowej cd ...

- Podstawa analizy wielorozdzielczej MRA

Wystarczy wziąć zbiór  $(2^j, 2^j k)$ ,  $j, k \in \mathbb{Z}$

i istnieją takie falki macierzyste  $\psi_0$ , że zbiór

$$\{\psi_{j,k}(u) = 2^{-j/2} \psi_0(2^{-j/2} u - k), j, k \in \mathbb{Z}\}$$

tworzy bazę ortonormalną (Riesza) w  $L^2(\mathbb{R})$



# Zarys analizy falkowej cd...

- Dyskretna transformata falkowa DWT

$$\left\{ \left\{ a_X(j, k), k \in \mathbb{Z} \right\}, \left\{ d_X(j, k), j = 1, 2, \dots, J, k \in \mathbb{Z} \right\} \right\}$$

gdzie  $a_X(j, k) = \langle X, \phi_{j,k} \rangle$   $d_X(j, k) = \langle X, \psi_{j,k} \rangle$

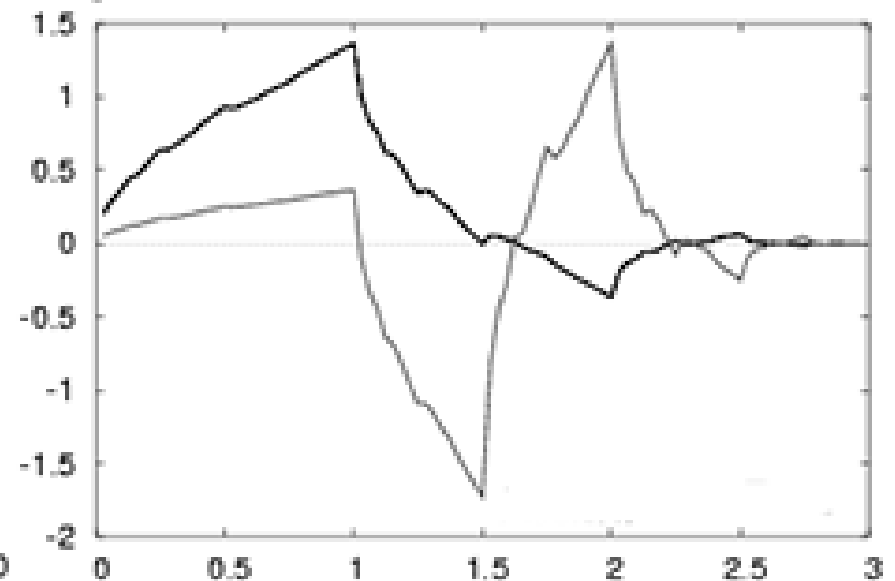
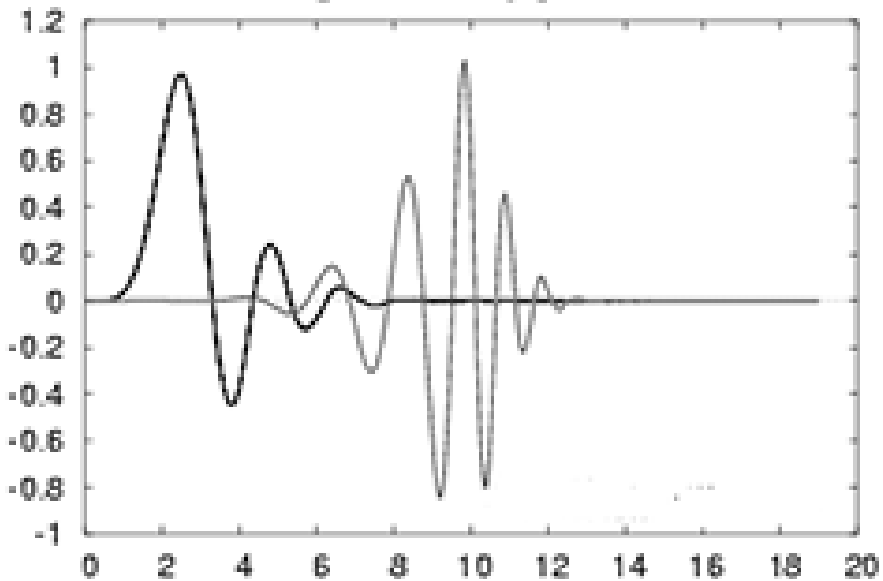
$$\left\{ \phi_{j,k}(u) = 2^{-j/2} \phi_0(2^{-j/2}u - k), k \in \mathbb{Z} \right\}$$

funkcja skalująca  $\phi_0$

# Zarys analizy falkowej cd...

— funkcje skalujące

— odpowiadające im falki macierzyste



# Metoda analizy falkowej

dla dyskretnej transformaty falkowej

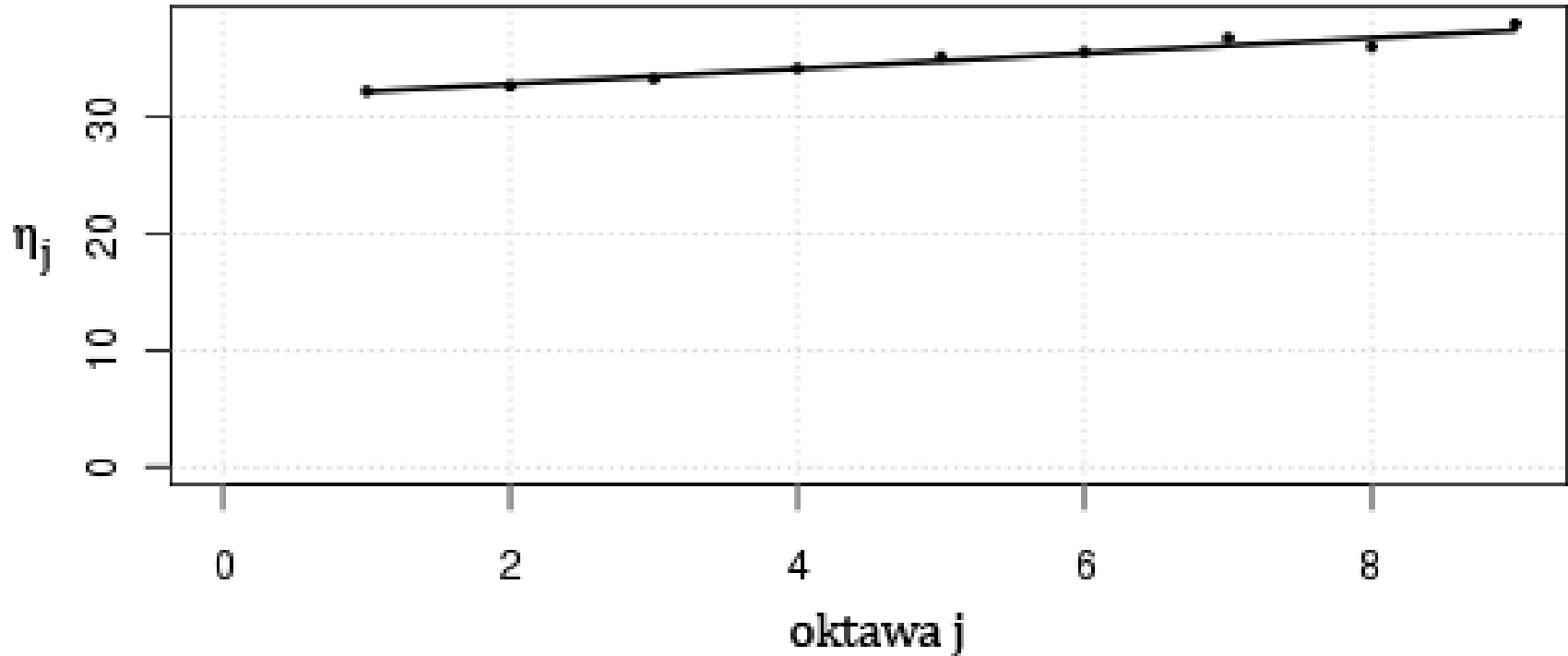
$$\eta_j = E[d_X(j, k)^2]$$

zachodzi

$$\log \eta_j \sim (2H - 1)j + c$$

gdzie  $c = \log C(H, \psi_0)$  to pewna stała zależna od wybranej falki oraz od parametru Hursta.

# Metoda analizy falkowej cd...



# Modelowanie przepływów w sieciach komputerowych

- Model ON/OFF z rozkładem Pareto
- Model oparty o ułamkowy szum Gaussa  $fGn$
- Model oparty o autoregresywne procesy FARIMA

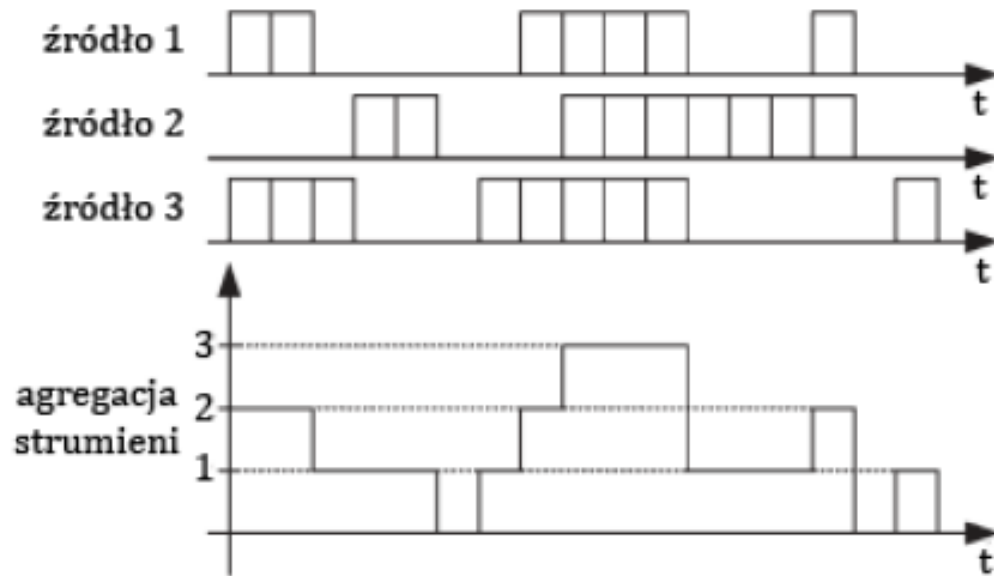
# Model ON/OFF i przepływy głosu VoIP

- zachodzą długoterminowe zależności
- czasy trwania połączeń oraz czasy pomiędzy nawiązaniem połączeń są charakteryzowane są poprzez rozkłady Pareto postaci

$$F(x) = 1 - (\beta/x)^{-\alpha}$$

gdzie  $\alpha > 1$  to parametr kształtu krzywej,  
a  $\beta$  to parametr skali.

# Model ON/OFF z rozkładem Pareto

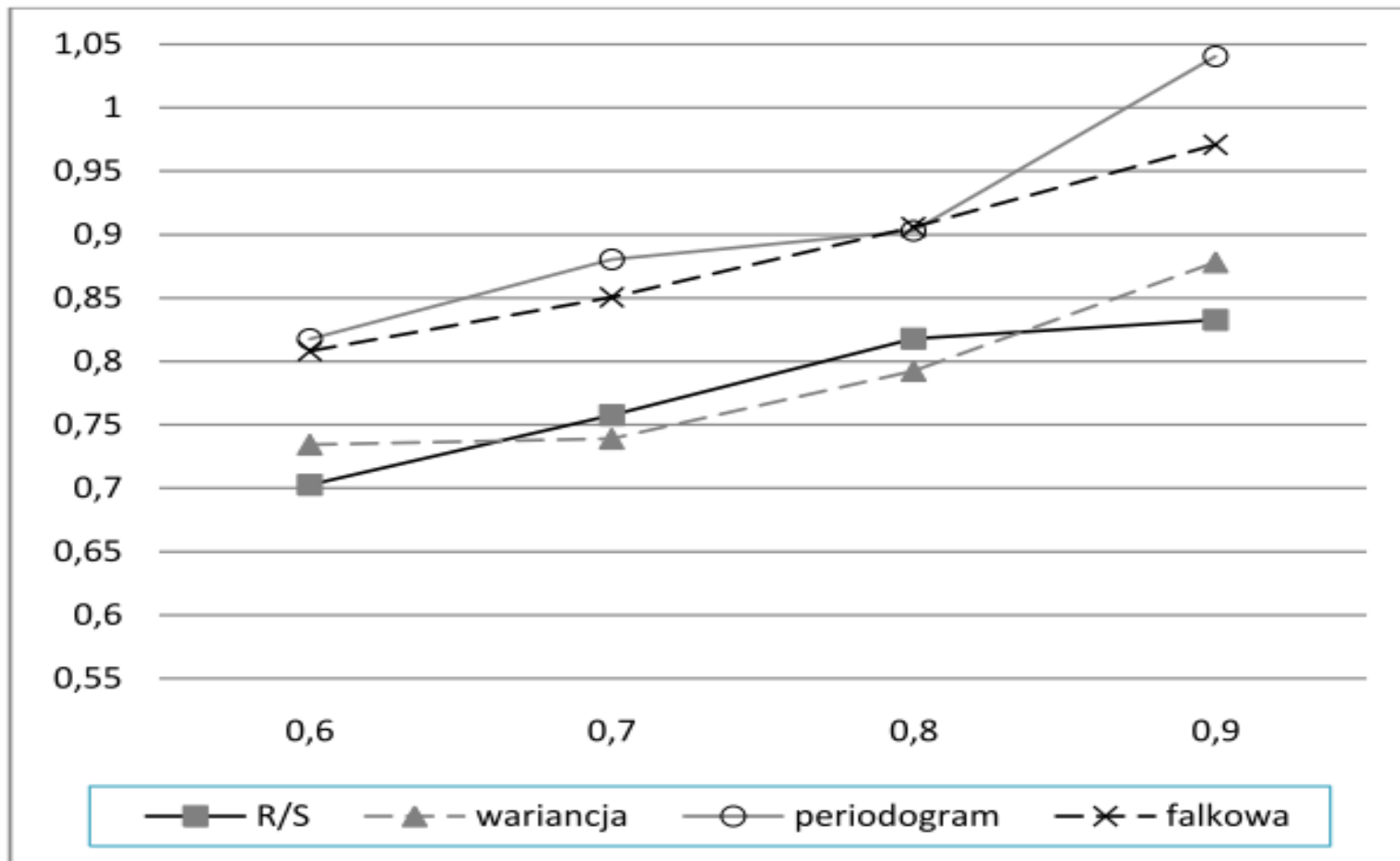


przy ilości agregowanych strumieni  $N \rightarrow \infty$   
zachodzi

$$H = \frac{3 - \alpha}{2}$$

# Badanie modelu ON/OFF

H wyznaczone na podstawie próbek



H docelowe, określone przy generowaniu próbek

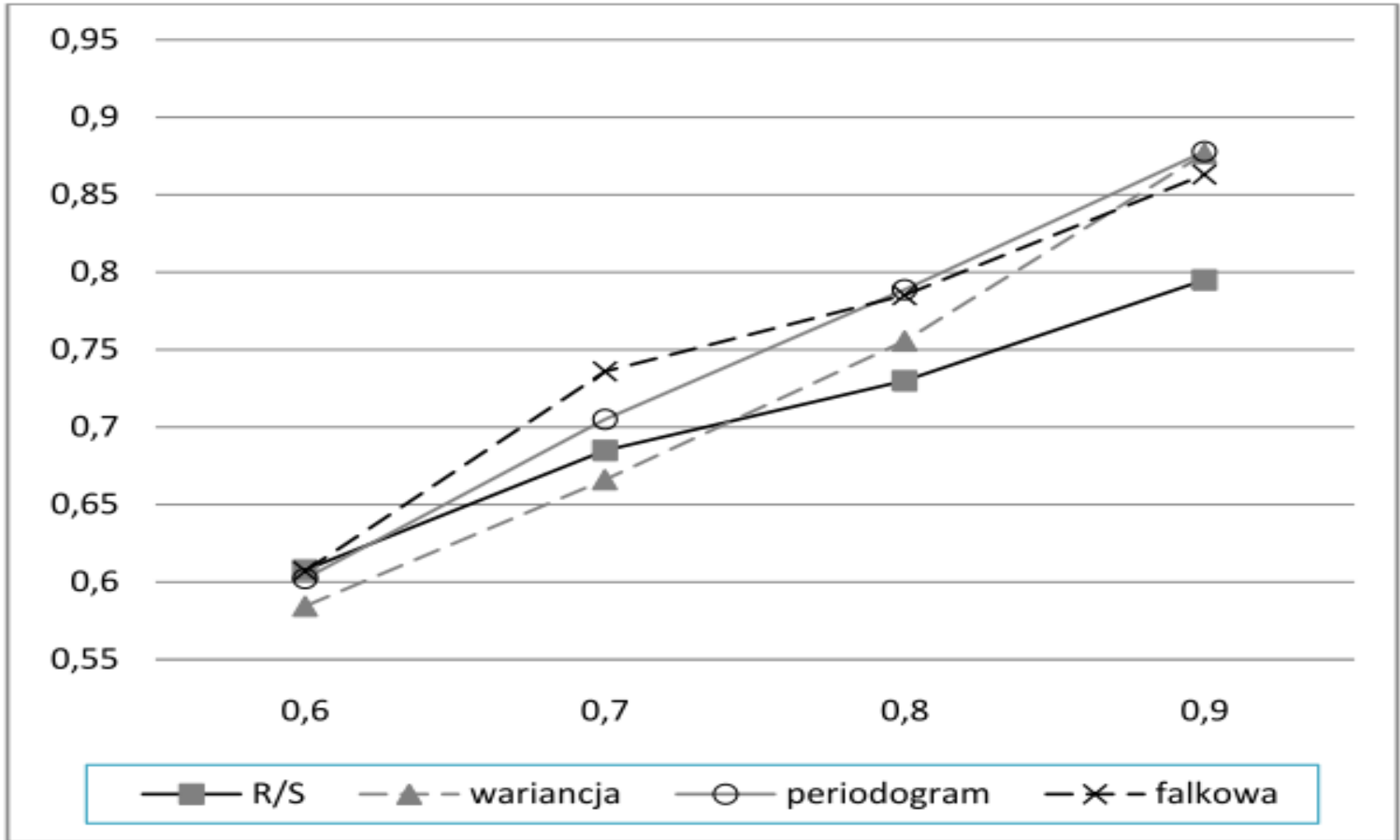


# Model $fGn$ i przepływy wideo VBR

- zachodzą długoterminowe zależności
- model oparty o ułamkowy szum Gaussa  $fGn$
- generowanie ruchu bazuje na szybkiej transformacie Fouriera (FFT) powiązanej z funkcją gęstości widma mocy

# Badanie metod 1

H wyznaczone na podstawie próbek



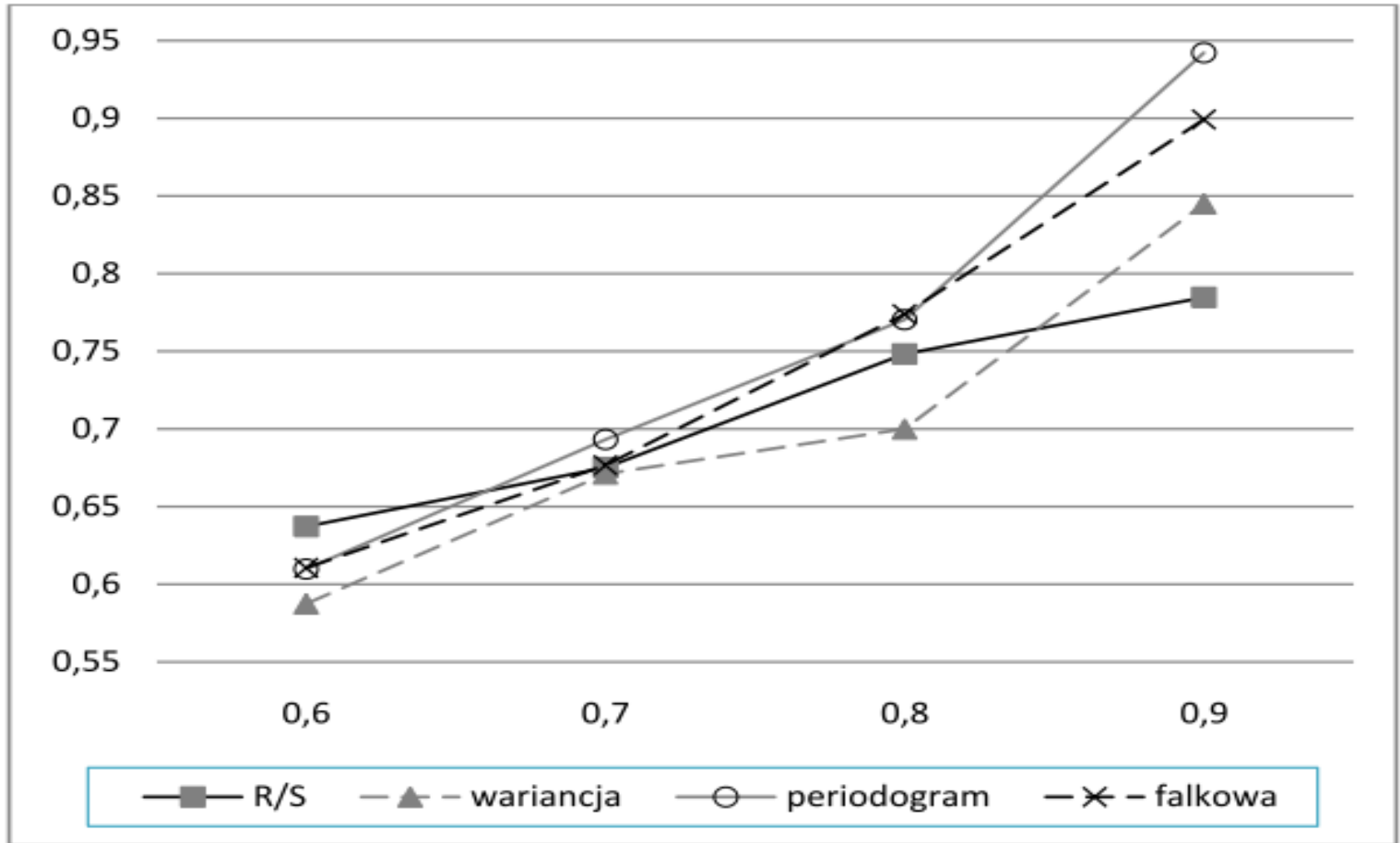
H docelowe, określone przy generowaniu próbek

# Model FARIMA i przepływy WWWW

- zachodzą długoterminowe zależności
- model oparty o autoregresywne procesy FARIMA, które wyznaczają następną zmienną losową rekurencyjnie na podstawie poprzednio wyznaczonych

# Badanie metod cd...

H wyznaczone na podstawie próbek



H docelowe, określone przy generowaniu próbek

# Podsumowanie i sugestie na przyszłość

- przepływy w Internecie często charakteryzowane są poprzez długoterminowe zależności (mają własność samopodobieństwa) i dlatego można je modelować za pomocą rozważanych technik
- krok dalej – procesy multifraktalne i teoria kolejek
- nieskończone podzielne kaskady – rozwinięcie, które zawiera w sobie zarówno charakterystyki monofraktalne, jak i multifraktalne
- natychmiastowe zastosowanie rozważanych technik do aktualnych potrzeb, analiz oraz badań

Dziękuję za uwagę.

Wykorzystano rysunki z badań autora oraz odpowiednio oznaczone z publikacji:

- [1] Willinger, W.; Paxson, V., “Where Mathematics Meets the Internet,” *Notices of the AMS*, vol. 45, 1998, s. 961-970
- [2] Sheluhin, O. I; Smolskiy, S. M.; Osin, A. V., *Self-similar processes in telecommunications*, John Wiley & Sons Ltd, 2007