Wavelet-based estimation of long-range dependence in MPEG video traces

N. Cackov, Ž. Lučić, M. Bogdanov,
and Lj. Trajković
{ncackov, zlucic, ljilja}@cs.sfu.ca, bogdanov@etf.ukim.edu.mk

Communication Networks Laboratory
http://www.ensc.sfu.ca/cnl
Simon Fraser University

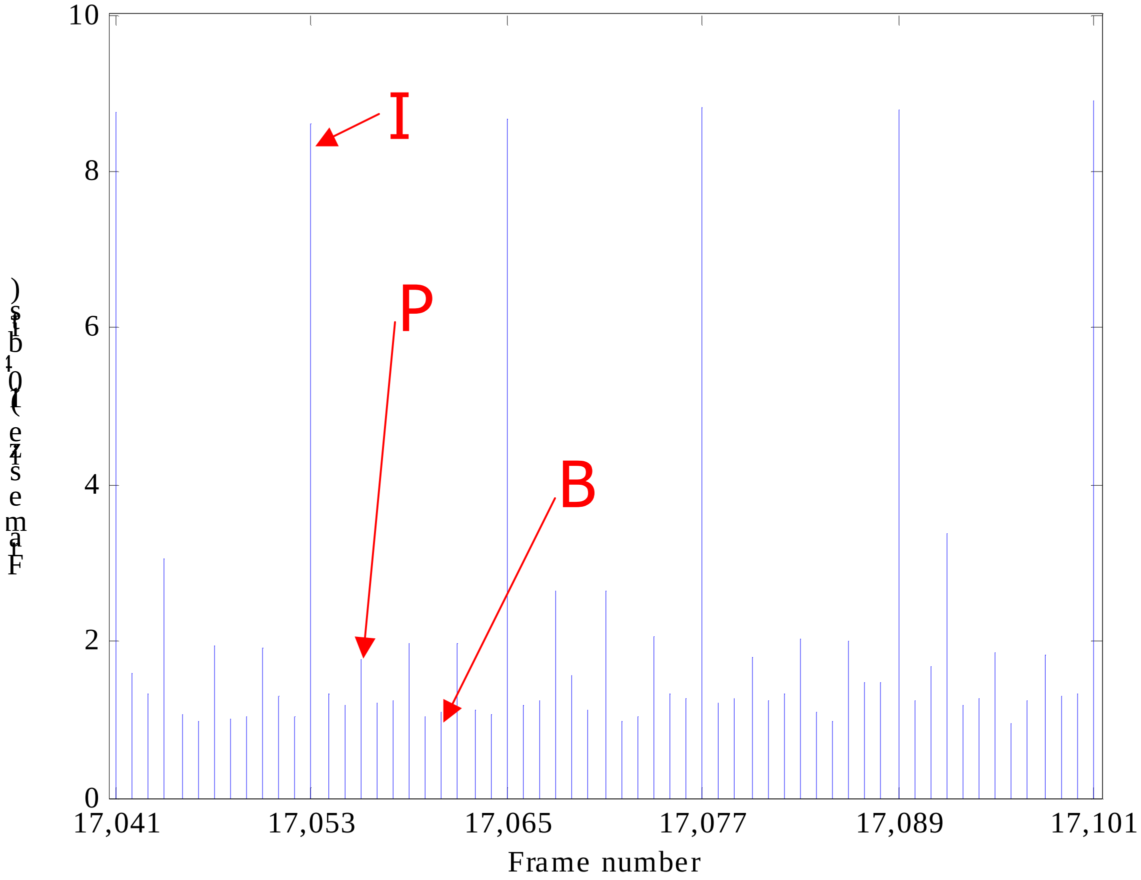
Road map

- Introduction
- Long-range dependence
- Wavelets and estimation of the Hurst parameter
- Analysis of MPEG video traces:
  - estimation of the Hurst parameter
  - sources of unreliability of the estimates
- Conclusion

Introduction

- Network and video traffic is long-range dependent (LRD)
- Hurst parameter characterizes the level of LRD
  - 0.5<H<1: LRD
- Estimators of H:
  - R/S, variance-time, periodogram, Whittle, wavelet-based
- Wavelet-based estimator proved:
  - suitable for network traffic
  - unreliable when applied to MPEG video traces and yields H>1
- We investigate several possible reasons for such behavior of the wavelet-based estimator

MPEG video trace: example



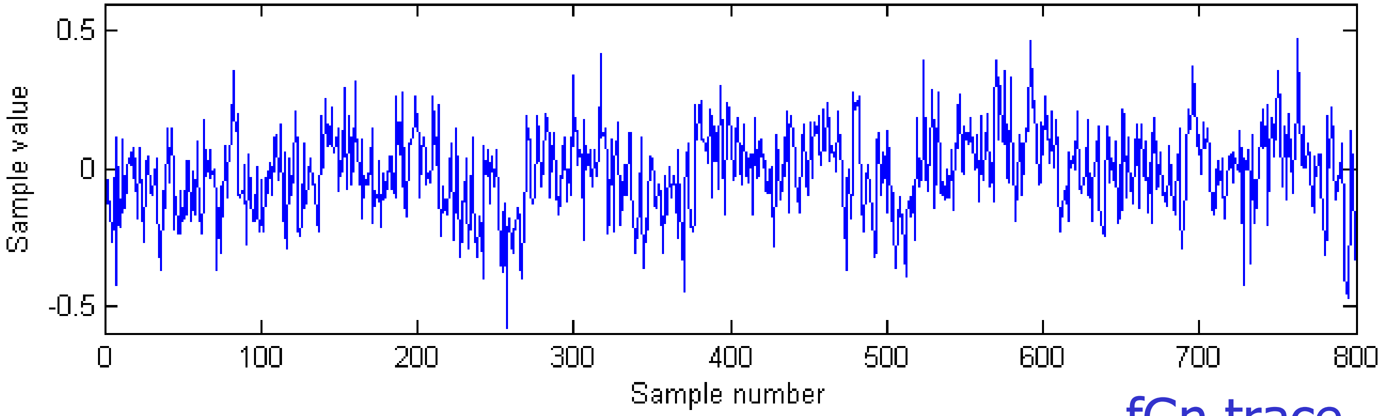
- Group of pictures: IBBPBBPBBPBB

Video traces

- MPEG-1 and MPEG-4 encoded video traces:
  - range: 22,498–89,998 frames per trace
  - coding: 25 frames/second
  - duration: 15–60 minutes of video
- Each sample represents the size of a frame in bits
- Traces:
  - University of Wuerzburg, index of MPEG traces: http://www-info3.informatik.uni-wuerzburg.de/MPEG/traces
  - University of Berlin, MPEG-4 video traces for network performance evaluation: http://www.tkn.ee.tu-berlin.de/research/trace/trace.html

Long-range dependence: properties

- High variability:
  - when the sample size increases, variance of the sample mean decays more slowly than expected
- Burstiness over a range of timescales:
  - long runs of large values followed by long runs of small values, repeated in aperiodic patterns



fGn trace

LRD: definition

Slow decay of the autocorrelation function $r(k)$ of a (wide-sense) stationary process $X(n)$:

$$\boxed{\sum_{k=-\infty}^{\infty} r(k) = \infty}$$

definition

$$\boxed{r(k) = c_r k^{-(2-2H)}, \;\; k \rightarrow \infty}$$

model

$$\boxed{f(\nu) = c_f \,|\nu|^{-\alpha}, \;\; \nu \rightarrow 0}$$

corollary

where $f(\nu)$ is the power spectral density of X(n), c_r and c_f are non-zero constants, and $0<\alpha<1$

LRD: long-range dependence

Wavelet coefficients

- Discrete wavelet transform of a signal $X(t)$:

$$d(j,k) = \int_{-\infty}^{\infty} X(t)\psi_{j,k}(t)dt$$

wavelet coefficients

where

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right)$$

- $\psi(t)$: mother wavelet
  - j : octave
  - k : translation
- Reconstruction formula:

$$X(t) = \sum_{j=0}^{\infty}\sum_{k} d(j,k)\psi_{j,k}(t)$$

LRD and wavelets

- Let $X(t)$ be LRD process (wide-sense stationary)
  - its power spectral density:

    $$f(\nu) \sim c_f \, |\nu|^{-\alpha}, \;\; \nu \to 0$$

- Mean square value of its wavelet coefficients on octave j satisfies:

$$\mathrm{E}\{d(j,k)^2\} = 2^{j\alpha} c_f C(\alpha,\psi)$$

where $C(\alpha,\psi) = \int |\nu|^{-\alpha} \left|\Psi(\nu)\right|^2 d\nu$ does not depend on j

D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. on Information Theory*, vol. 45, no. 3, pp. 878–897, April 1999.

LRD and wavelets

- Logarithm:

$$\log_2 \mathrm{E}\{d(j,k)^2\} = \alpha j + c$$

- Important property: for given *j*, *d(j,k)* does not exhibit long-range dependence (with respect to k)
  - with appropriately chosen mother wavelet
- Hence:
  - simple estimator for E{d(j,k)^2} is a sample mean:

$$\mathrm{E}\{d(j,k)^2\} = \frac{1}{n_j} \sum_{k=1}^{n_j} d(j,k)^2$$

  - n_j: number of wavelet coefficients at octave j

Estimation of α and H

- Logscale diagram: plot of $log_2E\{d(j,k)^2\}$ vs. j (octave)
- Linear relationship between $log_2E\{d(j,k)^2\}$ and j on the coarsest octaves indicates LRD
- Estimation of α:
  - linear regression of $log_2E\{d(j,k)^2\}$ on j in the linear region of the logscale diagram
- $H = 0.5 (\alpha + 1)$

Multifractal estimator

- Extension of the basic monofractal estimator
- Takes into account moments of arbitrary order

$$S_q(j) = \mathrm{E}\{d(j,k)^q\}$$

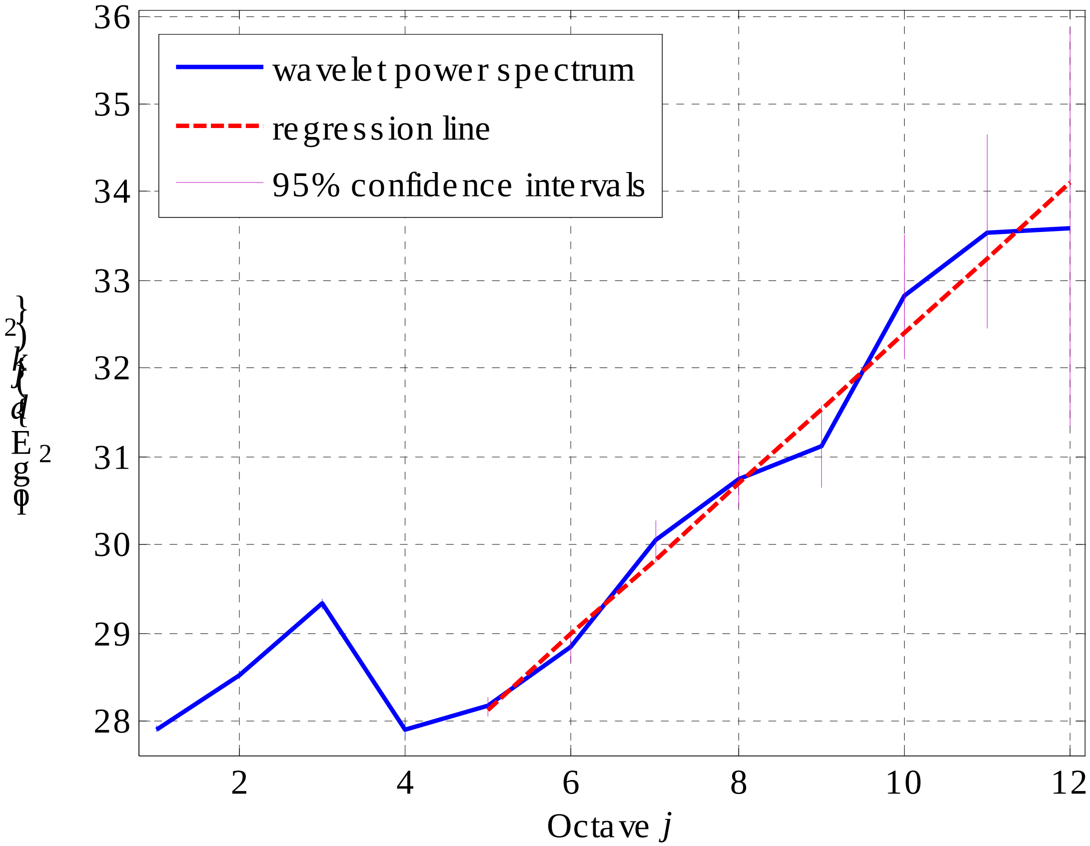
$$= \frac{1}{n_j}\sum_{k=1}^{n_j} d(j,k)^q \qquad \text{sample mean}$$

  - *q:* order of the moment
  - n_j: number of wavelet coefficients at octave j

$$\log_2 S_q(j) = \alpha_q j + c$$

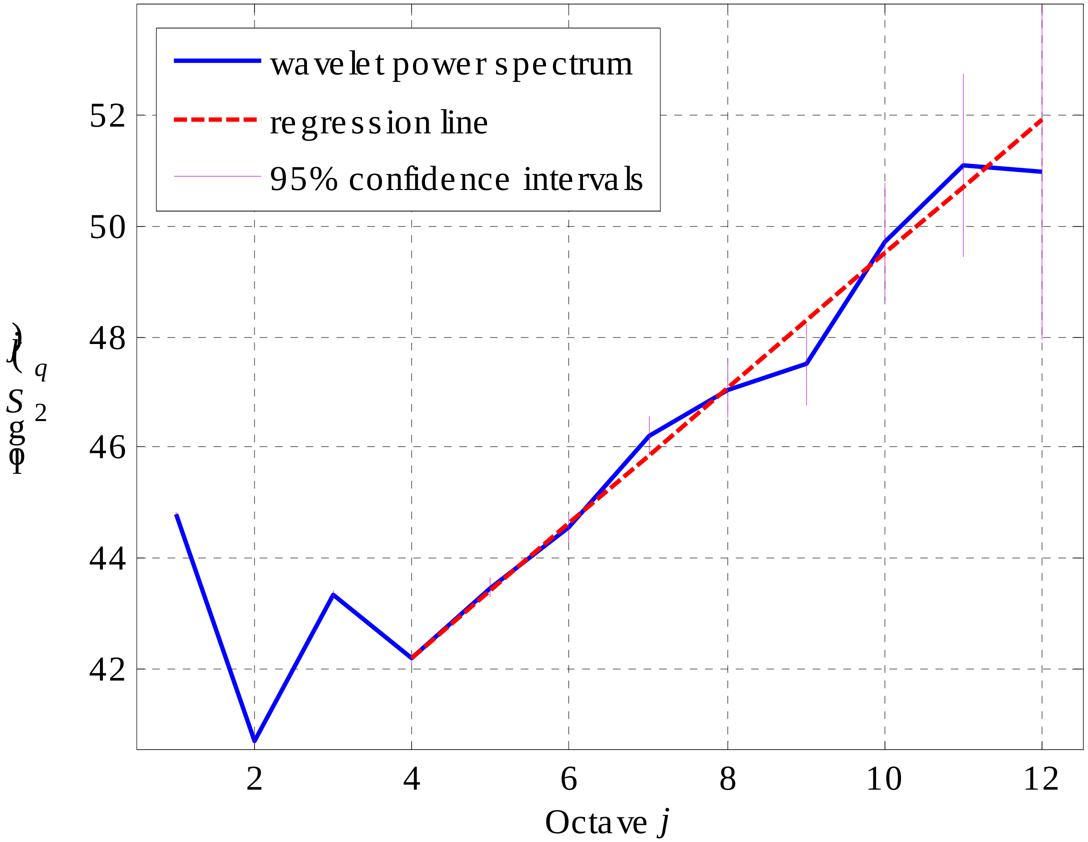
$$H = 0.5 + \alpha_q / q$$

Logscale diagram: monofractal



- Simpsons (MPEG-1)
- α=0.852, H=0.926 (octaves 5-12)

Logscale diagram: multifractal



- Simpsons (MPEG-1)
- α_3=1.218, H=0.906 (octaves 4-12)

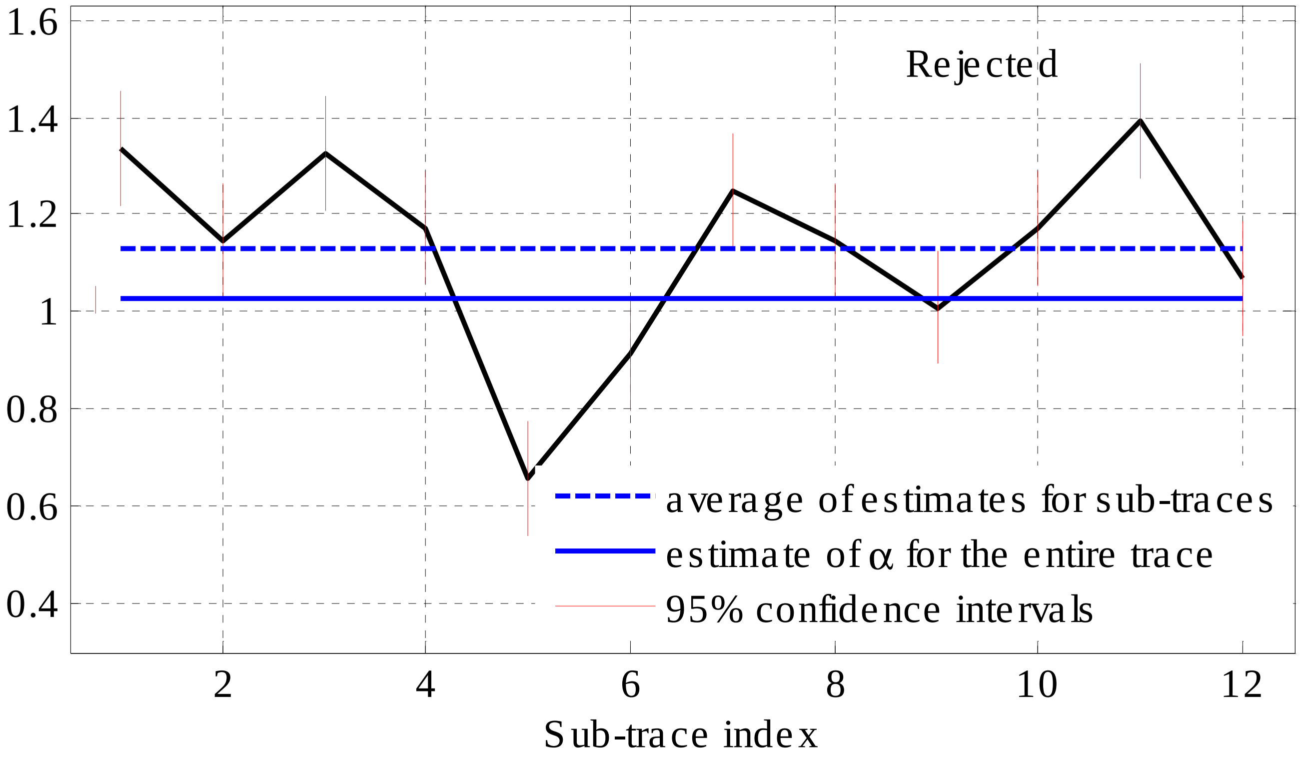
Assumptions

- For fixed j:
  $d(j,k)$ are stationary sequences of uncorrelated variables
- Processes $d(j,k)$ and $d(j',k)$ $(j \neq j')$ are uncorrelated
- Process $X(t)$ and, hence, the processes $d(j,k)$ are Gaussian

Test for time constancy of α

- $X(n)$: wide-sense stationary process
  - α does not depend on n
- Is α constant throughout the time series $X(n)$?
- Approach:
  - divide $X(n)$ into m blocks of equal lengths
  - estimate α for each block
  - compare the estimates
- If α varies significantly, estimating α for the entire time series is not meaningful

Test for constancy: example



- Trace is divided into 12 sub-traces of equal lengths
- Variation of the scaling exponent indicates that α is not constant

Star Wars IV (MPEG-4)

Estimates of H

Trace	Encoding	H (mono)	H (multi)	H (per)	H (R/S)
MTV	MPEG-1	0.959	0.937	0.992	0.89
Jurassic park	MPEG-1	1.096	1.012	1.191	0.88
Simpsons	MPEG-1	0.926	0.906	0.988	0.89
Mr. Bean	MPEG-1	1.214	1.258	1.295	0.85
Silence of the lambs	MPEG-1	1.130	1.152	1.171	0.89
Talk show	MPEG-1	1.084	1.132	1.174	0.89
ARD news	MPEG-4	1.382	1.310	1.310	0.967
Die hard III	MPEG-4	1.190	1.208	1.233	0.969
Formula 1	MPEG-4	1.189	1.169	1.216	0.867
Futurama	MPEG-4	0.943	0.909	1.064	0.877
From dusk till dawn	MPEG-4	1.139	1.138	1.186	0.909
First contact	MPEG-4	1.194	1.213	1.268	0.931
Star Wars IV	MPEG-4	1.013	1.051	1.138	0.903

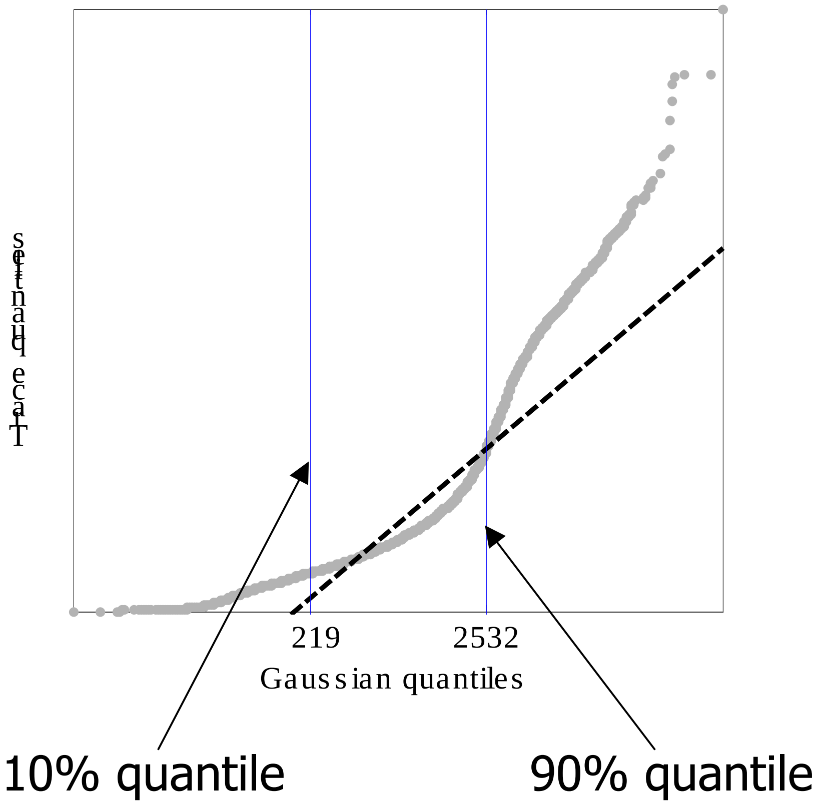
Unreliable estimates (MPEG)

- Wavelet-based estimator often yields H>1
- Possible causes:
  - non-Gaussianity of data
  - non-stationarity of α
  - LRD-SRD interactions
  - behavior of power spectral density
- Perform additional tests:
  - Gaussianity
  - time constancy of α
  - investigation of LRD and SRD components

SRD: short-range dependence

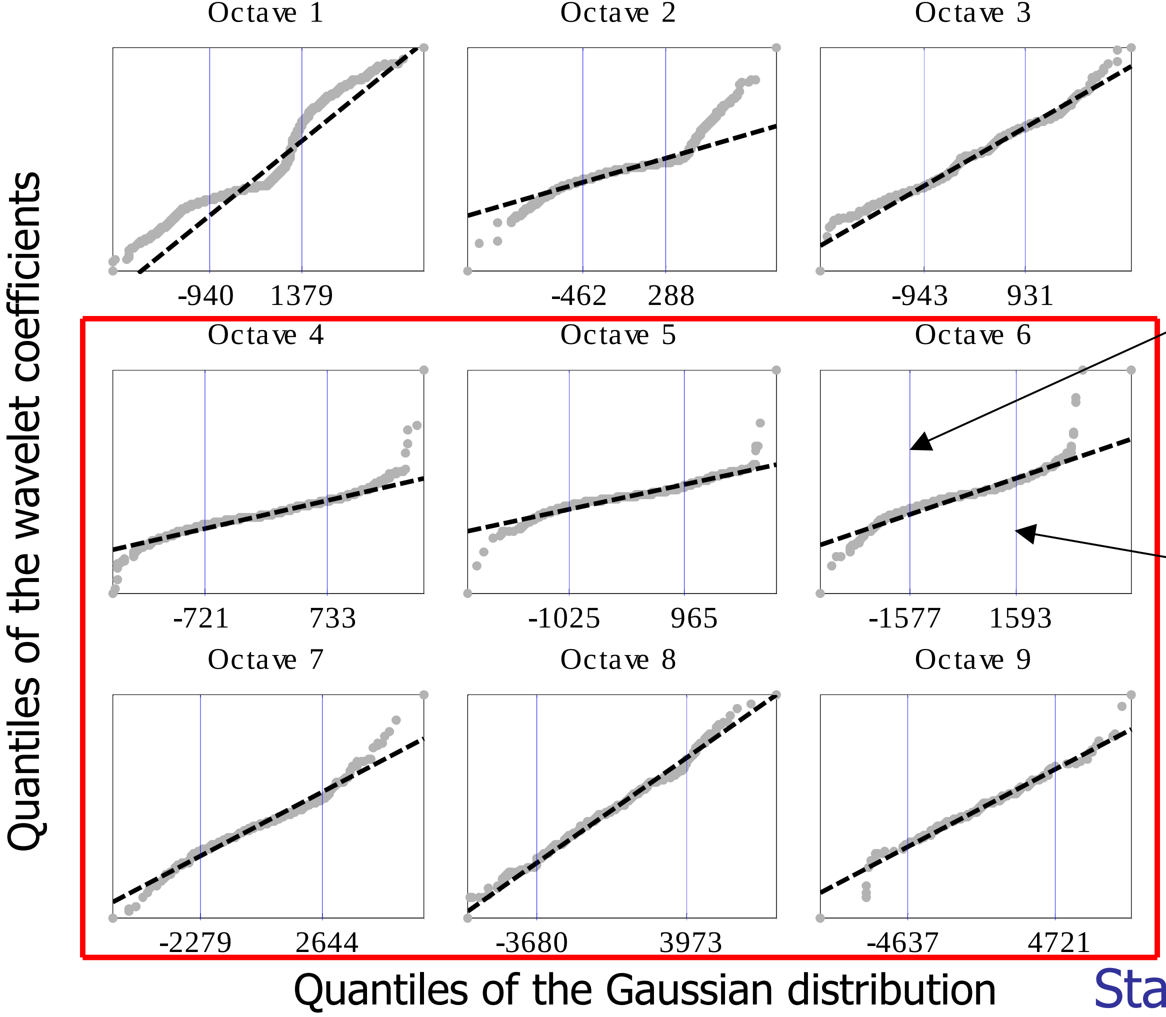
Star Wars IV (MPEG-4): comparison with Gaussian distribution

qq-plot



- Video frames were ordered based on the number of bits
- Gaussian distribution:
  - identical mean and variance as the video trace
  - samples were ordered in ascending order
- Video trace departs from the Gaussian distribution

Wavelet coefficients: comparison with Gaussian distribution



10% quantile

90% quantile

Star Wars IV (MPEG-4)

qq-plots: observations

- Analyzed traces are highly non-Gaussian
- Wavelet coefficients at the finest octaves deviate from Gaussian distribution
- Wavelet coefficients at the medium and coarse octaves have approximately Gaussian distribution

Test for time constancy of α

Trace	Encoding	failed	passed
MTV	MPEG-1	7	0
Jurassic park	MPEG-1	7	0
Simpsons	MPEG-1	2	5
Mr. Bean	MPEG-1	7	0
Silence of the lambs	MPEG-1	7	0
Talk show	MPEG-1	7	0
ARD news	MPEG-4	3	4
Die hard III	MPEG-4	7	0
Formula 1	MPEG-4	4	3
Futurama	MPEG-4	7	0
From dusk till dawn	MPEG-4	6	1
First contact	MPEG-4	7	0
Star Wars IV	MPEG-4	5	2

Time constancy: observations

- MPEG traces were tested for stationarity of α for various values of m (number of blocks)
- Two types of behaviour:
  - fail the test for all m’s
  - pass for some m, fail for others
- Failures of the test dominate
- Exceptions:
  - Simpsons (MPEG-1) and ARD news (MPEG-4)

LRD-SRD interactions

- R/S estimates of H are often greater than 0.9
    - strong LRD component
- Conjecture: traces posses strong SRD component
    - similarities within a single scene of video
- Wavelet-based estimator produces unreliable results when applied to traces with both:
    - strong SRD
    - strong LRD components

F. Xue and Lj. Trajković, “Performance analysis of a wavelet-based Hurst parameter estimator for self-similar network traffic,” in *Proc. SPECTS '2K*, Vancouver, BC, May 2000, pp. 294–298.

MPEG traces: discussion

- Wavelet coefficients over the coarsest scales have Gaussian distribution
- Generally α is not constant for the entire trace
- LRD-SRD interactions may cause unreliable estimates
- Wavelet- and periodogram-based estimators of H produce similar results
  - estimated power spectral density obeys a power law with exponent $\alpha > 1$

Conclusion

- We estimated H for several MPEG traces using the wavelet-based estimator
  - estimates of H were often greater than one
- Wavelet-based estimates of H:
  - agree with periodogram-based estimates
  - differ from those obtained by R/S estimator
- Most MPEG traces fail test for constancy of α
- Unreliability of the estimates may be attributed to the interactions of SRD and LRD components in the trace

References

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, “On the self-similar nature of Ethernet traffic (extended version),” *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.

[2] K. Park and W. Willinger, “Self-similar network traffic: an overview,” in Self-similar Network Traffic and Performance Evaluation, edited by K. Park and W. Willinger. New York: Wiley, 2000, pp. 1–38.

[3] O. Rioul and M. Vetterli, “Wavelets and signal processing,” *IEEE Signal Processing Magazine*, vol. 8, issue 4, pp. 14–38, Oct. 1991.

[4] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, “Wavelets for the analysis, estimation, and synthesis of scaling data,” in Self-similar Network Traffic and Performance Evaluation, edited by K. Park and W. Willinger. New York: Wiley, 2000, pp. 39–88.

[5] P. Abry and D. Veitch, “Wavelet analysis of long-range dependent traffic,” *IEEE Trans. Information Theory*, vol.44, no.1, pp.2–15, Jan. 1998.

[6] D. Veitch and P. Abry, “A wavelet-based joint estimator of the parameters of long-range dependence,” *IEEE Trans. Information Theory*, vol.45, no.3, pp.878–897, Apr. 1999.

[7] D. Veitch and P. Abry, “A statistical test for time constancy of scaling exponents,” *IEEE Trans. Signal Processing*, vol.49, no.10, pp.2325–2334, Oct. 2001.

[8] O. Rose, “Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems”, University of Wuerzburg, Institute of Computer Science, Report no. 101, Feb. 1995.

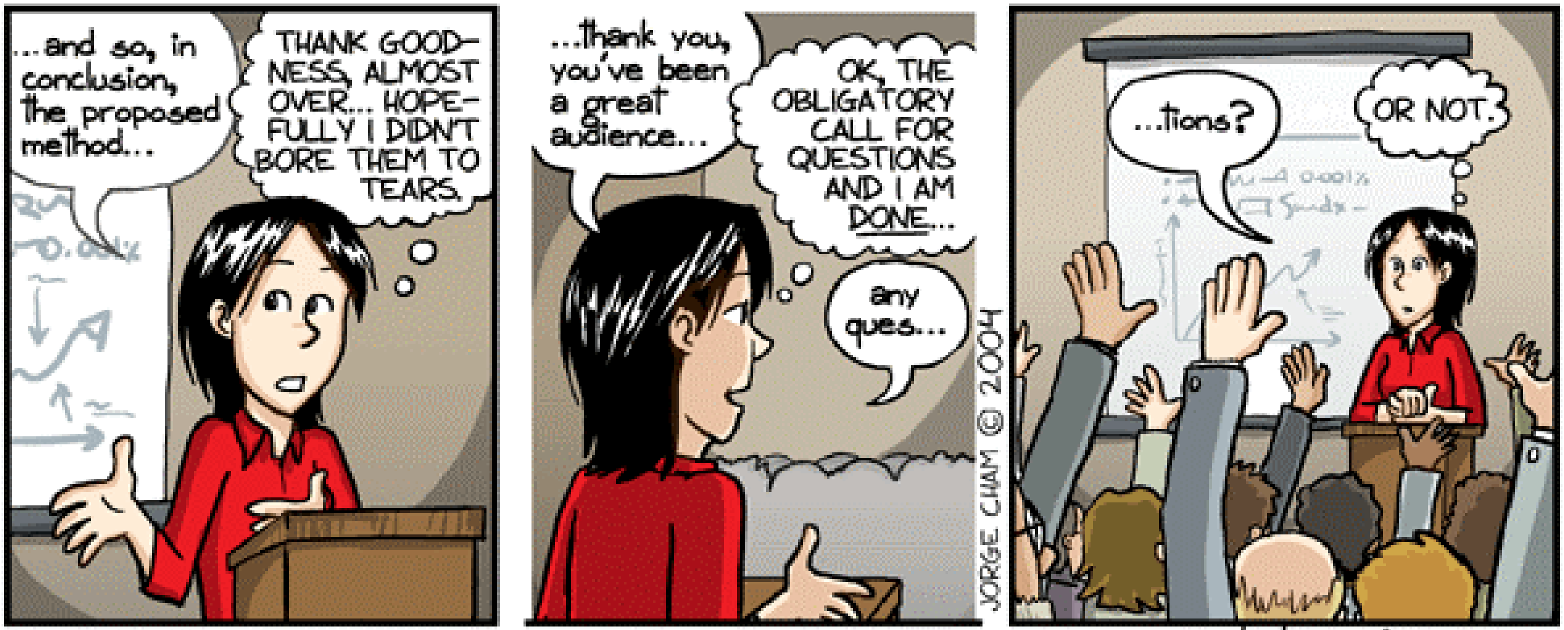
Web pages

[9] D. Veitch, MATLAB code for estimation of scaling exponents:
http://www.cubinlab.ee.mu.oz.au/~darryl/secondorder_code.html.

[10] University of Wuerzburg, Index of MPEG traces:
http://www-info3.informatik.uni-wuerzburg.de/MPEG/traces/.

[11] University of Berlin, MPEG-4 and H.263 video traces for network performance evaluation:
http://www-tkn.ee.tu-berlin.de/research/trace/trace.html.

Questions?



"Piled Higher and Deeper" by Jorge Cham

www.phdcomics.com

Second-order self-similarity

Let X(n) be wide-sense stationary discrete-time process with autocorrelation function r(k). Define:

$$X^{(m)}(i) = \frac{1}{m} \sum_{n=m(i-1)+1}^{mi} X(n), \quad m \in \mathrm{N}$$

- X(n) is second-order self-similar
  - exactly, if

$$r^{(m)}(k) = r(k) = \frac{\sigma^2}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H})$$

  - asymptotically, if

$$\lim_{m\to\infty} r^{(m)}(k) = \frac{\sigma^2}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H})$$

where $r^{(m)}(k)$ is autocorrelation function of $X^{(m)}(i)$

Second-order self-similarity

X(n): a wide-sense stationary discrete-time process with autocorrelation function ρ(k)

Define:

$$X^{(m)}(i) = \frac{1}{m} \sum_{n=m(i-1)+1}^{mi} X(n), \quad m \in \mathrm{N}$$

- X(n) is second-order self-similar if

$$\rho^{(m)}(k) = \rho(k) = \frac{1}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]$$

where $\rho^{(m)}(k)$ is autocorrelation function of $X^{(m)}(i)$

- H: Hurst parameter

Self-similarity and LRD

- X(n): second-order self-similar process

$$\rho(k) \sim H(2H-1)k^{2H-2},\ k \to \infty,\ 0 < H < 1$$

- X(n): LRD process

$$\rho(k) \sim c_{\rho} k^{\alpha-1},\ k \to \infty$$

- Second-order self-similarity implies LRD and vice-versa, with restriction
  - 0.5<H<1

with α=2H-1 or H=0.5(α+1)

- Important: X(n) is wide-sense stationary

Self-similarity: continuous time

- Y(t) is self-similar with parameter $H>0$ if:
    - $Y(0)=0$
    - Y(t) and $c^{-H}Y(ct)$ $(c>0)$ have the same finite-dimensional distributions
        - Y(t): non-stationary
        - H: Hurst parameter
- Y(t) is self-similar with stationary increments if:
    - Y(t) is self-similar
    - finite dimensional distributions of $\{Y(t+k)-Y(t)\}$ do not depend on t

Wavelet transform

- Continuous wavelet transform of $X(t)$:

$$w(a,b) = \int_{-\infty}^{\infty} X(t)\psi_{a,b}^{*}(t)dt$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

- $\psi(t)$ is called “mother wavelet”
  - a: scale factor, b: translation factor

Properties of wavelets

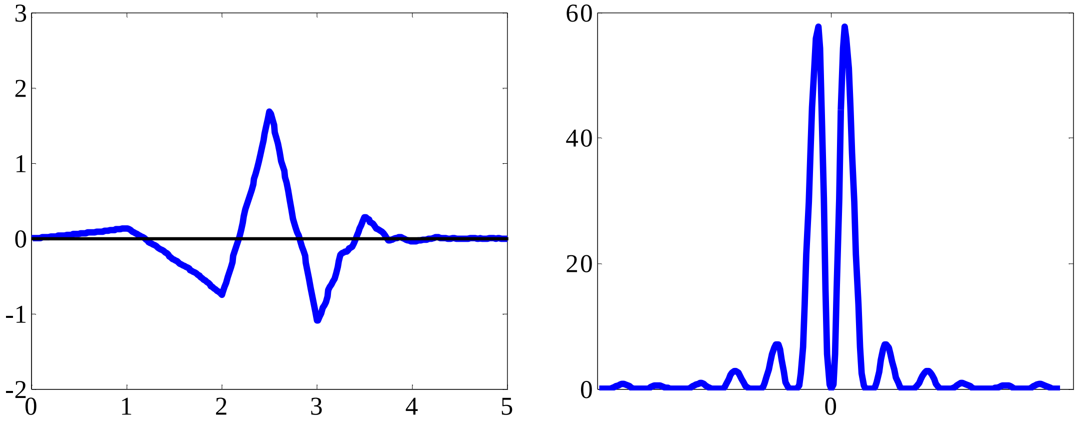
- Admissibility condition:

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \Leftrightarrow \left|\Psi(\omega)\right|^2\Big|_{\omega=0} = 0$$

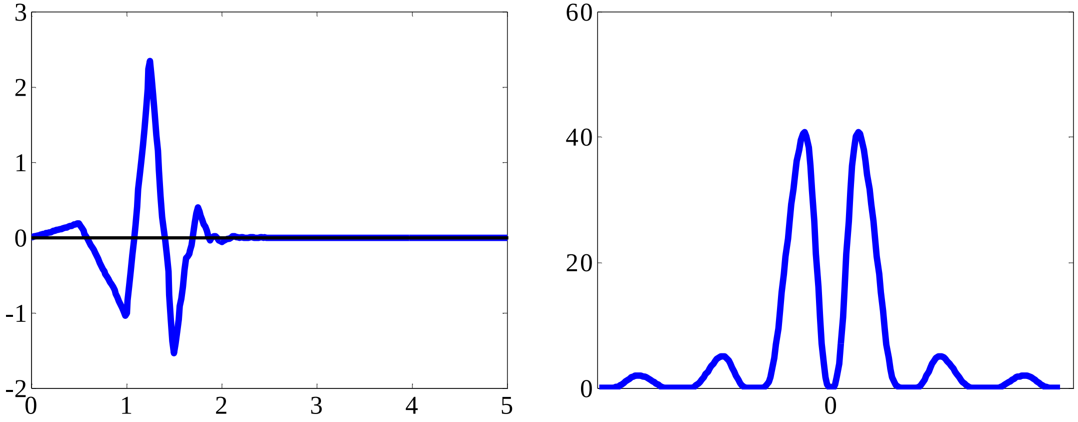
- Number of vanishing moments N:

$$\int_{-\infty}^{\infty} t^k \psi(t)dt = 0, \quad k = 0,1, \ldots, N-1$$

Scaling of wavelets

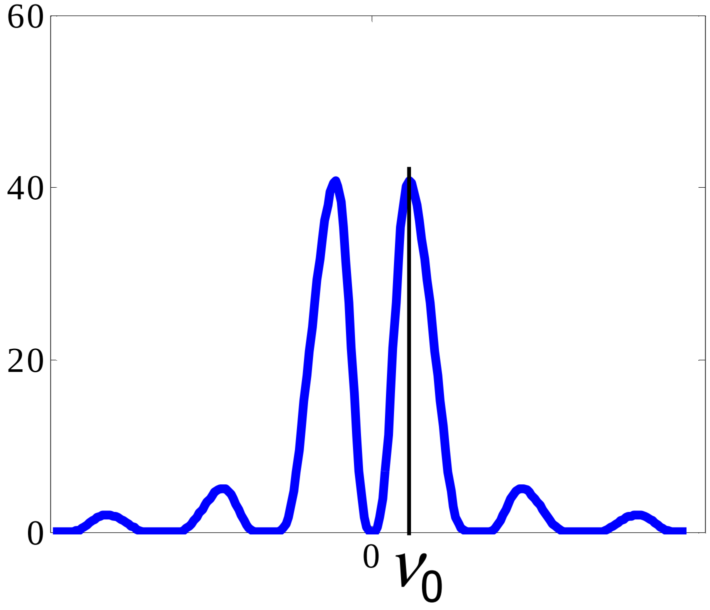


- Mother wavelet (Daubechies3) and its spectrum

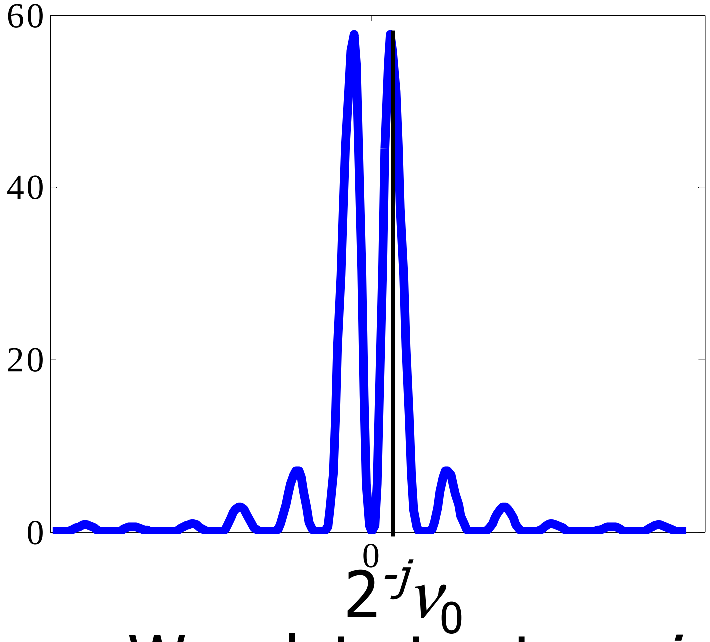


- Scaled wavelet (Daubechies3) and its spectrum (a=0.5)

Reminder



- Mother wavelet (spectrum)



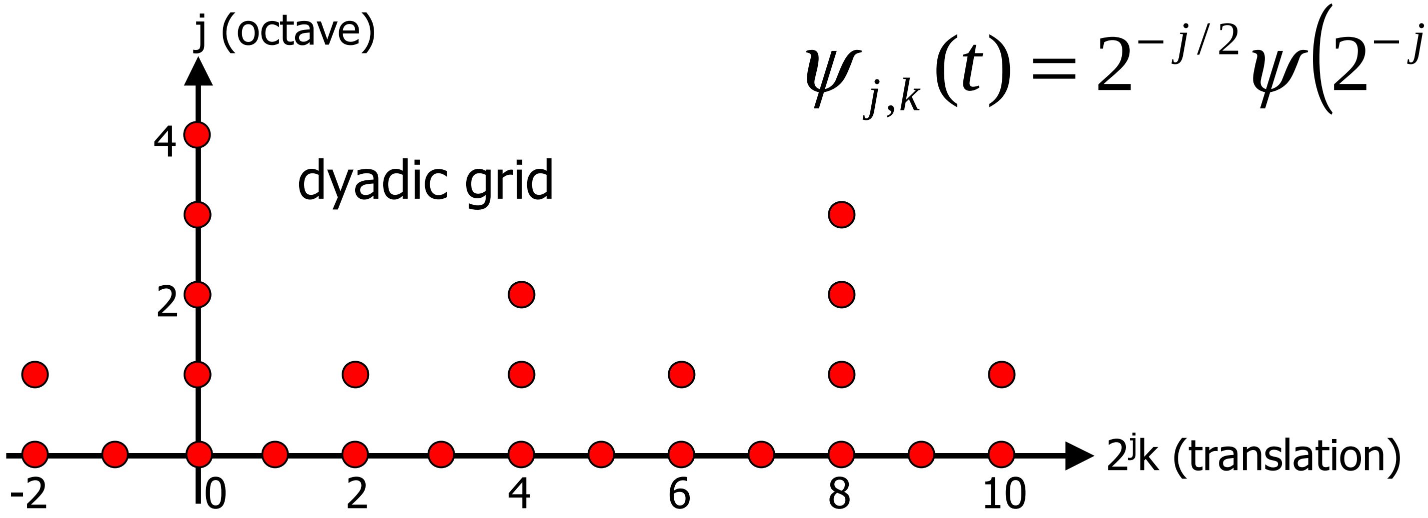
- Wavelet at octave *j* (spectrum)

octave = $\log_2$(scale)
scale ~ 1 / frequency
coarse scales = low frequencies

Discrete wavelets

- Dyadic sampling of the translation-scale plane
  - $a=2^j$, $b=2^jk$
    - j: positive integer (octave), k: integer (translation)
  - $d_{j,k} = w(2^j, 2^jk)$ wavelet coefficients

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t-k\right)$$



Wavelets and LRD (1)

- Let X(t) be LRD process (wide-sense stationary)
  - $d_{j,k}$: its wavelet coefficients
  - $\Gamma(\nu)$: power spectral density (PSD) of X(t)
  - $\Psi_j(\nu)$: spectrum of the wavelet at octave j
  - $\Psi_0(\nu)$: spectrum of the mother wavelet
- Average energy of X(t) around frequency $2^{-j}\nu_0$, where ν_0 is the "central" frequency of the mother wavelet:

$$\mu_j := \mathrm{E}\{d(j,k)^2\} := \int \Gamma(\nu)\,|\Psi_j(\nu)|^2\,d\nu = \int \Gamma(\nu) 2^j\,|\Psi_0(2^j\nu)|^2\,d\nu$$

Wavelets and LRD (2)

- Recall:

$$\Gamma(\nu) \sim c_f \, |\nu|^{-\alpha}, \; \nu \to 0$$

- Substituting yields:

$$\mu_j = 2^{j\alpha} \underbrace{c_f \int |\nu|^{-\alpha} \left|\Psi_0(\nu)\right|^2 d\nu}_{=e^c}$$

- Finally:

$$\log_2 \mu_j = \alpha j + c$$

P. Abry and D. Veitch, “Wavelet analysis of long-range dependent traffic,” *IEEE Trans. Information Theory*, vol.44, no.1, pp.2-15, Apr. 1998.

Wavelets and LRD: summary

- Conclusions:
  - $|\nu|^{-\alpha}$ in power spectral density of a signal X(t) implies linear dependence (with slope α) of $\log_2\mu_j$ on j
  - It may hold for all scales or only over a range of scales:
    - range of scales including the coarsest: LRD
  - logscale diagram: plot of $\log_2\mu_j$ vs. j (octave)

Linear regression

- Estimation of α: calculating the slope of the linear region in the log-log plot
- Method: linear regression of y_j on x_j
  - y_j: random variables
  - x_j: deterministic independent variables
- Hypothesis: $E\{y_j\} = bx_j + a$
- In our case
  - $y_j = \log_2 \mu_j$
  - $x_j = j$
  - $b = \alpha$

Good, but ...

- Generally:
  - $E\{\log_2(x)\} \neq \log_2(E\{x\})$
- Our case:
  - $E\{\log_2\mu_j\} \neq \log_2(E\{\mu_j\})=\alpha j + c$

hypothesis estimated

- Solution: introducing a correction factor g(j)
  - new random variable: $y_j = \log_2\mu_j - g(j)$
  - perform linear regression of y_j on j

Calculation of $\alpha(1)$

- Under the given assumptions:

$$g(j) = \frac{\Gamma'(n_j/2)}{\Gamma(n_j/2)\ln 2} - \log_2(n_j/2)$$

$$Var(y_j) := \sigma_j^2 = \zeta(2, n_j/2)/\ln^2 2$$

where $\Gamma(x)$ is the Gamma function and $\zeta(z,v)$ is a generalized Riemann Zeta function

Calculation of α (2)

- Finally (all sums are over j, for a given range of octaves):

$$S = \sum 1/\sigma_j^2$$

$$S_j = \sum j/\sigma_j^2$$

$$S_{jj} = \sum j^2/\sigma_j^2$$

minimum-variance unbiased estimator

$$\hat{\alpha} = \frac{\sum y_j (Sj - S_j)/\sigma_j^2}{SS_{jj} - S_j^2}$$

Estimation of α: summary

- Estimation of α: calculating the slope of the linear region in the logscale diagram
- Method: weighted linear regression of $\log_2\mu_j$ on j
  - $\log_2\mu_j$: random variables
  - j: deterministic independent variables
  - weight: 1/(variances of the estimates of $\log_2\mu_j$)
- $H = 0.5 (\alpha + 1)$

Periodogram

- Estimator of the PSD of a discrete signal X(n)

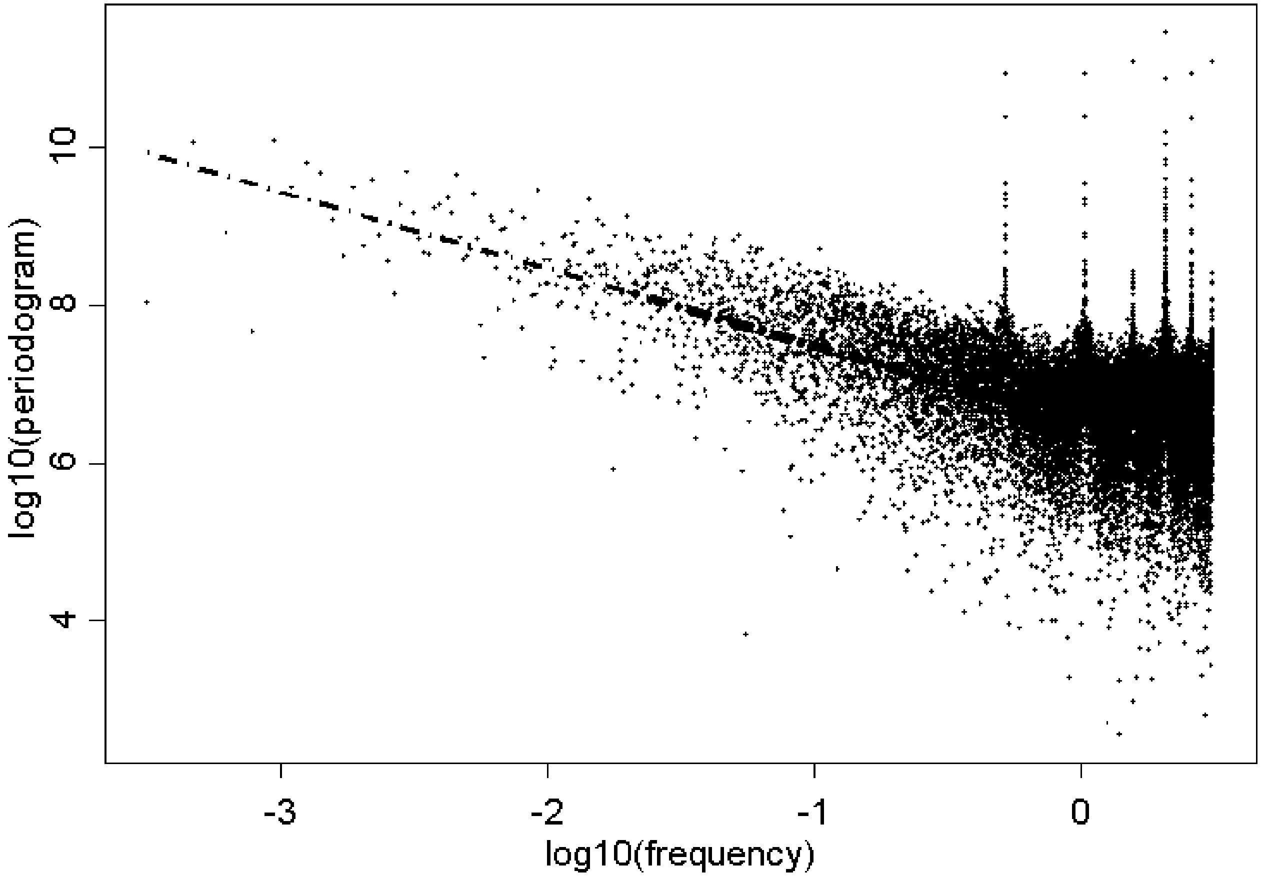
$$\hat{\Gamma}(\nu) = \frac{1}{2\pi N} \left| \sum_{k=1}^{N} X(k) e^{jk\nu} \right|^2$$

- Power spectral density of a LRD signal:

$$\Gamma(\nu) \sim c_f \, |\nu|^{-\alpha}, \; \nu \to 0$$

- α: slope of log-log plot of the periodogram
  - using the first 10% of the frequencies
- H = 0.5 (α + 1)

Periodogram: example



- Simpsons (MPEG-1)
- α=0.976, H=0.988

Rescaled adjusted range (R/S) plot

- X(n): discrete stochastic process, 1≤n≤N

$Y_j = \sum_{n=1}^{j} X(n)$ cumulative process

- k: lag (0≤k≤N); t: starting point (t+k≤N)

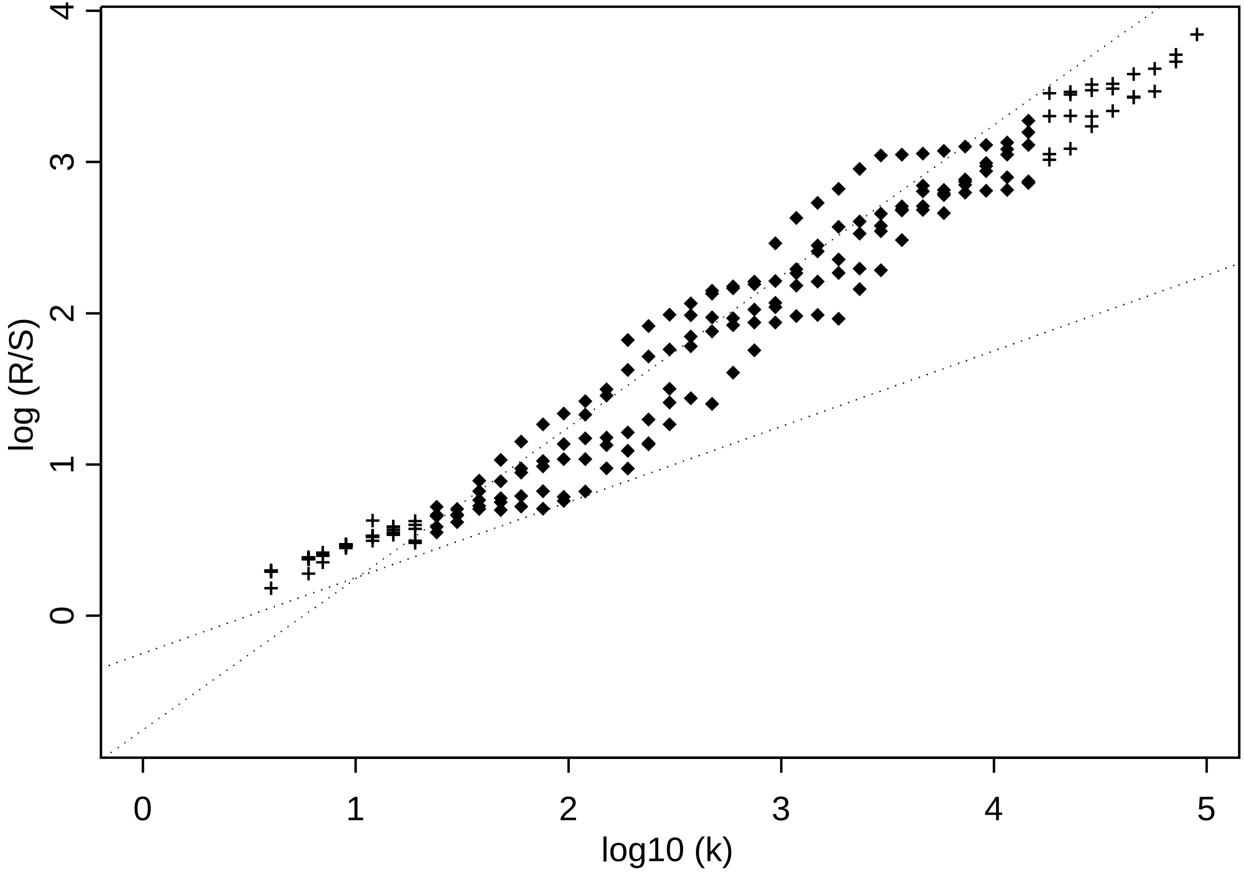
$$R(t,k) = \max_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right] - \min_{0\le i\le k}\left[Y_{t+i} - Y_t - \frac{i}{k}\left(Y_{t+k} - Y_t\right)\right]$$

$$S(t,k) = \sqrt{\frac{1}{k}\sum_{n=t+1}^{t+k}\left(X(n) - \overline{X}_{t,k}\right)^2} \qquad \overline{X}_{t,k} = \frac{1}{k}\sum_{n=t+1}^{t+k} X(n)$$

$$R/S(t,k) = \frac{R(t,k)}{S(t,k)}$$ rescaled adjusted range

$$\log \mathrm{E}\{R/S\} \approx a + H \log k$$

R/S plot: example

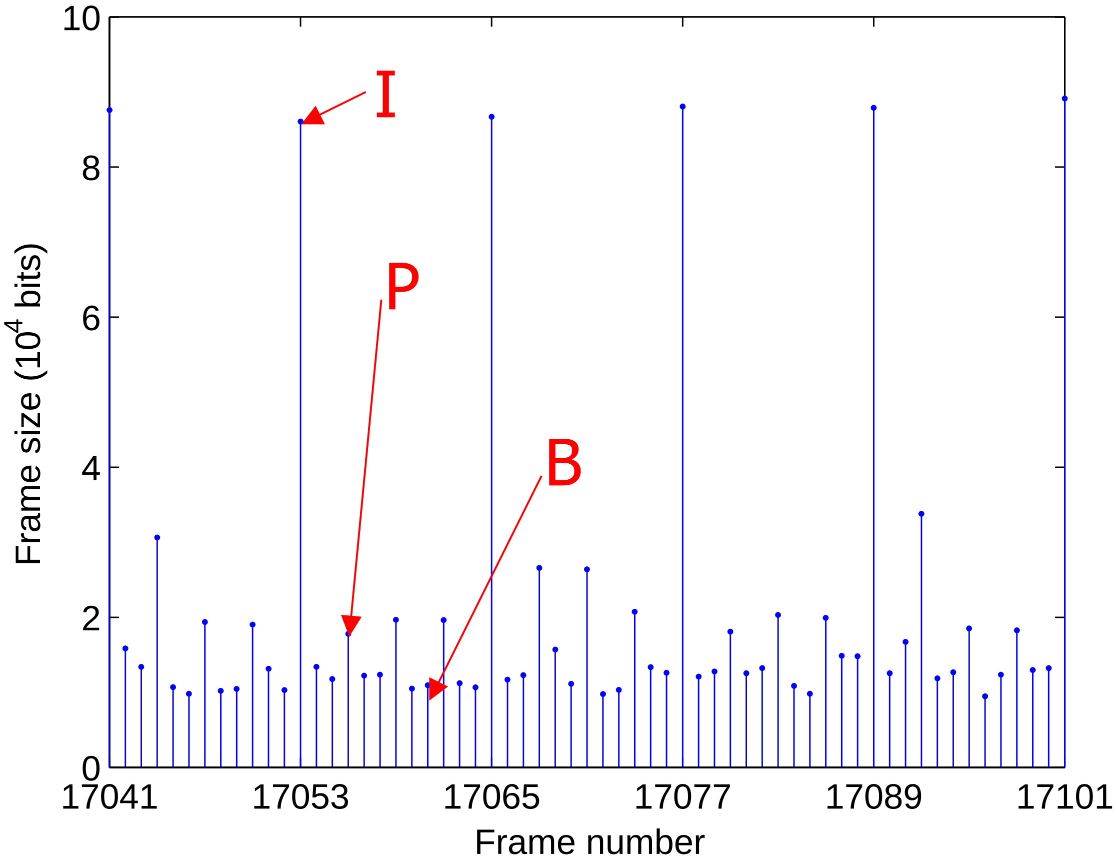


- Star Wars IV (MPEG-4)
- H=0.955

Structure of MPEG traces

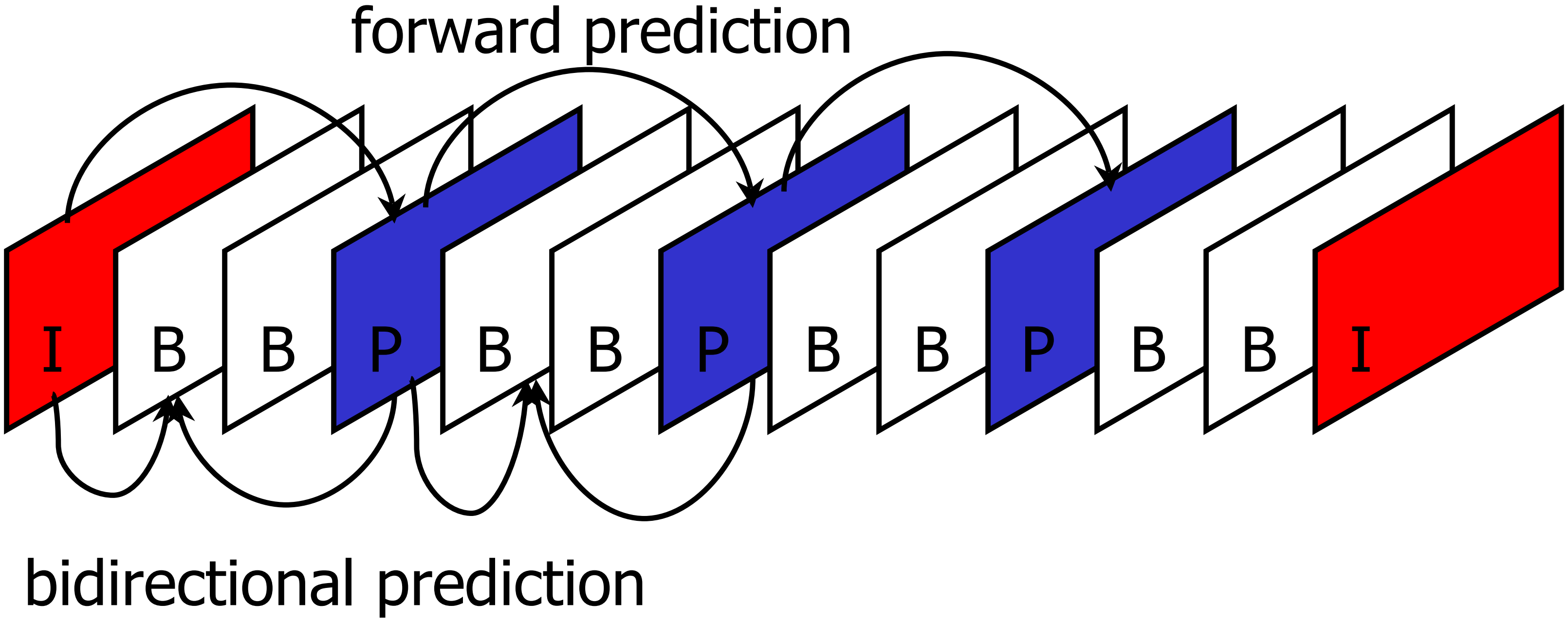
- 3 types of frames
  - I-frame: compressed original frame
  - P-frame: calculated by motion compensation with respect to the previous I- or P-frame
  - B-frame: calculated by motion compensation with respect to the previous I- or P- frame and the next I- or P- frame (interpolation between them)
- Group of pictures (GOP): arrangement of the frames in deterministic periodic sequence
  - IBBPBBPBBPBB (GOP length = 12)

MPEG video trace: example

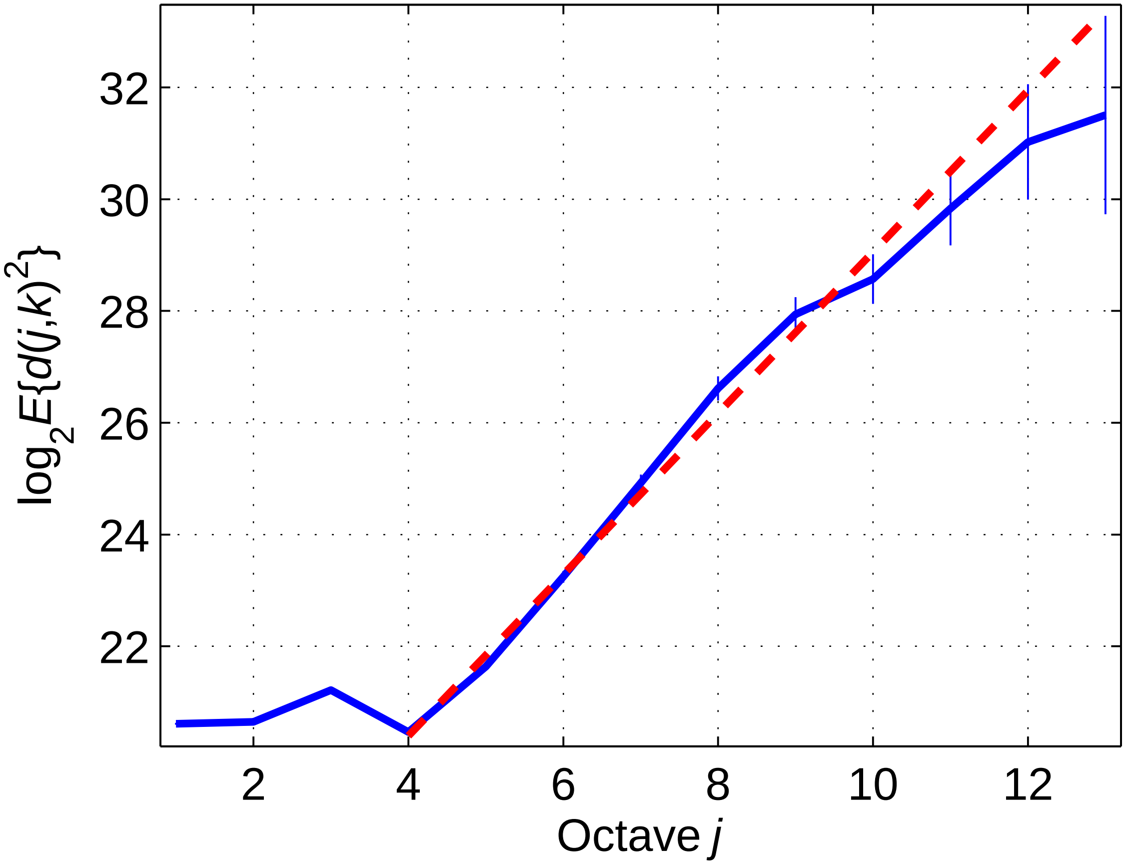


- Group of pictures: IBBPBBPBBPBB

GOP example


forward prediction
I
B
B
P
B
B
P
B
B
P
B
B
I
bidirectional prediction

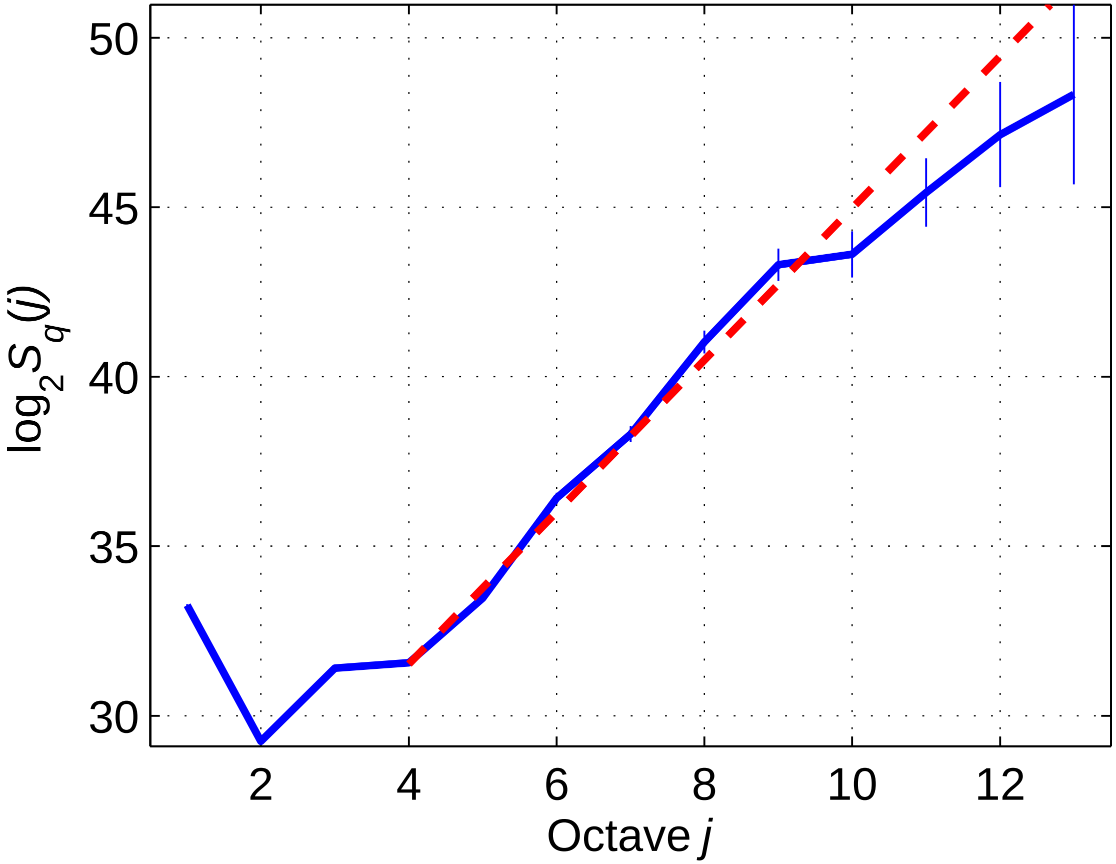
Logscale diagram: monofractal



- H=1.222, α=1.443 (octaves 4-13)
- R/S: H=0.973 (University of Berlin)

Jurassic park
(MPEG-4)

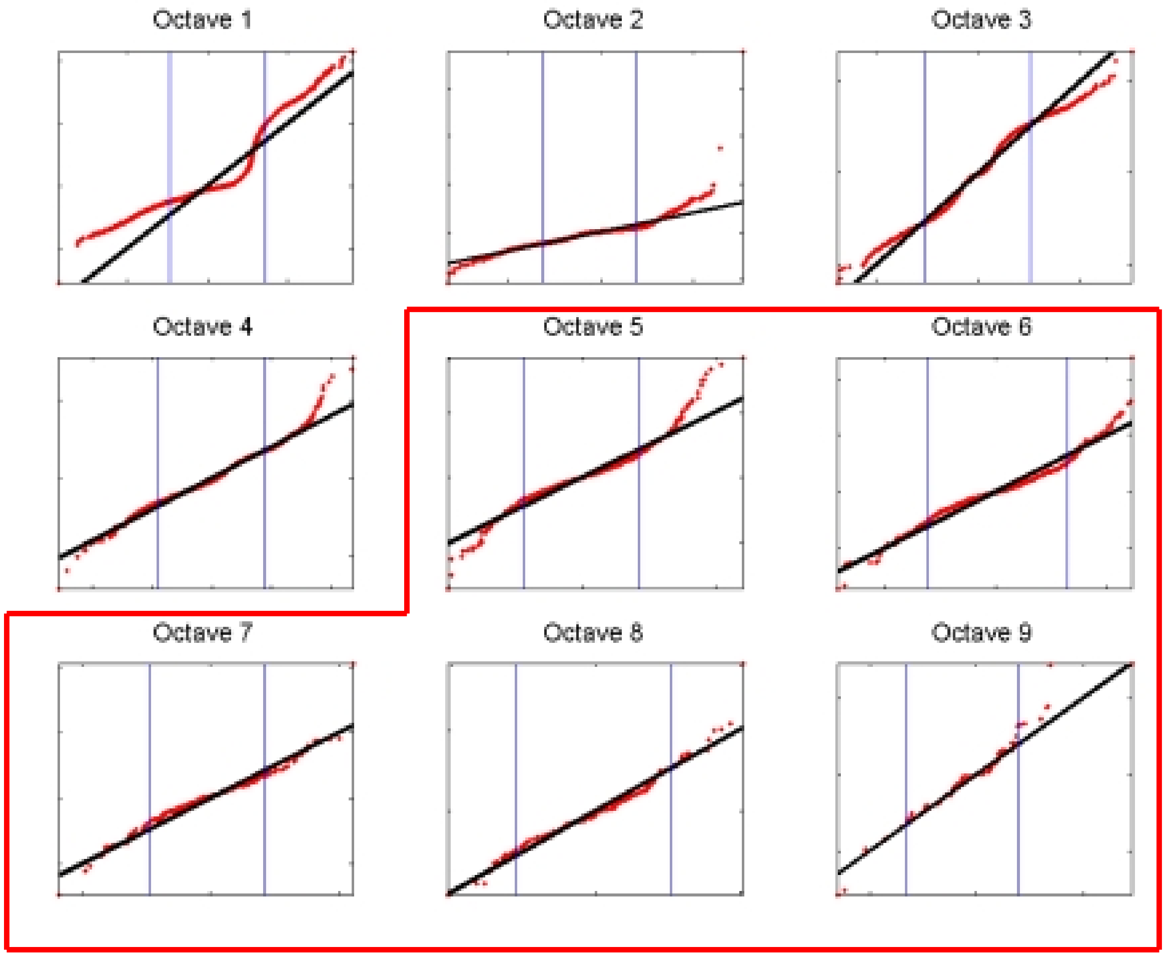
Logscale diagram: multifractal



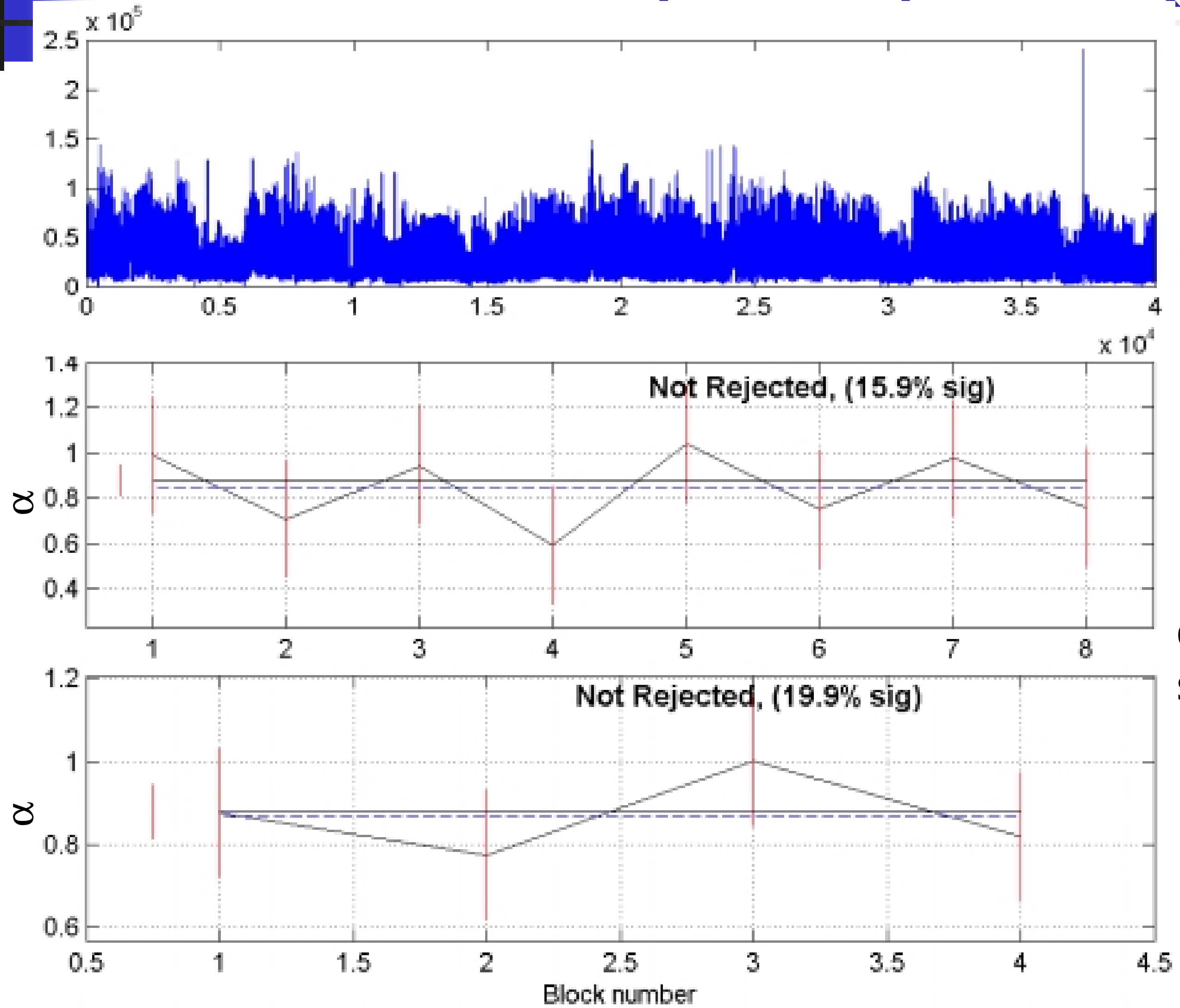
- H=1.247, α=2.241 (octaves 4-13)

Jurassic park
(MPEG-4)

qq-plots: Simpsons (MPEG-1)



Time constancy: Simpsons (MPEG-1)



dashed: average
solid: overall