Wavelets and scaling processes

Patrick Flandrin

CNRS & École Normale Supérieure de Lyon, France

thanks to Patrice Abry, Paulo Gonçalves, Darryl Veitch & Herwig Wendt

outline

1. Examples of “real life" scaling processes
2. Scaling concepts (self-similarity, fractality, long-range dependence, “$1/f$") and simple models (fractional Brownian motion, FARIMA, ON/OFF)
3. Wavelet basics
4. A wavelet-based framework for scaling processes (analysis and synthesis)
5. Extension to multifractal processes
6. References

random walk

Definition

Brownian motion is the zero-mean Gaussian process $\{X(t), t \geq 0; X(0) = 0\}$ *with increments* $\{\Delta_X(t,s) := X(t) - X(s), t \geq s\}$ *such that*

1. $\mathrm{var}\Delta_X(t, t-1) = 2D$, *for any t (stationarity)*
2. $\mathbb{E}\Delta_X(t_1, t_2)\Delta_X(t_3, t_4) = 0$, *if* $t_1 > t_2 \geq t_3 > t_4$ *(independence)*

Ordinary diffusion — We then have (Einstein, 1905):

$$\begin{aligned} \mathrm{var}X(t) &= \mathbb{E}[\Delta_X(t,t-1) + \Delta_X(t-1,t-2) + \ldots + \Delta_X(1,0)]^2 \\ &= t \times \mathrm{var}\Delta_X(1,0) \\ &= 2Dt \qquad \Rightarrow \qquad \underline{\sigma_X(t) := (\mathrm{var}X(t))^{\frac{1}{2}} \propto \sqrt{t}} \end{aligned}$$

credit: Leonardo da Vinci

disordered fluids

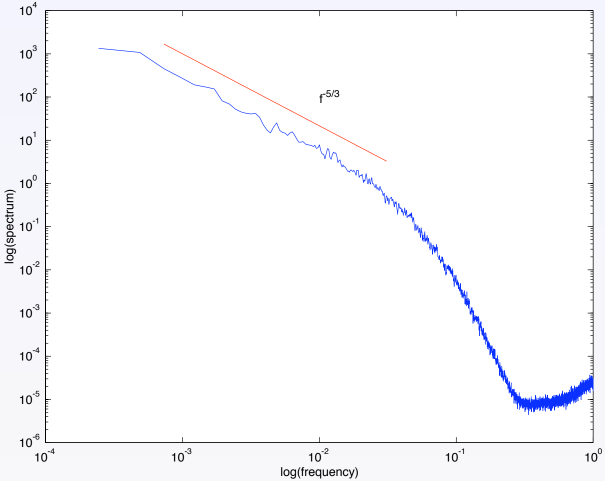
Prediction (Kolmogorov, ’41)

Assuming homogeneity, isotropy and a constant energy transfer rate in turbulent flows leads to the following (K41) predictions for the velocity field $v(x)$ and energy spectrum $E(k)$:

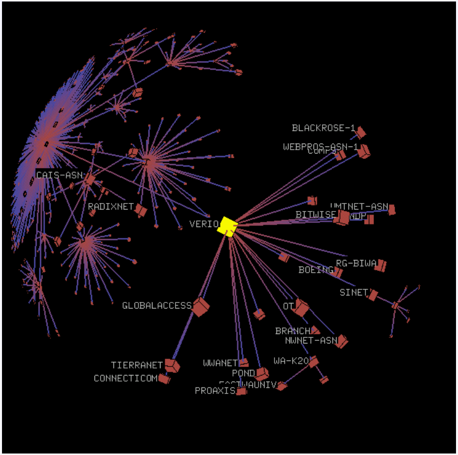
$$\langle |v(x+r) - v(x)|^q \rangle \propto r^{q/3} \quad \overset{(q=2)}{\Longrightarrow} \quad E(k) \propto k^{-5/3}$$

- *Experiments* — Observed *spectra* are “in $k^{-5/3}$", but *scaling laws* have *anomalous exponents* $\zeta_q \neq q/3$
- *Issues* — Reconsidering initial hypotheses, in terms of *statistics* (non-Gaussianity. . .) and *events* (coherent structures. . .)

velocity spectrum



complex systems



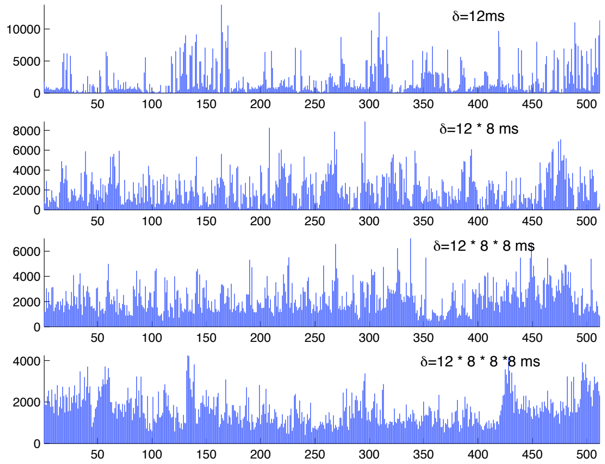
network traffic

Observation (Leland & *et al.*, '94)

Experimental evidence of scaling phenomena and long-range dependence in Ethernet traffic

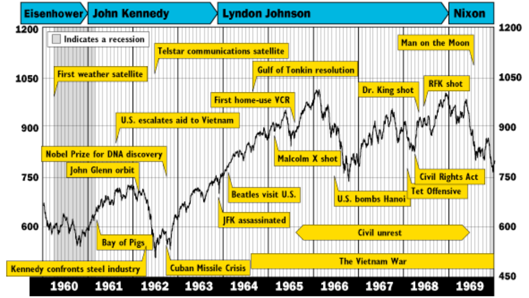
- *Extensions* — Similar behaviors observed in other types of traffic (Internet, VBR, WAN . . .).
- *Issues* — Traffic *control* (congestion, dimensioning, anomalies detection, prediction, protocols . . .)
  $\Rightarrow$ *modeling* and *analysis*.

IP Packets



credit: D. Veitch

Nature and finance!



scaling ubiquitous in many domains

- *Solid-state physics*
- *Hydrology*
- *Astrophysics*
- *Heart-rate variability*
- *Brain activity (fMRI)*
- *Earthquakes*
- *Social networks*
- *etc.*

scaling and power-laws

- *Power-law spectra* — Power-laws correspond to *homogeneous* functions:

$$\mathcal{S}(f) = C\,|f|^{-\alpha} \Rightarrow \mathcal{S}(kf) = C\,|kf|^{-\alpha} = k^{-\alpha}\mathcal{S}(f),$$

  for any $k > 0$

- *Fourier transform* — *Frequency* scaling carries over to the *time domain*. If we let $s(t) := (\mathcal{F}^{-1}\mathcal{S})(f)$, we get:

$$\int \mathcal{S}(kf)\, e^{i2\pi ft}\, df = k^{-1} \int \mathcal{S}(f')\, e^{i2\pi f'(t/k)}\, df' = s(t/k)/k$$

  It follows that $s(t/k) = s(t)/k^{\alpha-1} \Rightarrow$ *self-similarity*

credit: S. Kim

no characteristic scale

Definition

A process $\{X(t), t \in \mathbb{R}\}$ is said to be self-similar of index H (or "H-ss") if, for any $k > 0$,

$$\{X(kt), t \in \mathbb{R}\} \stackrel{d}{=} k^H \{X(t), t \in \mathbb{R}\}.$$

Interpretation

Any zoomed (in or out) version of an H-ss process looks (statistically) the same

Remark

If a process X is self-similar, it is necessarily nonstationary

zooming in on an *H*-ss process

credit: H. Wendt

stationary increments

Definition

A process $\{X(t), t \in \mathbb{R}\}$ is said to have stationary increments if and only if, for any $\theta \in \mathbb{R}$, the increment process:

$$\left\{X^{(\theta)}(t) := X(t+\theta) - X(t), t \in \mathbb{R}\right\}$$

has a distributional law which does not depend upon t

Remark

The concept of stationary increments can be naturally extended to higher orders ("increments of increments")

Definition

H-ss processes with stationary increments are referred to as "H-sssi" processes

covariance function of H-sssi processes

Theorem

The structure of the covariance function is the same for all H-sssi processes

Proof — Assuming that $X(t)$ is H-sssi, with $X(0) = 0$ and $X(1) \neq 0$, we have necessarily:

$$\begin{aligned}
\mathbb{E}X(t)X(s) &= \frac{1}{2}\left(\mathbb{E}X^2(t) + \mathbb{E}X^2(s) - \mathbb{E}\left(X(t) - X(s)\right)^2\right) \\
&= \frac{1}{2}\left(\mathbb{E}X^2(t) + \mathbb{E}X^2(s) - \mathbb{E}\left(X(t-s) - X(0)\right)^2\right) \\
&= \frac{\mathrm{var}X(1)}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)
\end{aligned}$$

asymptotic self-similarity

Definition

A stationary process $\{X(t), t \in \mathbb{R}\}$ is said to be asymptotically self-similar of index $\beta \in (0, 1)$ if

$$(\text{var}\, X(t))^{-1}\, \mathbb{E}X(t)X(t+\tau) \sim \tau^{-\beta}$$

when $\tau \to \infty$

- H-sssi processes are asymptotically self-similar of index $\beta = 2(1-H)$
- *non-summability* (and power-law decay) of the autocorrelation $\Rightarrow$ (power-law) *divergence* of the PSD at $f = 0$
- asymptotic self-similarity $\Rightarrow$ *long-range dependence* (LRD) (also referred to as *long memory*)

fractional Brownian motion

Definition 1

A process $B_H(t)$ is referred to as a fractional Brownian motion (fBm) of index $0 < H < 1$, if and only if it is H-sssi and Gaussian

- fBm has been introduced in (Mandelbrot & van Ness, '68), as an extension of the ordinary Brownian motion $B(t) \equiv B_H(t)|_{H=1/2}$ (*anomalous diffusion*)
- the index H is referred to as the *Hurst exponent*, and its limited range guarantees the *non-degeneracy* ($H < 1$) and the *mean-square continuity* ($H > 0$) of fBm

fractional Brownian motion

Definition 2

fBm admits the moving average representation:

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H + \frac{1}{2})} \left\{ \int_{-\infty}^{0} \left[(t - s)^{H - \frac{1}{2}} - (-s)^{H - \frac{1}{2}} \right] B(ds) + \int_{0}^{t} (t - s)^{H - \frac{1}{2}} B(ds) \right\}$$

- fBm results from a *"fractional integration"* of white noise
- no specific role attached to time $t = 0$

fractional Brownian motion

Definition 3

fBm admits the (harmonizable) spectral representation:

$$B_H(t) = C \int_{-\infty}^{+\infty} |f|^{-(H+\frac{1}{2})} \, (e^{i2\pi tf} - 1) \, W(df),$$

with $W(df)$ the Wiener measure

- the *"average spectrum"* of fBm behaves as $|f|^{-(2H+1)}$
- fBm is a widespread model for (nonstationary) Gaussian processes with a *power-law* (empirical) spectrum

fractional Gaussian noise

Definition

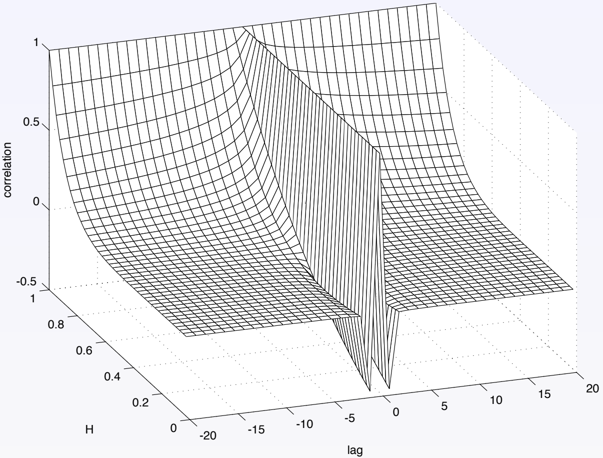
The (stationary) increment process $B_H^{(\theta)}(t)$ of fBm $B_H(t)$ is referred to as fractional Gaussian noise (fGn)

Autocorrelation — The (stationary) autocorrelation function of fGn, $c_H(\tau) := \mathbb{E} B_H^{(\theta)}(t) B_H^{(\theta)}(t+\tau)$, reads:

$$c_H(\tau) = \frac{\sigma^2}{2} \left(|\tau + \theta|^{2H} - 2|\tau|^{2H} + |\tau - \theta|^{2H} \right)$$

- if $\theta = 1$ and $H = \frac{1}{2}$, we have $c_H(k) = \sigma^2 \, \delta(k), k \in \mathbb{Z}$ (*discrete-time white noise*)
- for large lags τ, one has $c_H(\tau) \sim \sigma^2 \theta^2 H(2H-1)\tau^{2(H-1)}$ (subexponential, *power-law* decay)

fGn autocorrelation function



fGn spectrum

Result

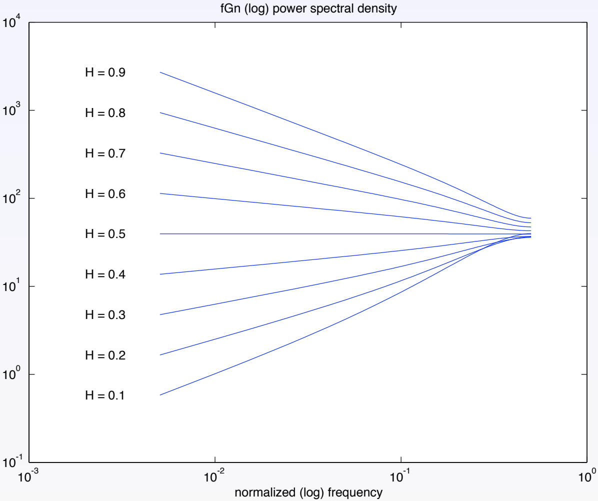
If $\theta = 1$, the power spectrum density of discrete-time fGn is given by:

$$\mathcal{S}(f) = C\,\sigma^2\,|e^{i2\pi f} - 1|^2 \sum_{k=-\infty}^{\infty} \frac{1}{|f+k|^{2H+1}},$$

with $-\frac{1}{2} \leq f \leq +\frac{1}{2}$

- if $H \neq \frac{1}{2}$, we have $\mathcal{S}(f) \sim C\,\sigma^2\,|f|^{1-2H}$ when $f \to 0$
- $0 < H < \frac{1}{2} \Rightarrow \mathcal{S}(0) = 0$
- $\frac{1}{2} < H < 1 \Rightarrow \mathcal{S}(0) = \infty$ (*spectral divergence*)

fGn power spectrum density



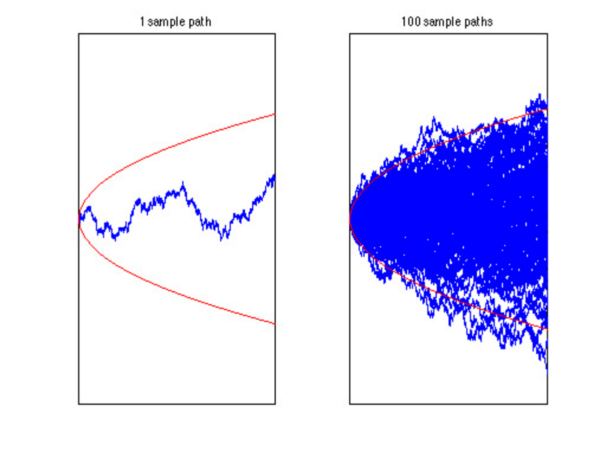
importance of fGn

1. *Aggregation* — Renormalized by T^{H-1}, fGn is *invariant* under the *recomposition rule*

$$X(n) \mapsto X^T(n) := T^{-1} \sum_{k=(n-1)T+1}^{nT} X(k)$$

2. *Attraction* — As $T \to \infty$, aggregating *any* asymptotically H-ss process ends up with a process whose covariance structure is that of fGn
3. *Long-range dependence* — fGn is *LRD* when $\frac{1}{2} < H < 1$

Bm sample paths



fBm sample paths 1

Definition

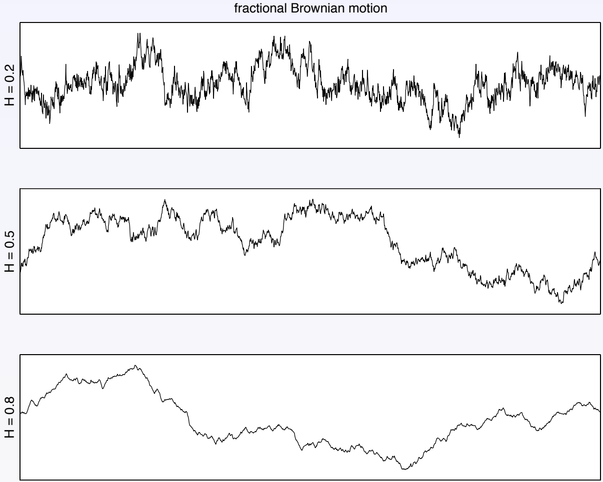
The local regularity of a signal $x(t)$ at a given point t_0 is measured by the Hölder exponent $h(t_0)$ defined as the supremum of α's such that $|x(t) - x(t_0)| < C\,|t - t_0|^{\alpha}$ when $|t - t_0| \to 0$ (the larger the exponent, the smoother the signal)

Result

For any (small enough) $\epsilon > 0$ and any $t \in \mathbb{R}$, we have $|B_H^{(\epsilon)}(t)| \leq C\,|\epsilon|^H$, with probability 1

- fBm is *everywhere continuous*, but *nowhere differentiable*
- sample paths have a *uniform* Hölder regularity $h = H$
- sample paths have a uniform (Haussdorf and box) *fractal dimension* $\dim_B \text{graph } B_H = 2 - H$

examples of fBm sample paths



correlation between increments

Interpretation

It follows from the covariance structure of fBm that, for any $t \in \mathbb{R}$,

$$C_H(\theta) := -\frac{\mathbb{E}B_H^{(-\theta)}(t)\,B_H^{(\theta)}(t)}{\mathrm{var}B_H^{(\pm\theta)}(t)} = 2^{2H-1} - 1$$

- $H = \frac{1}{2}$: no correlation (Brownian motion, $D = 1.5$)
- $H < \frac{1}{2}$: *negative* correlation (more erratic, $\lim_{H\to 0} D = 2$)
- $H > \frac{1}{2}$: *positive* correlation (less erratic, $\lim_{H\to 1} D = 1$)

Interpretation

H is a roughness measure of sample paths

"$1/f$" processes

Definition

A process is said to be of "$1/f$"-type if its empirical PSD behaves as $f^{-\alpha}$ ($\alpha > 0$) over some frequency range $[A, B]$

Depending on A and B, one can end up with:

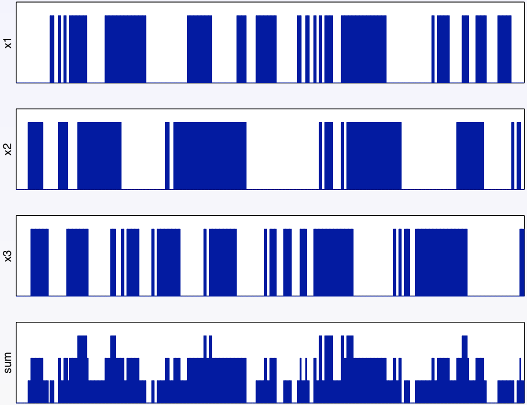
- *LRD*, if $A \to 0$ and $B < \infty$
- *scaling* in some "inertial range", if $0 < A < B < \infty$
- small-scale *fractality*, if $A < \infty$ and $B \to \infty$

Remark

In the fBm case, the only Hurst exponent H controls all 3 situations

ON/OFF Models 1

ON/OFF sources — Let $S_N(t)$ result from the superposition of N independent *ON/OFF* sources: $S_N(t) := \sum_{i=1}^{N} X_i(t)$



ON/OFF Models 2

Observation

Given N and T, the statistical properties of the cumulative process:

$$Y_N(Tt) := \int_0^{Tt} S_N(s)\, ds$$

depend on the pdf's of the ON and OFF periods τ

Theorem (Willinger *et al.*, '95)

Assuming that the ON and OFF periods are both heavy-tailed, i.e., that $\text{Prob}\{\tau > x\} \sim c\, x^{-\alpha}, \tau \to \infty$*, with* $1 < \alpha < 2$*, then (up to a linear trend)* $Y_N(Tt)$ *behaves asymptotically as fBm of Hurst exponent* $H = (3 - \alpha)/2$ *when N and T* $\to \infty$

FARIMA

Definition (Hosking, '81)

In the discrete-time case where $k \in \mathbb{Z}$, Fractional Auto-Regressive Integrated Moving Average (FARIMA) processes $X(k)$ are defined by

$$\Phi(z)(1 - z^{-1})^d X(k) = \Theta(z) W(k),$$

where z is the unit delay operator, Φ and Θ are polynomial of order p and q, respectively, $d \in]-1/2, +1/2[$ is the order of fractional differencing and $W(k)$ is white Gaussian noise

Interpretation

FARIMA(p, d, q) *processes generalize the classical random walk, which is a* FARIMA$(0, 1, 0)$

importance of FARIMA

1. FARIMA$(0, d, 0)$ — In the simplest case of a *fractionally integrated white noise*, the spectrum reads $\mathcal{S}(f) \propto |\sin \pi f|^{-2d}$ and

$$\mathcal{S}(f) \sim |f|^{-2d}, \quad f \to 0$$

2. *fGn-like* — At low frequencies, the spectrum is identical to that of fGn, with the identification $d \equiv H - 1/2$
3. *Long-range dependence* — FARIMA$(0, d, 0)$ processes are suitable models for *asymptotic self-similarity*, and *long-range dependence* when $d > 0$

evidencing scaling in data ?

Observation

- *Different and complementary signatures of scaling can be observed with respect to time (correlation, fractality,. . .) or frequency/scale (spectrum, zooming, . . .)*
- *Iterating aggregation reveals scale invariance*

Idea

- *Use explicitly an approach which combines time and frequency/scale*
- *Use explicitly a multiresolution approach*

$\Rightarrow$ *Wavelets* !

evidencing scaling in data ?

Observation

- *Different and complementary signatures of scaling can be observed with respect to time (correlation, fractality,. . .) or frequency/scale (spectrum, zooming, . . .)*
- *Iterating aggregation reveals scale invariance*

Idea

- *Use explicitly an approach which combines time and frequency/scale*
- *Use explicitly a multiresolution approach*

$\Rightarrow$ *Wavelets* !

rationale

Idea

"signal $=$ *(low-pass) approximation* $+$ *(high-pass) detail"*

$+$

iteration

- successive approximations (at coarser and coarser resolutions) $\sim$ *aggregated data*
- details (information differences between successive resolutions) $\sim$ *increments*

Multiresolution is a natural language for scaling processes

formalization

Definition (Mallat & Meyer, '86)

A MultiResolution Analysis (MRA) of $L^2(\mathbb{R})$ is given by:

1. *a hierarchical sequence of embedded approximation spaces $\ldots V_1 \subset V_0 \subset V_{-1} \ldots$, whose intersection is empty and whose closure is dense in $L^2(\mathbb{R})$*
2. *a dyadic two-scale relation between successive approximations*

$$X(t) \in V_j \Leftrightarrow X(2t) \in V_{j-1}$$

3. *a scaling function $\varphi(t)$ such that all of its integer translates $\{\varphi(t-n), n \in \mathbb{Z}\}$ form a basis of V_0*

multiresolution expansion

Definition

The wavelet $\psi(.)$ is constructed in such a way that all of its integer translates form a basis of W_0, defined as the complement of V_0 in V_{-1}

For a given resolution depth J, any signal $X(t) \in V_0$ can be expanded as :

$$\underbrace{X(t)}_{\text{signal}} = \underbrace{\sum_k a_X(J,k)\, \varphi_{J,k}(t)}_{\text{approximation}} + \underbrace{\sum_{j=1}^{J}}_{J\text{ octaves}} \underbrace{\sum_k \overbrace{d_X(j,k)}^{\text{wav. coeffs.}}\ \psi_{j,k}(t)}_{\text{details}},$$

with $\{\xi_{j,k}(t) := 2^{-j/2}\,\zeta(2^{-j}t-k), j \text{ and } k \in \mathbb{Z}\}$, for $\xi = \varphi$ and ψ

a simple construction

Example (Haar, 1911)

The simplest choice for a MRA is given by the Haar basis, attached to the scaling function $\varphi(t) = \chi_{[0,1]}(t)$ *and the wavelet* $\psi(t) = \chi_{[0,1/2]}(t) - \chi_{[1/2,1]}(t)$

Remark

When aggregated over dyadic intervals, data samples identify to Haar approximants

Interpretation

Wavelet analysis offers a refined way of both aggregating data and computing increments

two key properties

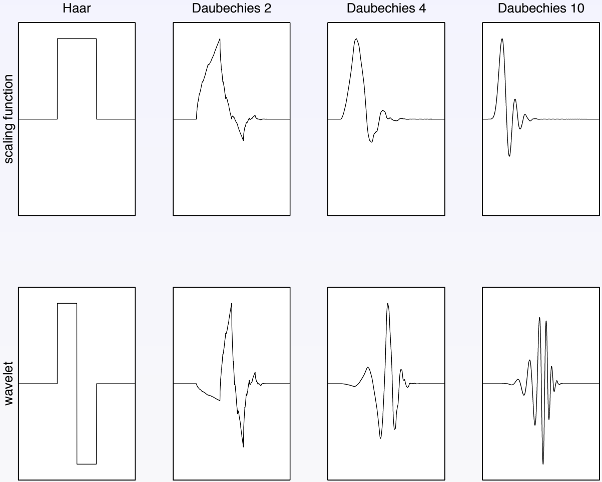
1. *Admissibility* — By construction, a scaling function (resp., a wavelet) is a low-pass (resp., high-pass) function $\Rightarrow$ an *admissible* wavelet $\psi(t)$ is necessarily *zero-mean*:

$$\Psi(0) := \int_{-\infty}^{+\infty} \psi(t)\, dt = 0$$

2. *Cancellation* — A further key property for a wavelet is the number of its *vanishing moments*, i.e., the integer $N \geq 1$ such that

$$\int_{-\infty}^{+\infty} t^k \, \psi(t)\, dt = 0, \ \text{for} \ k = 0, 1, \dots N - 1$$

the example of Daubechies wavelets



effective computation

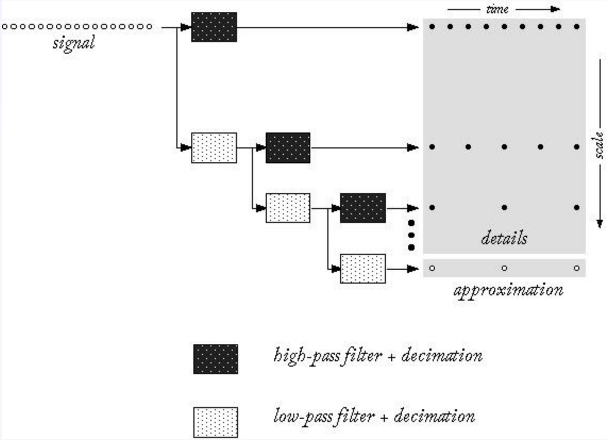
- *Theory* — The *wavelet coefficients* $d_X(j,k)$ are given by the inner products:

$$d_X(j,k) := \langle X, \psi_{j,k} \rangle$$

- *Practice* — They can rather be computed in a *recursive* fashion, via efficient *pyramidal algorithms* (faster than FFT)
  - *no need* for knowing explicitly $\psi(t)$
  - enough to characterize a wavelet by its *filter coefficients* $\{g(n) := (-1)^n \, h(1-n), n \in \mathbb{Z}\}$, with

$$h(n) := \sqrt{2} \int_{-\infty}^{+\infty} \varphi(t) \, \varphi(2t-n) \, dt$$

Mallat's algorithm



wavelets as filters

- *Input-output* — Given the statistics of the analyzed signal, statistics of its wavelet coefficients can be derived from imput-ouput relationships of *linear filters*
- *Stationary processes* — In the case of *stationary processes* with autocorrelation $\gamma_X(\tau) := \mathbb{E}X(t)X(t+\tau)$, stationarity carries over to wavelet sequences:

$$C_X(j,n) := \mathbb{E}d_X(j,k)d_X(j,k+n) = \int_{-\infty}^{+\infty} \gamma_X(\tau)\,\gamma_\psi(2^{-j}\tau+n)\,d\tau$$

$$\sum_{n=-\infty}^{\infty} C_X(j,n)\,e^{-i2\pi fn} = \Gamma_X(2^{-j}f)\,\times\,\sum_{n=-\infty}^{\infty} \gamma_\psi(n)\,e^{-i2\pi fn}$$

wavelets as stationarizers 1

Theorem (F., '89 & '92)

Wavelet admissibility ($N \geq 1$) guarantees that, if $X(t)$ has stationary increments, then $d_X(j,k)$ is stationary in k, for any given scale 2^j

Proof — Assuming that $X(t)$ is a s.i. process with $X(0) = 0$ and $\mathrm{var} X(t) := \rho(t)$, we have

$$\begin{aligned} \mathbb{E}X(t)X(s) &= \frac{1}{2}\left(\mathbb{E}X^2(t) + \mathbb{E}X^2(s) - \mathbb{E}\left(X(t) - X(s)\right)^2\right) \\ &= \frac{1}{2}\left(\mathbb{E}X^2(t) + \mathbb{E}X^2(s) - \mathbb{E}\left(X(t-s) - X(0)\right)^2\right) \\ &= \frac{1}{2}\left(\rho(t) + \rho(s) - \rho(t-s)\right) \end{aligned}$$

.../...

wavelets as stationarizers 2

and it follows that

$$
\begin{aligned}
\mathbb{E}d_X(j,n)d_X(j,m) &= \iint \mathbb{E}X(t)X(s)\,\psi_{jn}(t)\,\psi_{jm}(s)\,dt\,ds \\
&= \frac{1}{2}\int \rho(t)\,\psi_{jn}(t)\underbrace{\left(\int \psi_{jm}(s)\,ds\right)}_{=0}dt \\
&\quad +\frac{1}{2}\int \rho(s)\,\psi_{jm}(s)\underbrace{\left(\int \psi_{jn}(t)\,dt\right)}_{=0}ds \\
&\quad -\frac{1}{2}\iint \rho(t-s)\,\psi_{jn}(t)\,\psi_{jm}(s)\,dt\,ds \\
&= -\frac{1}{2}\int \rho(\tau)\,\gamma_\psi(2^{-j}\tau-(n-m))\,d\tau
\end{aligned}
$$

wavelets as stationarizers 3

- *Extension* — Stationarization can be extended to processes with stationary increments of *order* $p > 1$, under the vanishing moments condition $N \geq p$
- *Application* — Stationarization applies to H-sssi processes (e.g., fBm), with $\rho(t) = |t|^{2H}$

Remark

Nonstationarity is contained in the approximation sequence

wavelets and scale invariance

- *Self-similarity* — The *multiresolution* nature of wavelet analysis guarantees that, if $X(t)$ is H-ss, then

$$\{d_X(j,k), k \in \mathbb{Z}\} \stackrel{d}{=} 2^{j(H+1/2)} \{d_X(0,k), k \in \mathbb{Z}\}$$

  for any $j \in \mathbb{Z}$
- *Spectral interpretation* — Given a "$1/f$" process, the wavelet *tuning condition* $N > (\alpha - 1)/2$ guarantees that

$$\mathcal{S}_X(f) \propto |f|^{-\alpha} \Rightarrow \mathbb{E}d_X^2(j,k) \propto 2^{j\alpha}$$

wavelets as decorrelators 1

Theorem (F., '92; Tewfik & Kim, '92)

In the case where $X(t)$ is H-sssi, the condition $N > H + 1/2$ guarantees that

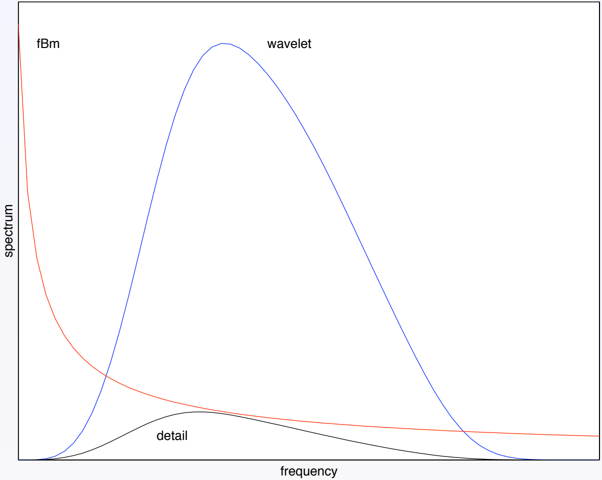
$$\mathbb{E} d_X(j,k) d_X(j,k+n) \sim n^{2(H-N)}, \;\; n \to \infty$$

Interpretation

Competition, at $f = 0$, between the (divergent) spectrum of the process and the (vanishing) transfer function of the wavelet:

$$\mathbb{E} d_X(j,k) d_X(j,k+n) \propto \int_{-\infty}^{+\infty} \frac{|\Psi(2^j f)|^2}{|f|^{2H+1}} \, e^{i2\pi nf} \, df$$

LRD and vanishing moments



wavelets as decorrelators 2

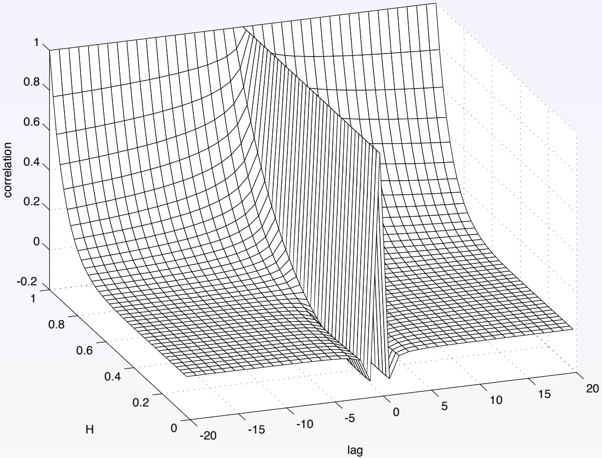
- *Consequence* — *Long-range dependence* (LRD) of a process X can be transformed into *short-range dependence* (SRD) in the space of its wavelet coefficients $d_X(j,.)$, provided that the number N of the vanishing moments is high enough

Remark

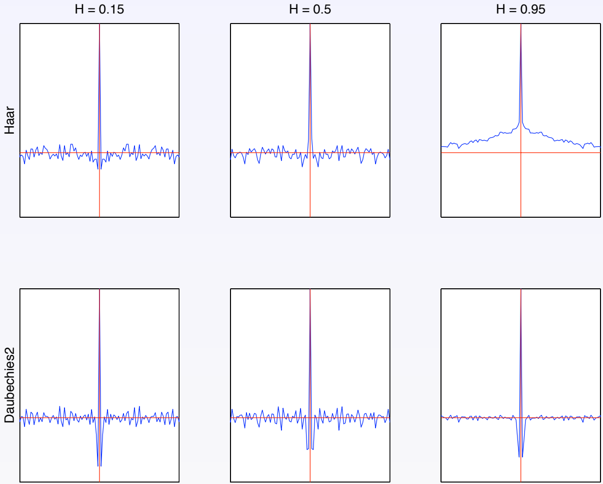
Residual LRD in the approximation sequence

- *The case of H-sssi processes* — LRD when $H > 1/2 \Rightarrow$ wavelet SRD needs $N > 1 \Rightarrow$ *Haar not suitable*

wavelet correlation of fBm in the Haar case



wavelet correlation and vanishing moments



multiresolution building blocks

- *Brownian motion* — The representation

$$B(t) = \int_0^t W(ds),$$

where $W(.)$ is white Gaussian noise, can be equivalently expressed as

$$B(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_W(j,k) \left(2^{-j/2} \int_{2^j k}^{2^j (k+t)} \psi(s)\, ds \right)$$

with *uncorrelated* details (*Karhunen-Loève expansion*).

- *Haar* — In the *Haar* case, this is *Lévy's construction* (1954) from the *Schauder* basis

approximate construction

- *Almost K-L expansion* — The *quasi-decorrelation* property of wavelets (with enough vanishing moments) suggests to approximate fBm by (Wornell, '90):

$$\tilde{B}_H(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \epsilon(j,k)\, \psi_{jk}(t),$$

  with *uncorrelated* weights such that $\mathrm{var}\,\epsilon(j,k) = \sigma^2\, 2^{j(2H+1)}$

- *Spectrum* — One can, however, only guarantee that

$$C_1\, |f|^{-(2H+1)} \leq \mathcal{S}_{\tilde{B}_H}(f) \leq C_2\, |f|^{-(2H+1)}$$

exact construction

- *Exact expansion* — Taking into account *low-pass contributions* leads to (Sellan *et al.*, '00):

$$B_H(t) = \sum_{k=-\infty}^{\infty} S(k)\phi^{(H)}(t-k) + \sum_{j=-\infty}^{0} \sum_{k=-\infty}^{\infty} \epsilon(j,k)\psi_{jk}^{(H)}(t) - b_0,$$

  with $S(k)$ a sum of k FARIMA$(0, H-1/2, 0)$ processes, $\epsilon(j,k)$ uncorrelated Gaussian variables with $\mathrm{var}\,\epsilon(j,k) = \sigma^2\, 2^{j(2H+1)}$, and b_0 a correcting term ensuring that $B_H(0) = 0$
- *Basis functions* — Both $\phi^{(H)}(t)$ and $\psi^{(H)}(t)$ have to be specifically designed for a given H

assessing scaling via "log-scale" diagrams 1

Idea

Given the variance $v_X(j) := \mathbb{E} d_X^2(j,k)$, scale invariance is revealed by the linear relation :

$$\log_2 \, v_X(j) = \alpha \, j + \text{Const.}$$

From theory to practice — The further properties of 1) *stationarization* and 2) *quasi-decorrelation* suggest to use as estimator of $v_X(j)$ the *empirical variance*

$$\hat{v}_X(j) := \frac{1}{N_j} \sum_{k=1}^{N_j} d_X^2(j,k),$$

where N_0 stands for the data size, and $N_j := 2^{-j} N_0$

assessing scaling via "log-scale" diagrams 2

- *Bias correction* — Given that "$\log \mathbb{E} \neq \mathbb{E} \log$", the effective estimator is $y_X(j) := \log_2 \hat{v}_X(j) - g(j)$, with

$$g(j) = \psi(N_j/2)/\log 2 - \log_2(N_j/2)$$

  and $\psi(.)$ the derivative of the Gamma function, so that $\mathbb{E} y_X(j) = \alpha j +$ Const. in the *uncorrelated* case
- *Variance* — Assuming *stationarization* and *quasi-decorrelation* guarantees furthermore that

$$\sigma_j^2 := \mathrm{var} y_X(j) = \zeta(2, N_j/2)/\log^2 2,$$

  where $\zeta(z, \nu)$ is a generalized Riemann function

scaling exponent estimation

- *From $y_X(j)$ to $\widehat{\alpha}$* — The slope α is estimated via a *weighted linear regression* in a log-log diagram:

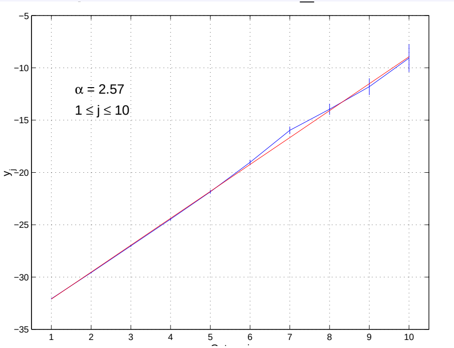
$$\widehat{\alpha} = \sum_{j=j_{\min}}^{j_{\max}} \frac{S_0\, j - S_1}{S_0\, S_2 - S_1^2}\, \frac{1}{\sigma_j^2}\, y_X(j),$$

with $S_k := \sum_j k/\sigma_j^2$, $k = 0, 1, 2$

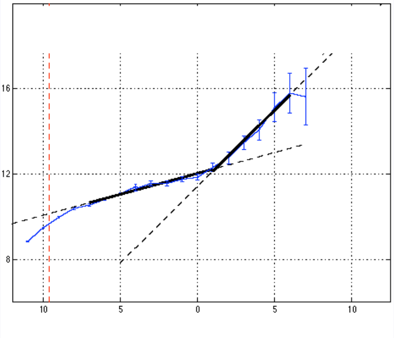
- *Bias and variance* — We have $\mathbb{E}\widehat{\alpha} \equiv \alpha$, by construction. Assuming *Gaussianity*, the estimator is moreover *asymptotically efficient* in the limit $N_j \to \infty$ (for any j), with

$$\mathrm{var}\widehat{\alpha} \sim 1/N_0$$

examples of log-scale diagrams



fBm



internet traffic

robustness

- *Detrending* — The *vanishing moments* condition

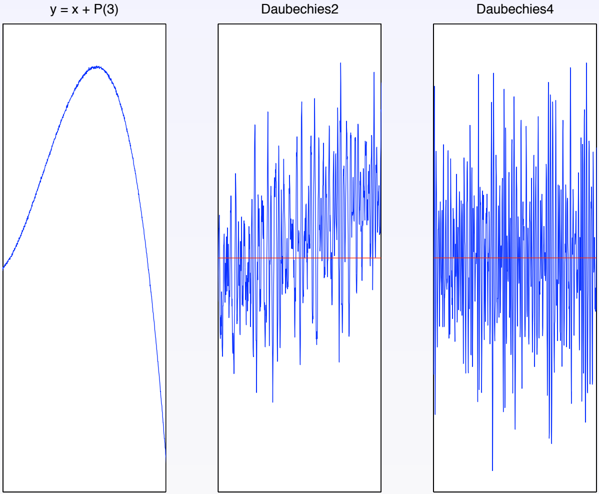
$$\int_{-\infty}^{+\infty} t^k \, \psi(t) \, dt = 0, \text{ for } k = 0, 1, \ldots N-1,$$

guarantees that $d_T(j, n) \equiv 0$ for *any* $T(t)$ of the form

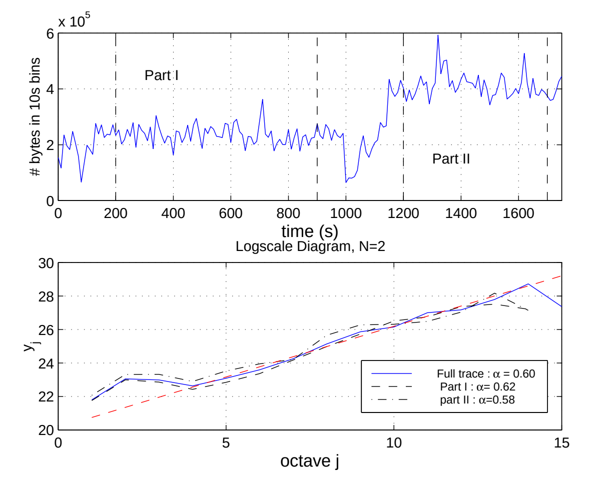
$$T(t) = \sum_{k=0}^{N-1} a_k \, t^k$$

In other words, a wavelet with enough vanishing moments makes the transform of $Z(t) := X(t) + T(t)$ *blind* to a superimposed polynomial trend

robustness to polynomial trends



robustness to “jumps”



related techniques aimed at scaling processes 1

- *Aggregation* — Wavelets offer a natural *generalization* to aggregation: Haar approximants $\mapsto$ Haar details $\mapsto$ wavelet details with higher N
- *Variogram* — Wavelets generalize as well *variogram* techniques (Matheron, 1967), which are based on the increment property $\mathbb{E}(X(t+\tau)-X(t))^2=\sigma^2|\tau|^{2H}$, since increments can be viewed as constructed on the “poorman’s wavelet":

$$\psi(t) := \delta(t+\tau)-\delta(t)$$

- *Structure functions* — Same relation to the generalization of the variogram beyond second order

related techniques aimed at scaling processes 2

- *Allan variance* — A refined notion of variance — introduced in the study of atomic clocks stability — is the so-called *Allan variance* (Allan, '66), defined by

$$\mathrm{var}_X^{(\mathrm{Allan})}(T) := \frac{1}{2T^2}\,\mathbb{E}\left[\int_{t-T}^{t} X(s)\,ds - \int_{t}^{t+T} X(s)\,ds\right]^2$$

  - in the case of H-ss processes, Allan variance is such that $\mathrm{var}_X^{(\mathrm{Allan})}(T) \sim T^{2H}$ when $T \to \infty$
  - when evaluated over dyadic intervals, Allan variance identifies to the variance of Haar details:

$$\mathrm{var}_X^{(\mathrm{Allan})}(2^j) = \mathrm{var} d_X^{(\mathrm{Haar})}(j,k)$$

related techniques aimed at scaling processes 3

- *Fano factor* — In the case of a *Poisson process* $P(t)$ of counting process $N(.)$, one can define the *Fano factor* as:

$$F(T) := \text{var} N(T)/\mathbb{E}N(T)$$

  - for a uniform density λ, we have $F(T) = 1$ for any T whereas, for a “fractal" density $\lambda(t) = \lambda + B_H^{(\theta)}(t)$, we have $F(T) \sim T^{2H-1}$ when $T \to \infty$
  - interpretation as *fluctuations/average* suggests the *wavelet generalization* given by:

$$F(T) \mapsto F_W(j) := 2^{j/2} \, \text{var} d_P(j,k)/\mathbb{E}a_P(j,k) \sim 2^{j(2H-1)}$$

when $j \to \infty$, and $F_W^{(\text{Haar})}(j) \equiv F^{(\text{Allan})}(2^j)$

beyond second order

Idea

Given the renormalized definition $T_X(a) := 2^{-j/2}\, d_X(j,n)\big|_{j=\log_2 a}$, *one can consider scaling laws which generalize second order behaviors:*

$$
\begin{array}{rcll}
\mathbb{E}|T_X(a)|^q \propto a^{Hq} & = & \exp\{Hq \log a\} & (\text{"monoscaling"}) \\
& & \downarrow & \\
& & \exp\{H(q) \log a\} & (\text{"multiscaling"}) \\
& & \downarrow & \\
& & \exp\{H(q)\, n(a)\} & (\text{"cascade"})
\end{array}
$$

beyond second order

Idea

Given the renormalized definition $T_X(a) := 2^{-j/2}\, d_X(j,n)\big|_{j=\log_2 a}$, *one can consider scaling laws which generalize second order behaviors:*

$$\mathbb{E}|T_X(a)|^q \propto a^{H(q)} = \exp\{H(q)\log a\} \quad (\text{“multiscaling”})$$

back to wavelets and regularity

Definition

The local regularity of a signal $x(t)$ at a given point t_0 is measured by the Hölder exponent $h(t_0)$ defined as the supremum of α's such that $|x(t) - x(t_0)| < C\,|t - t_0|^\alpha$ when $|t - t_0| \to 0$ (the larger the exponent, the smoother the signal)

$$W_x(a,t) := \frac{1}{\sqrt{a}} \int x(s)\, \psi\left(\frac{s-t}{a}\right) ds$$

Theorem (Jaffard, '88)

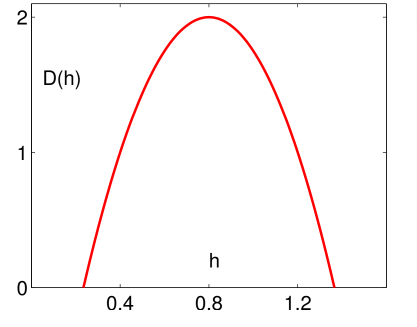
$$h(t_0) = \liminf_{a \to 0,\, t \to t_0} \frac{\log |W_x(a,t)|}{\log(a + |t - t_0|)}$$

multifractality

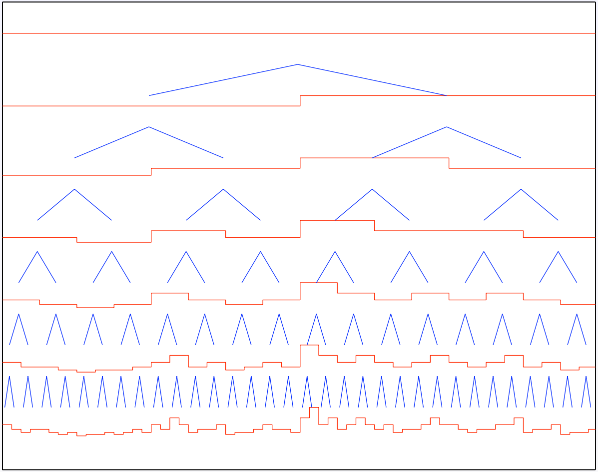
- some processes (e.g., fBm) are such that $h(t) = h$ for any t, but other ones exhibit *fluctuating* $h(t)$'s
- global characterization for $h(t)$ based on $E(h)$, the set of points t_h with the same Hölder exponent (i.e., such that $h(t_h) = h$)

Definition

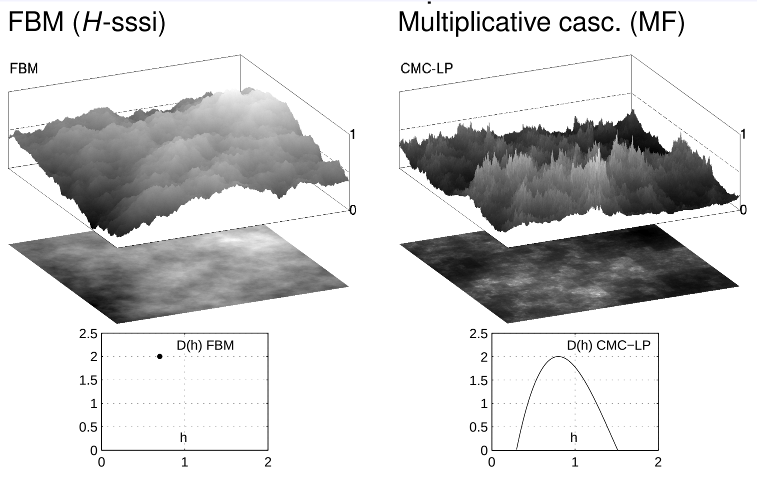
The multifractal spectrum is given by $D(h) = \dim_{\mathrm{H}} E(h)$*, where* $\dim_{\mathrm{H}}$ *stands for the Hausdorff dimension*



multiplicative cascades as an example



monofractal vs. multifractal



credit: P. Abry & H. Wendt

multifractal formalism (Parisi & Frisch, '85)

Given an observation $x(t) \in \mathbb{R}^d$:

1. start from some *multiresolution* quantity $M_x(a, t)$ (e.g., increments, aggregation, wavelet details, ...)
2. compute *structure functions*

$$S_x(a, q) = \langle |M_x(a, t)|^q \rangle_t \sim a^{\zeta(q)},\, a \to 0_+$$

3. estimate the *scaling exponents* as

$$\zeta(q) = \liminf_{a \to 0_+} \frac{\log S_x(a, q)}{\log a}$$

4. deduce the spectrum from a *Legendre transform*

$$D(h) = \min_{q \neq 0} \{d + qh - \zeta(q)\}$$

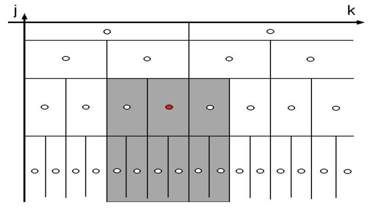
multifractal formalism and wavelets

Problem

1. *instability for $q \leq 0$*
2. *Legendre transform not valid in general*

Way out (Jaffard, '04)

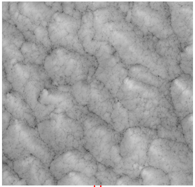
make use of hierarchical multiresolution quantities $\Rightarrow$ wavelet leaders



credit: P. Abry

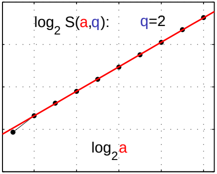
multifractal formalism and wavelet leaders

$X(t) \rightarrow d_X(a,t) \rightarrow L_X(a,t)$



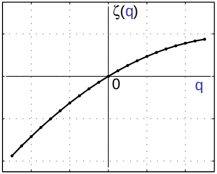
$\Downarrow$

$S(a,q) = \frac{1}{n_a} \sum_{k=1}^{n_a} |L_X(a,k)|^q$



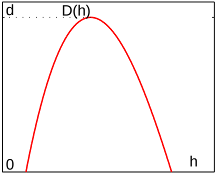
$\Longrightarrow$

$S(a,q) \simeq c_q a^{\zeta(q)}, \quad a \rightarrow 0$



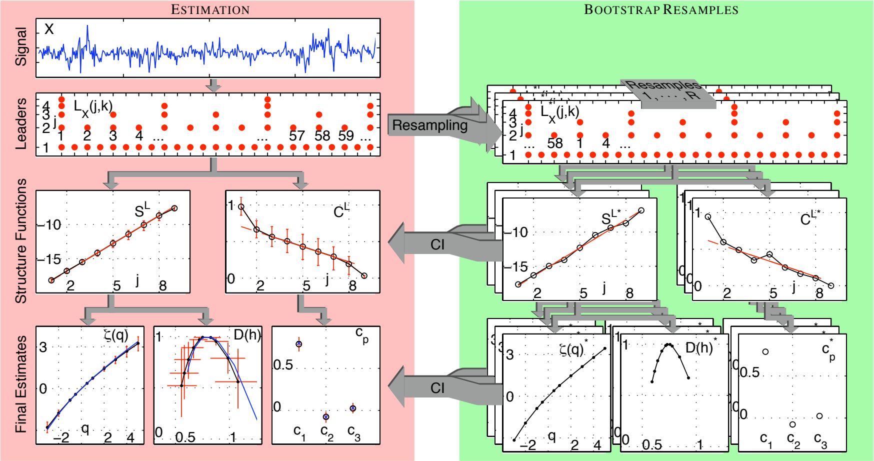
$\Uparrow$

$D(h) = \min_{q \neq 0} (d + qh - \zeta(q))$



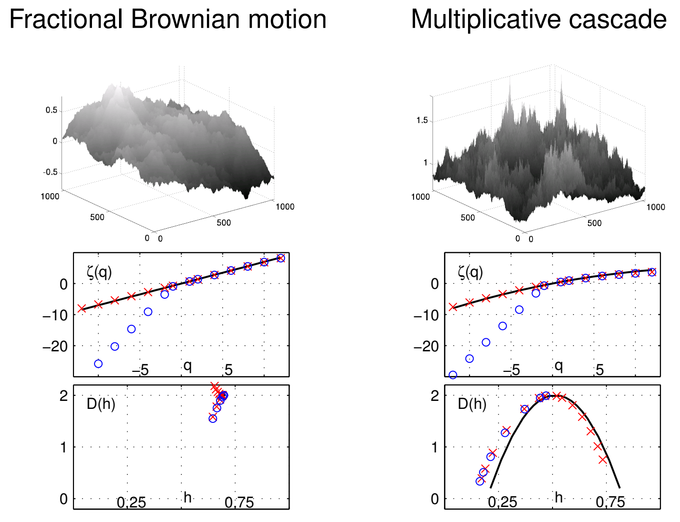
credit: P. Abry

bootstrap



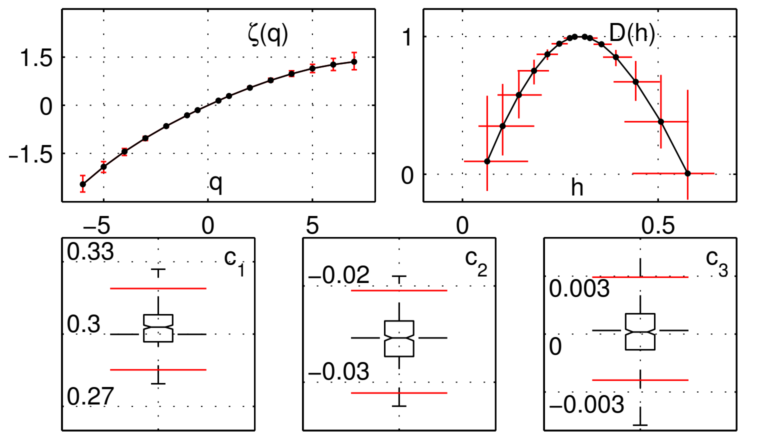
credit: P. Abry & H. Wendt

monofractal vs. multifractal



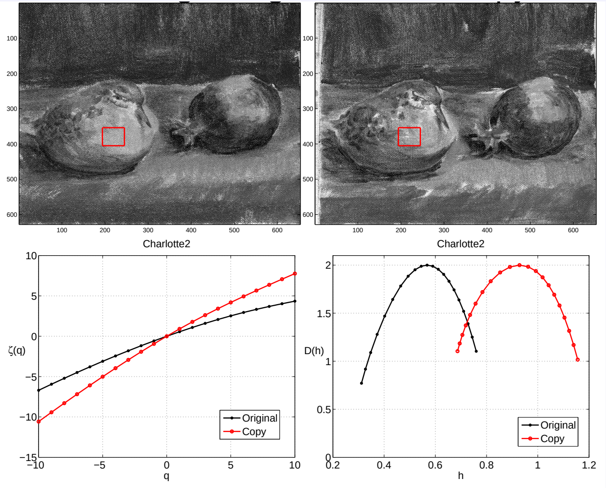
credit: P. Abry & H. Wendt

fully developed turbulence



credit: P. Abry & H. Wendt

VanGogh: original vs. copy



credit: P. Abry & H. Wendt

contact, (p)reprints & Matlab codes

- patrick.flandrin@ens-lyon.fr
- http://perso.ens-lyon.fr/patrick.flandrin/
- http://perso.ens-lyon.fr/patrice.abry/