# Wavelet Modeling: All That Scaling

Brani Vidaković

ISyE and BME @ Georgia Tech, Atlanta, USA

February 3, 2017 Mathematics & Statistics, USF, Tampa

# Overview

- Examples of Scaling.

# Overview

- Examples of Scaling.

- Why Scaling in Data is Important?

# Overview

- Examples of Scaling.
- Why Scaling in Data is Important?
- Measures of Scaling – Traditional and Wavelet Based

# Overview

- Examples of Scaling.

- Why Scaling in Data is Important?

- Measures of Scaling – Traditional and Wavelet Based

- Some New Results: Wavelet Transforms of Images, Robust Measures of Scaling, Scale-mixing Non-decimated and Packet Transforms.

# Overview

- Examples of Scaling.

- Why Scaling in Data is Important?

- Measures of Scaling – Traditional and Wavelet Based

- Some New Results: Wavelet Transforms of Images, Robust Measures of Scaling, Scale-mixing Non-decimated and Packet Transforms.

- Some Applications: Mammograms, DNA, LC-MS

# Scaling: It Started with Hurst and Nile Data

- Harold Edwin Hurst was a poor Leicester boy who made good, eventually working his way into Oxford, and later became a British "civil servant" in Cairo in 1906. He got interested in the Nile River.
- Hurst spent 62 years in Egypt mostly working on design and construction of reservoirs along the Nile.
- By inspecting historical data on the Nile flows, Hurst discovered phenomenon (now called Hurst effect).

# Hurst's Problem

Optimal reservoir capacity $R$ to accept the river flow in $N$ units of time, $X_1, X_2, \ldots X_N$, with a constant withdrawal of $\bar{X}$ per unit time. The optimal volume of the reservoir is adjusted range,

$$R = \max_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}) - \min_{1 \le k \le N} (X_1 + \cdots + X_k - k\bar{X}).$$

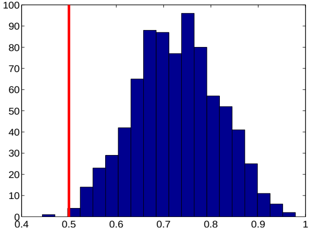
■ In order to compare 100 of years worth of data, Hurst standardized the adjusted ranges $R$, with sample standard deviation

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2} \,,$$

Dimensionless ratio $R/S$ - **rescaled adjusted range.**

# River Nile and Scaling

■ On basis of more than 800 records, Hurst found that quantity $R/S$ scales as $N^H$, for $H$ ranging from 0.46 to 0.93, with mean 0.73 and standard deviation of 0.09.




Easy: $H = 1/2$ for iid normal; Feller: any iid with finite variance; Barnard (1956): Markovian dependent variables.

■ Mandelbrot (1975), Mandelbrot and Van Ness (1968), and Mandelbrot and Wallis (1968) associated the Hurst phenomenon with the presence of long-memory ($\sum_i Cov(X_1, X_i) = \infty$).

# River Nile and Scaling

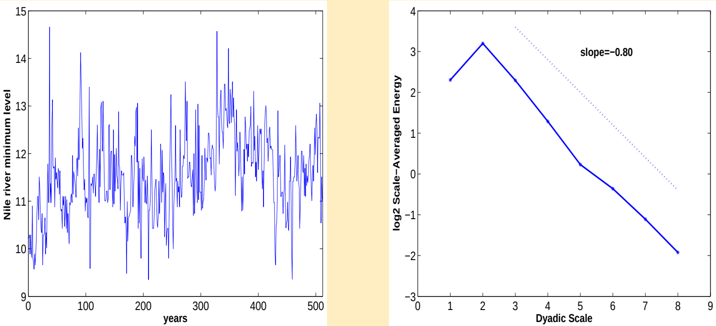



Figure: (a) Nile Yearly Minimum Level Data for $n = 512$ Consecutive Years from 622 A.D.; (b) **Wavelet Log-spectrum** $[0.80 = 2H - 1 \rightarrow H = 0.90]$

- But what is Wavelet Log-spectrum?

# What is Wavelet (Log)Spectrum?

- $\mathbf{y}$ - data, $n \times 1,\ n = 2^J$.
- $\mathbf{d} = W\mathbf{y}$ - discrete wavelet transform, $n \times 1,\ n = 2^J$.
- $\mathbf{d}$=$\{\mathbf{c}_{J-m}, \mathbf{d}_{J-m}, \mathbf{d}_{J-m+1}, \dots, \mathbf{d}_{J-2}, \mathbf{d}_{J-1}\}$.

# What is Wavelet (Log)Spectrum?

- $\mathbf{y}$ - data, $n \times 1,\ n = 2^J$.
- $\mathbf{d} = W\mathbf{y}$ - discrete wavelet transform, $n \times 1,\ n = 2^J$.
- $\mathbf{d}$=$\{\mathbf{c}_{J-m}, \mathbf{d}_{J-m}, \mathbf{d}_{J-m+1}, \dots, \mathbf{d}_{J-2}, \mathbf{d}_{J-1}\}$.

• **Wavelet (Log)Spectrum** $\left\{\left(j, \log_2[\frac{1}{2^j}\sum_k d_{j,k}^2]\right),\ J-m \le j \le J-1\right\}$
• SLOPE either $-(2H+1)$(cumulative) or $-(2H-1)$ (differenced). For example, Brownian Motion and White Noise both share $H = 0.5$.

# “Ubiquitous” – The Epithet of Scaling

## Atmospheric Turbulence

Wind velocity, temperature, concentration of $CO_2$, $O_3$,...

# "Ubiquitous" – The Epithet of Scaling

## Atmospheric Turbulence

Wind velocity, temperature, concentration of $CO_2$, $O_3$,...

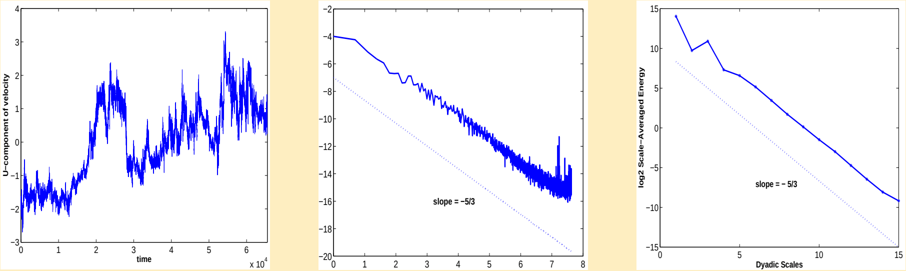



Figure: (a) $U$ Velocity Component; (b) Scaling in the Fourier Domain; (c) Scaling in the Wavelet Domain. $[5/3 = 2H + 1 \rightarrow H = 1/3]$

# DNA Scales

A DNA molecule consists of long complementary double helix of purine nucleotides (denoted as A and G) and pyrimidine nucleotides (denoted as C and T). $[A, G \to +1; \;\; C, T \to -1]$

# DNA Scales

A DNA molecule consists of long complementary double helix of purine nucleotides (denoted as A and G) and pyrimidine nucleotides (denoted as C and T). $[A, G \to +1; \quad C, T \to -1]$

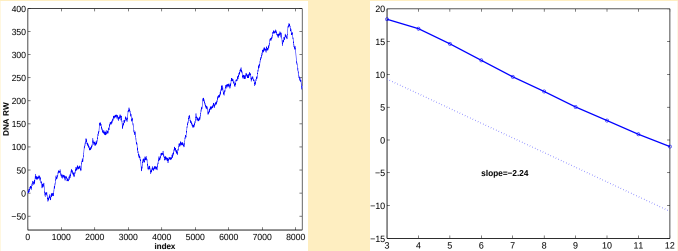


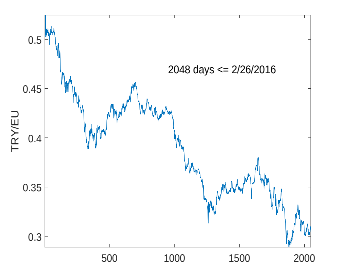

Figure: (a) 8196-long DNA Walk for Spider Monkey, from EMLB Nucleotide Sequence Alignment DNA Database; (b) Wavelet Scaling With Slope $-2.24$. $[2.24 = 2H + 1 \to H = 0.62]$

# Money scales

Daily rates of exchange of Turkish Lira (TRY) to Euro (€), as reported by the European Central Bank between September 22, 2010 and February 26, 2016, (`www.exchangerates.org.uk/`).

# Money scales

Daily rates of exchange of Turkish Lira (TRY) to Euro (€), as reported by the European Central Bank between September 22, 2010 and February 26, 2016, (`www.exchangerates.org.uk/`).



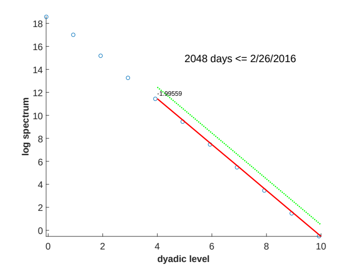



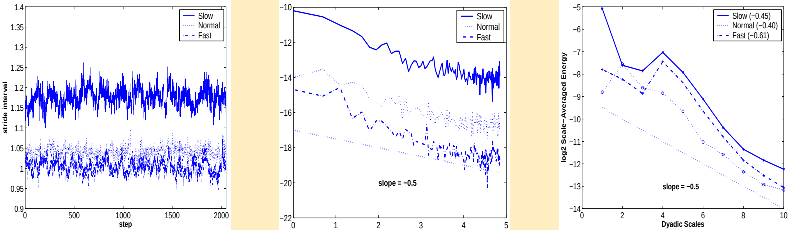

Figure: (a) Daily exchange Rates of TRY to € between 9/22/2010 and 2/16/2016 (Source ExchangeRates.org, UK) (b) Scaling behavior in the rates. $[2 \approx 2H + 1 \rightarrow H \approx 1/2]$

## The Stride Interval

Duration of the gait cycle in human walk. Measurements on a healthy subject who walked for 1 hour at normal, slow and fast paces.

# The Stride Interval

Duration of the gait cycle in human walk. Measurements on a healthy subject who walked for 1 hour at normal, slow and fast paces.




(Left) Gait timing for Slow, Normal and Fast Walk; (Middle) Scaling in the Fourier domain; (Right) In wavelet domain. Slow, normal, and fast stride intervals have slopes of -0.45, -0.4, and -0.61 respectively.

$[\{0.45|0.4|0.61\} = 2H - 1 \rightarrow H = \{0.725|0.7|0.805\}]$

# Other Examples

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)
- Internet Measurements (internet usage, package sizes, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)
- Internet Measurements (internet usage, package sizes, etc)
- Industrial Measurements (high speed manufacturing, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)
- Internet Measurements (internet usage, package sizes, etc)
- Industrial Measurements (high speed manufacturing, etc)
- Astronomy (spatial distributions of stars, astronomical measurements, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)
- Internet Measurements (internet usage, package sizes, etc)
- Industrial Measurements (high speed manufacturing, etc)
- Astronomy (spatial distributions of stars, astronomical measurements, etc)
- Medicine. Brain and Cancer Research (medical imaging, ultrasound, mass-spectrometry, psychomotor responses, etc)

# Other Examples

- Various Geophysical High Frequency Measurements (wind velocities, temperatures, chemical and pollutant concentrations, precipitation, etc.)
- Biometric Responses (brain potentials, ocular responses, etc)
- Economic Indices (stock markets, currency exchange, ultra-high frequency data, etc)
- Internet Measurements (internet usage, package sizes, etc)
- Industrial Measurements (high speed manufacturing, etc)
- Astronomy (spatial distributions of stars, astronomical measurements, etc)
- Medicine. Brain and Cancer Research (medical imaging, ultrasound, mass-spectrometry, psychomotor responses, etc)
- Art (music, paintings, writings, etc)

# Why Should One Care About Scaling?

## Because ...

# Why Should One Care About Scaling?

## Because ...

- If the observations look like "noise" – standard statistical modeling techniques could be inadequate.

# Why Should One Care About Scaling?

## Because ...

- If the observations look like "noise" – standard statistical modeling techniques could be inadequate.
- In denoising applications when there is a superposition of two colored noises, the scaling can guide filtering.

# Why Should One Care About Scaling?

## Because ...

- If the observations look like "noise" – standard statistical modeling techniques could be inadequate.
- In denoising applications when there is a superposition of two colored noises, the scaling can guide filtering.
- There is a need to simulate processes that scale (queueing systems, internet, surrogate data in geosciences, wavestrapping, bootstrap tests, etc.)

# Why Should One Care About Scaling?

## Because ...

- If the observations look like “noise” – standard statistical modeling techniques could be inadequate.
- In denoising applications when there is a superposition of two colored noises, the scaling can guide filtering.
- There is a need to simulate processes that scale (queueing systems, internet, surrogate data in geosciences, wavestrapping, bootstrap tests, etc.)
- Summaries of fractal/multifractal spectra as descriptors/data summaries.

# Types of Scaling

## Regular Scaling

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models
- Regular: Plot of **Log(AverageEnergy)** versus **Log(Scale)** is a straight line.

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models
- Regular: Plot of **Log(AverageEnergy**) versus **Log(Scale)** is a straight line.
- In Wavelet Terms: Plot of **Log2(LevelwiseMean(Wavelet Coefficients Squared))** versus **Level Index** is a straight line.

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models
- Regular: Plot of **Log(AverageEnergy**) versus **Log(Scale)** is a straight line.
- In Wavelet Terms: Plot of **Log2(LevelwiseMean(Wavelet Coefficients Squared))** versus **Level Index** is a straight line.

## Irregular Scaling

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models
- Regular: Plot of **Log(AverageEnergy**) versus **Log(Scale)** is a straight line.
- In Wavelet Terms: Plot of **Log2(LevelwiseMean(Wavelet Coefficients Squared))** versus **Level Index** is a straight line.

## Irregular Scaling

- Type I (Local Monofractals, Time- or Space-Dependent Hurst Exponent)

# Types of Scaling

## Regular Scaling

- Monofractals, Several Popular Theoretical Models
- Regular: Plot of **Log(AverageEnergy**) versus **Log(Scale)** is a straight line.
- In Wavelet Terms: Plot of **Log2(LevelwiseMean(Wavelet Coefficients Squared))** versus **Level Index** is a straight line.

## Irregular Scaling

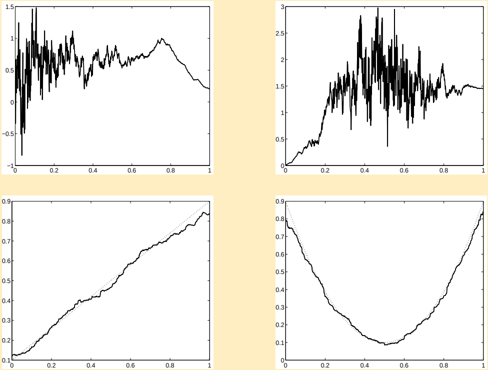
- Type I (Local Monofractals, Time- or Space-Dependent Hurst Exponent)
- Type II (Multifractals, Distribution of Irregularity Indices)

# Irregular Scaling: mfBm($H_t$)

Multifractional Brownian Motions generated with (a) $H_t = 0.8t + 0.1$; (b) $H_t = 0.1 + 3.2(t - 0.5)^2,\ 0 < t < 1$. $H_t$ is estimated by Local Quadratic Variations of sample paths (J.-F. Couerjolly, 2000).

# Irregular Scaling: mfBm($H_t$)

Multifractional Brownian Motions generated with (a) $H_t = 0.8t + 0.1$; (b) $H_t = 0.1 + 3.2(t - 0.5)^2,\ 0 < t < 1$. $H_t$ is estimated by Local Quadratic Variations of sample paths (J.-F. Couerjolly, 2000).

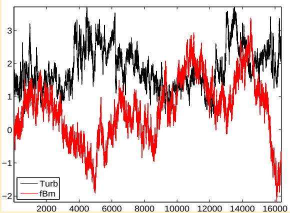


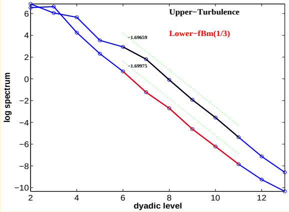
## Irregular Scaling: mf spectra

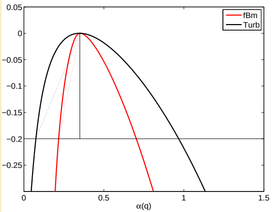
Turbulence (multifractal) and fBm with $H = 1/3$ (monofractal) are indistinguishable wrt the second order properties. Testing for Monofractality (Lee, 2010)

# Irregular Scaling: mf spectra

Turbulence (multifractal) and fBm with $H = 1/3$ (monofractal) are indistinguishable wrt the second order properties. Testing for Monofractality (Lee, 2010)



# Model

# Model

- Random process $X(t),\ t \geq 0$ is self-similar, with self-similarity index $H$ ($H$-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.

# Model

- Random process $X(t),\ t \geq 0$ is self-similar, with self-similarity index $H$ ($H$-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.
- Standard Brownian Motion $B(t)$ is 1/2-ss since the process $W(t) = 1/\sqrt{a}B(at)$ is standard Brownian motion, as well.

# Model

- Random process $X(t),\ t \geq 0$ is self-similar, with self-similarity index $H$ ($H$-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.
- Standard Brownian Motion $B(t)$ is 1/2-ss since the process $W(t) = 1/\sqrt{a}B(at)$ is standard Brownian motion, as well.
- Let $X(t), t \in R$ be $H - ss$ process. If its increments $X(t+h) - X(t)$ are stationary, $X(t)$ is called $H - sssi$ process.

# Model

- Random process $X(t),\ t \geq 0$ is self-similar, with self-similarity index $H$ ($H$-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.
- Standard Brownian Motion $B(t)$ is 1/2-ss since the process $W(t) = 1/\sqrt{a}B(at)$ is standard Brownian motion, as well.
- Let $X(t), t \in R$ be $H - ss$ process. If its increments $X(t+h) - X(t)$ are stationary, $X(t)$ is called $H - sssi$ process.
- $fBm(H)$ is a zero mean Gaussian process for which

$$\mathbb{E}B_H(t)B_H(s) = \frac{E|B_H(1)|^2}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right],$$

where $\mathbb{E}|B_H(1)|^2 = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}$.

# Model

- Random process $X(t),\ t \geq 0$ is self-similar, with self-similarity index $H$ ($H$-ss) if and only if there exists $H > 0$ such that for any $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$.
- Standard Brownian Motion $B(t)$ is 1/2-ss since the process $W(t) = 1/\sqrt{a} B(at)$ is standard Brownian motion, as well.
- Let $X(t), t \in R$ be $H - ss$ process. If its increments $X(t+h) - X(t)$ are stationary, $X(t)$ is called $H - sssi$ process.
- $fBm(H)$ is a zero mean Gaussian process for which

$$\mathbb{E} B_H(t) B_H(s) = \frac{E|B_H(1)|^2}{2} \left[ |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right],$$

  where $\mathbb{E}|B_H(1)|^2 = \frac{\Gamma(2-2H)\cos(\pi H)}{\pi H(1-2H)}$.
- $fBm(H) \equiv$ **unique** Gaussian $H - sssi$ process.

# fBm - Perfect Monofractal

Three fBm paths of length $2^{12}$ with $H = 0.3, 0.5$, and $0.7$.

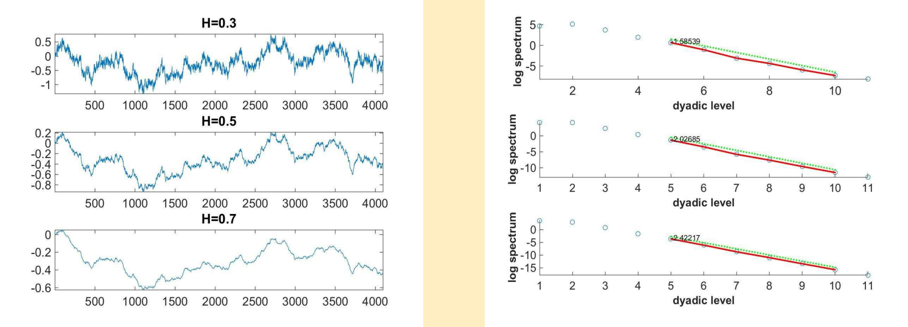



Figure: (a) 4096-long simulated fBm for $H = 0.3, 0.5$, and $0.7$.; (b) Corresponding wavelet spectra, slopes $s = -1.58539, -2.02685$, and $-2.42217$.

# Fractional Brownian Field (2-D fBm)

# Fractional Brownian Field (2-D fBm)

A 2-D fBm, $B_H(\mathbf{t})$, for $\mathbf{t} \in \mathbb{R}^+ \times \mathbb{R}^+$ and $H \in (0,1)$, is a process with stationary zero-mean Gaussian increments,

$$B_H(a\mathbf{t}) \stackrel{d}{=} a^H B_H(\mathbf{t}),$$

$$\mathbb{E}\left[B_H(\mathbf{t})B_H(\mathbf{s})\right] = \frac{\sigma_H^2}{2}\left(\|\mathbf{t}\|^{2H} + \|\mathbf{s}\|^{2H} - \|\mathbf{t}-\mathbf{s}\|^{2H}\right).$$

B. Pesquet-Popescu and J.L. Vehel, 2002

# Fractional Brownian Field (2-D fBm)

A 2-D fBm, $B_H(\mathbf{t})$, for $\mathbf{t} \in \mathbb{R}^+ \times \mathbb{R}^+$ and $H \in (0,1)$, is a process with stationary zero-mean Gaussian increments,

$$B_H(a\mathbf{t}) \stackrel{d}{=} a^H B_H(\mathbf{t}),$$

$$\mathbb{E}\left[B_H(\mathbf{t})B_H(\mathbf{s})\right] = \frac{\sigma_H^2}{2}\left(\|\mathbf{t}\|^{2H} + \|\mathbf{s}\|^{2H} - \|\mathbf{t}-\mathbf{s}\|^{2H}\right).$$

B. Pesquet-Popescu and J.L. Vehel, 2002

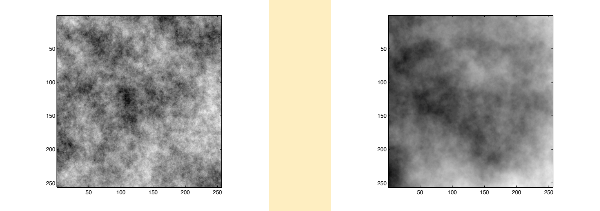



Figure: Fractional Brownian fields for (Left) $H = 1/3$ and (Right) $H = 3/4$.

# Estimation of $H$

# Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations
- Methods Based on Convex Rearrangements

## Multiscale Paradigm

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations
- Methods Based on Convex Rearrangements

## Multiscale Paradigm

- Identify a hierarchy of scales in the multiscale decomposition

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations
- Methods Based on Convex Rearrangements

## Multiscale Paradigm

- Identify a hierarchy of scales in the multiscale decomposition
- Fit the linear propagation of "log-energies" across the scales

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations
- Methods Based on Convex Rearrangements

## Multiscale Paradigm

- Identify a hierarchy of scales in the multiscale decomposition
- Fit the linear propagation of "log-energies" across the scales
- Transform the slopes of the fits to the regularity indices

## Estimation of $H$

- Methods Based on Contrasting Various Estimators of Variance
- Methods Based on Various Aspects of Fourier Spectra (Whittle, Periodogram, ...)
- Methods Based on Zero-Level-Crossings of Paths
- Methods Based on Quadratic Variations
- Methods Based on Convex Rearrangements

## Multiscale Paradigm

- Identify a hierarchy of scales in the multiscale decomposition
- Fit the linear propagation of “log-energies” across the scales
- Transform the slopes of the fits to the regularity indices

- “Waveletize” methods that use filtering (Quadratic Variations, Convex Rearrangements, Lorenzians, ...

## In the rest of the talk...

- 2-D Scale-Mixing Wavelet Transform and its Spectra
- Estimating $H$ from Scale-Mixing Spectra by Theil-Type Estimator
- Applications: Breast Cancer Diagnostics, DNA Exons/Introns, LC-MS Ovarian Cancer Data
- Spectra from 2-D Packet and Non-decimated Scale-Mixing Wavelet Transforms

# 1-D Wavelet Transform via a Matrix

- Given the wavelet basis (via its filter $\boldsymbol{h}$), form an orthogonal matrix $W$ of size $N \times N$ so that for signal $\boldsymbol{y}$, $\boldsymbol{d} = W \cdot \boldsymbol{y}$. (Strang, 1989)
- Matrix $W = W^{(J)}$ is defined iteratively ($J$ is the depth of transform)

$$W^{(1)} = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \quad W^{(2)} = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},$$

$$W^{(3)} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \ldots$$

# 1-D Wavelet Transform via a Matrix

- Given the wavelet basis (via its filter $\boldsymbol{h}$), form an orthogonal matrix $W$ of size $N \times N$ so that for signal $\boldsymbol{y}$, $\boldsymbol{d} = W \cdot \boldsymbol{y}$. (Strang, 1989)
- Matrix $W = W^{(J)}$ is defined iteratively ($J$ is the depth of transform)

$$W^{(1)} = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \quad W^{(2)} = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},$$

$$W^{(3)} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \dots$$

- Filter $\boldsymbol{h} = (h_0, h_1, \dots, h_M) \to H_k$ a matrix od size $N2^{-k} \times N2^{1-k}$, with $(i,j)$th element $h_s$, for $s = N + i - 2j$ modulo $N2^{1-k}$.

# 1-D Wavelet Transform via a Matrix

- Given the wavelet basis (via its filter $\boldsymbol{h}$), form an orthogonal matrix $W$ of size $N \times N$ so that for signal $\boldsymbol{y}$, $\boldsymbol{d} = W \cdot \boldsymbol{y}$. (Strang, 1989)
- Matrix $W = W^{(J)}$ is defined iteratively ($J$ is the depth of transform)

$$W^{(1)} = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \quad W^{(2)} = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix},$$

$$W^{(3)} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \ldots$$

- Filter $\boldsymbol{h} = (h_0, h_1, \ldots, h_M) \to H_k$ a matrix od size $N2^{-k} \times N2^{1-k}$, with $(i,j)$th element $h_s$, for $s = N + i - 2j$ modulo $N2^{1-k}$.
- Matrix $G_k$ formed as $H_k$ but with the QM filter $\boldsymbol{g}$ with $g_m = (-1)^m h_{M-m}$ .

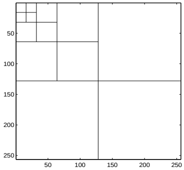
# Scale-mixing Transform

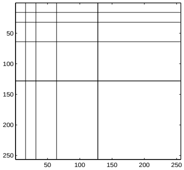
- Let $A$ be an image of power-of-two size, $2^n \times 2^n$. Form a wavelet matrix $W$ of the same size.
- The object $WA'$ represents a matrix in which columns are wavelet transformed rows of $A$. If the same is repeated with the rows of $WA'$ the result is

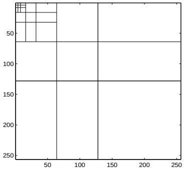
$$B = W(WA')' = WAW'.$$

  Matrix $B$ is called *scale-mixing* wavelet transform of $A$. (Also known as *hyperbolic* or *rectangular*).
- The inverse is straightforward: $A = W'BW$.
- "Energy" preserved: $E = \text{trace}(AA') = \text{trace}(BB')$.
- Easy to generalize to different wavelets (different left- and right-hand side matrices $W_1, W_2$), different transform depths, and to rectangular sizes of $A$.

# Tessellations by 2-D WT: Traditional, Scale-mixing, and a Mix of the Two




- Tiling images by (Left) Traditional 2-D Wavelet Transform; (Middle) Scale-mixing Wavelet Transform; and (Right) Generalized 2-D Wavelet Transform.
- (4,1), (1,4), and (3,2): $(s, t)$ – $s$ times perform a $t$-depth transform on a "smooth" submatrix.
- Any hierarchy of self-similar multiresolution subspaces leads to a spectra.
- For traditional wavelet transforms three spectra usually defined as: horizontal, vertical, and diagonal.

# Hierarchies in Scale-mixing 2-D WT

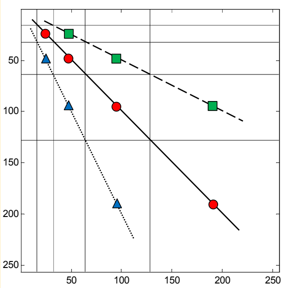



Figure: Hierarchies $(j, j+s)$ for $s = 0 \pm 1$

- Given an isotropic random field, all hierarchies $(j, j+s)$, for $s$ fixed, lead to the same power law.

## Result (Ramírez Cobo et al., 2011)

If $d_{(j,j+s)}$ $(= d_{(j,j+s;k_1,k_2)},\ j = j_0, \ldots, j_1;\ s$ fixed), is a wavelet coefficient in a scale-mixing decomposition of 2-D fBm

$$\log_2 \mathbb{E}\left[d^2_{(j,j+s)}\right] = -(2H+2)j + C_{\psi,s,H}$$

## Result (Ramírez Cobo et al., 2011)

If $d_{(j,j+s)}$ ($= d_{(j,j+s;k_1,k_2)},\ j = j_0, \dots, j_1;\ s$ fixed), is a wavelet coefficient in a scale-mixing decomposition of 2-D fBm

$$\log_2 \mathbb{E}\left[d^2_{(j,j+s)}\right] = -(2H+2)j + C_{\psi,s,H}$$

Hurst exponent can be estimated from the regression slope.

## Fitting the Linear Regression

- Average $d^2$ over the level, take the logs (mean-first) [Abry and collaborators] $(j, \log_2 \overline{d_j^2})$
- Take the logs of $d^2$, then average over the level (log-first) [Taqqu and collaborators] $(j, \overline{\log_2 d_j^2})$
- Average a few $d^2$, take the logs, take average (mean-log-mean) [Soltani and collaborators] $(j, \overline{\log_2 \overline{d_j^2}})$

## Estimating Slope: Approaches

- OLS – wrong methodology – but works OK.
- Weighted LS (Abry & Veitch, 1999)
  $y_j = \log \overline{d_j^2} + 1/(n_j \log 2), \quad n_j$ is the number of $d$s in level $j$.
  $\widehat{sl} = \sum_j w_j y_j$, where $w_j = (S_0 j - S_1)/(S_0 S_2 - S_1^2)$
  $S_k = \sum_j j^k/\sigma_j^2,\ k = 0, 1, 2$ and $\sigma_j^2 = 2/(n_j \log^2 2)$

## Estimating Slope: Approaches

- OLS – wrong methodology – but works OK.
- Weighted LS (Abry & Veitch, 1999)
  $y_j = \log \overline{d_j^2} + 1/(n_j \log 2)$, $n_j$ is the number of $d$s in level $j$.
  $\widehat{sl} = \sum_j w_j y_j$, where $w_j = (S_0 j - S_1)/(S_0 S_2 - S_1^2)$
  $S_k = \sum_j j^k/\sigma_j^2,\ k = 0, 1, 2$ and $\sigma_j^2 = 2/(n_j \log^2 2)$

## Theil-Type Estimator (Hamilton et al, 2016)

- Find all pairwise slopes $s_{ij} = (\log \overline{d_j^2} - \log \overline{d_i^2})/(j - i)$ generated by two points $(i, \log \overline{d_i^2})$ and $(j, \log \overline{d_j^2})$, $j_{min} \leq i < j \leq j_{max}$.
- Estimate the slope as a weighted average of pairwise slopes corrected for the bias.
- What are the optimal weights?

# Optimal Weights for Pairwise Slopes

- Variance of the pairwise slope $s_{ij}$ is

$$\mathbb{V}\mathbf{ar}\, s_{ij} = \frac{2}{(\log 2)^2} \cdot \frac{1/2^{2j} + 1/2^{2i}}{(j-i)^2}.$$

- Take weights $w_{ij}$ inverse-proportional to the variance of $s_{ij}$,

$$w_{ij} \propto (i-j)^2 \times HA(2^{2i}, 2^{2j}), \quad \sum_{i<j} w_{ij} = 1,$$

where $HA$ is the harmonic average of the two level sizes.

# Theil-Type Estimator:

## Bias-corrected slopes

$$w_{ij} \propto (i-j)^2 \times HA(2^{2i}, 2^{2j}),$$

$$s_{ij} = \frac{\log_2 \overline{d_j^2} - \log_2 \overline{d_i^2}}{j-i},$$

$$s_{ij}^* = s_{ij} + \frac{1}{(j-i)\log 2}\left(\frac{1}{2^{2j}} - \frac{1}{2^{2i}}\right)$$

# Theil-Type Estimator:

## Bias-corrected slopes

$$w_{ij} \propto (i-j)^2 \times HA(2^{2i}, 2^{2j}),$$

$$s_{ij} = \frac{\log_2 \overline{d_j^2} - \log_2 \overline{d_i^2}}{j-i},$$

$$s_{ij}^* = s_{ij} + \frac{1}{(j-i)\log 2}\left(\frac{1}{2^{2j}} - \frac{1}{2^{2i}}\right)$$

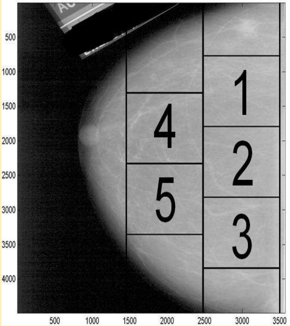
**Theil-Type Estimator:** Slope estimated by $\sum_{i<j} w_{ij} s_{ij}^*$

# Data: Mammogram Images

- Digitized mammograms from University South Florida Digital Database for Screening Mammography (DDSM)
- 45 normal and 79 cancer craniocaudal (CC) images
- Gold standard was biopsy
- HOWTEK scanner at 43.5 micron per pixel
- Five $1024 \times 1024$ subimages from each mammogram

# Data: Mammogram Images

- Digitized mammograms from University South Florida Digital Database for Screening Mammography (DDSM)
- 45 normal and 79 cancer craniocaudal (CC) images
- Gold standard was biopsy
- HOWTEK scanner at 43.5 micron per pixel
- Five $1024 \times 1024$ subimages from each mammogram



# Classification

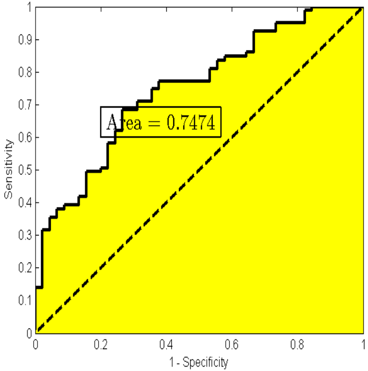



Figure: ROC curve for the logistic classifier. At the most distant point from the diagonal, sensitivity is 68.35% and the specificity is 73.33%

# Scaling in DNA

- We saw an example of DNA random walk
- Not invariant wrt assignment:
  Purine $\{A, G\} \longrightarrow -1$; Pyramidine $\{C, T\} \longrightarrow 1$.
  Weak $H$-bonds $\{A, T\} \longrightarrow -1$; Strong $H$-bonds $\{C, G\} \longrightarrow 1$;
  Amino $\{A, C\} \longrightarrow -1$; Keto $\{G, T\} \longrightarrow 1$.
- Assign each nucleotide one of the 4-D unit vectors $e_1 = \{1, 0, 0, 0\}, \ldots, e_4 = \{0, 0, 0, 1\}$.

$$GATCTCT \ldots \longrightarrow Y = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & \ldots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & \end{bmatrix}.$$

- $Y^*$ row-wise cumulative sum of $Y$.

# Scaling in DNA

$$Y^* = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & \\ 0 & 0 & 0 & 1 & 1 & 2 & 2 & \dots \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & \\ 0 & 0 & 1 & 1 & 2 & 2 & 3 & \end{bmatrix}.$$

$$Z = W_4 \times Y^* \times W_N', \quad \text{2-D scale-mixing wt of } Y^*.$$

- $Z_2 = Z(3:4, N/2+1:N)$ is the finest detail.
- $Z_1 = Z(2, N/4+1:N/2)$ is the coarse detail level.
- Only 2 levels of detail, $Z_1$ and $Z_2$.
- $$s = \log(\overline{Z_2^{*2}}) - \log(\overline{Z_1^{*2}}),$$

where $\overline{A^{*2}}$ is the mean of the squared entries of $A$.

# Scaling in DNA: Equivalence Classes

Let $ACGT$ be assigned to $\{e_1, e_2, e_3, e_4\}$ in this order and the resulting slope be $s$.
The same slope results in assignments $ACTG$, $CAGT$, $CATG$, $GTAC$, $GTCA$, $TGAC$, and $TGCA$ to $\{e_1, e_2, e_3, e_4\}$.

| $s_1$ | $s_2$ | $s_3$ |
|---|---|---|
| $ACGT$ | $AGCT$ | $ATCG$ |
| $ACTG$ | $AGTC$ | $ATGC$ |
| $CAGT$ | $GACT$ | $TACG$ |
| $CATG$ | $GATC$ | $TAGC$ |
| $GTAC$ | $CTAG$ | $CGAT$ |
| $GTCA$ | $CTGA$ | $CGTA$ |
| $TGAC$ | $TCAG$ | $GCAT$ |
| $TGCA$ | $TCGA$ | $GCTA$ |

• Three equivalence classes ($\frac{4!}{2!\ 2!\ 2!}$) of index-coding of nucleotides that lead to three different slopes, $s_1$, $s_2$ and $s_3$.

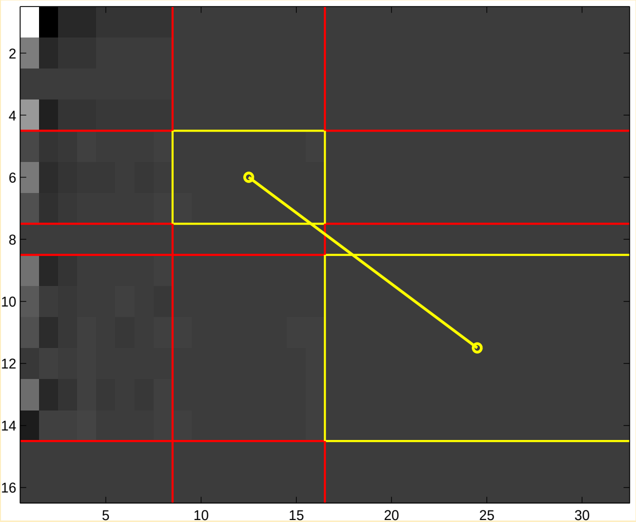
# Scaling in DNA

- The slope should not depend on the coding of nucleotides.
- Each nucleotide is assigned a vector of length 12.
- For example, if $ACGT$, $AGCT$, and $ATCG$ are used, $C$ would correspond to $(0\ 1\ 0\ 0\ \ 0\ 0\ 1\ 0\ \ 0\ 0\ 1\ 0)'$.
- Following this procedure, $Y$ and $Y^*$ become $12 \times N$ matrices.
- 4 rows of $Y^*$ are arbitrary (say 0's)

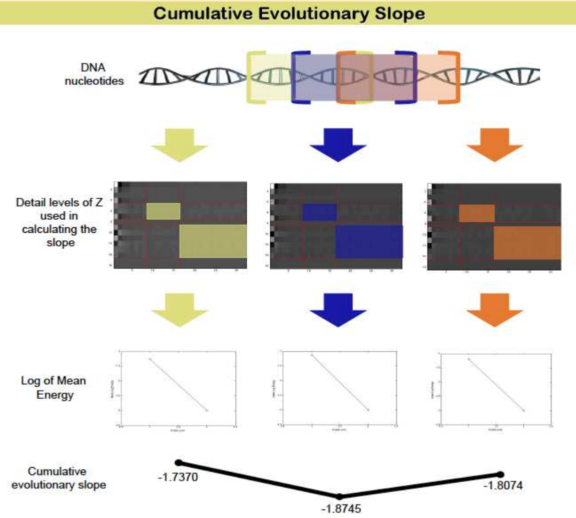
$$Z = W_{16} \times Y^* \times W_N'.$$

- Submatrices of $Z$ are $Z_2 = Z(9:14, N/2+1:N)$ and $Z_1 = Z(5:7, N/4+1:N/2)$
- The resulting slope $s = \log(\overline{Z_2^{*2}}) - \log(\overline{Z_1^{*2}})$ is assignment invariant.

# Finding Slope $s$

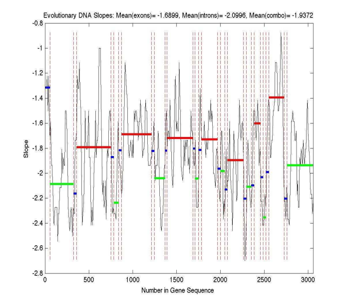


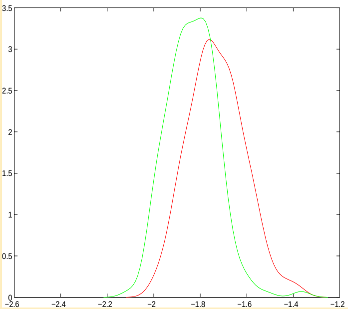
# Evolutionary Slope



# Scaling in DNA: Honey Bee Genome

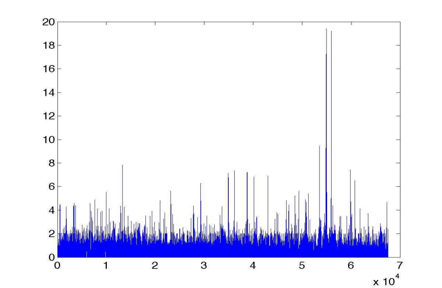
• Genome data of *Apis mellifera* (honey bee) from the NCBI Genome Database (http://www.ncbi.nlm.nih.gov/genome).
• DNA sequence from chromosome 1 contains a total of 29,893,408 base pairs. Exon/Intron (Red/Green) classification: Average accuracy of 65%.
• Woods et al. (2016). Characterizing Exons and Introns by Regularity of Nucleotide Strings. *Biology Direct*, **11**, 6, 1-17.



# Ovarian Cancer Diagnostic via LC-MS Spectra

- The data was obtained through the School of Biology at Georgia Institute of Technology.
- Samples of Liquid Chromatography-Mass Spectra (LC-MS) correspond to 40 cases (Stage III ovarian cancer) and 40 controls.



# Ovarian Cancer Diagnostic via LC-MS Spectra

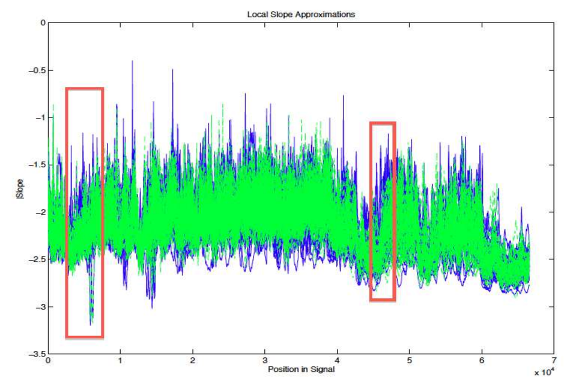



Figure: Regions of time-varying log-spectral slopes used to create classification inputs.

# Ovarian Cancer Diagnostic via LC-MS Spectra

- The results of 1,000 iterations of the classification procedure using these variables are summarized in terms of sensitivity, specificity, and overall accuracy rate.
- For each iteration, the data set is randomly split into a 70% training set (60 signals) and a 30% testing set (20 signals).
- The time-varying slope summaries and anisotropy measures are used as features to train a support vector machines (SVM) model with linear kernel. The mean accuracy achieved was 76.59%.
- Roberts et al. (2017). Wavelet-based Scaling Indices for Ovarian Cancer Diagnostics. Submitted to Stat in Med.

# Non-decimated Wavelet Transforms: Calculation and Spectra

## Non-decimated, Stationary, Maximum-Overlap, À Trous

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j(x-k))$$
$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j(x-k)),$$

insted of traditional

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k) = 2^{j/2}\phi(2^j(x - k/2^j))$$
$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) = 2^{j/2}\psi(2^j(x - k/2^j)),$$

$$\mathcal{W}^{(1)} = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, \mathcal{W}^{(2)} = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \mathcal{W}^{(3)} = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ G_1 \end{bmatrix}, \dots$$

The sizes of $H_p$ and $G_p$ for any dept $p \in \{1, 2, \dots\}$ are $m \times m$.
The entries of $H_p$ and $G_p$ are

$$\begin{aligned} a_{ij} &= \quad \frac{1}{\sqrt{2}} h_s^{[p-1]}, & s &= (N + i - j) \text{ modulo } m \\ b_{ij} &= \frac{1}{\sqrt{2}} (-1)^s h_{N+1-s}^{[p-1]}, & s &= (N + i - j) \text{ modulo } m, \end{aligned}$$

where $N$ is a shift parameter and $h_s^{[p-1]}$ is the $s^{th}$ element of a dilated wavelet filter $h$ with $p - 1$ zeros in between the original components $(h_1, h_2, \dots, h_t)$:

$$h^{[p-1]} = (h_1, \overbrace{0, \dots, 0}^{p-1}, h_2, \overbrace{0, \dots, 0}^{p-1}, h_3, \dots, \overbrace{0, \dots, 0}^{p-1}, h_t)$$

# Scaling of 2-D fBm

For an image $A$, a 2-D NDWT is

$$B = \mathcal{W}_1 \times A \times \mathcal{W}_2'$$

where $\mathcal{W}_1$ and $\mathcal{W}_2$ may use different wavelet filters and different transform depths $p$ and $q$.

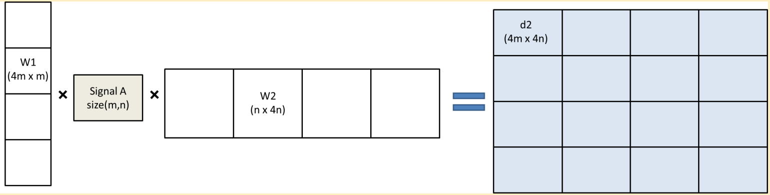
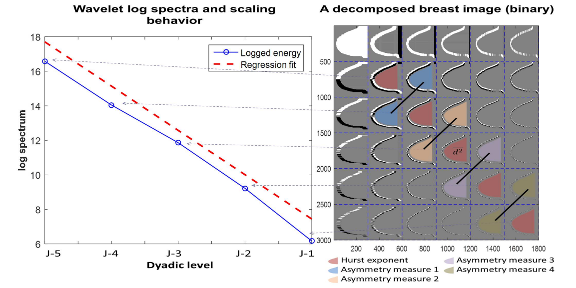



Figure: 2-D NDWT object with $p = q = 3$.

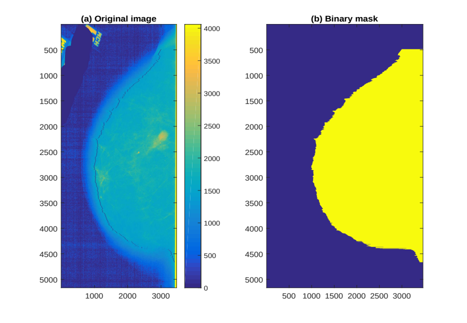
## Result

For fBf with Hurst exponent $H$: Let $d_j$ be a coefficient from diagonal hierarchy of subspaces,
$\log_2 \mathbb{E} d_j^2 = -(2H+2)j + C(\psi, H)$.

## Local Spectra



# Local Spectra




# Classification

Accuracy of approximately 80% using the local spectra and four anisotropy measures. Details in:

Kang et al. (2016). Non-decimated 2D Wavelet Spectrum and Its Use in Breast Cancer Diagnostics. Submitted, Manuscript available in ArXiV.

# 2-D Wavelet Packet Transform and its Spectrum

## 2-D Scale-Mixing Wavelet Packet Transforms

- Wavelet Packets

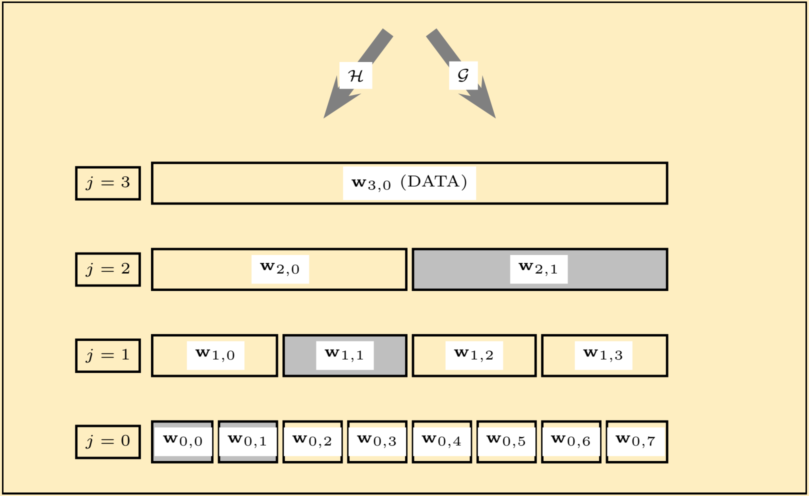
$$\mathbb{W}_{j,n,k}(t) = 2^{j/2}\mathbb{W}_n(2^j t - k),$$

  where $j$ is scale index, $n$ is oscillation index, and $k$ is shift index.

- Atoms $\mathbb{W}_{j,n,k}(t) = 2^{j/2}\mathcal{W}_n(2^j t - k)$, for $j = J-1, J-2, \ldots, J_0 = J - m,\ \ n = 0, 1, 2, \ldots, 2^{J-j} - 1$, and $k = 0, 1, \ldots, 2^j - 1$, form wavelet packet table.
- Let $f(t) \in V_J$.

$$f(t) = \sum_{n=0}^{2^{J-j}-1} \sum_{k=0}^{2^j-1} w_{j,n,k}\mathbb{W}_{j,n,k}(t),\ \ j \text{ fixed in } \{J_0, \ldots, J-1\}$$

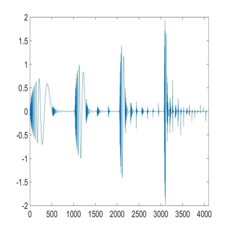
# Wavelet Packet Table

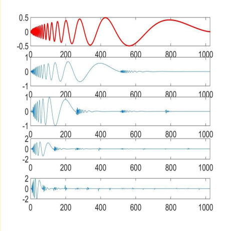



$$\mathbf{w}_{j,n} = (w_{j,n,0}, w_{j,n,1}, \ldots, w_{j,n,2^j-1}),\ 0 \le j \le J,\ 0 \le n \le 2^{J-j} - 1$$

$$W_1 = \begin{bmatrix} H_1 \\ G_1 \end{bmatrix}, W_2 = \begin{bmatrix} \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot H_1 \\ \begin{bmatrix} H_2 \\ G_2 \end{bmatrix} \cdot G_1 \end{bmatrix}, W_3 = \begin{bmatrix} \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot G_2 \end{bmatrix} \cdot H_1 \\ \begin{bmatrix} \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot H_2 \\ \begin{bmatrix} H_3 \\ G_3 \end{bmatrix} \cdot G_2 \end{bmatrix} \cdot G_1 \end{bmatrix}, \dots$$

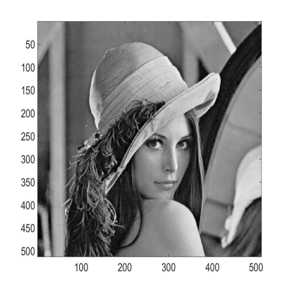
$\boldsymbol{W}_P = (W_1; W_2; \dots, W_k)_{nk \times n}$ $\quad$ $\boldsymbol{w} = \boldsymbol{W}_P \times \boldsymbol{y}$.

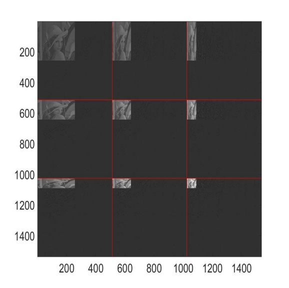


# 2-D Scale-Mixing Wavelet Packet Transform

- For an image $\boldsymbol{A}$, 2-D Scale-Mixing WP transform is $\boldsymbol{B} = \boldsymbol{W}_P \times \boldsymbol{A} \times \boldsymbol{W}_P'$
- Size $k^2$ times larger than the original, as $(nk \times n) \times (n \times n) \times (n \times nk)$.

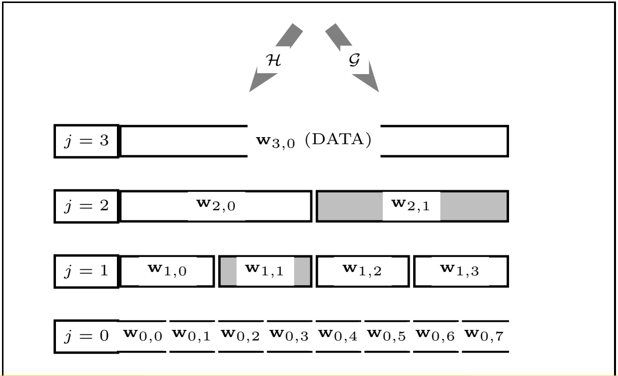


# Wavelet Packet Spectrum

## fBm(H)

$$\text{Average}_{k,\ n \text{ odd}} |w_{j,n,k}|^2 \sim C_1 \times 2^{-(2H+1)j}$$
$$\text{Average}_{k,\ n \text{ even}} |w_{j,n,k}|^2 \sim C_2 \times 2^{-j}$$

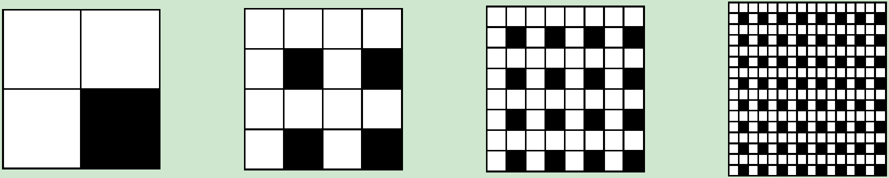


# Scaling of 2-D fBm

- In two dimensions the coefficients $w_{j_1,j_2,n_1,n_2,k_1,k_2}$ correspond to the atom $\mathbb{W}_{j_1,n_1,k_1}(x) \times \mathbb{W}_{j_2,n_2,k_2}(y)$.
- For fractional Brownian field with Hurst exponent $H$,

$$\text{Average}_{k_1,\,k_2,\ (n_1,\,n_2)\text{-shaded}}\ |w_{j,j,n_1,n_2,k_1,k_2}|^2 \sim C \times 2^{-2Hj}$$

## Diagonal Subspaces, Average over Shaded



# Some Simulations

- 1024 runs of $512 \times 512$ 2-D fBm's with given $H$.
- Bottom MSE is for $H_d$ form diagonal hierarchy of traditional 2-D WT.

| Haar 2: WP $\hat{H}$ for Level 4-6 | | | |
|---|---|---|---|
| $H$: | 0.3 | 0.5 | 0.7 |
| $Mean(\hat{H})$: | 0.298430 | 0.502753 | 0.704916 |
| $Var(\hat{H})$: | 0.000123 | 0.000234 | 0.000516 |
| $Bias^2(\hat{H})$: | 0.000003 | 0.000008 | 0.000024 |
| $MSE(\hat{H})$: | 0.000126 | 0.000242 | 0.000540 |
| $MSE(\hat{H}_d)$: | 0.003257 | 0.001848 | 0.001717 |

Ongoing work with graduate student I-Hsiang Lee.

# Some Ongoing "Scaling" Projects

- With Minkyoung Kang: Bayesian Estimation of Scaling: Start with $d_{j\cdot} \sim \mathcal{N}(0, 2^{sj}\sigma^2),\ s = -(2H + d)$. Prior on $s$, MAP estimation of $s$ via optimization.

- With Thelma Safadi: Use of scale mixing 1-D and 2-D NDW transforms in multiscale clustering of financial time series and classification of agricultural seeds wrt their quality.

- With Seonghye Jeon: Wavelet spectra with complex-valued wavelets. Use of phase information.

- With Parisa Yousefi: Use of Autocorrelation Shell hierarchy to estimate $H$. Bootstrapping ACS coefficients to reduce dependence. Estimators biased, but variance small.

# Conclusions

- A case for the importance od scaling assessment made
- When everything else fails, resort to scaling
- Any hierarchy of nested multiresolution subspaces appropriate for the definition of spectra
- Wavelet repository at `http://gtwavelet.bme.gatech.edu` for projects discussed
- Contact `brani@gatech.edu` for more info, data, and MATLAB software