Scale-free dynamics in Internet traffic - The benefits of multivariate analysis ?

Patrice Abry[1,2], **Romain Fontugne**[3], **Gustavo Didier**[4], **Herwig Wendt**[5], **Kensuke Fukuda**[2], **Kenjiro Cho**[3]

[1]CNRS, Ecole Normale Supérieure de Lyon, France
[2]National Institute of Informatics, Tokyo, Japan
[3]Internet Initiative Japan, Tokyo, Japan
[4]Tulane University, New Orleans, USA
[5]CNRS, IRIToulouse, France

Outline

Scale free Internet Traffic

Multivariate SelfSimilarity

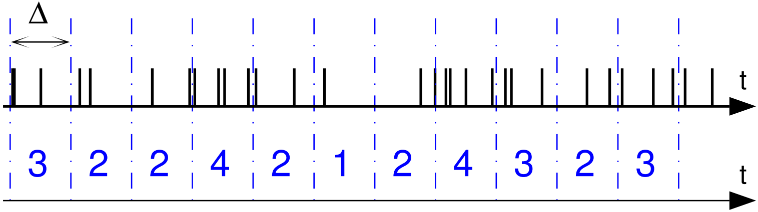
Multivariate Traffic

Anomaly detection

Conclusions

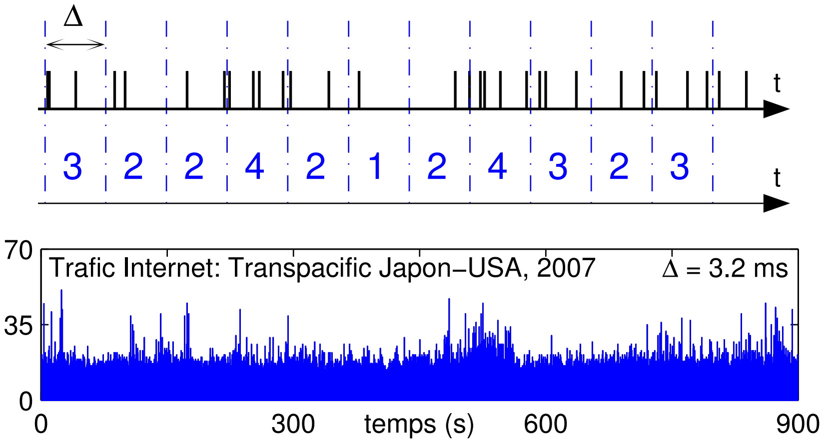
Aggregated Time Series

- Aggregation procedure
  - aggregation scale Δ,
  - Pkt or Byte counts.

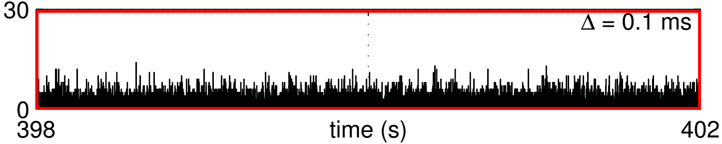


Aggregated Time Series

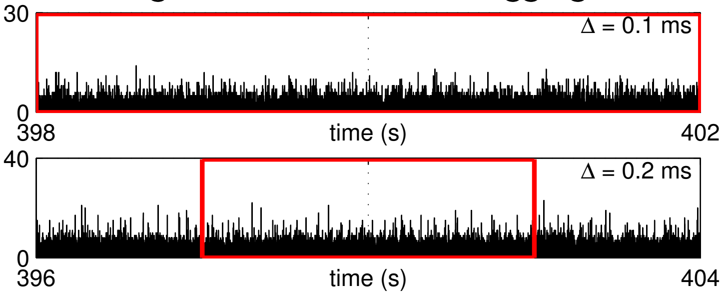
- Aggregation procedure
  - aggregation scale Δ,
  - Pkt or Byte counts.



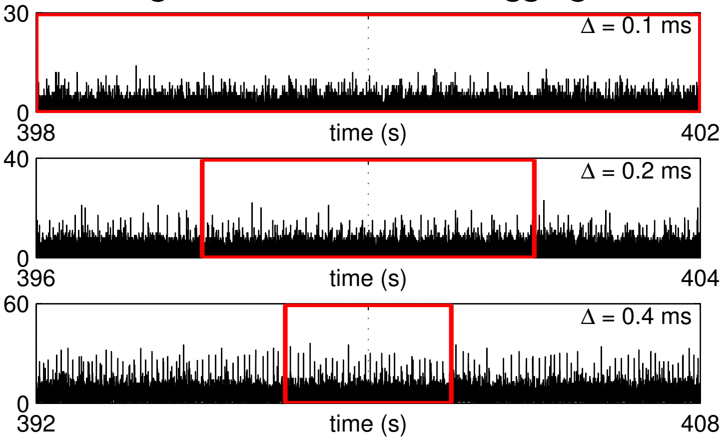
Scaling ? Covariance under aggregation



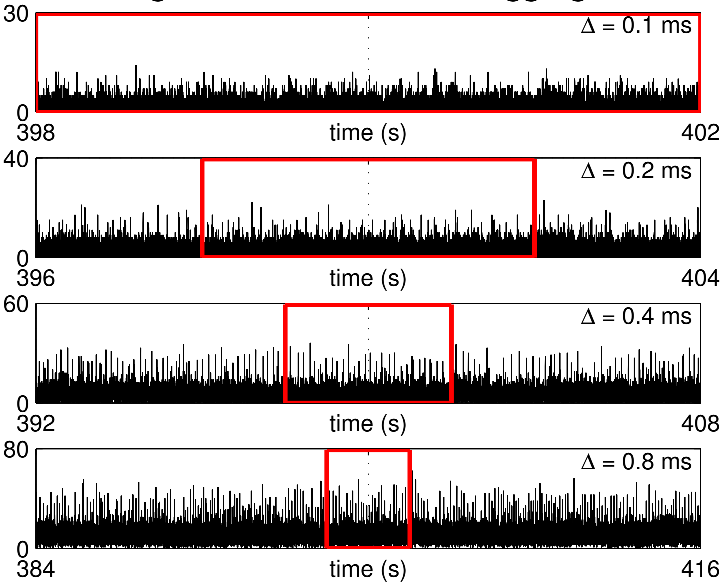
Scaling ? Covariance under aggregation



Scaling ? Covariance under aggregation

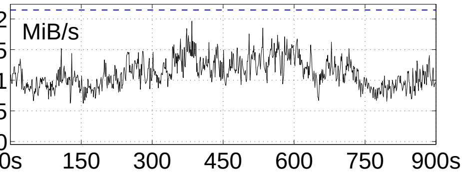


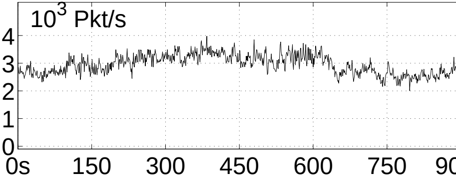
Scaling ? Covariance under aggregation



Statistical modeling of Internet traffic time series

- Aggregated time series: (aggregation levels Δ)
  - Packet counts, Byte counts, Flow counts,
  - Arrivals, durations, ...

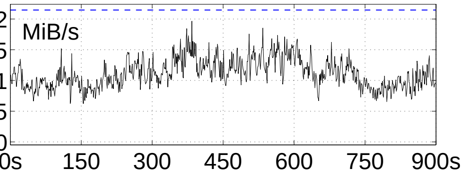


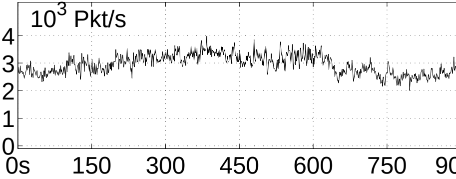


- Statistics: $\Rightarrow$ *Irregularity, Burstiness !*
  - Long Range Dependence (covariance functions) ▸ Definition
  - Heavy Tails (Marginal Distributions) ▸ Definition

Statistical modeling of Internet traffic time series

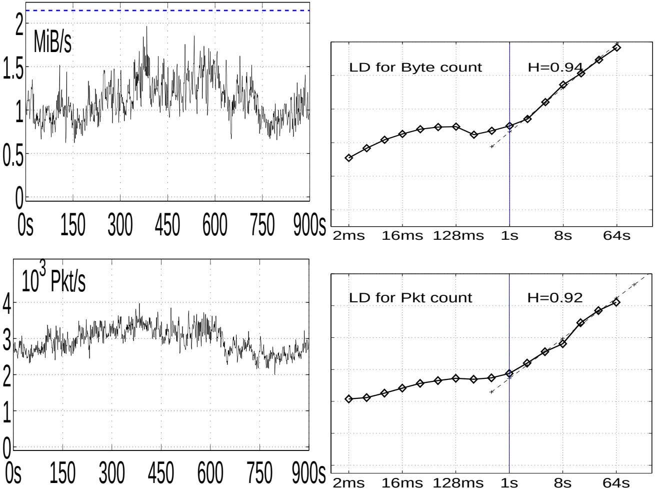
- Aggregated time series: (aggregation levels Δ)
  - Packet counts, Byte counts, Flow counts,
  - Arrivals, durations, ...





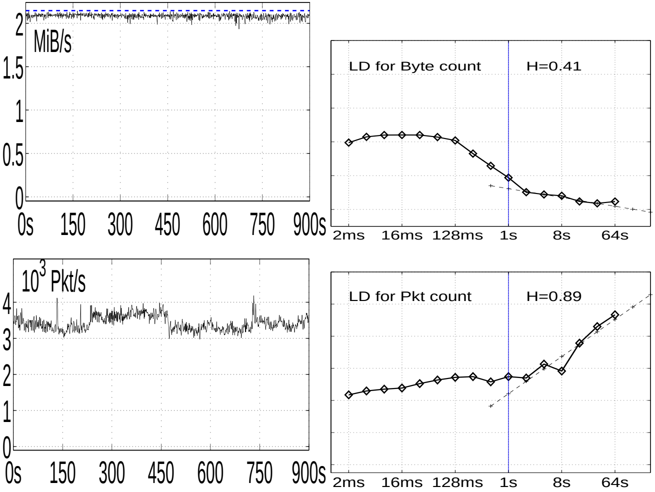
- Statistics: $\Rightarrow$ *Irregularity, Burstiness !*
  - Long Range Dependence (covariance functions) Definition
  - Heavy Tails (Marginal Distributions) Definition

MAWI data: **B**-US2Jp, 2005/07/11



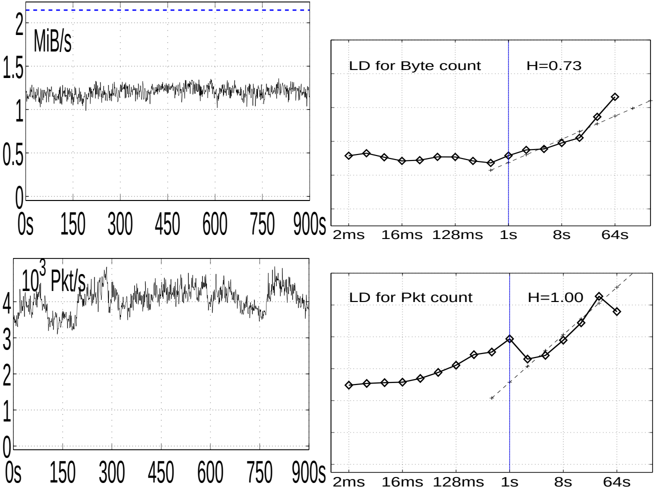
- Compares well with current knowledge and *theory/models*

MAWI data: **B**-US2Jp, 2003/06/03



- Congestion.

MAWI data: **B**-Jp2US, 2004/09/21



- Anomalies:

  network scan, spoofed flooding, attack on a Realserver

Random Projections or sketches

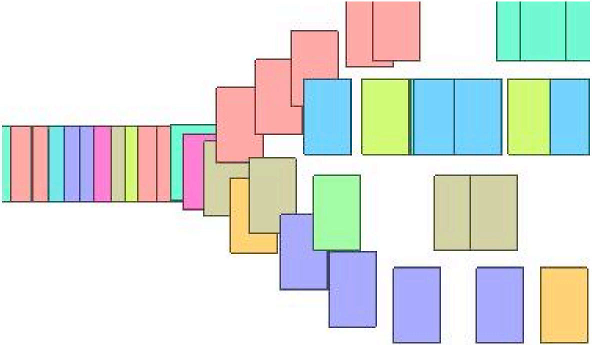
Sketches = ensemble of outputs of random hash table

[Muthukrishnan'03, Krishnamurty'03,...] [Abry+ SAINT'07, Dewaele+ Sigcomm LSAD'07]

- Random Hash Functions : h_n
  - $y = h(x)$,
  - $M-$ outputs: $y \in [1, \ldots, M]$,
  - $k-$ universal Hash functions.

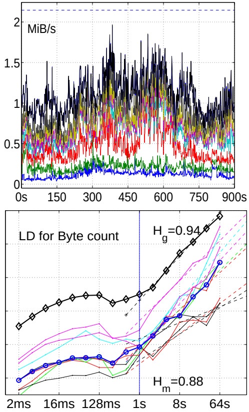
- Hash the Traffic :
  - Packet: $i-$th packet, $n-$tuple: $t_i, PTscr_i, PTdst_i, IPsrc_i, IPdst_i$
  - Choose one specific key: e.g., Destination Address
  - Hash according to this key: $m_i = h(IPdst_i) \in [1, \ldots, M]$,
  - All packets with same m_i = one sub-trace, sampled by random projection.
  - Aggregate traffic $\{t_i, m_i\}_{i \in I}$ into M series $X_\Delta^m(t)$, bins of Δs.

Sketched Traffic



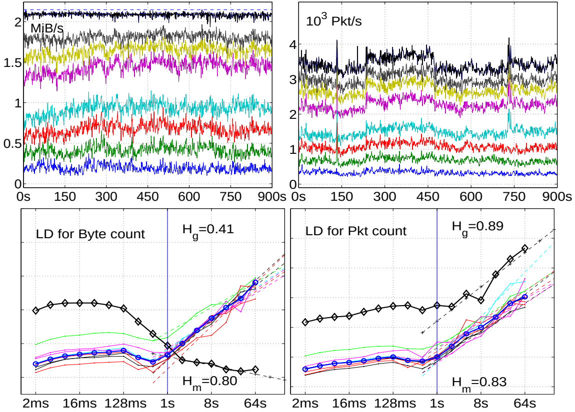
- Sketches = M sub-traces representing the total traffic
- Total of outputs = total trace (constrained sampling)
- Each sketched output = **random flow-sampling**

MAWI data: **B**-US2Jp, 2005/07/11



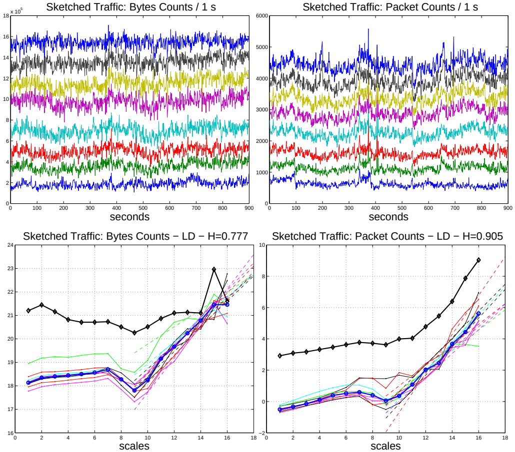
- All H_ms are consistent ! H_ms and H_g are consistent !
- LRDs on Bytes pr Pkts are consistent !
- Normal Traffic: no congestion (no anomaly ?)

MAWI data: **B**-US2Jp, 2003/06/03, Congestion



- $H_g^{\text{Byte}} \simeq 0.4$: no variability, no LRD, $H_g^{\text{Byte}} \neq H_g^{\text{Pkt}}$
- $H_m^{\text{Byte}} \simeq 0.9$, Flow variability, significant LRD, $H_m^{\text{Byte}} \simeq H_m^{\text{Pkt}}$

MAWI data: **B**-Jp2US, 2004/09/21, Anomalies

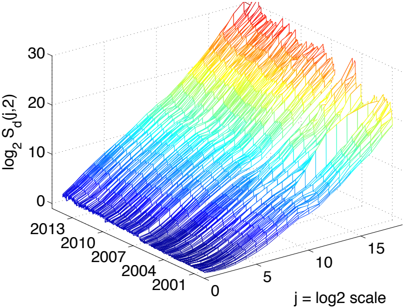


- $H_g^{\text{Byte}} \simeq 0.7$: LD ???, $H_g^{\text{Pkt}} \simeq 1$, ???
- $H_m^{\text{Byte}} \simeq 0.8$, LDs ok, significant LRD, $H_m^{\text{Byte}} \simeq H_m^{\text{Pkt}}$

Univariate Self-Similarity

Fontugne et al. 2017

- Long-Memory (Self-Similarity) at Coarse Scales, $H \simeq 0.9$.
- Multifractality like at Fine Scales
- Frontier scale around 1s, connected to RTT
- Random projections + Multiscale Analysis $\Rightarrow$ robust statistics, anomaly detection

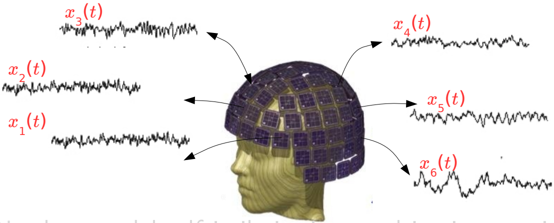


Another point of view?

Limitations

- Not versatile enough for data :
  - One-parameter model: $0 < H < 1$ - Jointly Gaussian
  - ⇒ Multifractal models (univariate) Mandelbrot 1974, Fontugne et al., 2017
  - ⇒ Non Gaussian asymptotically self-similar processes (univariate) Helgason et al., 2005
  - ⇒ Anisotropic SelfSimilar textures (univariate fields) Roux et al. 2013
- Data are naturally multivariate:

- Multivariate wavelet analysis: ‣ failure of univariate analysis



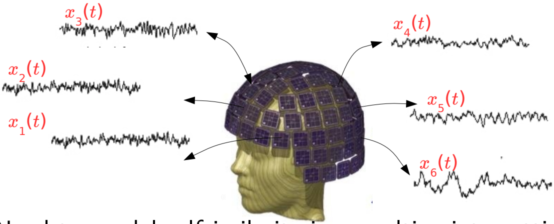
- Need to model selfsimilarity in a multivariate setting

Didier, Pipiras, 2011

Limitations

- Not versatile enough for data :
  - One-parameter model: $0 < H < 1$ - Jointly Gaussian
  - ⇒ Multifractal models (univariate) Mandelbrot 1974, Fontugne et al., 2017
  - ⇒ Non Gaussian asymptotically self-similar processes (univariate) Helgason et al., 2005
  - ⇒ Anisotropic SelfSimilar textures (univariate fields) Roux et al. 2013
- Data are naturally multivariate:

- Multivariate wavelet analysis: ▸ failure of univariate analysis

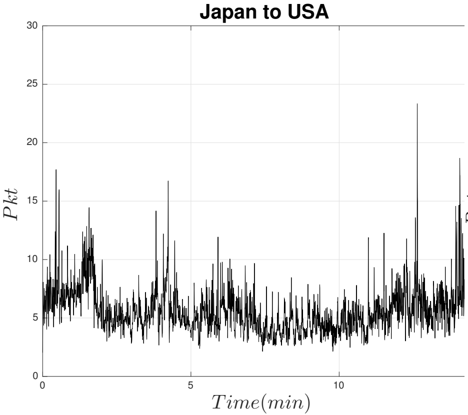


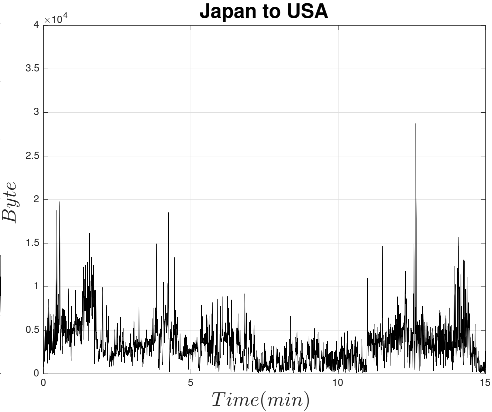
- Need to model selfsimilarity in a multivariate setting

Didier, Pipiras, 2011

Internet Traffic is naturally bivariate

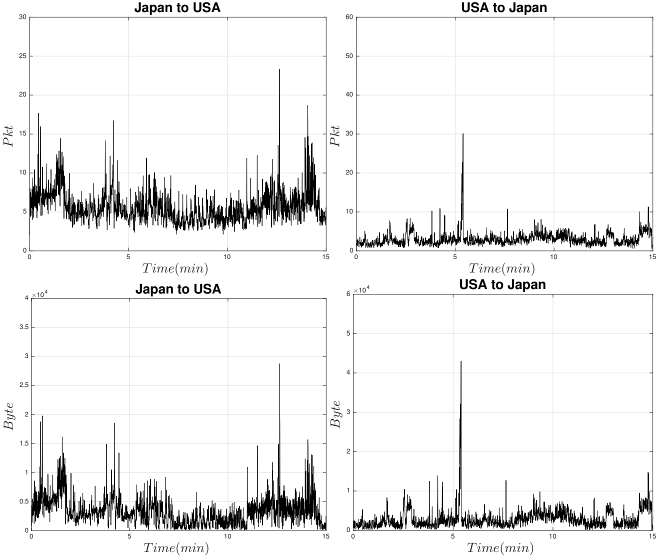
Fontugne et al. 2017, Abry, Didier 2017c





Internet Traffic is naturally 4-variate

Fontugne et al. 2017, Abry, Didier 2017c



Outline

Scale free Internet Traffic

Multivariate SelfSimilarity

Multivariate Traffic

Anomaly detection

Conclusions

Operator Fractional Brownian Motion (OFBM): Definition

Didier, Pipiras, 2011

- M-components: $\{B_{\underline{H},\underline{\underline{\Sigma}}}(t)\}_{t \in \mathcal{R}}$

  $$\{B_{\underline{H},\underline{\underline{\Sigma}}}(t)\}_{t \in \mathcal{R}} = \{B_{h_1}(t), \ldots, B_{h_m}(t), \ldots B_{h_M}(t)\}_{t \in \mathcal{R}}$$

  - M-correlated fBm each with Hurst parameter $0 < h_m < 1$
    $\underline{H} = \{h_1, \ldots, h_m, \ldots, h_M\}$
  - $\underline{\underline{\Sigma}}$: $M \times M$ point covariance (positive definite) matrix
- Linear mixing:
  - $\underline{\underline{W}}$: $M \times M$ invertible matrix (in $\mathcal{R}^M$)
- OFBM: $t \in \mathcal{R} \to B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}} \in \mathcal{R}^M$
  - $B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}(t) = \underline{\underline{W}} \cdot B_{\underline{H},\underline{\underline{\Sigma}}}(t)$
  - *Free* parameters:

    $\underline{H}, \underline{\underline{\Sigma}}, \underline{\underline{W}}$

    $M + M(M-1)/2 + M(M-1) = 3/2M^2 + M/2$

Properties

- Covariance:

  - $\Sigma_{B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}}(t,t') \equiv W \Sigma_{B_{\underline{H},\underline{\underline{\Sigma}}}}(t,t') W^*$

$$(\Sigma_{B_{\underline{H},\underline{\underline{\Sigma}}}}(t,t'))_{m,m'} = (\underline{\underline{\Sigma}})_{m,m'} \cdot \left(|t|^{h_m+h_{m'}} + |t'|^{h_m+h_{m'}} - |t-t'|^{h_m+h_{m'}}\right)$$

$$\Rightarrow \underline{\underline{\Sigma}} \equiv \Sigma_{B_{\underline{H},\underline{\underline{\Sigma}}}}(1,1)$$

- Existence:

  - Matrix $G \circ \underline{\underline{\Sigma}}$ has full rank (Hadamard matrix product)
    $G_{m,m'} = \Gamma(h_m + h_{m'} + 1) \sin((h_m + h_{m'})\pi/2)$

  $\Rightarrow$ constraints on *Free* parameters:

  $\Rightarrow$ $\underline{H}$ and $\underline{\underline{\Sigma}}$ cannot be chosen independently

  $\Rightarrow$ e.g., $M = 2$: $\rho_{12} = \underline{\underline{\Sigma}}_{1,2} / sqrt(\underline{\underline{\Sigma}}_{1,1} \underline{\underline{\Sigma}}_{2,2})$

$$\Gamma(2h_1+1)\Gamma(2h_2+1)\sin(\pi h_1)\sin(\pi h_2) - \rho_{12}^2 \Gamma(h_1+h_2+1)^2 \sin^2(\pi(h_1+h_2)/2) > 0$$

- Time Reversibility:

  By definition: $\Sigma_{B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}}(t,t') = (\Sigma_{B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}}(-t,-t'))^T$

  There exist more general definitions

Multivariate SelfSimilarity

- Selfsimilarity:
  - $\{B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}(t)\}_{t\in\mathcal{R}} \stackrel{fdd}{=} \{a^{\underline{\underline{H}}} B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}(t/a)\}_{t\in\mathcal{R}}, \forall a > 0$
    where $\stackrel{fdd}{=}$: equality of all finite dimensional distributions,
    with $\underline{\underline{H}} = W \cdot$ Diag $\underline{H} \cdot W^{-1}$, $M \times M$ matrix
    where $a^{\underline{\underline{H}}} := \exp(\log(a\underline{\underline{H}})) = \sum_{k>0} \frac{(\log a\underline{\underline{H}})^k}{k\ !}$.

- Mixture of Power-laws:
  - when $W \equiv I_M$

$$\{B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}(t)\}_{t\in\mathcal{R}} \stackrel{fdd}{=} \{a^{h_1} B_{h_1}(t/a), \ldots, a^{h_m} B_{h_m}(t/a), \ldots a^{h_M} B_{h_M}(t/a)\}_{t\in\mathcal{R}}, \forall a > 0$$

  - when $W \neq I_M$
    Multivariate SelfSimilarity $\Rightarrow$ Mixtures of power-laws

Multivariate (discrete) Wavelet Transform

- Wavelet Coefficients:

$$D_{\mathrm{y_m}}(j,k) = \int_{\mathbb{R}} 2^{-j/2}\psi(2^{-j}t-k)Y_m(t)\mathrm{d}t$$

- Vector of Coefficients

$$D_{\mathrm{y}}(j,k) \equiv (D_{\mathrm{y_1}}(j,k),\ldots,D_{\mathrm{y_m}}(j,k),\ldots,D_{\mathrm{y_M}}(j,k))^T$$

- Wavelet Spectrum

$$S(2^j) = \frac{1}{K_j}\sum_{k=1}^{K_j} D(2^j,k)D(2^j,k)^*, \quad K_j = \frac{N}{2^j}$$

$S(2^j)$ is $M \times M$ matrix for each scale 2^j

N: data sample size

$$\mathbf{S(2^j)} = \begin{pmatrix} S_{\mathbf{11}}(2^j) & S_{\mathbf{12}}(2^j) & \ldots & \ldots & \ldots & S_{\mathbf{1}M}(2^j) \\ S_{\mathbf{21}}(2^j) & S_{\mathbf{22}}(2^j) & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & S_{\mathbf{33}}(2^j) & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ S_{M\mathbf{1}}(2^j) & \ldots & \ldots & \ldots & \ldots & S_{MM}(2^j) \end{pmatrix},$$

Multivariate (discrete) Wavelet Transform and OFBM

Frecon et al. 2015, A., Didier 2017a, A., Didier 2017b, A., Didier 2017c,

- Short cuts:
  - Pre-Mixing: $X = B_{\underline{H},\underline{\underline{\Sigma}}}(t)\}_{t\in\mathcal{R}}$
  - Post-Mixing: $Y = B_{\underline{H},\underline{\underline{\Sigma}},\underline{\underline{W}}}(t)\}_{t\in\mathcal{R}}$
- Wavelet Coefficients

$$D_{\mathrm{y}}(j,k) = W2^{j(\underline{H}+I_M/2)}D_{\mathrm{x}}(0,k)$$

- Theoretical Wavelet Spectrum

$$\mathbb{E}D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^* = W2^{j(\underline{H}+I_M/2)}\mathbb{E}D_{\mathrm{x}}(0,k)D_{\mathrm{x}}(0,k)^*2^{j(\underline{H}+I_M/2)^*}W^*$$

$$(1)\ \mathbb{E}D_{\mathrm{y_m}}(j,k)D_{\mathrm{y_{m'}}}(j,k)^* = \sum_{p=1}^{M}\sum_{p'=1}^{M}A_{p,p'}^{(m,m')}(\underline{\underline{\Sigma}},\underline{\underline{W}})2^{j(h_p+h_{p'}+1)}$$

$\Rightarrow$ Mixtures of Power Laws

$\Rightarrow$ Identification: Non linear regression

Frecon et al. 2016, $M=2$, Branch and Bound Strategy

Multivariate analysis: Eigen Value Decomposition

Abry, Didier 2017a $M = 2$, Abry, Didier, Hui 2017b $\Sigma \equiv I_M$, Abry, Didier 2017c, $M \geq 2$

- For each scale j, :
  - Eigen Value Decomposition of $S(2^j)$:
    $$S(2^j) = U(2^j)\,\Lambda(2^j)\,U^*(2^j)$$

$$\mathbf{S}(\mathbf{2^j}) = \mathbf{U}(\mathbf{2^j}) \begin{pmatrix} \lambda_{\mathbf{1}}(S(2^j)) & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{\mathbf{2}}(S(2^j)) & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{\mathbf{3}}(S(2^j)) & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_M(S(2^j)) \end{pmatrix} \mathbf{U}(\mathbf{2^j})^{\mathbf{T}}$$

Multivariate analysis of SelfSimilarity

Abry, Didier 2017a $M = 2$, Abry, Didier 2017c, $M \geq 2$

- Assume:
  - $\forall (m, m'), m' \neq m, h_m \neq h_{m'}$
  - $0 < h_1 < \ldots < h_m < \ldots h_M < 1$

- Consistency:
  - $\lambda_m(S(2^{j(n)})) \to_{n \to +\infty} \xi_m 2^{2h_m j(n)}, \forall m = 1, \ldots, M$
  - $u_m \in \text{span}\{W_{\cdot,m}, W_{\cdot,m+1}, \ldots, W_{\cdot,M}\}, \quad 1 \leq m \leq M$

- Asymptotic Normality:

$$\sqrt{\frac{n}{2^{j(n)}}}\{\log_2 \lambda_m(S(2^{j(n)})) - \log_2 \lambda_m(\mathbb{E}S(2^{j(n)}))\}_{(m=1,\ldots,M, j_1(n) \leq j \leq j_2(n))} \to_{n \to +\infty} \mathcal{N}(0, \Sigma_\lambda)$$

Multivariate EVD estimation of Hurst exponents

- Multivariate estimators:

  $$\widehat{h}_m = \frac{1}{2}\sum_{j=j_1}^{j_2} w_j \log_2 \lambda_m(S(2^j))$$

- Asymptotic Normality:

  $$\sqrt{\frac{n}{2^j(n)}}\{\widehat{h}_m - h_m\}_{m=1,\dots,M} \rightarrow_{n\rightarrow+\infty} \mathcal{N}(0, M_{j_1,j_2}\Sigma_\lambda M^*_{j_1,j_2})$$

- Scaling range $(j_1(n), j_2(n)$

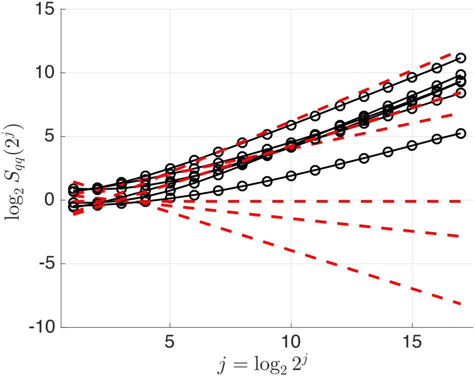
  $$(j_1(n), j_2(n) = (j_1^0 + f(n), (j_2^0 + f(n))$$ (see later)

- Univariate estimators:

  $$\widehat{h}_m^U = \frac{1}{2}(\sum_{j=j_1}^{j_2} w_j \log_2 S_{mm}(2^j) - 1)$$

Univariate (discrete) Wavelet Transform and OFBM

- Diagonal entries of $S_{m,m}(2^j)$:

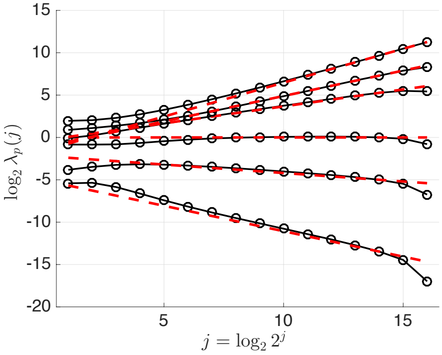


  - Mixture of Power-Laws
  - Dominant *h* only

$\Rightarrow$ Misleading conclusion: All *h* are equal

Multivariate (EVD) Wavelet Transform and OFBM

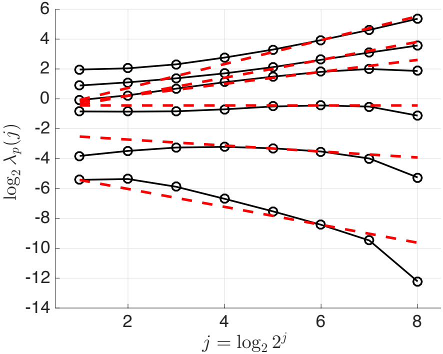
- Eigen Values of $S(2^j)$: λ_m



- Demixed Power-Laws
- All hs
- $\Rightarrow$ correct conclusion: All h can be different

Multivariate (EVD) Wavelet Transform and OFBM

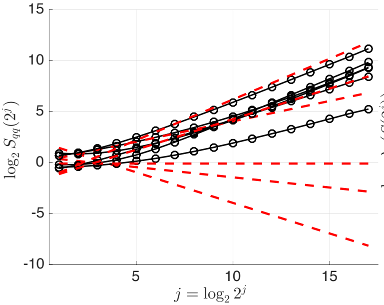
- Eigen Values of $S(2^j)$: λ_m



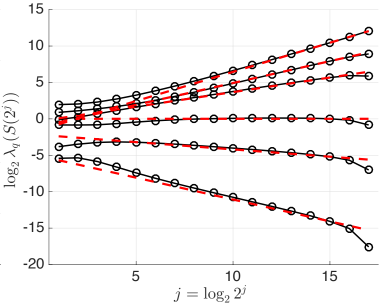
- Demixed Power-Laws
- All *h*s
- Even for very small sample size !

Multivariate (EVD) Wavelet Transform and OFBM

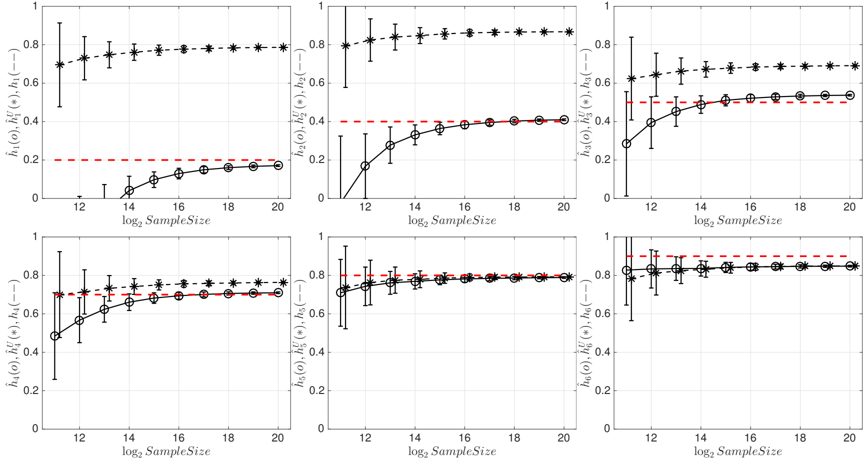
Diagonal entries of $S_{m,m}(2^j)$



Eigen Values of $S(2^j)$: λ_m



Estimation Performance: Bias $\rightarrow$ 0



▸ Proof

Outline

Scale free Internet Traffic

Multivariate SelfSimilarity

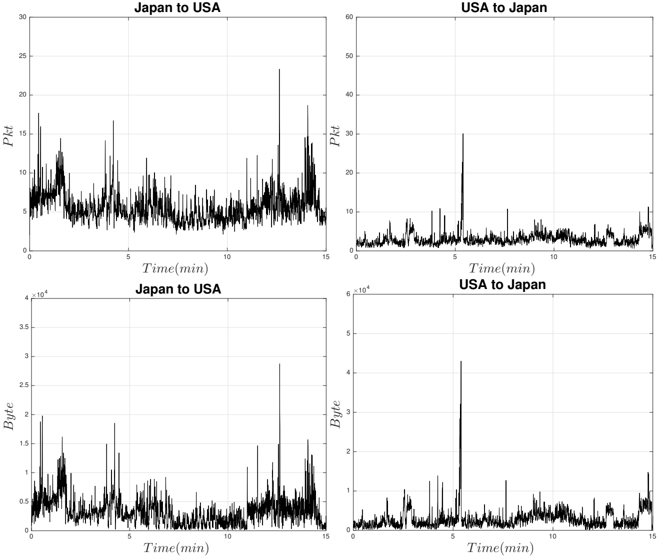
Multivariate Traffic

Anomaly detection

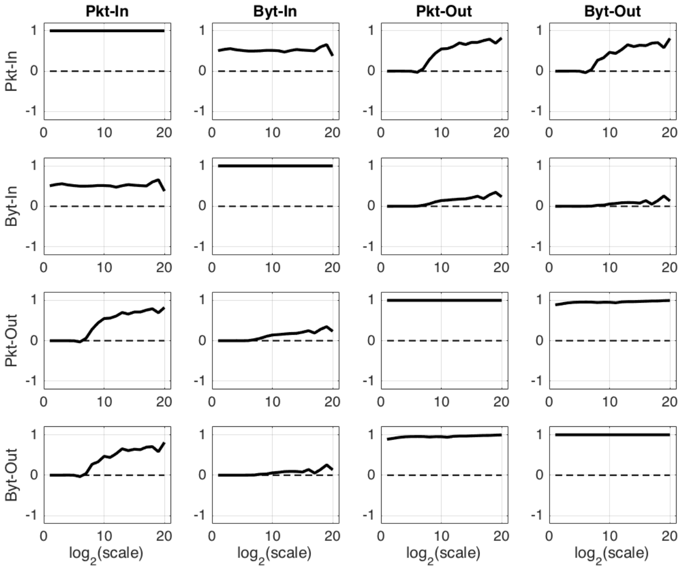
Conclusions

Internet Traffic - $M = 4$

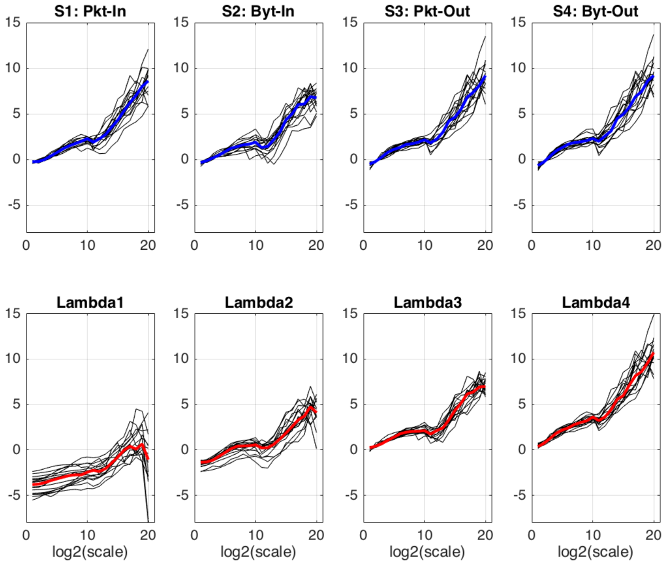
Fontugne et al. 2017, Abry, Didier 2017c



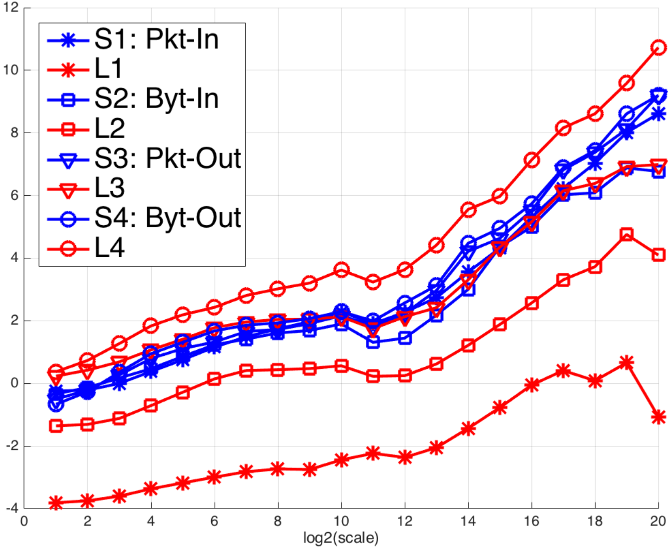
Wavelet Cross Coherence



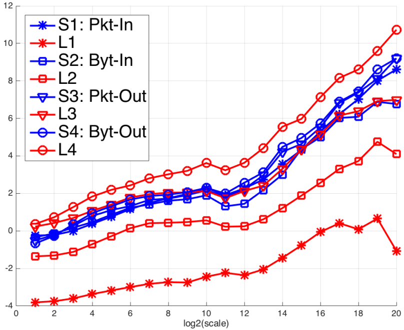
Wavelet Eigen Structure and Random Projections

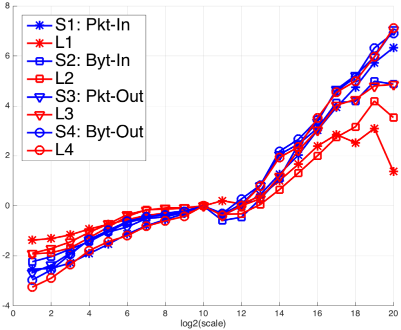


Multivariate (WavEigen) vs. Univariate Structures

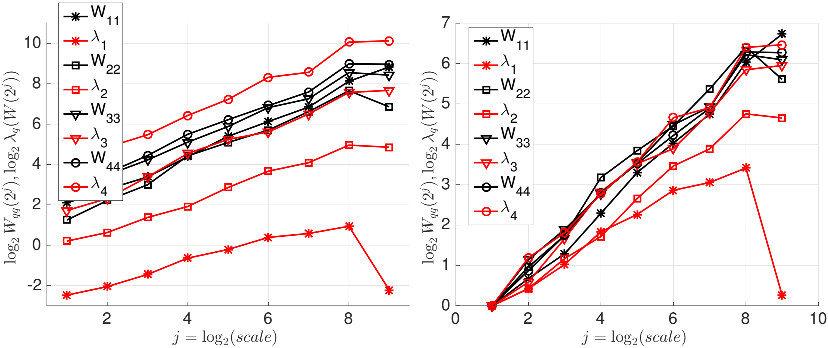


Multivariate (WavEigen) vs. Univariate Structures



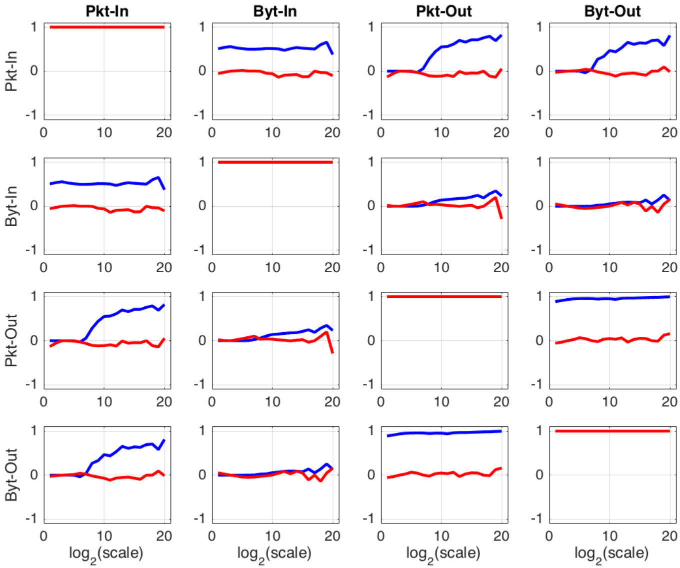


Long Memory at Coarse Scales

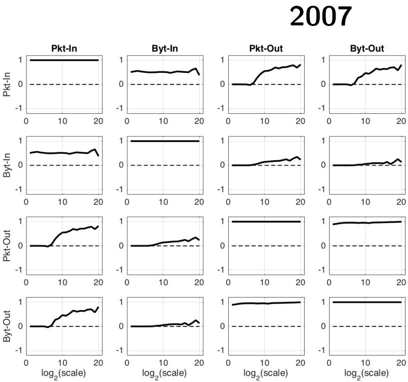


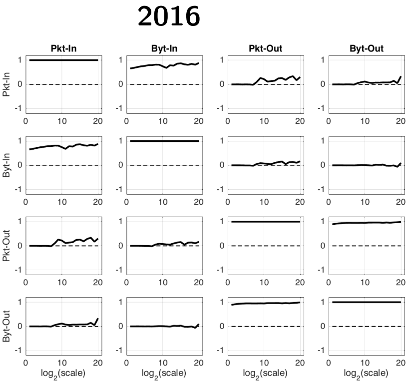
	$\widehat{H}_1$	$\widehat{H}_2$	$\widehat{H}_3$	$\widehat{H}_4$
univariate	0.85	0.86	0.86	0.90
multivariate	0.51	0.69	0.82	0.86

Demixing


Pkt-In
Byt-In
Pkt-Out
Byt-Out
Pkt-In
Byt-In
Pkt-Out
Byt-Out
log2(scale)

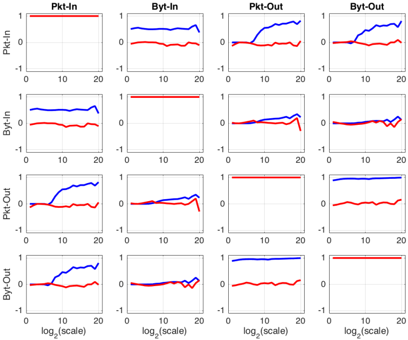
Wavelet Cross Coherence: 2007 vs 2016



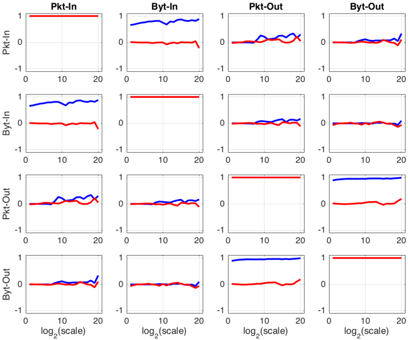


Demixing: 2007 vs 2016

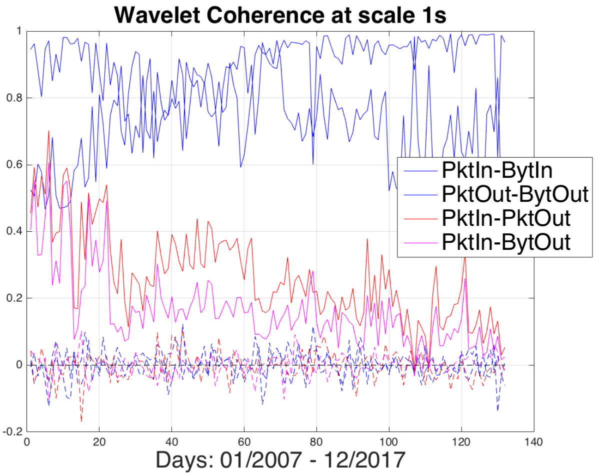
2007



2016

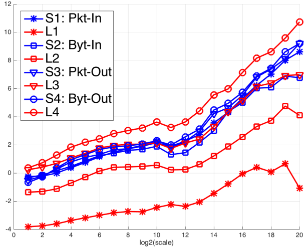


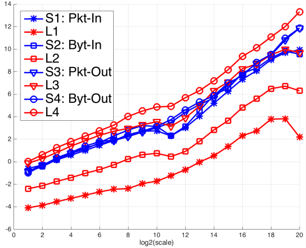
Wavelet Cross Coherence: from 2007 to 2017

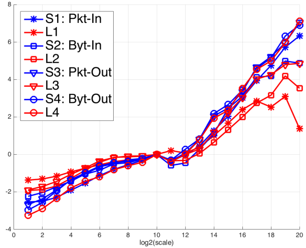


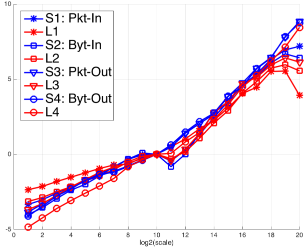
Multi vs. Uni Variate Structures: 2007 vs 2016

2007 | **2016**

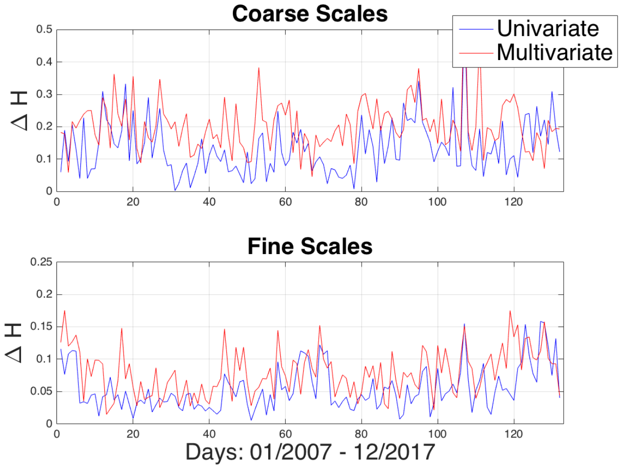








Multi vs. Uni Variate Structures: from 2007 to 2017



Outline

Scale free Internet Traffic

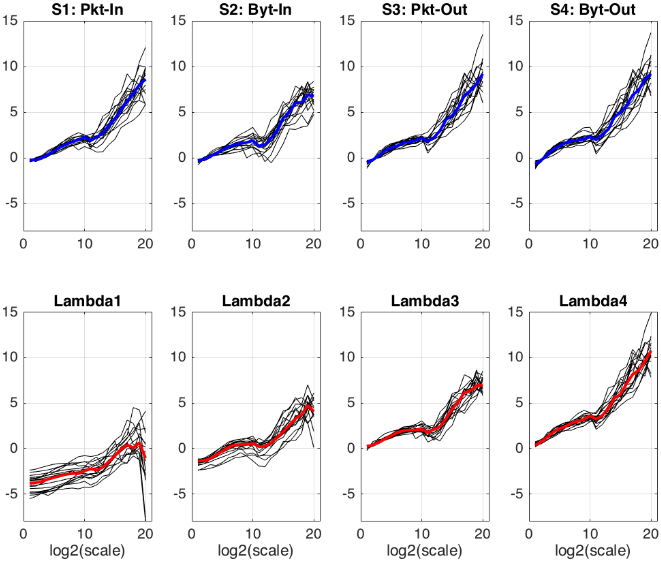
Multivariate SelfSimilarity

Multivariate Traffic

Anomaly detection

Conclusions

Principle


S1: Pkt-In
S2: Byt-In
S3: Pkt-Out
S4: Byt-Out
Lambda1
Lambda2
Lambda3
Lambda4
log2(scale)
15
10
5
0
-5
0
10
20

Case Study 1: Scan found by L1 only - Low Pkt

```
================================================
2007/02/15
scan found only by L1: src ip = "XXX.XXX.XXX.XXX"

1171515602.78838 (2007-02-15 14:00:02.078838)
1171516495.878712 (2007-02-15 14:14:55.878712)
nb_packets: 172
nb_bytes: 10664
src_addr: Nb different values: 1 (0 + 1)
Values:  "XXX.XXX.XXX.XXX" 172
dst_addr: Nb different values: 167 (162 + 5)
transport_portocol: Nb different values: 1 (0 + 1)
Values: tcp 172

TCP:
TCP: nb packets: 172
Src port: Nb different values: 164 (157 + 7)
Dst port: Nb different values: 2 (0 + 2)
Values: 139 74;5900 98
nb urg packets: 0
nb ack packet: 0
nb psh packet: 0
```

Case Study 2: Scan found by L2 only - Short Duration

```
=================================================
2007/01/15
scan found only by L2: src ip XXX.XXX.XXX.XXX

1168837324.539858 (2007-01-15 14:02:04.539858)
1168837473.391106 (2007-01-15 14:04:33.391106)
nb_packets: 1071
nb_bytes: 70686
src_addr: Nb different values: 1 (0 + 1)
Values: XXX.XXX.XXX.XXX 1071
dst_addr: Nb different values: 254 (0 + 254)
transport_portocol: Nb different values: 1 (0 + 1)
Values: tcp 1071

TCP:
TCP: nb packets: 1071
Src port: Nb different values: 480 (77 + 403)
Dst port: Nb different values: 2 (0 + 2)
Values: 139 549;445 522
nb urg packets: 0
nb ack packet: 0
nb psh packet: 0
nb rst packet: 0
```

Longitudinal Study

- Method:

  One day per month, from 2007 to 2017

  Trinocular: filtered out

- Detection:

  Top-5 most anomalous sketch,

  8 successive hash tables,

  Anomalous if suspicious in each hast table,

  Similarity index: $(|A \cap B|)/\min(|A|,|B|)$

Longitudinal Study - from 2007 to 2017

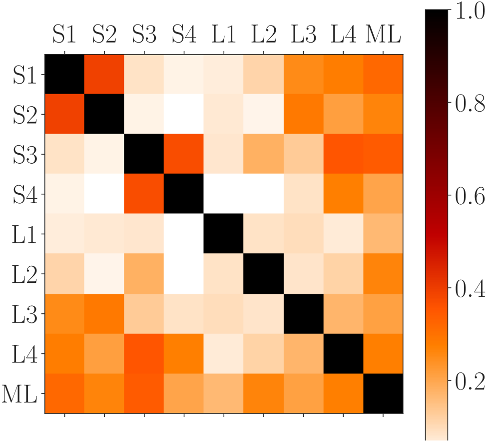
- MawiLab: ~ 142 detections per day on average
- Multiscale:

	1	2	3	4
S	~ 9	$\sim$8	~ 9	~ 8
L	~ 7	~ 7	~ 7	$\sim$9

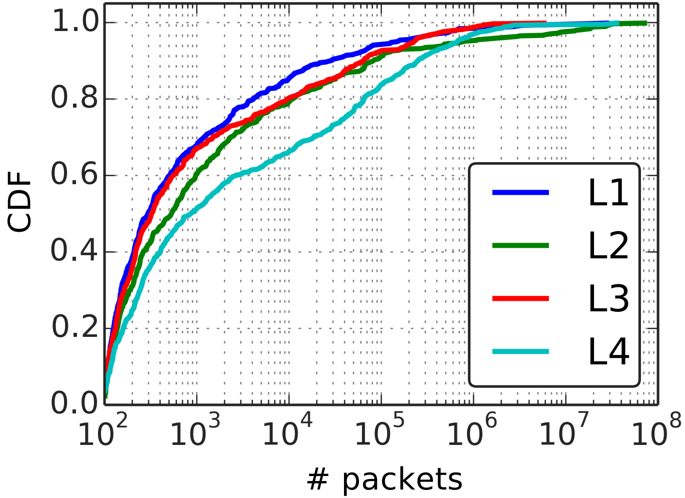
- Multiscale S $\cap$ L: 30% only !

Longitudinal Study

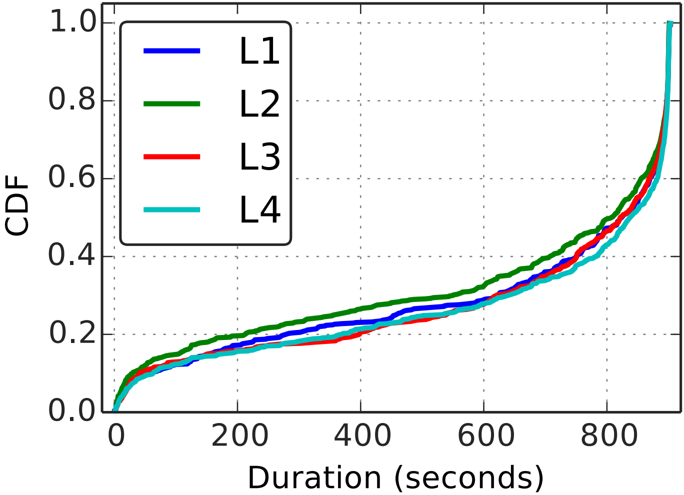
- Univariate: Same anomalies for Pkt & Byt in same direction
- Multivariate: L4 $\sim$ univariate, L1, L2, L3: different anomalies
- MawiLab: Univ. Pkt, then Univ. Byt
  much less in common with L1, L2, L3



Longitudinal Study - Low Pkt Anomalies



Longitudinal Study - Short Duration Anomalies



Outline

Scale free Internet Traffic

Multivariate SelfSimilarity

Multivariate Traffic

Anomaly detection

Conclusions

Conclusions and perspectives

- Scale-free dynamics:
  - Ubiquitous in applications
  - Well-modeled by SelfSimilarity
  - Efficiently analyzed with wavelets
- Multivariate SelfSimilarity:
  - But Data are multivariate
  - Multivariate SelfSimilarity model (OFBM)
  - Multivariate wavelet analysis: ▸ Change of Perspectives

    Univariate: Scales then Components

    Multivariate: Components then Scales

    ⇒ Efficient and robust estimation procedures
- Internet data:
  - Longitudinal study ? Demixing ? Interpretation ?
  - Multivariate statistical modeling ? Anomaly detection ?
- References:
  - patrice.abry@ens-lyon.fr ;
  - http://perso.ens-lyon.fr/patrice.abry/

Conclusions and perspectives

- Scale-free dynamics:
  - Ubiquitous in applications
  - Well-modeled by SelfSimilarity
  - Efficiently analyzed with wavelets
- Multivariate SelfSimilarity:
  - But Data are multivariate
  - Multivariate SelfSimilarity model (OFBM)
  - Multivariate wavelet analysis: ▸ Change of Perspectives

    Univariate: Scales then Components

    Multivariate: Components then Scales

    $\Rightarrow$ Efficient and robust estimation procedures
- Internet data:
  - Longitudinal study ? Demixing ? Interpretation ?
  - Multivariate statistical modeling ? Anomaly detection ?
- References:
  - patrice.abry@ens-lyon.fr ;
  - http://perso.ens-lyon.fr/patrice.abry/

Conclusions and perspectives

- Scale-free dynamics:
  - Ubiquitous in applications
  - Well-modeled by SelfSimilarity
  - Efficiently analyzed with wavelets
- Multivariate SelfSimilarity:
  - But Data are multivariate
  - Multivariate SelfSimilarity model (OFBM)
  - Multivariate wavelet analysis: Change of Perspectives

    Univariate: Scales then Components

    Multivariate: Components then Scales

    $\Rightarrow$ Efficient and robust estimation procedures
- Internet data:
  - Longitudinal study ? Demixing ? Interpretation ?
  - Multivariate statistical modeling ? Anomaly detection ?
- References:
  - patrice.abry@ens-lyon.fr ;
  - http://perso.ens-lyon.fr/patrice.abry/

Conclusions and perspectives

- Scale-free dynamics:
  - Ubiquitous in applications
  - Well-modeled by SelfSimilarity
  - Efficiently analyzed with wavelets
- Multivariate SelfSimilarity:
  - But Data are multivariate
  - Multivariate SelfSimilarity model (OFBM)
  - Multivariate wavelet analysis: ▸ Change of Perspectives

    Univariate: Scales then Components

    Multivariate: Components then Scales

    $\Rightarrow$ Efficient and robust estimation procedures
- Internet data:
  - Longitudinal study ? Demixing ? Interpretation ?
  - Multivariate statistical modeling ? Anomaly detection ?
- References:
  - patrice.abry@ens-lyon.fr ;
  - http://perso.ens-lyon.fr/patrice.abry/

References - Theory

- Multivariate SelfSimilarity Analysis:

- Frecon 2016: J. Frecon, G. Didier, N. Pustelnik, P. Abry, Non-Linear Wavelet Regression and Branch and Bound Optimization for the Full Identification of Bivariate Operator Fractional Brownian Motion, IEEE Transactions on Signal Processing, 64(15):4040-4049, 2016. .pdf
- Abry, Didier, 2017a: Abry, P. and Didier, G., Wavelet estimation for operator fractional Brownian motion, Bernoulli, to appear, 2017. .pdf
- Abry, Didier, Hui, 2017b: Abry, P., Didier, G., Hui L., Two-step wavelet-based estimation for mixed Gaussian fractional processes, Preprint, 2017. .pdf
- Abry, Didier, 2017c: Abry, P. and Didier, G., Wavelet eigenvalue regression for *n*-variate operator fractional Brownian motion, preprint 2017. .pdf
- H. Wendt, G. Didier, S. Combrexelle, P. Abry, Multivariate Hadamard self-similarity: testing fractal connectivity, Signal Processing, 2017. .pdf
- G. Didier, H. Helgason, P. Abry, Demixing Multivariate-Operator Self-Similar Processes, IEEE Int. Conf. on Acoust., Speech and Signal Proc., ICASSP 2015, Brisbane (AU), 20-24 april 2015 .pdf

References - Applications

- Internet Traffic:

- Fontugne et al. 2017: R. Fontugne, P. Abry, K. Fukuda, D. Veitch, K. Cho, P. Borgnat, H. Wendt, Scaling in Internet Traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections, IEEE Trans. on Networking, 2017 .pdf
- P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, D. Veitch, Multiscale Network Traffic Analysis, Modeling, and Inference Using Wavelets, Multifractals, and Cascades, IEEE Signal Processing Magazine 19(3):28–46, May 2002. .pdf

- Neurosciences:

- Ph. Ciuciu, P. Abry, B. He., Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks, NeuroImage, 95:248-263, 2014. .www .pdf

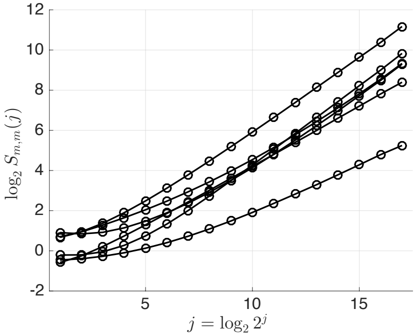
- Art Investigations:

- P. Abry, S. G. Roux, H. Wendt, P. Messier, A. G. Klein, N. Tremblay, P. Borgnat, S. Jaffard, B. Vedel, J. Coddington, L. Daffner, Multiscale Anisotropic Texture Analysis and Classification of Photographic Prints: Art scholarship meets image processing algorithms, IEEE Signal Processing Mag., vol. 32, no. 4, pp. 18-27, July 2015. .pdf

Thank you !

Univariate analysis is dangerous !

Univariate versus Multivariate Analyses

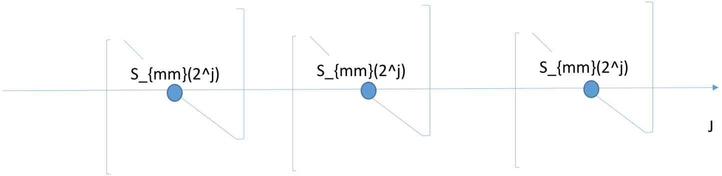


- Diagonal entries of $S_{m,m}(2^j)$:
  - Mixture of Power-Laws

$\Rightarrow$ Misleading conclusion: All h are equal

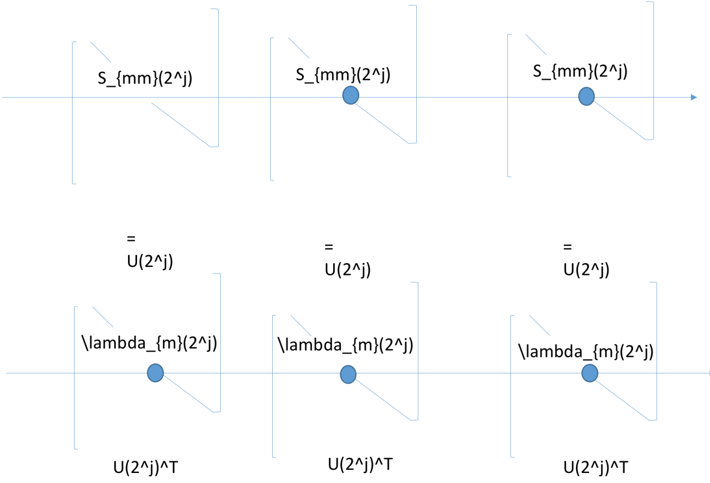
back

Univariate versus Multivariate Analyses



For each m
Study S_mm(2^j)
As a function of j
Then scan m

Univariate versus Multivariate Analyses



For each component m
Study S_mm(2^j)
As a function of j
Then scann
across components

J

For each scale 2^j
Diagonalize S(2^j)
Then study
\lambda_{m}(2^j)
As a function of j

j

◂ back

Long Range Dependence (or covariance)

- Theory: Y 2nd order stationary process
  - Definition:

    Spectrum: $\Gamma_Y(\nu) \sim C_\Gamma |\nu|^{-\gamma}$, $|\nu| \to 0$, $0 < \gamma < 1$,
    Covariance: $\gamma_Y(\tau) \sim C_\gamma |\tau|^{-(1-\gamma)}$, $|\tau| \to \infty$
  - Self-similarity:

    X is H-ss, $\{X(t), t \in \mathcal{R}\} \stackrel{fdd}{=} \{a^H X(t/a), t \in \mathcal{R}\}$, $a > 0$,
    if stationary increments, $Y(k) = X(k+1) - X(k)$
    and $1/2 < H < 1$,
    then Y is LRD with $\gamma = 2H - 1$.
  - Fractional Brownian motion $B_H(t)$:
    Gaussian H-ss, with stationary increments

Long Range Dependence (or covariance)

- Theory: Y 2nd order stationary process
  - Definition:

    Spectrum: $\Gamma_Y(\nu) \sim C_\Gamma |\nu|^{-\gamma}$, $|\nu| \to 0$, $0 < \gamma < 1$,
    Covariance: $\gamma_Y(\tau) \sim C_\gamma |\tau|^{-(1-\gamma)}$, $|\tau| \to \infty$
  - Self-similarity:

    X is H-ss, $\{X(t), t \in \mathcal{R}\} \stackrel{fdd}{=} \{a^H X(t/a), t \in \mathcal{R}\}$, $a > 0$,
    if stationary increments, $Y(k) = X(k+1) - X(k)$
    and $1/2 < H < 1$,
    then Y is LRD with $\gamma = 2H - 1$.
  - Fractional Brownian motion $B_H(t)$:
    Gaussian H-ss, with stationary increments

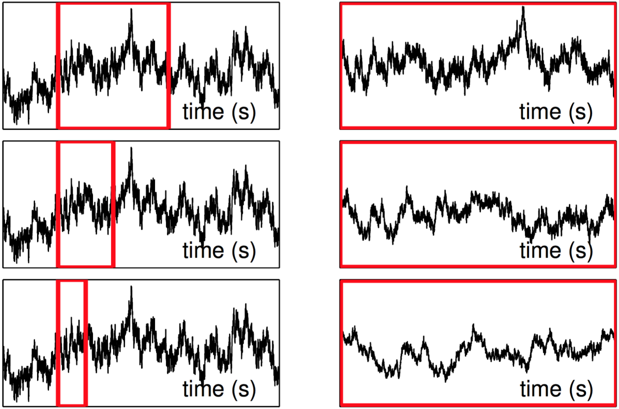
Long Range Dependence (or covariance)

- Theory: Y 2nd order stationary process
  - Definition:

    Spectrum: $\Gamma_Y(\nu) \sim C_\Gamma |\nu|^{-\gamma}$, $|\nu| \to 0$, $0 < \gamma < 1$,
    Covariance: $\gamma_Y(\tau) \sim C_\gamma |\tau|^{-(1-\gamma)}$, $|\tau| \to \infty$
  - Self-similarity:

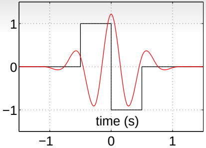
    X is H-ss, $\{X(t), t \in \mathcal{R}\} \stackrel{fdd}{=} \{a^H X(t/a), t \in \mathcal{R}\}$, $a > 0$,
    if stationary increments, $Y(k) = X(k+1) - X(k)$
    and $1/2 < H < 1$,
    then Y is LRD with $\gamma = 2H - 1$.
  - Fractional Brownian motion $B_H(t)$:
    Gaussian H-ss, with stationary increments

Scale-free dynamics: Intuition



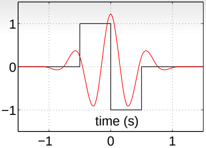
- Covariance under Dilation (Change of Scale),
- The Whole and the SubPart (Statistically) Undistinguishable,
- No Characteristic Scale of Time

LRD and Wavelets



- Wavelets: ▸ Wavelet Transform
  - Mother-Wavelet ψ: Oscillating pattern,
  - Number of vanishing moments N_ψ: $\forall k = 0, ..., N-1$, $\int_{\mathcal{R}} t^k \psi_0(t) dt \equiv 0$ and $\int_{\mathcal{R}} t^N \psi_0(t) dt \neq 0$.
  - Basis: $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\}$,
  - Coefficients of Y: $d_Y(j,k) = \langle \psi_{j,k}, Y \rangle$
- Wavelets and 2nd order stationary process:
  - $\mathbf{E}|d_Y(j,k)|^2 = \int_{\mathcal{R}} \Gamma_Y(\nu) 2^j |\tilde{\Psi}_0(2^j \nu)|^2 d\nu$,
- Wavelets and LRD:
  - $\mathbf{E}|d_Y(j,k)|^2 \sim C 2^{j(2H-1)}$ for $2^j \to +\infty$,
  - $S(j) = \frac{1}{n_j} \sum_k |d_Y(j,k)|^2$,
  - Logscale Diagram: $\log_2 S(j)$ vs. $\log_2 2^j = j$,
  - $\hat{H} = \frac{1}{2}\left(1 + \sum_{j=j_1}^{j_2} w_j \log_2 S(j)\right)$.

LRD and Wavelets

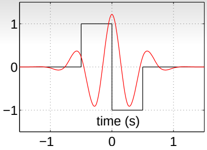


- Wavelets: ▸ WaveletTransform
  - Mother-Wavelet ψ: Oscillating pattern,
  - Number of vanishing moments N_ψ: $\forall k = 0, ..., N-1$, $\int_{\mathcal{R}} t^k \psi_0(t)dt \equiv 0$ and $\int_{\mathcal{R}} t^N \psi_0(t)dt \neq 0$.
  - Basis: $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\}$,
  - Coefficients of Y: $d_Y(j,k) = \langle \psi_{j,k}, Y \rangle$

- Wavelets and 2nd order stationary process:
  - $\mathbf{E}|d_Y(j,k)|^2 = \int_{\mathcal{R}} \Gamma_Y(\nu) 2^j |\tilde{\Psi}_0(2^j\nu)|^2 d\nu$,

- Wavelets and LRD:
  - $\mathbf{E}|d_Y(j,k)|^2 \sim C 2^{j(2H-1)}$ for $2^j \to +\infty$,
  - $S(j) = \frac{1}{n_j}\sum_k |d_Y(j,k)|^2$,
  - Logscale Diagram: $\log_2 S(j)$ vs. $\log_2 2^j = j$,
  - $\hat{H} = \frac{1}{2}\left(1 + \sum_{j=j_1}^{j_2} w_j \log_2 S(j)\right)$.

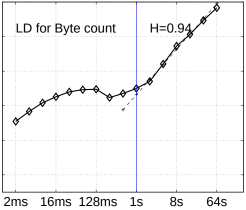
LRD and Wavelets



- Wavelets: ‣ WaveletTransform
  - Mother-Wavelet ψ: Oscillating pattern,
  - Number of vanishing moments N_ψ: $\forall k = 0, ..., N-1$, $\int_{\mathcal{R}} t^k \psi_0(t)dt \equiv 0$ and $\int_{\mathcal{R}} t^N \psi_0(t)dt \neq 0$.
  - Basis: $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)\}$,
  - Coefficients of Y: $d_Y(j,k) = \langle \psi_{j,k}, Y \rangle$

- Wavelets and 2nd order stationary process:
  - $\mathbf{E}|d_Y(j,k)|^2 = \int_{\mathcal{R}} \Gamma_Y(\nu) 2^j |\tilde{\Psi}_0(2^j \nu)|^2 d\nu$,

- Wavelets and LRD:
  - $\mathbf{E}|d_Y(j,k)|^2 \sim C 2^{j(2H-1)}$ for $2^j \to +\infty$,
  - $S(j) = \frac{1}{n_j} \sum_k |d_Y(j,k)|^2$,
  - Logscale Diagram: $\log_2 S(j)$ vs. $\log_2 2^j = j$,
  - $\hat{H} = \frac{1}{2}\left(1 + \sum_{j=j_1}^{j_2} w_j \log_2 S(j)\right)$.



Wavelet Transform

- Let ψ_0 denote an elementary mother wavelet,
- Shifted and dilated templates of ψ_0:
  $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$,
- Wavelet Coefficients: $d_{X_\Delta}(j,k) = \langle \psi_{j,k}, X_\Delta \rangle$.

◂ Back

Sketch of Proof - 1

- $\log_2 \lambda_m(S(2^j(n))) \to 2jh_m$?

- Courant-Fischer principle:
  - Let $\mathcal{U}_m$ such that $\dim \mathcal{U}_m = m$

$$\lambda_m(S(2^j)) = \inf_{\mathcal{U}_m} \sup_{x \in \mathcal{U}_m \cap S_{\mathbb{C}}^{M-1}} x^* S(2^j) x$$

- Hence:
  - Study $x^* S(2^j) x$

Sketch of Proof - 2

- Wavelet Spectrum:
  $\mathbb{E}D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^* = W2^{j(\underline{H}+I_M/2)}\mathbb{E}D_{\mathrm{x}}(0,k)D_{\mathrm{x}}(0,k)^*2^{j(\underline{H}+I_M/2)^*}W^*$

$$\mathbb{E}D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^* = W2^{j(\underline{H}+I_M/2)}\underbrace{W^{-1}\mathbb{E}D_{\mathrm{Y}}(0,k)D_{\mathrm{y}}(0,k)^*(W^*)^{-1}}_{B_{W,\Sigma}(0)}2^{j(\underline{H}+I_M/2)^*}W^*$$

$$S(2^j) = W2^{j(\underline{H}+I_M/2)}\underbrace{W^{-1}D_{\mathrm{Y}}(0,k)D_{\mathrm{y}}(0,k)^*(W^*)^{-1}}_{\hat{B}_{W,\Sigma}(0)}2^{j(\underline{H}+I_M/2)^*}W^*$$

- OFBM is well-defined:
  - $\Rightarrow$ $B_{W,\Sigma}(0)$ has bounded eigen values
  - $\Rightarrow$ $\hat{B}_{W,\Sigma}(0)$ has bounded eigen values
  - $\Rightarrow$ $0 < A \leq \lambda_m(\hat{B}_{W,\Sigma}(0)) \leq B < \infty$
- Hence:
  $0 < A \cdot x^*WDD^*W^*x \leq x^*S(2^j)x \leq B \cdot x^*WDD^*W^*x < \infty$

  with $D = \mathrm{Diag}\{2^{jh_1}, \ldots, 2^{jh_m}, \ldots, 2^{jh_M}\}$

Sketch of Proof - 4

- $D = \text{Diag}\{2^{jh_1}, \dots, 2^{jh_m}, \dots, 2^{jh_M}\}$
- When $W = I_M$:

$$0 < A \cdot x^* WDD^* W^* x \leq x^* S(2^j) x \leq B \cdot x^* WDD^* W^* x < \infty$$

$$0 < A \cdot x^* DD^* x \leq x^* S(2^j) x \leq B \cdot x^* DD^* x < \infty$$

$$\forall m = 1, \dots, M,\ 0 < A \cdot \lambda_m(DD^*) \leq \lambda(S(2^j)) \leq B \cdot \lambda_m(DD^*) < \infty$$

$$\forall m = 1, \dots, M,\ 0 < A \cdot 2^{2jh_m} \leq \lambda_m(S(2^j)) \leq B \cdot 2^{2jh_m} < \infty$$

$$\Rightarrow\ \forall m = 1, \dots, M,\ \log_2 \lambda_m(S(2^j))) \to 2h_m \cdot j$$

Sketch of Proof - 5

- When $W \neq I_M$:
  $$0 < A \cdot x^* W D D^* W^* x \leq x^* S(2^j) x \leq B \cdot x^* W D D^* W^* x < \infty$$

$$x^* W D D^* W^* x = \frac{x^* W}{||W^* x||} D D^* \frac{W^* x}{||W^* x||} \times ||W^* x||^2$$

$$0 < A' \cdot \leq ||W^* x||^2 \leq B' < +\infty \text{ since } W \text{ invertible}$$

$$0 < A' \cdot \frac{x^* W}{||W^* x||} D D^* \frac{W^* x}{||W^* x||} \leq x^* W D D^* W^* x < B' \cdot \frac{x^* W}{||W^* x||} D D^* \frac{W^* x}{||W^* x||} < +\infty$$

$$\forall m = 1, \ldots, M, \; 0 < A \cdot A' \cdot \lambda_m(D D^*) \leq \lambda(S(2^j)) \leq B \cdot B' \cdot \lambda_m(D D^*) < \infty$$

$$\forall m = 1, \ldots, M, \; 0 < A \cdot A' \cdot 2^{2jh_m} \leq \lambda_m(S(2^j)) \leq B \cdot B' \cdot 2^{2jh_m} < \infty$$

$\Rightarrow$ $\forall m = 1, \ldots, M, \; \log_2 \lambda_m(S(2^j))) \to 2h_m \cdot j$
$\Rightarrow$ $W \neq I_M$ does not create difficulties compared to $W \equiv I_M$
$\Rightarrow$ key step in proof: $B_{W,\Sigma}(0)$ has bounded eigen values

Sketch of Proof - 6

- $M = 2$
  $$\lambda_1(2^j) = 2\frac{\det(\mathbf{E}S)}{\mathbf{E}S_{22}(1)}\left(2^{2jh_1} + \frac{\mathbf{E}S_{22}(1)}{F(\mathbf{E}S_{11}(1),\mathbf{E}S_{22}(1))}2^{2j(h_1-h_2)}\right)$$

- Scaling range $(j_1(n), j_2(n)$
  $$(j_1(n), j_2(n) = (j_1^0 + f(n), (j_2^0 + f(n))$$
  $$\beta \log_2 n \leq f(n) \leq (1-\epsilon)\log_2 n,\ \epsilon > 0$$
  $$\beta = \frac{1}{1+2\max(h_1, \min_{1\leq m<m'\leq M}(h_{m'}-h_m))}$$
  Bias-Variance trade-off

◂ back