Testing fractal connectivity in multivariate long memory processes

Herwig Wendt, Antoine Scherrer, Patrice Abry, Sophie Achard

TESTING FRACTAL CONNECTIVITY IN MULTIVARIATE LONG MEMORY PROCESSES

H. Wendt, A. Scherrer, P. Abry,

S. Achard

Physics Dept., ENS Lyon, CNRS,
46, Allée d'Italie 69364 Lyon Cedex 07, France
firstname.lastname@ens-lyon.fr

GIPSA Lab, UMR CNRS 5216,
Grenoble, France
sophie.achard@gipsa-lab.inpg.fr

ABSTRACT

Within the framework of long memory multivariate processes, *fractal connectivity* is a particular model, in which the low frequencies (coarse scales) of the interspectrum of each pair of process components are determined by the autospectra of the components. The underlying intuition is that long memories in each components are likely to arise from a same and single mechanism. The present contribution aims at defining and characterizing a statistical procedure for testing actual fractal connectivity amongst data. The test is based on Fisher's Z transform and Pearson correlation coefficient, and anchored in a wavelet framework. Its performance are analyzed theoretically and validated on synthetic data. Its usefulness is illustrated on the analysis of Internet traffic Packet and Byte count time series.

Index Terms— Fractal connectivity, Long memory, Wavelet transform, Statistical test, Internet traffic

1. INTRODUCTION

Sensor network deployment is nowadays common for system monitoring in many different applications, such as medicine, biology, environment, ... to name but a few (cf. e.g., [1] and reference therein). Therefore, data to be analyzed often consist of multivariate time series, conveying in a potentially redundant and correlated manner the information practitioners are interested in. Measuring the amount of information shared amongst such data, or their (inter-)correlation levels (or functions), are often key issues in multivariate data analysis and processing. Moreover, in a large number of applications, the time series to be analyzed are characterized by long range dependence [2]: Their autocorrelation functions have extremely slow (algebraic power law type) decay, significantly complicating analyses aiming at establishing whether - and to which extent - the different time series embody identical or complementary information. A fundamental question arising from the analysis of long memory multivariate data is whether there is a *unique mechanism* in the system controlling long range dependence on the different data components, or whether there are different mechanisms at work, producing *unconnected* long memory properties. The goal of the present contribution is to propose a statistical test providing practitioners with elements of answer to such questions. Our study is based on the so called *fractal connectivity* model, recently introduced in [3] and consisting of multivariate long memory time series (recalled in Section 2): *Fractal* is used to refer to long memory while *connectivity* implies that the interspectrum is proportional to the product of the autospectra in the limit of coarse analysis scales (equivalently, low frequencies), i.e, the coherence function goes to a non zero constant in the limit $|f| \rightarrow 0$. It is now well-known that long range dependence phenomena are relevantly and accurately analyzed in a wavelet framework [4] (briefly re-sketched in Section 3). Therefore,

to test the fractal connectivity model, we propose a statistical procedure, based on the discrete wavelet transform coherence function and Fisher's Z transform. It is explained and defined in Section 4. Its performance are devised analytically and validated numerically by application to synthetic bivariate long memory time series with and without fractal connectivity (cf. Section 5). It is then applied to Internet traffic packet and byte count time series, collected very recently on a major transpacific backbone (cf. Section 6).

2. BIVARIATE FRACTAL CONNECTIVITY

Long memory. *Long memory* (LM), or long range dependence, is defined, for a process X, as a power law behavior of its spectrum $\Gamma_X(f)$ at the origin [2]:

$$\Gamma_X(f) \sim C|f|^{-\alpha}, \ |f| \rightarrow 0, \ \text{with } 0 < \alpha < 1. \quad (1)$$

This property has been widely observed for many different data in various research domains and its relevant analysis is important because it is known to strongly impair parameter estimation and to degrade the performance of a system, e.g., the amount of buffer needed on an Internet link.

Bivariate long memory model. For simplicity of notation, and without loss of generality, we restrict presentation here to the bivariate case only. The proposed test can be straightforwardly extended to the multivariate case by considering time series pairwise. Let $\mathbf{Z} = \{\mathbf{Z}(t)\}_{t\in\mathbb{Z}} = \{[X(t), Y(t)]\}_{t\in\mathbb{Z}}$ be a real-valued bivariate discrete time series. Following [3], $\mathbf{Z}$ is called a *bivariate long memory process* with parameters α_X, α_Y, α_1 and α_2 if its N-th order difference process $\tilde{\mathbf{Z}}(t) = \delta^N \mathbf{Z}(t)$ is stationary and has spectral and inter-spectral densities ($-\pi \leq f \leq \pi$), where the $\Omega.$ consist of arbitrary positive multiplicative factors:

$$\Gamma_{\tilde{W}}(f) = \Omega_{\tilde{W}} \left|1 - e^{-jf}\right|^{-2\alpha_W} \Gamma_{\tilde{W}}^*(f), \quad W = X \text{ or } Y \quad (2)$$

$$\Gamma_{\tilde{X}\tilde{Y}}(f) = \Omega_{12} \left(1 - e^{-jf}\right)^{-\alpha_1} \left(1 - e^{jf}\right)^{-\alpha_2} \Gamma_{12}^*(f). \quad (3)$$

The parameters $\alpha_{(.)}$ are confined to the range $[0, 0.5]$. The functions $\Gamma_{(.)}^*(f)$ are non-negative, symmetric, with limit 1 at the origin, hence modeling *short memory* (SM) properties at high frequencies, without affecting the spectral densities around the origin. Let us define $\alpha_{XY} = \alpha_1 + \alpha_2$. By definition, the coherence function: $C_{\tilde{X}\tilde{Y}}(f) = \frac{|\Gamma_{\tilde{X}\tilde{Y}}(f)|}{\sqrt{\Gamma_{\tilde{X}}(f)\Gamma_{\tilde{Y}}(f)}}$, has to be between 0 and 1. Also, it behaves asymptotically, in the limit $f \rightarrow 0$, as [3]:

$$C_{\tilde{X}\tilde{Y}}(f) \sim_{f\rightarrow 0} C_0|f|^{-(\alpha_{XY} - \alpha_X - \alpha_Y)}. \quad (4)$$

This shows that the model is well defined only if $\alpha_{XY} \leq \alpha_X + \alpha_Y$ and that $C_0 = |\Omega_{12}|/\sqrt{\Omega_{\tilde{X}}\Omega_{\tilde{Y}}}$ is a constant, controlling the global level of correlation of the series X and Y.

Fractal connectivity. *Fractal connectivity* is theoretically defined as the special case where $C_0 \neq 0$ and $C_{\tilde{X}\tilde{Y}}(f)$ exactly reduces to a non-zero constant over a range of coarse scales (low frequencies). Equivalently, this implies, $C_0 \neq 0$ and:

$$\alpha_{XY} = \alpha_X + \alpha_Y. \tag{5}$$

Essentially, the test described below aims at testing this equality. The intuition underlying fractal connectivity is that a same and single mechanism in the system uniquely controls the independent and joint LM properties of the multivariate data components. This test can hence avoid practitioners the burden of erroneously searching for different causes for LM.

3. DISCRETE WAVELET TRANSFORM

Discrete wavelet transform. A mother wavelet $\psi_0(t)$ is a reference pattern with narrow supports in both time and frequency domains. It is characterized by its number of vanishing moment $N_\psi \geq 1$: $\forall k = 0, 1, \ldots, N_\psi - 1, \int_{\mathbb{R}} t^k \psi_0(t) dt \equiv 0$ and $\int_{\mathbb{R}} t^{N_\psi} \psi_0(t) dt \neq 0$. Also, it is such that the $\{\psi_{j,k}(t) \equiv 2^{-j/2} \psi_0(2^{-j}t - k), j \in \mathbb{N}, k \in \mathbb{N}\}$ form a basis of $L^2(\mathbb{R})$. The discrete wavelet transform (DWT) coefficients of X are defined as: $d_X(j,k) = \langle X, \psi_{j,k} \rangle$. For further details, readers are referred to e.g., [5].

Stationary processes. Let $\tilde{X}$ and $\tilde{Y}$ denote second order stationary processes. It is straightforward to show that [4]:

$$\mathbb{E}d_{\tilde{W}}(j,k)^2 = \int \Gamma_{\tilde{W}}(f) 2^j |\Psi_0(2^j f)|^2 df, \ \ \tilde{W} = X \text{ or } Y, \tag{6}$$

$$\mathbb{E}d_{\tilde{X}(j,k)} d_{\tilde{Y}(j,k)} = \int \Gamma_{\tilde{X}\tilde{Y}}(f) 2^j |\Psi_0(2^j f)|^2 df, \tag{7}$$

where Ψ_0 stands for the Fourier transform of ψ_0 and $\mathbb{E}$ for the mathematical expectation. Following [4, 6], relevant wavelet based estimators for the auto- and inter-spectra of $\tilde{X}$ and $\tilde{Y}$ are defined as:

$$S_W(2^j) = \frac{1}{n_j} \sum_{k=1}^{n_j} d_W(j,k)^2, \quad W = \tilde{X} \text{ or } \tilde{Y}, \tag{8}$$

$$S_{\tilde{X}\tilde{Y}}(2^j) = \frac{1}{n_j} \sum_{k=1}^{n_j} d_{\tilde{X}}(j,k) d_{\tilde{Y}}(j,k), \tag{9}$$

where n_j is the number of coefficients available at scale 2^j. Qualitatively, the scale 2^j acts as the inverse of the frequency, $f \sim f_0/2^j$, with f_0 a constant depending on ψ_0.

Bivariate long memory processes. For bivariate LM processes **Z**, on condition that $N_\psi > N$, the wavelet coefficients d_X and d_Y at scale j form stationary sequences, and Eqs. (6-7) translate to: $\mathbb{E}d_X(j,k)^2 \sim c_X 2^{2j(\alpha_X+N)}$, $\mathbb{E}d_Y(j,k)^2 \sim c_Y 2^{2j(\alpha_Y+N)}$ and $\mathbb{E}d_X(j,k)d_Y(j,k) \sim c_{XY} 2^{2j(\alpha_{XY}+N)}$, when $2^j \rightarrow +\infty$. Also, it has been proven that the $d_X(j,k)$ and $d_Y(j,k)$ are freed from LM, so that the time averages S_X, S_Y, S_{XY} provide efficient and robust estimators of the spectra of **Z**, and of their power law exponents [4]. This implies that the *wavelet coherence function* [6] behaves, in the limit of coarse scales, as (with $\gamma_0 = c_{XY}/\sqrt{c_X c_Y}$):

$$\hat{\gamma}_{XY}(2^j) = \frac{S_{XY}(2^j)}{\sqrt{S_X(2^j)S_Y(2^j)}} \simeq \gamma_0 2^{j(\alpha_{XY}-\alpha_X-\alpha_Y)}. \tag{10}$$

Fractal connectivity. If fractal connectivity, Eq. (5), is valid, Eq. (10) above implies that $\hat{\gamma}_{XY}(2^j)$ takes a quasi constant non zero value over a range of coarse scales $2^j \geq 2^{J_1}$:

$$\hat{\gamma}_{XY}(2^j) \simeq \gamma_0 \neq 0, \tag{11}$$

while it decreases to 0 as $\gamma_0 2^{j(\alpha_{XY}-\alpha_X-\alpha_Y)}$ otherwise. This serves as the key ingredient for the design of a test for fractal connectivity.

4. TESTING FRACTAL CONNECTIVITY

Test formulation. The $\hat{\gamma}_{XY}(2^j)$ can be read as the Pearson product-moment correlation coefficient of the series $d_X(j,\cdot)$ and $d_Y(j,\cdot)$. It is known that, for many distributions F, $d_X(j,\cdot) \stackrel{d}{\sim} F$, the Fisher's Z statistic $\hat{z}_{XY}(2^j)$ of $\hat{\gamma}_{XY}(2^j)$ is asymptotically Normal (e.g. [7]):

$$\hat{z}_{XY}(2^j) = \frac{1}{2} \ln \frac{1+\hat{\gamma}_{XY}(2^j)}{1-\hat{\gamma}_{XY}(2^j)} \stackrel{d}{\sim} \mathcal{N}(z_{XY}(2^j), \sigma(2^j)), \tag{12}$$

with $z_{XY}(2^j) = \frac{1}{2} \ln \frac{1+\gamma_{XY}(2^j)}{1-\gamma_{XY}(2^j)}$ and variance $\sigma^2(2^j) = \frac{1}{n_j-3}$, where $\gamma_{XY}(2^j) = \mathbb{E}d_X(j,k)d_Y(j,k)/\sqrt{\mathbb{E}d_X(j,k)^2 \mathbb{E}d_Y(j,k)^2}$. Therefore, testing fractal connectivity can be formulated as a test of the *equality of means of Gaussian r.v.s* with known but different variances, i.e., of the null hypothesis:

$$H_0: z_{XY}(2^{J_1}) \equiv z_{XY}(2^{J_1+1}) \equiv \cdots \equiv z_{XY}(2^{J_2}), \tag{13}$$

where the scale range $j \in [J_1, J_2]$ is discussed below. Let $J = J_2 - J_1 + 1$. The test statistic for the UMPI test of equality of means of Gaussian r.v.s is given by [8]:

$$\hat{V}_J = \sum_{j=J_1}^{J_2} \frac{1}{\sigma^2(2^j)} \left(\hat{z}_{XY}(2^j) - \frac{\sum_{j=J_1}^{J_2} \hat{z}_{XY}(2^j)/\sigma^2(2^j)}{\sum_{j=J_1}^{J_2} 1/\sigma^2(2^j)} \right)^2. \tag{14}$$

Under H_0, idealizing the quasi-decorrelation of the wavelet coefficient into exact independence [4], one expects $\hat{V}_J$ to follow a $\chi^2_{(J-1)}$ distribution. Consequently, the $(1-\alpha)$ significance test for fractal connectivity can be formulated as:

$$\hat{d}_J = 1 \text{ if } \hat{V}_J > C_\chi, \ \hat{d}_J = 0 \text{ otherwise}, \tag{15}$$

where C_χ is the upper $(1-\alpha)$ percentile of the $\chi^2_{(J-1)}$ distribution. Similarly, the p-value of the observed test statistic $\hat{V}_J$ is given by: $\hat{p}_J = 1 - \tilde{\chi}^2_{(J-1)}(\hat{V}_J)$, where $\tilde{\chi}^2_{(J-1)}$ denotes the cumulative $\chi^2_{(J-1)}$ distribution function.

Power of the test. When H_0 is not true, $\hat{V}_J$ follows a non-central $\chi^2_{(J-1),V_J}$ distribution, where V_J is given by Eq. (14) with $z(j)$ replacing $\hat{z}(j)$. This enables to evaluate explicitly the power of the test against a specific alternative hypothesis $\alpha_{XY} - \alpha_X - \alpha_Y < 0$.

Scale range. Selecting the range of scales $j \in [J_1, J_2]$ where to perform the test results from a standard trade-off: J_1 needs to be chosen large enough so that SM (controlled by the $\Gamma^*_{(.)}$) no longer contribute (Type I error); however a too large J_1 decreases effective sample size and the power of the test (Type II error). J_2 is naturally limited by the available sample size ($J_2 \simeq \log_2 n$).

5. TEST PERFORMANCE

Numerical simulations. To evaluate the test performance, we apply it to a large number $N_{MC} = 1024$ of realizations of length $n = 2^{18}$ of Gaussian bivariate 2nd order stationary (hence $N = 0$, cf. Section 2) long memory processes, with prescribed auto- and inter-spectra according to Eqs. (2-3), implemented by ourselves following Chambers' algorithm [9]. The constant C_0 is varied within $0.5 \leq C_0 \leq 0.9$, implying a significant global correlation between X and Y, be they fractally connected or not. Parameters are set to $\alpha_X = 0.1$, $\alpha_Y = 0.3$. Under H_0, $\alpha_{XY} = 0.4$, while under H_1,
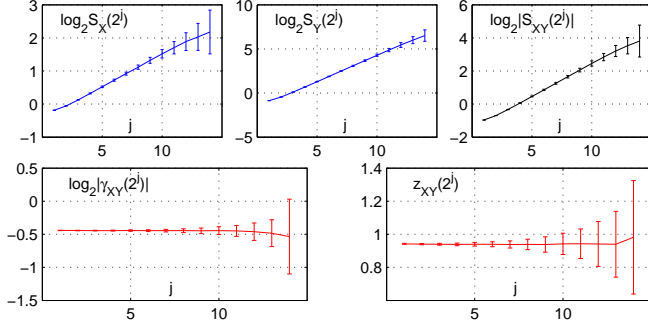
Fig. 1: **Wavelet spectra.** Average (and 95% CI for) wavelet spectra and inter-spectrum of X and Y (top), wavelet correlation and Fisher Z statistics (bottom) under H_0 without SM ($C_0 = 0.7$).
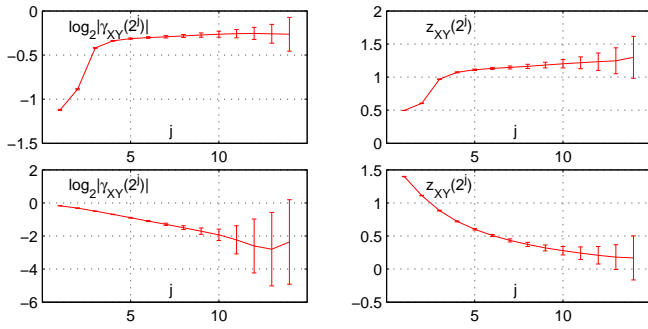

Fig. 2: **Short memory and alternative hypothesis.** Average (and 95% CI for) wavelet correlation (left column) Fisher's Z statistic (right column) as a function of scales ($C_0 = 0.7$) under H_0 with ARMA(1,1) SM (top) and under H_1 (bottom, $\alpha_{XY} = 0.2$).

α_{XY} is varied from 0 to $\alpha_X + \alpha_Y = 0.4$, so as to evaluate test powers. Short memory (SM) properties are modeled with ARMA(1,1) processes (parameters $\{0.4, -0.3\}$ for X and $\{-0.2, 0.1\}$ for Y). The significance level is set to $\alpha = 0.1$. Test performance are assessed by the mean rejection rates $\bar{d}_J = \hat{\mathbb{E}}_{MC}\hat{d}_J$ and mean p-values $\bar{p}_J = \hat{\mathbb{E}}_{MC}\hat{p}_J$, where $\hat{\mathbb{E}}_{MC}$ stands for the mean over N_{MC} Monte Carlo realizations: Ideally, under H_0, $\bar{d}_J$ should reproduce the preset significance level α, whereas the p-value should be uniformly distributed on $[0, 1]$, hence $\bar{p}_J$ should equal 0.5; Under H_1, the test should reject H_0, hence the larger (smaller) $\bar{d}_J$ ($\bar{p}_J$), the better.

Wavelet spectra and coherence, Fisher Z statistics. Fig. 1 shows, under H_0, the average over realizations (and corresponding 95% asymptotic confidence intervals (CI)) of the (log of the absolute val-
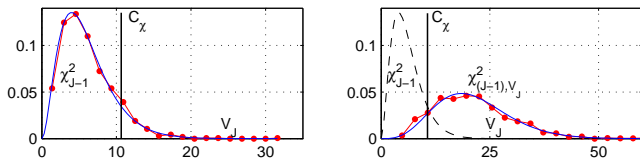

Fig. 3: **Test statistic.** Histogram of test statistic $\hat{V}_J$ (points, $C_0 = 0.7$) and the corresponding theoretical distribution (line), under H_0 (left) and H_1 (right, $\alpha_{XY} = 0.2$). The vertical line indicates the 10% significance critical value C_χ.

[J_1, J_2] = [7, 13] - 10% significance					
Test decisions - mean rejection rates $\bar{d}_J$ (in %)					
α_{XY}	**0.4**	0.35	0.3	0.25	0.2
$C_0 = 0.5$	**9.4**	22.7	42.9	54.9	56.8
$C_0 = 0.7$	**10.3**	51.9	83.0	89.2	89.3
$C_0 = 0.9$	**8.7**	99.3	99.9	99.9	99.9
Mean p-value $\bar{p}_J$					
α_{XY}	**0.4**	0.35	0.3	0.25	0.2
$C_0 = 0.5$	**0.51**	0.38	0.25	0.18	0.16
$C_0 = 0.7$	**0.50**	0.18	0.06	0.04	0.04
$C_0 = 0.9$	**0.50**	0.00	0.00	0.00	0.00

Table 1: **Test performance.** Mean test decisions (top) and p-values (bottom) for different values of C_0 and α_{XY}.

$\alpha_{XY} = 0.4$ - [J_1, J_2] = [$\cdot$, 13] - 10% significance						
Gaussian no short memory - $C_0 = 0.7$						
J_1	5	6	7	8	9	10	11
$\bar{d}_J$	11.0	9.6	9.1	8.6	9.3	9.0	8.5
$\bar{p}_J$	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Gaussian ARMA(1,1) - $C_0 = 0.7$						
J_1	5	6	7	8	9	10	11
$\bar{d}_J$	66.7	33.5	21.5	14.4	11.2	8.9	9.2
$\bar{p}_J$	0.12	0.29	0.38	0.45	0.50	0.51	0.52

Table 2: **Scale range.** Mean test decisions (in %) and p-values for different values of J_1, without (top) and with (bottom) SM.

ues of the) wavelet (inter-)spectra (top) and coherence, and Fisher's Z statistics (bottom), when no SM are present (i.e., $\Gamma^*_{(\cdot)} \equiv 1$). It indicates that the estimates $\hat{\gamma}_{XY}(2^j)$ and $\hat{z}_{XY}(2^j)$ are quasi constant over the entire range of available scales 2^j, as predicted by the model. Fig. 2 illustrates $\hat{\gamma}_{XY}(2^j)$ and $\hat{z}_{XY}(2^j)$ under H_0 ($C_0 = 0.7$) when SM are present (top row), and under H_1 (bottom row, $\alpha_{XY} = 0.2$). The plots indicate that: i) Under H_0, the existence of SM has a clear impact on fine scales (below $2^j \leq 2^{J_1} = 2^7$). Yet, at coarse scales, both $\hat{\gamma}_{XY}(2^j)$ and $\hat{z}_{XY}(2^j)$ are quasi-constant; ii) Under H_1, both $\hat{\gamma}_{XY}(2^j)$ and $\hat{z}_{XY}(2^j)$ display a significantly non-constant behavior with scales 2^j. These preliminary investigations clearly validate the test as formulated in Eqs. (13-15).

Test performance: Significance. Tab. 1 summarizes mean test decisions and p-values when α_{XY} and C_0 are varied and no SM is present. Under H_0 ($\alpha_{XY} = 0.4$, left column), the targeted 10% significance level is closely reproduced and mean p-values are close to the expected value 0.5, regardless of the precise value of C_0, indicating that the test statistic $\hat{V}_J$ Eq. (14) accurately follows the predicted $\chi^2_{(J-1)}$ distribution. This is further confirmed in Fig. 3 (left), showing the histogram of $\hat{V}_J$ under H_0.

Test performance: Power. Tab. 1 (col. 3-6) clearly indicates that the test is powerful in rejecting H_0 when $\alpha_{XY} < \alpha_X + \alpha_Y$ and an alternative hypothesis is true: With increasing discrepancy between α_{XY} and $\alpha_X + \alpha_Y$, mean rejection rates (p-values) increase (decrease). Also, the larger C_0, the more powerful the test: For large global correlation $C_0 = 0.9$, the test rejects fractal connectivity with probability close to 1 already for discrepancy in exponents α_{XY} and $\alpha_X + \alpha_Y$ as small as 0.05. Fig. 3 (right) confirms that the test statistic $\hat{V}_J$ closely follows the predicted non central $\chi^2_{(J-1)}$ distribution under an alternative hypothesis.
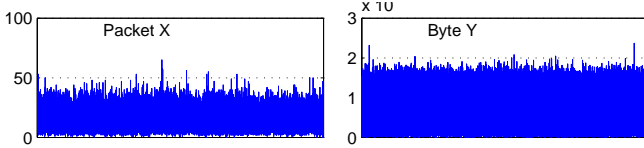
Fig. 4: **Internet time series.** Packet (left) and byte (right) counts as functions of time, $\Delta_0 = 10$ms.
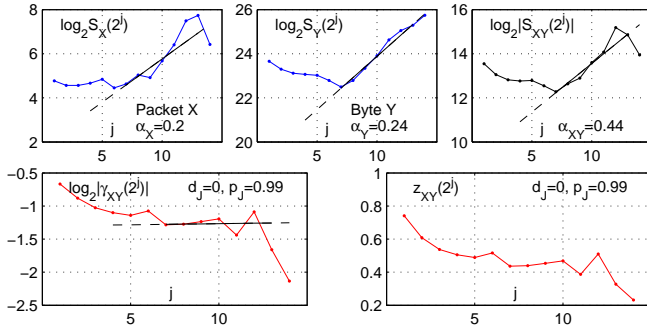

Fig. 5: **Wavelet spectra.** Wavelet (inter-)spectra for Packet and byte counts (top), wavelet coherence and Z statistics (bottom).

Scale range and short memory. Tab. 2 compares the impact of the choice of J_1 on test performance under H_0, without (top) and with (bottom) SM. Without SM, the targeted 10% significance level (and the corresponding $p = 0.5$ average p-value) are systematically well reproduced, regardless of the choice of J_1. In contrast, with SM, the scale range must be restricted to coarse scales $j \geq J_1 = 9$ to reproduce nominal performance, resulting in less powerful tests and increased difficulty in rejecting the fractal connectivity hypothesis.

6. INTERNET TRAFFIC TIME SERIES

Internet monitoring and security nowadays constitute major tasks and challenges, often achieved by means of traffic flows statistical characterization. Commonly, Internet traffic is analyzed as aggregated time series, consisting of either IP packet (Pkt) or bytes (Byt) counts within bins of selected duration Δ_0. It is naturally suspected that such time series should be correlated but their precise interrelation is still an open debate. Moreover, Internet traffic is well known to be characterized by a strong LRD property [4], and we intend here to illustrate the usefulness of the proposed test to analyze LRD jointly in Pkt or Byt counts and hence to contribute to the ongoing debate: Should practitioners concentrate on the analysis of Pkt or Byt counts? Data analyzed are part of the MAWI data set (1GB public repository available at http://mawi.wide.ad.jp/, cf. [10]). Results are reported here for data collected on March 3rd, 2006, between 7.30am and 7.45am (Tokyo time), on a transpacific OC3 backbone link (time series shown in Fig. 4). Fig. 5 indicates that both Pkt and Byt counts present LRD, at coarse scales $2^j \geq 2^{J_1} = 2^7$, i.e., roughly, for time lags larger than 0.125s. It also reveals a power law behavior for the wavelet inter-spectrum and a quasi-constant wavelet coherence function at coarse scales. When applied to data, the test (with a 10% targeted significance level) validate the hypothesis of fractal connectivity, with an output p-value of $p = 0.99$, indicating that data show absolutely no evidence for rejection. Equivalent results are obtained from numerous different time series within the MAWI data sets. This leads to conclude that

LRD in Pkt and Byt time series result from a single and same network mechanism and are hence not created from different network sources. However, for a number of data, wavelet inter-spectra differ significantly from the one shown here (while individual auto spectra do present LRD). When applied, the test indicates no correlation at coarse scales and significantly rejects fractal connectivity, suggesting that LRD in Pkt and Byt may result from different and independent causes. Manual inspections tend to indicate that such cases are related to the occurrence of anomalies in the traffic, be they legitimate or not. This requires further validation and is under current investigation. This opens new perspectives for Internet traffic monitoring and for automated anomaly detection, where inter-spectra are rarely considered, despite the multivariate nature of traffic data.

7. DISCUSSION AND CONCLUSION

A statistical test for fractal connectivity in bivariate time series has been defined, analyzed and assessed. Its extension to multivariate data is straightforward, by considering each pair of data components. The test relies on wavelet based estimations of the autospectra, interspectra and coherence functions of the data. It has been shown to present satisfactory practical performance. To our knowledge, this is the first procedure for testing joint long memory properties of multivariate data practically available in the literature. It provides practitioners with elements of answers to questions regarding the independence or not of the mechanisms at work in the production of the data, notably of their long range dependence property. Preliminary attempts for the analysis of Internet traffic and various biological and biomedical data indicate promising perspectives.

8. REFERENCES

[1] Z.-Q. Luo, M. Gastpar, J. Lui, and A. Swami, "Distributed signal processing in sensor networks," *IEEE Signal Proc. Mag.*, vol. 23, no. 4, pp. 14–15, 2006.

[2] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.

[3] S. Achard, D.S. Bassett, A. Meyer-Lindenberg, and E. Bullmore, "Fractal connectivity of long-memory networks," *Phys. Rev. E*, vol. 77, no. 3, pp. 036104, 2008.

[4] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "Multiscale nature of network traffic," *IEEE Signal Proc. Mag.*, vol. 19, no. 3, pp. 28–46, 2002.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[6] B. Whitcher, P. Guttorp, and B. D. Percival, "Wavelet analysis of covariance with application to atmospheric time series," *J. Geophys. Res. [Atmos.]*, vol. 105, pp. 14941–14962, 2000.

[7] D.L. Hawkins, "Using u statistics to derive the asymptotic distribution of Fisher's z statistic," *American Statistical Association*, vol. 43, no. 4, pp. 235–237, 1989.

[8] E.L. Lehmann, *Testing Statistical Hypotheses*, Wiley, New York, 1959.

[9] M.J. Chambers, "The simulation of random vector time series with given spectrum," *Mathematical and Computer Modelling*, vol. 22, no. 6, pp. 1–6, July 1995.

[10] K. Cho, K. Mitsuya, and A. Kato, "Traffic data repository at the WIDE project," in *USENIX 2000 Annual Technical Conference: FREENIX Track*, June 2000, pp. 263–270.