INFINITELY DIVISIBLE CASCADE ANALYSIS OF NETWORK TRAFFIC DATA

D. Veitch

Software Engineering Research Centre
GPO Box 2476V, Melbourne
Victoria 3001, - Australia
darryl@serc.rmit.edu.au

P. Abry, P. Flandrin and P. Chainais

CNRS, UMR 5672 - Laboratoire de Physique
Ecole Normale Supérieure de Lyon
46, Allée d'Italie 69364 LYON Cedex 07, France
pabry,flandrin,pchainai@ens-lyon.fr

ABSTRACT

Infinitely Divisible Cascades are a model class recently introduced in the field of turbulence to describe the statistics of velocity fields. In this paper, using a wavelet reformulation of the cascades, we investigate their ability to analyze and model scaling properties of data and compare their fundamental ingredients to those of other scaling model classes such as self-similar and multifractal processes. We also propose an estimation procedure for the propagator or kernel of the cascades. Finally the cascade model is successfully applied to describe Internet TCP network traffic data, bringing new insights into their scaling properties and revealing a pitfall in existing techniques.

1. MOTIVATION

The ubiquity of scaling phenomena in natural and man-made systems is nowadays a well-recognized fact, the most prominent example being perhaps that of communication networks [8, 10]. Much work has been carried out for analyzing and modeling signals produced by such systems, and it has been amply demonstrated that wavelet-based tools happen to play a key role in this context [1]. Nevertheless, a number of issues still remain open, especially in terms of *model validation*. One of the central problems is that the idea of scaling is generally associated to some form of linearity in a well-chosen log-log diagram, a behaviour which has to be validated. Another crucial issue is that, in general, scaling (if any) only occurs in a given range of scales and, depending on this range, different models have to be advocated (e.g., long-range dependence at large scales *vs.* multifractality at small scales). It follows that estimating relevant scaling parameters is somehow linked to the choice of some *a priori* model whose validation may prove difficult. What we propose here is to circumvent this difficulty by (*i*) making use of a general and versatile model based on a notion of a *multiplicative cascade*, and (*ii*) validating its use and estimating its parameters on a statistical basis.

Partially supported by the CNRS grant TL97035, Programme Télécommunications and and the Bede Morris French Embassy Fellowship, 1999. Data generously supplied by Professor J. Cleary, WAND, University of Waikato, New Zealand, and special thanks to Jörg Micheel at WAND and Li Dong Huang of SERC for time series extraction.

More precisely, it has been shown [1] that wavelets can be considered as "matched" to self-similar processes in the sense that wavelet coefficients exactly reproduce, from scale to scale, the self-replicating statistical structure of such processes. In the very special case of a strictly self-similar process $\{X(t), t \in \mathbb{R}\}$ of index $H \in]0,1[$, and for a proper normalization of its wavelet coefficients $\{d_X(j,k), (j,k) \in \mathbb{Z}^2\}$, the key scaling relation is that $\mathbb{E}|d_X(j,k)|^q \propto \exp\{qH\ln(2^j)\}$ for any $q \in \mathbb{R}$. The main quality of self-similarity (SS) with respect to scaling analysis is its simplicity: the moments of the wavelet coefficients all behave as power-laws of the scale $a = 2^j$, controlled by a single scaling exponent, the self-similarity parameter H. This simplicity is also its major drawback since the model may not be versatile enough to model actual empirical data. It has therefore been proposed that the unique scaling parameter H be replaced by a collection of exponents $H(q)$, allowing much greater freedom to fit to data, in other words that $\exp\{qH\ln(2^j)\} \to \exp\{H(q)\ln(2^j)\}$. This model will be referred to as *multi-scaling* (MS). (One can remark that, when such a situation is encountered in the small scales limit ($j \to -\infty$), $H(q)$ is nothing but the classical ζ_q function of the multifractal formalism [11].) The main limitation of this model, however, is that power-laws for the moments must be observed. A second level of generalization is therefore possible, according to $\exp\{H(q)\ln(2^j)\} \to \exp\{H(q)n(2^j)\}$, where $n(.)$ does not necessarily reduce to the logarithm function, with is associated to a notion of strict scale invariance. This corresponds to what is called an *infinitely divisible cascade* (IDC) model. While giving up the requirement that moments behaves as power-laws of scales, such a model maintains, however, a fundamental feature in common with SS and MS : the *separability* of the moments structure in the variables q (order of the moment) and 2^j (scale). In summary, we have the following relations between SS, MS and IDC :

$$\begin{array}{lllll} \text{SS} & \mathbb{E}|d_X(j,k)|^q = & C_q(2^j)^{qH} & = C_q\exp(qH\ln(2^j)) \\ \text{MS} & \mathbb{E}|d_X(j,k)|^q = & C_q(2^j)^{H(q)} & = C_q\exp(H(q)\ln(2^j)) \\ \text{IDC} & \mathbb{E}|d_X(j,k)|^q = & & = C_q\exp(H(q)n(2^j)). \end{array} \quad (1)$$

Section 2 will present the IDC model in more detail and will then address the question of validating the model and estimating its two ingredients, the functions $H(q)$ and $n(a)$.

Section 2.2 will finally illustrate the usefulness of the approach in an application to Internet data traffic.

2. INFINITELY DIVISIBLE CASCADES

2.1. Definitions

The concept of IDC was first introduced by B. Castaing in [3, 4] and rephrased in the wavelet framework in [2]. We now briefly recall its intuition, definition, consequences and relations to other models. Starting again from the self-similar case, one can write the probability density function (pdf) of the wavelet coefficients at scale $a = 2^j$, as a dilated version of the pdf of the wavelet coefficients at a larger scale a': $p_a(d) = (1/\alpha_0)\, p_{a'}(d/\alpha_0)$ where the dilation factor is unique : $\alpha_0 = (a/a')^H$. In the cascade model, the key ingredient is that there is no longer a unique factor but a collection of dilation factors α ; consequently p_a will result from a weighted sum of dilated incarnations of $p_{a'}$:

$$p_a(d) = \int G_{a,a'}(\ln\alpha)\,\frac{1}{\alpha}\,p_{a'}\left(\frac{d}{\alpha}\right) d\ln\alpha.$$

The function $G_{a,a'}$ is called the kernel or the *propagator* of the cascade. Obviously, if $G_{a,a'}$ is a Dirac function, $G_{a,a'}(\ln\alpha) = \delta(\ln\alpha - H\ln(a/a'))$, IDC reduces to SS, therefore understood as a special case. The definition of the cascade above shows that the pdf's of $\underline{p}_a$ and $\underline{p}_{a'}$ of the wavelet log-coefficients $\ln|d|$ are related by a convolution with the propagator :

$$\begin{aligned}\underline{p}_a(\ln\alpha) &= \int G_{a,a'}(\ln\alpha)\,\underline{p}_{a'}(\ln|d| - \ln\alpha)\, d\ln\alpha \\ &= (G_{a,a'} * \underline{p}_{a'})(\ln\alpha). \qquad (2)\end{aligned}$$

If cascades exist between scales a'' and a' and between scales a and a'', then a cascade between scales a and a' exists, and the corresponding propagator results from the convolutions of the two propagators: $G_{a,a'} = G_{a,a''} * G_{a'',a'}$. Infinite divisibility (also called continuous self similarity) means that no scale between a and a' plays any characteristic role (i.e., a'' in the above statement can be any scale between a and a'). Infinite divisibility therefore implies that the propagator consists of an elementary function G_0 convolved with itself a number of times, where that number depends on a and a':

$$G_{a,a'}(\ln\alpha) = [G_0(\ln\alpha)]^{*\left(n(a)-n(a')\right)}.$$

Using the Laplace transform $\tilde{G}_{a,a'}(q)$ of $G_{a,a'}$, this can be rewritten as $\tilde{G}_{a,a'}(q) = \exp\{H(q)(n(a) - n(a'))\}$, with $H(q) = \ln\tilde{G}_0(q)$ and $a := 2^j$; this implies that $\mathbb{E}|d_X(j,k)|^q = C_q \exp\{H(q)n(a)\}$, thus validating eq. (1). The main consequences of IDC (in other words, of the separability of the variables q and a), read therefore :

$$\ln\mathbb{E}|d_X(j,k)|^q = H(q)n(a) + K_q \qquad (3)$$

$$\ln\mathbb{E}|d_X(j,k)|^q = \frac{H(q)}{H(p)}\ln\mathbb{E}|d_X(j,k)|^p + \kappa_{q,p}. \qquad (4)$$

This last equation implies that moments behave as power-laws relative to each other. Such relations are sometimes called, in turbulence mainly, "extended self-similarity". Note that, in the relation (3) above, there is an arbitrary element, indeed:

$$\begin{aligned}H(q)n(a) + K_q &= \left(\tfrac{H(q)}{\beta}\right)(\beta n(a) + \gamma) + (K_q - \tfrac{H(q)\gamma}{\beta}) \\ &= H'(q)n'(a) + K'_q\end{aligned}$$

where $\beta \neq 0$ and γ are arbitrary constants. It clearly indicates that H is defined up to a multiplicative constant while n is defined up to multiplicative and additive constants.

If it is moreover required that the function $n(a) \equiv \ln a$, the IDC is called *scale invariant* (SIIDC) and this implies that :

$$\tilde{G}_{a,a'}(q) = (a/a')^{\ln\tilde{G}_0(q)} \text{ and } \mathbb{E}|d_X(j,k)|^q = (2^j)^{\ln\tilde{G}_0(q)},$$

proving that SIIDC reduces to MS. If, moreover, the power-laws are observed in the limit of small scales ($a = 2^j \to 0$), then, MS is equivalent to *multifractal* (MF), and the exponents ζ_q—from which the Legendre MF spectrum is obtained through a Legendre transform [11]—are related to the propagator through $\zeta_q = H(q) = \ln\tilde{G}_0(q)$. In this framework, MF is therefore understood as a special case of IDC. The stochastic multiplicative cascades introduced by Mandelbrot [9], constitute the canonial example of such situations. In a SIIDC, one can also inquire as to whether ζ_q is a linear function of q or not, in which case the cascade reduces to the even more special case of SS. It is, therefore, natural to consider the function $\zeta_q/q = H(q)/q$ and test its constancy (see next section).

2.2. Analysis

When one intends performing a cascade type analysis of actual data, the key point is the estimation of the $\mathbb{E}|d_X(j,k)|^q$. Due to the fact that wavelet coefficients of wide classes of scaling processes are stationary and exhibit weak statistical dependences [1, 7], $S_q(j) = 1/n_j \sum_k^{n_j} |d_X(j,k)|^q$, an average over time, will constitute reliable estimators of the $\mathbb{E}|d_X(j,k)|^q$ (n_j is the number of available coefficients at scale 2^j). From the properties of the wavelet coefficients and under mild hypothesis, it is moreover possible to estimate $\mathbb{E}\ln S_q(j)$ and $\mathrm{Var}\ \ln S_q(j)$. For instance, when the random variable $|d_X(j,k)|^q$ has finite variance, we use, in the estimation, the asymptotic results:

$$\begin{aligned}\mathbb{E}\ln S_q(j) &\simeq \ln\mathbb{E}|d_X(j,k)|^q + \mathrm{Cste}/n_j \\ \mathrm{Var}\ \ln S_q(j) &\simeq \mathrm{Cste}/n_j.\end{aligned}$$

To test the adequacy of the IDC model on data, one can check that its main consequence, extended self-similarity (relation (4)), is verified. The hypothesis testing procedure therefore consists in plotting the $\ln S_q(j)$ versus $\ln S_p(j)$ diagrams and testing whether they are straight lines or not. To be meaningful, this test must be performed taking into account the variance of the $\ln S_q(j)$. The next step in the analysis is to estimate the parameters of the cascade, that is, identify its propagator, or equivalently, the functions $H(q)$ and $n(2^j)$). The $H(q)$ are estimated from weighted linear regressions in the $\ln S_q(j)$ versus $\ln S_p(j)$ plots,

$$\hat{H}(q)/H(p) = \mathrm{slope}_{q,p}, \qquad (5)$$

the weights of the linear fit being related to the Var $\ln S_q(j)$. To estimate n, as proposed in [5], we start again from the separability of the variables q and 2^j and (3), from which we propose the following estimators:

$$\begin{aligned} \hat{K}_q &= \langle \ln S_q(a) - \text{slope}_{q,p} \ln S_p(a) \rangle_a \\ \hat{n}(a) &= \frac{1}{H(p)} \left\langle \frac{1}{\text{slope}_{q,p}} \left(\ln S_q(a) - \hat{K}_q \right) \right\rangle_q + K_p, \end{aligned} \quad (6)$$

where $\langle . \rangle_q$ (resp., $\langle . \rangle_a$) means average over the values of q (resp., the values of a). In these estimation procedures, we naturally encounter again the arbitrary components in the definitions of H and n. These are removed by fixing p at some arbitrary q value, which in practice means that one selects arbitrary values for $H(p)$ and K_p. In other words, H can only be known up to a multiplicative constant $1/H(p)$, and n up to a multiplicative, $1/H(p)$, and additive, K_p, constant. From a practical point of view however, the choice of p is not entirely arbitrary. It should be chosen as either the smallest or the largest of the available q values. Indeed, the quantities $\ln S_q(j)$ and $\ln S_p(j)$ are not independent, although intuitively the farther from 1 the ratio q/p is, the weaker the dependences will be, an important feature for both model testing and estimation procedures.

This cascade analysis has been applied in turbulence leading to interesting conclusions [5], the next section illustrates the benefit of the cascade analysis for the study of computer network traffic data.

3. APPLICATION TO NETWORK TRAFFIC

The existence of scaling in telecommunications traffic is now well established, and recent work has shown the relevance of multifractal models [6, 12]. We now illustrate the benefits of an IDC approach by examining a two hour long set of TCP data collected on a link at the University of Auckland between 6pm - 8pm, Thursday July 8th 1999, by the WAND group at the University of Waikato. TCP is the protocol used for reliable data transfer over the Internet, including Web based data retrieval. An understanding of the traffic it generates is a key problem in modern networking. The capture hardware developed at WAND enables TCP/IP traffic carried by *asynchronous transfer mode* technology at 155 Mbits/s to be measured with $\sim 0.1\mu$s timestamp accuracy and no losses.

From a given set of raw data many different time series can be extracted. Here we consider a time indexed series $X(k)$ of length $n = 720,000$, corresponding to the number of new TCP connections arriving per 10 ms interval. A second order scaling analysis is presented in the top left plot in figure 1. Biscaling is clearly observed, that is two separate scaling ranges, 'small scales': $j \in [3, 8]$, and 'large scales': $j \in [8, 16]$, joined at the characteristic scale $j = 8$ (~ 2.56 s). This time scale is of roughly the same order of magnitude as round trip times in the Internet and may thus be related to network control feedback mechanisms, perhaps that of the TCP protocol itself as suggested in [6]. Biscaling with a similar breakpoint has been observed by the authors in *end-to-end* Internet data [1]. (Note that $j = \{1, 2\}$ are not included in the small range. This is associated with the fact that, as the link is lightly loaded, $X(k)$ is sparse, being 90.2% zeros, and therefore contains little information at these scales). Logscale Diagrams were examined at orders $q = \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6\}$, and similar results were found, for example $\zeta_6(j)$ is displayed in the top right in figure 1. In addition, informal tests for stationarity were performed by dividing the series into blocks and examining the Logscale Diagrams at different orders over each. The biscaling, as well as the estimated exponents, remain remarkably constant, and we therefore proceed under the assumption of stationarity. Since, over each range separately, scaling is observed at many orders, it is not unreasonable to apply an independent multiscale analysis over each range. The resulting ζ_q/q plots are shown in the bottom row of figure 1. Over the small scales (left hand plot) a horizontal alignment is clearly not observed, corresponding to a non-trivial multiscaling model. In contrast at large scales both the form of the curve and the larger confidence intervals suggest a simple multiscaling model, such as a self-similar model.

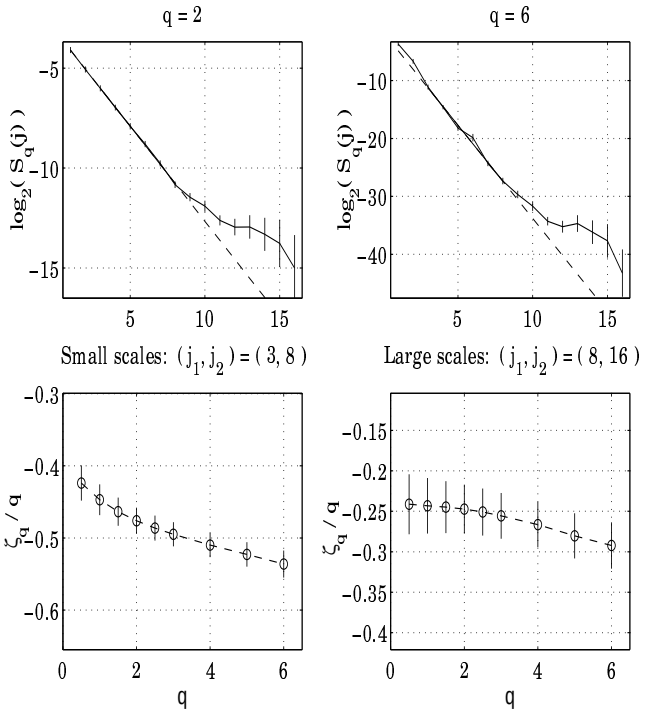

Figure 1: **Multiscale analysis, the ζ_q.** Biscaling is observed in both second order (top left) and higher order (right) Logscale Diagrams. A resulting multiscale analysis over small scales (bottom left) shows non-trivial multiscaling (curve is not horizontal) whereas at large scales (right) a trivial multiscaling is apparently observed.

Since multiscaling models are a special case of the IDC model, implicitly two scale invariant IDC cascades, one per range, have also been identified in the above analysis. A natural question is, are these part of one and the same cascade? To test this, we perform a cascade analysis over the combined 'full range': $j \in [3, 16]$. Over the same set of q values, $\{\ln S_q, \ln S_1\}(j)$ plots were examined. To within the confidence intervals, in each plot a well defined slope was found, confirming the hypothesis of separability, and thus

the presence of a single cascade over the full range. Examples for $q = \{3, 6\}$ are given in the top row of figure 2. The functions $H(q)$ and $n(a) = n(2^j)$ can then be estimated and appear in the bottom row, together with comparisons of results from the multiscale analyses. In the bottom right plot $n(a)$ is seen to **not** be $\ln(a)$, so indeed a single multiscale analysis over the full range is not possible. However $n(a)$ is piecewise logarithmic about the octave $j = 8$ (see figure 2, bottom right), justifying the multiscale analysis performed over the small and large ranges. In the bottom left plot $H(q)$ is compared with ζ_q/ζ_1 from the multiscale analyses, and close agreement is found for each range, consistent with the conclusion that a single cascade model underlies them both. The new fact that arises from the cascade analysis is that the functions $\zeta(q)$ from the two ranges are not independent but in fact equivalent, being simple multiples of each other, as could be checked from figure 1 and as is plotted on figure 2, bottom left. This can be explained using the IDC model by setting $n_s(a) = b_s \ln(a) + c_s$ (resp., $n_l(a) = b_l \ln(a) + c_l$) over the small (resp., the large) scales. Identifying $H(q) = \zeta(q)$ in the MS interpretation, indeed yields : $\zeta_s(q)/\zeta_l(q) = b_s/b_l = \text{cste}$. Thus the conclusions from figure 1, that the analyses are independent and the models of a different nature, is incorrect. This new insight eliminates a misguided application and arbitrary fitting of unrelated scaling models over the two ranges. Instead, it allows the full set of data across all scales to be used to estimate the key quantity $H(q)$, reducing estimation errors. Furthermore it clearly identifies the nature of the change at octave j, as being captured in $n(a)$ only, and not $H(q)$.

For sake of simplicity in this paper, results were presented concerning the number of new TCP connections per $10ms$ only. Similar cascade analyses have been performed on time series such as the durations of the TCP connections, the number of TCP connections active at $10ms$ intervals, the number of TCP packets per $10ms$, the number of TCP bytes per $10ms$, ... and similar conclusions drawn: the IDC describes the data very well, and biscaling is observed with a function $n(a)$ which in most cases is "piecewise log" (figure 2).

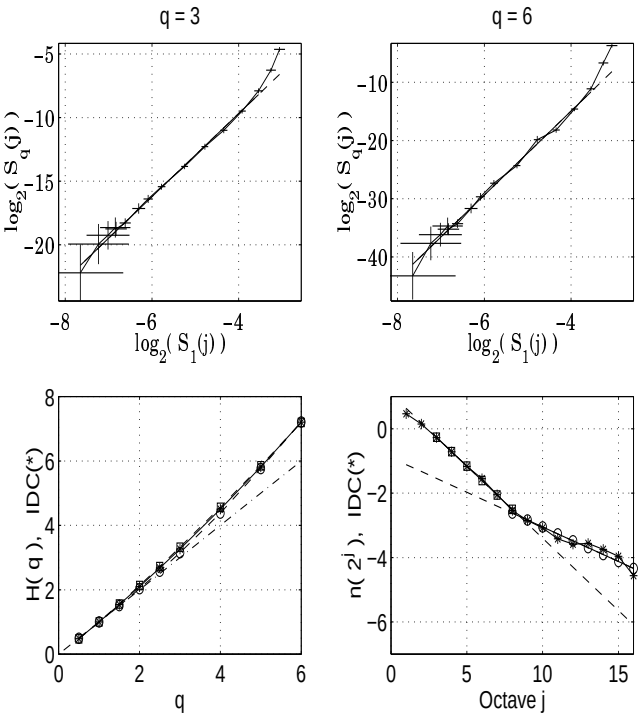

Figure 2: **ID Cascade analysis, and Comparison.** A common slope is seen in $(\ln S_3, \ln S_1)(j)$ (top left), and $(\ln S_6, \ln S_1)(j)$, (right), validating the IDC hypothesis over the full range. Excellent agreement (bottom left) is seen between estimates of $H(q)$ from the cascade $(*)$, and corresponding values from both the multiscaling model at small (squares) and large scales (circles), consistent with the 'piecewise log' nature of the estimate of $n(a)$ (bottom right).

4. REFERENCES

[1] P. Abry, P. Flandrin, M.S. Taqqu and D. Veitch. Wavelets for the analysis, estimation and synthesis of scaling data. To appear in [10]

[2] A. Arnéodo, J.F. Muzy, S.G. Roux, Experimental analysis of self-similar random cascade processes: application to fully developped turbulence, *J. Phys. II France*, 7:363–370, 1997.

[3] B. Castaing, Y. Gagne, E. Hopfinger, Velocity probability density functions of high Reynolds number turbulence. *Physica D*,46:177, 1990.

[4] B. Castaing. The temperature of turbulent flows. *J. Phys. II France*, 6:105–114, 1996.

[5] P. Chainais, P. Abry, and J.F. Pinton, "Intermittency and coherent structures in a turbulent flow: a wavelet analysis of joint pressure and velocity measurements," *Phys. Fluids*, Vol. 11, no 11, pp. 3524-3539, 1999.

[6] A.C. Gilbert, W. Willinger, A. Feldmann, Scaling analysis of random cascades, with applications to network traffic, *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications, 45(3):971–991, 1999.

[7] P. Gonçalvès and R. H. Riedi. Wavelet analysis of fractional Brownian motion in multifractal time. Proc. 17ème Colloque GRETSI, Vannes, France, 1999.

[8] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, On the self-similar nature of Ethernet traffic, Extended-Version, *IEEE/ACM Trans. on Networking*, 2:1–15, 1994.

[9] B.B. Mandelbrot, Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier, J. of Fluid Mech., 62(2):331–358, 1974.

[10] *Self-Similar Network Traffic and Performance Evaluation.* K. Park and W. Willinger, eds. Wiley Interscience, to appear, 1999.

[11] R. Riedi. Multifractal processes. 1999. preprint.

[12] R. Riedi, M.S. Crouse, V.J. Ribeiro, R.G. Baraniuk, A Multifractal Wavelet Model with Application to Network Traffic, *IEEE Trans. on Info. Theory*, Special Issue on Multiscale Statistical Signal Analysis and its Applications, 45(3):992–1018, April, 1999.