Fractals, Vol. 6, No. 4 (1998) 305–311

(max, +) DYNAMIC SYSTEMS FOR MODELING TRAFFIC WITH LONG-RANGE DEPENDENCE

FRANÇOIS CHAPEAU-BLONDEAU
Laboratoire d'Ingénierie des Systèmes Automatisés (LISA)
Université d'Angers, 62 avenue Notre Dame du Lac, 49000 ANGERS, FRANCE

Received July 13, 1998; Accepted September 18, 1998

Abstract

A simple first-order recurrence in a (max, +) dynamic system is numerically investigated and shown to exhibit statistical long-range dependence, characterized by slowly decaying aggregated variances and power-law evolutions of the autocorrelation and spectrum. We propose this model as a basis for a very parsimonious modeling of some long-range dependent processes such as data traffic.

1. INTRODUCTION

Several types of processes arising in complex systems of various kinds have recently been shown to exhibit fractal statistical properties characterized by long-range dependence and self-similarity. For instance, the results of Ref. 1 (and references there in) are concerned with highway traffic and report various quantities, both from models and measurements, that fluctuate like $1/f^{\alpha}$ noises, such as the flow of cars or the size of traffic jams. References 2–5 exhibit long-range correlations in texts or computer programs or files: usually a numerical code is assigned to each character and the series of characters forming the text or file is transformed into a random walk from which long-range correlations are quantified. Other important examples are provided by data traffics in communication networks, where long-range dependence has been measured on various quantities, such as the travelling times between nodes in the network,[6] or the number of data packets incoming on a node per unit time,[7–8] or the arrival or interarrival times of connections,[9] or the number of bytes to encode successive video frames in variable-bit-rate video traffic.[10]

Accurate modeling and simulation of such long-range dependent processes are key issues for analysis, control and performance evaluation of complex engineering systems like modern communication networks. Classical models like Poisson-based processes or linear models (such as autoregressive moving average (ARMA) processes) are known to be unable to generate long-range dependence.[7,9,11] More elaborate processes have been proposed for modeling long-range dependent evolutions, including fractional Gaussian noises and fractional ARIMA (autoregressive integrated moving average) processes.[7–8,12] Yet, the practical implementation of such processes to generate synthetic traces is usually uneasy and demanding on computational resources. Generating a single data point of the synthetic trace often requires processing a large number of auxiliary points, usually infinite in principle if one closely adheres to the definitions of the processes. It is the case for the fractional Gaussian noise obtained by convolution (over an infinite time horizon in the past) of a white noise with an hyperbolic kernel,[11] the case also with the realization of fractional ARIMA processes with linear filters of infinite order.[13] Also, most of the time the synthesized data do not come through a simple (low-order) recurrence allowing "on-line" generation, in contrast to the scheme we are about to propose.

In another area of signals and systems science, (max, +) dynamic systems have been introduced, mainly in a deterministic context, for modeling discrete events and their synchronization, as they occur for instance in manufacturing lines or transportation networks.[14–16] We propose here a novel application of (max, +) systems, in a stochastic context, with a simple instance that we show capable of generating, through a parsimonious first-order recurrence, statistical long-range dependence and usable as a basis for traffic modeling.

2. THE MODEL AND ITS PROPERTIES

Consider the system defined by:

$$U(k) = U(k-1) + u(k)\,, \tag{1}$$

$$Y(k) = \max[Y(k-1),\ U(k)]\,, \tag{2}$$

$$y(k) = Y(k) - Y(k-1)\,, \tag{3}$$

for integers $k > 0$, with the initial condition $U(0) = Y(0) = 0$. For all $k > 0$, the quantities $u(k)$ forming the input sequence are independent and identically distributed random variables with zero mean. We are interested in the statistical properties of the output sequence $y(k)$ and shall show that it exhibits long-range dependence.

The stochastic process described by Eqs. (1)–(3) has a very concise writting under the form of a simple first-order recurrence allowing direct numerical implementation. It makes use of the max and + operators which form the basis of the so-called (max, +) dynamic systems.[15–16] This process

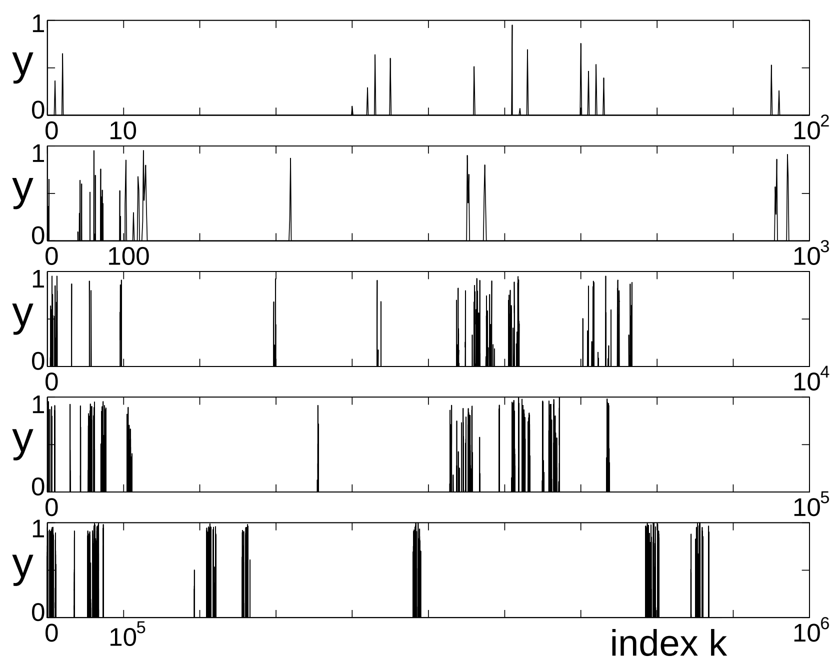

Fig. 1 A typical evolution of $y(k)$ from Eqs. (1)–(3) when $u(k)$ is uniform over $[-1, 1]$, displayed over intervals of increasing lengths and revealing a self-similar structure.

can also be interpreted in the context of random walks:[17] it represents the stochastic process formed by the successive increments of the running maximum of a random walk with arbitrary increment distribution. Yet, such an unconventional process is usually not considered as such in classical texts on random walks, especially for the property of long-range dependence that we shall now address.

A typical evolution of the sequence $y(k)$ is shown in Fig. 1. It is formed by bursts where $y(k) > 0$, separated by intervals where $y(k) = 0$. The values of the increment $y(k)$ at zero happen each time the sequence $U(k)$ lags behind $Y(k)$. When $U(k)$ catches up on $Y(k)$, the increment $y(k)$ gets above zero. Every time $U(k)$ catches up on $Y(k)$, it is just as if the system was restored in its initial condition. The resulting evolution of $y(k)$ displays self-similarity, with bursts where $y(k) > 0$ separated by intervals where $y(k) = 0$ occurring with a similar appearance at any time scale, as visible in Fig. 1.

To characterize this self-similarity, we consider the aggregated process $y^{(m)}(\ell)$ formed by averaging $y(k)$ over non-overlapping successive bins of a fixed length m, i.e. $y^{(m)}(\ell) = m^{-1} \sum_{k=\ell m+1}^{\ell m+m} y(k)$. From a very long segment of data, we have estimated the empirical variance of the aggregated process according to $\mathrm{var}[y^{(m)}] = N^{-1} \sum_{\ell=1}^{N} [y^{(m)}(\ell)]^2 - [N^{-1} \sum_{\ell=1}^{N} y^{(m)}(\ell)]^2$ and for different values of the aggregation level m. The variation shown by the log-log plot of Fig. 2 is quite consistent with a power law of the form $\mathrm{var}[y^{(m)}] \sim m^{-\beta}$ with $\beta \approx 0.5$. This slow decay of $\mathrm{var}[y^{(m)}]$, with an exponent $\beta < 1$, is characteristic of a process with long-range dependence.[7–8]

We have also performed the empirical estimation of the autocorrelation function of the sequence $y(k)$ according to $R_{yy}(\ell) = N^{-1} \sum_{k=1}^{N} y(k)y(k+\ell)$. Figure 3 shows a typical evolution of $R_{yy}(\ell)$ which is quite consistent with a power law of the form $R_{yy}(\ell) \sim \ell^{-\beta}$ with $\beta \approx 0.5$. This slow power law decay of the autocorrelation function is again characteristic of a process with long-range dependence.[7,8]

Fourier transforming the autocorrelation function $R_{yy}(\ell)$ gives access to an estimation of a frequency spectrum $P_{yy}(f)$ for the sequence $y(k)$. For such an empirically evaluated Fourier pair, the

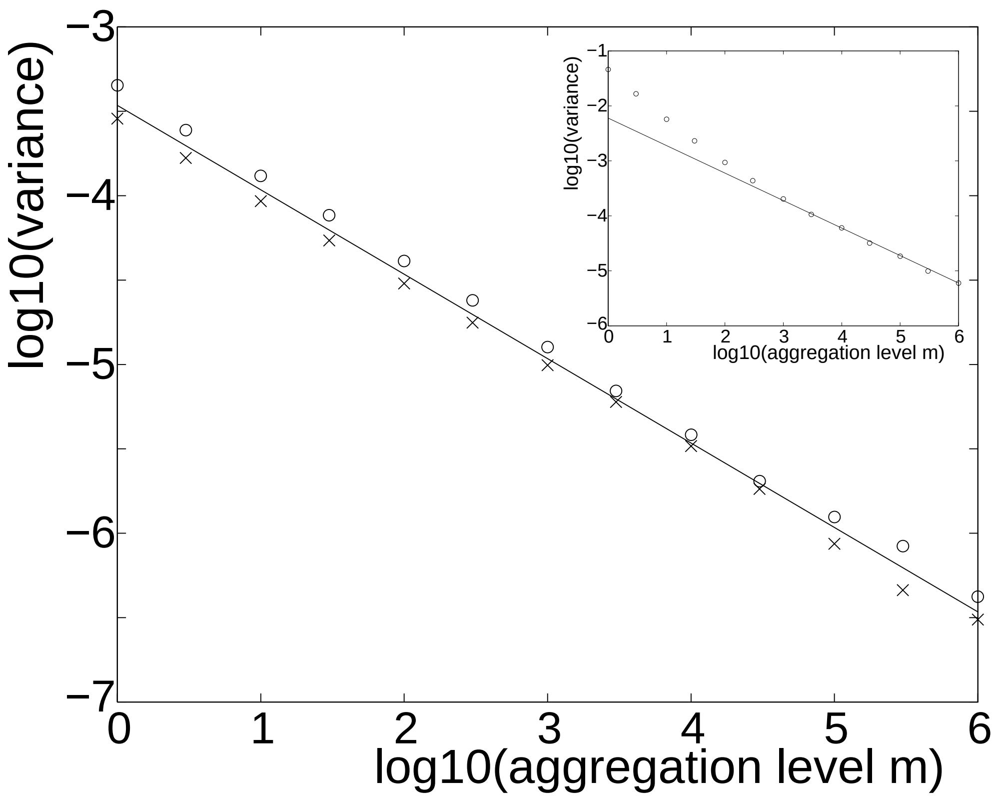

Fig. 2 Variance $\mathrm{var}[y^{(m)}]$ of the aggregated process $y^{(m)}(\ell)$ versus m, from Eqs. (1)–(3) when $u(k)$ is: ($\times$) uniform over $[-1, 1]$ and ($\circ$) zero-mean unit-variance Gaussian. The inset concerns $s(k)$ defined in Fig. 5. In both graphs, the solid line has the slope -0.5.

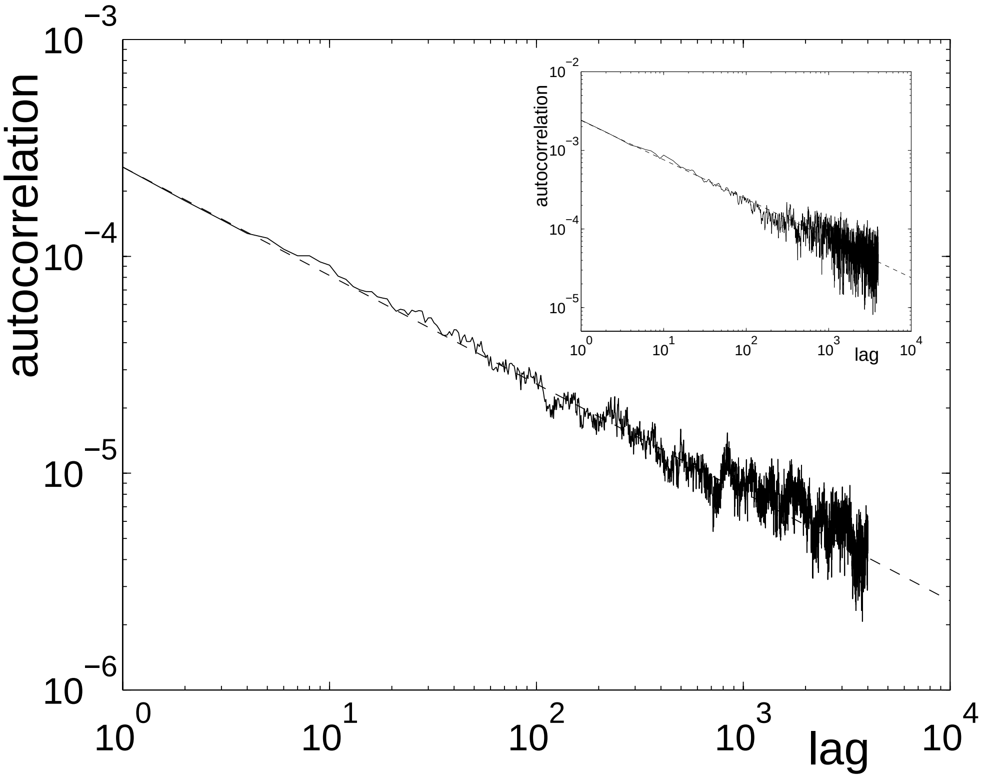

Fig. 3 Autocorrelation $R_{yy}(\ell)$ versus lag ℓ, from Eqs. (1)–(3) when $u(k)$ is zero-mean unit-variance Gaussian. The inset concerns $s(k)$ defined in Fig. 5. In both graphs, the dashed line has the slope -0.5.

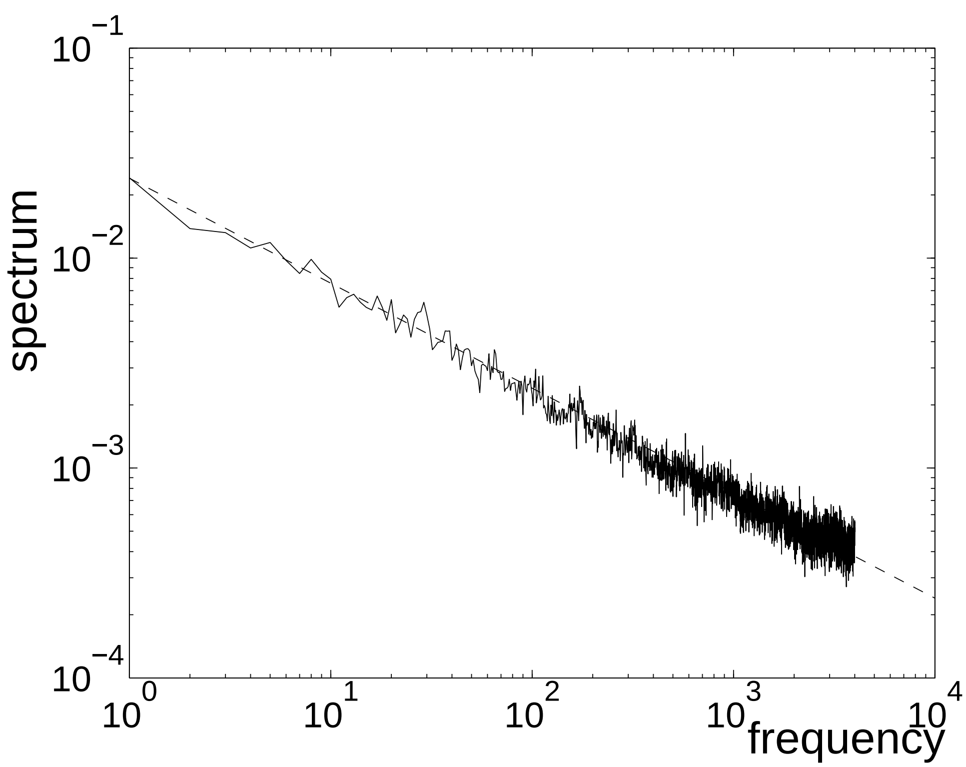

Fig. 4 Spectrum $P_{yy}(f)$ from a Fourier transform of the autocorrelation of Fig. 3. The dashed line has the slope -0.5.

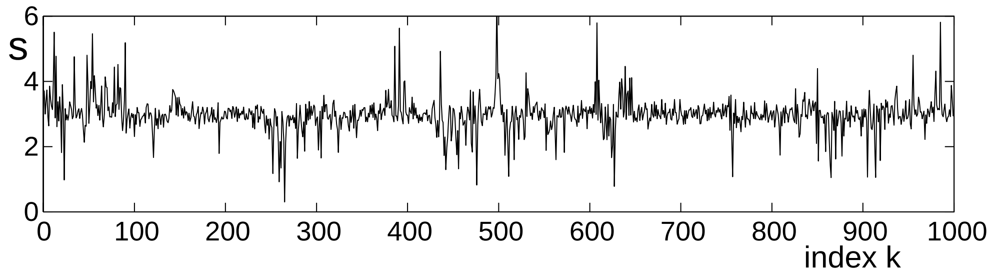

Fig. 5 An evolution of the process $s(k)$ (see text) with the $y_i(k)$ driven in Eqs. (1)–(3) by zero-mean unit-variance Gaussian independent $u_i(k)$, $s_0 = 3$ and $\eta(k)$ a zero-mean Gaussian white noise with standard deviation 0.2.

theory predicts[11] that with a statistically self-similar signal, the autocorrelation function will follow a power law with exponent β, and then the associated spectrum will follow a power law with exponent $\alpha = 1 - \beta$. We have verified that our process $y(k)$ realizes this property, as exemplified by Fig. 4 which shows an evaluation of the spectrum quite consistent with a power law of the form $P_{yy}(f) \sim f^{-\alpha}$ with $\alpha = 1 - \beta \approx 0.5$. This power law behavior identifies the stochastic process $y(k)$ in the class of $1/f^{\alpha}$ noises.[8,11,18–19]

We have also verified that the autocorrelation function that can be estimated in the same way for the aggregated process $y^{(m)}(\ell)$ and the corresponding spectrum, exhibit the same power law dependences with identical values for β and α.

The power laws empirically observed in Figs. 2–4 allow one to assign to the process $y(k)$, a self-similarity property which can be quantitatively identified by the self-similarity parameter $H = 1 - \beta/2 \approx 0.75$.[11] It is this type of procedure that is adopted, for instance in Refs. 7, 8 and 10 to characterize empirical data traffics as self-similar and to assign to them a self-similarity parameter H. For our process of Eqs. (1)–(3), we have observed that the long-range dependence property and its characteristic values $\beta = 0.5$, $\alpha = 0.5$ and $H = 0.75$, remained unchanged when tested with various distributions of the zero-mean input process $u(k)$ (Gaussian, uniform, exponential, discrete distributions were tested). This robust behavior seems to be a universal property of the long-range statistics of the (max, +) system of Eqs. (1)–(3), unaffected by short-range features conveyed by the distribution of $u(k)$.

Similar long-range dependence with comparable values for the self-similarity parameter has recently been observed in different types of data traffic.[7–8,10] We propose the process $y(k)$ of Eqs. (1)–(3) as a possible basis for modeling such long-range dependent traffics, or other long-range dependent processes with a self-similarity parameter $H \approx 0.75$. The process $y(k)$ alone has a peculiar structure with bursts of positive activity separated by intervals at zero. This offers a possibility for representing burstiness in traffics. More realism can be introduced in the synthetic traffic traces by using several independent sequences $y_i(k)$ as building blocks, in order to induce more local variability in the traces while preserving long-range dependence. For instance, we form the sequence $s(k) = s_0 + \sum_{i=1}^{3} y_i(k) - \sum_{i=4}^{6} y_i(k) + \eta(k)$, where the $y_i(k)$ are driven in Eqs. (1)–(3) by independently and identically distributed inputs $u_i(k)$. $\eta(k)$ is a zero-mean white noise (or a noise with short-range correlation) whose role is to describe fluctuations with short-range correlation in the traffic, as observed for instance in Refs. 7, 9 and 10. $s(k)$ will fluctuate around the constant s_0 and because of the statistical independence of the components forming $s(k)$, these fluctuations will display long-range dependence with the same power law behaviors as the individual $y_i(k)$. An illustration is given in Fig. 5 and in the insets of Figs. 2 and 3.

3. DISCUSSION

The present results show that the process of Eqs. (1)–(3) is able to generate synthetic traces that, when tested the way empirical data traffics are, exhibit comparable long-range dependence properties. This forms the basis for our proposing the process of Eqs. (1)–(3) for the modeling of traffic with long-range dependence. The type of modeling we have in mind here is essentially "black-box" modeling. We are not trying to represent, with Eqs. (1)–(3), actual internal mechanisms taking place in traffic systems. Rather, we propose a formal model that is able to reproduce some important statistical properties (long-range dependence) that are observed in the "external" behavior measured on various traffic processes. A large amount of measured data have recently become available which point to long-range dependence properties in traffic. Their modeling have important implications for modern engineering systems, yet no simple models are now available to reproduce these long-range dependence properties. Our formal model offers a possibility to this aim.

It is possible to go a little further toward a "physical" grounding of our model. As evoked in Sec. 1, (max, +) dynamic systems are used in some areas of systems science to model the evolution of delayed or synchronized processes as they occur for instance in manufacturing lines or transportation networks. They manage to represent in a natural way, signals or events that are coming in asynchronously into a system possessing internal processing times regulating the signals or events that are put out.[14–16] These elementary ingredients, when organized in a sufficiently elaborated way, may be at the root of some specific mechanisms taking place in networks of traffic and leading to statistical properties such as long-range dependence. It is thus an existing possibility, that Eqs. (1)–(3) could appear as building blocks in models describing actual internal mechanisms in networks of traffic. But such models referring to actual networks would require careful elaboration; they would be more specific to a given type of traffic; and also their inherent complexity could somehow obscure the simple origin of the long-range dependence property, that we show clearly present as soon as the elementary process described by Eqs. (1)–(3).

We thus prefer in the present study to keep a "black-box" status to our model of Eqs. (1)–(3). Elaborations on Eqs. (1)–(3) to show if they can be meaningfully incorporated in realistic models of specific networks of traffic may constitute possible developments. Here, we concentrate on showing that intrinsically, Eqs. (1)–(3) can generate long-range dependence and on establishing them as a very parsimonious "black-box" model for long-range dependent traffics, which we think interesting and useful in its own right.

A process like $y(k)$, with a $1/f^\alpha$ spectrum, is usually non-stationary.[11] Empirical traffic data are measured and analyzed in the absence of any proof that they are stationary and given the complexity uncovered in actual traffic, it is probable that it is not stationary. Also, from a theoretical point of view, a strict self-similar process cannot be stationary.[20] In practice, empirical estimators are employed which allow a characterization, in measured data, of self-similarity and long-range dependence properties[7–10]. It is this type of empirical estimators that we implement here to characterize the long-range dependence in our process. The empirical autocorrelation function and spectrum used here are of this type. For non-stationary processes, Ref. 21 explains how such empirical measures can be interpreted as averages of time-dependent quantities. For instance, a time-dependent spectrum which reduces to the conventional power spectral density if the signal happens to be stationary.

The main purpose for establishing Eqs. (1)–(3) as a possible model for long-range dependent traffic is related to the observation reported, that when our synthetic traces are tested the way actual traces are, with the same empirical estimators, they exhibit comparable long-range dependence properties.

Finally, we note that our scheme of Eqs. (1)–(3) does not provide control on the parameter H characterizing the long-range dependence. Nevertheless, the invariant value $H = 0.75$ is close to values of H currently observed for actual data traffics, and it can serve as a useful approximation, given the simplicity of Eqs. (1)–(3) that allow on-line generation. In fact, Eqs. (1)–(3) involving a simple first-order recurrence on a scalar variable with a straightforward on-line implementation, constitute to our knowledge the simplest scheme that has ever been proposed for the generation of a non-trivial $1/f^\alpha$ noise. Our scheme realizes $\alpha = 0.5$ as a universal property, much like the white noise realizes $\alpha = 0$ and the Brownian motion $\alpha = 2$. Further, various extensions could be investigated to gain additional control on the process, such as the consideration of higher-order (max, +) systems, or the introduction of correlation in the input sequence $u(k)$, all of which are currently under study.

REFERENCES

1. M. Paczuski and K. Nagel, "Self-Organized Criticality and $1/f$ Noise in Traffic," in *Traffic and Granular Flow*, eds. D. E. Wolf, M. Schreckenberg and A. Bachem (World Scientific, Singapore, 1996), pp. 73–85.
2. A. Schenkel, J. Zhang and Y. C. Zhang, "Long Range Correlation in Human Writings," *Fractals* **1**, 47–57 (1993).

3. M. Amit, Y. Shmerler, E. Eisenberg, M. Abraham and N. Shnerb, "Language and Codification Dependence of Long-Range Correlations in Texts," *Fractals* **2**, 7–13 (1994).
4. M. M. Novak, "Correlations in Computer Programs," *Fractals* **6**, 131–138 (1998).
5. G. F. Zebende, P. M. C. de Oliveira and T. J. P. Penna, "Long-Range Correlations in Computer Diskettes," *Physical Review E* **57**, 3311–3314 (1998).
6. I. Csabai, "$1/f$ Noise in Computer Network Traffic," *Journal of Physics A* **27**, L417–L421 (1994).
7. W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Transactions on Networking* **2**, 1–15 (1994).
8. W. Willinger, M. S. Taqqu, W. E. Leland and D. V. Wilson, "Self-Similarity in High-Speed Packet Traffic: Analysis and Modeling of Ethernet Traffic Measurements," *Statistical Science* **10**, 67–85 (1995).
9. V. Paxson and S. Floyd, "Wide Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking* **3**, 226–244 (1995).
10. J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Transactions on Communications* **43**, 1566–1579 (1995).
11. G. Wornell, *Signal Processing with Fractals* (Prentice Hall, New York, 1996).
12. J. Beran, "Statistical Methods for Data with Long-Range Dependence," *Statistical Science* **7**, 404–427 (1992).
13. C. W. J. Granger and R. Joyeux, "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis* **1**, 15–29 (1980).
14. F. L. Baccelli, G. Cohen, G. J. Olsder and J. P. Quadrat, *Synchronization and Linearity: An Algebra for Discrete Event Systems* (Wiley, New York, 1992).
15. C. G. Cassandras, S. Lafortune and G. J. Olsder, "The Max-Plus Algebra Approach to Discrete Event Systems," in *Trends in Control*, ed. A. Isidori (Springer, Berlin, 1995), pp. 252–267.
16. S. Gaubert, "Performance Evaluation of (max, +) Automata," *IEEE Transactions on Automatic Control* **40**, 2014–2025 (1995).
17. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971).
18. B. Mandelbrot, "Some Noises with $1/f$ Spectrum, a Bridge Between Direct Current and White Noise," *IEEE Transactions on Information Theory* **13**, 289–298 (1967).
19. S. B. Lowen and M. C. Teich, "Generalised $1/f$ Shot Noise," *Electronics Letters* **25**, 1072–1074 (1989).
20. M. S. Taqqu, V. Teverovsky and W. Willinger, "Is Network Traffic Self-Similar or Multifractal?," *Fractals* **5**, 63–73 (1997).
21. P. Flandrin, "On the Spectrum of Fractional Brownian Motions," *IEEE Transactions on Information Theory* **35**, 197–199 (1989).